

CERIDAP

RIVISTA INTERDISCIPLINARE SUL
DIRITTO DELLE
AMMINISTRAZIONI PUBBLICHE

Fascicolo Speciale n. 6 (2026)

Swan song of the anthropocentric
administration in Europe

DIREZIONE: Prof.ssa Diana-Urania Galetta

Iscritta con delibera 16.06.2020 nell'Elenco Speciale dei direttori responsabili di periodici o riviste a carattere scientifico annesso all'Albo dei Giornalisti della Lombardia

Inserita nell'elenco delle Riviste scientifiche di classe A con Delibera ANVUR n. 147/2023

Rivista indicizzata nella [Directory of Open Access Journals](#) e in [SCOPUS](#).

REFERENTE DELLA DIREZIONE E RESPONSABILE TECNOLOGICO: Prof. Gherardo Carullo

COMITATO DI DIREZIONE: Prof.ssa Elisa Arcioni, Prof.ssa Monica Bonini, Prof.ssa Barbara Boschetti, Prof.ssa Elena Buoso, Prof. Gherardo Carullo, Prof.ssa Elena D'Orlando, Prof.ssa Cristina Fraenkel-Haeberle, Prof. Grégory Kalfèlche, Prof.ssa Roberta Lombardi, Prof. Gianpaolo Maria Ruotolo, Prof.ssa Renata Spagnuolo Vigorita

RESPONSABILE DEL COMITATO DI REDAZIONE: Prof.ssa Clara Napolitano

COMITATO DI REDAZIONE: Stefano D'Ancona, Massimo Giavazzi, Micaela Lottini, Alessia Monica, Paolo Provenzano

RESPONSABILI DELLA SEGRETERIA DI REDAZIONE: Pierandrea Corleto, Leonardo Scuto

SEGRETERIA DI REDAZIONE: Filippo Borriello, Emanuela Furiosi, Angelo Maria Gallo, Benedetto Marsili, Giulia Re Ferrè, Pier Marco Rosa Salva, Stefano Rossa, Margherita Sias, Corso Tozzi Martelli, Davide Vaira

COMITATO SCIENTIFICO:

- Prof. Margaret Allars (University of Sydney, Australia)
- Prof. Ricardo Alonso (Universidad Complutense de Madrid, España)
- Prof. Jean-Bernard Auby (Sciences-Po, Paris, France)
- Prof. Patrick Birkinshaw (University of Hull, UK) †
- Prof. David Capitant (Université Paris 1 Panthéon-Sorbonne, France)

- Prof. Mario Pilade Chiti (Università degli Studi Firenze, Italia)
- Prof. Stefano Cognetti (Università degli Studi di Macerata, Italia)
- Prof. Lorenzo Cotino Hueso (Universidad de Valencia, España)
- Prof. Guido Corso (Università degli Studi di Roma III, Italia)
- Prof. Paul Craig (University of Oxford, UK)
- Prof. Mary Crock (University of Sydney, Australia)
- Prof. José Luís da Cruz Vilaça (Universidade de Lisboa, Portugal e Corte di Giustizia dell'EU, Lussemburgo)
- Prof. Pedro Cruz Villalón (Universidad Autónoma de Madrid, España)
- Prof.ssa Daria de Pretis (Università degli Studi di Trento e Corte costituzionale, Italia)
- Prof. Rosario Ferrara (Università degli Studi di Torino, Italia)
- Prof. Guido Greco (Università degli Studi di Milano, Italia)
- Prof. Constance Grewe (Université de Strasbourg, France)
- Prof. Ian Harden (University of Sheffield, UK)
- Prof. Carol Harlow (London School of Economics)
- Prof. Herwig Hofmann (Université du Luxembourg)
- Prof. Martin Ibler (Universität Konstanz)
- Prof.ssa Julia Iliopoulos-Strangas (Università Nazionale e Kapodistrian di Atene, Grecia)
- Prof.ssa Maria Immordino (Università degli Studi di Palermo, Italia)
- Prof. Vittorio Italia (Università degli Studi di Milano, Italia) †
- Prof. Jean-Paul Jacqué (Université de Strasbourg, France et Service juridique du Conseil de l'UE, Bruxelles)
- Prof. Irena Lipowicz (Università Cardinale Stefan Wyszyński di Varsavia, Polonia)
- Prof. Mario G. Losano (Università degli Studi di Milano e del Piemonte orientale, Italia)
- Prof. Roberto Mastroianni (Università degli Studi di Napoli, Italia e Tribunale dell'UE, Lussemburgo)
- Prof. Jacqueline Morand-Deville (Université Paris 1 Panthéon-Sorbonne, France)
- Prof. Bruno Nascimbene (Università degli Studi di Milano, Italia)
- Prof. Luke Nottage (University of Sydney, Australia)

- Prof. Angelika Nussberger (Universität Köln, Deutschland e Corte Europea dei diritti dell'uomo, Strasbourg, France)
- Prof. Giorgio Orsoni (Università Ca' Foscari Venezia, Italia)
- Pres. Alessandro Pajno (Presidente emerito del Consiglio di Stato)
- Prof. Vasco Pereira da Silva (Universidade de Lisboa, Portugal)
- Prof. Ricardo Perlingeiro Mendez da Silva (Universidade Federal Fluminense e Tribunal Regional Federal de Rio de Janeiro, Brasil)
- Prof. Gloria Regonini (Università degli Studi di Milano, Italia)
- Prof. Wojciech Sadurski (University of Sydney, Australia)
- Prof. Daniel Sarmiento (Universidad Complutense de Madrid, España)
- Prof. Jens-Peter Schneider (Universität Freiburg, Deutschland)
- Prof.ssa Maria Alessandra Sandulli (Università degli Studi di Roma Tre, Italia)
- Prof. Eberhard Schmitt-Aßmann (Universität Heidelberg, Deutschland)
- Prof. Jürgen Schwarze (Universität Freiburg, Deutschland) †
- Prof. Karl-Peter Sommermann (Universität Speyer, Deutschland)
- Prof. Klaus Stern (Universität Köln, Deutschland) †
- Prof. Albrecht Weber (Universität Osnabrück, Deutschland)
- Prof. Jacques Ziller (Università degli Studi di Pavia, Italia; Université Paris 1 Panthéon-Sorbonne, France; Istituto Universitario Europeo di Fiesole-IUE)

CERIDAP è una rivista ad accesso libero sul sito ceridap.eu, aggiornata on-line quotidianamente e fascicolata trimestralmente. Alla fine di ogni trimestre gli articoli riprodotti nel sito sono tutti pubblicati nel fascicolo del trimestre di riferimento.

CERIDAP è la Rivista nata nel 2020 dall'iniziativa del Direttore dell'omonimo Centro di Ricerca Interdisciplinare sul Diritto delle Amministrazioni Pubbliche dell'Università degli Studi di Milano per un'azione amministrativa più efficiente e più imparziale, anche in attuazione del principio digital first. La Rivista CERIDAP, pur indipendente dall'omonimo Centro, è la sede principale in cui vengono pubblicate le ricerche svolte in seno al Centro stesso e si è avvalsa sin da principio della collaborazione, all'interno del suo Comitato di Direzione, di un gruppo di studiosi dell'area giuspubblicistica provenienti da vari Atenei, sia italiani che stranieri, che ne condividono l'impostazione di fondo e la

filosofia ispiratrice.

Obiettivo della Rivista CERIDAP è di fornire un contributo di analisi, approfondimento e proposta avente come focus principale le Pubbliche Amministrazioni: intese sia nella loro dimensione istituzionale, organizzativa e di attività, che nella loro dimensione procedurale. E con una naturale apertura prospettica anche al diritto comparato, europeo ed internazionale nonché alla discussione interdisciplinare che è favorita dalla presenza, nel suo Comitato scientifico, di autorevoli studiosi provenienti da prestigiosi Atenei di tutto il Mondo.

Tutte le collaborazioni organizzative ed editoriali sono a titolo gratuito, e agli autori non sono imposti costi di elaborazione e pubblicazione. La Rivista si uniforma agli standard internazionali definiti dal Committee on Publication Ethics (COPE), e fa proprie le relative linee guida.

CERIDAP è interessata a pubblicare contributi relativi a tutti gli ambiti scientifico-disciplinari che sono riferibili al c.d. Diritto Pubblico, inteso nella sua accezione più ampia e che ricomprende anche aree di analisi al confine fra la scienza giuridica e le scienze c.d. politologiche. CERIDAP è inoltre aperta all'approccio tipico dell'analisi economica del diritto. Per questa ragione si è dotata di un "Comitato dei Revisori" che include anche esperti provenienti da aree di expertise diverse da quella del Diritto Pubblico tradizionale (costituzionale ed amministrativo).

I materiali pubblicati su CERIDAP sono oggetto di licenza CC BY-NC-ND 4.00 International. Il lettore può riprodurli e condividerli, in tutto o in parte, con ogni mezzo di comunicazione e segnalazione anche tramite collegamento ipertestuale, con qualsiasi mezzo, supporto e formato, per qualsiasi scopo lecito e non commerciale, conservando l'indicazione del nome dell'autore, del titolo del contributo, della fonte, del logo e del formato grafico originale (salve le modifiche tecnicamente indispensabili). Il testo completo della licenza è consultabile su <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

Per la citazione dei contributi presenti in questo fascicolo di CERIDAP, si consiglia di utilizzare la forma di seguito esemplificata: N. Cognome, *Titolo del contributo*, in *CERIDAP*, speciale n. 6/2026, 2026, p. 35 ss.

Per l'invio di contributi destinati alla pubblicazione: ceridap.eu/rivista/invio-contributi

INDICE SOMMARIO

ARTICOLI

ADRIEN ALBERINI, YANNICK CABALLERO CUEVAS, <i>Supervision of Financial Markets Based on Artificial Intelligence</i>	1
MARC-OLIVIER BUSSLINGER, <i>Trusting the use of AI by the authorities, the Swiss terms and conditions at the forefront</i>	25
OSCAR CAPDEFERRO VILLAGRASA, <i>Automated Administrative Sanctioning and the Rule of Law</i>	58
PAVLÍNA HUBKOVÁ, <i>Technology-driven harmonization? Automation in administrative decision-making at the national level and the role of the EU Charter</i>	91
LUKÁŠ JANČÁT, DIANA REPIŠČÁKOVÁ, <i>Automation of Administrative Proceedings in Slovakia: Current State and Future Perspectives</i>	117
JULIO PONCE SOLÉ, <i>Artificial Intelligence and Interaction Between Machines and Humans: Discretionary Powers, Human Reserve in the Public Sector and the EU Regulation of 2024</i>	140
SYLVA ŘEZNÍKOVÁ, <i>Regulating AI - borders of freedom and sovereignty</i>	165
VLADIMÍR SHARP, <i>Artificial Intelligence and discretionary powers in public administration</i>	180
KRISTÝNA ŠOPKOVÁ, <i>Legal Framework for Artificial Intelligence in Outer Space</i>	208

RASSEGNE E COMMENTI

KAMILA BALOUNOVÁ, <i>From Human to Robot Interpreters: Legal Reflections on the Draft Amendment to the Foreign Nationals Residence Act</i>	230
NATALIYA ISAYEVA, LILIIA SERHIICHUK, <i>AI as a tool for ensuring the</i>	

CERIDAP

<i>housing rights of internally displaced persons in Ukraine</i>	247
RADKA SEMANCOVÁ, <i>AI Administration in Construction: Vienna's</i> <i>Innovative Solution</i>	267

HANNO COLLABORATO A QUESTO NUMERO:

ADRIEN ALBERINI, Professore Associato di Diritto Pubblico dell'economia nell'Università di Ginevra. Avvocato e Of Counsel presso Walder Wyss Ltd, Ginevra, Svizzera.

KAMILA BALOUNOVÁ, Dottoressa di ricerca in Diritto Amministrativo e Ricercatrice nell'Università "Karlova" di Praga.

MARC-OLIVIER BUSSLINGER, Dottorando di Ricerca in Diritto Pubblico nell'Università di Ginevra, Svizzera.

YANNICK CABALLERO CUEVAS, Associate presso Walder Wyss Ltd, Losanna, Svizzera.

OSCAR CAPDEFERRO VILLAGRASA , Professore Associato di Diritto Amministrativo nell'Università di Barcellona.

PAVLÍNA HUBKOVÁ, Ricercatore nell'Università di Exeter, Regno Unito.

NATALIYA ISAYEVA, Dottoranda di ricerca nell'Università Karlova di Praga, Repubblica Ceca.

LUKÁŠ JANČÁT, Ricercatore Senior nell'Università di Košice, Slovacchia.

JULIO PONCE SOLÉ, Professore Ordinario di Diritto Amministrativo nell'Università di Barcellona, Spagna.

DIANA REPIŠČÁKOVÁ, Ricercatrice Senior nell'Università di Košice, Slovacchia.

RADKA SEMANCOVÁ, Dottoranda di ricerca nell'Università di Trnava, Slovacchia.

LILIA SERHIICHUK, Dottoranda di Ricerca in Diritto Pubblico nell'Università Karlova di Praga, Repubblica Ceca.

VLADIMÍR SHARP, Assistant Professor nell'Università Karlova di Praga, Repubblica Ceca.

SYLVA ŘEZŇÍKOVÁ, Dottoressa di ricerca in Law and Linguistics nell'Università di Lancaster.

KRISTÝNA ŠOPKOVÁ, Dottoranda di ricerca nell'Università Karlova di Praga, Repubblica Ceca.

Introduction

The contributions published in the Special Issue n.6/2026 of CERIDAP are revised and expanded versions of papers presented at the *Swan song of the anthropocentric administration in Europe* round table, which took place at the Law Faculty of the Charles University in Prague, Czech Republic, on 27 and 28 November 2026. This round table was organised jointly by the Department of Administrative Law at the Law Faculty of the Charles University and the Interdisciplinary Research Centre for Public Administration (CERIDAP) at the Law Faculty of the University of Milan.

The progressive transformation of public administration in the age of artificial intelligence has been examined from a variety of legal perspectives. One widely recognised feature of this evolution is the emergence of public administration as a virtual organisation, increasingly relying on digital technologies and artificial intelligence to achieve its objectives and perform its functions, rather than on the physical presence of civil servants and their interaction with citizens.

Against this background, the Articles collected in this Special Issue explore the legal implications of AI for public administration from complementary perspectives. Several *Articles* focus on the conditions for building trust in the use of AI by public authorities (See, M.-O. Busslinger, *Trusting the Use of AI by the Authorities, the Swiss Terms and Conditions at the Forefront*), the relationship between automated administrative sanctioning and the rule of law (See, O. Capdeferro Villagrasa, *Automated Administrative Sanctioning and the Rule of Law*), the digitalisation and automation of administrative decision-making at both national and EU level (See, P. Hubáč, *Technology-driven Harmonization? Automation in Administrative Decision-making at the National Level and the Role of the EU Charter*; L. Jančát and D. Repiščáková, *Automation of Administrative Proceedings in Slovakia: Current State and Future Perspectives*), the use of artificial intelligence in the supervision of financial markets (See, A. Alberini and Y. Caballero Cuevas, *Supervision of Financial Markets Based on Artificial Intelligence*), the interaction between machines and human decision-makers in the exercise of public powers (See, J. Ponce Solé, *Artificial Intelligence and Interaction Between Machines and Humans: Disciplinary Powers, Human Review in the Public Sector and the EU Regulation of 2024*), regulatory approaches to AI beyond national borders (See, S. Řezníková, *Regulating AI – Borders of Freedom and Sovereignty*), and the emerging legal framework governing artificial intelligence in outer space (See, K. Šopková, *Legal Framework for Artificial Intelligence in Outer Space*).

The *Reviews and Comments* complement these analyses by addressing the use of robot interpreters in migration procedures (See, K. Balounová, *From Human to Robot Interpreters: Legal Reflections on the Draft Amendment to the Foreign Nationals Residence Act*), the role of artificial intelligence in ensuring the housing rights of internally displaced persons in Ukraine (See, N. Isayeva and L. Serhiichuk, *AI as a Tool for Ensuring the Housing Rights of Internally Displaced Persons in Ukraine*), and the application of AI in construction administration (See, R. Semancová, *AI Administration in Construction: Vienna's Innovative Solution*). Taken together, these contributions offer a comparative perspective on the legal

challenges posed by the growing integration of artificial intelligence into public administration and regulatory governance, while questioning whether the traditional anthropocentric paradigm of European administrative law is still capable of responding to the profound transformations brought about by digital technologies.

Introduzione

Gli articoli pubblicati nel numero speciale n. 6/2026 del CERIDAP sono versioni riviste e ampliate dei contributi presentati alla tavola rotonda “*Il canto del cigno dell’amministrazione antropocentrica in Europa*”, tenutasi presso la Facoltà di Giurisprudenza dell’Università Karlova di Praga, il 27 e 28 novembre 2026. La tavola rotonda è stata organizzata congiuntamente dal Dipartimento di Diritto Amministrativo della Facoltà di Giurisprudenza dell’Università di Praga e dal Centro di Ricerca Interdisciplinare per la Pubblica Amministrazione (CERIDAP) della Facoltà di Giurisprudenza dell’Università degli Studi di Milano.

La progressiva trasformazione della pubblica amministrazione nell’era dell’intelligenza artificiale è stata esaminata da una varietà di prospettive giuridiche. Una caratteristica ampiamente riconosciuta di questa evoluzione è l’emergere della pubblica amministrazione come organizzazione virtuale, che fa sempre più affidamento sulle tecnologie digitali e sull’intelligenza artificiale per raggiungere i propri obiettivi e svolgere le proprie funzioni, piuttosto che sulla presenza fisica e sull’interazione personale tra funzionari e cittadini.

In questo contesto, i contributi raccolti in questo numero speciale esplorano le implicazioni giuridiche dell’IA per la pubblica amministrazione da prospettive complementari. Diversi Articoli si concentrano sulle condizioni per costruire la fiducia nell’uso dell’IA da parte delle autorità pubbliche (cfr. M.-O. Busslinger, *Trusting the Use of AI by the Authorities, the Swiss Terms and Conditions at the Forefront*), il rapporto tra sanzioni amministrative automatizzate e lo Stato di diritto (cfr. O. Capdeferro Villagrasa, *Automated Administrative Sanctioning and the Rule of Law*), la digitalizzazione e l’automazione del processo decisionale amministrativo sia a livello nazionale che europeo (cfr. P. Hubáč, *Technology-driven Harmonization? Automation in Administrative Decision-making at the National Level and the Role of the EU Charter*; L. Jančát e D. Repiščáková, *Automation of Administrative Proceedings in Slovakia: Current State and Future Perspectives*), l’uso dell’intelligenza artificiale nella vigilanza sui mercati finanziari (cfr. A. Alberini e Y. Caballero Cuevas, *Supervision of Financial Markets Based on Artificial Intelligence*), l’interazione tra le macchine e la decisione umana (Cfr. J. Ponce Solé, *Artificial Intelligence and Interaction Between Machines and Humans: Disciplinary Powers, Human Review in the Public Sector and the EU Regulation of 2024*), gli approcci normativi all’IA oltre i confini nazionali (Cfr. S. Řezníková, *Regulating AI – Borders of Freedom and Sovereignty*) e il quadro giuridico emergente che disciplina l’intelligenza artificiale nello spazio (Cfr. K. Šopková, *Legal Framework for Artificial Intelligence in Outer Space*).

La sezione *Recensioni e commenti* integra queste analisi affrontando il tema dell’uso di interpreti robot nelle procedure di immigrazione (Cfr. K. Balounová, *From Human to Robot*

Interpreters: Legal Reflections on the Draft Amendment to the Foreign Nationals Residence Act), il ruolo dell'intelligenza artificiale nel garantire il diritto all'alloggio degli sfollati interni in Ucraina (Cfr. N. Isayeva e L. Serhiichuk, *AI as a Tool for Ensuring the Housing Rights of Internally Displaced Persons in Ukraine*) e l'applicazione dell'IA nella gestione dei cantieri nelle aree urbane (Cfr. R. Semancová, *AI Administration in Construction: Vienna's Innovative Solution*). Nel loro insieme, questi contributi offrono una prospettiva comparata sulle sfide giuridiche poste dalla crescente integrazione dell'intelligenza artificiale nella pubblica amministrazione e nella governance normativa, mettendo al contempo in discussione se il tradizionale paradigma antropocentrico del diritto amministrativo europeo sia ancora in grado di rispondere alle profonde trasformazioni determinate dalle tecnologie digitali.

Milan-Prague/Milano-Praga, July 2026/Luglio 2026
the Editors/I curatori: Jakub Handrlica, Alessia Monica.

Supervision of Financial Markets Based on Artificial Intelligence

Adrien Alberini, Yannick Caballero Cuevas

DOI: 10.13130/2723-9195/2026-3-29

Il presente contributo esamina l'impiego dell'intelligenza artificiale nella vigilanza sui mercati finanziari, concentrandosi sui benefici, sui rischi e sulle garanzie giuridiche connessi all'integrazione di sistemi automatizzati nei procedimenti delle autorità di supervisione. Sebbene l'uso dell'intelligenza artificiale da parte degli operatori finanziari abbia già ricevuto significativa attenzione sul piano regolatorio, il suo utilizzo da parte delle autorità pubbliche di vigilanza rimane ancora relativamente poco esplorato. Lo studio analizza il quadro delineato dall'AI Act europeo, soffermandosi sulla possibile qualificazione dei sistemi impiegati nella supervisione finanziaria come sistemi ad alto rischio e sui conseguenti obblighi in materia di gestione dei rischi, trasparenza, tracciabilità e controllo umano. Il contributo sostiene che la supervisione umana non possa ridursi a un adempimento meramente formale, ma debba consentire alle autorità di comprendere, contestualizzare e, ove necessario, contestare gli esiti probabilistici prodotti dai sistemi di intelligenza artificiale. Attraverso un caso di studio relativo alla manipolazione del mercato nell'ordinamento svizzero, l'analisi evidenzia inoltre come l'obbligo di motivazione possa costituire una garanzia essenziale di intelligibilità, controllo e legittimità delle decisioni adottate con il supporto dell'intelligenza artificiale.

This paper examines the use of artificial intelligence in financial-market supervision, focusing on the benefits, risks, and legal safeguards associated with the integration of automated systems into the activities of supervisory authorities. While the use of artificial intelligence by financial institutions has already attracted significant regulatory attention, its deployment by public supervisory bodies remains comparatively underexplored. The study analyses the framework established by the EU AI Act, with particular regard to the potential classification of AI systems used in financial supervision as high-risk systems and to the resulting obligations concerning risk management, transparency, traceability, and human oversight. It argues that

human oversight cannot be reduced to a mere formal requirement but must enable supervisory authorities to understand, contextualise, and, where necessary, challenge the probabilistic outputs generated by artificial intelligence systems. Drawing on a case study concerning market manipulation under Swiss law, the paper further shows that the duty to state reasons can operate as a central safeguard of the intelligibility, accountability, and legitimacy of decisions adopted with the support of artificial intelligence.

Summary: 1. Introduction.- 2. Integration of AI into Financial Supervision.- 2.1. Functions of AI in Supervisory Practice.- 2.2. Risks and Challenges of AI-Based Supervision.- 3. Financial Supervision Based on AI: Regulatory Framework and Qualifications.- 3.1. Supervisory Authorities Subject to the EU AI Act.- 3.2. Supervisory AI Tools as Potential High-Risk Systems.- 4. Regulatory Requirements Applicable to Financial Supervision Based on AI.- 4.1. No prohibition; Overview of Applicable Requirements.- 4.2. Human Oversight: Justification and Implications.- 4.3. Interpreting the Output: Need to Consider the Probabilistic Results.- 4.4. Decision-Making through AI?- 5. Transparency Requirements Applicable to Financial Supervision Based on AI?- 5.1. No Transparency Based on the EU AI Act.- 6. Case Study: Trade-Based Manipulation and Swiss Financial Market Supervision.- 6.1. Swiss Regulatory Framework.- 6.2. Case 2C_315/2020: Spoofing and Layering Acts.- 6.3. AI-Based Supervision against Manipulative Behaviours.- 7. Conclusion.

1. Introduction

The deployment of artificial intelligence (“AI”) tools by financial institutions has attracted significant scholarly, regulatory, and public attention in recent years, prompting the emergence of the first dedicated regulatory frameworks^[1]. By contrast, the use of similar technologies by financial supervisory authorities has remained comparatively underexplored^[2]. This relative silence is striking: AI systems hold considerable promise for enhancing the oversight of financial markets, and representatives of supervisory bodies have already acknowledged – albeit informally – their growing reliance on such tools. Against this backdrop,

important questions arise regarding the risks associated with the use of AI by supervisory authorities and the appropriate regulatory framing of AI-enabled financial supervision^[3].

Situated within the broader context of *The Swan Song of the Anthropocentric Administration in Europe*, this contribution seeks to illuminate the evolving balance between human and machine in administrative decision-making, particularly in the field of financial supervision. It does so through a legal analysis grounded primarily in European law, complemented by a case study drawn from Swiss law. While anchored in these two legal orders, the reflections developed here resonate more broadly and offer insights relevant to jurisdictions beyond Europe and Switzerland.

The chapter proceeds in six parts. Section 2 examines the functions of AI in financial supervision and identifies the risks inherent in its integration into supervisory processes. Section 3 outlines the European regulatory framework applicable to AI and assesses the extent to which it governs AI-based market supervision. Section 4 turns to the substantive requirements for AI-enabled financial oversight, with particular emphasis on the necessity of maintaining meaningful human control. Section 5 complements this analysis by exploring transparency obligations and the constitutional duty to provide reasons, understood as a core component of the right to a fair trial. Section 6 illustrates these issues through a concrete case study involving market manipulation. The chapter concludes in Section 7 with a synthesis of the main findings and the broader implications for the future of technologically augmented supervision.

2. Integration of AI into Financial Supervision

2.1. Functions of AI in Supervisory Practice

The integration of AI into financial supervision reflects the growing complexity and digitalisation of financial markets. Supervisory authorities operate in an environment characterised by high-frequency trading, algorithmic decision-making, complex financial products and cross-border flows of capital. Traditional analytical tools – often manual, fragmented and resource-intensive – struggle to keep pace with such developments. AI-based systems may offer a

significant enhancement in supervisory capacity by enabling the systematic processing and interpretation of vast quantities of structured and unstructured data^[4].

In our view, AI tools relevant for financial supervision can essentially be divided into two main categories^[5].

First, AI tools can improve data management and information processing, particularly in the context of auditing and reporting obligations. Auditors and supervisory authorities receive extensive reporting documentation from regulated institutions, often in diverse formats and with varying degrees of structure. Natural language processing and automated classification systems can support the organisation, extraction and interpretation of relevant information from these submissions. This facilitates a more coherent and timely supervisory overview and reduces administrative burdens associated with manual data processing^[6].

A second and more sophisticated area in which AI may contribute to supervisory practice is support to decision-making by providing structured insights:

- Typically, this includes the detection of anomalous or fraudulent market behaviours. Machine-learning systems can analyse extensive datasets relating to transaction flows, market movements and trading patterns, thereby identifying indicators of insider trading, market manipulation or other similar illicit behaviours. Unlike traditional rule-based monitoring tools, which rely on predefined scenarios, machine-learning models can detect emerging patterns that were not explicitly coded *ex ante*. This adaptability is particularly valuable in a context where illicit strategies evolve rapidly and often exploit novel market conditions^[7].
- Similarly, AI may contribute to systemic risk analysis and stress-testing. Machine-learning algorithms can simulate complex market dynamics by incorporating heterogeneous datasets, including macroeconomic indicators, liquidity profiles, and behavioural responses of financial institutions. These models can support the development of predictive tools that help identify vulnerabilities within the financial system before they materialise. In doing so, AI-based systems enhance supervision by allowing authorities to assess a broader range of potential risk scenarios

than would be feasible through traditional methods^[8].

2.2. Risks and Challenges of AI-Based Supervision

The deployment of AI in financial market supervision not only offers significant advantages but also introduces a series of risks that may challenge the integrity, reliability and legitimacy of supervisory action. These risks must be carefully analysed to understand the limits of AI-assisted oversight and to identify the legal safeguards necessary to maintain compliance with administrative law standards and fundamental rights.

A first category of concerns relates to operational risks^[9], which arise from the technical and functional characteristics of machine-learning systems. Unlike deterministic rule-based tools, machine-learning models may evolve over time in response to changing data inputs, which can lead to model drift or unexpected variations in output quality. Similarly, the performance of such systems depends heavily on the quality, representativeness and completeness of training data. Biased, incomplete or outdated datasets may produce erroneous risk assessments or distort supervisory priorities. Operational risks also encompass explainability limitations: AI systems may generate outputs that are difficult to interpret, particularly when based on deep learning architectures. This opacity complicates the integration of AI outputs into supervisory processes that require traceability and reasoned decision-making^[10].

Second, the use of AI raises significant bias and fairness concerns. Machine-learning systems identify patterns based on historical data, which may encode systemic biases or reflect pre-existing market structures. If such biases are inadvertently reproduced or amplified, supervisory decisions based on AI outputs may disproportionately affect certain categories of institutions or transactions. This risk is particularly problematic for public authorities, which are bound by principles of equality and non-discrimination. Unchecked algorithmic bias may thus translate into differentiated supervisory burdens, inconsistent risk classifications or unequal treatment of market participants^[11].

A third category, which is related to the preceding ones, concerns ethical and reputational risks. Financial supervisory authorities play a central role in maintaining market confidence and ensuring the proper functioning of the

financial system. If AI-assisted assessments produce inconsistent or discriminatory results, or if they appear arbitrary to supervised institutions, the credibility of supervisory action may be undermined. The mere perception that authorities rely excessively on opaque algorithms can also erode trust, particularly if market participants fear that automated tools may misinterpret legitimate behaviour as suspicious. Such reputational vulnerabilities have direct implications for the legitimacy of supervisory interventions and may expose authorities to criticism or public scrutiny^[12].

A fourth – and somewhat distinct – risk category concerns cybersecurity and data protection^[13]. Supervisory authorities handle highly sensitive financial information, often subject to strict confidentiality requirements. The integration of AI systems introduces new vulnerabilities linked to data access, model security and the integrity of digital infrastructures. Adversarial attacks, data poisoning or model inversion techniques may compromise the reliability of AI outputs or expose confidential supervisory information. Moreover, the use of large datasets and automated analysis tools raises concerns under data protection law, particularly regarding profiling, automated processing and the storage of granular financial data.

3. Financial Supervision Based on AI: Regulatory Framework and Qualifications

3.1. Supervisory Authorities Subject to the EU AI Act

The EU AI Act establishes a regulatory framework applicable to both private and public entities^[14], including supervisory authorities. The EU AI Act applies to public authorities because it is designed as a horizontal regulatory framework that governs all actors who develop, place on the market, or use AI systems within the EU – regardless of whether they are private companies or public bodies. Public authorities frequently deploy AI in areas that directly affect individuals' rights, such as social services, law enforcement, migration, and public administration^[15].

In this connection, one of the critical aspects is the distinction between providers and deployers, which determines the set of obligations applicable to supervisory authorities^[16]. As a rule, financial supervisory authorities qualify as deployers, since

they use AI systems without placing them on the market. However, they may exceptionally be considered providers when they develop bespoke tools or substantially modify an existing system^[17].

3.2. Supervisory AI Tools as Potential High-Risk Systems

The EU AI Act is a comprehensive regulation which applies to all AI systems, defined in essence as adaptable machine-learning systems^[18]. Furthermore, it is a risk-based regulation, making a critical distinction between high-risk AI systems and lower risk AI systems. High-risk AI systems are essentially defined as such when they may have a significant harmful impact on the health, safety and fundamental rights of persons. More specifically, Article 6 EU AI Act, which provides classification rules for high-risk AI systems, refers to annexes to this Act^[19].

In relation to financial supervision, Annex III, based on Article 6(2) EU AI Act, is of particular relevance:

- High-risk AI systems comprise AI systems in the area of administration of justice, namely AI systems intended to be used by a judicial authority^[20] or on their behalf to assist a judicial authority in researching and interpreting facts and the law and in applying the law to a concrete set of facts, or to be used in a similar way in alternative dispute resolution^[21].
- The reason for inclusion is due to the potentially significant impact on democracy, rule of law, individual freedoms as well as the right to an effective remedy and to a fair trial. With respect to the nature of the potential harm, the risks are potential bias, errors and opacity^[22].
- Regarding the exceptions, «*[t]he classification of AI systems as high-risk should not, however, extend to AI systems intended for purely ancillary administrative activities that do not affect the actual administration of justice in individual cases, such as anonymisation or pseudonymisation of judicial decisions, documents or data, communication between personnel, administrative tasks*»^[23].

In light of the above, the qualification of AI systems deployed by financial supervisory authorities under the EU AI Act depends on the specific purpose and

context of their use, rather than on the mere fact that an authority deploys them. The Act does not automatically classify AI tools used by supervisory bodies as high-risk. Instead, the categorisation results from an assessment of the system's intended function, operational role, and potential impact on the legal position of supervised entities.

Some supervisory use cases – for example, systems designed to assess compliance, identify potentially unlawful behaviours, or produce risk scores that may inform enforcement measures – may fall within the category of as high-risk systems. In such situations, AI-generated outputs may contribute to supervisory interventions or form part of the factual basis upon which regulatory actions are taken. This functional connection to decision-making processes increases, in our view, the likelihood that the system will be regarded as high-risk.

Conversely, certain AI tools used by supervisory authorities may fall within the category of limited-risk systems. In our opinion, these include systems employed primarily for exploratory analysis, pattern detection without immediate regulatory consequences, internal workflow optimisation or preliminary data structuring. When AI outputs do not directly influence supervisory decisions, and when the system's role remains confined to assisting internal processes without shaping legal outcomes, the justification for categorising the system as high-risk becomes less evident.

Accordingly, the qualification of an AI system used by supervisory authorities requires a case-by-case assessment. Authorities must examine the system's operational purpose, its degree of influence on supervisory decisions and the potential impact of its outputs on supervised entities. This nuanced analysis reflects the risk-based logic of the EU AI Act and ensures that regulatory obligations are proportionate to the system's actual role within supervisory practice.

4. Regulatory Requirements Applicable to Financial Supervision Based on AI

4.1. No prohibition; Overview of Applicable Requirements

As a first and important note, the use of AI systems by financial supervisory

authorities is not prohibited, as it does fall in the category of prohibited systems under the EU AI Act ^[24] .

High-risk AI systems are subject to detailed requirements under Chapter II of the EU AI Act, it being specified that these obligations generally fall on providers; however, deployers must ensure that the systems they use comply with these requirements, which include the following: risk management systems ensuring continuous identification, mitigation and monitoring of risks ^[25] ; standards on data and data governance, including training data quality, representativeness and relevance ^[26] ; technical documentation requirements ensuring traceability and auditability ^[27] ; record-keeping obligations, including automated logging of system interactions ^[28] ; transparency obligations, ensuring that deployers understand the system's capabilities and limitations ^[29] ; human oversight mechanisms, enabling effective supervision, override and intervention ^[30] ; requirements on accuracy, robustness and cybersecurity ^[31] .

Supervisory authorities, acting as deployers, are subject to additional obligations, which aim to ensure that the deployment of AI tools does not compromise procedural fairness, fundamental rights or public accountability in supervisory processes: Implementation of appropriate technical and organisational measures ^[32] ; Designation of competent individuals to exercise human oversight ^[33] ; Maintenance of log records to ensure traceability ^[34] ; Registration duties for high-risk systems used by public authorities ^[35] ; Mandatory fundamental rights impact assessments ^[36] , expressly required for public-sector deployers.

4.2. Human Oversight: Justification and Implications

Among the above-mentioned requirements, human oversight deserves particular attention in relation to financial supervision based on AI:

- The fact that human oversight is required when high-risk AI systems are used is a manifestation of the human centric approach that the EU takes to regulating AI and demonstrates that there is a lack of trust towards machines. Hence, «*human oversight is an attempt to counterbalance autonomy, which refers to the increasing capacity of the technology to act without human control or supervision*» ^[37] .

- The lack of trust towards the machine does not reach the level of prohibiting the use of AI systems in the context of financial supervision but, as it stems from the concept of human oversight, forces to implement the proper measures to ensure that humans remain in control.

As per the EU AI Act, human oversight means notably that the high-risk AI system shall be provided to the deployer in such a way that natural persons to whom human oversight is assigned are enabled, as appropriate and proportionate:

- (a) to properly understand the relevant capacities and limitations of the high-risk AI system and be able to duly monitor its operation, including in view of detecting and addressing anomalies, dysfunctions and unexpected performance;
- (b) to remain aware of the possible tendency of automatically relying or over-relying on the output produced by a high-risk AI system (automation bias), in particular for high-risk AI systems used to provide information or recommendations for decisions to be taken by natural persons;
- (c) to correctly interpret the high-risk AI system's output, taking into account, for example, the interpretation tools and methods available; and
- (d) to decide, in any particular situation, not to use the high-risk AI system or to otherwise disregard, override or reverse the output of the high-risk AI system^[38].

In this regard, it is worth noting that, prior to deploying high-risk AI systems – and subject to certain exceptions – public-sector deployers are required to conduct an impact assessment that includes a detailed account of the human oversight mechanisms put in place^[39]. This requirement underscores the centrality of meaningful human oversight in the use of high-risk AI systems by public authorities.

4.3. Interpreting the Output: Need to Consider the Probabilistic Results

As indicated above, human oversight means notably that the natural person using the high-risk AI system should be in a position to correctly interpret the high-risk AI system's output, taking into account, for example, the interpretation tools and methods available.

In this context, it is essential to recall that AI-based systems operate on probabilistic inferences rather than deterministic conclusions^[40]. As a result, an AI-generated indication that enforcement action should be prioritized, or that potentially fraudulent market behavior has occurred, may rest on a probability of 51% or, alternatively, 90%, two scenarios that are qualitatively distinct. This raises the question of the extent to which supervisory authorities should assess such probabilistic information.

Several observations can be made in this regard:

- First, it is somewhat surprising that the EU AI Act remains silent on the concept of probabilistic results, despite its emphasis on human oversight^[41].
- Second, with respect for instance to decisions concerning the prioritization of supervisory or enforcement activities, supervisory authorities traditionally enjoy a broad margin of discretion. Given this margin of discretion, there appears to be limited justification for requiring detailed assessment of the probabilistic outputs generated by AI systems in relation to such decisions.
- Third, by contrast, once supervisory authorities move beyond the mere organization or prioritization of their activities and enter the realm of adopting formal enforcement measures in individual cases, the nature of their responsibilities changes fundamentally. At this stage, the epistemic quality of AI-generated findings becomes critical. Assume the use of an AI-based tool to conduct stress tests on a financial institution with significant exposure to interest-rate fluctuations. If the stress tests indicate that the institution is undercapitalized in light of the simulated scenarios, the supervisory authority may require the institution to increase its capital buffers. In such a context, it becomes essential to understand the degree of confidence with which the AI system reaches its conclusions. A supervisory decision based on an output associated with a confidence level of 51% is of a fundamentally different nature from one supported by a confidence level of 90%. The legitimacy and legal defensibility of the resulting enforcement measure depend, at least in part, on the authority's ability to assess and interpret this confidence level appropriately.
- Although AI systems do not, as a rule, provide explicit confidence levels

or explanations for their outputs, supervisory authorities should nevertheless interrogate these systems to understand why certain facts, indicators, or assumptions were weighted or disregarded. Hence, human oversight must not be reduced to a formalistic requirement but should instead entail a substantive capacity to question, challenge, and contextualize AI-generated outputs.

4.4. Decision-Making through AI?

A final question arising in connection with human oversight concerns whether a decision may, in principle, be produced entirely by a machine. Although it is important to recall that the degree of autonomy attributed to AI systems is frequently overstated^[42] – and that such a scenario remains largely theoretical in the current context of financial supervision – the issue warrants careful consideration:

- The wording of Article 6 EU AI Act does not appear to prohibit this possibility. The provision requires only that the resulting decision be subject to human oversight; it does not mandate that humans must themselves generate each component of the decision-making process.
- Recital 61 EU AI Act adopts a seemingly more restrictive tone, stating that «*[t]he use of AI tools can support the decision-making power of judges or judicial independence, but should not replace it: the final decision making must remain a human driven activity*»^[43]. At first sight, this might suggest that the involvement of AI must be limited to an auxiliary role.
- Yet, in our view, the two sources can be reconciled. There is no compelling reason – assuming it is technically feasible – why an authority could not delegate the entire preparatory phase of a decision to an AI system, provided that the authority subsequently exercises diligent and substantive control over the outcome. Such control necessarily presupposes that the system is capable of providing a complete account of its reasoning, not in algorithmic terms, but in a form that can be translated into a legally cognizable methodology.
- Ultimately, the decision is, in any event, formally adopted by the

competent authority. What matters, therefore, is not whether the machine drafts the decision, but whether the human decision-maker is able to understand, verify, and endorse the reasoning on which it rests. This ensures compliance with both the letter and the spirit of the EU AI Act's human-oversight requirements.

5. Transparency Requirements Applicable to Financial Supervision Based on AI?

5.1. No Transparency Based on the EU AI Act

Ultimately, the central issue is ensuring that interventions by supervisory authorities are both substantively justified and intelligible to the financial institutions and other actors subject to their oversight. In other words, supervisory decision-making must remain reasoned, transparent, and amenable to external scrutiny.

These considerations raise the question of whether the use of AI systems – and the manner in which human oversight has been exercised – should be disclosed by supervisory authorities.

No such requirement arises from the provisions governing human oversight, nor from any other provisions of the EU AI Act. More specifically, the transparency obligations set out in Article 13 EU AI Act aim solely to ensure that deployers understand the system's capabilities and limitations^[44]. Likewise, the transparency obligations under Article 50(1) require only that natural persons interacting directly with an AI system be made aware of that fact^[45].

B. Duty to State Reasons as Sufficient Safeguard

More fundamentally, the issue turns on the constitutional guarantee embodied in the duty to state reasons:

- This guarantee requires that the addressee of an administrative decision be able to understand both the decision itself and the reasoning that underpins it^[46]. In and of itself, the guarantee does not require the authority to indicate whether it relied on a computational system to establish certain facts or to assess them from a legal perspective.

- One may nonetheless ask whether an evolution of the duty to state reasons is desirable when supervisory authorities in the financial sector use AI systems, on the grounds that such use may affect the transparency or legitimacy of administrative action. Under this view, the authority would be required to incorporate explanations regarding how AI systems were used, how their outputs were evaluated, and how potential shortcomings or uncertainties were addressed^[47].
- In our view, however, these elements – aside from the mere information that an AI system was used, which is of limited relevance in itself – add nothing beyond what the duty to state reasons already requires. Whether the assessment of a case is performed by a human or by a machine, the decision must contain a pertinent and convincing legal rationale. In reality, the duty to state reasons operates as an additional safeguard that, in practice, reinforces the concept of human oversight: if the authority is unable to understand and evaluate the reasoning produced by the system, it will be unable to reproduce that reasoning in its decision. This would either violate the duty to state reasons or result in a substantively weak decision that is highly susceptible to challenge by the addressee.

6. Case Study: Trade-Based Manipulation and Swiss Financial Market Supervision

6.1. Swiss Regulatory Framework

Art. 143(1)(b) FinMIA prohibits any form of trade-based manipulation involving securities admitted on trading on a trading venue or a DLT trading facility with its registered office in Switzerland^[48]. This provision seeks to safeguard the integrity of the Swiss financial market by sanctioning transactions capable of impairing its proper functioning. It covers transactions that generate, or are likely to generate, false or misleading signals as to market conditions. Practices such as *painting the tape*, *wash trades*, *matched orders*, or *spoofing* are typically behaviours that may mislead investors^[49].

In the application of Art. 143(1)(b), the decisive criterion is the capacity of the transaction to convey a false or misleading signal to the market. The concept of

signal implicitly refers to the reasonable investor test^[50]. This latter, developed by the United States Supreme Court^[51] and subsequently endorsed by Swiss case law^[52], requires an assessment of whether the transaction is likely to influence the investment decision of a reasonable investor^[53]. This figure serves as an objective standard for determining whether the transaction produces a market signal. According to doctrinal approaches, the reasonable investor either refers to a rational and idealized investor or to the market as a whole^[54]. However, it appears that recent Swiss doctrine, as well as decisions of the Federal Criminal Court, seem to adopt the interpretation of the reasonable investor as representing the market as a whole^[55].

The reasonable investor test constitutes an objective and qualitative assessment conducted from an ex ante perspective. It therefore requires consideration of all relevant circumstances of the case in order to determine whether the transaction was likely to influence the investment decision of a reasonable investor. In this respect, actual market reactions may nonetheless be taken into account to assess the materiality threshold, that is, the point at which a transaction conveys a signal to the market.

FINMA supervises the conduct of market participants in the Swiss financial market in order to ensure the integrity of price formation. It is responsible for enforcing Art. 143(1) FinMIA. The supervisory instruments provided for under FINMASA apply to any person engaging in market manipulation^[56]. The monitoring of securities transactions is delegated to trading venues, which are required to report suspicious transactions to FINMA. In Switzerland, two stock exchanges – namely SIX Swiss Exchange and BX Swiss Exchange – are authorized by FINMA.

6.2. Case 2C_315/2020: Spoofing and Layering Acts

In a judgment of 7 October 2020, the Federal Supreme Court upheld FINMA's decision against a company and its sole shareholder, acting as its governing body, which had, inter alia, breached Art. 143(1)(b) FinMIA^[57]. The Federal Supreme Court relies on the conduct of the perpetrators to establish the existence of market manipulation. In this respect, it applies FINMA Circular 13/8 to determine whether the conduct in question was manipulative.

The supervisory authority of SIX Swiss Exchange examined the company's trading conduct for the period from 3 December 2012 to 5 August 2013. On 26 March 2014, it forwarded its investigation report to FINMA, which subsequently initiated an enforcement proceeding on the basis of substantiated indications of systematic market manipulation. In its final decision of 20 June 2017, FINMA found that the company had seriously breached regulatory requirements. According to FINMA, the company and its shareholder engaged in market manipulation by employing spoofing, layering and ramping strategies. In the present case, on all analyzed trading days on which the complainants traded derivative instruments with the corresponding shares as underlying assets (approximately 100 samples), his trading behavior was characterized by the systematic creation of a misleading surplus of orders in the underlying securities (*spoofing*), predominantly placed at multiple price levels (*layering*), followed by the immediate cancellation of those orders^[58]. The complainant thereby derived economic benefits in the value of the related derivative instruments. He further switched sides of the order book with high frequency, repeatedly reproducing the same trading pattern. In nearly all instances, a persistent inconsistency was observed between executed trades and submitted orders^[59]. Within the sample examined, only a marginal proportion of the total recorded order volume was ultimately executed.

According to the Swiss Federal Supreme Court, the short interval between the placement and cancellation of orders, together with their high volume, indicates the absence of any genuine economic rationale for the transactions. Virtually all analyzed samples reflect a systematic and disproved trading pattern. Thus, the complainants' conduct was aimed at generating impermissible income through manipulative trading strategies, of which he must have been aware^[60].

As evidenced by the random samples, the complainant repeatedly created the appearance of selling a certain volume of securities. Other market participants followed the price movement thereby induced, generating selling pressure on the securities concerned^[61]. The complainant subsequently cancelled his sell orders, waited a few second, and replenished his holding at lower prices. Following the acquisition of the securities, market prices rose again, enabling the complainant to resell the securities at a profit^[62].

In its defense, the complainant contends that it was engaged in market-making

and activities^[63]. However, market makers ensure continuous bids and ask quotations for securities with broadly symmetrical volumes on both sides of the order book, with aim of profiting from the spread between bid and ask prices. Hedging, by contrast, seeks to mitigate price risk, typically by offsetting exposure in the underlying asset through derivative positions. According to the Swiss Federal, nothing in the present case suggests that the complainant acted in either capacity. Liquidity was provided only on a large scale for very short periods and in a unilateral manner, creating one-side order book surpluses that were regularly withdrawn almost immediately. Such conduct is inconsistent with that of a market maker. Moreover, the extent to which the complainant engaged in hedging is neither apparent nor substantiated. This conclusion is further supported by the fact that the positions taken in the underlying assets were generally of limited economic significance, while the leverage of the derivative positions was comparatively high, militating against any genuine neutralization of market risk^[64].

6.3. AI-Based Supervision against Manipulative Behaviours

SIX Swiss Exchange relies on an AI-based surveillance tool, known as Prometheus, to monitor trading activity:

- This system identifies patterns suggesting insider trading and market manipulation. The technology underlying this system incorporates, inter alia, machine learning techniques capable of identifying both known and previously undiscovered forms of market manipulation. Its effectiveness stems in particular from its ability to process large volumes of trading data and to establish correlations between data points that may not have been identified through traditional surveillance methods^[65].
- In this process, the system identifies suspicious transactions and generates an alert, which is then transmitted to the team responsible for conducting a more in-depth investigation of the conduct in question^[66]. Where SIX considers the conduct to be manipulative, it subsequently transmits all relevant documentation to FINMA.

It follows from the above that the process implemented by SIX reflects the

following division of tasks: the quantitative component – namely the identification of correlations within large datasets – is performed by the machine, whereas the reasonable-investor test, which constitutes a qualitative assessment, is carried out by humans. In other words, the signal generated by the system serves merely as an indication that must subsequently be corroborated through a qualitative human analysis. This balance is broadly consistent with the requirements governing human oversight under the EU AI Act.

Assuming that AI tools were, in the future, capable of performing the qualitative assessment as well, such an evolution would, in our view, remain acceptable – taking into account the considerations developed earlier in this Chapter – provided that human oversight does not become a mere formality. Human reviewers would need to engage meaningfully with the probabilistic nature of AI outputs, questioning the system to understand why certain parameters were weighted more heavily than others, or why some were disregarded altogether. Ultimately, the reasoning must be clearly intelligible and reflected in the authority's final decision.

In our view, the use of AI systems should not alter the applicable legal test. Their deployment cannot exempt SIX or FINMA from the obligation to assess conduct through a qualitative examination grounded in the reasonable-investor standard. Both institutions must remain capable of substantiating – where appropriate with the assistance of AI tools – that the transactions at issue were capable of influencing the investment decision of a reasonable investor. While model-based methods may be employed, the assessment of the signalling effect must remain qualitative and must take into account all relevant circumstances.

7. Conclusion

As a preliminary note, while the use of AI tools by regulated entities has already attracted considerable regulatory attention, the deployment of such tools by financial supervisory authorities remains comparatively underexplored. Against this backdrop, the main findings and arguments advanced in this contribution may be summarized as follows:

- The initial issue that arises concerns the applicable legal framework.

Within the European Union, it is acknowledged that the use of AI by public authorities entails risks comparable to those associated with private actors. Consequently, the EU legislature has opted to subject both categories of users to the EU AI Act. Moreover, it is recognized that AI-enabled public authority activities may pose heightened risks to individual rights and freedoms – a category that, in our view, may encompass supervisory functions in the financial sector.

- A further question concerns the limits and requirements governing such use, particularly with respect to the human–machine relationship. In line with its human-centric approach, the EU AI Act unequivocally requires the presence of human oversight, while remaining silent on the precise modalities through which such oversight must be exercised. Given the inherently probabilistic nature of AI systems, supervisory authorities must be able to interrogate these systems in order to understand why certain facts, indicators, or assumptions were weighted or disregarded. Human oversight therefore cannot be reduced to a merely formal requirement. Moreover, the EU AI Act is ambiguous as to whether a decision may be fully prepared through AI-based processes. In our view, this should indeed be permissible for supervisory authorities – although it is doubtful that current AI tools are capable of achieving such a result – provided that the final output is subjected to meaningful human review.
- The final issue concerns the degree of transparency required of supervisory authorities that rely on AI. The EU AI Act itself does not impose explicit transparency obligations in this regard. However, this absence is of limited practical consequence, as supervisory authorities are already bound by constitutional requirements to provide reasons for their decisions. This obligation ensures that the authority’s reasoning must be articulated clearly. At the same time, it exerts a disciplining effect on authorities, reinforcing the necessity of critically assessing and validating the output generated by AI systems.

In light of the foregoing, the balance struck between machine and human involvement in financial-market supervision appears acceptable under the EU’s AI-specific regulatory framework, particularly when read together with the

constitutional guarantee of the duty to state reasons. This balance may also serve as a reference for other jurisdictions, including Switzerland.

1. See, for instance, Swiss Financial Market Supervisory Authority FINMA Guidance 08/2024, of 18 December 2024, *Governance and risk management when using artificial intelligence*, available at <https://www.finma.ch/en/news/2024/12/20241218-mm-finma-am-08-24/>; Turksen, Benson, Adamyk, *Legal implications of automated suspicious transaction monitoring: enhancing integrity of AI*, in *Journal of Banking Regulation*, 2024, Vol. 25, pp. 359 *et seq.*
2. On 14 October 2025, the Swiss Federal Audit Office has issued a report titled “*Use of artificial intelligence in supervision*” which focuses on the said use by the Swiss Financial Market Supervisory Authority (FINMA); this report is available at <https://www.efk.admin.ch/en/audit/use-of-artificial-intelligence-in-supervision/>. In September 2025, the U.S. Securities and Exchange Commission (SEC) has issued the “*SEC’s 2025 AI Compliance Plan*” which is available at <https://www.sec.gov/ai>. On this topic, see also, for instance, Bains et al., *AI Projects in Financial Supervisory Authorities – A Toolkit for Successful Implementation*, IMF Working Paper WP/25/199, October 2025, available at <https://www.imf.org/-/media/files/publications/wp/2025/english/wp25199-source-pdf.pdf>; Toronto Centre, *The Supervisor of the Future*, June 26, 2025, available at <chrome-extension://efaidnbmninnibpcjpcglclefindmkaj/https://www.imf.org/-/media/files/publications/wp/2025/english/wp25199-source-pdf.pdf> https://torontocentre.org/index.php?option=com_content&view=article&id=600&Itemid=99#:~:text=AI%20can%20assess%20systemic%20risks,Behavioural%20analytics; Kasireddy, *The Ethical Implications of AI in Financial Market Surveillance: Are We Over-Monitoring traders?*, in *European Journal of Accounting, Auditing and Finance Research*, 2025, Vol.13, No. 4, pp.17 *et seq.*
3. This contribution focuses on the use of AI and does not extend to other related topics, such as real time automated surveillance, which raises additional critical issues.
4. In April 2025, a new Integrated Risk Expertise division has been created within the Swiss Financial Market Supervisory Authority (FINMA). This cross-divisional unit provides the supervisory divisions with expertise in data-driven analysis, analysis of financial and non-financial risks and operational resilience through several competence centers. It is also responsible for the planning, coordination, implementation and quality assurance of on-site supervisory reviews. A central function of this new unit includes, among other things, the design, further development, and operation of AI systems for risk and market monitoring. See Swiss Federal Audit Office, *Use of artificial intelligence in supervision*, pp. 12 *et seq.*; this report is available at <https://www.efk.admin.ch/en/audit/use-of-artificial-intelligence-in-supervision/>.
5. Tools that merely aim to improve the authority’s internal organization and administrative functioning are set aside in this paper. In this respect, it is rather surprising that the SEC

- seems to rely only on such limited tools; see the excel sheet titled “2025 SEC Consolidated AI Use Cases” available at <https://www.sec.gov/ai>.
6. See notably Rohara, *The Impact of Artificial Intelligence on Financial Auditing Practices*, April 21, 2025, available at <https://ssrn.com/abstract=5246748>. It may be expected that AI systems may also play a significant role in prioritisation of supervisory tasks. Authorities typically face resource constraints preventing uniform monitoring of all market participants. AI-driven classification and clustering techniques can assist in identifying entities or behaviours that warrant closer inspection. By generating prioritisation models, AI supports a more efficient allocation of supervisory resources and contributes to a risk-based approach to oversight, which is widely recognised as a cornerstone of modern financial regulation.
 7. See the case study in Section 6 of this paper.
 8. Toronto Centre, *The Supervisor of the Future*, June 26, 2025, available at https://torontocentre.org/index.php?option=com_content&view=article&id=600&Itemid=99#:~:text=AI%20can%20assess%20systemic%20risks,Behavioural%20analytics. See also Clarke, Moreau, *AI-Powered Stress Testing and Scenario Analysis for Financial Stability Assessment*, available online at https://www.researchgate.net/publication/399284947_AI-Powered_Stress_Testing_and_Scenario_Analysis_for_Financial_Stability_Assessment.
 9. Specifically, the risks associated with the use of AI in transaction surveillance lie primarily in false negatives and false positives. While a certain degree of false negatives – namely, undetected manipulative conduct – may be tolerable, false positives are more problematic, as they may trigger investigations into conduct that is in fact lawful.
 10. With respect to the risk of error, see Recital 61 EU AI Act.
 11. With respect to the risk of bias, see Recital 61 EU AI Act.
 12. The concept of trust is at the heart of the report issued by Swiss Federal Audit Office titled “*Use of artificial intelligence in supervision*” which focuses on the said use by the Swiss Financial Market Supervisory Authority (FINMA); this report is available at <https://www.efk.admin.ch/en/audit/use-of-artificial-intelligence-in-supervision/>.
 13. On this topic more generally in connection with the use of AI in the public administration, see for instance Swiss Confederation, Strategy – Use of AI systems in the Federal Administration, 11 December 2025, p. 3; this document is available at <https://www.bk.admin.ch/bk/en/home/digitale-transformation-ikt-lenkung/vorgaben/sb021-strategie-einsatz-von-ki-systemen-in-der-bundesverwaltung.html>.
 14. See notable Recital 13 EU AI Act.
 15. Recitals 3, 58 and 60 EU AI Act.
 16. Articles 3(3) and (4) EU AI Act.
 17. Articles 3(3) and 25 EU AI Act.
 18. Article 3(1) EU AI Act.
 19. Recital 7 EU AI Act. For more details regarding the various categories under the EU AI Act, see Caballero Cuevas, *Systèmes d’intelligence artificielle – Les catégories de la*

réglementation européenne, 2024. Available at <https://cdbf.ch/1390/>.

20. In our view, the notion of “judicial authority” should be interpreted broadly to encompass administrative bodies – or at least those administrative bodies empowered to impose measures and sanctions – since the use of AI tools by such authorities raises issues comparable to those arising from their use by courts.
21. Annex III(8)(a) EU AI Act. Potentially, Annex III(6) EU AI Act, which relates to law enforcement, may also be relevant in the context of financial supervision, it being however specified that «*AI systems specifically intended to be used for administrative proceedings by [...] financial intelligence units carrying out administrative tasks analysing information pursuant to Union anti-money laundering law should not be classified as high-risk AI systems used by law enforcement authorities for the purpose of prevention, detection, investigation and prosecution of criminal offences*» (Recital 59 EU AI Act).
22. Couneson, *Classification Rules for High-Risk AI Systems*, in Pehlivan, Forgo, Valcke (eds), *The EU Artificial Intelligence (AI) Act – A Commentary*, Wolters Kluwer, 2025, p. 213, referring to Recital 61 EU AI Act.
23. Recital 61 EU AI Act.
24. Article 5 EU AI Act.
25. Article 9 EU AI Act.
26. Article 10 EU AI Act.
27. Article 11 EU AI Act.
28. Article 12 EU AI Act.
29. Article 13 EU AI Act.
30. Article 14 EU AI Act.
31. Article 15 EU AI Act.
32. Article 26(1) EU AI Act.
33. Article 26(2) EU AI Act.
34. Article 26(6) EU AI Act.
35. Article 26(8) EU AI Act.
36. Article 27 EU AI Act.
37. Panezi, *Human Oversight*, in Pehlivan, Forgo, Valcke (eds), *The EU Artificial Intelligence (AI) Act – A Commentary*, Wolters Kluwer, 2025, p. 365.
38. Article 14(3) EU AI Act. To be further noted that human oversight must be connected with transparency requirements, which aim at addressing black box issues. See Corrêa, *Transparency and Provision of Information to Deployers*, in Pehlivan, Forgo, Valcke (eds), *The EU Artificial Intelligence (AI) Act – A Commentary*, Wolters Kluwer, 2025, p. 343.
39. Article 27(1) EU AI Act. More comprehensively, this assessment shall consist of the following: (a) a description of the deployer’s processes in which the high-risk AI system will be used in line with its intended purpose; (b) a description of the period of time within which, and the frequency with which, each high-risk AI system is intended to be used; (c) the categories of natural persons and groups likely to be affected by its use in the specific context; (d) the specific risks of harm likely to have an impact on the categories of

natural persons or groups of persons identified pursuant to point (c) of this paragraph, taking into account the information given by the provider pursuant to Article 13; (e) a description of the implementation of human oversight measures, according to the instructions for use; (f) the measures to be taken in the case of the materialisation of those risks, including the arrangements for internal governance and complaint mechanisms.

40. Krause, Hübotter, *Probabilistic Artificial Intelligence*, ETH Zürich, 7 February 2025, available at <https://arxiv.org/abs/2502.05244?>.
41. Under the EU AI Act, the critical concept is the fact that a system has the ability to “infer” results from the input it receives, as the ability of an AI system to “infer” transcends basic data processing, enables learning, reasoning and modelling. See Feiler, König, *Definitions*, in Pehlivan, Forgo, Valcke (eds), *The EU Artificial Intelligence (AI) Act – A Commentary*, Wolters Kluwer, p. 58, 2025, referring to Recital 12 EU AI Act.
42. Panezi, *Human Oversight*, in Pehlivan, Forgo, Valcke (eds), *The EU Artificial Intelligence (AI) Act – A Commentary*, Wolters Kluwer, 2025, p. 365.
43. Recital 61 EU AI Act.
44. Corrêa, *Transparency and Provision of Information to Deployers*, in Pehlivan, Forgo, Valcke (eds), *The EU Artificial Intelligence (AI) Act – A Commentary*, Wolters Kluwer, 2025, p. 343.
45. Gils, *Transparency Obligations for Providers and Deployers of Certain AI Systems*, in Pehlivan, Forgo, Valcke (eds), *The EU Artificial Intelligence (AI) Act – A Commentary*, Wolters Kluwer, 2025, p. 780 *et seq.*
46. Bovet, Popadic, *Art. 35 PA - Motifs et indication des voies de recours*, N 5, in Bellanger, Candrian, Hirsig-Vouilloz (eds), *Commentaire romand de la loi fédérale sur la procédure administrative*, Helbing Lichtenhahn, 2024.
47. On this issue in general (not limited to financial supervision), see Hendrickx, *The Judicial Duty to State Reasons in the Age of Automation? The Impact of Generative AI Systems on the Legitimacy of Judicial Decision-Making*, September 20, 2024. Available at <https://ssrn.com/abstract=5043685>.
48. SR 958.1 (Swiss federal law register).
49. Circ.-FINMA 2013/8, Cm 19-30.
50. Federal Tribunal, 2C_315/2020, § 7.2.3 ; Federal Administrative Tribunal, B-2370/2022, § 4.1.1 ; Federal Council, 2011, *Message sur les abus de marché*, p. 6358.
51. See *Basic, Inc. v. Levinson*, 485 U.S. 224 (1988), p. 231 s., in particular «*a substantial likelihood that the disclosure of the omitted fact would have been viewed by the reasonable investor as having significantly altered the ‘total mix’ of information made available*» (emphasis omitted).
52. See e.g. Federal Tribunal, ATF 145 IV 407, § 3.4.1, JdT 2020 IV 163 ; Federal Criminal Tribunal, CA.2023.27, § 1.4.5.6 ; Federal Criminal Tribunal, SK.2022.30, § 2.2.4.6.
53. Darbellay, Caballero Cuevas, *The Materiality of Sustainability Information under Capital Markets Law*, in : RSDA 2023, vol. 95, No 1, p. 46.
54. Remund, *L’exploitation d’informations d’initiés selon les articles 154 et 142 LIMF : une*

étude de droit suisse et comparé, thèse Lausanne, Zurich, Bâle, Genève 2021, pp. 377 *et seq.* ; Kumpan, Misterek, *Der verständige Anleger in der Marktmissbrauchsverordnung : Zu den Eigenschaften der Massstabsfigur für Insiderinformationen*, in *ZHR*, 2020, vol. 184, pp. 193 *et seq.*; Klöhn, *Artikel 7*, N 268 *et seq.*, in Klöhn Lars, Brellochs Michael, Schmolke Klaus Ulrich, Semrau Stephan, *Marktmissbrauchsverordnung : Verordnung (EU) Nr. 596/2014 über Marktmissbrauch*, 2nd ed., Munich, 2023.

55. Federal Criminal Tribunal, CA.2023.27, § 1.4.5.6.
56. Art. 145 FinMIA.
57. Federal Tribunal, 2C_315/2020.
58. Federal Tribunal, 2C_315/2020, § 7.3.1.
59. Federal Tribunal, 2C_315/2020, § 7.3.1.
60. Federal Tribunal, 2C_315/2020, § 7.3.1 *et seq.*
61. Federal Tribunal, 2C_315/2020, § 7.3.1 *et seq.*
62. Federal Tribunal, 2C_315/2020, § 7.3.1 *et seq.*
63. Federal Tribunal, 2C_315/2020, § 7.3.4
64. Federal Tribunal, 2C_315/2020, § 7.3.4.
65. Leybold, Müller, *SER bekämpft Marktmissbrauch mit KI-basierter Technologie*, 2022.
Available at <https://www.pwc.ch/de/insights/digital/six-prometheus.html>.
66. Leybold, Müller, *SER bekämpft Marktmissbrauch mit KI-basierter Technologie*, 2022.
Available at <https://www.pwc.ch/de/insights/digital/six-prometheus.html>.

Trusting the use of AI by the authorities, the Swiss terms and conditions at the forefront

Marc-Olivier Busslinger

DOI: 10.13130/2723-9195/2026-2-42

È generalmente riconosciuto che la fiducia nelle autorità riveste un ruolo fondamentale nel comportamento dei cittadini. Inoltre, il diritto costituzionale impone alle autorità pubbliche l'obbligo di agire in buona fede e di comportarsi in modo affidabile. Allo stesso tempo, le autorità mirano ad implementare nuove tecnologie e sistemi di intelligenza artificiale per fornire servizi pubblici migliori e sperano di ottenere un aumento dell'efficienza. I chatbot sono stati tra i primi sistemi di intelligenza artificiale ad essere stati implementati dalle autorità svizzere. Per cercare di limitare la propria responsabilità riguardo all'uso di questi sistemi, le autorità amministrative, a tutti i livelli di governo, hanno scelto di pubblicare termini e condizioni per informare il pubblico e limitare l'uso di questi chatbot. Sebbene la qualificazione giuridica di questi testi e la validità del loro contenuto siano discutibili, la nostra ricerca analizza il contenuto e la formulazione di questi termini e condizioni e le ampie clausole di esclusione di responsabilità in essi contenute per determinarne il probabile impatto sulla fiducia. La tesi qui sostenuta è che, rifiutando qualsiasi responsabilità, le autorità pubbliche presentano questi servizi, e di conseguenza se stesse, come inaffidabili. Ciò, a sua volta, potrebbe avere un impatto più ampio sulla percezione sia dell'IA per i servizi pubblici sia dello sviluppo dei servizi di eGovernment in generale.

It is generally accepted that trusting authorities plays a key role in citizens' behaviour. Moreover, constitutional law places a duty on authorities to act in good faith and to behave in a trustworthy manner. At the same time, authorities wish to implement new technologies and AI systems to provide better public services and hope for efficiency gains. Swiss authorities were among the first to implement chatbots as AI systems. To try to limit their liability concerning the use of these systems, authorities on all levels of government chose to publish terms and conditions to advise the public and curtail the use of these chatbots. Although the legal qualification of these texts and the validity of

their content are questionable, our research analyses the content and wording of these terms and conditions and the broad disclaimers they contain to determine their probable impact on trust. We argue that by refusing any accountability, the authorities present these services, and themselves by extension, as untrustworthy. This, in turn, might have a broader impact on the perception of AI in public services and the overall development of e-government services.

Summary: 1. Introduction.- 2. Trust.- 2.1. A definition of Trust.- 2.2. The need to act and to be perceived as trustworthy when providing public services.- 2.3. The probable impact of digital transformation and AI systems implementation on trust.- 3. The implementation of AI systems by the Swiss authorities.- 3.1 A quick overview of the adoption of AI systems by public administrations.- 3.2 The implementation of Chatbots in particular.- 4. The terms and conditions and their influence on Trust.- 4.1 The use of terms and conditions surrounding public chatbots in Switzerland.- 4.2 The content of the terms and conditions surrounding the use of such chatbots.- 4.3 The legality of such disclaimers.- 4.4 The influence of these disclaimers on public trust.- 5. Possible pathways to foster or restore Trust in the testing and implementation of AI systems.- 6. Conclusion.

1. Introduction

AI systems are rapidly developing in public administrations, and we believe that the first systems to be implemented and proposed for use by the public can give precious insight on trends that might generalise in the development of such systems. In this sense, chatbots based on LLMs are a perfect terrain to study, as “pioneer” project used to test such technologies. We believe that the way authorities act, communicate and try to regulate such novel AI systems can have an influence on the public sentiment around these technologies in the trust in the authorities’ actions. Thus, studying the content of the texts supposed to inform the users of such AI systems about their rights allows us to both get insight into the authorities’ concrete view of these technologies and analyse their possible impact on public trust^[1].

In this sense, this study doesn't approach the technical aspects of the reliability of the AI systems implemented^[12], nor conducts a psychological study of the perception by the users of these systems^[13]. We chose to concentrate on one type of "cue"^[14] of the trustworthiness of public administration AI that are the rules the authorities communicate to the people using their chatbots. To provide a practice-oriented analysis, we decided to limit the scope of this case study to terms and conditions of chatbots present on Swiss public law institutions. These texts are a tool used by the authorities to communicate about the degree of accountability^[15] – or in this case lack thereof – they are willing to assume. Although similarities can be found in other countries that implemented similar texts with similar contents, we had to limit ourselves to a national context for practical reasons. This paper analyses the terms and conditions of selected public chatbots and AI systems. Switzerland being a federal system with a high degree of decentralisation, three levels of state coexist, and we searched for examples from each level of Government and each linguistic region. Since there is no definitive list of existing systems^[16], we based our analysis on systems either present on these curated lists, that had had media coverage or that were present in the literature. Since our study focuses on the use by the authorities of such systems, we chose chatbots implemented directly on the homepage of an administration, or the ones present on a page of an administrative body. Since our focus is on the trust the citizens should place in the use of AI by the authorities, we chose only publicly available chatbots^[17]. For our analysis to be rooted in public law, we chose systems that were implemented by the authorities on their official webpage or that were reachable by means advertised on their websites^[18]. These systems are in this sense illustrative of a common trend both in the use of such terms and conditions as well as in their wording.

The broader subject of "trust in the public authorities" concerns the three levels of government, each in their own sphere of responsibility. For the public, however, the subject isn't as clear, digitisation of state has been a major factor for the development of "one stop government" that tends to blur the line between the areas of competencies of the different authorities.

This paper chose to focus on the disclaimers present in these terms and conditions, these exclusions of responsibility that wish to exonerate the authorities for the information given. We primarily focus on the rules pertaining

to data protection that have already been studied^[9]. We will approach the way these data protection rules differ from the rules surrounding the content produced by the chatbots and the influence of these differences on public trust. The systems implemented by public administration are constantly evolving, the nature of the terms of conditions allow for their evolution which further complicates the analysis. For example between the beginning of our study in September 2025 and the completion thereof in March 2026, one website underwent a complete overhaul in January and the chatbot is now unavailable and points to an error^[10] on the webpage of the authority advertising it^[11]. In this sense, our study is example based and qualitative in nature. We argue that the relatively large scope of entities considered, on all level of government, from all regions, allow us to show a trend in the way authorities across Switzerland implement terms and conditions to try to limit their responsibility.

2. Trust

2.1. A definition of Trust

Trust is one of the building blocks of society and law. Each discipline studying trust have different approaches and views the definition of trust, and there still doesn't seem to be a definitive consensus across scientific fields about this notion^[12].

Trust is often defined as the «*willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control that other party*»^[13].

This makes trust something that is perceived by people, is not necessarily rational and is influenced by the interaction, context and the relationship between the parties^[14]. The nature of trust places it outside of the direct control of an entity. It is not possible to command trust, but only to behave in such a way to generate trust in one's counterpart. By behaving in a trustworthy manner, an actor might be able to inspire trust.

From a legal standpoint, the need for trust is both a requirement in the relationship between the actors, and a core principle underlying the whole legal

environment^[15]. For rules to be effective there needs to be a shared belief that these rules will be respected and enforced.

2.2. The need to act and to be perceived as trustworthy when providing public services

The relationship between the individual and the authorities is asymmetric and there is a definite verticality and power relationship between the individuals and the authorities^[16]. The individuals have administrative obligations towards the authorities, while authorities are bound by the rule of law and administrative principles^[17].

But trust isn't outside of the equation. By behaving in a trustworthy manner, the authorities will gain in legitimacy^[18]. The perceived trustworthiness of the authorities might have an influence on the behaviour of the citizens. The development of good governance principles^[19] often refers to the trustworthiness of the authorities to increase their legitimacy or as a justification for the implementation of certain legislation or principles.

Although this assumption is challenged by the literature, transparency of the authorities is considered a «*key value for trustworthy governments*»^[20]. To create trust in the institutions, empirical studies conducted have mostly focused their attention on the right to ask for transparency the rights to demand access to government documents. These studies couldn't find a definitive impact of such "rights for transparency" on the trust the individuals put into the government, the successful interaction with a government agency having more of an effect^[21]. Some studies have even found that the knowledge of the existence of a latent right to transparency, knowing one has a right to ask for information from the state, wouldn't necessarily produce higher trust in the government itself^[22].

Since trust is a building block of constitutional democracies and a big factor in the legitimacy of the government^[23]. Every actor is therefore required to act in good faith^[24] (art. 5 al. 3 Cst.^[25]). Which is then specified by art. 9 Cst. «*Every person has the right to be treated by state authorities in good faith and in a non-arbitrary manner*». In the German version, this good faith element is presented as a «*duty to be trustworthy*»^[26]. This prerequisite is also enshrined in the art. 2 of the Civil code^[27] for the relationships between individuals and enterprises.

2.3. The probable impact of digital transformation and AI systems implementation on trust

The implementation of new eGovernment services by the authorities presents a “new” challenge for trust. Indeed, since digital transformation is mostly an iterative process, most digitisation processes go through a trial period before transitioning entirely, and suppressing existing analog or pre-existing digital systems. In this period their adoption by users might have a particular impact on their future adoption. In the first phase of their implementation, the users need to make the decision to “try out” the new system, to adopt a new way to do a process and transition from a mostly analog process to the new digital one.

For this reason, the public discourse often presents trust as being a core component of the acceptance of eGovernment systems^[28]. Trust also appears as a central principle in strategies on the implementation of AI systems^[29]. Studies have focused on the influence of trust of the citizens in the State for the uptake of such systems^[30], on the trust in the technologies themselves^[31] or even the trust in the use governments would do of these technologies^[32]. Legal studies have discussed the concept of trustworthy AI, its origin and its influence on AI regulation^[33]. Most of these studies seem to have found a positive effect of trust in the adoption of the eGovernment services^[34].

The importance of trust in the implementation of digital technologies is a developing field of study. Some focus on the changes AI could represent in the definition of trust itself^[35], while other research ways to implement certain pieces of technology to foster trust^[36]. While these approaches are either technical or philosophical, our study concentrates on the legal framework and the way the authorities communicate about such systems and the influence these texts could have on trust. We hope in this sense that our paper, coming from a legal background, will inspire more experimental studies for researcher focusing on the philosophical and psychological aspects of trust in eGovernment and AI.

3. The implementation of AI systems by the Swiss authorities

3.1 A quick overview of the adoption of AI systems by public administrations

The evolution of technologies and their adoption by private companies is encouraging public authorities to follow suit to «*keep up with the times*»^[37]. The adoption of technology is in this sense not a new process^[38], the development of internet and more recently of AI technology empowered Swiss authorities to start a digitisation process of their internal administrative processes wishing for such processes to be «*digital first*»^[39]. This modernisation isn't based on competitive mechanisms, as it would be the case in the private sector^[40]. ICT is often seen by public authorities as a tool for achieving effectiveness gains and efficiency gains^[41]. By computerising data collection and processing, the state can provide better services at a lower cost^[42]. The automation of certain computerised processes constitutes a logical next step in this quest. Some researchers even consider that automation is the result of a fundamental dynamic in public law resulting from the desire to limit arbitrariness and to ensure that the law is applied as uniformly as possible, which would thus lead to the development of a “mechanical” law and hence lead to the automation of state processes^[43].

While AI is not yet widespread in the public sector, some sector-specific systems are being developed^[44]. Their adoption by the public authorities follows a similar logic of seeking effectiveness and efficiency^[45]. The hoped-for efficiency gains, and hence improvements in the quality of the public service, are generally aimed at improving output, that is, the service provided to the public. This improvement can be achieved by reducing the risk of errors, by improving service availability as well as the processing speed of cases. Improvements to the public service can also be achieved by changing the processes themselves, or by using more and better data^[46]. Numerous AI applications wish to simplify the work of the agents responsible for managing files within the administration^[47] or envision to help combat fiscal fraud or crime^[48]. Efficiency gains, and therefore potential savings, seem to be expected in payroll; by automating tasks, the authorities hope to be able to do more, and better, with the same amount or even fewer people; thus they hope to reduce the financial burden on the state budget. Few studies have been carried out to date to calculate the real cost-benefit *ratio* of the

implementation of AI systems. This might be explained by the novelty of the phenomenon and the rather prospective aspects of this research^[49]. It is not unlikely that the high development costs of these specialised systems and the additional costs associated with maintaining IT systems might, in many cases, prevent authorities from achieving real efficiency gains.

3.2 The implementation of Chatbots in particular

Following the rapid uptake of generative AI on the market, many administrative bodies took upon themselves to “try out” the implementation of such systems by creating “chatbots” that would help better communicate with the public^[50]. The main reason for implementing chatbots in public administrations is to improve administrative efficiency^[51]. Some researchers believe that chatbots could represent an «*exploratory application of AI in such organisations*»^[52]. These types of systems are the most common type of artificial intelligence system implemented in public administrations in Switzerland^[53]. These systems facilitate access to government information, as these conversational systems are available to citizens 24 hours a day, whereas physical counters or telephone counters are only available during fixed hours, often office hours^[54]. These chatbots also improve access to information for the authorities by lowering the barriers of access. Indeed, these systems are directly accessible on the authorities’ web page and it is not necessary for a person looking for information to go through multiple steps to find it. The question of the quality of the information provided is open to debate. While citizens seem relatively ready to believe the information provided by these systems^[55], the systems implemented to date have several limitations partly due to the way LLMs operate^[56]. Indeed LLM’s have, for the moment at least, technical limitations and are prone to “hallucinations”, which in layman’s term means that AI might give out wrong or incomplete answers. We appreciate the term of “careless answers” to approach the notions of particularly wrong or dangerous answers^[57]. This notion allows for an adaptation of the threshold for exactitude depending on the sensitivity of the system. In this sense, public chatbots should be held to higher exactitude standards. It is still unclear if the technological development will really be able to guarantee the total absence of errors. For chatbot systems, more than a public interest, the authorities have an

obligation, based in public law, to inform the public about their activities^[58]. On the federal level for instance, the Government is supposed to «*inform the general public fully and in good time about its activities*» (art. 180 al. 2 Cst.^[59], implemented by the Government and Administration Organisation Act^[60] (art. 10), which embodies for the federal administration an *active* duty to be transparent. Similar laws were also adopted on cantonal level^[61].

The right to ask for access to documents, often described as a passive transparency principle, is prescribed on the federal level by the Federal Act on Freedom of Information in the Administration (FoIA)^[62]. The same types of obligations of transparency are implemented in cantonal laws^[63]. For the moment, there doesn't seem to be any chatbots implemented to fulfil this type of transparency. Indeed, the procedural requirements are pretty narrow, and most current laws wouldn't allow chatbots to fulfil the requirements needed to give access to specific documents.

Chatbots partly embody that “active” transparency principle, in the sense that they give a new way for the administration to fulfil their obligation of information towards the public. These chatbots might also have an organisational function that can be linked with every special procedure they support. The authorities have an obligation to inform the public about the way in which administrative proceedings can be conducted^[64], they might choose chatbots to do so. It might then be less of a generic “information” of the public, and more a tool used to fulfil an obligation resulting from the proceedings themselves. For example, the chatbot established by Basel land^[65] from unemployed people finds its origin in the general competence of the cantonal office for industry, trade and labour to execute the federal law on the matter^[66] (§1 AVLG-BL^[67]) and to operate contact, information and training centres for jobseekers (§11 AVLG-BL). Since this chatbot was implemented notably to inform the individuals about the proceedings needed to grant them a right for unemployment benefice, this technology is at the border between a general duty to actively inform the public and a specific obligation linked to a specific administrative procedure. If most chatbots discussed in this paper are closer to the general duty to inform, the development of AI and the adoption of such systems could represent a shift towards the development of chatbots as means to conduct administrative proceedings.

If multiple types of systems might have been called chatbots by the authorities, starting from mere interactive information panels without any user inputs, the majority of chatbots developed in the last few years use AI and LLMs to process textual inputs from the users and give “answers” to their questions.

The industry predicts the evolution of “virtual assistants” chatbots^[68], that would, based on a conversation with the user automate some real-life processes. We couldn’t find an example of such technologies being implemented on Swiss public websites to date^[69]. Some websites used forms that had components that in some ways are akin to interactive questionnaires, with an array of options to choose from^[70]. The way these forms operate forbids assimilating them to real chatbots. Moreover, such systems weren’t presented as “chatbots” by the authorities, presumably to differentiate between the conversational – and informational – agent and the form being part of an actual administrative proceeding. Although we didn’t find literature on the subject, it doesn’t seem improbable to see AI chatbots soon be implemented into, or maybe even totally replace forms and form like systems to conduct administrative proceedings.

The chatbots implemented by the authorities either use ChatGPT^[71] or LLMs implemented by DeepLink^[72] a swiss company that seems to have developed their own AI system. We also found one vocal assistant that is based on Spitch^[73].

4. The terms and conditions and their influence on Trust

4.1 The use of terms and conditions surrounding public chatbots in Switzerland

Terms and conditions are present either in the chatbot itself, for example in an information checkbox at the top of the chat window, with a message in the chat, or even with a link in the footnotes of the webpage giving access to the chatbot itself.

For example, the chatbot of the migration office of the canton Basel land^[74], has a message that asks the users to consent to the use of data with a link to a pdf document containing data protection information^[75]. Since the chatbot is on the canton’s website, their general terms and conditions^[76], available through a link at the bottom of each page named «*disclaimer*», also apply to the use of the

chatbot. The chatbot of the local police in Morges, in Canton Vaud, only has a link at the bottom of the website named «*terms and conditions*», which contains both rules for the content and the data protection^[77]. The chatbot ALVA from Canton Basel-Stadt^[78] has an info box that gives a general disclaimer and a link to the «*data protection rules*»^[79] that also contain disclaimers about the content and rules for the use of the system and asks the user to accept them to start the chat. Some systems have different documents available in their interface, separating the “disclaimer” from the rules surrounding the data processed^[80]. The chatbot of the federal archives^[81] that starts the discussion by a short text informing the user about some data protection rules and a short disclaimer, with links to a pdf document with more detailed rules^[82], and a button asking the user to «*accept*» these principles. This pdf then sends the user to the data protection rules for the whole archive’s website^[83], which themselves send the user to the terms and conditions for the federal administration website^[84], thus layering the information in multiple interconnected documents. The chatbot of the Swiss army^[85], includes an info logo at the top of the windows that gives access to four pop-up windows, one giving general information about the chatbot, one about the use thereof – which actually contains information about the data provided and the suppression of identifying data and metadata – one titled «*disclaimer*»^[86], and one «*legal notice*»^[87].

If most chatbots are implemented on public authorities’ websites, with texts indicating that the services are provided by the authorities themselves, it is worth noting that the chatbot of the municipality of Hochdorf, canton Lucern refers to the chatbot present on their website^[88] as a service provided by a private company^[89] which seem to have set their own terms and conditions in place^[90].

It seems important to note that although most systems presented terms and conditions, we couldn’t find any such provisions for the Vocal assistant of the commune of Kreuzlingen. Although the system «*is not permitted to provide legally binding information*» according to its implementors^[91], nor the specialised administrative webpage indicating the contact number to the assistant^[92], nor the terms and conditions of the communal webpage^[93] give any information about this system. This is particularly striking considering how detailed these terms and conditions are, and the sensitive nature of the voice recording that are processed through Amazon Cloud Services^[94].

As we exemplified, two types of rules are presented or linked to these chatbots. First these texts contain information or rules concerning the data protection of the users. Second rules concerning the content that is published and the texts that will be provided by the chatbots themselves when “answering” questions and the possible responsibility of the authorities for these texts.

4.2 The content of the terms and conditions surrounding the use of such chatbots

Before the terms and conditions themselves, some chatbots inform their users about their use for example Alva «*If you activate me, I will help you find the right information*»^[95], the municipal police state, «*I will assist you in finding information about our services*» before listing different types of information they can provide, in particular «*Getting help if you are experiencing domestic violence*» or «*Obtain a parking permit*»^[96]. The Swiss army chatbot asks «*How can I help you?*»^[97] as a first question in the chat.

These introductory texts also make the user attentive to the limitations of the system; this is for example the case for the BarBot of the federal archives «*It is possible that some information may be incorrect. Please check important information*»^[98]. The Swiss army chatbot includes a warning at the bottom of the chatbot window: «*[t]he Swiss Army chatbot may make mistakes. Please check any important information*»^[99]. Alva states «*[m]y answers are generated by artificial intelligence and are therefore not always correct*»^[100], while the police chatbot indicates, «*I am just starting out and learning every day*»^[101]. The “HofdererBot” of the Municipality Hochdorf indicates at the bottom of the chatpage «*[a]ll information provided without guarantee*»^[102].

The Swiss army chatbot is the only one that clearly stated the experimental nature of the system: «*[t]he current version is a prototype and is being used to carry out further user testing*»^[103], and provided an email address to allow users to provide commentaries and feedbacks.

Some chatbots require the users to accept the terms and conditions by pressing a button titled «*accept*»^[104], «*agreed*»^[105] or «*activate*»^[106] before being able to use the system. While others don’t demand such an explicit consent^[107], the telephonic system of the city of Kreuzlingen doesn’t even inform the caller that the call is

recorded.

The structure of these disclaimers varies widely between websites and services, but most of them starts with a general statement about the goal of the terms themselves^[108], or even with the disclaimers themselves^[109].

In comparison to the rules about data protection, the rules concerning the information and texts produced by the chatbot are scarce. If most chatbots include disclaimers or responsibility waivers, these texts sometimes include indications on the way to ask for human help or review^[110]. There are, however, no indications of any rights by the users concerning the content and texts produced by the users, that means that a right for a human review isn't included, and that these texts don't provide any possibility to get access to past conversations. In the same sense, if some data protection texts gave information about the laws governing their implementation^[111], the same isn't true for the disclaimers and responsibility exemption clauses that gave no information about the public law that allowed such exceptions to be put in place.

The content of the disclaimers is always similar, the authorities reject all and every responsibility, both for the information provided, for any actions taken by the individuals based on the texts provided by the chatbots and even for any direct harm that the chatbots could cause.

The canton Basel-Stadt is particularly precautionous, concerning Alva they state: «*The Canton of Basel-Stadt does not guarantee the accuracy, completeness or timeliness of the information provided by Alva. The information is for general information purposes only and is not legally binding. We ask that you verify the information provided by Alva on the official websites of the canton*»^[112], although the same terms and conditions, under disclaimer^[113] indicate that if they take care to «*ensure that the information on its websites is accurate and up to date*», the content of the website can't be trusted: «*Nevertheless, neither the canton nor individual offices or departments can guarantee the accuracy, reliability or completeness of the information*»^[114]. They then include two disclaimers, one specific to Alva: «*Alva is based on artificial intelligence and may therefore make mistakes or provide inaccurate information. In particular, answers may appear correct but be incorrect in terms of content. The Canton of Basel-Stadt accepts no liability for the consequences resulting from the use of the information provided by Alva*»^[115], which completes the general disclaimer for the whole canton's website:

«The canton accepts no liability for damage or consequential damage resulting from access to its websites or individual parts thereof (such as downloaded documents), their use (or the inability to access or use them) or links to other websites»^[116].

The disclaimer for the chatbot of the federal archives is the most comprehensive we could find. After stating the goal of the chatbot and stating that the chatbot couldn't replace personal advice, the disclaimer waives every responsibility for the system: *«The use of the chatbot and any consequences thereof are your responsibility. The Federal Archives do not guarantee the truthfulness, accuracy, completeness or reliability of the answers provided by the robot. False, inaccurate or invented answers cannot be ruled out. You should always verify the information and sources provided. The Federal Archives are in no way responsible for actions taken by users or their decision not to act on the basis of information provided by the chatbot, nor for any consequences thereof. Any liability claims are expressly excluded»^[117].*

The chatbot of the Swiss army, which intends to *«serve[...] as a key point of contact for military personnel. It handles standard enquiries immediately and forwards more complex requests to the relevant departments»^[118]*, indicates in its disclaimer that the system *«was developed to provide general information and general details about the Swiss army»^[119]*. It then went on to assert that this information doesn't constitute personalised legal advice and that the users should seek advice towards the hotline of the army. This provision was then followed by the broadest disclaimer found to date: *«Use of the chatbot is at your own risk and peril. The Swiss army accepts no liability for the correctness, accuracy, completeness and reliability of the answers provided. False, incorrect and fictitious answers must be expected. Please check each answer, including any sources cited. The Swiss army is not liable if the user takes or fails to take action in reliance on the correctness and completeness of the answers and suffers disadvantages or damages as a result. Any liability claims caused by the use of this website are expressly excluded»^[120]*. A similar provision is still present on the terms and conditions of the canton Basel Land which starts with a disclaimer: *«The canton makes every effort to ensure the secure operation of the website. Nevertheless, defects such as data loss or falsification, virus attacks, operational interruptions, etc. cannot be ruled out. Access to the website is therefore at your own risk and responsibility»^[121]*. Which is then followed by a

exclusion of responsibility clause «*[t]he publications on the websites of the Canton of Basel-Landschaft are for informational purposes only. The Canton of Basel-Landschaft expressly reserves the right to change or delete all or part of the content at any time without notice, or to discontinue publication temporarily or permanently. Liability claims against the Canton of Basel-Landschaft for material or immaterial damage arising from access to or use or non-use of the published information or from technical faults are rejected*»^[122].

In contrast, the municipal police of Morges terms of uses don't contain a real "disclaimer" but informs the user on the unreliability and dangers of using the service: «*[i]n addition, [the association] endeavours to provide accurate and up-to-date information on its website. However, it cannot guarantee the accuracy, reliability or completeness of the information contained therein or referred to therein. Furthermore, system failures such as interruptions, virus attacks or data loss cannot be ruled out*»^[123]. The disclaimers convey multiple messages; first the unreliable nature of the information conveyed through the chatbot and the risk for the system to make errors and provide false advice^[124]. This indicates that the authorities are aware that the risks of the systems they implemented aren't trivial, to the point the users should "expect" them.

Second these texts inform the users that, because of these errors, the users shouldn't trust these texts and can't use these texts to make decisions. Since these chatbots were implemented specifically to give information to the users, it is rather surprising to see such warnings surrounding their use. The disclaimers convey the message that the authorities consider the individuals to be the sole responsible for any damage resulting from the use of their systems^[125].

Third, the exclusion of responsibility clauses implies that the authorities explicitly try to exclude the right of the users to hold the authorities accountable for the information their system conveyed or the damage their systems could have caused. If the occurrence of such harms isn't yet recorded on authorities' chatbots, it is not entirely impossible for such a chatbot to defame someone or to create harm to the personality of the users or of public figures.

To make an analogy, these disclaimers are like warning signs that would be put up on a critical piece of public service. Such warning signs are often times needed, to warn that a walk way could be dangerous^[126], or an old mountain source be untested and thus, not rated for human consumption. But e would be shocked if

the authorities spent money to provide more access to water on a public square, installed a new water fountain and publicised this event as a great development in the provision of such an essential public service, before putting warning signs saying that there are risks for people to poison themselves sometimes, and that each user should test the water before drinking. Information provided to the users by the administration has a real influence on their lives, missing a deadline or failing to fill the proper administrative paperwork could have real consequences for these persons. These authorities, by choosing to provide a “new” public service all while refusing to take any responsibility are at least sending the wrong signals about their trustworthiness and clearly underline the untrustworthiness of the technology they implement.

4.3 The legality of such disclaimers

In a private law systematic, where the website and the person consulting it are on an equal level, terms and conditions are generally considered to be of a contractual nature even if extra-contractual liability may arise from information available online^[127], these terms and conditions are part of the contract that is or will be passed between the user and the website to either access to content, to buy goods or services^[128]. With the recent developments of data protection legislation, most websites even have an explicit consent form for the individuals to accept the data processing made by the website and the use of cookies present on it. We should note that too broad disclaimers are generally forbidden in Swiss private law^[129] which forces private actors to set “reasonable” limitations in their responsibility.

The situation for public authorities is widely different. If a definitive analysis of this question is outside of the bound of the present study, we found no reasons for the relationship between the individuals and the authorities to be placed in the realm of private law, and thus no reason to consider these terms and conditions not to be of a public law nature.

The behaviour of the authorities when publishing or using software to provide public information and administrative services to the public are ruled by public law. Indeed, to determine if the matter is of a public or a private law nature, Swiss doctrine considers a multiplicity of criteria that, combined, assert the nature of

the relationship and of the acts adopted in it^[130]. In this case, public actors communicate with privates, to realise public tasks in a clear public interest. There is a clear subordination relationship, and the individuals are in no place to negotiate, the fact that the user disposes of a choice, to use the chatbot or not, doesn't change the relationship between the user and the state. Which means that these terms and conditions have to be analysed from a public law standpoint.

The principle of legality^[131] will in this sense be of the utmost importance to consider the legal standing of these texts. Such an analysis will need to be conducted for each system separately, which explains why we decided to leave it out of the present study. Indeed, the legality of each system will vary depending on the legal framework in place. The relevant laws should also be examined for the possible delegation of authority allowing the administration to enact such texts and the possible legal exceptions tailored for such systems.

It isn't simple to define which type of administrative act these terms and conditions are, to the point that it might be easier to state what these texts aren't.

The terms and conditions are abstract texts, presenting rules and principles applicable to all users, in this sense they can't be administrative decisions as they miss the needed individuality^[132]. These disclaimers are fixed for all users of the service; there is no individual aspect at play, and thus, no administrative decision. These terms and conditions can't be part of an administrative contract, because of the missing bilaterality of the agreement, of the effects and of the missing legal basis authorising such contracts to be passed^[133].

These texts aren't formal Laws or legislative ordinances since they weren't approved nor published according to the constitutional and legal obligations in place^[134].

In this sense, we believe these texts constitute administrative ordinances^[135], although these texts should theoretically only be addressed to agents of the administrations and shouldn't contain any rule of law^[136]. Since these texts aren't considered as rules of law, they shouldn't directly deploy any effect on the individuals^[137]. In a particular case, these texts would then only constitute hints about the way the authorities considered their information, but by elements permitting to infer the rights, or lack thereof, of the individuals.

If these terms and conditions, in the hierarchy of norms, really shouldn't be able to create any rights and obligations for the individuals, we argue that a judge

could consider these texts as void and without value for a specific case, thus reinstating the liability the authorities tried to avoid.

It is worth noting that previous studies were already conducted on the rather weak legal standing of such disclaimers in the protection of the trust principle the individuals should be able to put into advice given by the authorities chatbots^[138] and websites^[139].

We decided not to conduct such an analysis since providing definitive answer to this question would require, for each disclaimer an analysis of their wording as well as of the topical federal or cantonal laws. If the principle of the protection of trust in the information given by the authorities is present in multiple procedural laws, special provisions might be added to exempt certain chatbots or webpages from these provisions. The same is true for the clauses exempting the authorities from any liability for harm, each level of Government has put in place a specific legal framework around “state’s liability”^[140], if these laws don’t include exemptions for these systems, there is a high probability that the disclaimers and responsibility exemptions we analysed in this paper have little to no legal standing^[141].

4.4 The influence of these disclaimers on public trust

We argue that by implementing such disclaimers and pointing towards the unreliable nature of such technical systems, the authorities ask their users not to trust the system they implemented. If further studies are needed to evaluate the subjective impact of such terms and conditions on perceived trustworthiness of authorities, we can already affirm that these texts don’t comply with the usual standards of “good faith” actions^[142]. Nor with the federal recommendations on the use of AI by the authorities that specifically demand accountability^[143].

Having “terms and conditions” available with the chatbot itself is a positive element in the relationship between State and citizens. A specific information about the rights of the users, about their redress mechanisms are mechanisms of transparency that profits to the users. If the existence of privacy statement could have a positive impact on the trust of the users, the same effect might be true on the terms and conditions framing the information provided by such AI systems. Although data protection texts included links to the legal basis governing their

implementation, most disclaimers or terms and conditions about the content of the information provided by the chatbots, that we examined in this study presented no link of references to the applicable public laws^[144]. Although their legal standing is questionable at best, a non-experienced user might reasonably^[145] consider the clauses present in these terms and conditions to be binding.

Since most of these texts present clear disclaimers in favour of the authorities, because they affirm that the user is using the service at their own risks^[146]. A reasonable user might, rationally, consider these clauses to be an admission by the authorities of their untrustworthiness.

These broad disclaimers in favour of the authorities can't, from a rational point of view participate in creating trust in the eGovernment systems. Why use the chatbot function if «*False, incorrect and fictitious answers must be expected*»^[147]. Why use such a system if the terms and conditions ask the users to «*verify the information provided by [the chatbot] on the official websites*»^[148]. We believe the systematic use of disclaimers on Swiss public websites might convey the wrong message about the trustworthiness of State as a whole, and in particular about the eGovernment systems that the authorities implement.

The contrast between the content of the data protection rules and the responsibility for the content and answers provided by the chatbots is striking. The former present clear pathways for citizens to exert their rights or to ask questions about their personal data processing. These provisions don't include the right to ask for a review from the information given by the authorities themselves. Even in texts that mix both types of rules, the provisions for the rights of the users don't include any rights for the users concerning the texts produced by the chatbots^[149].

5. Possible pathways to foster or restore Trust in the testing and implementation of AI systems

Considering the critics we expressed in this study, we see multiple pathways that could be considered and combined to foster trust in the experimentation and implementation of AI technologies for public services.

First, future projects should be cleared up both from a technical point of view as well as from a legal perspective before such technologies are implemented^[150]. By

deciding to test unsafe technologies in a “production” environment, on a public facing website, the authorities take a risk to destroy trust from their users that might be difficult to gain back. Such prerequisites should in our view be generalised with internal and external audits to ensure that the solutions put in place are both safe and functional. As recently said in the examination of the future Swiss eID system, «*there will be no second chance to win the trust of the Swiss citizens*»^[151], we believe that this affirmation is applicable for every technological solution implemented by any authorities.

Second, authorities should develop clearly indicated and transparent “testing” framework for each new system. These tests would be clearly communicated and apparent for the users and be subject to specific regulation and testing requirements^[152]. For these tests, clear indications would be present on the systems themselves to differentiate these systems still undergoing testing from systems that are “in production”. Most of the chatbots we considered in this study are presented as fully functional piece of technology providing public services, the chatbot of the Swiss army indicates it’s prototype status in a separate popup, in the middle of a text on the goal of the service. An indication about the “testing phase” should be present from the start window.

A real testing framework, put in place on the government’s level the new algorithm is developed for, would have multiple advantages. Such systems could be implemented with ad hoc legislation that would define before the start of the experimentation both the duration of and the evaluation^[153] of the test. A legal framework would be put in place that both limit the risks and guarantee the accountability the authorities are taking upon themselves during the testing. A system being implemented for testing purposes should be clearly recognisable by the users, for example, with specific pages and warnings about the system. To better differentiate between testing and implemented technology, the chatbots still in the testing phase probably shouldn’t be present on the landing page of the authorities’ websites. Special mechanism would also be put in place to ensure a possibility for the users to signal errors and bugs, and to make sure their rights are guaranteed. In the case of a chatbot implemented by the authorities, we think for example of a button allowing in a simple way to ask for a human review of the advice given by the machine. The implementation of transparency and evaluation mechanisms would then ensure that the decision process to continue

the development to allow for a better implementation of such technologies.

Third, the authorities should take accountability for the systems they implement. From a legal and moral point of view, authorities have a duty to act in good faith^[154], to act in a trustworthy manner. This means that they shouldn't try to "hide" behind broad disclaimers and responsibility waivers. The validity and legal standing of such disclaimers being questionable at best, their implementation should really be put into question. The probable impact on public trust of such disclaimers far outweighs the protection the authorities could gain from them in their current redaction. The general principles surrounding the protection of trust already provide protection to the authorities for their actions on the web. Disclaimers should in this sense be reserved for systems still in their testing phase, with the necessary provision to ensure the rights of the users.

6. Conclusion

In this paper, we exemplified how the broad use disclaimer in terms and conditions by Swiss authorities in their implementation of AI chatbots is contrary to the duty of the authorities to act in a trustworthy manner. By implementing technology that they know is faulty, by saying that the service implemented shouldn't be trusted, and by refusing to take accountability for the answers given by these chatbots, the administrations are asking the citizens not to trust these systems.

In this sense, the administrations' practices surrounding their implementation of terms and conditions on their internet platforms are creating an appearance of untrustworthiness. Such an appearance creates unnecessary risks, both for future systems adoption and a risk for public backlash that could threaten a project that would be technically sound.

Since disclaimers might be unavoidable in certain testing situations, further research is needed to analyse the impact on the users of such disclaimers, their placement and the way they operate.

Although this point isn't central in the present study, the authorities should, in our opinion be more cautious and question the need to implement such technologies^[155]. AI technologies present whole new sets of challenges, many of which have not yet been solved. From ecological concerns^[156], to unclear social

impacts^[157], the unclear impact of AI should, in our opinion, warrant more caution in their implementation by public administration. It is always difficult to know what the future holds, uncertainties about the regulation of AI, the promises for the development of such systems and geopolitical tensions might have an impact on the generalisation of such systems. We encourage the authorities to review their use of these disclaimers and to accept to be accountable for their use of technologies and, thus, to take action to ensure that they continue to maintain a high level of public trust in their use and implementation of technologies.

1. This text was written by the author without the help of LLMs for the redaction or the brainstorming. DeepL was used to streamline the translations, and antidote was used to proofread the text. All links were checked on the 22.03.2026.
2. For a discussion on the matter, see for example E. Toreini, M. Aitken, K. Coopamootoo et. al., *The relationship between trust in AI and trustworthy machine learning technologies, Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, Barcelona Spain, in ACM, FAT* '20: Conference on Fairness, Accountability, and Transparency, 27 janvier 2020, p. 272-283; Also in a broader sense T. M. Lechterman, *The Concept of Accountability in AI Ethics and Governance*, in J.B. Bullock, Y. Chen, J. Himmelreich et. al. (dir.), *The Oxford Handbook of AI Governance*, Oxford University Press, Oxford, 2024.
3. For a discussion on the factors influencing trust in the relationship, see for example N. Schlicker, K.. Baum, A. Uhde et. al., *How do we assess the trustworthiness of AI? Introducing the trustworthiness assessment model (TrAM)*, in *Comput. Hum. Behav.* 170, 2025, p. 2 ss.
4. Multiple cues - information pieces - will be used to make an assessment about the trust to place in an agent E. J. De Visser, M. Cohen, A. Freedy et. al., *A Design Methodology for Trust Cue Calibration in Cognitive Agents*», in R. Shumaker, S. Lackey (dir.), *Virtual, Augmented and Mixed Reality. Designing and Developing Virtual and Augmented Environments*, Springer International Publishing, Cham, 2014, p. 251-262; a concept used by N. Schlicker et. al., *How do we assess the trustworthiness of AI?*, 3, 7 ss.
5. On the question of accountability as an ethical requirement in the implementation of AI, see for example D. Leslie, *Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector*, The Alan Turing Institute, 2019, p. 23 ss. the authorities in these disclaimers refuse to implement any «remedial accountability», on the topic p. 25 s. The focus is here on the accountability for their implementation of the systems, more than of the accountability that the systems can achieve, for a discussion of the concept of AI accountability and their use see for example M. Busuioc, *Accountable Artificial Intelligence: Holding Algorithms to Account* in *Public Adm. Rev.*, 81-5, 2021 pp.

6. Only partial lists exist which are sometimes not updated regularly. The list provided by Algorithm watch dates from 2023 *Atlas of Automation Switzerland, Algorithm Watch CH*, online on the website www.algorithmwatch.ch/en/atlas; and the list from the CNAI lists only systems implemented by the federal administration, *Projektsteckbriefe Kompetenznetzwerk CNAI*, Geschäftsstelle CNAI - OFS, 2026, online on the website www.bk.admin.ch.
7. This means purely internal chatbots are excluded from this study, for an internal AI system on the federal level see Office fédéral de l'informatique et de la télécommunication OFIT, *RoBIT apporte son soutien en cas de problèmes d'informatique*, online on the website www.bit.admin.ch.
8. This excludes the AI systems implemented by private entities that were trained on public data «*GenevaGPT*», online on the website www.geneva.gpt.liip.ch.
9. See in particular P. Glass, *Datenschutzrecht für künstliche Intelligenz in der öffentlichen Verwaltung*, in *Datenschutz Rechtliche Schnittstellen*, Dike Verlag, Zürich, 2023, p. 252 for chatbots.
10. The chatbot “Frag Esi”, online on the website <https://www.fragesi.ch>, totally changed its presentation at the end of 2025 and as of march 2026 gives an “error 403”; An archived version of november 2025 is still available at the internet archive, online on the website www.web.archive.org.
11. «*Esi – le chatbot de l'ASF*», online on the website www.esa.admin.ch.
12. For a deep reflection about trust, in particular in law see K. Sutter, *Vertrauen im Recht: eine Theorie für den demokratischen Verfassungsstaat*, Dike, Zürich, 2020 for the definitions of trust in particular p. 177 ss; or also M. Hartmann, *Die Praxis des Vertrauens*, Suhrkamp, Berlin, 2022, p. 20; for an analysis of the literature around the implementation of digital technologies by public authorities, see also L. Alzahrani, W. Al-Karaghoul, V. Weerakkody, *Analysing the critical factors influencing trust in e-government adoption from citizens' perspective: A systematic review and a conceptual framework*, in *Int. Bus. Rev.* 26-1, 2017, p. 165 ss.
13. R. C. Mayer, J. H. Davis, F. D. Schoorman, *An Integrative Model of Organizational Trust*, in *Acad. Manage. Rev.*, 20-3, 1995, p. 712.
14. We refer the reader to the large literature on the different meaning of trust, its impact and the factor that influence it, for example R. Campagna, A. Mislin, K. Dirks, H. A. Elfenbein, *The (Mostly) Robust Influence of Initial Trustworthiness Beliefs on Subsequent Behaviors and Perceptions*, in *SSRN Scholarly Paper*, Rochester, NY, Social Science Research Network, 18 février 2021; or D. J. Dekker, D. Krackhardt, P. H. Franses, *Dynamic Effects of Trust and Cognitive Social Structures on Information Transfer Relationships*, in *SSRN Scholarly Paper*, Rochester, NY, Social Science Research Network, 3 décembre 2002.
15. K. Sutter, *Vertrauen im Recht*, p. 411 ss.
16. On the tension between the authorities and individuals, see for example G. Malinverni, M. Hottelier, M. Hertig Randall et. al., *Droit constitutionnel suisse Vol. I: L'Etat, 4ème*,

- Stämpfli, Bern, 2021, p. 5 ss.; On the the regulation of the exercise public power by the interdiction of arbitrary actions see for example P. Moor, A. Flückiger, V. Martenet, *Droit administratif I* □: *les fondements*, vol. 1, 2012, p. 892 ss.
17. On the notion of legality see for example C. Grisel Rapin, *La légalité*, in *Les grands principes du droit administratif*, Schulthess, 2022, online on the website: www.archive-ouverte.unige.ch.
 18. On the difference and interdependencies of the two notions see for example J. Jackson, J. Gau, *Carving Up Concepts? Differentiating between Trust and Legitimacy in Public Attitudes Towards Legal Authority*, in *SSRN Scholarly Paper*, Rochester, NY, Social Science Research Network, 20 février 2015.
 19. For analysis on the notion in EU law, see for example J. Mendes, *Good Administration in EU Law and the European Code of Good Administrative Behaviour*, in *SSRN Scholarly Paper*, Rochester, NY, Social Science Research Network, 2009, p. 4 mentions trust explicitly; or J. Ponce, *The Right to Good Administration and the Role of Administrative Law In Promoting Good Government*, in *SSRN Scholarly Paper*, Rochester, NY, Social Science Research Network, 24 janvier 2016.
 21. For example, P. Rodríguez et al., «*How the exercise of the right to information (RTI) affects trust in political institutions*», online on the website www.sciencedirect.com.
 22. S. G. Gimmelikhuijsen, S. J. Piotrowski, G. G. Van Ryzin, *Latent transparency and trust in government: Unexpected findings from two survey experiments*, in *Gov. Inf. Q.*, 37- 1, 2020.
 23. K. Sutter, *Vertrauen im Recht*, p. 226 ss.
 24. On this notion F. Bernard, *La protection de la bonne foi*, in *Les grands principes du droit administratif*, Schulthess, 2022, online on the website www.archive-ouverte.unige.ch.
 25. *Constitution fédérale de la Confédération suisse (Cost.)*, du 18 avril 1999, RS 101.
 26. «*Treu und Glauben*»
 27. *Code civil suisse (CC)*, du 10 décembre 1907, RS 210.
 28. *Stratégie suisse de cyberadministration 2020–2023*, 2019, pp. 8273-8282.
 29. *Stratégie Utilisation de systèmes d'IA dans l'administration fédérale*, du 21 mars 2025, SB021, p. 6.
 30. Kantar, *eGovernment MONITOR 2023*, Initiative D21 e. V, Technical University of Munich, p. 15.
 31. *Ibid.*
 32. *Ibid.*; H. Amberg, C. Büchler, M. Buess, *Étude nationale sur la cyberadministration 2025. Rapport ciblé sur l'IA*, Lucerne, Adlingenswil, Interface Politikstudien Forschung Beratung AG et DemoSCOPE AG, 2025, p. 9 ss.
 33. The notion of “trustworthy AI” plays a key role as a justification for the AI Act of the EU an act that isn’t applicable to Switzerland *Regulation (EU) 2024,1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence*, du 13 juin 2024; for a commentary on the notion see for example J. Eichenhofer, *Trustworthy AI and Fundamental Rights*, in *Eur. Rev. Public Law*, 2024.

34. L. Alzahrani, W. Al-Karaghoul, V. Weerakkody, *Analysing the critical factors influencing trust in e-government adoption from citizens' perspective*, vol. 12; M. Z. I. Lallmahomed, N. Lallmahomed, G. M. Lallmahomed, *Factors influencing the adoption of e-Government services in Mauritius*, in *Telemat. Inform.* 34 – 4, 2017, p. 67; R. Fakhoury, B. Aubert, *Citizenship, trust, and behavioural intentions to use public e-services: The case of Lebanon*, in *Int. J. Inf. Manag.* 35-3, 2015, p. 348.
35. See for example S. W. Palmer, *The Role of Trust in Facilitating Human-Artificial Intelligence Relationships*, Aarhus University, 2024.
36. For such a study, see for example on the topic of “health data spaces” P. Daniore, F. Zavattaro, F. Gilles, *Public Trust in a Swiss Health Data Space*, Zürich, University of Zürich - Digital Society Initiative, 2024, online on the website www.health.dsi.uzh.ch.
37. In 2002, *Technological developments have a significant influence on the drivers of eGovernment* (free translation of 1st K. Schedler, L. Summermatter, *Was treibt das eGovernment?*, in D. Spahni (dir.), *eGovernment 2: Perspektiven und Prognosen*, P. Haupt, Bern. 2002 p. 120; The expectation from the users still support this orientation *eGovernment MONITOR 2025*, Initiative D21 e. V, Technical University of Munich, p. 14.
38. As an example, the federal administration established a report in 1979 on the electronic data processing in the federal administration *Bericht über den Einsatz der elektronischen Datenverarbeitung in der Bundesverwaltung*, Berne, Office federal de l'organisation, 1979.
39. Digital first is at the center of the vision for the Swiss digital strategy 2026 *Stratégie Suisse numérique 2026*, du 12 décembre 2025, 39, p. 2.
40. L. Stenberg, S. Nilsson, *Factors influencing readiness of adopting AI: A qualitative study of how the TOE framework applies to AI adoption in governmental authorities*, 2020, p. 55.
41. The “vision” of the federal administration centers around optimisation, helping the employees and developing “new” solutions *Stratégie Utilisation de systèmes d'IA dans l'administration fédérale*, vol. 29, p. 3; N. Braun Binder, M. Spielkamp, C. Egli et al., *Einsatz Künstlicher Intelligenz in der Verwaltung: rechtliche und ethische Fragen*, 2021, p. 15.
42. L. Stenberg, S. Nilsson, *Factors influencing readiness of adopting AI*, vol. 40, p. 42 s.; N. Braun Binde et al., vol. 41, p. 16.
43. P. Ringeisen, A. Bertolosi-Lehr, L. Demaj, *Automatisierung und Digitalisierung in der öffentlichen Verwaltung: digitale Verwaltungsassistenten als neue Schnittstelle zwischen Bevölkerung und Gemeinwesen*, in *Yearb. Swiss Adm. Sci.* 9-1, 2018, p. 51 ss.
44. For an independent overview of automation projects in Switzerland see the Atlas of Algorithm watch, vol 6 it is possible to select only government authorities in the searchbox; For an actual list of AI project on the federal level, see the *AI Projektdatenbank* of the CNAI 01.26, vol. 6.
45. With organisational performance, these systems also search to improve the effectiveness of the State by delivering better public service O. Neumann, K. Guirguis, R. Steiner, *Exploring artificial intelligence adoption in public organizations: a comparative case study*,

- in *Public Manag. Rev.*, 2022, p. 5; This point is even a domain of action in the last federal strategy *Stratégie Utilisation de systèmes d'IA dans l'administration fédérale*, vol. 29, p. 7.
46. N. Braun Binder et al., vol. 41, p. 16 ss.
 47. For a summary of the expectations of public administration see *Des algorithmes dans les administrations publiques* 30 mars 2021, online on the website www.ictjournal.ch.
 48. Although fraud detection doesn't need AI see A. Bănărescu, *Detecting and Preventing Fraud with Data Analytics*, in *Procedia Econ. Finance*, 32, 2015, pp. 1827-1836; The developments of such AI systems for tax purposes in Switzerland in this domain seem to have only begun *Automatisation et systèmes de gestion des risques au sein des administrations fiscales cantonales – swissprivacy.law*, online on the website www.swissprivacy.law, M. Nufer, *Künstliche Intelligenz in der Steuerveranlagung*, in *ASA*, 2019, p. 88; For a french outlook see M. Wolf, S. Dale, *Les technologies numériques au service de l'administration de l'impôt*, in *ERA Forum*, 19-3, 2019, pp. 457-464.
 49. In a field in constant evolution, it can be hard for the scholar to frame a subject at a given point, see for example W. G. de Sousa, E. R. P. de Melo, P. H. De Souza Bermejo et al., *How and where is artificial intelligence in the public sector going? A literature review and research agenda*, in *Gov. Inf. Q.*, 36-4, 2019.
 50. On the federal level for example there are at least 10 chatbots projects underway as of 2026 *AI Projektdatenbank 01.26*, vol. 6, p. 7, 17, 48, 49, 69, 107, 114, 118, 119, 121.
 51. For example, of uses of chatbots worldwide, see for example Y. Wang, N. Zhang, X. Zhao, *Understanding the Determinants in the Different Government AI Adoption Stages: Evidence of Local Government Chatbots in China*, in *Soc. Sci. Comput. Rev.*, 40-2, 2022, p. 537.
 52. O. Neumann, K. Guirguis, R. Steiner, *Exploring artificial intelligence adoption in public organizations*, vol. 45, p. 18.
 53. Chatbots (general) in Public Administration, *Algorithm Watch CH*, online on the website www.algorithmwatch.ch.
 54. P. Ringeisen, A. Bertolosi-Lehr, L. Demaj, *Automatisierung und Digitalisierung in der öffentlichen Verwaltung*, vol. 43, p. 54.
 55. H. Amberg, C. Büchler, M. Buess, vol. 32, p. 10.
 56. On this see for example V. Habib Lantyer, *Hallucination by Design: The Hidden Incentives of AI*, in *SSRN Scholarly Paper*, Rochester, NY, Social Science Research Network, 8 septembre 2025.
 57. S. Wachter, B. Mittelstadt, C. Russell, *Do large language models have a legal duty to tell the truth?*, (in *R. Soc. Open Sci.*, 11-8, 2024).
 58. See T. Poledna, *Staatliche Informationspflichten (Grundversorgung mit elektronischen Daten)*, in *Tag. 2003 Für Inform. Recht Vom 29 August 2003 Bern Journ. 2003 Inform. Jurid. 29 Août 2003 À Berne*.
 59. Federal Constitution (Cst.) of the 18 April 1999, RS 101.
 60. Of the 21 March 1997, RS 172.010.
 61. See for example for canton of Geneva *Loi [genevoise] sur l'information du public, l'accès*

- aux documents et la protection des données personnelles (LIPAD)*, du 5 octobre 2001, RS GE A 2 08.
62. Of 17 December 2004, RS 152.3.
 63. For an analysis of these regulations see V. Mabillard, M. Pasquier, *Les lois d'accès à l'information en Suisse. Mise en perspective internationale*, 2022, pp. 325-344.
 64. T. Poledna, vol. 58, p. 88.
 65. The project started in 2022 and is offline as of 2026 *RAV Baselland startet einen Chatbot-Pilot - Kanton Basel-Landschaft*, online on the website www.baselland.ch.
 66. Loi fédérale sur l'assurance-chômage obligatoire et l'indemnité en cas d'insolvabilité (Loi sur l'assurance-chômage, LACI), of the 25 June 1982, RS 837.0. The federal law charges the cantons of the execution of the proceedings in matter of unemployment. In particular, the registration of unemployed people (art. 17 al. 2 LACI and art. 85 LACI).
 67. Gesetz über die Arbeitsvermittlung und die Arbeitslosenversicherung (AVLG), of the 25.03.1999, RS-BL 837.
 68. IBM, *What is a chatbot*, in *IBM explainers blog*, online on the website www.ibm.com.
 69. Although the terminology "digital administrative assistant" is sometime used to describe conversational agents, see P. Ringeisen et al., *Automatisierung und Digitalisierung in der öffentlichen Verwaltung*, vol. 43, p. 57 ss.; These types of hopes are already mentioned in descriptions of future project on the federal level, see for example the CFC- Copilot , *Suisse numérique*, online on the website www.digital.swiss.
 70. Such guided forms can be found on numerous e-government proceedings. For example in order for a person to announce their address changes to the authorities via www.eumzug.swiss, to create a new company on easygov visit the website www.easygov.swiss. Similar forms are available on the cantonal and communal level and are, to our knowledge always considered as eGovernment forms, more than chatbots.
 71. This is the case for Alva form canton Basel Stadt and the BarBot of the federal archive (that indicates using GPT-4o).
 72. AI Chatbot, *Deeplink.ai* (16 février 2026), online on the website www.deeplink.ai.
 73. Due to a lack of public informations, see the publicity from the enterprise that installed the solution "Voicebot from UMB", online on the website www.umb.ch; their system uses the vocal solution from "Plateforme Conversationnelle Omnicanale Spitch", *Spitch*, online on the website www.spitch.ai.
 74. Amt für Migration, Integration und Bürgerrecht - Kanton Basel-Landschaft, online on the website www.baselland.ch.
 75. *Datenschutzerklärung Chatbot - KI-Bot Basel-Landschaft*, du 25 avril 2025.
 76. Disclaimer, *baselland.ch*, online on the website www.baselland.ch.
 77. Conditions générales d'utilisation - Police Region Morges, *Conditions générales d'utilisation - Police Region Morges*, online on the website www.prm-vd.ch.
 78. The chatbot is available at the bottom of the webpages and on it's dedicated page Alva, Kanton Basel-Stadt, online on the website, www.bs.ch.
 79. The link is called «*datenschutzbestimmungen*» and point to *Datenschutzerklärung*,

- Kanton Basel-Stadt* (16 mai 2024), online on the website www.bs.ch.
80. the chatbot of the Swiss army, online on the website www.armee.ch. In the infobox on the right corner of the chat has one “disclaimer” (*Clause de non responsabilité*) and one document about the “*Information about the use of the chatbot*” that contains rules about the timeline of the suppression of the metadata. From a historical standpoint, the FragEsi chatbot [V.n. 10] had two separate documents one named «*data protection*», and one «*terms and conditions*».
 81. Archives fédérales suisses AFS, online on the website www.bar.admin.ch.
 82. Archives fédérales suisses AFS, *Clause de non-responsabilité, déclaration de protection des données et conditions d'utilisation du chatbot*, that can be downloaded online on the website www.bar.admin.ch.
 83. *Idem*, p. 5 which contains a link to the «*Conditions of use and data protection policy*», online on the website www.recherche.bar.admin.ch.
 84. «*Terms and conditions*», online on the website www.admin.ch.
 85. The chatbot is available only with a Swiss internet connection on the bottom right corner of the page, online on the website www.armee.ch.
 86. Free translation of «*Clause de non responsabilité*».
 87. Free translation of «*Mentions légales*», the popup contains indications about the authority in charge of the chatbot and means to contact them (address, email and telephone).
 88. The HofdererBot present with a button with the logo of the municipality at the bottom left corner *Home – Gemeinde Hochdorf*, 19 février 2026, online on the website www.hochdorf.ch.
 89. Section about the Integration of third-party services and content, *Einbindung von Diensten und Inhalten Dritter*, in *Datenschutzerklärung*, online on the website www.hochdorf.ch.
 90. The link to this text is present after the name of the company, but the document is hosted on the municipality's website *Datenschutzrichtlinie_für_den_Chatbot_der_CAUSE_Systems*, online on the website www.hochdorf.ch.
 91. *Voicebot UMB*, vol. 73.
 92. Stadt Kreuzlingen, *Einwohneramt*, online on the website www.kreuzlingen.ch.
 93. Stadt Kreuzlingen, *Datenschutzerklärung*, online on the website <https://www.kreuzlingen.ch/datenschutzerklaerung>.
 94. *Voicebot UMB*, vol. 73.
 95. Free translation of the indication present on the chatbot before it's "activation" by the user «*Wenn Sie mich aktivieren, unterstütze ich Sie bei der Suche nach den richtigen Informationen*», Startseite, *Kanton Basel-Stadt*, online on the website <https://www.bs.ch>.
 96. Free translation of the first messages of the chatbot «*je vais vous accompagner dans la recherche d'informations sur nos services*» in *Accueil - Police Region Morges*, online on the website www.prm-vd.ch.

97. Free translation of «*Comment puis-je vous aider?*».
98. Free translation of «*Es kann sein, dass einzelne Informationen nicht korrekt sind. Bitte überprüfen Sie wichtige Informationen*», text of the chatbot present on the webpage of the federal archives Archives fédérales suisses AFS, vol. 84.
99. Free translation of «*le chatbot de l'armée suisse peut commettre des erreurs. Vérifiez les informations importantes*».
100. Free translation of «*Meine Antworten werden von einer künstlichen Intelligenz generiert und sind deshalb nicht immer korrekt*», vol. 98.
101. Free translation of «*je suis à mes débuts et j'apprends tous les jours*».
102. Free translation of «[a]ll information provided without guarantee» at the bottom of the chat window, vol. 91.
103. Free translation of «*[l]a version actuelle est un prototype et sert à effectuer d'autres tests utilisateurs*».
104. Kanton Basel Stadt migration office chatbot.
105. Bar Bot of the federal archives.
106. Alva.
107. So the Swiss army and Morges municipal police chatbot.
108. *Data protection Basel Stadt*, vol. 79.
109. Conditions générales chatbot archives, ,vol. 85, p. 2.
110. Clause 5 indicates an assistance and contact point Conditions générales chatbot archives, p. 5; The Municipal police chatbot includes a contact email in the first messages, vol. 99; While Alva includes an email adress for «*further information of questions about these terms and conditions*»□: «*weitere Informationen oder Fragen zu diesen Nutzungsbedingungen*» (n.8 of the KI ALVA terms) *Data protection Basel Stadt*, vol. 79 In contrast, the general disclaimer of the Kanton Basel Stadt gives no contact informations *baselland.ch>disclaimer*, vol. 76.
111. See for example the numerous links to legal texts in the *AFS data protection policy*, vol. 86.
112. Free translation of «*[d]er Kanton Basel-Stadt übernimmt keine Gewähr für die Richtigkeit, Vollständigkeit oder Aktualität der von Alva zu Verfügung gestellten Informationen. Die Informationen dienen lediglich zu allgemeinen Informationszwecken und sind rechtlich nicht bindend. Wir bitten Sie, die von Alva zu Verfügung gestellten Informationen auf den offiziellen Websites des Kantons zu überprüfen*» *Data protection Basel Stadt*, vol. 79.
113. «*Gewährleistungsausschluss*», *Idem*.
114. Free translation of «*Der Kanton verwendet all seine Sorgfalt darauf, dass die Informationen auf den Websites zum Zeitpunkt der Veröffentlichung korrekt und aktuell sind und keine Rechte Dritter (insbesondere Rechte des Persönlichkeitsschutzes) verletzen. Dennoch können weder der Kanton noch einzelne Ämter oder Dienststellen eine Gewährleistung hinsichtlich der Genauigkeit, Zuverlässigkeit oder Vollständigkeit der Information machen*», *Idem*.
115. Free translation of point 2 of the conditions for Alva: «*Alva basiert auf künstlicher*

- Intelligenz und kann dabei Fehler machen oder ungenaue Informationen bereitstellen. Insbesondere können Antworten korrekt erscheinen, inhaltlich jedoch falsch sein. Für die Folgen, die aus der Nutzung der von Alva bereitgestellten Informationen resultieren, wird vom Kanton Basel-Stadt keine Haftung übernommen», Idem.*
116. Free translation of the exclusion of liability clause «*Der Kanton lehnt jegliche Haftung für Schäden oder Folgeschäden ab, die sich aus dem Zugriff auf seine Websites beziehungsweise auf einzelne Teile davon (wie z.B. heruntergeladene Dokumente) deren Benutzung (bzw. aus der Unmöglichkeit des Zugriffs oder der Benutzung) oder Links auf andere Websites ergeben» Idem.*
 117. Free translation of «*L'utilisation du chatbot et ses éventuelles conséquences relèvent de votre responsabilité. Les Archives fédérales ne garantissent en aucun cas la véracité, l'exactitude, l'exhaustivité et la fiabilité des réponses fournies par le robot. Des réponses fausses, inexactes ou inventées ne sont pas exclues. Vous devez toujours vérifier les informations et les sources indiquées. Les Archives fédérales ne sont en aucun cas responsables des actions entreprises par les utilisateurs ou de leur décision de ne pas agir sur la base des informations fournies par le chatbot, ni de leurs quelconques conséquences. Toute prétention en matière de responsabilité est expressément exclue*», Terms and conditions of BarBot, colo. 85, p. 2.
 118. Free translation of «*le chatbot est un interlocuteur central pour les militaires. Il traite sans délai les questions standard et transfère les requêtes plus complexes aux services compétents*» «*Sur armee.ch, un chatbot répond aux questions 24h, 24 et 7j,7*», online on the website <https://www.vtg.admin.ch>.
 119. Free translation of «*Le chatbot de l'armée suisse a été développé pour fournir des informations et des renseignements généraux sur l'Armée Suisse*», the disclaimer of the Swiss army chatbot seems directly inspired from the now defunct FragEsi chatbot that had a very similarly worded introduction: «*Esi wurde entwickelt, um allgemeine rechtliche Informationen und Auskünfte rund um die ESA bzw. das Schweizer Stiftungsrecht bereitzustellen*».
 120. Free translation of «*L'utilisation du chatbot se fait à vos propres risques et périls. L'Armée suisse ne garantit pas l'exactitude, la précision, l'exhaustivité et la fiabilité des réponses fournies. Il faut s'attendre à des réponses fausses, erronées ou inventées de toutes pièces. Vérifiez chaque réponse, y compris les éventuelles sources indiquées. L'Armée suisse n'est pas responsable si l'utilisateur ou l'utilisatrice entreprend ou s'abstient d'entreprendre des actions en se fiant à l'exactitude et à l'exhaustivité des réponses et qu'il en résulte des inconvénients ou des dommages. Toute prétention en responsabilité causée par l'utilisation est expressément exclue*» infobox «*Clause de non-responsabilité*», on the chatbot of the Swiss army website; this disclaimer uses the same exact wording of the now defunct Fragesi Chatbot (with the exception of the name of the authorities).
 121. Free translation of «*Der Kanton unternimmt alle Anstrengungen, um einen sicheren Betrieb der Webseite sicherzustellen. Dennoch können Mängel wie z.B. Datenverlust oder - verfälschung, Virenbefall, Betriebsunterbruch u. a. nicht ausgeschlossen werden. Der Zugriff auf die Webseite erfolgt deshalb auf eigene Gefahr und in eigener Verantwortung*».

122. Free translation of «*Die Publikationen auf den Webseiten des Kantons Basel-Landschaft weisen lediglich Informationscharakter auf. Der Kanton Basel-Landschaft behält sich ausdrücklich vor, Inhalte jederzeit ohne Ankündigung ganz oder teilweise zu ändern, zu löschen oder die Veröffentlichung zeitweise oder endgültig einzustellen. Haftungsansprüche gegen den Kanton Basel-Landschaft wegen Schäden materieller oder immaterieller Art, welche aus dem Zugriff auf die oder der Nutzung bzw. Nichtnutzung der veröffentlichten Informationen oder durch technische Störungen entstanden sind, werden abgelehnt*», Idem.
123. Free translation of «*[d]e plus, [l'association] s'efforce de donner des informations exactes et actualisées sur son site. Néanmoins, elle ne peut garantir l'exactitude, la fiabilité et l'exhaustivité des informations qui y figurent ou auxquelles il est fait référence. En outre, des défaillances du système telles que des interruptions, des attaques par des virus ou des pertes de données ne peuvent pas être exclues*» *Conditions d'utilisation - police Morges*, vol. 64.
124. It is worth noting that these unreliabilities already made their way to the press *À Morges, l'IA de la police locale affirme qu'on peut rouler à 58 km/h sans amende, in 24 heures*, 3 septembre 2025, online on the website www.24heures.ch.
125. On this point in US law, see for example L. Y. G. Tenzer, *Defamation in the Age of Artificial Intelligence*, in *SSRN Scholarly Paper*, Rochester, NY, Social Science Research Network, 18 août 2023.
126. The authorities have a certain responsibility to inform the public and to close off dangerous ways, on the matter see G. Seferovic, *Die Haftung des Gemeinwesens für Schäden durch Naturgefahren auf Wanderwegen zwischen Werkeigentümer- und Staatshaftung*, in *Sicherheit & Recht*, 1-2018, p. 54 ss.
127. R. G. Briner, *Haftung für Informationen auf Websites*, in *Sic!*, 2002, p. 236 ss.
128. For special types of terms and conditions on the internet, see M. Dill, *Internet-Verträge*, in *Aktuelle Juristische Prax*, 2000, p. 1513 ss; For online platforms see for example A.-C. Fornage, *Les plateformes en ligne et la protection du consommateur ... vers un changement de paradigme*, in Pichonnaz, Werro (dir.), *La place du consommateur au quotidien*, Schulthess, Genève - Zurich - Bâle, 2022, p. 165-195, online on the website www.iris.unil.ch.
129. Briner, *Sic!*, 2002 p. 231, vol. 130.
130. P. Moor et al., *Droit Administratif I*, vol. 16. p. 95.
131. G. Malinverni et al., *DCS I*, vol. 16, p. 1812 ss; See also C. Grisel Rapin, *La légalité*, vol. 17.
132. P. Moor, E. Poltier, *Les actes administratifs et leur contrôle, Entièrement revue, mise à jour et augmentée*, coll. Droit administratif Vol. 2, Stämpfli, Berne, 2011, p. 12.
133. *Ibid.*, p. 415 ss.
134. G. Malinverni et al., *DCS I*, vol. 16, p. 1532 ss.
135. For a general presentation see for example *Ibid.*, p.1670 ss.
136. *Ibid.*, p. 1674.
137. P. Moor et al., *Droit Administratif I*, vol. 16, p. 420 ss.
138. N. Braun Binder, L. Obrecht, G. Wittmer, *Vertrauensschutz bei fehlerhaften*

- Behördenauskünften durch Chatbots*, in *IusNet Digit, Recht Datenrecht*, 2024.
139. J. Klemperer, P. Hauenstein, *Vertrauensschutz bei behördlichem Webauftritt*, in *Aktuelle Juristische Prax.* 2023, pp 871-880.
 140. On the federal level, the *Loi fédérale du 14 mars 1958 sur la responsabilité de la Confédération, des membres de ses autorités et de ses fonctionnaires (Loi sur la responsabilité, LRCF)*, du 14 mars 1958, RS 170.32.
 141. According to F. Uhlmann, J. Kaspar, *Schweizerisches Staatshaftungsrecht*, Dike, Zürich, 2017, p.149 there should be an exam for each disclaimer if applicable laws allow for private agreement.
 142. J. Klemperer, P. Hauenstein, vol. 142.
 143. The fourth line of action reads: «*In order to determine and establish liability in the event of damage, accident or breach of law, it is necessary to clearly establish responsibility when using AI*» (free translation) in *Intelligence artificielle – lignes directrices pour la Confédération*, 2020, online on the website www.sbfi.admin.ch., p. 5.
 144. Since the law is both at the basis and the limit of the authorities' actions (art. 5 al. 1 Cst.), referencing the applicable legal basis for a public activity is at least a "good practice" since it allows users to confirm the information they are given. The presence of such texts could also empower the users to exercise their rights.
 145. Because of the trust he has in the authorities to follow the core constitutional principles of legality and good faith (art. 5 al. 1 and al. 3 Cst.).
 146. Baselland.Ch>Disclaimer, *Der Zugriff auf die Webseite erfolgt deshalb auf eigene Gefahr und in eigener Verantwortung*, vol. 76.
 147. Haftungsausschluss *armee.ch*, vol.123.
 148. Free translation from the point 2 of the terms of uses of the KI Assistent Alva, *Data protection Basel Stadt*, vol. 79.
 149. For example only the the data protection statement includes a section «your rights» (vos droits) *Terms and Conditions* of BarBot, vol. 85, p. 4.
 150. We note that this point is part of the guiding principles of the federal strategy for the implementation of AI in the federal administration *Stratégie Utilisation de systèmes d'IA dans l'administration fédérale*, vol. 29, p. 3.
 151. *Zur Erlangung des Vertrauens der Schweizer Bevölkerung in diese vom Bund bereitgestellte Lösung wird es keine zweite Chance geben*, in *Prüfung des Schlüsselprojektes*, Eidgenössische Finanzkontrolle (EFZ), Bern, 2025, p. 25.
 152. For and example of such testing framework, see for example the way the eVoting systems are tested – for twenty years in Switzerland, with both ad hoc legislation and constant evaluations see the dossier about the trials online on the website www.bk.admin.ch.
 153. On the federal level, a framework for pilot projects is envisioned, which doesn't include provisions in favor of the users, see art. 15 *Loi fédérale sur l'utilisation des moyens électroniques pour l'exécution des tâches des autorités (LMETA)*, du 17 mars 2023, RS 172.019; For legistical consideration see for example A. Flückiger, *(RE)FAIRE LA LOI: traité de légistique à l'ère du droit souple*, Stämpfli Éditions, Genève, 2018, p. 645 ss., for

evaluation of legislation, and 653 ss for experimental legislation in particular. For a specific approach on sandboxes see for example S. Volz, *KI Sandboxen für die Schweiz?*, in *SZW RSDA* 1,2022, p. 51.

154. For similar recommendations, see Amberg, Bächler, Buess, vol. 32, p. 14.
155. We see similar call for caution for example in Amberg, Bächler, Buess, vol. 32, p. 14.
156. See for a study on the environmental impact, for example T. Le Goff, *Recommandations pour une action publique en faveur d'une IA générative respectueuse de l'environnement*, 2023, p. 4 ss.
157. For a discussion on the different cost of the implementation of AI see D. Acemoglu, *Harms of AI*, in J.B. Bullock, Y. Chen, J. Himmelreich et al. (dir.), *The Oxford Handbook of AI Governance*, Oxford University Press, Oxford, 2024, p. 0.

Automated Administrative Sanctioning and the Rule of Law

Oscar Capdeferro Villagrasa

DOI: 10.13130/2723-9195/2026-2-11

Il presente contributo analizza la crescente automazione dei poteri sanzionatori delle pubbliche amministrazioni, con particolare attenzione all'impiego di sistemi basati sull'intelligenza artificiale per l'individuazione delle violazioni e l'irrogazione delle sanzioni. Muovendo da esempi empirici tratti dall'esperienza spagnola – in particolare dall'applicazione automatizzata delle zone a basse emissioni a Barcellona e dalle procedure di ispezione del lavoro – il lavoro esamina le implicazioni di tali pratiche per i principi fondamentali del diritto amministrativo. La tesi che si sostiene è che gli attuali quadri normativi non affrontino adeguatamente i rischi specifici connessi all'automazione dell'esercizio del potere sanzionatorio, quali i falsi positivi, la rigidità nell'applicazione delle sanzioni e l'erosione delle garanzie procedurali. Si propone perciò lo sviluppo di strumenti di tutela giuridica che siano specificamente calibrati ed idonei a conciliare l'efficienza tecnologica con le esigenze fondamentali dello Stato di diritto e la tutela dei diritti dei singoli.

This paper analyses the growing automation of administrative sanctioning powers in public administrations, with particular attention to the use of AI-driven systems for detecting legal infringements and issuing penalties. Drawing on empirical examples from Spain – especially the automated enforcement of low-emission zones in Barcelona and labour inspection procedures – it examines the implications of these practices for fundamental principles of administrative law. The paper argues that current regulatory frameworks insufficiently address the specific risks posed by automated sanctioning, such as false positives, rigidification of penalties, and erosion of procedural guarantees. It therefore advocates the development of tailored legal safeguards capable of reconciling technological efficiency with the core requirements of the rule of law and the protection of individual rights.

Summary: 1. Introduction.- 2. Two Spanish Rules in Force Allowing Automated Controls and Penalty Proposals.- 2.1. Barcelona's Ordinances on Low-Emission Zone.- 2.2. Labour Inspectorate Procedures for Labour Law Violations.- 3. New Law for a New Era: Moving Forward or Moving Backwards?- 4. Automation and AI: the regulatory framework and the need for safeguards in automated punishment systems.- 5. Concluding Remarks: Automation, Sanctioning Power and the Future of Administrative Law.- 5.1. The growing tension between efficiency-driven automation and classical legal safeguards.- 5.2. False positives and mass enforcement: the problem of systemic error.- 5.3. Fragmentation and insufficiency of current regulatory frameworks.

1. Introduction

The automation of tasks and processes, specially fostered by artificial intelligence, have expanded rapidly across the public sector. For years, public administrations have successfully deployed chatbots to provide information and guidance to citizens^[1], adopted automated systems to detect and punish traffic infringements^[2], or used algorithmic tools for fraud detection^[3], to name only some of the most paradigmatic and widely documented examples.

Moreover, the European Commission's comprehensive assessment of 940 AI use cases across Europe illustrates the breadth and maturity of these deployments^[4].

This finding is also reinforced by *Grimmelikhuijsen & Tangi*^[5], whose study draws on insights from over 500 public managers across various European countries. Their results highlight a paradigm shift whereby AI is no longer perceived as merely an aspirational tool but rather as an operative component within contemporary public services, increasingly integrated across a wide range of activities.

Within this broader trend, automation has begun to transform one of the most sensitive spheres of public administration powers: the monitoring of legal compliance and the imposition of administrative penalties. Using a variety of technical tools – video surveillance systems, speed radars, environmental sensors, automated number plate recognition^[6], large-scale data analytics, and algorithmic decision systems^[7] – administrations can now conduct continuous monitoring,

automatically verify potential infringements, and even issue sanctioning decisions – or penalties proposals – without human involvement. This transformation has been associated with the rise of Smart City infrastructures, leading Sharp to coin the term “smart administrative punishment”^[8].

Empirical evidence reveals how widespread these practices have become. Examples include the automated control of polluting vehicles entering Barcelona’s low-emission zone^[9]; algorithmic infringement-detection systems within Spain’s Labour Inspectorate^[10]; automated speeding sanctions in many countries but significantly in France^[11]; and successive legislative reforms in Lithuania since 2018 expanding automated administrative punishment across traffic and tax domains^[12]. A review of 250 EU cases by Van Noordt & Misuraca^[13] further documents computer-vision systems used in the Netherlands and Belgium to detect mobile phone use while driving.

Automated detection and punishment systems appear to improve rule compliance, as they address traditional administrative limitations such as limited human monitoring capacity or the low perceived likelihood of punishment^[14].

This paper examines the legal implications of this transformation in the functioning of public administrations, whose operations are becoming increasingly automated, particularly with regard to their supervisory and sanctioning activities.

In this context, a preliminary clarification is necessary. It should be noted that this study will focus exclusively on the administrative sanctioning powers, understood as the authority of public administrations to punish breaches of (administrative) law that are classified as administrative infringements, rather than criminal offences. Accordingly, the paper only reflects on the legal and regulatory framework governing administrative sanctions or penalties, most commonly imposed in the form of fines by public officials or administrative bodies, as opposed to criminal penalties imposed by judicial authorities in response to the commission of a crime.

According to the Council of Europe Recommendation No R(91)1 of the Committee of Ministers to member states on administrative sanctions, those can be defined as «*administrative acts which impose a penalty on persons on account of conduct contrary to the applicable rules, be it a fine or any punitive measure, whether pecuniary or not*». As for context, the same recommendation states that

«administrative authorities enjoy considerable powers of sanction as a result of the growth of the administrative state as well as a result of a marked tendency towards decriminalisation».

Naturally, interactions may arise between criminal and administrative sanctioning powers. The most paradigmatic case of concern in regard of these interactions is overlapping the sanctioning powers, raising the risk of punishment through both administrative and criminal channels^[15]. Undoubtedly, an automated system of administrative surveillance may detect conduct amounting to a criminal offence. In such cases, however, the administrative inspectorate will refer the file to the Public Prosecutor's Office or to the competent judicial authority, and any subsequent investigation and, where appropriate, sanctioning will be conducted in accordance with the safeguards and rules of criminal procedure, avoiding the risk of overlapping penalties, according to the appropriate national or regional rules.

Nevertheless, the focus of this article lies exclusively on the manner in which public administrations are automating both the detection of legal infringements and the imposition of sanctions, leaving aside the interactions with criminal investigations and sanctions, since in those cases the regulatory framework would be different. Accordingly, the instruments and procedures discussed will be limited to those operating within the administrative sphere of the State's exercise of its *ius puniendi*.

With these considerations in mind, the paper begins by illustrating how, in Spain, the automation of administrative supervisory and enforcement activities is already envisaged within the legal system. While a general regulatory framework governing automated administrative action exists, certain specific areas are subject to more detailed rules that have significantly developed the regulation of automated control functions. Two such examples are the Barcelona's Metropolitan Area ordinances on low-emission zones and the Labour Inspectorate procedures for labour law violations.

In both cases, AI-based digital systems are capable of autonomously identifying regulatory infringements and, without any human intervention, generating proposals for administrative sanctions, which are subsequently notified to the alleged offender.

After presenting these two paradigmatic examples of the growing automation

within Spanish public administrations in section 2, the paper reflects, in section 3, on the implications of a regulatory framework adapted to automated sanctioning practices, particularly in light of the rights and principles governing administrative procedures. Special attention is paid to those guarantees that operate within the context of the exercise of administrative sanctioning powers.

Section 4 focuses on presenting the argument that the current legal framework regulating AI and automated administrative action does not establish specific guarantees for the automation on the area of administrative sanctioning power, and yet, in view of experiences such as those in Spain described in section 2, it is necessary to move forward in regulating certain specific minimum safeguards for digital systems that are capable of identifying infringements and proposing appropriate administrative sanctions for such breaches. That regulatory proposal would be drafted under the basis of the traditional safeguards and guarantees that are at risk according to what will be presented in section 3.

Finally, section 5 reflects on the growing tension between efficiency-driven automation in public administration and legal safeguards, with some concluding remarks that highlight the need of further legal developments in order to deal with the problems inherent to technology in the use of administrative sanctioning powers.

In this paper, in order to establish a more neutral term, the word “penalty” will be the term preferred, but “sanction” and even “fine” are used.

The first two words are treated as synonyms, while fine will be used having into consideration that fines are, only, those sanctions consisting on the payment of an amount of money to the public administration.

2. Two Spanish Rules in Force Allowing Automated Controls and Penalty Proposals

This section addresses Spanish regulations at different levels that provide for fully automated actions to detect infringements and propose applicable penalties. Specifically, a selection has been made of two legal instruments: a municipal ordinance and a state law. Those legal instruments foresee the possibility of carrying out administrative control actions without human intervention.

In the two selected cases, a high degree of automation is clearly observed, unlike

other provisions in the Spanish legal system which, although regulating occasional automated functions, reserve a final decision-making role for the competent public official or authority. One such case could be, for example, the innovative Valencian Law 22/2018 of 6 November, on the General Inspection of Services and the Alert System for the Prevention of Misconduct, in which a regulation of a digital red-alert system is provided, but with a very limited automation degree.

In contrast, in the two areas presented below, the regulated actions can be fully automated, without human intervention, and may have legally binding sanctioning effects on infringing individuals. The two selected areas are the control of low-emission zones in Barcelona and the Labour Inspectorate procedures for labour law violations. These areas are currently regulated by quite recent rules that show the improvements and advancements in automating the administrative sanctioning powers, significantly regarding the detection of infringements and issuing sanctioning proposals. However, this automation race, as will be seen expressively in the case of the low-emission zone, involves risks.

2.1. Barcelona's Ordinances on Low-Emission Zone

Spanish Law 7/2021 of 20 May, on climate change and energy transition, establishes in Article 14 that municipalities meeting certain characteristics, which will be addressed below, were required to establish, before 2023, low-emission zones within their respective sustainable urban mobility plans, introducing mitigation measures aimed at reducing emissions arising from mobility.

The same Article 14 of Law 7/2021 defines low-emission zones in the following terms: *«the area delimited by a public authority, in the exercise of its powers, within its territory, of a continuous nature, in which restrictions on vehicle access, circulation, and parking are applied in order to improve air quality and mitigate greenhouse gas emissions, in accordance with the classification of vehicles by their emission levels as established in the current General Vehicle Regulations»*.

Pursuant to Article 14.3 of the aforementioned Law 7/2021, this obligation applies, as a general rule, to all municipalities with more than 50,000 inhabitants and to island territories. Exceptionally, it may also apply to municipalities with more than 20,000 inhabitants, but only when the limit values of the pollutants

regulated in Royal Decree 102/2011 of 28 January, on the improvement of air quality, are exceeded.

Some of the most densely populated areas in Spain, such as Madrid and the metropolitan area of Barcelona, had already begun to restrict the circulation of polluting vehicles in urban centers, establishing low-emission zones by means of municipal ordinances in 2019, anticipating what would later be imposed as a general obligation throughout Spain by the 2021 Law aforementioned.

Those initial ordinances were challenged, and, essentially, the courts ultimately annulled, in whole or in part, the local ordinances establishing the low-emission zones due to deficiencies and flaws in the reasoning or in the procedure followed (see the judgments of the High Courts of Justice of Madrid of 27 July 2020, ECLI:ES:TSJM:2020:9774 and ECLI:ES:TSJM:2020:8151; and of Catalonia of 21 March 2022, ECLI:ES:TSJCAT:2022:1576)^[16]. In what follows, references will be made exclusively to Barcelona, which is where the automated control system Chronos-Eco is used, to which I will refer later in this section.

Following that adverse judicial ruling from March 2022, the Barcelona City Council worked on a new ordinance, which was finally approved in 2023. This new 2023 Barcelona low-emission zone ordinance was also challenged before the High Court of Justice of Catalonia, which, in a judgment of 24 November 2025 (ECLI:ES:TSJCAT:2025:6428), dismissed the appeal against the Ordinance of the Barcelona City Council of 27 January 2023, which establishes the criteria for vehicle access, circulation, and parking within the Barcelona Low-Emission Zone and promotes zero-emission mobility, thereby upholding the legality of the new regulation.

Article 10 of the 2023 Ordinance establishes the prohibition on circulation within the low-emission zone on specific days and hours of the week: «1. *With the aim of improving air quality and mitigating greenhouse gas emissions, as well as mitigating the impacts of climate change and protecting public health, access to and circulation within the Barcelona low-emission zone by the most polluting vehicles is prohibited.* 2. *The measure established in section 1 applies from Monday to Friday on working days, between 07:00 and 20:00*». There is the possibility of obtaining occasional access authorizations, as well as justified situations in which access by polluting vehicles would be permitted, as detailed in other articles (Articles 12, 14, and 15).

For the purposes of this study, it is particularly relevant to refer to Article 16, as it provides the legal basis for the use of an automated system to monitor compliance with access restrictions, and to identify and sanction offenders. Its literal wording is as follows: *«Compliance with the provisions of this Ordinance shall be monitored through a license-plate recognition system and the City Council's technological platform using automated means, without prejudice to the powers assigned to the Urban Guard and in compliance with the applicable regulations on the capture and use of images and the protection of personal data»*.

In these respects, the 2019 and 2023 ordinances are equivalent, and the same automated technological control solution has been used in the metropolitan area of Barcelona from 2019 to the present. Therefore, despite the change in the regulatory framework, it can be said that the Chronos-Eco system has been monitoring access to the Barcelona low-emission zone since 2019.

The technical solution used to carry out this monitoring consists of cameras installed around the perimeter of the low-emission zone, connected to a system that processes the images in order to identify the license plates of vehicles entering the zone. These data are cross-checked against the database containing the traffic environmental labels assigned according to the pollution levels produced by the vehicle associated with that license plate, so as to determine automatically whether the obligation not to access the zone with a polluting vehicle is being complied with; and, in the event of non-compliance, to automatically validate the penalty by identifying the corresponding infringement and sanction, so that the proposal of penalty can be notified to the registered owner of the vehicle^[17].

So, what happens if an entry into the low-emission zone is detected by a vehicle that is prohibited from accessing it due to its high level of pollution? In this regard, the Ordinance establishes a system of infringements and penalties. Article 20 provides that *«failure to comply with the prohibition established in Article 10 constitutes a serious infringement, punishable by a fine of 200 euros»*. Furthermore, if air quality is worse, exceptionally, the penalty is set at 260 euros, in accordance with the literal wording of section 2 of Article 20: *«Where the infringement is committed during an air pollution episode declared by the competent authority of the Government of Catalonia, the amount of the fine shall be 260 euros»*.

These penalties are also supported by Articles 80 and 81 of Royal Legislative

Decree 6/2015 of 30 October, approving the consolidated text of the Law on Traffic, Motor Vehicle Circulation, and Road Safety, which generally provides for fixed-amount fines (200 euros, with a 30% increase in cases of greater severity).

Since its inception, the automated monitoring and enforcement system of the low-emission zone has been controversial. Thus, from the very early stages of its operation, it was common for proposed fines to be issued for access to the low-emission zone by polluting vehicles that had been detected by the cameras through license-plate recognition, even though they were not in breach of the access restriction regime, as the vehicle was being towed by a tow truck, on its way to a repair workshop or to a scrapyard. Therefore, the vehicle was not circulating within that area, polluting the air, since it was not being propelled by its own engine. At the time, in 2020, the press described these fines as anomalous (“insólita”) fines^[18].

This problematic situation, far from being resolved, appears to have persisted, which ultimately led, in 2023, to the Barcelona City Council publicly requesting tow truck operators to cover up license plates in order to avoid these erroneous and anomalous automated penalties.^[19] It is, however, striking that the burden of taking action to mitigate the shortcomings of the public administration’s automated systems has been shifted onto private operators (tow truck drivers), instead of adopting internal administrative measures to resolve the problem, such as assigning a public officer to manually review and verify all proposed sanctions. Moreover, a case of an erroneous automated fine was brought to the attention of the local ombudsman, the *Sindicatura de Greuges of Barcelona*, who resolved the citizen’s complaint in May 2024. In that case, a fine proposal was automatically issued to a car owner for an alleged infringement for driving within the low-emission zone with a polluting car, but, again, in the case it was the municipal tow truck that was transporting the vehicle within the zone.

This was the picture included in the fine proposal:



Source: Sindicatura de Greuges de Barcelona (2024).

As could be clearly seen, the polluting vehicle has no driver; it is being towed by a tow truck and, furthermore, if it was moving by itself, it would be travelling in the wrong direction, as indicated by the directional arrows painted on the road surface.

Throughout the entire process, from image capture to the issuance of the proposed sanction, as confirmed by the municipal Sindicatura de Greuges, no employee of the City Council viewed the captured images. As a result, the system amounted to fully automated enforcement, in which errors also occurred that would have been evident to a human reviewer had the images been examined. This situation, according to the Sindicatura de Greuges de Barcelona, is considered against the principles of good administration and effectiveness of public service.

To conclude this section, I would like to point out that this case, concerning the low-emission zone, may be paradigmatic in highlighting two aspects: the tendency toward uniformity – or the suppression, if not disregard, of the role of discretion – in the determination of sanctions, through fixed amounts that can be applied in a fully automated manner; and the risk of digital false positives, which, although potentially avoidable as they would be evident to a human observer or controller, are multiplied by the use of high-volume automated

systems (such as traffic surveillance) and could have widespread consequences. To prevent these problems, it would be necessary for the regulations to establish additional safeguards. Surprisingly, however, the legislation provides no such additional guarantees against these risks (loss of adaptation of the sanction to the specific case; risk of receiving an unjust sanction due to a false positive), nor did the public administration implement organizational or procedural measures to meaningfully address these issues, allowing improper sanctions to occur for years, as we have seen.

2.2. Labour Inspectorate Procedures for Labour Law Violations

In the sectoral field of labour, which, together with tax control and traffic control, is one of the public administration areas where automation has developed most extensively in Spain, regulations can be found since 2009 that specify actions that may be carried out using automated decisions. In particular, regarding unemployment benefits, this possibility was introduced through Royal Decree-Law 10/2009 of 13 August, which regulates the temporary unemployment protection and insertion program, in its first final provision^[20].

However, the most significant development in automation, at least from the perspective relevant here, related to administrative monitoring and enforcement functions, occurred in 2021. Indeed, in 2021 there was an amendment in the Spanish rules regarding the Labour and Social Security Inspectorate. Article 53 of Royal Legislative Decree 5/2000 of 4 August, approving the consolidated text of the Law on Infringements and Sanctions in the Social Order, provides, following that reform in 2021^[21], that infringement reports issued by the Labour and Social Security Inspectorate may be generated in an automated manner.

In the same vein, Royal Decree 928/1998 of 14 May, approving the General Regulations on Procedures for the Imposition of Sanctions for Social Order Infringements and for Social Security Contribution Settlement Files, was also amended in 2021 in order to elaborate further on automated action, with a specific regulatory framework in Articles 43 et seq., incorporating notable safeguards regarding human intervention in cases of disagreement with the automated system's criteria.

It is envisaged that the automated system of the Inspectorate may carry out the checks automatically and even issue the infringement reports on its own (Articles 44 and 45 Royal Decree 928/1998), which are notified and, if neither the indicated fine has been paid nor explicit disagreement with the content of the report has been expressed, may automatically generate the decision containing the proposed penalty, so that the subsequent steps of the labour administrative sanctioning procedure may be pursued before the competent labour authority.

Accordingly, this Royal Decree points out that «*If allegations are made invoking facts or circumstances different from those stated in the record, inadequacy of the factual account in said record, or unfairness for any reason*» to the automated infringement report, it will be then when «*the case must be assigned to an officer with inspection functions, to report on them*». In consequence, in this case, the role of the human inspector seems to be relegated to that of a mere reviewer of the automated system, but limited to intervening only in the event of the recipient's disagreement with the automated control received.

However, it is not specified which inspection and control activities, with respect to compliance with labour and Social Security obligations, will be carried out in an automated manner, as this task is deferred to a subsequent administrative act. In this regard, it is stated that this will be determined by «*a resolution of the Director of the State Body for Labour and Social Security Inspectorate, to be published in the electronic headquarters, specifying the cases in which such automated action will be used*» (Article 43 of Royal Decree 928/1998).

This is the formal, legal description found in the legislation regulating labour inspection activities. However, this is not sufficient to determine whether these provisions have had any practical effectiveness or implementation, that is, whether labour inspection and sanctioning activities are actually being automated. Unfortunately, very little information is currently available in the public domain on this matter.

The European Labour Authority (ELA) made a study visit to Spain on the 6th February 2024, in order to study the automated procedure on formal infringements by the Spanish Labour and Social Security Inspectorate. ELA published the corresponding review, and, according to that paper, the Spanish inspectorate has, indeed, a pilot automated penalty system for employers that commit formal infringements, but at the time of the visit it was running still

under human oversight, despite being designed to work autonomously^[22].

According to the description in the ELA's executive summary of the *Study Visit to Spain: Automated Procedure on Formal Infringements*, the main objective of the project is to leverage existing data sources, such as Social Security registration and declaration information, to automatically and proactively identify corporate behaviors that breach uncontested formal obligations, such as notifications of worker registrations within established deadlines. These types of non-compliances, referred to as formal infringements, do not require complex or subjective investigation by an inspector at the initial stage, making them particularly suitable for automated processing.

Operationally, the pilot system integrated by the Labour Inspectorate uses algorithms and technological tools to cross-check data from various official sources to detect indications of formal non-compliance. When the presence of such an infringement is confirmed, the system automatically generates a proposed sanction, which can be translated into a notification electronically sent to the offending employer. In accordance to the ELA's paper, «*Infringement notices are mass-generated, and individuals are promptly notified, receiving an invoice that offers a 40% penalty reduction. The electronic notifications speed up the process. Notified individuals can either pay with the reduction and end the procedure or submit allegations. If they submit allegations, the files are transferred to an inspector or sub-inspector. If no action is taken, an automated resolution proposal is forwarded to the competent authority for a final decision*»^[23].

Although the process is designed to become fully automated, at the current stage the sanctions generated are reviewed by human inspectors before being communicated, thereby ensuring a space for human supervision to address potential errors or misinterpretations that could arise from exclusive reliance on digital computational tools.

This regulatory framework governing labour inspection activities thus reveals the importance of providing agile mechanisms enabling alleged offenders to challenge the soundness of the automated outcome before a human inspector. As seen before, sometimes AI mistakes and false positives can be really obvious mistakes for a human being.

This solution can, at least partially, mitigate the problem of false positives through this additional safeguard, although it will be necessary to ensure that the

submission of such arguments is a very simple procedure and does not impose an excessive burden on the alleged infringer. Even so, the caution with which the system is being deployed is evident, since, despite having had legal authorization since 2021 to operate with full autonomy, in 2024 the system was still operating in a pilot phase with limited autonomy, subject to final human supervision of each sanctioning decision proposed by the digital system.

3. New Law for a New Era: Moving Forward or Moving Backwards?

Crucially, this digitally-driven transformation in administrative controls is also altering the regulatory architecture of public powers on punishing law infringements. New punishing regimes designed for automated enforcement display features that depart from long-standing administrative traditions. The ordinance establishing low-emission zones in Barcelona exemplifies this shift: traditional sanction ranges with discretionary minimum and maximum amounts were replaced by fixed, uniform penalties, eliminating the individualised case-by-case assessment.

This form of normative rigidification will be expanded to other areas in order to allow automated action, as suggests the developments in Denmark's digitally ready legislation strategy, which seeks to draft legal rules in ways that facilitate automation. As Gøtze^[24] has argued, this digitally ready legislation tends to reduce administrative discretion and constrains the contextual evaluation reducing the space needed for case-specific assessment, which, one can add, has been traditionally central in the common punishing practices, under the principle of proportionality. Indeed, these changes raise serious legal and constitutional concerns.

Automated sanctioning may be incompatible with core components of proportionality doctrine, particularly the obligation to tailor sanctions to the circumstances of the specific case and to prevent arbitrary or excessive penalties. Indeed, by way of example in the Spanish case, it may be noted that automated administrative punishment entails a direct conflict with the principle of proportionality in at least two respects that the Spanish courts have upheld in connection with the necessary individualisation of penalties to the specific case,

in order to avoid disproportionate sanctions or, more generally, administrative decisions adopted without duly taking into account the circumstances present in a given case.

Firstly, Spanish Constitutional Court Ruling 199/2014, of 15 December (ECLI:ES:TC:2014:199), held that the administration had violated the principle of proportionality in sanctioning by regulating fixed amounts of fines since penalties need to be individualized. As stated in that ruling: *«the application of this key determined a fixed amount of the fine, disregarding the principle of proportionality, with no reference whatsoever to the individualisation of the sanction»*.

A second interesting ruling takes into account another relevant safeguard in relation to administrative punishment: the Spanish Constitutional Court has developed the protection of the punished subject against potential arbitrariness of the public administration when punishing an administrative law infringement, stating that penalties must be adapted to the specific case by providing adequate reasoning as to why the amount imposed is appropriate in each concrete instance, and that constitutional review of the decision is possible. As stated in a 2021 ruling: *«indeed, the individualisation of the penalty, within a range that does not infringe the requirement of predetermination, corresponds to the punishing administration, and the role of this Court in an eventual appeal proceeding must be limited to verifying the constitutional correctness of the reasoning developed to justify the specification of the administrative penalty in light of the circumstances present in the case, expressly taking into account that the mere passage of time may cause the amount of penalties to lose relevance, in proportion to the conduct that may be subject to punishment»*^[25].

A third relevant ruling is found in the Supreme Court Judgment of 5 June 2025 (ECLI:ES:TS:2025:2596), which established that it is possible to apply the penalty corresponding to a less serious category of infringement when, if the penalty applicable to the formally classified offence were imposed, the outcome would be excessive in light of the circumstances of the case. As stated by the Supreme Court: *«the Chamber considers that there are sufficient arguments to regard this as a new criterion for modulating the imposition of administrative sanctions in justified cases where the strict application of the legal classification of a predetermined sanction may lead to a disproportionate result»*. In this regard, the

principle of proportionality is given precedence over the principle of typification, understood as an element of the broader principle of legality in sanctioning matters.

In short, the Constitutional Court has established that standardising punishments in the form of fixed sanctioning amounts infringes the principle of proportionality by preventing their individualised application to the specific case, and it affirms that such case-specific application must be supported by adequate reasoning capable of dispelling any trace of arbitrariness. Going a step further, the Supreme Court even introduces a new requirement derived from the proportional application of administrative penalties on a case-by-case basis, consisting in assessing whether a less serious category of infringement should be recognised in the particular case in order to adjust the penalty accordingly.

Looking at other legal systems, further issues linked to proportionality can also be identified which clearly require a case-by-case assessment before deciding on the sanction. For instance, in the case of the Czech Republic, constitutional doctrine established in 2003 implies that the fine cannot be imposed in such an amount that would ultimately be liquidating for the sanctioned entrepreneur^[26].

Moreover, at the European Union level, it has been said that since Greek Maize case in 1988/9^[27], there has been a solid trend to even through secondary law instruments, consider that sanctions shall be, both by criminal or administrative proceedings, «*effective, proportionate and dissuasive*» (Judgment of the Court of 21 September 1989, Case C-68/88, Commission v. Greece, ECLI:EU:C:1989:339, paragraph 24). In this regard, there is a solid and more advanced construction nowadays regarding solely the “principle of proportionality”, insisting, among other aspects, in the extent of the principle also to the assessment of the factors which may be taken into account in the fixing of the fines (Judgment of the Court (Sixth Chamber) of 19 June 2025, Case C-671/23, M v Lietuvos bankas, ECLI:EU:C:2025:457, paragraphs 45-63). From a rights-based perspective, automated sanctioning by issuing a fine proposal when an infringement is automatically detected threatens two rights that have been considered relevant safeguards on the common core of Administrative Laws in Europe, such as the right to be heard and the duty to give reasons when an administrative decision affects someone, especially in a detrimental manner^[28]. Moreover, both rights are considered to be recognised in common rights charters

in Europe, since the European Court of Human Rights confirmed their inclusion under the right to a fair trial under article 6 of the European Convention of Human Rights^[29], and are both, also, expressly considered a part of the right to good administration under Article 41 (1) of the EU Charter of Fundamental Rights.

Moreover, the full automation of the sanctioning procedure would not, in principle, be compatible with data protection rights.

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016, known as the General Data Protection Regulation (GDPR), establishes in Article 22 the right not to be subject to a decision based solely on automated processing, including profiling, when such a decision produces legal effects concerning an individual or similarly significantly affects them.

However, this right may be limited, and full automation could be permitted when provided for by European Union law or by the laws of a Member State, although in any case suitable measures must be established to safeguard the data subject's rights and freedoms and legitimate interests. Nevertheless, no further details are provided, in the case of this exception, as to when and how to meet the protection standard required to counterbalance the exemption from the right not to be subject to fully automated processing, and there has also been discussion on whether or not the automated administrative action frameworks comply with Art. 22 of the GDPR^[30], which may even be resolved in the negative in those cases where legislation on administrative automation is scarcely regulated and citizen safeguards underdeveloped or not provided, as considered by Reichel in relation to the clause in Article 28 of the Swedish Administrative Procedures Act^[31].

In short, it seems that these requirements, which have been developed over decades to limit the arbitrary use of administrative sanctioning power, are at risk of losing their validity, in parallel with the expansion of automated administrative sanctioning.

In accordance with the guarantees set out above, in order to determine the amount of a fine, the public administration must first exercise its discretion in order to propose the appropriate penalty in view of the circumstances of the specific case, and it cannot do so without hearing the person who is to be punished.

Faced with this contradictory situation, several lines of thought and action are possible. Firstly, it could be argued that the implementation of digital tools that automate administrative control and sanctions operates informally, that is, outside administrative procedures in the strict sense, and that the fines generated are merely proposed penalties with no legal effect, as they can be voluntarily accepted or challenged. Well, that could indeed be a valid line of thinking, although it overlooks a fundamental issue: often, in practice, these proposed sanctions have the capacity to become firm administrative penalties with full legal effect simply by the passage of time without having been subject to allegations. Furthermore, they also overlook the fact that, for the person to whom the penalty is proposed, submitting arguments may require the same technical advice as filing a formal administrative appeal. Therefore, it seems that this line of thinking relies excessively on formal categories, without taking into account that, on a practical level, the difference between a proposed penalty and a formal administrative sanction is not so great for the person being punished.

Secondly, it may be thought that those principles (in special, proportionality) were designed for a society that was very different from the present one with the advent of AI and high possibilities for automation, and that technological progress should also transform the way in which administrative oversight is carried out and the way in which administrative sanctioning powers are exercised. After all, its overall impact, from the perspective of effectiveness, seems positive in reducing law infringements, as illustrated by some papers^[32], which could justify, to some extent, an optimistic view regarding digitalization and automation within public administrations. Consequently, this view would maintain that those traditional safeguards no longer fit within the new material or practical context, and should therefore be reframed in order to also validly encompass automated decision-making, reformulating those safeguards in light of the new reality of administrative oversight possibilities. This more utilitarian view prioritizes outcomes, efficacy, and efficiency, at the expense of other aspects that it would consider less relevant once weighed against them. According to the aforementioned survey conducted by Grimmelikhuijsen & Tangi^[33], it is relatively common among public managers to believe or expect that AI would help their organisation to reap benefits. However, from a legal standpoint, this position cannot be upheld, since the guarantees I have described regarding the

exercise of administrative sanctioning powers are linked to fundamental rights, and it is not possible for lower-ranking rules to limit or even conflict with these guarantees.

Thirdly, one may adopt a rigid formal legal stance, arguing that, if fully automated administrative sanctioning action does not currently fit within the legal safeguards outlined above, it is unfeasible and no public administration should employ it for the time being, until automated systems allow for fair assessments adaptable to the specific case and for effective interaction with the offending individual.

This approach, the most protective of legal guarantees and the most rigid from a legal standpoint, would not accommodate current techniques of automated administrative oversight, since these systems are generally based on fixed-amount fines (which could be against the principle of proportionality as applied to the sanctioning power) and on infringements detected through isolated objective parameters, without interpretation of the overall context of the case or of aspects subject to subjective assessment (for example, they take into account only whether a car's license plate crossed a specific point where passage was prohibited). Moreover, they do not allow for any dialogue with the system, so that the person who is intended to be fined may be heard before the sanction proposal is issued and notified.

In light of all these possible approaches, and the risks they involve, is there room for technologies that automate certain aspects of administrative oversight and the exercise of sanctioning powers, while still providing citizens with sufficient legal safeguards and ensuring an adequate level of legal certainty for administrative action?

My personal view at this point is that automated administrative sanctioning may ultimately become legally valid and consistent with the core safeguards that have been developed in legal doctrine. However, the path is not straightforward, and much progress remains to be made within legal systems, since, given the current state of technology, it will be necessary to establish additional safeguards to compensate for the limitations these systems may have. Indeed, as would be seen in the following section, regulatory frameworks are not fully aligned in regulating administrative automated action, and it is possible to identify room for further developing a coherent set of safeguards to address the issues identified.

4. Automation and AI: the regulatory framework and the need for safeguards in automated punishment systems

Despite the scale and speed of these transformations, regulatory responses remain fragmented and conceptually underdeveloped. In fact, legal approaches across Europe – and even among different regions within the same country – are disparate and sometimes even diametrically opposed.

In this regard, the starting point is that automated decision-making does not always require the use of AI. As has been widely analysed and accepted, automated public action will in some cases rely on AI, while in other cases it will be based on less sophisticated systems that do not fall within the definition of AI^[34].

Starting from this premise, the fragmentation of the regulatory framework governing automated administrative sanctioning systems becomes evident: state-level legislation – or, where applicable, regional or local rules – on automated decision-making will always apply, while AI-specific regulations (European, national and, where relevant, regional or local) may additionally apply, but only when the system qualifies as AI.

Moreover, another element of controversy lies in the lack of a unified theoretical and definitional framework concerning automated administrative action. Thus, when referring to automated sanctioning, at least three distinct dimensions can be identified: automation in detection (e.g., cameras, recognition algorithms); automation in the legal qualification of facts (systems that classify a conduct as an infringement); and automation in decision-making (where the digital system determines the penalty applicable to the infringement). Even the generation of the document to be notified to the punished person may be subject to automation.

In the legal traditions of certain countries, such as Spain^[35] or Germany^[36], referring to automated administrative action entails only fully automated processes, without human involvement, which, in our area of activity would equal to covering all automatable stages of the cycle, from the detection of non-compliance to the drafting of the proposed sanction. Consequently, these systems, that can work without human oversight and can deliver decisions with negative effects on individuals, seems to require a higher and more specific set of

safeguards in comparison to those systems that only automate partially that control and punishing activity.

With regard to the regulation of automated administrative systems, with or without AI, legislative responses vary significantly. In Spain, basic state legislation allows automation in all cases, subject to certain safeguards that must be observed^[37]. However, in some Autonomous Communities (names of Spanish regions), the use of automated systems is expressly limited. Thus, Article 44.2 of Catalan Law 26/2010 of 3 August, on the legal regime and administrative procedure of the public administrations of Catalonia, provides that *«only those acts that can be adopted through programming based on objective criteria and parameters are susceptible to automated administrative action»*. In a similar vein, Balearic Law 7/2024 of 11 December, on urgent measures for the administrative simplification and rationalisation of the public administrations of the Balearic Islands, provides in Article 66.2 that activities involving value judgments may not be carried out by means of automated administrative actions.

Likewise, in Germany, since 2017 the possibility of adopting fully automated administrative acts is allowed, but with a limit: when using automated means, *«the administrative authority must not have any discretion»* in the activity carried out (Article 35a VwVfG).

As an intermediate solution, we can consider the Lithuanian case. Paužaitė-Kulvinskienė & Strikaitė-Latušinskaja present developments in the Code of Administrative Offences of the Republic of Lithuania No. XII-1869 2015, which permits automated sanctioning in certain specifically listed cases that are considered *«typically clear, objective, and indisputable»*, significantly across traffic and tax domains^[38].

Furthermore, where such automated administrative sanctioning systems are based on AI, additional rules, specifically tailored to rule AI, apply. As an interesting example, we can consider Galician Law 2/2025 of 2 April, on the development and promotion of artificial intelligence in Galicia. As established in Article 1.2 of this Galician law, it is a statute that lays down *«the legal framework governing the use of artificial intelligence by the Galician public sector in the exercise of its activities and in its relations with citizens, businesses, entities, and other public administrations»*.

Among other noteworthy aspects, attention may be drawn to the recognition of

the right to request the review of decisions made by AI systems on the grounds of technical deficiencies, unreliable methods, or bias (Article 25 of Galician Law 2/2025); as well as the establishment of safeguards to prevent bias in AI systems, and the limitation on the use of AI systems for the adoption of automated administrative acts, such that, in principle, they may not be employed when such acts require a subjective assessment of the relevant circumstances or legal interpretation (Article 12 of Galician Law 2/2025).

However, the possibility of resorting to the use of AI in such cases is envisaged, provided that the additional safeguards established in Article 12.5 of Galician Law 2/2025 are implemented. Among these safeguards is the requirement to design the system *«so that it does not allow unsupervised alteration of the system's or model's operation and provides simple and easily understandable information about its functioning, enabling affected individuals to comprehend and challenge the outcome»*.

Moreover, going further on AI-specific rules, we can consider the safeguards for AI considered as high-risk systems under the EU AI Act (Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828).

The AI Act defines high-risk systems in Article 6, distinguishing between security-related systems and those listed in Annex III. While some systems used by public authorities are considered high-risk, the types analysed in this paper generally perform tasks that may not fall under this classification. Indeed, Annex III refers to Law enforcement functions, but according to its definitions, this concerns to criminal investigations. Indeed, section 6 refers to law enforcement, and to functions carried out by law enforcement authorities, generally in relation to criminal activities (criminal offenses). Additionally, the term “law enforcement authority” is defined in Article 3(45) and the subject of “law enforcement” could be found in Article 3(46), emphasizing their connection with the detection and prosecution of crimes, as well as the enforcement of criminal sentences. In consequence, since administrative authorities using these systems lack jurisdiction to prosecute crimes, limiting their primary purpose to

detect and punish administrative infringements, these systems cannot fall under that type of high-risk systems. While such systems could incidentally detect criminal activity, this is not their main function.

On the other hand, Section 5 in Annex III refers to AI-systems related to the access to and enjoyment of public services and benefits. In particular, subsection (a) contemplates systems used by public authorities to grant or revoke essential benefits or services. Obviously, AI based control systems could be used to detect fraud in public assistance, such as the systems used by the Labour Inspectorate in Spain to detect whether a person is fraudulently receiving subsidies or benefits. This could lead us to consider, at least hypothetically, whether that clause might be interpreted broadly to apply high-risk AI safeguards to systems that may produce such indirect consequences, even if their primary purpose is different, such as detecting fraud. Hence, if the system detects a fraud, and the consequence of the fraud is to revoke the benefit, are we dealing with a high-risk AI system?

In this regard, despite not being highly clear, we must consider that subsection (b) specifically excludes from its high-risk status systems that are used to detect financial fraud, which, to some extent, can have the revoking effect (for benefits or subsidies) as a consequence of detecting a fraud.

Consequently, we can, at the present moment, say that systems used for administrative punishing in this context currently do not clearly meet the criteria on Annex III and are unlikely to be classified as high-risk. This points out to a recommendation of including the systems based on AI and that are used to detect, and/or punish administrative infringements, since their consequence on citizens can be really detrimental.

Finally, to sum up, when considering together automated action rules and AI regulations, some common ground emerges: the limitation of digital means in relation to tasks or activities implying discretion or the need of some kind of relevant subjective judgment in appreciation. In this vein, Ponce argues in favour of a «*legal reservation of the exercise of discretionary powers in favour of human beings*», in order to ensure «*the human exercise of empathy in the diligent weighing of facts, rights, and interests that constitutes the proper administration of discretionary powers*», which he has termed as a reservation of humanity («*reserva de humanidad*»)^[39].

Sanctioning activity is particularly subject to discretion and requires individualized assessment, as established by the principle of proportionality. For this reason, it seems difficult to conceive that it could be fully automated, given the legal safeguards surrounding the use of technology.

However, as some examples have shown, it is likely that not all sectoral areas require the highest level of human rigor when detecting infringements, since, as shown by the approach followed in Lithuanian Code or in the “formal infringement” approach by the Spanish Labour Inspectorate, it could be possible to identify a small number of infringements that can be detected automatically without requiring contextual interpretation, paired with fixed, low-value penalties that are intended to have minimal impact on the offender’s assets. Nevertheless, it will always be necessary to balance the safeguards lost when the individualized logic of justice is bypassed by not tailoring the sanction to the specific case. Moreover, another problem emerges, as shown in the low-emission zone in Barcelona: even when the infringement is objective and easy to detect with simple data-crossing, the system can cause false positives, both by bugs and by misunderstanding the context (e.g. the polluting car being carried by the tow truck).

In consequence, even when only some specific infringements are isolated to automated compliance control, extra safeguards need to be in place in response to this phenomenon. These safeguards should, at least, ensure that individuals are informed about the automated nature of control activities, and guarantee the possibility of promptly appealing any automated decision to a human authority/inspector.

A positive example can be found in the aforementioned Royal Decree 928/1998 of May 14, as amended in 2021. Firstly, automatically generated infringement reports must explicitly indicate that they result from an automated administrative process and must specify the means used to verify the supporting facts (Article 45). Secondly, any objections to these reports must be reviewed by human inspection staff (Article 47.3), ensuring human oversight. These objections can be submitted without legal representation and at no cost to the individual, although seeking legal counsel would entail personal expenses not recoverable from the administration.

Awareness that an alleged infraction and a penalty were identified and generated

automatically can be highly valuable when preparing an objection. The individual may simply argue that the system made an error, attaching relevant documentation to support their claim that no infraction occurred, or, simply, pointing out the incoherence between the legal reasoning and the facts observed in the data input (e.g. the polluting car being towed through the low-emissions perimeter). Human review of the objection opens the door to identifying false positives, annulling improperly proposed penalties, and ideally prompting internal review of the AI system to address the source of the error.

Furthermore, the right to request a review of AI-based decisions is explicitly recognized in certain legislation, such as Article 25 of the Galician Law 2/2025. This provision allows individuals to challenge decisions based on technical deficiencies, unreliable methods, or biases in the system, strengthening the protection of rights against errors or unfairness generated by automated processes.

However, discouragingly, these safeguards related to the automation of public administration are implemented in a sector-specific rather than a general, cross-European manner. Consequently, their application is limited to certain sectoral control activities or to specific digital tools.

In general, legal scholarship has made significant efforts to study AI, and also there has been some interesting developments in AI regulation, such as the EU AI Act, and some national rules. However, the study and the development of legal systems regarding the administrative sanctioning use of these automated digital tools remains incomplete. This is partly due to the lack of a comprehensive development of administrative sanctioning powers in Europe, and especially because the treatment of systems that may be automated without AI has not been addressed at a European level, despite their growing practical application.

Administrative sanctioning law, in general terms, has historically lacked its own distinct identity, having been largely constructed with inspiration from criminal law, treating both punishing sources as complementary components of a single *ius puniendi*, governed by equivalent rules and principles.

Nonetheless, for more than a decade now, scholars have begun to speak of an administrative sanctioning law that is freeing itself «*from the traditional constraints imposed by criminal law*»^[40]. Moreover, it has been observed – sometimes with a degree of caution – that in several areas sanctioning is also a

way to administrate^[41], highlighting the clear lack of alternative means to ensure compliance with legal norms. This occurs alongside the expanding use of sanctioning powers by public administrations in diverse areas, aimed at deterring potential infractions^[42].

In light of this landscape, marked by regulatory fragmentation and the insufficiency of current safeguards, it is essential to move towards a coherent and harmonised system for the regulation of digital systems that automate administrative punishing powers in Europe.

Firstly, a reform of national administrative procedure laws is required to establish a common minimum standard of safeguards for any automated administrative action with sanctioning effects, regardless of whether it employs AI or simpler digital systems. This standard should include, at least, 1) a “reservation of humanity” that limits full automation to infringements of an objective and rule-bound nature, excluding those that require contextual or discretionary assessments, and reducing to a minimum the use of fixed sanctions (avoiding rigidification of penalties), and always providing reasons for their appropriateness, since they cannot be adapted to the specific circumstances of each case and therefore place essential elements of the principle of proportionality under strain. And 2), the right to meaningful human oversight, allowing any citizen to challenge an automated decision through a simple, free procedure that leads to effective review by an inspector empowered to correct “false positives” and, where appropriate, to promote the reprogramming or updating of the digital system to prevent future errors.

Secondly, this procedural standard should be connected with AI regulations, particularly the EU AI Act, through the explicit inclusion in its Annex III of AI systems used for the detection of administrative infringements and the imposition of sanctions as high-risk systems. This classification would entail the application of enhanced requirements for transparency, human oversight, risk management, and traceability throughout the system’s entire lifecycle.

5. Concluding Remarks: Automation, Sanctioning Power and the Future of Administrative Law

5.1. The growing tension between efficiency-driven automation and classical legal safeguards

The progressive automation of administrative monitoring and sanctioning functions marks a profound transformation in the exercise of public power. This digital transformation is not marginal or speculative. As demonstrated by the empirical examples analysed in this paper, automated administrative sanctioning is already a concrete reality. These developments illustrate how automation increasingly substitutes human judgment in decisive procedural stages, including the identification of infringements, the qualification of conduct, and the proposal of penalties.

One of the main points of interest of technology aiming automated controls is that it make it possible to implement effective enforcement tools. They even can allow to introduce new legal duties that could not otherwise be effectively imposed, due to the practical impossibility of properly monitoring their enforcement. For example, in the case of low-emission zones, it would be possible to carry out occasional random checks on vehicles entering low-emission restricted areas, with police officers stopping some vehicles in order to verify, through a database, if their license plates are in the list of low polluting vehicles. However, traditional (or mainly human) enforcement of such an obligation may require a large number of officers and may cause significant disruption for road users, since it would further hinder mobility. Indeed, it should also be noted that this type of automated monitoring involves fewer burdens or inconveniences for those subject to control, as the system performs the verification of infringements without interrupting the activity being carried out. Thus, access-control systems for low-emission zones do not require vehicles to stop, as mentioned before, and likewise, the Labour Inspectorate system do not involve inspectors entering the offices of companies, nor do they interrupt staff activity in order to provide documentation or respond to other requests from inspection personnel.

These benefits, at first glance, appear positive. However, as noted earlier, there is a clear tension with certain rights and principles, particularly the principle of proportionality in the context of sanctioning powers, as well as the rights to be heard and not to be subjected to fully automated decision-making. In doing so, automation challenges long-established assumptions about discretion, procedural

guarantees, and individualised justice that have traditionally structured administrative sanctioning on the presuppose deliberation, contextual evaluation and meaningful human judgment.

Automated sanctioning systems, however, operate on fundamentally different logics. They are designed to process large volumes of data, apply predefined parameters, and generate standardised outcomes with minimal or no human intervention. This technological architecture favours uniformity, speed, and scalability, but systematically constrains contextual sensitivity, discretionary balancing, and dialogical interaction.

This tension becomes particularly visible in the rigidification of sanctioning frameworks, as illustrated by the replacement of sanction ranges with fixed penalties in order to enable full automation. While such normative choices may facilitate technological deployment, they risk undermining constitutional doctrines of proportionality. In this respect, automation does not merely affect procedural techniques but reshapes the very substance of administrative sanctioning law, and additional safeguards should be in place in order to reestablish a balance.

5.2. False positives and mass enforcement: the problem of systemic error

The empirical cases examined in this paper also reveal that automated sanctioning systems significantly amplify the risks associated with false positives and systemic error. In traditional enforcement models, human oversight and limited enforcement capacity functioned as implicit filters that contained the scale of erroneous sanctions. By contrast, automated systems are capable of producing a lot of sanction proposals daily, meaning that even low error rates may translate into substantial numbers of unjust penalties.

The Barcelona low-emission zone example illustrates this phenomenon. Automated licence-plate recognition systems repeatedly generated fines against vehicles that were objectively not circulating, but merely being towed. These errors, which would have been immediately apparent to a human observer, persisted over time due to the absence of systematic human verification. The result was a form of algorithmically produced administrative injustice, where

individuals bore the burden of contesting manifestly incorrect penalties.

This dynamic highlights a structural shift in the allocation of procedural burdens. In automated systems, the default assumption increasingly becomes the correctness of the machine-generated outcome, while individuals are expected to activate complex objection mechanisms in order to rectify errors. Such an inversion of procedural logic raises serious concerns from the perspective of good administration and effective protection of rights, especially when sanctions are relatively small but massively replicated. In that regard, it is important to recall that the *Sindicatura de Greuges* de Barcelona highlighted that avoiding control and oversight of the digital system outcomes is against the citizen's right to good administration^[43].

5.3. Fragmentation and insufficiency of current regulatory frameworks

Another major conclusion of this study concerns the fragmentation and inadequacy of the current legal framework governing automated administrative sanctioning. While general provisions on automated administrative action and AI regulation exist at both national and European levels, they remain largely disconnected from the specificities of sanctioning procedures. As a result, they fail to provide targeted safeguards proportionate to the intensity of interference involved in administrative penalties.

Neither the general administrative automation clauses nor the GDPR's framework on automated decision-making sufficiently address the concrete procedural challenges posed by automated sanctioning. Key issues such as mandatory human review and meaningful participation rights remain under-regulated.

Moreover, the coexistence of general automation law, sector-specific enforcement rules, and emerging AI regulations produces a highly complex normative environment. This fragmentation undermines legal certainty for both public authorities and individuals, complicates judicial review, and hinders the development of coherent standards of protection. The absence of a dedicated regulatory framework for automated administrative sanctioning thus emerges as one of the most pressing challenges identified in this paper.

1. See A.G. Larsen, A. Følstad. *The impact of chatbots on public service provision: A qualitative interview study with citizens and public service providers*, in *Government Information Quarterly*, vol. 41(2), Elsevier BV, Amsterdam, 2024; and A. Brioscú, et al. *A new dawn for public employment services: Service delivery in the age of artificial intelligence*, in *OECD artificial intelligence papers*, 19, 2024, pp. 34-37.
2. Among many others, see L. Carnis, *Automated Speed Enforcement: What the French Experience Can Teach Us*, in *Journal of Transportation Safety & Security*, 3(1), 2011; A.Snow. *Automated Road Traffic Enforcement: Regulation, Governance and Use: A review*, RAC Foundation, London, 2017; and J. M. Martínez Otero. *Hipervigilancia administrativa y supervisión automatizada: promesas, amenazas y criterios para valorar su oportunidad*, in *Revista Española de Derecho Administrativo*. 231, 2024.
3. See S. Ranchordás & Y. Schuurmans. *Outsourcing the Welfare State: The Role of Private Actors in Welfare Fraud Investigations*, in *European Journal of Comparative Law and Governance*, 7:1, 2020, p. 6.
4. European Commission. *Public Sector Tech Watch. Mapping Innovation in the EU Public Services*, in Publications Office of the European Union, Luxembourg, 2024, available on www.european-union.europa.eu. 3
5. See S. Grimmelikhuijsen and L. Tangi. *What factors influence perceived artificial intelligence adoption by public managers*, Publications Office of the European Union, Luxembourg, 2024, available on www.publications.jrc.ec.europa.eu.
6. See J. Tang, et al., *Automatic number plate recognition (ANPR) in smart cities: A systematic review on technological advancements and application cases*, in *Cities*, 2022, p. 129.
7. A relevant discussion, based on the Spanish legal system and administrative practice with these type of automated systems, can be found in M. Casino Rubio, *La automatización de las actuaciones administrativas sancionadoras*, in M. Vaquer Caballería, & J. Pedraza Córdoba (eds.). *La Actuación Administrativa Automatizada: sus claves jurídicas*, Tirant Lo Blanch, Valencia, 2025, pp. 559-560.
8. See V. Sharp, *Smart Administrative Punishment: a Slippery Slope of Automated Decision-Making and its Economic Incentives in Public Law*, in *CERIDAP*, 4/2025.
9. See section 2.1 of this paper. Also, a fine proposal by this kind of systems is analysed in Sindicatura de Greuges de Barcelona. *Resolució de la Sindicatura de Greuges. Queixa relativa al dret a una bona administració (gestió i recaptació)*, available on www.sindicaturabarcelona.cat.
10. See ITSS (Inspección de Trabajo y Seguridad Social), *Plan Estratégico ITSS 2025-2027*, available on www.oeitss.gob.es; and also see J. M. Goerlich Peset. *Decisiones administrativas automatizadas en materia social: algoritmos en la gestión de la Seguridad Social y en el procedimiento sancionador*, in *Labos*, vol. 2(2), 2021.
11. Some of the most representative papers on that matter are L. Carnis,, *Automated Speed Enforcement: What the French Experience Can Teach Us*, in *Journal of Transportation*

- Safety & Security*, 3(1), 2011; and L. Carnis, *Automated Speed Detection and Sanctions System: Application and Evaluation in France*, in *Journal of Intelligent Transportation Systems*, 12(2), 2008.
12. See, on the Lithuanian case, J. Paužaitė-Kulvinskienė, G. Strikaitė-Latušinskaja. *Regulating automation: the legal landscape of 'automated administrative orders' in Lithuania*, in *Italian Journal of Public Law*, vol. 17, issue 2/2025.
 13. See C. Van Noordt & G. Misuraca. *Artificial intelligence for the public sector: results of landscaping the use of AI in government across the European Union*, in *Government Information Quarterly*, 2022, vol. 39(3), Elsevier BV, Amsterdam.
 14. In this regard, empirical evidence is presented in Z. A. Cheng, Z. Dong, M. S. Pang, *Automated Enforcement and Traffic Safety*, in *Management Science*, 71(12), 2025.
 15. This set of cases is covered by the well-known issue of *ne bis in idem*, considered the principle establishing that the same person cannot be sanctioned twice for the same conduct under the same grounds, which is generally prohibited. For a European take, see the European Court of Human Rights Judgment, 10 February 2009, case of Sergey Zolotukhin v. Russia. Moreover, Principle 3 in the Council of Europe Recommendation No R(91)1 of the Committee of Ministers to member states on administrative sanctions considers this issue, stating that «*A person may not be administratively penalised twice for the same act, on the basis of the same rule of law or of rules protecting the same social interest*», in Spain, Article 31 Law 40/2015 states that «*Acts [meaning, in this context, acts that are qualified as infringements] that have already been sanctioned criminally or administratively may not be punished again in cases where there is an identity of the subject, the act, and the legal basis*».
 16. A reflection on this last judgment can be found on M. J. Montoro Chiner, M. J. *Ordenanza relativa a la zona de bajas emisiones para preservar la calidad del aire en la ciudad de Barcelona: Análisis del instrumento ambiental, reflexión sobre el alcance de las restricciones impuestas y consideraciones sobre las sentencias del Tribunal Superior de Justicia de Cataluña, de 21 de marzo de 2022*, in *Cuadernos de derecho regulatorio*, Vol. 1(1), 2023.
 17. See AMB. *Guía técnica para la implementación de zonas de bajas emisiones*. Barcelona, AMB-FEMP, 2021, pp. 33-38, available on www.femp.femp.es.
 18. See www.lavanguardia.com.
 19. See www.elperiodico.com.
 20. On this matter, with reference to other developments in the labour and social security sector, J. M. Goerlich Peset. *Reglamento de inteligencia artificial e intervención pública en las relaciones laborales*, in *Labos*, vol. 5, 2024, pp. 230–233.
 21. J. M. Goerlich Peset. *Decisiones administrativas automatizadas en materia social: algoritmos en la gestión de la Seguridad Social y en el procedimiento sancionador*, in *Labos*, vol. 2(2), 2021, pp. 22-25.
 22. The following lines, from that paper, define the kind of infringements that can be automatedly detected (what they call “formal infringements”) and the current state of its usage: «*Formal infringements' is a type of non-compliance by employers in the Social*

Security system which is uncontroversial and does not require further investigation. It can be detected without direct intervention from an inspector at that stage. For example, if employers fail to notify workers' registrations on time, this delay is automatically recorded in the social security's files. The Spanish Labour and Social Security Inspectorate (ITSS) has a semi-automated procedure for detecting and sanctioning formal infringements in different areas, such as the improper use of bonuses in labour contracts by employers. The sanction is automatically generated, but it is still reviewed by inspectors before it is notified to the employers. The system is ready to be moved to a fully automated process but has not been implemented yet» (ELA, Study Visit to Spain: Automated Procedure on Formal Infringements, 2024, p. 1, available on www.ela.europa.eu).

23. ELA, *Study Visit to Spain: Automated Procedure on Formal Infringements*, 2024, pp. 4-5, available on www.ela.europa.eu.
24. See M. Gøtze. *Danish Digital Design and the Gradual Erosion of Technology Neutral Administrative Law*, in *Central European Public Administration Review*, 22(2), 2024.
25. Spanish Const. Court Ruling, jud. 28 January 2021, n.4.
26. See T. Krabec & R. Čížinská R. *Measuring the impact of an administrative fine on a company and its future survival: a case study from the Czech Republic*, in *Financial Internet Quarterly*, 17 (4), 2021.
27. H. Korkka-Knuts, S. Melander. *Contours of a principled corporate sanction policy in the EU: Exploring a constitutionally justified balance between criminal and administrative sanctions*, in *New Journal of European Criminal Law*, 16(1), 2025, p. 34.
28. See G. della Cananea, *The Common Core of European Administrative Laws: Retrospective and Perspective*, Brill: Leiden, 2023, pp. 98-99.
29. In this regard, see A. Andrijauskaitė. *The Principles of Administrative Punishment under the ECHR*. Doctoral dissertation, Vilnius, 2022, available on www.dopus.uni-speyer.de.
30. In this regard, see J. P. Schneider & F. Enderlein. *Automated Decision-Making Systems in German Administrative Law*, in *CERIDAP*, n. 1, 2023, pp. 105-106.
31. See J. Reichel. *Regulating Automation of Swedish Public Administration*, in *CERIDAP*. Issue 1, p. 82.
32. See Z. A. Cheng, Z. Dong, *Automated Enforcement and Traffic Safety*, in *Management Science* 71(12), 2025.
33. See S. Grimmelikhuijsen, L. Tangi, *What factors influence perceived artificial intelligence adoption by public managers*, Publications Office of the European Union, Luxembourg, 2024, available on www.publications.jrc.ec.europa.eu.
34. For all, see I. Martín Delgado, *El impacto de la IA en la toma de decisiones administrativas. Una visión general a la luz del ordenamiento jurídico español*, in Zegarra Valdivia (ed.), *La inteligencia artificial y la actividad de la administración pública*. CIACJ, Lima. 2025.
35. See Article 41 Law 40/2015 of 1 October, on the Legal Regime of the Public Sector (*Régimen Jurídico del Sector Público*), and also I. Martín Delgado. *El impacto*, op. cit. .
36. See article 35a of the General Administrative Procedures Act (*Verwaltungsverfahrensgesetz – VwVfG*). Also, see J. P. Schneider & F. Enderlein. *Automated Decision-Making Systems*

- in *German Administrative Law*, in *CERIDAP*, n. 1, 2023; C. Fraenkel-Haeblerle. *Fully Digitalized Administrative Procedures in the German Legal System*, in *European Review of Digital Administration & Law – Erdal*. Vol. 1, Issue 1-2, 2020; and E. Buoso. *Fully Automated Administrative Acts in the German Legal System*, in *European Review of Digital Administration & Law – Erdal*, vol. 1 issue 1-2.
37. See Article 41 Spanish Law 40/2015, and also see A. Cerrillo i Martínez, *Lección 21. Actuación automatizada, robotizada e inteligente*, in: *Manual de Derecho administrativo* (2n ed.), Marcial Pons, Madrid, 2024.
 38. See J. Paužaitė-Kulvinskienė & G. Strikaitė-Latušinskaja. *Regulating automation: the legal landscape of ‘automated administrative orders’ in Lithuania*, in *Italian Journal of Public Law*, vol. 17, issue 2/2025, p. 22.
 39. See J. Ponce Solé, J. *Inteligencia artificial, Derecho administrativo y reserva de humanidad: algoritmos y procedimiento administrativo debido tecnológico*, in *Revista General de Derecho Administrativo*, 50, 2019.
 40. A. Nieto. *Derecho administrativo sancionador*, Tecnos, Madrid, 5th ed., 2012, p. 15.
 41. In this regard, see A. Huergo Lora. *Diferencias de régimen jurídico entre las penas y las sanciones administrativas que pueden y deben orientar su utilización por el legislador, con especial referencia a los instrumentos para la obtención de pruebas*, in A. Huergo Lora (ed.), *Problemas actuales del Derecho administrativo sancionador*, Iustel, Madrid, 2018, p. 29.
 42. See H. Korkka-Knuts, et al. *Reassessing Deterrence: The Effectiveness of Administrative Sanctions across three EU Regulatory Frameworks*, in *Helsinki Legal Studies Research Paper* No. 90, Forthcoming in the *European Law Review* (2025), available on: www.ssrn.com/abstract=5169603 or www.dx.doi.org.
 43. See Sindicatura de Greuges de Barcelona, *Resolució de la Sindicatura de Greuges. Queixa relativa al dret a una bona administració (gestió i recaptació)*, 2024, available on www.sindicaturabarcelona.cat.

Technology-driven harmonization? Automation in administrative decision-making at the national level and the role of the EU Charter

Pavína Hubková

DOI: 10.13130/2723-9195/2026-1-134

Lo sviluppo delle tecnologie digitali consente alle autorità amministrative nazionali di tutta l'UE di utilizzare strumenti di decisione automatizzata o persino sistemi basati sull'intelligenza artificiale nell'esercizio dei loro poteri pubblici in tutti i settori del diritto amministrativo. Sebbene alcuni di questi settori rientrino nell'ambito di applicazione del diritto dell'UE e altri rimangano formalmente di competenza interna, gli strumenti e i processi digitali utilizzati dalle amministrazioni sono sempre più disciplinati dalla normativa dell'UE, in particolare dal GDPR e dall'AI Act. Quando le autorità nazionali si avvalgono di tali tecnologie regolamentate, esse agiscono nell'ambito di applicazione del diritto dell'UE, indipendentemente dal settore di policy sostanziale interessato. Ciò comporta implicazioni di natura costituzionale. Agendo sulla base della legislazione digitale dell'UE, le autorità nazionali determinano l'applicabilità della Carta dei diritti fondamentali dell'Unione europea, che vincola gli Stati membri ogniquale volta essi attuino il diritto dell'Unione. La digitalizzazione e l'automazione dei procedimenti amministrativi ampliano pertanto la portata pratica dei diritti fondamentali dell'UE. Gli standard europei in materia di diritti fondamentali acquisiscono così la capacità di incidere sulla progettazione e sullo sviluppo degli strumenti di ADM (Automated Decision-Making) o di intelligenza artificiale, nonché sul ragionamento giuridico e sul controllo giurisdizionale delle decisioni generate dall'ADM o assistite dall'IA. Il contributo sostiene che ciò favorisca una forma di armonizzazione delle prassi amministrative nazionali guidata dalla tecnologia, che emerge indirettamente dalla regolazione degli strumenti digitali, e ne analizza le implicazioni sia per le autorità amministrative sia per i giudici nazionali.

The development of digital technologies enables national administrative authorities

across the EU to use automated decision-making tools or even systems based on artificial intelligence in the exercise of their public powers across all fields of administrative law. While some of these fields fall within the scope of EU law and others remain formally domestic, the digital tools and processes used by administrations are increasingly regulated by EU legislation, most prominently the GDPR and the AI Act. When national authorities rely on such regulated technologies, they act within the scope of EU law, irrespective of the substantive policy field concerned. This has constitutional implications. By acting under EU digital legislation, national authorities trigger the applicability of the EU Charter, which binds Member States whenever they implement EU law. Digitalisation and automation in administrative proceedings thus expand the practical reach of EU fundamental rights. The EU standard of fundamental rights thus has the capacity to shape harmonisation in the design and development of ADM or AI tools, as well as the reasoning and judicial review of ADM-generated or AI-assisted decisions. The article argues that the EU standard fosters a form of technology-driven harmonisation of national administrative practices that emerges indirectly from the regulation of digital tools, and it analyses its implications for both administrative authorities and national courts.

Summary: 1. Introduction.- 2. Delimitating the scope of application of EU law and of the EU Charter.- 3. Digital legislation pulling the national administrative action into the scope of the EU Charter.- 4. Automation or AI in national administrative action and the relevant EU Charter provisions.- 4.1. Substantive rights.- 4.1.1. Right to the protection of personal data.- 4.1.2. Non-discrimination.- 4.2. Fundamental procedural rights.- 4.2.1. Right to good administration.- 4.2.2. Right to effective judicial protection.- 5. Conclusion.

1. Introduction

The development of modern technologies and the ever-increasing capacity of digitalization enable national administrative authorities across the EU to use automated decision-making (ADM) tools or even systems based on artificial intelligence (AI) in the exercise of their public powers. These tools can be used for all administrative activities, including administrative decision-making, in a

variety of fields, such as taxation, social security, market regulation, competition and migration. While some of these fields are governed by EU law, others remain outside its scope and are governed solely by national law.

Digital tools, systems and processes that can be used within national administrations are subject to ever-expanding EU legislation regulating data protection, digital governance and AI. The most prominent examples of this legislation are currently the GDPR and the AI Act. When a national administrative authority uses technology subject to such regulations, it begins to act under EU law regardless of the actual field of administrative law in which it decides.

This has constitutional implications, because using the regulated technology or processes in administrative cases pulls such cases into the scope of application of the Charter of Fundamental Rights of the EU (EU Charter). The EU Charter is applicable in national proceedings only when the authority is implementing EU law, and acting in accordance with the digital legislation governing ADM or AI tools constitutes such implementation. Thus, the increasing digitalization, technological development and automation of administrative practices opens the door to the application of the EU standard of fundamental rights across all policies, reinforcing the relevance of the EU Charter throughout the lifecycle of modern tools – from their development and design to the reasoning and review of ADM-generated or AI-assisted decisions. Furthermore, the mandate for national administrative authorities to comply with digital legislation and the EU Charter creates potential for the harmonization of certain administrative practices and procedures. This harmonization does not stem from actual procedural rules but emerges indirectly through the regulation of the technology used.

Against this background, this article examines the implications of this “technology-driven harmonization” of certain aspects of national administrative law, with regard to practice of both national administrative authorities and the national courts that review their actions. The article proceeds in three steps. First, it summarizes the ways in which the scope of application of EU law, and therefore the EU Charter in national proceedings is defined. Second, it demonstrates how digital legislation, specifically the GDPR and the AI Act, contributes to extending the scope of application of the EU Charter by

regulating technology, devices, and automated processes employed by national authorities. Third, it analyses the most relevant provisions of the EU Charter with regard to automation and the deployment of AI in national administrative decision-making, examining how and to what extent these provisions can influence national administrative practices and contribute to their harmonization.

2. Delimitating the scope of application of EU law and of the EU Charter

The provisions of the EU Charter – and the EU standard of fundamental rights protection more generally – do not apply universally and unconditionally within the EU Member States. As Article 51(1) of the EU Charter explicitly states, the EU Charter is binding on the Member States «*only when they are implementing Union law*». The meaning of implementing has been clarified by the case law of the Court of Justice saying that Member States are required to respect fundamental rights defined by EU law only when acting within the scope of EU law, and that «*the applicability of European Union law entails applicability of the fundamental rights guaranteed by the Charter*»^[1]. The underlying rationale is that Member States, when implementing EU law, function as “agents” of the Union and are therefore subject to the EU’s standard of fundamental rights^[2].

However, it may be challenging to determine in concrete cases whether a national authority is acting within the scope of EU law and thus whether it is required to apply the EU Charter at the national level. In its case law, the Court of Justice has generally identified two situations in which the EU’s standard of fundamental rights is applicable to Member States: first, when they are fulfilling an obligation arising from EU law and acting as required by it (“*Wachauf*”^[3] type of situation or the “agency situation”); and second, when they are derogating from EU law and must nonetheless respect its fundamental rights obligations (“*ERT*”^[4] type of situation or the “derogation situation”)^[5]. Nonetheless, even this categorisation is neither entirely clear nor beyond dispute, and legal scholars have already suggested several ways how to work with an alternative and more nuanced categorisation^[6].

Even the sole agency situation can be divided into sub-categories based on the

type of the EU source of obligations: First, a national authority applies an EU provision that is directly applicable, typically a provision of an EU regulation. Second, the authority also acts as an agent of EU law when it indirectly gives effect to obligations stemming from EU law through national implementing measures, most commonly by applying national acts that are transposing EU directives^[7].

The definition of the first type of the agency situation is rather unequivocal: the national authority acts within the scope of the application of EU law insofar as the given case is covered by a directly applicable EU provision, and that includes both the situation where the authority actually applies that provision and where it fails to do so. However, the second type, i.e. identifying situations involving the indirect implementation of EU measures through national legislation, can be rather challenging, especially when the case involves the application of multiple legal sources from both the EU and national levels, or when a substantive EU rule is enforced through a national procedural norm. The Court of Justice has already clarified that the indirect application of EU law also encompasses situations where a Member State adopts measures within the discretionary powers conferred by an EU act^[8]. As was explicitly stated in the landmark *Fransson* case, indirect implementation of an EU commitment also occurs when a Member State enforces substantive EU law by applying its own procedural or sanctioning rules^[9]. In concrete cases, however, it may still be unclear whether the applicable national provisions are indeed giving effect to a commitment stemming from EU law or how far the reach of EU law extends, and where, or in which respects, the matter remains governed solely by national law and hence by the national catalogue of fundamental rights.

As the Court of Justice frequently emphasises, in order to determine whether a case falls within the scope of the EU Charter, national authorities must assess *«whether the national legislation at issue is intended to implement a provision of European Union law, what the character of that legislation is, and whether it pursues objectives other than those covered by European Union law, even if it is capable of indirectly affecting that law, and also whether there are specific rules of European Union law on the matter or capable of affecting it»*^[10]. It also notes that the implementation of EU law *«presupposes a degree of connection between an act of EU law and the national measure at issue which goes beyond the matters*

referred to or the indirect effects of one of the matters on the other, having regard to the assessment criteria laid down by the Court»^[11]. In other words, the Court of Justice requires a sufficient link of a national measure to the original EU commitment or a certain degree of proximity between them. For example, in *Ispas*^[12], the Court of Justice concluded that EU law also applies to the issue of access to the administrative file in proceedings concerning the collection of VAT, because the VAT framework is governed by EU law, and thus the national enforcement of VAT rules must comply with the general principles of EU law, including the rights of defence. The national authorities must comply with such principles, even if the applicable EU legislation «*does not expressly provide for such a procedural requirement*»^[13]. In other words, the mere fact that EU VAT rules are being enforced establishes a sufficient link to the national procedural rules applied for that purpose, meaning that those procedural rules must comply with EU fundamental rights standards.

Even though the cases might have been complex, the logic of the reasoning is rather clear. However, what can become difficult to assess are situations where a national authority applies EU procedural rules that themselves serve to implement fundamental rights guaranteed by the EU Charter. This is notably the case with procedural directives in areas such as asylum or criminal law. A good example is Directive (EU) 2016/343 on the strengthening of certain aspects of the presumption of innocence^[14], which elaborates upon the right to effective judicial protection (Article 47 of the EU Charter), since it aims to «*to enhance the right to a fair trial in criminal proceedings by laying down common minimum rules concerning certain aspects of the presumption of innocence and the right to be present at the trial*»^[15]. In such cases, the applicability of the EU Charter becomes self-reinforcing: Applying an EU procedural rule (directive) that is itself designed to implement a fundamental right (Article 47 of the EU Charter) triggers the applicability of the EU Charter as a whole.

Interestingly, rules stemming from procedural directives apply even to cases that are purely national in terms of their subject matter and do not contain any cross-border elements or other aspects of EU law. For example, as the Court of Justice confirmed in *Moro*^[16], the provisions of Directive 2012/13 on the right to information in criminal proceedings^[17] are applicable in all cases, whether or not they entail a cross-border element. Criminal procedural directives are designed to

support the system of mutual recognition and to ensure that procedural rights are generally safeguarded, so that when mutual recognition is triggered, the authority in the second Member State can trust the fairness of the proceedings in the first. While the primary purpose of these directives is to lay the groundwork for potential cross-border situations, they apply even in purely domestic cases with no cross-border element. In any case, the harmonization they pursue is limited to specific aspects of criminal procedure. This means that the mere claim that the information obligation has been breached does not render the entire criminal case subject to EU law. Criminal proceedings as a whole are governed by national law, and most applicable rules, both substantive and procedural, are unaffected by EU law. EU law only intervenes in partial aspects related to providing information through a specific provision rooted in an EU procedural directive. Therefore, only these aspects fall within the scope of EU law and the EU Charter ^[18].

As the extensive case law illustrates, the precise meaning of “implementing EU law” remains subject to ongoing contestation and clarification. However, the most complex and problematic cases dealt by the CJEU concern national measures transposing EU directives. In these cases, doubts and uncertainty arise from the fact that legal obligations stemming from EU law are implemented indirectly through national provisions, and this makes it difficult to determine whether a national rule, which is applied by a national authority primarily, is sufficiently close to the original commitment required by EU law.

However, this uncertainty should not arise when national authorities find themselves in the first type of the agency situation, namely when they act under EU regulations that are directly applicable in Member States based on Article 288 TFEU. When a national authority applies a regulation, this action falls clearly within the scope of EU law and automatically triggers the applicability of the EU Charter. But even this category may be difficult where the directly applicable provision of a regulation primarily concerns procedural elements of national decision-making. In such cases, the provision affects the domestic process only indirectly or in a limited manner. The resulting ambiguities and implementation challenges then resemble those typically associated with procedural directives.

3. Digital legislation pulling the national administrative action into the scope of the EU Charter

The issue of scope of application of EU law and hence of the EU Charter arises also with regard to regulations governing the digital sphere and in the context of the digitalisation of national administrations which is covered by these regulations: In order to secure more efficient, faster and more objective decision-making, national administrative authorities are increasingly relying on tools that enable the automated processing of data or are even employing more sophisticated AI tools. The use of these modern tools is governed by two key EU regulations: the General Data Protection Regulation (GDPR)^[19] and the AI Act (AIA)^[20].

The GDPR sets out the conditions for processing personal data, and it is applicable both to private and public actors. It follows that if a national administrative authority processes personal data, it is considered either a “controller” [Art. 4(7) GDPR] or a “processor” [Art. 4(8) GDPR] and therefore falls within the scope of this directly applicable EU legislative act. More specifically, the GDPR applies to situations in which personal data is processed automatically, and this leads to a decision that significantly affects an individual. Article 22 GDPR prohibits automated individual decision-making, including administrative decision-making, unless authorized by EU or national law. If authorized, such decision-making must meet criteria that safeguard the individual’s rights, freedoms, and legitimate interests. Furthermore, individuals that are subject to fully or partially automated decision-making must be informed that the automation has taken place and provided with meaningful information about the logic of automated processes^[21].

Article 22 GDPR is thus the crucial point here: Whenever a national authority employs automated processing of personal data, it acts under this provision. As a result, automated processing of personal data in national administrative decision-making means that a national authority acts under EU law.

Similarly, whenever a national administrative authority deploys a high-risk AI system as part of its activities, it qualifies as a “deployer” in accordance with Article 3(4) AIA. If the authority develops its own AI system, it is considered a “provider” pursuant Article 3(3) AIA. It follows that in both cases, the authority

is subject to the requirements and duties of this regulation, such as transparency, record-keeping, human oversight, robustness or accuracy. If a high-risk AI system is used to generate a decision that significantly affects an individual, such a person has a right to get clear and meaningful explanation of what role the AI system played in the decision-making and how it affected the result (Article 86(1) AIA]. Again – acting under the obligations under the AIA, a national authority acts within the scope of EU law.

It does not, however, mean that any automated processes or use of any types of AI would necessarily fall under these regulations. The GDPR governs automated processing of personal data, which excludes processing of data related to legal persons, synthetic data, or data of a technical or economic nature. Similarly, the AIA contains rules for the deployment of high-risk AI system, and it is foreseeable that some AI systems used by administrative authorities do not qualify for this category. Anyway, using those digital tools that are governed by the said regulations in administrative decision-making would mean acting under these regulations, and hence “implementing EU law”, which also triggers the applicability of the EU Charter.

This introduces a significant conceptual shift: when a national administrative authority engages in automated processing of personal data or deploys a high-risk AI system in its decision-making, the case inevitably falls within the scope of EU law – and thus the EU Charter – regardless of the substantive area of the case at hand. It follows that the applicability of the EU Charter can be triggered not only by the substantive legal domain that is covered by EU law (such as VAT, asylum, or labour law), but also by the use of tools governed by directly applicable EU regulations, namely the GDPR and the AIA. As a result, the EU Charter becomes applicable even in cases that would otherwise be seen as falling outside the substantive scope of EU law, including areas such as social security in purely domestic contexts or direct taxation.

However, in such situations, the EU Charter provisions will only apply to those aspects of the case that are directly related to the use of ADM or AI tools covered by the relevant digital legislation, not to the substance of the decision. In other words, the commitments stemming from the GDPR and the AIA are not about requiring the effective enforcement of substantive EU rules through national implementing legislation or procedural law. Rather, they impose duties on

national authorities to use certain digital technologies, such as automated processing of personal data or AI systems, in conformity with EU standards, irrespective of whether the underlying administrative activity is based on EU or national law.

Here, the logic would be comparable to that applied by the Court of Justice in the *Moro* case^[22] and others involving EU criminal procedural directives, as mentioned above: While EU law does not govern the substantive layer of domestic criminal proceedings, it does require national procedures to comply with EU procedural legislation and, consequently, with fundamental rights as they are guaranteed by EU law. Aimed ultimately at establishing standards that facilitate mutual recognition, these directives partially harmonize aspects of national criminal procedure. In a similar vein, the harmonization introduced by the GDPR and the AIA aims to set standards for the use of certain technology, and it also targets only some elements of the national administrative activity.

However, unlike the harmonization of certain aspects of criminal procedure through dedicated procedural directives, the harmonization through digital legislation emerges indirectly. This legislation regulates the use of digital tools rather than imposing explicit procedural rules to be followed by administrative authorities. Neither the GDPR nor the AIA are procedural codes addressed to national administrations, but they establish rules for processing personal data and for deploying AI across all sectors, including public administration. Consequently, implications of a rather procedural nature – such as transparency requirements, the provision of meaningful information or explanations, and human review and oversight – are applicable to all technology-driven decision-making, whether the technology is used by a private entity for business purposes or by an administrative authority to produce administrative decisions or similar outputs.

Arguably, this normative framework results in a partial harmonization of administrative procedure, which emerges not through general procedural legislation, but through the regulation of specific digital tools. Accordingly, in the context of automation and AI, it is the use of digital tools regulated by EU law that brings national administrative actions within the scope of the EU Charter. More specifically, the design and use of such tools, the enforcement of rights under the GDPR (including requirements of transparency and human

oversight), and the fulfilment of obligations under the AIA at the national administrative level must comply with the relevant Charter rights.

Ultimately, the employment or deployment of modern tools in national administrative decision-making creates a layered system of the protection of fundamental rights. National law and its catalogue of fundamental rights must be observed with regard to the substance of the case, while EU law and the EU Charter govern the way how the tool is used.

4. Automation or AI in national administrative action and the relevant EU Charter provisions

It follows from the preceding analysis that when national administrative action relies on ADM tools or involves the deployment of high-risk AI systems, national authorities must comply with the EU Charter, insofar as the use of these technologies is governed by EU law. While the EU Charter applies as a whole, certain rights are particularly relevant in this context due to their close connection with digital legislation and the types of risks that are involved therein. These include overarching substantive rights such as the right to privacy and data protection, and the right to non-discrimination, which are especially critical when administrative action is mediated or supported by technology. In addition, specific procedural rights come to the forefront, especially those concerning the operation of national administrative authorities and courts. These are the right to good administration (which includes the duty to give reasons), and the right to effective judicial protection. Although both sets of rights are important, they affect the use of digital technologies in administrative decision-making processes differently. Substantive rights are likely to influence the parameters of ADM or AI tools, while procedural rights will govern the subsequent actions of administrative and judicial authorities.

4.1. Substantive rights

The fundamental rights most likely to be affected by automation and the deployment of AI in administrative practice are the right to the protection of personal data (Article 8) and the principle of non-discrimination (Article 21).

4.1.1. Right to the protection of personal data

The applicability of Article 8 of the EU Charter in the context of the GDPR seems circular. It operates on a logic similar to that of the procedural directives in criminal law, as previously explained: Applying the GDPR that gives a more concrete expression a fundamental right to the protection of personal data anchored in the EU Charter^[23] triggers the applicability of the EU Charter itself. Compliance with the GDPR is therefore considered respect for this fundamental right, and vice versa, a breach of the GDPR can constitute a violation of this right. Consequently, in the context of automated decision-making under the GDPR, invoking the right to the protection of personal data can appear to be conceptually redundant. Nevertheless, a question arises of whether the right to the protection of personal data (potentially in combination with the right to privacy enshrined in Article 7) can play a distinct role alongside the provisions of the GDPR, and whether invoking this right could add substantial value to a legal dispute^[24].

As the CJEU insists, fundamental rights review cannot be reduced into the review of legislative compliance, because the EU Charter serves as the ultimate constitutional benchmark against which all legislative and administrative action must be assessed. As emphasized by the CJEU, e.g. in *Digital Rights Ireland*^[25] or *Schrems I*^[26], secondary legislation such as the GDPR must be interpreted in light of Articles 7 and 8 of the EU Charter. It follows that the right to the protection of personal data enshrined in the EU Charter retains autonomous normative significance even where the administrative action formally complies with the GDPR. In practice, there may be a situation in which an authority or a court interprets a GDPR provision in light of the Charter's rights or, when the GDPR text appears ambiguous or incomplete, it attempts to address a potential normative lacuna through fundamental rights.

In case of the use of ADM in national administrative decision-making, it is possible to imagine a scenario where narrow reading of Article 22 GDPR and the associated safeguards for the automated processing of personal data, including the right to explanation, do not offer adequate protection of privacy and personal data of the individual, and therefore, these provisions should be interpreted more broadly to provide such protection^[27]. Article 8 of the EU Charter can also be

invoked if a national authority interprets Article 22 GDPR too narrowly, arguing that it applies only to fully automated administrative decisions, which would exclude partially automated decisions from the given safeguards, leaving individuals without effective legal protection. There might also be a claim that, although the automated decision-making appears to be formally in compliance with the GDPR, it interferes with the fundamental rights to privacy and to the protection of personal data in a disproportionate manner or even violates the essence of these rights. The Charter provision could also be used to challenge the legality of Article 22 GDPR and other relevant provisions. In all these scenarios, the right to the protection of personal data constitutes not only the legal basis for the GDPR, but also a benchmark for assessing the legality of its provisions and a standard for interpreting how automated decision-making should be employed in administrative practice^[28]. Furthermore, the EU Charter embeds a duty of a proportionality assessment that goes beyond formal compliance with the GDPR^[29]. Even if the processing of personal data complies with the legal basis under Articles 6 and 9 GDPR and the automation respects safeguards stated by Article 22 GDPR, there is still a possibility that such an action would interfere with Article 8 EU Charter in a disproportionate manner^[30]. As emphasised by the CJEU in *Digital Rights Ireland*^[31], any limitations on the right to data protection must be strictly necessary and proportionate.

In contrast to the GDPR, the AIA does form a legal basis for processing personal data. Nonetheless, as recital 69 AIA reminds, «[t]he right to privacy and to protection of personal data must be guaranteed throughout the entire lifecycle of the AI system». This means that the compliance with Article 8 of the EU Charter must be ensured from initial design and development of the tool to its deployment and ongoing use within national administrative decision-making. In practice, this requires that using AI tools is in compliance with the right to the personal data protection, but also with the GDPR that further implements that right.

At the initial design and development stage, the right to data protection influences the purpose for which the AI tool is created, the categories of data used, and the entire model architecture. For instance, the categories of data incorporated into the tool must adhere to the principles of purpose limitation and data minimisation (Article 5 of the GDPR), and sensitive data may only be

used if it is strictly necessary^[32]. The entire system must also be traceable to enable data subjects to exercise their rights of access, rectification and objection. During the training stage, it is necessary to ensure data cleaning and bias testing and to document the whole process. If large datasets are reused, compliance with Article 6(4) GDPR is required, which stipulates that the original purpose of data collection must not be exceeded disproportionately. Finally, when the AI tool is deployed within actual administrative decision-making, the right to personal data protection requires that individuals affected by AI-assisted decisions have meaningful safeguards. This includes the right to be informed about the automated data processing through an AI tool in accordance with Articles 13, 14, 15 and 22 GDPR. Furthermore, it must be possible to contest such a decision and to alter the outcome of the processing based on human oversight. Therefore, if the national authority uses a simpler ADM tool to process personal data, or deploys a more sophisticated AI tool, the right to the protection of personal data, as enshrined in the EU Charter, will act as a normative constraint throughout the entire process, from developing the tool to making the final decision. This development is likely to lead to a certain degree of harmonization in the national procedures and technology-related practices of administrative authorities. Regardless of the area of administrative law in question (and whether it is within the scope EU law), national authorities must comply with the normative constraints arising from the EU Charter. The fundamental right to the protection of personal data therefore has horizontal effect. All ADM or AI tools must be developed in a manner that ensures compliance with this right, which must also be reflected by the procedural frameworks within which the ADM-based or AI-assisted outputs are generated, assessed and relied upon. Subsequently, national courts are responsible for supervising this alignment and verifying that the protection guaranteed by the EU Charter is effectively realized in practice.

4.1.2. Non-discrimination

Unlike the right to the protection of personal data, the principle of non-discrimination (Article 21 of the EU Charter) is not fully imprinted into the GDPR let alone the AIA. While the principle of non-discrimination is reflected

in several provisions of these regulations, its incorporation remains partial. For instance, Article 9 of the GDPR prohibits the processing of special categories of personal data that may reveal characteristics commonly associated with discrimination, such as racial or ethnic origin, political opinions, religious beliefs, or sexual orientation. In the context of automated decision-making, data controllers are encouraged to implement appropriate safeguards to prevent discriminatory outcomes, as noted in Recital 71. The AIA is more explicit in this regard because its recitals repeatedly affirm that its rules must be non-discriminatory and applied in a non-discriminatory manner. Moreover, Article 77 grants national authorities charged with protecting fundamental rights – including the principle of non-discrimination – access to the technical documentation required by the AIA. However, even though both regulations seek to align with the principle of non-discrimination, compliance with them does not exclude the possibility that this principle will be violated anyway. There is a particular risk that ADM tools may be designed in a way that enables discriminatory outcomes^[33]. Similarly, AI systems that are trained using ostensibly neutral data may nevertheless produce discriminatory effects due to historical biases encoded in the training data or its labelling^[34].

This creates an opportunity for greater engagement with Article 21 of the EU Charter when administrative authorities at the national level use automation. In such cases, the EU Charter can serve as an important normative source for safeguarding the non-discriminatory outcomes of automated administrative decision-making, particularly when there is a risk that automation or the use of AI could have a negative and disproportionate impact on certain individuals based on their race, gender, or sexual orientation. Article 21 may not only be invoked *ex post* – in a review procedure in order to correct an allegedly discriminatory administrative decision, but it may and should be used *ex ante* as a standard governing the design and parameters of the ADM or AI tool. The EU Charter will therefore intervene into the otherwise technical dimension of the algorithm and setting of the specific software used as an ADM tool or of the training and functioning of an AI system.

In its case law, the Court of Justice has already highlighted the legal relevance of the sole design and programming of tools used for automated processing of data when it insisted that «*the pre-established models and criteria on which that type of*

data processing are based should be, first, specific and reliable, making it possible to achieve [intended] results»^[35]. This implies that both programming an ADM tool and building an AI system have a considerable normative significance, because the way how these tools are designed predetermines the outcome of the process^[36]. Consequently, when developing ADM or AI tools, the principle of non-discrimination must be embraced sensitively, since flaws in design or training can induce bias in the outcome of the decision-making process, which will render the administrative decision in violation of the EU Charter. As scholars tend to agree, ADM tools are highly susceptible to discrimination because it is practically impossible to eliminate all potential biases^[37]. AI tools based on machine learning techniques, in particular, can absorb and replicate real-world discrimination. From a technical point of view, problems with the tool can arise at various stages of its development. For example, if the sample is unrepresentative, the patterns identified by the tool will also be unrepresentative. If incorrect features are assigned to data, the tool will learn from these errors. Ultimately, problems will emerge if characteristics within the data that should not influence the outcome are given incorrect or biased labels^[38]. The technical development of the tool and the quality of the training data are therefore of a high legal relevance. If a national administrative authority decides to deploy an ADM or AI tool within its decision-making, it must ensure that the tool – regardless of whether it has been developed by the authority itself or procured from a private company – complies with all necessary rules to avoid discrimination. As the CJEU has already highlighted, automated analysis cannot use sensitive personal data, which are defined in Article 9 of the GDPR, such as racial or ethnic origin, political opinions, religious or philosophical beliefs, trade-union membership, or data concerning health and sex life or sex preferences, because such automated processing would violate Articles 7, 8 and 21 of the EU Charter^[39]. Consequently, this should apply not only to the actual processing of data, but also to the designing and programming of the tool, including training of the AI tool. Furthermore, the CJEU insists on regular assessment of the tool and its results *«to ensure that those pre-established models and criteria and the databases used are reliable and up to date»*^[40]. This requires that an authority relying upon ADM or AI tools performs their consistent monitoring and evaluation to avoid discriminatory outputs.

As with the right to personal data protection, compliance with the right to non-discrimination will probably lead to converging standards across Member States regarding system design, transparency, oversight and remedies for individuals affected by automated decisions.

4.2. Fundamental procedural rights

The EU Charter and its procedural standards constitute an extension of legislation, influencing administrative procedures at a national level. The administrative authority must comply with the right to good administration enshrined in Article 41 of the EU Charter whenever automated processing of personal data takes place within administrative decision-making or a high-risk AI system is deployed. Although this provision formally addresses only EU institutions, the Court of Justice considers the principle of good administration to be part of the general principles of EU law^[41]. Therefore, the content of Article 41 should also apply to national authorities. Furthermore, administrative decisions that rely on ADM tools or AI systems under the GDPR or the AIA will be subject to judicial review in accordance with Article 47 of the EU Charter, which enshrines the right to effective judicial protection.

4.2.1. Right to good administration

The right to good administration as enshrined in Article 41 of EU Charter guarantees that every person has the right to have their affairs handled impartially, fairly, and within a reasonable time. It also includes specific procedural entitlements, such as the right to be heard, the right of access to one's file, and the duty of the administration to give reasons for its decisions and the duty of care, which means that administrative authorities must «*examine carefully and impartially all the relevant aspects of the individual case*»^[42].

Overall, this right extended to the national authorities through general principles establishes the expectation of procedural fairness whenever national authorities implement EU law. In the context of automated or AI-assisted decision-making, this requires the individual to be informed that an ADM or AI tool has been used^[43]. Furthermore, all relevant facts must be carefully assessed, and the standards must not be lower than those applied when the entire case is processed

by human officials. The right to good administration also requires human oversight of the ADM-based or AI-assisted outcome^[44]. This oversight must be effective, in the sense that the outcome must be modified if inaccuracies are detected^[45].

A part of the right and the principle of good administration, which is particularly important in the context of automation and AI, is the duty of public authorities to give reasons for their decisions. Knowing the reasons on which the administrative decision is based is not only important for the affected individuals, but it is also crucial at the stage of judicial review of such a decision. As the CJEU puts it, the duty to give reasons is «*a corollary of the principle of respect for the rights of the defence, which is a general principle of EU law*»^[46]. An individual affected by an administrative decision must be able to understand its grounds, so the reasons for the decision must be specific and concrete^[47]. In parallel, a court examining legality of that decision must also be able to reconstruct the reasoning that led to that decision, because a substantive review can be performed only if such a reasoning is comprehensible for the court. In the words of the CJEU, the duty to give reasons «*is an essential procedural requirement that must be distinguished from the question whether the reasoning is well founded, which is a matter of the substantive legality of the contested act*»^[48].

The duty to give reasons is closely linked to the concept of explanation^[49], which is enshrined both in the GDPR and the AIA^[50]. Under these regulations, the administrative authority must provide meaningful information of how the tool or algorithm used for automated data processing works^[51]. As the Court of Justice put in case *Dun and Bradstreet*, the right to meaningful information must be understood «*as a right to an explanation of the procedure and principles actually applied in order to use, by automated means, the personal data of the data subject with a view to obtaining a specific result*», and such explanation «*must be provided by means of relevant information and in a concise, transparent, intelligible and easily accessible form*»^[52]. However, technical explanation is not enough, and it is necessary to provide “legal explanation”, i.e. information which can be assessed from a legal point of view^[53]. In the words of the CJEU, the duty to provide meaningful information «*cannot be satisfied either by the mere communication of a complex mathematical formula, such as an algorithm, or by the detailed description of all the steps in automated decision-making, since none of*

those would constitute a sufficiently concise and intelligible explanation»^[54]. Therefore the requested standard should be the “human standard for explanation”^[55]. This means that any output generated by or with the help of ADM or AI tools must be legible and meaningful to a human being, who must then be able to assess its legality.

However, as remarked by Hildebrandt, «*an explanation is not the same as a justification*»^[56]. While the GDPR and the AIA require sufficient transparency of the criteria, formulas, parameters and other components of the ADM tool, the requirement for proper justification of the decision goes further. Simply making the decision-making process transparent will not guarantee that the decision will clearly explain the reasons behind it. Therefore, an ADM or AI-assisted administrative decision must also contain a traditional legal justification. This justification must go beyond an explanation of how the tool works or how the data is processed. It must present what facts were considered, how they were assessed, how they are linked to legal provisions and how they support the final decision. In other words, although the duty to provide an explanation and the duty to give reasons may overlap in certain practical respects, they are not equivalent. The requirement to state reasons for an administrative decision entails a more demanding standard of legal justification than merely explaining the technical functioning of the algorithm or the tool.

Conceptually, it is also useful to distinguish between two levels of automated output, which may overlap in practice but impose different requirements in terms of justification and, consequently, in the assessment of the legality of administrative decisions. Firstly, the tool generates either a partial result that forms part of the final decision, or the final decision itself. At this level, it is required to provide information on the input data, the tool's parameters and the processing method. Ultimately, a court will evaluate whether these are in accordance with the law.

At the second level, the tool generates a text containing the reasoning and legal justification requirements, including the presentation of relevant facts and their evaluation, the interpretation of applicable legal provisions, and a description of how these lead to the final decision. At this stage, the court will conduct a traditional review to assess whether the automatically generated text is coherent and understandable, and free from internal contradictions or illogical

considerations. The court will treat the text in the same way, regardless of whether it is generated based on pre-set rules or through machine learning AI.

4.2.2. Right to effective judicial protection

As already mentioned above, the requirements for the right to effective judicial protection are closely linked to the principle of good administration. For judicial review of an ADM-based or AI-assisted administrative decision to be effective, a certain quality of the administrative output must be ensured^[57]. This requires traceability of automated decision-making, auditability of algorithms, preservation of documentation and intelligible reasoning that can be reconstructed by a judge. Therefore, the demands of judicial review under EU law, stemming from Article 47 of the EU Charter, reinforce the demands on administrative decision-making assisted by new digital tools^[58].

At the same time, effective judicial protection is a right that individuals affected by EU law can invoke directly. Article 47 therefore highlights the necessity for those individuals to be able to meaningfully challenge ADM-based or AI-assisted administrative decisions. This is also linked to the requirements of good administration, since affected persons must be provided with all relevant information about how the tool contributed to the final decision, and they must have an effective means of contesting any inaccuracies that may have arisen regard to both input data and the method how they were processed.

In other words, the effectiveness of judicial protection is still dependent on the quality of explanations and reasoning provided in the administrative decision^[59]. Both the affected individual and the judge conducting the review must be able to understand the administrative decision and its basis, regardless of whether it was made entirely by a human official or with the help of digital tools. In this regard, Article 47 puts further pressure on administrative authorities to use systems that comply with explainability requirements, to maintain safeguards for effective human oversight of the process, allowing for necessary corrections, and to ensure that the final decision is accompanied by proper legal justification^[60].

In this regard, the right to effective judicial protection, enshrined in the EU Charter, plays an important systemic and constitutional role^[61]. While secondary legislation, such as the GDPR and the AIA, provides for obligations mostly

related to the technical aspects of automation in administrative decision-making, Article 47 sets a higher standard, filling potential normative gaps and guiding the interpretation of the rules embedded in the legislation. It also invalidates practices that could render judicial review ineffective. Overall, the requirements of effective judicial protection should ensure that national administrative action employing ADM or AI tools complies with the rule of law^[62].

Using ADM or AI tools in national administrative decision-making processes can not only lead to the convergence of administrative processes and administrative adjudication across Member States. It can also have a spillover effect. Since the deployment of technology governed by EU law extends the scope of EU law to otherwise internal situations, the category of cases in which national courts may submit preliminary references according to Article 267 TFEU is enlarged. National courts can pose questions related to the rules on the proper use of technology as set out in the relevant EU legislation. Nonetheless, they can also seek guidance on how national settings and conditions should be reconciled with the requirements of effective judicial protection. This could create a new channel for rule-of-law-oriented preliminary references, particularly in Member States where judges seek clarification on their role under EU law and its institutional implications^[63].

5. Conclusion

The current or future use of ADM or AI tools by national administrative authorities cannot escape the normative force of the EU Charter. Whenever a national authority deploys modern technologies that are governed by EU legislation, it is acting within the scope of EU law. And whenever a national authority acts within the scope of EU law, it must also comply with the EU Charter. Consequently, regardless of whether the authority is dealing with cases that are substantively governed by EU law, or with so-called purely internal cases, the deployment of certain modern technologies in the decision-making process triggers the applicability of the EU Charter. The EU Charter is therefore apt to influence the technological and procedural aspects of such decision-making, such as the design of ADM or AI tools, the modalities of data storage and processing, the automated generation of outputs and the structuring of legal reasoning

produced by these tools or with their assistance. The EU Charter also affects the standard of the duty of administrative authorities to give reasons and the conditions under which administrative decisions are subject to judicial review.

Therefore, the EU Charter has the potential to be a catalyst for the technology-driven harmonization of administrative practices across the Member States, since it sets standards for the design and development of ADM and AI tools, promotes transparency in processes and emphasizes the requirement of reviewability of ADM-generated or AI-assisted administrative decisions. The EU Charter hence not only constrains national administrations, but it also influences the emerging model of automated administrative decision-making across the EU.

Conceptually, this produces a specific form of unification of administrative rules and practices that is not based on traditional techniques of integration. It is not harmonization through comprehensive procedural codes or partial procedural rules. It also differs from harmonization through mutual recognition or through coordination among national authorities. It is more accurate to speak of a technology-driven convergence of rules that flows rather indirectly from EU legislation governing the technologies used in administrative decision-making and from the accompanying duty to respect the EU Charter. In order to comply with both sets of rules, national authorities must inevitably adapt their processes, institutional settings and potentially their entire decision-making culture. In this way, the extension of the scope of application of the EU Charter through the deployment of digital tools governed by EU law can lead to a partial unification of administrative processes at the national level, which is derived from the obligation to comply with fundamental rights and principles anchored in EU law.

1. Court of Justice, judgment of 26 February 2013, C-617/10, Åkerberg Fransson, ECLI:EU:C:2013:105, p. 21.
2. Cf. J. Weiler, *The constitution of Europe: «do the new clothes have an emperor?» and other essays on European integration*, Cambridge University Press, Cambridge; New York, 1999, pp. 120–121.
3. Court of Justice, judgment of 13 July 1989, 5/88, *Wachauf*, ECLI:EU:C:1989:321.
4. Court of Justice, judgment of 18 June 1991, 260/89, *ERT*, ECLI:EU:C:1991:254.
5. The terms “agency situation” and “derogation situation” used e.g. in K. Lenaerts, J. A. Gutierrez Fons, *The Place of the Charter in the EU Constitutional Edifice*, in S. Peers et

- al.*, *The EU Charter of Fundamental Rights: A Commentary*, Hart, Oxford, 2014, pp. 1567–1568.
6. L. Besselink, *The Member States, the National Constitutions and the Scope of the Charter*, in *Maastricht Journal of European and Comparative Law*, 8, 2001, pp. 68–80; A. Cuyvers, *The Scope, Nature and Effect of EU Law*, in E. Ugirashebuja *et al.*, *East African Community Law*, Brill, Nijhoff, 2017, pp. 161–181; B. De Witte, *The scope of application of the EU Charter of Fundamental Rights*, in M. González Pascual *et al.*, *The right to family life in the European Union*, Routledge, London, 2017; M. Dougan, *Judicial review of Member State action under the general principles and the Charter: Defining the scope of Union law*, in *Common Market Law Review*, fasc. 52, Issue 5 2015, pp. 1201–1245; B. Pirker, *Mapping the Scope of Application of EU Fundamental Rights: A Typology*, in *European Papers - A Journal on Law and Integration*, 3, 2018, pp. 133–156.
 7. Cf. Court of Justice, Opinion of AG Bobek of 7 September 2017, C-298/16, *Ispas*, ECLI:EU:C:2017:650, p. 32.
 8. Court of Justice, judgments of 21 December 2011, C-411/10 and C-493/10, *N.S. and Others*, ECLI:EU:C:2011:865, paragraphs 65 to 68; and of 13 June 2017, C-258/14, *Florescu and Others*, ECLI:EU:C:2017:448, p. 48.
 9. Court of Justice, judgment of 26 February 2013, C-617/10, *Åkerberg Fransson*, ECLI:EU:C:2013:105.
 10. Court of Justice, judgment of 8 November 2012, C-40/11, *Iida*, ECLI:EU:C:2012:69, p. 79, referring to C-309/96, *Annibaldi*, ECLI:EU:C:1997:631, paragraphs 21 to 23. See also judgment of 5 May 2022, C-83/20, *BPC Lux 2*, ECLI:EU:C:2022:346, p. 27; and of 22 January 2020, C-177/18, *Baldonado Martín*, ECLI:EU:C:2020:26, p. 59, and the case-law cited.
 11. Court of Justice, judgment of 28 October 2021, C-319/19, *Komisija za protivodejstvie na korupcijata i za otnemane na nezakonno pridobitoto imuštstvo*, ECLI:EU:C:2021:883, p. 44; of 22 April 2021, C-485/19, *Profi Credit Slovakia*, ECLI:EU:C:2021:313, p. 37; of 22 January 2020, C-177/18, *Baldonado Martín*, ECLI:EU:C:2020:26, paragraphs 57 to 59; and of 16 July 2020, C-686/18, *Adusbef and Others*, ECLI:EU:C:2020:567, paragraphs 51 and 52.
 12. Court of Justice, judgment of 9 November 2017, C-298/16, *Ispas*, ECLI:EU:C:2017:843.
 13. *Ibid.*, p. 26. Cf. Judgment of 17 December 2015, C-419/14, *WebMindLicenses*, ECLI:EU:C:2015:832, p. 84; and of 22 October 2013, C-276/12, *Sabou*, ECLI:EU:C:2013:678, p. 38.
 14. Directive (EU) 2016/343 of the European Parliament and of the Council of 9 March 2016 on the strengthening of certain aspects of the presumption of innocence and of the right to be present at the trial in criminal proceedings, OJ L 65, 11.3.2016, p. 1-11.
 15. *Ibid.*, recital 9.
 16. Judgment of 13 June 2019, C-646/17, *Gianluca Moro*, ECLI:EU:C:2019:489.
 17. Directive 2012/13/EU of the European Parliament and of the Council of 22 May 2012 on the right to information in criminal proceedings, OJ L 142, 1.6.2012, p. 1-10.

18. As to the opposite argument, cf. Opinion of AG Campos Sánchez Bordona from 10 July 2019, C-467/18, *Rayonna prokuratura Lom*, ECLI:EU:C:2019:590, p. 1 (not endorsed by the Court of Justice).
19. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation), OJ L 119, 4.5.2016, pp. 1–88.
20. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act), OJ L, 2024/1689, 12.7.2024.
21. Articles 14(2)(g) and 15(1)(h) GDPR.
22. Judgment of 13 June 2019, C-646/17, *Gianluca Moro*, ECLI:EU:C:2019:489.
23. Court of Justice, judgments of 29 July 2019, C 40/17, *Fashion ID*, ECLI:EU:C:2019:629, p. 66; and of 28 April 2022, C 319/20, *Meta Platforms Ireland*, ECLI:EU:C:2022:322, p. 73.
24. Cf. I. Bantekas, V. Bratsiakou, *Automated Decision-Making Systems and Black Box Challenges Under European Union Administrative Law*, in *Fordham International Law Journal*, 49, 1/2026, pp. 28–29.
25. Court of Justice, judgment of 8 April 2014, C-293/12 and C-594/12, *Digital Rights Ireland*, ECLI:EU:C:2014:238.
26. Court of Justice, judgment of 6 October 2015, C-362/14, *Schrems I*, ECLI:EU:C:2015:650.
27. Cf. A. Roßnagel, P. Richter, *Art. 5. Principles relating to processing of personal data*, in I. Spiecker genannt Döhmann *et al.*, *General Data Protection Regulation: Article-by-Article Commentary*, Nomos Verlagsgesellschaft mbH & Co. KG, Baden-Baden, 2023.
28. Cf. G. Rudolf, P. Kovač, *The Role of Automated Decision-Making in Modern Administrative Law: Challenges and Data Protection Implications*, in *Central European Public Administration Review*, 22, 2/2024, pp. 83–108, p. 92.
29. H. C. H. Hofmann, *Automated Decision-Making (ADM) in EU Public Law*, in H. C. H. Hofmann, F. Pflücke, *Governance of automated decision making and EU law*, Oxford University Press, Oxford, 2024, pp. 1–32, p. 18.
30. Cf. L. Enqvist, M. Naarttijärvi, *Discretion, Automation, and Proportionality*, in M. Suksi, *The Rule of Law and Automated Decision-Making*, Springer International Publishing, Cham, 2023, pp. 147–178.
31. Court of Justice, judgment of 8 April 2014, C-293/12 and C-594/12, *Digital Rights Ireland*, ECLI:EU:C:2014:238.
32. Cf. G. Rudolf, P. Kovač, *The Role of Automated Decision-Making in Modern Administrative Law: Challenges and Data Protection Implications*, p. 91, cit.
33. Cf. U. Franke, *First- and Second-Level Bias in Automated Decision-making*, in *Philosophy*

- § *Technology*, 35, 2/2022, p. 21; A. Vetrò *et al.*, *A data quality approach to the identification of discrimination risk in automated decision making systems*, in *Government Information Quarterly*, 38, 4/2021; R. Xenidis, *Beyond bias: algorithmic machines, discrimination law and the analogy trap*, in *Transnational Legal Theory*, 14, 4/2023, pp. 378–412.
34. A.Z. Huq, *Constitutional Rights in the Machine Learning State*, in *Cornell Law Review*, 105, 2020, pp. 1875–1954, pp. 1924–1925. Referring to S. Corbett-Davies *et al.*, *The Measure and Mismeasure of Fairness*, in *Journal of Machine Learning Research*, 24, 2023, pp. 1–117, pp. 17–19.
 35. Court of Justice, judgment of 6 October 2020, C-511/18 – C-520/18, *La Quadrature du Net*, ECLI:EU:C:2020:791, p. 180, referring to Opinion of 26 July 2017, C-1/15, *EU-Canada PNR Agreement*, ECLI:EU:C:2017:592, p. 172.
 36. Cf. O. Mir, *Algorithms, Automation and Administrative Procedure at EU Level*, in H. C. H. Hofmann, F. Pflücke, *Governance of automated decision making and EU law*, Oxford University Press, Oxford, 2024, pp. 53–78, p. 58.
 37. E. Abrusci, R. Mackenzie-Gray Scott, *The questionable necessity of a new human right against being subject to automated decision-making*, in *International Journal of Law and Information Technology*, 31, 2/2023, pp. 114–143, p. 124; J. Adams-Prassl *et al.*, *Directly Discriminatory Algorithms*, in *The Modern Law Review*, 86, 1/2023, pp. 144–175, p. 144.
 38. A.Z. Huq, *Constitutional Rights in the Machine Learning State*, pp. 1924–1925, cit. Referring to S. Corbett-Davies *et al.*, *The Measure and Mismeasure of Fairness*, pp. 17–19, cit.
 39. Court of Justice, judgment of 6 October 2020, C-511/18 – C-520/18, *La Quadrature du Net*, ECLI:EU:C:2020:791, p. 165.
 40. Ibid., p. 182; referring to Opinion of 26 July 2017, Opinion 1/15, *EU-Canada PNR Agreement*, ECLI:EU:C:2017:592, paragraphs 173 and 174.
 41. Court of Justice, judgment of 8 May 2019, C-230/18, *PI v Landespolizeidirektion Tirol*, ECLI:EU:C:2019:383, paragraphs 56–58.
 42. Court of Justice, judgment of 21 November 1991, C-269/90, *Technische Universität München*, ECLI:EU:C:1991:438, p. 14.
 43. Cf. Recital 71 GDPR.
 44. Cf. O. Mir, *Algorithms, Automation and Administrative Procedure at EU Level*, pp. 63–65, cit.
 45. Court of Justice, judgment of 6 October 2020, C-511/18 – C-520/18, *La Quadrature du Net*, ECLI:EU:C:2020:791, p. 182.
 46. Court of Justice, judgments of 22 November 2012, C-277/11, *M.*, ECLI:EU:C:2012:744, p. 88; and of 11 December 2014, C-249/13, *Boudjlida*, ECLI:EU:C:2014:2431, p. 38.
 47. Ibid.
 48. Court of Justice, judgments of 5 May 2022, C-54/20 P, *Commission v Missir Mamachi di Lusignano*, ECLI:EU:C:2022:349, p. 69; of 10 March 2022, C-167/19 P and C-171/19 P, *Commission v Freistaat Bayern and Others*, ECLI:EU:C:2022:176, p. 77; or of 30

- November 2016, C-486/15 P, *Commission v France and Orange*, ECLI:EU:C:2016:912, p. 79.
49. Cf. J. C. Covilla, *Artificial Intelligence and Administrative Discretion: Exploring Adaptations and Boundaries*, in *European Journal of Risk Regulation*, 16, 1/2025, pp. 36–50.
 50. Article 13(2)(f), Article 14(2)(g), Article 15(2)(h) GDPR; Article 86(1) AIA. Cf. S. Wachter *et al.*, *Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation*, in *International Data Privacy Law*, 7, 2/2017, pp. 76–99, pp. 82–83.
 51. O. Mir, *The AI Act from the Perspective of Administrative Law: Much Ado About Nothing?*, in *European Journal of Risk Regulation*, 16, 1/2025, pp. 63–75.
 52. Court of Justice, judgment of 27 February 2025, C-203/22, *Dun & Bradstreet Austria*, ECLI:EU:C:2025:117, p. 58.
 53. H. P. Olsen *et al.*, *What's in the Box? The Legal Requirement of Explainability in Computationally Aided Decision-Making in Public Administration*, in H.-W. Micklitz, *Constitutional Challenges in the Algorithmic Society*, Cambridge University Press, Cambridge, 2022, p. 222. See also Z. Zódi, *Algorithmic explainability and legal reasoning*, in *The Theory and Practice of Legislation*, 10, 1/2022, pp. 67–92.
 54. Court of Justice, judgment of 27 February 2025, C-203/22, *Dun & Bradstreet Austria*, ECLI:EU:C:2025:117, p. 59.
 55. H. P. Olsen *et al.*, *What's in the Box? The Legal Requirement of Explainability in Computationally Aided Decision-Making in Public Administration*, p. 226, cit.
 56. M. Hildebrandt, *Algorithmic regulation and the rule of law*, in *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 376, 2018.
 57. H. C. H. Hofmann, *Automated Decision-Making (ADM) in EU Public Law*, p. 20, cit.
 58. Cf. J.C. Covilla, *Artificial Intelligence and Administrative Discretion: Exploring Adaptations and Boundaries*, pp. 42, 47, cit.
 59. Cf. R. Williams, *Rethinking Administrative Law for Algorithmic Decision Making*, in *Oxford Journal of Legal Studies*, 42, 2/2022, pp. 468–494, p. 481.
 60. Cf. O. Mir, *The impact of the AI Act on public authorities and on administrative procedures*, in *CERIDAP - Rivista Interdisciplinare Sul Diritto Delle Amministrazioni Pubbliche*, 4, 2023, p. 249.
 61. Cf. D.-U. Galetta, H. C. Hofmann, *Evolving AI-based automation-the continuing relevance of good administration*, in *European Law Review*, 48, 6/2023, pp. 617–635, p. 628.
 62. Cf. V. Roeben, *Judicial Protection as the Meta-norm in the EU Judicial Architecture*, in *Hague Journal on the Rule of Law*, 12, 1/2020, pp. 29–62.
 63. See, e.g., Court of Justice, judgment of 20 March 2020, C-558/18 and C-563/18, *Miasto Łowicz*, ECLI:EU:C:2020:234.

Automation of Administrative Proceedings in Slovakia: Current State and Future Perspectives

Lukáš Jančát, Diana Repiščáková

DOI: 10.13130/2723-9195/2026-2-35

L'automazione dei procedimenti amministrativi, compreso l'uso dell'intelligenza artificiale, è da tempo un argomento ampiamente dibattuto in numerosi Stati europei. Nella Repubblica Slovacca, tuttavia, un dibattito accademico e professionale approfondito e sistematico su tali questioni è emerso solo di recente. Lo scopo di questo contributo è dunque di contribuire al dibattito accademico emergente, presentando l'attuale quadro giuridico che disciplina la automazione dei procedimenti amministrativi nella Repubblica Slovacca, visto attraverso la lente della legislazione esistente e proposta, nonché della giurisprudenza fondamentale dei tribunali di ultimo grado.

The automation of administrative procedures, including the use of artificial intelligence, has long been a widely discussed topic in numerous European states. In the Slovak Republic, however, a thorough and systematic scholarly and professional debate on these issues has only emerged relatively recently. The aim of this article is to contribute to the emerging scholarly debate by presenting the current legal framework governing the automation of administrative proceedings in the Slovak Republic, viewed through the lens of existing and proposed legislation, as well as the key case law of the highest judicial authorities.

Summary: 1. Introduction.- 2. Status quo in the Slovak Republic.- 2.1. Political visions vs. normative reality.- 2.2. eKasa judgment as the flagship of Slovak case law regarding automated decision-making.- 2.2.1. Lack of legality.- 2.2.2. Lack of transparency.- 2.2.3. Data protection concerns.- 3. AI and the new legal framework in Slovakia: Legal basis for automation on the horizon?- 3.1. Draft Act on the Organisation of State Administration in the Field of Artificial Intelligence (2025).- 3.2 Draft Act on Artificial Intelligence and European Data Regulation

(2026).- 4. *Conclusions.*

1. Introduction

The automation of administrative proceedings^[1], including the use of artificial intelligence (hereinafter referred to as “AI”)^[2], is currently a widely debated topic within legal scholarship across Europe^[3]. In response to governmental efforts aimed at automating decision-making, legal scholars have intensified their research activities and have drawn attention to a number of legal challenges associated with the implementation of algorithms in decision-making processes within public administration. Among the most frequently discussed issues are, in particular, the absence of a sufficient legal basis for the automation of decision-making processes^[4]; the need to comply with the guarantees of the right to a fair trial, the right to privacy, and the right to the protection of personal data, as enshrined not only in national law but also at the supranational and international levels of protection^[5]; as well as the requirements of controllability, transparency, and non-discrimination of the algorithms employed^[6].

The Slovak Republic, as a Member State of the European Union, the Council of Europe, and other international organisations from which binding obligations arise, is not immune to these challenges. Unlike in some other states, however, a thorough and systematic scholarly debate on these issues has only recently begun in the Slovak context^[7]. This delayed development is likely related to the fact that, in Slovakia, despite political declarations, it has thus far not been possible to identify any fully automated decision-making process within public administration, nor any legislation aimed at establishing a legal framework enabling the automation of administrative proceedings.

The aim of this article is to contribute to the emerging scholarly debate by presenting the current legal framework governing the automation of administrative proceedings in the Slovak Republic, viewed through the lens of existing and proposed legislation, as well as the key case law of the highest judicial authorities. In order to achieve this objective, the authors employed traditional research methods commonly used in legal scholarship. The authors primarily relied on doctrinal legal analysis of legislation, legislative proposals, explanatory

memoranda, case law, and academic literature, followed by synthesis and generalisation of the key findings.

2. Status quo in the Slovak Republic

2.1. Political visions vs. normative reality

As of the beginning of 2026, the Slovak Republic does not have a dedicated strategy for artificial intelligence. Coordination of AI policy initiatives between ministries and agencies is governed by several mechanisms and tools^[8]. Since 2019, the key strategic document, which is comprehensively focused on guiding the digital transformation of the Slovak Republic in the current decade, is the 2030 Digital Transformation Strategy for Slovakia (hereinafter referred to as “*Digital Strategy*”). The Digital Strategy can be described as a comprehensive cross-sectoral government framework strategy that defines the policy and particular priorities of Slovakia in the context of currently ongoing digital transformation of economy and society under the influence of innovative technologies and global megatrends of the digital era. Thus, it puts primary emphasis on current innovative technologies such as AI, Internet of Things, 5G technology, big data and analytical data processing, blockchain, and high-performance computing. The Digital Strategy identifies the area of public administration as one of the key areas in which it will be necessary to multiply the potential through the aforementioned technologies^[9].

In connection with the digital transformation of public administration, the digital strategy, among other things, declares that by 2030 the Slovak Republic will become a modern country with an effective public administration ensuring smart use of the territory and infrastructure. The key to fulfilling this vision is a significant improvement of the use of data and application of methods such as impact assessment, risk analysis, automated evaluation of cases or applications, as well as predictive planning of future public service capacities^[10]. Despite the fact that the digital strategy for fulfilling the vision explicitly assumes the implementation of a method of automated assessment of cases or applications in the field of public administration, the Slovak legislator has not yet taken any specific legislative steps to create a legal basis for the automation of administrative

proceedings, including the use of AI.

De lege lata, the Slovak general regulation governing administrative proceedings (hereinafter referred to as “*Administrative Procedure Code*”)^[11], as well as special legislation regulating individual types of administrative proceedings^[12], do not contain any specific regulation of automated administrative decision-making, including the use of AI in decision-making. Moreover, the Administrative Procedure Code explicitly links decision-making activity to a human actor.

It should be noted that this legal regulation was systematically developed in the 1960s and, despite subsequent amendments partially reflecting technological progress, it is conceptually based on an anthropocentric understanding of decision-making. This approach is reflected in individual procedural institutions that are tied to an official person who performs procedural acts on behalf of the administrative authority. Examples include the institution of disqualification of an employee of an administrative authority^[13] or the institution of a record (minutes), which presupposes that it is drawn up by a public servant conducting the proceedings^[14]. Another significant manifestation of this anthropocentric concept is the fact that an official stamp and a signature indicating the name, surname, and official position of the authorised person constitute mandatory elements of an administrative decision^[15].

At the same time, the Slovak Republic is a state governed by the rule of law, in which public authorities may act, in accordance with the principle of legality, only within the scope and in the manner prescribed by law^[16]. Furthermore, as a Member State of the European Union, the Slovak Republic is bound by EU law, including the obligation arising from Article 22(1) and Article 22(2)(b) of the General Data Protection Regulation^[17] (hereinafter referred to as “*GDPR*”). This obligation requires that fully automated decision-making in the field of public administration which produces legal effects concerning a data subject or similarly significantly affects them must be based on an explicit legal basis laid down in EU law or in the national law of the Slovak Republic, while simultaneously providing for appropriate safeguards ensuring the protection of the rights, freedoms, and legitimate interests of the data subject^[18].

In view of the anthropocentric nature of the legal regulation of administrative proceedings and taking into account the absence of explicit regulation of automated decision-making in Slovak administrative law, it is possible, in

accordance with a constitutionally conforming and EU-conforming interpretation of the principle of legality, to conclude that fully automated decision-making in the field of public administration is currently not permissible in the Slovak Republic due to the lack of a legal basis.

Under the current legal framework, however, the automation of certain parts of administrative proceedings, or the automation of specific procedural acts within administrative proceedings, is permissible. Nevertheless, even such semi-automated administrative proceedings must always operate within the limits of the principle of legality. Jakab states that if the general psychological model of decision-making is adapted to a legal case, in Slovak conditions automation occurs mainly at the stage of problem recognition and at the stage of fact finding. Specifically, as an example, he states that: *«For example, the information system will automatically assess which new property owners have not filed a tax declaration based on data from the Land Registry. This identified the problem, i.e., failure to file a tax declaration in case the person was required to do so. Thus, the prerequisite for the initiation of sanction proceedings was given. As a result, it can be concluded that under the current legal situation, “partial” automation of decision-making processes in public administration is possible, with automation taking place at some stage of the process, while preserving the interference of human influence on the overall outcome of the decision-making process»*^[19].

2.2. eKasa judgment as the flagship of Slovak case law regarding automated decision-making

To date, Slovak courts have not substantively addressed the issue of the automation of administrative proceedings, including the use of AI in decision-making processes within public administration. The sole notable exception in this regard is the Judgment of the Constitutional Court of the Slovak Republic, 10 November 2021, PL. ÚS 25/2019, commonly referred to in analytical literature as the *“eKasa Judgment”*^[20].

In this judgment, the Constitutional Court of the Slovak Republic (hereinafter referred to as *“Constitutional Court”*) primarily examined the compatibility of the automated collection and processing of personal data with the constitutional and Union legal framework for the protection of the fundamental right to

privacy, private life, and the right to the protection of personal data. The Constitutional Court specifically assessed the compatibility of a machine learning risk assessment algorithm called AIS-R implemented by the Slovak Financial Administration, which processed data from the e-Kasa system^[21] to predict VAT fraud risks.

In the context of this article, the judgment of the Constitutional Court makes it possible to identify three categories of issues that have a fundamental impact on the possibility of automating administrative proceedings in the Slovak Republic. These concern deficiencies in legality, transparency, and the protection of personal data.

2.2.1. Lack of legality

The first doctrinal problem identified by the Slovak Constitutional Court was the insufficient legal basis for the implementation of the AIS-R algorithm. In this context, the Constitutional Court stated that: *«A “silent statute” cannot serve as a legal basis for the automated assessment of an individual, irrespective of whether such assessment results in a formal decision or in administrative inaction. Automated risk assessment cannot be merely the outcome of a managerial decision taken by a public authority; rather, it must be the subject of public debate reflected in the legislative process. The absence of explicit authorisation by the legislature is also manifested in the lack of regulation governing the oversight infrastructure applicable to such assessments. (...)»*^[22].

In the case under review, the Constitutional Court thus identified a violation of the principle of legality. This conclusion of the Constitutional Court further reinforces our conviction that a legal basis for fully automated administrative proceedings is lacking. Accordingly, its findings may also be understood as a warning to the Slovak legislature to ensure, prior to the implementation of automated tools in decision-making processes within public administration, the creation of, or verification of the existence of, an adequate legal basis for both fully automated and semi-automated decision-making.

2.2.2. Lack of transparency

Another problem identified by the Constitutional Court concerned the transparency of automated decision-making, including decisions formally taken by a human official based on profiling carried out by an AIS-R algorithm, the functioning of which is not sufficiently known or understood by the data subject and, in some cases, not even by the public official who issued the decision. In this context, the Constitutional Court expressed the view that the application of technological progress in public administration must not result in an impersonal state whose decisions are inexplicable, unreviewable, and devoid of accountability. As the sole actor capable of doing so, the State must, through specific statutory provisions, ensure that the criteria, models, and interconnected databases used in the context of automated assessments are up to date, reliable, and non-discriminatory. Beyond the general safeguards that are indispensable in the processing of personal data, the Constitutional Court therefore requires the establishment of specific safeguards in the form of: (i) transparency, (ii) individual protection, and (iii) collective oversight^[23]. These safeguards must take into account the particular nature of the procurement and deployment of automated systems whose operation affects individuals and their fundamental rights and freedoms^[24].

First and foremost, the Constitutional Court identified as a specific safeguard the legislative enshrinement of the right to be informed that the conduct of a public authority, e.g. its decision, has been influenced by the use of an automated system. The individual concerned must have the right to be aware of the existence, scope, and consequences of the automated assessment of his or her person, whether through public registers, specific notifications, or other appropriate means. Only where the addressee is clearly informed that he or she is subject to automated assessment to a certain extent can he or she effectively challenge potential errors. In this regard, the Constitutional Court further emphasized that where the technical solution is provided to the State by a third party, the State remains bound by the same transparency requirements. Thus, intellectual property rights, trade secrets, or system security considerations cannot serve as grounds for denying effective access to the information necessary for the protection of individual rights^[25].

Secondly, the Constitutional Court held that the use of automated systems must be subject to independent collective oversight operating both *ex ante* and *ex post*. Such oversight must allow for an assessment of the systems quality, its components, error rates, and shortcomings before it is put into operation, as well as after deployment, e.g. through audits. If the system contains weaknesses or blind spots, public officials working with it must demonstrably be aware of them, as ignorance of such deficiencies may itself give rise to errors. The use of the system must also be properly documented and logged to facilitate both collective oversight and the effective enforcement of individual rights^[26].

Lastly, in the opinion of the Constitutional Court, the guarantee of oversight over automated systems must enable individuals to effectively challenge errors generated by such systems in such a manner as to allow the individual to obtain a modification of an automated decision. It is for the legislature to determine the precise procedural mechanisms to be adopted. However, such mechanisms must be effective, accessible, and entrusted to independent supervisory authorities^[27].

While the matter assessed by the Constitutional Court related specifically to profiling conducted through an AI algorithm in the field of tax law, the Courts reasoning clearly indicates that its implications are considerably broader and likewise apply to the transparency safeguards that must be observed in automated decision-making processes in the sphere of public administration in Slovakia.

2.2.3. Data protection concerns

Last but not least, in the case under review the Constitutional Court identified a problem concerning the extensive collection of personal data in the operation of automated systems used by the Slovak Financial Administration. Although a legal basis for the collection of such data formally existed, it was not entirely clear to what extent and for what purposes public authorities other than the Financial Administration were entitled to use the personal data thus collected. At the same time, the Constitutional Court found certain aspects of the collection of selected data relating to the affected entities to be unconstitutional, on the grounds that such interference with their right to privacy and protection of personal data lacked sufficient legitimacy^[28].

When introducing automated systems into decision-making processes within

public administration, the Slovak legislature should therefore, in the future, ensure not only that an adequate and sufficiently precise legal basis is established for the collection and processing of personal data by automated systems used in administrative proceedings, but also that the legitimacy and proportionality of the scope and purpose of such data collection are duly assessed in light of their impact on the fundamental rights of the data subjects concerned.

3. AI and the new legal framework in Slovakia: Legal basis for automation on the horizon?

According to the Plan of Legislative Tasks of the Government of the Slovak Republic for 2025, a legislative task entitled «*A draft Act regulating the institutional framework, the competences of authorities, and the rights and obligations of relevant entities in connection with the use of artificial intelligence systems*» was scheduled for April 2025^[29]. However, an attempt to locate such a draft reveals that, within the legislative process, only a preliminary information notice was published^[30]. The pre-legislative (preparatory) phase commenced on 1 March 2025 and was completed on 13 October 2025^[31].

From 1 August 2025 to 21 August 2025, an interdepartmental comment procedure was conducted in respect of the draft Act on the organisation of state administration in the field of artificial intelligence^[32]. According to publicly available information, the assessment of the comments has been ongoing since 22 August 2025^[33]. In total, 269 comments were submitted, of which 75 were classified as substantive^[34], which makes amendments to the draft highly likely. Such amendments would in any event be desirable, as the draft revealed a number of deficiencies that require correction. The Act was intended to enter into force on 1 January 2026, however, this did not ultimately occur. To date, the Act has not even been published in the Collection of Laws.

In January 2026, a further Slovak draft Act relating to AI was published, namely the draft Act on Artificial Intelligence and European Data Regulation and on the Amendment and Supplementation of Certain Acts. The interdepartmental comment procedure in respect of this draft was conducted from 27 January 2026 to 5 February 2026. Given the need to ensure the timely implementation of selected European Union regulations, the procedure was conducted in an

accelerated form^[35]. A total of 574 comments were submitted, of which 224 were classified as substantive^[36]. Since 5 February 2026, the comments have been under review. The legislative process concerning this draft Act will be closely monitored, as given the volume of comments submitted – it may reasonably be expected that the final version of the Act will differ from the draft. The new Act is intended to enter into force on 1 May 2026.

3.1. Draft Act on the Organisation of State Administration in the Field of Artificial Intelligence (2025)

According to Section 1 of the draft Act, the Act governs the organisation of state administration in the field of AI pursuant to a special legal instrument, namely the AI Act. The draft Act on the organisation of state administration in the field of artificial intelligence thus implements provisions of the European AI Act. The general part of the explanatory memorandum indicates that, at the national level, it is necessary to legislate in particular for the institutional framework for market surveillance in the field of AI. This includes the designation of the competent national authorities responsible for market surveillance in relation to the use of, in particular, high-risk AI systems and the definition of their competences, the establishment of a single point of contact, the designation of a notifying authority, as well as the conditions for setting up a regulatory sandbox. It is likewise necessary to lay down the powers of the national market surveillance authority, the framework for the imposition of sanctions, and the obligation of cooperation between market surveillance authorities^[37].

An exception is Section 6 of the draft Act, which – contrary to the stated subject matter of the draft – regulates «*certain obligations of operators of high-risk AI systems*». This provision does not fit within the overall concept of the draft Act, as the draft is intended to address the organisation of state administration in the field of AI rather than to impose obligations on operators of high-risk AI systems. It therefore constitutes a criticised deficiency in the delineation of the draft Acts subject matter.

The draft Act does not implement Article 5 of the AI Act, which regulates so-called prohibited AI practices. Pursuant to Article 5(5) of the AI Act, Member States are afforded discretion as to whether, for the purposes of law enforcement,

they will provide for the possibility of fully or partially authorising the use of real-time remote biometric identification systems in publicly accessible spaces. In the Slovak context, the legislative response in this area falls within the remit of the Ministry of the Interior of the Slovak Republic^[38].

The draft Act likewise does not address the protection of fundamental rights in connection with the use of AI systems under Article 77 of the AI Act. The AI Act requires the designation and notification of such bodies vis-à-vis the European Commission. Within the Slovak legal framework, bodies for the protection of fundamental rights and freedoms had already been designated and, moreover, already possessed the necessary competences; therefore, no specific national implementing legislation was considered necessary. The publication and notification of these bodies, namely the Personal Data Protection Office and the Public Defender of Rights, was carried out on 2 November 2024 in accordance with Article 77(2) of the AI Act^[39].

According to the preliminary information notice concerning the legislative task entitled «*A draft Act regulating the institutional framework, the competences of authorities, and the rights and obligations of relevant entities in connection with the use of artificial intelligence systems*», it was envisaged that this list would be expanded in the future to include other relevant institutions, such as the Slovak National Centre for Human Rights^[40]. The same preliminary information also indicated that the envisaged draft Act would regulate in greater detail, at national level, the competences and tasks of the above-mentioned bodies under Article 77 of the AI Act, including the framework for the imposition of sanctions^[41]. It is noteworthy that the draft Act for which only a preliminary information notice was published contemplated a more detailed regulation of the competences of fundamental rights bodies, whereas the subsequently published draft Act on the organisation of state administration in the field of artificial intelligence no longer addresses this issue.

Pursuant to Section 2 of the draft Act, state administration in the field of AI is carried out by:

- (a) the general market surveillance authority,
- (b) the notifying authority, and
- (c) the Ministry of Investment, Regional Development and Informatization of the Slovak Republic (hereinafter referred to as “*the Ministry of Investment*”).

Let us take a closer look at which specific state authorities are to exercise competences in the field of AI. The institutional design is more complex than it may appear at first glance. Under Section 3 of the draft Act, the general market surveillance authority is, as a rule, the Ministry of Investment, unless the exercise of competence is entrusted by special legislation to a sectoral authority. Pursuant to the same provision, the general market surveillance authority also serves as the single point of contact. Section 4 of the draft Act further delineates the competences of the Ministry of Investment as a specific authority, beyond its role as the general market surveillance authority and the single point of contact.

However, the normative text of the draft Act itself does not specify which bodies will constitute the sectoral authorities. Section 5 merely provides that a sectoral market surveillance authority is a public authority designated as such by special legislation. In a footnote, the legislator offers only two illustrative examples of such legislation, pursuant to which the sectoral authority could be, for instance, the Personal Data Protection Office or the National Security Authority. This enumeration is not exhaustive, and the number of such authorities may increase in the future. The explanatory memorandum further indicates that sectoral market surveillance authorities may also include the bodies listed in Section 26 of Act No. 56/2018 Coll. on conformity assessment, the making available of designated products on the market, and on amendments to certain acts, such as the Slovak Trade Inspection, the Transport Authority, or the State Institute for Drug Control^[42].

As regards the question of which body is to act as the notifying authority, the situation becomes genuinely complex. The legislator did not include this information in the normative text of the draft Act, nor does the draft expressly designate any authority as the notifying authority, which required further inquiry. Ultimately, it was identified that the draft Act, *inter alia*, amends Act No. 56/2018 Coll. on conformity assessment, the making available of designated products on the market, and on amendments to certain acts, as amended. Section 3(2) of that Act already provides that the Office for Standards, Metrology and Testing of the Slovak Republic acts as the notifying authority. However, according to the relevant footnote (which is not legally binding), the legal basis for this designation is not (at least not yet) the AI Act. The draft Act proposes to amend the footnote so that the AI Act will also be included among the relevant

legal bases^[43].

Such a legislative approach would also be consistent with Article 28(2) of the AI Act, pursuant to which Member States may decide that the assessment and monitoring referred to in paragraph 1 shall be carried out by the national accreditation body within the meaning of, and in accordance with, Regulation (EC) No 765/2008^[44]. In the Slovak legal framework, this role is performed precisely by the Office for Standards, Metrology and Testing of the Slovak Republic.

Under the AI Act, Member States were required, by 2 August 2025, to designate the notifying authority. It is therefore debatable whether the Slovak Republic has complied with this obligation. In our view, the Slovak Republic has not fulfilled this requirement or, if it has, it has done so only inadvertently since the Act on conformity assessment and the making available of designated products on the market does provide that the Office for Standards, Metrology and Testing of the Slovak Republic acts as the notifying authority, yet not on the basis of the AI Act.

To summarise, the Slovak legal framework is intended, going forward, to distinguish between the general market surveillance authority, namely the Ministry of Investment, and sectoral authorities, such as the Personal Data Protection Office or the National Security Authority. The notifying authority is to be the Office for Standards, Metrology and Testing, and the draft Act further regulates the competences of the Ministry of Investment as a specific authority beyond its role as the general market surveillance authority. In our view, this decentralised model is complex, coordination-intensive, and may in practice lead to conflicts of competence. In order to clarify this intricate legal framework, we have prepared the following table:

Section 7 of the draft Act indicates that the Act is also intended to implement Article 57 of the AI Act, which regulates regulatory sandboxes for AI. Pursuant to Article 3(55) of the AI Act, a regulatory sandbox for AI is defined as: «*a controlled framework set up by a competent authority which offers providers or prospective providers of AI systems the possibility to develop, train, validate and test, where appropriate in real world conditions, an innovative AI system pursuant to a sandbox plan for a limited time under regulatory supervision*». Under Section 7 of the draft Act, the Ministry of Investment shall ensure the establishment of a

regulatory sandbox for AI and its operation and development. The Ministry of Investment is also designated as the competent authority for the purposes of exercising the rights and fulfilling the obligations related to the regulatory sandbox for AI.

Section 8 of the draft Act sets out the competences of the general market surveillance authority. The general market surveillance authority carries out market surveillance by:

- (a) handling complaints,
- (b) conducting inspections,
- (c) imposing interim measures pursuant to Section 10,
- (d) adjudicating administrative offences and imposing fines, and
- (e) exercising other powers under special legislation and under this Act, the purpose of which is to ensure compliance with the conditions laid down in the relevant special legislation.

At this point, particular attention should be paid to the imposition of fines, as the draft reveals a deficiency in the legal framework. Under Section 12, the general market surveillance authority adjudicates administrative offences and imposes fines for breaches of obligations laid down in the special legal instrument – namely the AI Act. As can be seen, however, the Slovak legislator does not specify which conduct is to be sanctioned. The relevant footnote merely refers to the AI Act as such, without even indicating specific provisions or articles. In our view, this approach is inconsistent with the principle of *nullum crimen sine lege certa*.

This is only one of several shortcomings of the draft Act. In addition, the draft is, in our view, drafted in an unclear and insufficiently structured manner, and the organisation of state administration in the field of AI in the Slovak Republic would, under this proposal, be excessively complex and may give rise to practical difficulties in its application.

3.2 Draft Act on Artificial Intelligence and European Data Regulation (2026)

If we compare the draft Act on Artificial Intelligence and European Data Regulation (2026) with the draft Act on the organisation of state administration

in the field of artificial intelligence (2025), it can be concluded that the new draft is based on an entirely different conceptual approach. Whereas the draft Act from the previous year introduced a complex decentralised model, the new draft proceeds from the premise that, in the Slovak context, the most effective arrangement is a model centred on a single central authority combined with already existing state administration bodies^[45]. The new draft Act establishes the Office for Digital Integrity as a new central state administration authority.

At this point, it is necessary to highlight the first problem that arises: as noted above, the new Act is intended to enter into force as early as 1 May 2026. In our view, the *vacatio legis* is insufficient, and the proposed date of entry into force appears unrealistic, particularly in light of the establishment of a new central state administration authority. This deficiency was also repeatedly raised in the comments submitted during the interdepartmental consultation. For the sake of completeness, it should be added that the new draft Act likewise does not implement Article 5 or Article 77 of the AI Act.

Pursuant to Section 3(1) of the draft Act, state administration in the field of artificial intelligence is carried out by:

- (a) the Office for Digital Integrity,
- (b) the market surveillance authorities pursuant to Sections 6 and 7, and
- (c) the notifying authority, namely the Office for Standards, Metrology and Testing of the Slovak Republic.

The Office for Digital Integrity is to independently manage and coordinate the performance of state administration in the field of AI and in the field of data regulation under the relevant EU regulations. In general terms, this means that it carries out supervision and oversight, acts as the single point of contact, serves as the data coordinator, imposes fines for breaches of the EU regulations and of this Act, issues methodological guidance within its remit, and performs other tasks laid down in this Act^[46]. With regard to specific competences in the field of AI, the new Office is to act as the general market surveillance authority pursuant to Article 70 of the AI Act and as the single point of contact. It is also responsible for establishing and operating the regulatory sandbox^[47].

Pursuant to Section 6 of the draft Act, the Office for Digital Integrity acts as the general market surveillance authority where the relevant designated product is an AI system, to the extent that market surveillance competences are not entrusted

to a sectoral market surveillance authority. Under Section 7 of the draft Act, the sectoral market surveillance authorities are the Personal Data Protection Office of the Slovak Republic, in respect of market surveillance over AI systems pursuant to special legislation (the AI Act), and the Ministry of Justice of the Slovak Republic, in respect of market surveillance over AI systems used by courts in the exercise of judicial authority.

The new draft Act thus establishes a new central state administration authority, namely the Office for Digital Integrity. It clearly defines the market surveillance authorities, which are, as a rule, the Office for Digital Integrity, and, where applicable, the sectoral market surveillance authorities expressly designated by the Act. Finally, the new draft Act also explicitly specifies which body acts as the notifying authority. It may therefore be concluded that, compared to last years draft, the new proposal is significantly more transparent, and it is immediately apparent which authority serves as the sectoral market surveillance authority and which acts as the notifying authority. In order to clarify the structure of state administration bodies in the field of AI, we have prepared the following table:

It should be noted that, although the legislative text of the new draft Act is significantly more transparent, the risk of conflicts of competence remains. The new draft establishes a complex structure involving multiple authorities, yet without providing a clear procedural mechanism. During the interdepartmental consultation, a recurring concern related to overlapping competences and the unclear institutional design of the supervisory framework. Several stakeholders pointed out that the proposal creates a fragmented model of state administration while lacking an explicit coordination mechanism and rules for resolving conflicts of competence. Such an arrangement may result in duplicative controls, legal uncertainty and unpredictability for regulated entities, including uncertainty as to which authority is empowered to impose sanctions. In this context, the need to introduce a “single entry point” mechanism was emphasised, in order to reduce administrative burdens and strengthen legal certainty^[48].

As regards fines, they are no longer defined as abstractly as they were under the 2025 draft Act. Nevertheless, the interdepartmental consultation repeatedly raised criticism of the sanctions regime, which a number of stakeholders considered disproportionately strict and overly broad. The comments focused in particular on the need to differentiate the level of fines according to the

seriousness of the breach and the nature of the obligation concerned, while emphasising that enforcement should primarily aim at achieving compliance and remediation rather than immediate repression. At the same time, stakeholders pointed to the need for a transitional period during which supervisory authorities would prioritise methodological guidance and corrective measures over the imposition of sanctions^[49].

Comments concerning insufficient procedural safeguards in the exercise of supervision and the imposition of measures also recurred prominently. Several stakeholders objected in particular to the exclusion of the application of the Administrative Procedure Code, the short time limits for submitting objections, and the possibility of immediate enforceability of measures without an adequate opportunity for the affected party to defend itself. Emphasis was also placed on the need for legal certainty in the service of decisions and for reasonable time to respond, as the proposed procedures may result in interferences with the rights of regulated entities before they are able to avail themselves of effective procedural remedies^[50].

Comments on the regulation of the AI regulatory sandbox also appeared repeatedly, in particular with regard to its practical feasibility and the absence of clear operational rules. Several stakeholders noted that the draft lacks sufficiently articulated principles of transparency, participation and data protection, while also pointing to the absence of adequate financial coverage for the technical infrastructure, without which the sandbox would remain merely a formal instrument. Sector-specific comments further emphasised the need for mandatory cooperation with specialised institutions (for example, in the healthcare sector with the National Health Information Centre), as without access to relevant data and expertise the experimental environment cannot be meaningfully operationalised^[51].

4. Conclusions

The Slovak Republic currently lacks an effective legal framework governing the automation of administrative proceedings. Moreover, the Slovak general regulation on administrative procedure is premised upon an anthropocentric conception of the decision-making process. In observance of the principle of

legality, it must therefore be concluded that, at present, there exists no legal basis in Slovakia for the full automation of administrative proceedings. Semi-automated administrative proceedings are, in principle, permissible; however, they must at all times respect the limits imposed by the principle of legality.

The absence of a sufficient legal basis for the automation of administrative proceedings, including the use of artificial intelligence, may also be inferred from the sole doctrinally significant decision addressing automation in the Slovak Republic, commonly referred to as the “*eKasa Judgment*”. In that decision, the Constitutional Court emphasized not only the necessity of compliance with the principle of legality, but also the requirement to satisfy guarantees of transparency and personal data protection. Although the judgment concerned the assessment of the compatibility of an AI algorithm designed for profiling in the field of tax law, its conclusions are, in our view, undoubtedly applicable in a broader sense, including to the automation of decision-making processes in public administration.

Overall, the Slovak legislative developments in the field of AI demonstrate a clear shift from an initially decentralised and conceptually fragmented approach towards a more centralised model of governance. While the January 2026 draft Act is, at first glance, more coherent and institutionally transparent than the 2025 proposal, the consultation process revealed that significant structural concerns remain unresolved. In particular, recurring comments point to the risk of overlapping competences, insufficient procedural safeguards, and an overly rigid sanctions regime, as well as uncertainty surrounding the practical implementation of regulatory sandboxes. These issues are not merely technical in nature: they directly affect the predictability, legitimacy and enforceability of the emerging AI governance framework.

From the perspective of automation, the ongoing legislative process indicates that Slovakia is moving towards the establishment of a domestic legal basis capable of supporting more advanced forms of AI deployment in both the public and private sectors. However, the effectiveness of such a framework will ultimately depend on whether the final legislation succeeds in ensuring legal certainty, meaningful coordination between supervisory authorities, and procedural guarantees proportionate to the potential impact of AI systems. In this sense, the Slovak approach may be characterised as a legal basis for automation “*on the*

horizon” rather than a fully consolidated and operational framework at the time of writing. Nevertheless, the proposed legislation does not constitute any tangible progress toward fulfilling the vision of automating certain decision-making processes in public administration, as envisaged by the Digital Strategy. Consequently, the establishment of a clear legal basis for the automation of administrative proceedings in the Slovak Republic remains a matter for future development.

1. This article was prepared with the support and is the output of a research project supported by the Scientific Grant Agency VEGA no. 1/0062/25 entitled Automatisation of decision-making processes in public administration.
2. By the automation of administrative proceedings, we understand procedures leading to administrative actions, including the issuance of an individual administrative act, where elements of the procedure are either fully or partially carried out without direct human intervention by means of sophisticated computer software, including AI tools. A similar definition is provided by H.C.H. Hofmann, *Comparative Law of Public Automated Decision-Making. An Outline*, in *CERIDAP*, 1, 2023.
3. See, for example, the special issue of the journal *CERIDAP* 1/2023 or *IJPL* 2/2025, the contents of which are devoted to contributions by authors from across Europe analysing the current state of automation and the use of artificial intelligence in decision-making processes within public administration in individual European states. In addition to these articles, see, for instance E. Csatlós, *Hungarian administrative processes in the digital age: An attempt at a comprehensive examination*, in *Intersections. East European Journal of Society and Politics*, 10(1), 2024, or J. Jakubek-Lalik, *The Challenges of AI in Administrative Law and the Need for Specific Legal Remedies: Analysis of Polish Regulations and Practice*, in *Central European Public Administration Review*, 22(2), 2024. Other authors address this issue from the perspective of European Union law. See, e.g. H.C.H. Hoffman (eds), *Governance of Automated Decision-Making and EU Law*, Edward Elgar Publishing, Cheltenham/Northampton, 2022.
4. For a general discussion of the issue, see F. Geburczyk, *Automated administrative decision-making under the influence of the GDPR – Early reflections and upcoming challenges*, in *Computer Law & Security Review*, 41, 2021. From a national perspective, see, e.g. V. Sharp, J. Nešpor, E. Klimentová, *Automation of Administrative Proceedings in the Czech Republic: Critical Reflections on the Draft “ADM Amendment” to the Administrative Procedure Code*, in *STUDIA IURIDICA Cassoviensia*, 13, Special Issue, 2025, D.U. Galetta, G. Pinotti, *Automation and Algorithmic Decision-Making Systems in the Italian Public Administration*, in *CERIDAP*, 1, 2023, J.-P. Schneider, F. Enderlein, *Automated Decision-Making Systems in German Administrative Law*, in *CERIDAP*, 1, 2023, or J. Reichel, *Regulating Automation of Swedish Public Administration*, in *CERIDAP*, 1, 2023.

5. M.C.H. Decker, L. Wegner, C. Leicht-Scholten, *Procedural fairness in algorithmic decision-making: the role of public Engagement*, 27:1, 2025; J.C. Covilla, *Artificial Intelligence and Administrative Discretion: Exploring Adaptations and Boundaries* in *European Journal of Risk Regulation*, 16, 2025; F.J. Zuiderveen Borgesius, *Strengthening legal protection against discrimination by algorithms and artificial intelligence*, in *Int. J. Hum. Rights*, 24(10), 2020, or G. Malgieri, *Automated decision-making in the EU Member States: The right to explanation and other “suitable safeguards” in the national legislation*, in *Computer Law & Security Review*, 5, 2019.
6. See, e.g. P. Hubková, *EU Administrative Decision-Making Delegated to Machines – Legal Challenges and Issues*, in *Acta Universitatis Carolinae – Iuridica*, 70(2), 2024; J. Nešpor, *Automated Administrative Decision-Making: What is the Black Box Hiding?*, in *Acta Universitatis Carolinae – Iuridica*, 70(2), 2024, or F.J. Zuiderveen Borgesius. *Strengthening legal protection against discrimination by algorithms and artificial intelligence*, in *Int. J. Hum. Rights*, 24(10), 2020.
7. Among the relevant sources, reference may be made to R. Jakab, *National Report on Automation in Decision-Making in Public Administration in Slovakia*, in *Acta Universitatis Carolinae – Iuridica*, 70(2), 2024; L. Jančát, *Automation of Processes in Public Administration in Selected Member States of the European Union*, in *STUDIA IURIDICA Cassoviensia*, 13, Special Issue, 2025; A. Lengyel, *Automated Decision-Making in Administrative Penal Matters*, in *STUDIA IURIDICA Cassoviensia*, 13, Special Issue, 2025, or Z. Hamuláková, *Právo na spravodlivý proces v kontexte automatizovaných rozhodnutí vo verejnej správe*, in M. Maslen (eds.), *Elektronizácia a digitalizácia verejnej správy*, Typi Universitatis Tyrnaviensis, Trnava, 2024.
8. Their list is available on the OECD website. See The Organisation for Economic Co-operation and Development. 2026. OECD homepage, Online. [cited 2026-02-10]. Available at: https://www.oecd.org/en/publications/progress-in-implementing-the-european-union-coordinated-plan-on-artificial-intelligence-volume-1_6d530a88-en/slovak-republic_5b824a40-en.html.
9. Ministry of Investments, Regional Development and Informatization of the Slovak Republic, 2026. MIRRI homepage. Online. [cited 2026-02-10]. Available at: <https://mirri.gov.sk/wp-content/uploads/2019/10/SDT-English-Version-FINAL.pdf>. See p. 3 and p. 5.
10. Ministry of Investments, Regional Development and Informatization of the Slovak Republic, 2026. MIRRI homepage. Online. [cited 2026-02-10]. Available at: <https://mirri.gov.sk/wp-content/uploads/2019/10/SDT-English-Version-FINAL.pdf>. See p. 26 and p. 31.
11. Act No. 71/1967 Coll. on Administrative Procedure (Administrative Procedure Code), as amended.
12. For example, tax proceedings are regulated in Act No. 563/2009 Coll. on Tax Administration (Tax Code) and on Amendments and Supplements to Certain Acts, as

amended, or proceedings in matters of social security in Act No. 461/2003 Coll. on Social Insurance, as amended, while the application of the Administrative Code is expressly excluded by these acts. However, in the case of most proceedings, e.g. proceedings on misdemeanors regulated in Act No. 372/1990 Coll. on Misdemeanors, as amended, or construction proceedings regulated in Act No. 25/2025 Coll. on Construction Act, as amended, their specificities are regulated in these statutes in Slovakia, while to the extent not regulated by these special regulations, the Administrative Procedure Code is applied subsidiary to the given proceedings.

13. See § 9 of the Administrative Procedure Code.
14. See § 22 of the Administrative Procedure Code.
15. See § 47 par. 5 of the Administrative Procedure Code.
16. According to Art. 2 para. 2 of the Constitution No. 460/1992 Coll. Constitution of the Slovak Republic, as amended: «*State bodies may act only on the basis of the Constitution, within its limits and to the extent and in the manner established by law*».
17. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).
18. Although the GDPR applies to the protection of personal data, it will be relevant to all decision-making processes, as their addressees are always the persons holding their personal data as the addressees. See R. Hučková, *Zásady spracovania osobných údajov podľa GDPR so zreteľom na oblasť poskytovania zdravotnej starostlivosti*, in *STUDIA IURIDICA Cassoviensia*, 8(2), 2020.
19. R. Jakab, *National Report on Automation in Decision-Making in Public Administration in Slovakia*, in *Acta Universitatis Carolinae – Iuridica*, 70(2), 2024.
20. JuLIA, 2025. julia-project homepage. Online. [cited 2026-02-10]. Available at: https://www.julia-project.eu/sites/default/files/2025-05/Final%20Handbook_%20AI%20and%20Public%20Administration_%20The%20%28legal%29%20limits%20of%20algorithmic%20governance.docx.pdf, p. 91.
21. The eKasa system can be defined as an electronic cash register system that directly connects entrepreneurs cash registers with the Slovak Tax Office in real time. See Act No. 384/2025 Coll. on Revenue Recording and on Amendments and Supplements to Certain Acts.
22. Cons. C. of the Slovak Republic, 10 November 2021, PL. ÚS 25/2019, point 123 and 142.
23. In this context, the Constitutional Court referred to the following sources on which it relied: BVerfG, sp. zn. 1 BvR 142/15, point 101, CJEU Opinion No. 1/15, para. 168 et seq., as well as Recommendation CM/Rec(2020)1, part B, points 4 and 5.
24. Cons. C. of the Slovak Republic, 10 November 2021, PL. ÚS 25/2019, point 127, 129 and 132.
25. Cons. C. of the Slovak Republic, 10 November 2021, PL. ÚS 25/2019, point 133 and 135.

26. Cons. C. of the Slovak Republic, 10 November 2021, PL. ÚS 25/2019, point 136.
27. Cons. C. of the Slovak Republic, 10 November 2021, PL. ÚS 25/2019, point 138 and 139.
28. Cons. C. of the Slovak Republic, 10 November 2021, PL. ÚS 25/2019, points 88 to 99 and point 144.
29. Plan of Legislative Tasks of the Government of the Slovak Republic for 2025. Online. [cited 2026-02-10]. Available at: https://www.vlada.gov.sk/site/assets/files/1771/pluv-sr_2025.pdf.
30. Legislative process PI/2025/2. Online. [cited 2026-02-12]. Available at: <https://www.slov-lex.sk/elegislativa/legislativne-procesy/SK/PI/2025/2>.
31. Legislative process PI/2025/2. Online. [cited 2026-02-12]. Available at: <https://www.slov-lex.sk/elegislativa/legislativne-procesy/SK/PI/2025/2>.
32. Legislative process LP/2025/401. Online. [cited 2026-02-12] Available at: <https://www.slov-lex.sk/elegislativa/legislativne-procesy/SK/LP/2025/401>.
33. Legislative process LP/2025/401. Online. [cited 2026-02-12]. Available at: <https://www.slov-lex.sk/elegislativa/legislativne-procesy/SK/LP/2025/401>.
34. Legislative process LP/2025/401. Online. [cited 2026-02-12]. Available at: <https://www.slov-lex.sk/elegislativa/legislativne-procesy/SK/LP/2025/401>.
35. Legislative process LP/2026/17. Online. [cited 2026-02-12]. Available at: <https://www.slov-lex.sk/elegislativa/legislativne-procesy/SK/LP/2026/17>.
36. Legislative process LP/2026/17. Online. [cited 2026-02-12]. Available at: <https://www.slov-lex.sk/elegislativa/legislativne-procesy/SK/LP/2026/17>.
37. The explanatory memorandum to the draft Act. Online. [cited 2026-02-12]. Available at: <https://www.slov-lex.sk/elegislativa/legislativne-procesy/SK/LP/2025/401>.
38. The explanatory memorandum to the draft Act. Online. [cited 2026-02-12]. Available at: <https://www.slov-lex.sk/elegislativa/legislativne-procesy/SK/LP/2025/401>.
39. The explanatory memorandum to the draft Act. Online. [cited 2026-02-12]. Available at: <https://www.slov-lex.sk/elegislativa/legislativne-procesy/SK/LP/2025/401>.
40. Preliminary information notice concerning the draft Act. Online. [cited 2026-02-12]. Available at: <https://www.slov-lex.sk/elegislativa/legislativne-procesy/SK/PI/2025/2>.
41. Preliminary information notice concerning the draft Act. Online. [cited 2026-02-12]. Available at: <https://www.slov-lex.sk/elegislativa/legislativne-procesy/SK/PI/2025/2>.
42. The explanatory memorandum to the draft Act. Online. [cited 2026-02-12]. Available at: <https://www.slov-lex.sk/elegislativa/legislativne-procesy/SK/LP/2025/401>.
43. See § 3(1) of the draft Act. Online. [cited 2026-02-12]. Available at: <https://www.slov-lex.sk/elegislativa/legislativne-procesy/SK/LP/2025/401>.
44. Regulation (EC) No 765/2008 of the European Parliament and of the Council of 9 July 2008 setting out the requirements for accreditation and market surveillance relating to the marketing of products and repealing Regulation (EEC) No 339/93.
45. The explanatory memorandum to the draft Act. Online. [cited 2026-02-12]. Available at: <https://www.slov-lex.sk/elegislativa/legislativne-procesy/SK/LP/2026/17>.

46. The explanatory memorandum to the draft Act. Online. [cited 2026-02-12]. Available at: <https://www.slov-lex.sk/elegislativa/legislativne-procesy/SK/LP/2026/17>.
47. The explanatory memorandum to the draft Act. Online. [cited 2026-02-12]. Available at: <https://www.slov-lex.sk/elegislativa/legislativne-procesy/SK/LP/2026/17>.
48. Comments submitted on the draft Act. Online. [cited 2026-02-12]. Available at: <https://www.slov-lex.sk/pripomienky/legislativne-procesy/SK/LP/2026/17/pripomienky/zobraz>.
49. Comments submitted on the draft Act. Online. [cited 2026-02-12]. Available at: <https://www.slov-lex.sk/pripomienky/legislativne-procesy/SK/LP/2026/17/pripomienky/zobraz>.
50. Comments submitted on the draft Act. Online. [cited 2026-02-12]. Available at: <https://www.slov-lex.sk/pripomienky/legislativne-procesy/SK/LP/2026/17/pripomienky/zobraz>.
51. Comments submitted on the draft Act. Online. [cited 2026-02-12]. Available at: <https://www.slov-lex.sk/pripomienky/legislativne-procesy/SK/LP/2026/17/pripomienky/zobraz>.

Artificial Intelligence and Interaction Between Machines and Humans: Discretionary Powers, Human Reserve in the Public Sector and the EU Regulation of 2024

Julio Ponce Solé

DOI: 10.13130/2723-9195/2026-1-120

La presente analisi esamina il rapporto tra esseri umani e intelligenza artificiale con riferimento a un profilo attinente al diritto a una buona amministrazione e all'equità delle decisioni amministrative discrezionali: ciò che può essere definito come «riserva di umanità» o «riserva umana».

Con riguardo a tale questione, lo studio affronta i temi del limite, dell'empatia umana, delle tipiche inferenze abduttive proprie dell'essere umano e delle garanzie procedurali (che devono comprendere l'audizione e l'obbligo di motivazione), quali elementi rilevanti per l'applicazione del principio di precauzione allo stato attuale della tecnologia e per l'adozione di scelte normative volte a escludere l'intelligenza artificiale dalle decisioni amministrative discrezionali automatizzate, al fine di preservare la correttezza e l'equità delle decisioni. L'intelligenza artificiale dovrebbe pertanto svolgere esclusivamente una funzione di supporto al decisore umano nell'esercizio dei poteri amministrativi discrezionali. Sono infine esaminate le normative vigenti dell'Unione europea e degli Stati membri che prevedono tale riserva, nonché le possibili evoluzioni normative future in questo ambito.

This analysis considers the relationship between human beings and Artificial Intelligence in reference to one aspect linked to the right to good administration and fair discretionary administrative decisions: what we might call the “reserve of humankind” or “human reserve”. In relation to this issue, the study analyses the questions of fettering, human empathy, typical human abductive inferences and due administrative procedures (which must include hearing and giving reasons) as

relevant reasons to apply the precautionary principle in the current state of technology and to establish legal decisions to exclude artificial intelligence from automated discretionary administrative decisions in order to preserve good and fair decisions. Artificial intelligence should act only as a support for the human decision-maker when developing discretionary administrative powers. Current EU and Member State regulations establishing such a reservation are discussed, as well as possible future regulations in this field.

Summary: 1. Introduction: object of this analysis.- 2. Humans and artificial intelligence: fettering, empathy and abductive inferences.- 2.1. The problem of fettering.- 2.2. The problem of empathy and abductive inferences in treating human beings fairly.- 2.2.1. Machines do not have empathy and therefore they cannot treat human beings fairly.- 2.2.2. Machines cannot develop abductive inferences.- 2.3. Fair treatment, empathy and abductive inferences.- 3. The right to good administration, due procedure and hearing: audi alteram parte and the duty to give reasons.- 4. The “reserve of humankind” or “human reserve” in the field of Artificial Intelligence.- 5. Final thoughts and the need for further reflection: promoting fair and good discretionary administrative decisions.

1. Introduction: object of this analysis

The interaction between humans (developers, users) and machines can exist (at the time) when the system is being created (conception design) as well as when it is functioning (implementation)^[1]. In relation to the latter, the human supervision is sometimes described as the human in the loop (semi-autonomous systems that need a final human decision), human on the loop (the human can supervise the system and intervene if necessary) and human out of the loop (autonomous systems in which there is no possibility of human intervention after its deployment)^[2]. In the following lines, of all the possible topics in this human-artificial intelligence *interface*, I will select one in particular: the role of humans in limiting the operation of AI in practice.

Indeed, in previous studies I have had the opportunity to highlight how I considered that it was necessary to limit, in principle, the use of AI in relation to

the exercise of discretionary administrative powers, on the grounds that it lacks the human empathy necessary to make decisions that may affect citizens in an appropriate manner. This is what I have called a legal «*human reserve*»^[5] It means that it is humans, through the mechanisms of the democratic state and the rule of law, who decide, in a given time and territory, to prohibit the use of AI in certain areas, for various reasons, as will be seen.

The law can decide, as will be seen, to prohibit it in certain cases, for certain reasons, or to accept it, with human supervision. The following may therefore be a relevant question: why would it be in the law's interest to prohibit automated decision-making? In order to answer this question, I will also consider the EU Regulation on Artificial Intelligence passed by the European Parliament in 2024 (hereinafter AI Act^[6]).

2. Humans and artificial intelligence: fettering, empathy and abductive inferences

The advantages of AI seem unquestionable: a much greater processing capacity than humans (whose memory is far more limited and is affected by fatigue), the ability to make predictions through correlations, greater effectiveness, efficiency and, therefore, good management^[7], the possibility of avoiding the cognitive biases of humans and of doing away with noise, i.e. the existence of undesired and undesirable differences as regards decisions in similar conditions^[8], and the ability to generate new related occupations.

However, be that as it may, we must not forget the problems that AI can also generate: from the replacement of humans in private and public tasks with the consequent unemployment^[9], to the fact that correlations can give rise to *statistical hallucinations* and that these correlations have a *conservative* tendency, as the *status quo* takes precedence, since predictions are drawn from what happened in the past, but the past, and people, can change. The errors can also exist in programming (*bugs*), which can have a large-scale impact, affecting many more people than a decision made by a human.

Furthermore, it is not true that AI is free of biases (since biases exist in the data used to train the decision-making AI, together with statistical biases). AI is also *infected* by human programmers' cognitive biases^[10], such as the *availability bias*,

regarding the use of certain data used to train the machine, or the *confirmation bias*, for example, when selecting such data.

In addition, AI produces a specific and well-studied human cognitive bias: the *automation bias*, whereby, in general, humans tend to trust what the systems employed tell us and rarely question it, which can lead to extreme scenarios, such as what are known as *GPS deaths*: people lost in the desert who die because they follow the wrong directions of the navigation system^[9]. On the other hand, humans may erroneously value human judgement over algorithmic recommendations (under-reliance).

What is more, as is well known, automatic and deep learning systems present the problem of what are known as *black boxes*. The complexity of these decision-making systems makes it difficult to explain the steps followed until the final decision is made, which raises legally relevant questions regarding judicial control of the monitoring of due process, in terms of the prohibition of arbitrariness (e.g. 9.3 Spanish Constitution), i.e. the prohibition of irrational or unmotivated decisions, and also of good administration, i.e. the requirement of decisions resulting from respect for the legal obligation of due care or due diligence in making the final decision, which is required to have considered all relevant factors and discarded irrelevant ones.

It is sometimes claimed that AI does not present a larger black box than the brain of a human decision-maker, for example, a public manager, from a legal point of view. But this is not really the case; jurisprudence at the international (e.g. American hard look doctrine), European and Spanish levels has made an enormous effort in recent decades to unveil the *human black box*, requiring compliance with the legal obligations applicable to the human cognitive process of gathering information and weighing it, through the careful analysis of the administrative file, of the legal bases offered for the final decision, and of the coherence between them.

It should be noted that this is a notable advance, which cannot go unnoticed, the result of the right to good administration, a typical European legal concept, included in Article 41 of the European Charter of Fundamental Rights, which is important for digital administration as well^[10].

The principle of good administration has been quoted and developed by the ECJ since the 1950s, in thousands of cases, well before its inclusion as a right in the

ECHR. The first reference to the right was included in case T-54/99, *Max.mobil v. Commission, Judgment of the Court of First Instance (Second Chamber, Extended Composition), 30 January 2002*^[11]. This principle and right to good administration is recognized in Member States as well. That is the case of Italy, with the legal concept of *buon andamento* included in Article 97 of the Italian Constitution, or Spain, with the concept of *buena administración*, implicitly included in the Spanish Constitution according to case law and explicitly in various regional Statutes of Autonomy, as well as in the legislation on transparency and good governance (e.g. Spanish Act 19/2013) with many judicial decisions from the Spanish Supreme Court interpreting this right to good administration^[12].

The right to good administration implies a *legal duty of due diligence or due care*, as interpreted by the case law of the European Court of Justice^[13]: due diligence or due care in considering all the relevant factors before deciding, as an expression of a *fair* treatment which is imposed by Article 41 of the European Charter of Fundamental Rights.

Although the concept of fairness is complicated and can be understood in different ways from different perspectives, a common element of fairness can be identified: it has to do with equity, which is not the same as equality. As the Merriam-Webster dictionary emphasizes: «*The idea that sometimes sameness of treatment (equality) does not result in proportional fairness (equity) is one way that these words are distinguished from each other, even in similar contexts*»^[14]. Fair treatment allows and obliges the decision-maker to take into account all the relevant factors in a situation, including differences among the people involved and deciding in accordance with them (e.g. applying positive actions to compensate for inequality).

In my opinion, which will be developed in this article, the human limitation of the use of AI should be based on the right to good administration and the correlative idea of fairness in connection with something much deeper still. The legal doctrine and regulations dealing with the subject mention the concepts of human dignity and possible violation of rights, as does Article 14 of the AI Act, in this second case, at an advanced stage of the approval process, to justify human “distrust” of AI: human oversight is needed for «*preventing or minimising the risks to health, safety or fundamental rights that may emerge when a high-risk AI*

system is used in accordance with its intended purpose or under conditions of reasonably foreseeable misuse [...]». A similar question to the previous one can be asked, although changing the perspective: why can an algorithmic system affect the dignity of the person or violate their rights in a specific way that a human cannot?

To answer this question, it is necessary to make some distinctions that will be important for drawing conclusions. These are three pairs of concepts that, as we shall see, are interrelated:

- a) Types of AI participation in formalized administrative decision-making (obviating nuances for the moment): such participation may provide support and assistance for the making of a human administrative decision, which will be semi-automated, as opposed to automated decisions which involve the adoption of the final decision that closes the procedure by AI.
- b) Types of administrative powers exercised: which may be either binding or discretionary.
- c) Types of AI: as opposed to symbolic AI, which develops deductive inferences (good old-fashioned AI or GOFAI), there is non-symbolic, connectionist or statistics-based AI (machine learning, deep learning), which makes inductive inferences. Now, article 3 AI Act defines an AI system as a machine-based system designed to operate with varying levels of autonomy, that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments ^[15].

We should now consider how these three perspectives combine and what conclusions can be drawn from them. I consider that, depending on the type of AI, the answers to the question of limitation of AI are different, as we will see immediately.

2.1. The problem of fettering

In the case of rule-based systems (a type of expert system that uses a set of predefined rules to make decisions or provide solutions to problems, developing deductive inferences), the problem is a legal one: the prohibition of fettering.

According to the right to good administration and its legal obligation of due care and due diligence, established by the ECJ and national case law, there is a duty to develop discretionary powers considering all the relevant factors of a case and not applying rigid rules in an identical way in different possible situations (France, *Conseil d'État*, case *Piron*, 1942, in a similar way: Article 3 Dutch *Algemene wet bestuursrecht* of 1994).

This is known in the Common Law as a prohibition of fettering. In that sense, the UK judicial case *British Oxygen v. Board of Trade*, 1970, establishes that «*the general rule is that anyone who has to exercise a statutory discretion must not “shut his ears” to an application*»^[16].

But this prohibition according to Craig «*may be problematic where decisions are made by AI systems. This is because such systems are normally geared towards uniformity of output, rather than considerations that pertain to the specifics of an individual case. This is especially so where a public body relies exclusively or unthinkingly on an algorithm when exercising its discretionary power*»^[17].

2.2. The problem of empathy and abductive inferences in treating human beings fairly

In the case of AI using statistics and developing inductive inferences (e.g. machine learning), the problems in relation to discretionary administrative powers are different, as we will see immediately.

2.2.1. Machines do not have empathy and therefore they cannot treat human beings fairly

Machines lack empathy (although they can imitate it); therefore, they cannot make good decisions in those cases where there is a margin of administrative assessment, while maintaining due respect for citizens' right to good administration. Indeed, this absence of empathy is a relevant factor in relationships between human beings, which is linked to their dignity and personality (see Article 10.2 Spanish Constitution).

We can now define empathy very simply as a human cognitive component that allows us to understand others, including their difficulties, vulnerabilities and

problems, and to adapt our behaviour accordingly^[18].

Obviously, not all humans have the same level of empathy, but almost all have the ability not only to have it but also to train it. I say almost all because, precisely, a trait that characterizes psychopaths, who make up 1% of the population, is their lack of it (which does not mean that they cannot imitate it so as to manipulate other human beings in their favour)^[19].

On the other hand, empathy should not be confused with sympathy or compassion. Empathy includes feelings similar to those felt by the other person, but not feelings for what the other person feels. Empathy is a first step that may or may not lead to sympathy and compassion or to an emotional contagion that may be of anguish, oriented not to the other, but to oneself. Consequently, empathy should not be confused with a simple emotion that will lead to the adoption of partial and non-objective decisions in the legal field: on the contrary, it will be that empathy, that capacity to consider different perspectives that will lead us to impartiality and objectivity, by weighing all the factors^[20].

Algorithmic systems and artificial intelligence have no empathy^[21]. The media are accustomed to using metaphorical language, *humanizing* machines, and so, for example, they have reported on *Norman*, a system, trained with specifically negative data, which has been called the first psychopathic AI system^[22].

I am not in favour of using such metaphorical language with machines, but if we follow the game now, actually the media were completely wrong: *Norman* was not the first psychopathic AI system, for the simple reason that *all of them have been, are and in principle, we'll say more about that later, will be psychopaths*. They all lack empathy, they cannot understand or put themselves in the place of a human being, because machines do not get sick, age or die; they do not suffer and therefore cannot empathize with a human being, although they can mimic empathy (like psychopathic humans), which is even more chilling and can lead to humans being manipulated...

In this regard, the analysis carried out by Weizenbaum, a scientist and programmer of German origin at MIT (*Massachusetts Institute of Technology*), who, in 1966, as is well known, designed the first *chatbot* in history, which he named Eliza, in honour of the character in George Bernard Shaw's play *Pygmalion*, is noteworthy. Eliza was designed and programmed to act as a Rogerian therapist, entering into communication with people through questions

and answers, which were posed using a keyboard. Weizenbaum found that people, including his secretary, came to trust the program, which they humanized and trusted, giving rise to the so-called *Eliza effect*, of which, by the way, in 2022 we have had a new example with the LaMDA AI developed by Google, a much more sophisticated system^[23].

In 1976 Weizenbaum published an important book in which he warned about the danger of attributing tasks that demanded genuine human empathy to algorithmic systems, alluding among them to some, such as those made by judges or police officers^[24].

Along the same lines, Atienza has recently wondered whether AI machines can be or become moral agents, in whom some degree of dignity should be recognized. His answer is negative, «*since they do not have the capacity to feel pleasure and pain*», although «*they can perhaps behave as if they did, but that is another matter*»^[25].

Moving on from these general reflections to the field of administrative law and public management, it should be noted that the literature has insisted on the importance of empathy, a value of public service and an element to be considered in the elaboration and application of the law, which is crystallized, for example, in the principle of legal equity^[26].

In fact, empathy is embedded in the aforementioned principle and the right to good administration, since, as is well known, good administration requires due diligence and due care in taking into consideration all relevant factors before adopting an administrative decision and, undoubtedly, the specific situation of the person who is legally related to the Administration (for example, his or her vulnerability due to poverty, disability, life difficulties, human error...) is a relevant factor to be taken into account. It is linked with the idea «*that [e]very person has the right to have his or her affairs handled...fairly*». Fairness implies a due administrative procedure where proper consideration to all relevant factors, including the human component, must exist.

As Brennan-Marquez and Henderson argue^[27], in a liberal democracy, there must be an aspect of “role-reversibility” to judgement. Those who exercise judgement should be vulnerable, reciprocally, to its processes and effects. As Pasquale underlines when commenting on their work: «*the problem with an avatar judge, or even some super-sophisticated robot, is that it cannot experience punishment the*

way that a human being would». Role-reversibility is necessary for «*decision-makers to take the process seriously, respecting the gravity of decision-making from the perspective of affected parties*»^[28].

Empathy cannot mean sympathy or partiality. Obviously, there can be a risk of confusion between the two situations, which must be supervised and controlled in order to avoid biases or conflicts of interests.

2.2.2. Machines cannot develop abductive inferences

On the other hand, algorithmic systems cannot develop abductions^[29]. According to Pierce's point of view, abductions are one form of inference, based on formulating hypotheses and guessing, the other two being deduction and induction. Such guesses are the only form of inference that gives rise to knowledge. Deduction reiterates what we know, and induction tests or generalizes knowledge that we already have^[30]. AI cannot make abductions, because we are a long way from artificial general intelligence^[31].

2.3. Fair treatment, empathy and abductive inferences

Consequently, it is this lack of empathy and abductions on the part of machines that makes them unsuitable for making automated decisions with a margin of judgement that affects human beings. Fair and good decisions require rationality, but not only rationality, as the Portuguese neuroscientist carrying out research in the USA, Antonio Damasio, has pointed out, for example in his book *Descartes' Error*^[32]. In the legal field, as in other professional fields, the belief persists, inherited from the Enlightenment, that a good decision should not incorporate either emotions or feelings, but should simply use *cold rationality*. However, the neuroscientific studies of recent decades show us that this is a crude and outdated concept: we are sentient emotional beings who, because of this and thanks to this, make rational decisions. It seems as if nature had built the apparatus of rationality from and with the body.

Law professors Reis de Albuquerque and De Brito Machado Segundo have highlighted how reason and emotion, body and mind intertwine in human decision-making, something that AI is far from achieving, if it ever will^[33].

Ultimately, good administrative decision-making requires respect for the right to

good administration, which includes the obligation of fairness. This implies the legal obligation to consider the relevant factors before deciding with due diligence or due care in the context of the world. These include being able to be aware of the situation of the persons who relate to the Administration, to understand it and to act accordingly, applying equity, if necessary, in order to guarantee due administrative procedure, the dignity of the individual and the free development of his or her personality. In that sense, considering again the issue of empathy, there is a high risk in the absolute lack of empathy when deciding: the absence of fairness because empathy is a human ability and a key element that allows fair decisions. It is possible to say that empathy leads to fairness, and its absence is an obstacle to fair decisions^[34].

This is emphasized by both the Council of Europe and the Belgian Federal Ombudsman, for example. In the first case, the Council of Europe stresses that empathy is an important determinant of moral behaviour and a necessary element in building moral communities, because it leads us to understand the interests, needs and points of view of others when deciding^[35]. In the second case, the Belgian Ombudsman emphatically states that empathy is the key to the basis of good public service delivery and a necessary condition for a human public service that is adapted to the individual situation of each citizen. It favours changes of perspective, determines the way we look at a person or a group, and has a great impact on the way we perceive and respond to others and to life events^[36].

None of that can be done today by algorithmic systems and AI^[37].

3. The right to good administration, due procedure and hearing: *audi alteram parte* and the duty to give reasons

Beyond the limits described above, in some cases, human intervention will be imposed by Law during administrative procedures. This is the case of the existence of the right to be heard (*audi alteram parte*) and the duty of giving reasons, components of the right to good administration established by Article 41 of the European Charter of Fundamental Rights and a part of the constitutional traditions of the member states.

A classic component of the rule of law is the impossibility of taking decisions

affecting a person without first hearing the person concerned. This makes it impossible to fully automate decision-making, since it requires prior human intervention in the administrative decision-making procedure that diligently considers, in accordance with good administration, the arguments of the person concerned.

Full automation of decision-making, both discretionary and binding powers, would not be possible when the party concerned must be heard, since the administrative procedure must include human intervention that considers and integrates the input provided by the citizen into the final decision, accepting or rejecting the allegations, but taking them into consideration.

Consequently, the use of AI for the fully automated exercise of discretionary administrative powers (i.e. all those cases in which there is a margin of assessment) should, in principle, be ruled out, in application of the principle of good administration, unless there is a different express legal decision, and therefore restricted to the exercise of binding powers, and even in these, full automation will not be possible if there is an obligation to open a hearing procedure.

On the other hand, the duty to give reasons is connected to transparency and reasonableness, in order to avoid arbitrariness. Now, Article 86 of the AI Act recognises a right to explanation of individual decisions, establishing that any affected person subject to a decision which is taken by the deployer on the basis of the output from a high-risk AI system listed in Annex III, with the exception of systems listed under point 2 thereof, and which produces legal effects or similarly significantly affects that person in a way that they consider to have an adverse impact on their health, safety or fundamental rights shall have the right to obtain from the deployer clear and meaningful explanations of the role of the AI system in the decision-making procedure and the main elements of the decision taken.

In the case of statistical AI, if there is no way to give reasons for justifying the decision, beyond explaining the way in which the system is deployed, there is a serious legal problem of fairness. In that sense, we could say that if a result cannot be justified in accordance with the facts, law and criteria used, then the system that produced such a result should not be used^[38].

4. The “reserve of humankind” or “human reserve” in the field of Artificial Intelligence.

All the previous findings consequently lead to the use of one regulatory technique (the “reserve of humankind”), which we must understand to be based on the precautionary principle as applied to AI, to limit and monitor it in those areas that a society considers particularly sensitive, given the intrinsic limitations of the technology already discussed.

In the following pages, I will only point out certain essential questions, which require further research.

As for the concept of the “reserve of humankind” or the “human reserve”, this could be defined as human decision-making concerning certain areas in which human beings, through the bodies legitimized to do so, exclude the application of AI because its use is considered inappropriate as a result of being unfair.

The idea of a “reserve of humankind” is comparable, on another level, to the idea of reserving the exercise of certain powers for civil servants, currently provided for in Spain in Article 9.2 of Royal Legislative Decree 5/2015, of October 30, approving the revised text of the Law on the Basic Statute of the Public Employee. In this context, the Spanish legislator considers that administrative powers should only be exercised by public employees who, due to their statutory relationship with the Administration, which protects their impartiality and objectivity and offers them resistance to political or private lobbying pressures, are in the best situation to serve general interests well.

In the field of AI, I believe a similar reasoning implicitly lies behind the legal rules already in place or in the process of being developed that require that only humans should (be allowed to) make certain decisions (thus preventing fully automated decisions) due to various reasons, among which is the capacity for empathy that machines lack.

Thus, Article 22 of the General Data Protection Regulation (hereinafter, GDPR^[39]) configures, in fact, a “reserve of humankind” as a principle when establishing that the data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her. The same article includes some exceptions to that general principle.

Among them we can find the case in which a regulation specifically allows automated decisions. This regulation in the Spanish context, for example, should be a rule with the rank of law, by application of Article 18.4 Spanish Constitution («*The law shall limit the use of information technology to ensure the honour and personal and family privacy of citizens and the full exercise of their rights*»)^[40].

Likewise, with a narrower focus, affecting only discretionary powers, from another approach, various laws prohibit the use of AI except in the exercise of regulated powers. Thus, German legislation (from my translation: «*An administrative act may be issued fully automatically, provided that this is permitted by law and that there is neither discretion nor margin of appreciation*», (*Verwaltungsverfahrensgesetz, VwVfG, § 35a Vollständig automatisierter Erlass eines Verwaltungsaktes*) or Catalan (law 26/2010, Article 44.2, from my translation: «*Only acts that can be adopted with a programming based on objective criteria and parameters are susceptible to automated administrative action*»). Still more clearly, law 2/2025, of April 2, for the development and promotion of artificial intelligence in Galicia, Spain, from my translation):

«Article 12. Humanity reserve and human supervision.

2. In the event of the use of artificial intelligence systems to support or inform the adoption of administrative acts or decisions, the necessary safeguards shall be adopted to mitigate any bias on the part of the competent decision-making body. Under no circumstances shall such actions employing artificial intelligence systems constitute administrative decisions or acts in themselves without validation by the head of the competent body.

3. In cases of use of artificial intelligence systems that serve for the adoption of formalized administrative acts, both procedural and decisive, in an automated manner without direct human intervention, in accordance with the provisions of Article 76 of Law 4/2019, of July 17, it must be administrative acts that do not require a subjective assessment of the concurrent circumstances or a legal interpretation».

The rationale of these texts, I believe, must be found in the aforementioned prohibition of fettering (rule-based AI) and in the lack of empathy of machines and the limits in relation to abductions (AI using statistics, e.g. machine learning): the exercise of discretion requires fairness, that is the use of empathy

when weighing all relevant elements in making the decision (obligation of due diligence or due care inherent in the right to good administration), and the application, where appropriate, of equity and experience and common sense. Only humans can decide in this way. A different matter is the automated application of regulated powers without any margin of appreciation or the use of AI as a supporting device in human decisions.

To sum up, when exercising discretionary administrative powers, machines cannot decide by themselves because they would produce unfair decisions violating the right to good administration. From a legal point of view, it is not possible to have an effective but unfair public administration: it must be fair *and* effective^[41]. In the current state of technology, I advocate the application of the precautionary principle in this field, reserving discretionary decisions to human beings^[42]. Human administrative decision-makers can use AI as a supporting tool to help them to make better decisions. From this perspective, our human reserve coincides partially, from the perspective of public administration, with Pasquale's proposed «*first new law of robotics*»: robotic and artificial intelligence systems should complement professionals, or members of a profession, and not replace them^[43].

In accordance with those conclusions, I think that the AI Act, in its Article 5, when deciding to exclude several instances of AI use, including, unless specifically authorized, biometric identification, should also make reference to the prohibition of fully automated discretionary administrative decisions. In fact, Annex III of the AI Act establishes that high-risk AI uses include «*AI systems intended to assist competent public authorities for the examination of applications for asylum, visa and residence permits and associated complaints with regard to the eligibility of the natural persons applying for a status*». In that case it could imply that AI systems can only assist decision-makers, not replace them. But in other references in Annex III this distinction is not so clear (e.g. «*Migration, asylum and border control management*»). I think it is possible to extract a principle from Annex III interpreting words such as «*assist*», «*evaluate*», «*establish priorities*» in the sense that machines should be a supporting tool provided there is discretion, but not replace human decision-making.

Although this is not the topic of this study, the reasons I have considered should also lead to a prohibition of fully automated decisions in relation to the judicial

and legislative branches. Actually, in relation to judges, Annex III of the AI Act states that high-risk AI uses include:

«8. *Administration of justice and democratic processes:*

(a) AI systems intended to assist a judicial authority in the investigation and interpretation of facts and law, and in the application of the law to a particular set of facts».

The text uses the verb ‘to assist’. The question to be clarified here is whether the use of this verb excludes fully automated judicial decision-making with a margin of appreciation. Along these lines, in the amendments tabled by the European Parliament to the proposal of regulation, recital 40 was modified by amendment 71, which stated the following:

«The use of artificial intelligence tools may support decision-making but should not replace the decision-making power of judges or judicial independence, since final decision-making must remain a human activity and a human decision».

This is the origin of current recital 61 AI Act, which indicates that *«The use of AI tools can support the decision-making power of judges or judicial independence but should not replace it: the final decision-making must remain a human-driven activity».*

Whatever the case, I think that such a prohibition of replacing the final discretionary human decision should be explicitly included in the future in Article 5 of the AI Act to make it clear and to cover other cases in which there is no specific high risk but rather administrative discretion, the use of which can, in itself, constitute a high risk of maladministration and violation of citizens’ rights. This point of view is supported by the European Ombudsman and by different decisions from the European Court of Justice.

The European Ombudsman, launched an investigation to understand how the European Commission decides on and uses AI in relation to the right to good administration, making several suggestions in 2024 and noting that *«good administration implies being humane and humane. When human beings are removed from the equation of service delivery, it is clear that problems can arise»*^[44]. The European Ombudsman, citing the CJEU judgment of 22 January 2014, Case C-270/12, *UK v. Parliament and Council*^[45], has indicated that the use of AI cannot constitute a prohibited delegation of discretion, stressing the need for legislators to specify these limits^[46]. AI could only be used to fully automate

regulated powers, or to support and assist human decision-making in both regulated and discretionary matters.

In the case of motivation, the CJEU ruling of June 21, 2022 (case C817/19) stated in relation to Directive (EU) 2016/681 of the European Parliament and of the Council of April 27 on the use of Passenger Name Record (PNR) data (PNR) for the prevention, detection, investigation, and prosecution of terrorist offenses and serious crime, given the opacity with which PNR systems operate (PNR Directive), it might be impossible to understand the reason why a given program arrived at a positive identification. This CJEU ruling established that the need for predetermined criteria regarding the use of passenger name records, imposed by the PNR Directive, meant that it was legally impossible to use self-learning artificial intelligence technology (machine learning)^[47].

5. Final thoughts and the need for further reflection: promoting fair and good discretionary administrative decisions

As has been argued, the fact that AI is technically available, now or in the future, to be applied in any area of administrative operation does not mean that it is always legal, necessary, and convenient to apply it. A danger of unfairness exists when applying discretionary administrative powers leading to the violation of the right to good administration, which in my opinion requires current AI systems to be prohibited from taking discretionary administrative decisions in application of the precautionary principle.

In the same way that there must be a necessary evaluation of the need or not to contract goods and services by Public Administrations which must accredit the need, suitability and efficiency of the contracting, there should be, even before going on to evaluate any AI system, a prior administrative assessment of whether the use of AI systems is necessary, suitable and efficient in a given area, considering costs, benefits and risks. This evaluation should be legally required. The starting point, I understand, should be consider the costs and benefits (from a social, environmental and economic point of view) when deciding if development of public functions will be carried out by humans or machine. Only after a justified weighted evaluation based on the administrative file, the use of AI

will make sense if it is necessary, suitable and efficient.

The current limits of AI have been underlined here may make it advisable to prohibit its use in certain areas, a prohibition that I believe should be applied by establishing a nudge, a default option (*opt out*) in favour of not using automated decision-making, unless a regulation allows it.

In accordance with the reasons that have been explained, I think that Article 5 of the EU regulation should include a reference to the prohibition of fully automated administrative decisions^[48]. Although this is not the topic of this analysis, the same reasons should lead to a similar prohibition in the judicial and legislative branches.

Another issue related to the AI Act will be its impact on the national legal system. The AI Act will offer a European common starting point, but it is worth reflecting on whether it is necessarily a national ceiling. The Explanatory Memorandum of the AI Act project stated that the choice of a regulation as a legal instrument (in accordance with Article 288 TFEU) is justified «*by the need for a uniform application of the new rules, such as definition of AI, the prohibition of certain harmful AI enabled practices and the classification of certain AI systems*». This is without prejudice to the fact that «*the provisions of the regulation are not overly prescriptive and leave room for different levels of Member State action for elements that do not undermine the objectives of the initiative, in particular the internal organization of the market surveillance system and the uptake of measures to foster innovation*».

Thus, it is worth considering whether the EU Member State regulators could establish options that, while respecting the AI Act, could complement it. In relation to the human reserve, I think that a prohibition of replacing the final discretionary human decision can be established by Member States respecting EU powers and should be explicitly included in the future in Article 5 of the AI Act in accordance with all the reasons explained.

The application of the AI Act and the future national regulations should try to obtain the best of the machine and human worlds and not the worst to ensure that we do not end up having in our societies the (non-existent) empathy and compassion of unfair AI together with human cognitive limits and errors transferred to AI, thereby hindering its proper functioning.

1. The analysis is based in J. Ponce *Artificial Intelligence, Automated Administrative Decisions and Discretionary Powers: The “Human Reserve” and the Human in the Loop in the 2024 European Union Regulation*, in J. Ponce, A. Cerrillo-i-Martínez (Eds.), *The EU Artificial Intelligence Act and the Public Sector - Humans and AI Systems in Public Administration in the light of the European Regulation on Artificial Intelligence of 2024*, 2025, EPLO, (with Foreword by C. Coglianese). However, the text has been completely reviewed, introducing different changes and new content.
2. C.P. Trumbull IV, *Autonomous Weapons: How Existing Law Can Regulate Future Weapons*, in *Emory International Law Review*, vol. 34, 2, 2020, p. 539 ff.
3. See for the first time J. Ponce, *Inteligencia artificial, Derecho administrativo y reserva de humanidad: algoritmos y procedimiento administrativo debido tecnológico*, in *Revista General de Derecho Administrativo*, 1, January 2019. Following this construction, see in the Italian jurisprudenc G. Gallone, *Riserva di Umanità e Funzione Amministrativa*, Wolters Kluwers, CEDAM, 2023.
4. European Parliament legislative resolution of 13 March 2024 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on Artificial Intelligence (Artificial Intelligence Act) and amending certain Union Legislative Acts (COM (2021)0206 – C9-0146/2021 – 2021/0106(COD))
5. C. Coglianese, A. Lai, *Algorithm vs. Algorithm*, in *Faculty Scholarship at Penn Law*, 2795, 2022.
6. D. Kahneman, O. Sibony, C. Sunstein, *Noise: A failure of human judgment*, Hachette Book Group USA, 2021.
7. A. Pastor, P. Nogales, *El futuro del trabajo en la administración pública ¿estamos preparados?*, in *Revista Vasca de Gestión de Personas y Organizaciones Públicas*, extra-3, 2019, pp. 34-51.
8. J. Ponce, *Nudging’s Contributions to Good Governance and Good Administration – Legal Nudges*, in *Public and Private Sectors*, EPLO, 2022.
9. See in this regard, and in general, the interesting book by J. Bridle, *The new dark ages: technology and the end of the future*, Verso books, 2018.
10. This right includes the obligation of deciding fairly: «1. Every person has the right to have his or her affairs handled impartially, fairly and within a reasonable time by the institutions, bodies, offices and agencies of the Union. 2. This right includes: (a) the right of every person to be heard, before any individual measure which would affect him or her adversely is taken; (b) the right of every person to have access to his or her file, while respecting the legitimate interests of confidentiality and of professional and business secrecy; (c) the obligation of the administration to give reasons for its decisions. 3. Every person has the right to have the Union make good any damage caused by its institutions or by its servants in the performance of their duties, in accordance with the general principles common to the laws of the Member States. 4. Every person may write to the institutions of the Union in one of the languages of the

Treaties and must have an answer in the same language».

11. In this ruling, the Court stated that: «*Since the present action is directed against a measure rejecting a complaint, it must be emphasised at the outset that the diligent and impartial treatment of a complaint is associated with the right to sound administration which is one of the general principles that are observed in a State governed by the rule of law and are common to the constitutional traditions of the Member States. Article 41(1) of the Charter of Fundamental Rights of the European Union proclaimed at Nice on 7 December 2000 (OJ 2000 C 364, p. 1, hereinafter 'the Charter of Fundamental Rights') confirms that '[e]very person has the right to have his or her affairs handled impartially, fairly and within a reasonable time by the institutions and bodies of the Union'.*».

It is necessary to consider article 6.3 Treaty of the European Union: «*Fundamental rights, as guaranteed by the European Convention for the Protection of Human Rights and Fundamental Freedoms and as they result from the constitutional traditions common to the Member States, shall constitute general principles of the Union's law.* See also article 52.4 European Charter of Fundamental Rights: «*In so far as this Charter recognises fundamental rights as they result from the constitutional traditions common to the Member States, those rights shall be interpreted in harmony with those traditions*».

12. J. Ponce, *La lucha por el buen gobierno y el derecho a una buena administración mediante el estándar jurídico de diligencia debida*, UAH-Defensor del Pueblo, 2019.
13. H.C.H. Hofman, *The Duty of Care in EU Public Law-A Principle Between Discretion and Proportionality*, in *Review of European Administrative Law*, vol. 13, 2, 2020, pp. 87-112.
14. Dictionary Merriam-Webster, *Equality vs. Equity: What is the Difference?*, Merriam-Webster: «*Sometimes this distinction is explained with an illustration showing people of different heights using boxes to stand on in order to see over a fence; equality is if all the boxes are identical, but equity is if the boxes are different sizes to permit the people, regardless of their height, the ability to see over the fence*».
15. Recital 12 indicates that: «*The notion of 'AI system' in this Regulation should be clearly defined and should be closely aligned with the work of international organisations working on AI to ensure legal certainty, facilitate international convergence and wide acceptance, while providing the flexibility to accommodate the rapid technological developments in this field. Moreover, it should be based on key characteristics of AI systems that distinguish it from simpler traditional software systems or programming approaches and should not cover systems that are based on the rules defined solely by natural persons to automatically execute operations. A key characteristic of AI systems is their capability to infer. This capability to infer refers to the process of obtaining the outputs, such as predictions, content, recommendations, or decisions, which can influence physical and virtual environments, and to a capability of AI systems to derive models or algorithms from inputs or data. The techniques that enable inference while building an AI system include machine learning approaches that learn from data how to achieve certain objectives, and logic- and knowledge-based approaches that infer from encoded knowledge or symbolic representation of the task to be solved. The capacity of an AI system to infer transcends basic data processing, enables*

learning, reasoning or modelling. The term ‘machine-based’ refers to the fact that AI systems run on machines. The reference to explicit or implicit objectives underscores that AI systems can operate according to explicit defined objectives or to implicit objectives. The objectives of the AI system may be different from the intended purpose of the AI system in a specific context. For the purposes of this Regulation, environments should be understood to be the contexts in which the AI systems operate, whereas outputs generated by the AI system reflect different functions performed by AI systems and include predictions, content, recommendations or decisions. AI systems are designed to operate with varying levels of autonomy, meaning that they have some degree of independence of actions from human involvement and of capabilities to operate without human intervention. The adaptiveness that an AI system could exhibit after deployment, refers to self-learning capabilities, allowing the system to change while in use. AI systems can be used on a stand-alone basis or as a component of a product, irrespective of whether the system is physically integrated into the product (embedded) or serve the functionality of the product without being integrated therein (non-embedded)».

16. In that sense, Craig underlines that: *«The general principle is that a public body endowed with statutory discretionary powers cannot adopt a policy or rule that allows it to dispose of a case without any consideration of the merits of the individual applicant. The dominant line of authority allows the body to apply its rule provided only that the individual is granted the opportunity to contest its application to the particular case. The policy must be legitimate given the statutory framework within which the discretion is exercised. It must be based on relevant considerations and must not pursue improper purposes. These controls are necessary since otherwise a public authority could escape the normal constraints on the exercise of discretion by framing general policies. There is, however, also a principle of consistency, which creates a presumption that a public body will follow its own policy. If it seeks to depart from that policy then there must be good reasons for the departure and these must be given to the applicant».* P. Craig, *Administrative Law*, 9th Edition, Sweet &Maxwell, Thompson Reuters, 2021, paragraph 10-025.
17. P. Craig, *Administrative Law*, 9th Edition, Sweet &Maxwell, Thompson Reuters, 2021, paragraph 10-025.
18. In Spain, for example, the Dictionary of the Royal Spanish Academy defines empathy (the translation to English is mine) as the *«feeling of identification with something or someone»* and as the *«capacity to identify with someone and share their feelings»*. It also includes in one of the entries of the word humanity the following: *«sensitivity, compassion for other people’s misfortunes»*. In turn, the word human is defined in one of its entries as *«sympathetic, sensitive to the misfortunes of others»*.
19. We are facing a serious, dangerous disease, despite attempts to configure it as an evolutionary advantage (!), instead of what it is: a serious personality disorder, called an antisocial personality disorder by the DSM-5 (the Diagnostic and Statistical Manual of Mental Disorders, edited by the American Psychiatric Association).
20. S. Ranchordas, *Empathy in the digital administrative state*, in *Duke Law Journal*,

- Forthcoming, University of Groningen Faculty of Law Research Paper, 13, 2021.
21. In the field of war, see C.P. Trumbull IV, C.P., op.cit. p. 552, underlines how the use of autonomous weapons may eliminate the last vestiges of compassion or honor in warfare, quoting a former U.N. Special Rapporteur who argues that «*machines lack morality and mortality, and should as a result not have life and death powers over humans*» and an academic who notes that autonomous weapons may contribute to the «*dehumanization of killing*».
 22. The BBC, for example, speaks of the psychopathic algorithm; see *Are you scared yet? Meet Norman, the psychopathic AI* - BBC News.
 23. LaMDA is the acronym for Language Model for Dialogue Applications. In 2022 a Google engineer, Blake Lemoine, suffered from the Eliza effect, claiming that LaMDA was a conscious, sentient being, which he even provided with legal help through a lawyer. See “*The Google engineer who thinks the company’s AI has come to life*”, The Washington Post, June 11, 2022.
 24. J. Weizenbaum, *Computer power and human reason: From judgment to calculation*, W. H. Freeman & Co, 1976.
 25. M. Atienza, *Sobre la dignidad humana*, Trotta, Madrid Trumbull 2022. The translation from Spanish is mine.
 26. Article 3 of the Spanish Civil Code.
 27. K. Brennan-Marquez, S. Henderson, *Artificial Intelligence and Role-Reversible Judgment*, in *J. Crim. L. & Criminology*, 109, 2019.
 28. F. Pasquale, *Empathy, Democracy, and the Rule of Law*, JOTWELL (May 8, 2019) (reviewing Kiel Brennan-Marquez & Stephen E. Henderson, *Artificial Intelligence and Role-Reversible Judgment*, *J. of Crim. L. and Criminology*, available at SSRN: <https://cyber.jotwell.com/empathy-democracy-and-the-rule-of-law/>).
 29. For example, abduction is defined in Artificial Intelligence – foundations of computational agents – 5.6 Abduction (artint.info) as: «*a form of reasoning where assumptions are made to explain observations. For example, if an agent were to observe that some light was not working, it can hypothesize what is happening in the world to explain why the light was not working. An intelligent tutoring system could try to explain why a student gives some answer in terms of what the student understands and does not understand*».
- The term abduction was coined by Peirce (1839-1914) to differentiate this type of reasoning from deduction, which involves determining what logically follows from a set of axioms, and induction, which involves inferring general relationships from examples.
30. J. Brent, *Charles Sanders Pierce. A life*, Indiana University Press, 1993, p. 349.
 31. E. J. Larson, *The Myth of Artificial Intelligence: Why Computers Can’t Think the Way We Do*, Belknap Press: An Imprint of Harvard University Press, 2021.
 32. A. Damasio, *Descartes’ error*, Putnam, 1994.
 33. Could machines acquire real empathy in the future and therefore rethink some of the reflections I am going to put forward here? In an interesting article co-authored with Man, appearing in 2019 in Nature (K. Man, A. Damasio, *Homeostasis and soft robotics in the*

design of feeling machines, in *Nature Machine Intelligence*, vol. 1, 10, 2019, pp. 446-452) and in his latest book (*Feeling and Knowing*, Pantheon, 2021). Damasio does not close the door to it: he proposes working along the lines of giving machines a sense of survival, simulating in them a biological property of humans, that of homeostasis, i.e. the ability of organisms to remain in acceptable conditions for life (e.g. temperature). If we were able to endow machines with a sense of vulnerability and self-preserving behaviour, by means of sensors that teach them which factors play a role in their own survival (connected wires, adequate amount of electricity...), it is argued, this could lead to the generation of feelings (restlessness, satisfaction...) and from there, perhaps, to empathy. In any case, this is a hypothesis. This study prepared in 2024 is based on the current situation, which does not seem likely to change in the short or medium term.

34. K. M. Page, M. A. Nowak, *Empathy leads to fairness*, in *Bull Math Biol.*, 6, 2002, pp. 1101-16.
35. <https://www.coe.int/fr/web/digital-citizenship-education/ethics-and-empathy>.
36. Médiateur Fédéral, Rapport d'activités, 2021: Rapport annuel 2021.pdf (federaalombudsman.be).
37. McCann, who is also against the use of AI to adopt discretionary decisions, adds an interesting complementary consideration: the exercise of discretion by machines would entail the displacement of the human moral commitment that the exercise of discretion should entail, so that public servants would be (very humanly) tempted to hide behind machine decisions, not facing their ultimate responsibility for them. S. McCan, *Discretion in the Automated State*, in *The Canadian Journal of Law & Jurisprudence*, 2023, pp. 1 ff.
38. J. Palairt, *Reason-Giving in the Age of Algorithms*, in *Auckland University Law Review*, vol. 26, 2020, pp. 92 ff.
39. «Article 22 Automated individual decision-making, including profiling.1. The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.2. Paragraph 1 shall not apply if the decision:1. is necessary for entering into, or performance of, a contract between the data subject and a data controller;
2. is authorised by Union or Member State law to which the controller is subject and which also lays down suitable measures to safeguard the data subject's rights and freedoms and legitimate interests; or
3. is based on the data subject's explicit consent.
3. In the cases referred to in points (a) and (c) of paragraph 2, the data controller shall implement suitable measures to safeguard the data subject's rights and freedoms and legitimate interests, at least the right to obtain human intervention on the part of the controller, to express his or her point of view and to contest the decision.
4. Decisions referred to in paragraph 2 shall not be based on special categories of personal data referred to in Article 9(1), unless point (a) or (g) of Article 9(2) applies and suitable measures to safeguard the data subject's rights and freedoms and legitimate interests are in

place».

40. We could say, paraphrasing the well-known expression coined in the American Revolution: No Automation Without Representation. Along the same lines, although without specific reference to a norm with the status of law, as required by Article 18 Spanish Constitution, is Article XVIII.6.d of the Spanish Charter of Digital Rights (*«That the adoption of discretionary decisions be reserved to individuals, unless the adoption of automated decisions with adequate guarantees is provided for by law»*). A different question is whether it is considered that the current Article 41 of the 40/2015 Spanish Act, of October 1, on the Legal Regime of the Public Sector (LRJSP), with its very limited regulation, covers said reservation of law. The translation from Spanish is mine:

«Article 41. Automated administrative proceedings.

1. An automated administrative action is understood to be any act or action carried out entirely by electronic means by a Public Administration within the framework of an administrative procedure and in which a public employee has not intervened directly.

2. In the case of an automated administrative action, the competent body or bodies, as the case may be, for the definition of the specifications, programming, maintenance, supervision and quality control and, where appropriate, auditing of the information system and its source code, shall be established beforehand. Likewise, the body to be held responsible for the purposes of contestation shall also be indicated».

41. P. Váckzi, *Fair and Effective Public administration*, in *Institutiones Administrationis-Journal of Administrative Science*, Vol. 2, 1, 2022, pp. 161-170.
42. . C. Som, L. M. Bilty, A. R. Köhler, *The Precautionary Principle as a Framework for a Sustainable Information Society*, in *Journal of Business Ethics*, 85, 2009, pp. 493-505.
43. F. Pasquale, *New Laws of Robotics: Defending Human Expertise in the Age of AI*, Harvard University Press, 2020.
44. https://www.ombudsman.europa.eu/es/doc/closing-note/es/196934#_ftn18.
45. <https://curia.europa.eu/juris/document/document.jsf?text=&docid=146621&pageIndex=0&dclang=en&mode=lst&dir=&occ=first&part=1&cid=7732888>.
46. In that sense, in the case of Finland see the Kela case: <https://oikeuskansleri.fi/en/-/automated-decision-making-in-kela>.
47. Paragraph 194: this requirement *«precludes the use of artificial intelligence technologies in the context of machine learning systems, which may alter, without human intervention or oversight, the evaluation process and, in particular, the evaluation criteria on which the result of the application of the process is based, as well as the weighting of those criteria»*. The French Conseil Constitutionnel has ruled along the same lines in its decision 2018-765 of June 12, stating (our translation) that: *«71. Finally, the data controller must ensure that it has control over the algorithmic processing and its development, so that it can explain in detail and in an intelligible manner to the data subject how the processing has been applied in relation to him or her. Consequently, algorithms capable of reviewing the rules they apply on their own, without the control and validation of the data controller, cannot be used as the sole basis for an individual administrative decision»*.

48. On the other hand, the approval of the AI Act raises some questions. One is the possible contradiction between it and the GDPR in the matter at hand. The 22nd GDPR, as we already know, establishes a generalized “reserve of humankind”, which can be excepted by legal rule and replaced in such a case by human supervision, in any case. Article 5 of the AI Act, for its part, establishes the prohibition of AI only in certain cases and the aforementioned Article 14, as regards human supervision, refers only to high-risk cases (in the other uses of AI that do not involve high risk, the REUIA does not provide for either a “human reserve” or human supervision).

Regulating AI - borders of freedom and sovereignty

Sylva Řezníková

DOI: 10.13130/2723-9195/2026-1-152

Il presente contributo esamina l'ambito di applicazione del quadro normativo dell'Unione europea in materia di intelligenza artificiale nei confronti di soggetti non europei. L'analisi combina un approccio giuridico-dottrinale con una prospettiva socio-giuridica. I principali atti normativi europei, insieme alla più recente giurisprudenza, sono analizzati mediante il metodo della "black-letter law". Gli sviluppi che hanno condotto all'adozione di una normativa complessiva volta a limitare i rischi dell'IA sono altresì confrontati a livello globale. Vengono inoltre esaminate, in chiave socio-giuridica, recenti situazioni determinate da diversi attori sociali e politici, tra cui X di Elon Musk e l'amministrazione statunitense di Donald Trump, con particolare attenzione al possibile impatto della normativa europea su soggetti stranieri o su imprese stabilite all'estero.

La conclusione è nel senso che, nonostante l'indubbia necessità di regolamentare l'uso dell'intelligenza artificiale e di tutelare i consumatori europei, l'applicazione dei principi della legislazione europea a soggetti non europei può generare conflitti e costituire un ostacolo alla competitività del mercato tecnologico europeo. D'altro canto, consentire alle richieste provenienti dall'esterno di influenzare la legislazione europea rappresenterebbe una minaccia per la sovranità europea.

This paper examines the scope of the applicability of the European Union's regulatory framework on artificial intelligence to non-European subjects. The analysis combines doctrinal legal analysis with a socio-legal approach. We assess core pieces of European legislation, along with recent case law, using black-letter methodology. We compare worldwide developments that lead to comprehensive legislation aimed at limiting the risks of AI. Recent situations created by several social and political actors, including Elon Musk's X and the U.S. Donald Trump's administration, are dealt with in a

socio-legal analysis, which focuses on the possible impact of European legislation on foreign nationals or foreign-residing companies. It is concluded that despite the undoubtable necessity to regulate the use of AI and to protect European consumers, the application of European legislation's principles on non-European subjects may lead to conflicts and form obstacles to the European technological market's competitiveness. On the other hand, allowing foreign demands to influence European legislation would pose a threat to European sovereignty.

Summary: 1. Introduction.- 2. The European regulatory framework, the race to AI legislation, and relevance to foreign subjects.- 3. Case studies.- 3.1. The Digital Services Act and recent sanctions to X.- 3.2. Grok and disinformation spread through system prompts.- 3.3. GDPR violations sanctions and relation to AI regulation.- 4. Conclusions.

1. Introduction

The European Union aims to lead the way worldwide in regulating artificial intelligence (AI). The purpose of such regulation is to avoid risks and solve in advance ethical issues such as using AI by public administration to identify individuals remotely or based on biometrical data and to practice real-time surveillance. The purpose of regulating AI is also to establish a competitive and fair digital market and to protect consumers therein. Nonetheless, the aim of AI legislation is also to prevent spread of disinformation at hate speech, while the use of AI, large language models, and chatbots can lead to radicalizing certain social groups.

The EU's regulation of AI has significant effect on subjects residing outside the EU, especially high-tech companies of outside the EU entering the European market. The EU's digital market has over 300 million consumers and it is one of the usages of these consumers' data is one of the main concerns of legislators. With the global impact of AI technologies, international borders become less significant, and it is increasingly common that domestic laws are applied extraterritorially^[1].

Recent developments have shown that the EU's aspiration to protect its

consumers may be paid for by growing animosity between the EU and the U.S. While the EU aims to regulate all subjects trading in its territory equally, the U.S.'s administration has recently demonstrated that they will back U.S.'s giant companies violating EU laws.

These developments raise questions about the EU's sovereignty and capacity to legislate in its territory and about the possible fragility of the EU's aspiration to act as a pioneering subject in regulating the use of AI and diminishing its possible risks.

2. The European regulatory framework, the race to AI legislation, and relevance to foreign subjects

The EU has developed a comprehensive piece of legislation to regulate the use of AI - the Artificial Intelligence Act^[2] ("AIA" hereinafter). Such regulation serves many purposes, based on avoiding risks of manipulating with vulnerable groups, social scoring, biometric identification and categorization of people, and real-time identification of people in public spaces.

While the EU has introduced a regulatory regime for the use of AI, other jurisdictions are attempting to do the same. One of the driving forces to such attempts to legislate is the increasing visibility of the fact that the technology is carrying dangers regarding legality and ethics. Together with the ongoing «*race to AI*»^[3], which is motivated by securing a competitive place on the global market, governments are also perceiving the risks of using AI and at the same time taking part in a «*race to AI regulation*»^[4]. Recent developments have on one hand shown that there may be vast differences in the preferences for the application of AI regulations, caused by economical, political, and cultural factors. On the other hand, recent events have also shown that regulatory frameworks in various jurisdictions have started to converge, potentially leading to harmonizing AI regulation worldwide.

The EU regulatory framework for the use of AI is not only significant as a role-model for other jurisdictions. According to the concept of «*extraterritoriality*» stated in international law, countries are permitted to exercise their jurisdictions over foreign subjects in their territory, be it individuals or corporations^[5].

Any regulation aimed at limiting the risks of using AI is based on introducing

means to influence the behaviour of actors using AI, be it individuals and companies. Lessing defines four modalities, law (regulatory modality) being one of them, the other three modalities being social norms, the market, and architectural design of technological applications^[6]. Legal regulation of AI has several roles, the most prominent ones being a protective role (*i.e.* steering behaviour towards minimising adverse impacts) and an enabling role (*i.e.* stimulation of beneficial innovation). There is no clear distinction between the two and they overlap^[7]. Recent case studies discussed below, however, show, that in the recently exacerbated conflict between the U.S. and the EU over regulating American tech companies on the European market, the EU's efforts to protect its consumers through the emerging legislation framework may lead to the hinderance of economic growth and the decrease in the competitiveness of European-based high tech companies.

While the regulation of AI in the EU has mostly happened through the introduction of new laws, it may as well come to place through amending or derogating the existing laws, or introducing soft-law measures^[8]. The EU aims to lead the way with AI regulation, as it successfully did with data protection legislation, but is now clear that other jurisdictions target at different outcomes and take different legislative means. While the EU regulation is mostly focused on outcomes, other approaches are centred around regulating processes, and some combine both aspects^[9].

One of the difficulties posed to any efforts in adopting a coherent and universally applicable piece of legislation regulating AI is the fact that there is no single, universally accepted definition of AI as of today. The lacking definition of AI is posing complications to governments aiming to adopt a leading legislation. Many risks of AI may also arise with other technologies^[10]. Despite attempts to define it by various bodies such as European Commission's High-Level Expert Group on AI, and the scope of applications counted as falling under the umbrella term of AI constantly changes^[11]. However, such lack of a universally accepted and legally binding definition has formerly proven not to be an obstacle in the application of a sound regulatory framework. The EU's GDPR^[12] is an example of technology that is neutral towards specific technologies, because it is focused on the aims to be achieved^[13]. Similarly to the GDPR, the AI regulation can affect domains as distant from each other as tax law, public procurement law, consumer

protection, health law, tort law, privacy and data protection law, consumer protection law, and many other areas^[14].

Apart from the EU, other legislators worldwide are attempting to regulate AI and to minimize its risks and adverse effects on consumers or markets. These include the U.S., Canada, China, and other countries. As part of participation in the race to AI regulation, China has, similarly to the EU, put AI in the centre of its own economic strategy and it aims to create an international regulatory framework governing the use of AI^[15]. It is at the same time clear that the legislators worldwide are aware of risks of regulating AI ranging from deepening inequalities to existential risks^[16].

The EU and China serve as two examples of vastly different approaches to extraterritoriality of the AI laws. The EU shows an explicit territorial extension, whereas China is an example of vertical regulation with a narrower territorial scope of application^[17]. The EU aims to be the global leader and the shaper of the international debate on regulating the AI and maintaining European values and it also aims to apply the AIA to global AI markets^[18]. Chinese regulation, on the other hand, focuses on Chinese citizens, and does not primarily aim at their protection, which is evident from the fact that the regulation was adopted without a significant societal debate and there is therefore a risk that the resulting policies will not align with citizens' interests^[19].

In addition to the EU and China, there are other international standards for using and developing AI systems. One of them comes from the Council of Europe which Established its *Framework Convention on AI* in 2024, with the purpose of making sure that AI activities follow human rights protection^[20]. There are also some non-binding instruments, such as the UNESCO's *Recommendation on the Ethics of Artificial Intelligence*^[21] and the OECD's *AI Principles*^[22]. Some AI regulating instruments are also derived from the Northern America, such as the U.S.'s proposed *Algorithmic Accountability Act* in 2019^[23], followed by an AI Bill of Rights in 2022^[24]. Canada established the so-called *Algorithmic Impact Assessment Tool*^[25] in 2023^[26]. Additionally, China published the Beijing *Artificial Intelligence Principles*^[27] in 2019^[28].

The need for AI regulation also comes from the fact that new technologies are also transforming governments' work^[29]. This is posing dangers because of AI's capacity to cause error or bias, which is mostly problematic in areas where

governments have traditionally practices discretions, and thus such areas are traditionally connected with the society's tolerance towards limited legal liability of the governments and decision-making bodies^[30].

Despite the global ongoing race to AI regulation, the EU AI Act is a pioneering piece of legislation in its coercive nature, including banning some applications altogether on moral grounds. The risk-based approach can be viewed as grounded in constructivist and critical political economy^[31]. It also has foundations in cultural political economy, where the EU Commission is seen to employ the risk heuristics to enact the EU's vision of rights-oriented and at the same time globally competitive common market^[32].

The EU legal standards and policies such as the GDPR, the Digital Markets Act^[33] ("DMA" hereinafter), and the DSA^[34] ("DSA" hereinafter) influence international tech giants. These regulations aim and succeed in preventing international tech companies from engaging in anticompetitive behaviour. On the other hand, such regulations lead to prolonged disputes and contribute to Europe's falling behind the tech developments. Because of the data protection laws, European companies gain access to large language models much later than U.S. companies and companies in many other states in the world. These delays pose obstacles in European innovation, and critics argue that data protection is not worth such delays^[35].

3. Case studies

3.1. The Digital Services Act and recent sanctions to X

The acute need for regulating digital technologies together with the problematic nature of cross-border and transatlantic working of such regulation has recently become evident with the European Commission's decision^[36] to impose sanctions on Elon Musk's X over violating the Digital Services Act.

The DSA, as well as the general concept of the regulation of digital technologies and social media platforms, has recently been challenged, especially by the U.S. administration. Such concerns are driven by the fears ranging from those of unnecessary economic hinderance to those of limiting freedom of expression. Preceding the AIA, the DSA is within these discussions often regarded as a

central component of an emerging regulatory framework^[37]. Even though the DSA is an important step in the European Union's regulation of online services, as it addresses issues such as online disinformation and harmful content, its effectiveness is largely dependent on consistent and robust enforcement throughout the EU^[38] and, as we have recently witnessed, even beyond the EU's borders. As there is increasing attention paid to the DSA's enforcement framework, it has now become obvious that with the digital services market outreaching the EU's borders, we have just so far seen a tip of the iceberg as for the possible contested issues of cross-border application.

On 5 December 2025, the European Commission imposed a fine of €120 million on X over violations of the DSA. Specifically, the violations were seen firstly, in the deceptive design which exposes users to scams; secondly, in the lack of transparency over X's ads depository; and thirdly, in the failure to provide researchers access to public data^[39].

On 6 December 2025, Elon Musk responded to the fine and suggested dissolving the European Union^[40]. On 7 December 2025, Trump administration issued an official statement in support for Musk, and the United States imposed visa bans on five individuals, including Thierry Breton, who had had a leading role in the development and implementation of the DSA. The visa ban was rejected by the European Parliament^[41].

The official position of the U.S, and the Trump administration can be seen as an appeal on the EU to allow the U.S, companies to trade within the European market under the same conditions as would be applicable to them in the United States. However, such approach would disable the EU to apply its laws within its territory and as a result, the EU would be disallowed to protect its consumers. In short, granting exemptions to American companies would lead to clashes with the established principles of jurisdiction and regulatory sovereignty.

While the U.S. government has clearly expressed its view on European regulation of digital technologies – *i.e.* apprehending EU regulations as disproportional burdens, the EU maintains that its laws apply equally to all subjects. The visa bans to certain individuals to enter the U.S. indicate that the Trump administration is willing to defend the interests of American companies. The EU's stance, on the other hand, indicates that enforcement measures are based on criteria such as company size and market impact, rather than the company's country of origin.

While Musk's X as well as some media, including *The Guardian*, are presenting the issue as a limitation to freedom of expression^[42], the wording of the European Commission's decision makes it clear that the DSA application and EU's policies overall are aimed at non-discriminatory application to all subjects^[43].

The recent case has exacerbated the debate about jurisdictional claims and discrimination. While the DSA is highly relevant to the United States, as the U.S. is a home to many large and impactful tech companies, and while the EU aims and succeeds in shaping global legislative standards for technological regulation, it is more than clear that any limitations imposed on the free digital market can lead to the hinderance of economic growth in the EU and lowering the EU's competitiveness in comparison to other global high tech markets.

While there have not been any cases of sanctions to companies based on the violations on the AIA, it is obvious that the implementation of the AIA will lead to similar problems, as it, similarly to the DSA, aims at regulating digital platforms and technologies, albeit addressing different aspects of online services. The DSA together with the AIA present complimentary parts of the EU's digital governance framework.

3.2. Grok and disinformation spread through system prompts

The European Commission on X over violating the DSA has not been the first issue when X has been regarded as violating EU's legislation as well as the global standards of freedom of expression and its justifiable limitations. Another case of X, specifically xAI's chatbot Grok, seen as transgressing European regulations happened in May 2025, while at the same time there were no means to impose sanctions or fines by the European Commission.

The AIA prohibits the use of «*subliminal techniques beyond a person's consciousness or purposefully manipulative or deceptive techniques, with the objective, or the effect of materially distorting the behaviour of a person*»^[44]. It is also prohibited to «*put into service an AI system that exploits any of the vulnerabilities of a natural person or a specific group of persons due to their age, disability or a specific social or economic situation, in a manner which would cause harm to such vulnerable persons*»^[45].

In May 2025, the xAI's chatbot Grok has been observed by many users bringing up the issue of "white genocide" in South Africa, while such mentions were unprompted and were presented by the chatbot together with unrelated responses to other issues^[46]. Following this event, X has issued Grok's system prompt to increase transparency. This event has revealed more about Grok's workings than what was expected from the "white genocide" unprompted mentions: It was revealed that Grok is programmed to «*be extremely sceptical*» in its responses and not to «*defer to mainstream narratives*», while it is encouraged to «*always seek alternative truth*»^[47].

Such way of programming the AI large language models, *i.e.* to spread disinformation, false narratives, or hate speech, is clearly in violation with the wording of the AIA. Even though the AIA is not applicable in this case, the above issues of cross-border applicability of EU's standards of AI may soon become relevant to similar cases. The European Commission has identified disinformation as a threat to society^[48]. The digital sphere enables actors to spread disinformation targeted to prone individuals and to targeted groups with precision^[49]. While disinformation is not a new phenomenon, online social media and large language models have accelerated the spread of disinformation^[50]. Because disinformation is exacerbated in social bubbles, people become increasingly disinformed through algorithms sharing information among like-minded individuals^[51]. As a result, people are increasingly prone to half-truth and value judgements^[52]. These processes can force people into rabbit holes, *i.e.* make them get into deeper misinformed states, resulting in radicalisation^[53].

Up to recently, regulating freedom of expression has been seen as granting one of the essential human rights and posing some limits, and such human-centred approach to free expression and criminalization of hate speech would certainly have been applicable to foreign nationals in the EU's territory. These concepts were however only familiar when we had in mind naturally occurring language. When we come to large language models and chatbots, there is much less transparency about how discourses are being produced^[54] and the legislation standards are not yet ready to work in an equally efficient way as they do with naturally occurring language. At the same time, it has proven to be extremely difficult to regulate the production of speech by large language models, as they are constant processes rather than finalized products^[55].

We are entering a post-human future, while we are also inheriting certain dangerous codes from our modern past^[56], and while the residues of modern racisms and sexism have not yet been overcome, they are being perpetuated through generative AI. There may be cultural variations among jurisdictions on how the discourses of hate speech are criminalized, prosecuted, and punished^[57], and the introduction of a universal regulatory mode applicable across borders is even more so problematic.

3.3. GDPR violations sanctions and relation to AI regulation

The AIA entered into force on 1 August 2024. It has a phased implementation schedule, and thus its most salient provisions regarding enforcement will enter into force gradually, some as late as in 2027. Therefore, we have not yet witnessed any sanctions imposed upon the AIA violations, similar to the penalties according to the DSA discussed above.

The case law of the Court of Justice of the EU related to the AIA so far only concerns interpretation and application issues^[58]. In practice, some of the recent violations of the AIA provisions or principles have been dealt with by domestic courts in EU Member States under the regime of GDPR violations, as these constituted violations of data-processing rules by AI systems^[59].

4. Conclusions

The EU has recently become part of tensions over several issues with the U.S. With the ongoing peace negotiations and the rapid development of technologies offered by American companies on the European market, such tensions are exacerbating. The late 2025 has seen a major conflict between Trump administration and the European Commission over a fine imposed on Elon Musk's X over violations of the DSA. This incident did not, however, arrive out of the blue and had been preceded by other cases of violations of European regulations by global technological companies, while this raised questions about the applicability of European law on foreign subjects and allegations of inequality and discrimination.

With the European Commission's fine over DSA violations, a discussion has

been instigated about limitations on freedom of expression and this has been framed by some media, such as *The Guardian*, as a «war for freedom of speech». The reaction of Trump's administration, including entry bans to five European individuals, was centred around claims of the EU's "censorship of online platforms". Therefore, it is unsurprising that one of the five *personae non gratae* was Thierry Breton, a former European Commissioner and one of the leading figures who had overseen the EU regulation of digital technologies and the regulation of companies on the EU market.

Some large U.S. and other global tech firms, while supported by other actors such as the U.S. government, have accused the EU of discrimination against them on the European market. The EU has, on the contrary, introduced legislation and rules for companies trading with digital technologies on its market, in order to avoid discrimination and apply unified standards to all subjects within this realm. Europe presents a digital market of more than 300 million consumers. Therefore, any decisions taken in terms of regulating the use of digital technologies are vital to foreign companies. Limiting the EU's regulatory initiatives and giving up to the U.S.'s pressure would weaken the autonomy the EU has so far had in terms of regulating its market.

1. J. Lee, *Artificial intelligence and international law*, Singapore, Springer, 2022.
2. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act).
3. N. A. Smuha, *From a 'race to AI' to a 'race to AI regulation': regulatory competition for artificial intelligence*, in *Law, innovation and technology*, 1, 2021.
4. *Ibid.*
5. R. Arnold, J. Cremades, *Rule of Law, Technology and Environment: Contributions to the World Law Congress 2023*, Cham, Springer Nature, 2025.
6. L. Lessig, *The Law of the Horse: What Cyber Law Might Teach*, in *Harvard Law Review*, 1, 1999.
7. Smuha, *op.cit.*
8. W. Hoffmann-Riem, *Artificial Intelligence as a Challenge for Law and Regulation*, in *Regulating Artificial Intelligence*, eds. Wischmeyer, T. and Rademacher, T., Springer, 2020.
9. P. Frantz, N. Instefjord, *Regulatory Competition and Rules/Principles-Based Regulation*, in

- Journal of Business Finance and Accounting*, 1, 2018.
10. Smuha, *op.cit.*
 11. M. Haenlein, A. Kaplan, *A Brief History of Artificial Intelligence: On the Past, Present, and Future of Artificial Intelligence*, in *California Management Review*, 1, 2019.
 12. Regulation 2016/679/EU of the European Parliament and of the Council of 27 April 2016 concerning the protection of individuals with regard to the processing of personal data, as well as the free circulation of such data and which repeals Directive 95/46/EC (General Data Protection Regulation).
 13. Smuha, *op.cit.*
 14. *Ibid.*
 15. E. Gibney, *China wants to lead the world on AI regulation - will the plan work?*, in *Nature*, 1, 2025.
 16. *Ibid.*
 17. Y. Wang, *Do not go gentle into that good night: The European Union's and China's different approaches to the extraterritorial application of artificial intelligence laws and regulations*, in *The computer law and security report*, 53, 2024.
 18. E. Thelisson, H. Verma, *Conformity Assessment under the EU AI Act General Approach*, in *AI and Ethics*, 4, 2024.
 19. H. Roberts, et al., *Governing Artificial Intelligence in China and the European Union: Comparing Aims and Promoting Ethical Outcomes*, in *Information Society*, 39, 2023.
 20. Council of Europe. *Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law*, Council of Europe Treaty Series No. 225 (opened for signature 5 September 2024; adopted 17 May 2024). Council of Europe, Strasbourg. ISBN 978-92-871-9621-7 (bilingual French/English edition).
 21. United Nations Educational, Scientific and Cultural Organization (UNESCO). *Recommendation on the Ethics of Artificial Intelligence*. Adopted 23 Nov. 2021, UNESCO Digital Library, 2022.
 22. Organisation for Economic Co-operation and Development (OECD). *Recommendation of the Council on Artificial Intelligence*, OECD/LEGAL/0449, adopted 22 May 2019 (revised 3 May 2024).
 23. Algorithmic Accountability Act of 2019, H.R. 2231, 116th Cong. (U.S. House of Representatives April 10, 2019), introduced by Rep. Yvette Clarke (D-NY), referred to the House Committee on Energy and Commerce. U.S. Congress, Library of Congress Congress.gov.
 24. The White House Office of Science and Technology Policy. *Blueprint for an AI Bill of Rights: Making Automated Systems Work for the American People*. Washington, DC: The White House, October 2022.
 25. Treasury Board of Canada Secretariat. (2023). *Algorithmic Impact Assessment Tool (AIA)*. Government of Canada.
 26. R. Paul, *European artificial intelligence 'trusted throughout the world': Risk-based regulation and the fashioning of a competitive common AI market*, in *Regulation &*

- governance*, 4, 2024.
27. Beijing Academy of Artificial Intelligence, Peking University, Tsinghua University, Institute of Automation (Chinese Academy of Sciences), Institute of Computing Technology (Chinese Academy of Sciences), & AI Industry Innovation Strategy Alliance (AITISA). (2019, May 25). Beijing Artificial Intelligence Principles. Beijing AI Principles. Retrieved from linking-ai-principles.org.
 28. *Ibid.*
 29. C. Coglianese, D. Lehr, *Transparency and algorithmic governance*, in *Administrative Law Review*, 1, 2018.
 30. D. F. Engstrom, A. Haim, *Regulating Government AI and the Challenge of Sociotechnical Design*, in *Annual review of law and social science*, 1, 2023.
 31. R. Paul, *op.cit.*
 32. J. Bareis, C. Katzenbach, *Talking AI into being: The narratives and imaginaries of national AI strategies and their performative politics*, in *Science, Technology, & Human Values*, 5, 2022.
 33. Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act), Official Journal of the European Union L 265, 12 October 2022, pp. 1–66.
 34. Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act).
 35. N. Kshetri, *Navigating EU Regulations: Challenges for U.S. Technology Firms and the Rise of Europe's Generative AI Ecosystem*, in *Computer*, 10, 2024.
 36. Commission Decision, C(2025) 8630 final, decision of 5.12.2025 pursuant to Articles 73(1), 73(3) and 74(1) of Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market for Digital Services and amending Directive 2000/31/EC (Digital Services Act) Cases DSA.100101, DSA.100102 and DSA.100103 – X (formerly Twitter).
 37. M. Fasel, *Is the Digital Services Act Here to Protect Users? Platform Regulation and European Single Market Integration*, in *Nordic Journal of European Law*, 4, 2024.
 38. P. Mattioli, *Navigating the Complexities of the DSA's Enforcement Framework: Sincere Cooperation in Action?*, in *Utrecht Law Review*, 1, 2023.
 39. Commission Decision, C(2025) 8630 final, *op.cit.*
 40. Elon Musk's X account, 6 December, 2025, available from <https://x.com/elonmusk/status/>, date accessed 15 February 2026.
 41. European Parliament, *EP leaders reject the visa ban imposed on former Commissioner Breton*, 22 January 2026, available from <https://www.europarl.europa.eu/news/cs/press-room/20260122IPR32567/ep-leaders-reject-the-visa-ban-imposed-on-former-commissioner-breton>, date accessed 15 February 2026.

42. H. Horton, *Elon Musk says UK wants to suppress free speech as X faces possible ban*, in *The Guardian*, 10 January 2026, available from <https://www.theguardian.com/technology/2026/jan/10/elon-musk-uk-free-speech-x-ban-grok-ai>, date accessed 15 February 2026.
43. Commission Decision, C(2025) 8630 final, *op.cit.*
44. AIA, *op.cit.*, Article 5(1)a).
45. AIA, *op.cit.*, Article 5(1)b).
46. H. Murphey, C. Criddle, *Elon Musk's AI chatbot shared 'white genocide' tropes on X*, in *Financial Times*, 14 May 2025, available from <https://www.ft.com/content/37416a0e-8f35-45af-9ace-2cf4c973daa5>, date accessed 15 February 2026.
47. E. Steedman, *For hours, chatbot Grok wanted to talk about a 'white genocide'. It gave a window into the pitfalls of AI*, in *Australian Broadcasting Corporation*, 24 May 2025, available from <https://www.abc.net.au/news/2025-05-25/grok-ai-accuracy-doubts-after-white-genocide-claims-fixation/105325028>, date accessed 15 February 2026.
48. T. Enarsson, *Countering 'lawful but awful' disinformation online: EU-regulations targeting disinformation on major social media platforms*, in *Nordic Journal of European Law*, 3, 2025.
49. E. De Blasio, D. Selva *Who Is Responsible for Disinformation? European Approaches to Social Platforms Accountability in the Post-Truth Era*, in *American Behavioral Scientist*, 2021.
50. W. Mbioh, *Beyond Echo Chambers and Rabbit Holes: Algorithmic Drifts and the Limits of the Online Safety Act, Digital Services Act, and AI Act*, in *Griffith Law Review*, 3, 2024.
51. S. Steinert, M. J. Dennis, *Emotions and Digital Well-Being: On Social Media's Emotional Affordances*, in *Philosophy & Technology*, 1, 2022.
52. C. Diaz Ruiz, T. Nilsson, *Disinformation and Echo Chambers: How Disinformation Circulates on Social Media Through Identity-Driven Controversies*, in *Journal of Public Policy & Marketing*, 1, 2023.
53. B. Lewis, *Rabbit Hole: Creating the Concept of Algorithmic Radicalization*, in J. Farkas, M. Maloney (eds), *Digital Media Metaphors*, Routledge, 2024.
54. M. Gillings, T. Kohn, G. Mautner, *The rise of large language models: Challenges for Critical Discourse Studies*, in *Critical Discourse Studies*, 2, 2024.
55. N. S. Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism*, 2018, New York University Press.
56. A. Adib-Moghaddam, *Is Artificial Intelligence Racist? The Ethics of AI and the Future of Humanity*, 2023, Bloomsbury.
57. E. Siapera, *AI content moderation, racism and (de) coloniality*, in *International journal of Bullying Prevention*, 1, 2022.
58. Case C-806/24: A preliminary ruling request referred to the CJEU on 25 November 2024 by the Sofia District Court (Bulgaria).

CERIDAP

59. Autoriteit Persoonsgegevens. (2024, May 16). Administrative decision imposing a €30.500.000 fine on Clearview AI Inc. for violations of the General Data Protection Regulation (GDPR) (Decision, Netherlands). Autoriteit Persoonsgegevens; Commission nationale de l'informatique et des libertés (CNIL). (2022, October 19). Sanction imposing a €20.000.000 fine on Clearview AI for GDPR violations. CNIL.

Artificial Intelligence and discretionary powers in public administration

Vladimír Šarp

DOI: 10.13130/2723-9195/2026-3-23

Il presente articolo esamina il rapporto tra Intelligenza Artificiale e discrezionalità amministrativa nell'ambito della pubblica amministrazione, e sostiene che la diffusa convinzione secondo cui le decisioni discrezionali non possano essere automatizzate si fondi su una concezione eccessivamente semplificata e binaria della discrezionalità. Muovendo da un'analisi della dottrina e della prassi amministrativa, l'articolo dimostra che la discrezionalità si manifesta secondo gradi e modalità differenti ed è spesso strutturata attraverso linee guida, metodologie e prassi amministrative consolidate. In molti casi, poteri discrezionali formalmente ampi operano, nella pratica, come una forma di "pseudo-discrezionalità", nella quale criteri standardizzati limitano significativamente la gamma dei possibili esiti. L'articolo suggerisce, pertanto, che la questione centrale non sia stabilire se l'Intelligenza Artificiale possa esercitare, in quanto tale, la discrezionalità, bensì di individuare quale tipo di discrezionalità venga in rilievo. Mentre le decisioni che richiedono una valutazione autenticamente individualizzata possono rimanere compiti intrinsecamente umani, le forme di discrezionalità strutturata o di pseudo-discrezionalità potrebbero, in linea di principio, essere riprodotte mediante sistemi automatizzati.

This article examines the relationship between Artificial Intelligence and administrative discretion in public administration. It argues that the widespread assumption that discretionary decisions cannot be automated rests on an overly simplified, binary understanding of discretionary power. Drawing on doctrinal analysis and administrative practice, the article demonstrates that discretion operates in different degrees and modalities and is often structured by guidelines, methodologies, and established administrative practice. In many cases, formally broad discretionary powers function in practice as what the article describes as "pseudo-discretion", where standardised criteria significantly constrain the range of

possible outcomes. The article, therefore, suggests that the key issue in the debate is not whether Artificial Intelligence may exercise discretion as such, but rather what type of discretion is involved. While decisions requiring genuine individualised judgement may remain inherently human tasks, forms of structured or pseudo-discretion may, in principle, be capable of being replicated through automated systems.

Summary: 1. Introduction.- 2. The concept of administrative discretion and its raison d'être.- 3. Administrative discretion in practice.- 4. Modalities of administrative discretion.- 5. Artificial intelligence and the replication of structured discretion.- 6. Conclusions.

1. Introduction

In recent years, the rapid development of artificial intelligence has given a somewhat new meaning to debates about the automation of administrative decision-making^[1]. Public administrations across Europe increasingly try to explore digital tools to process large volumes of cases, streamline administrative procedures, and improve efficiency. At the same time, the prospect of delegating administrative tasks to automated systems raises fundamental questions about legality, accountability, and the role of human judgment in public administration. Administrative discretion (understood primarily as the authority's power to choose among legally permissible outcomes) represents one of the most frequently invoked limits to automation and use of artificial intelligence. It is commonly assumed that while routine and strictly rule-based tasks may be automated, decisions involving discretion must remain within the exclusive domain of human officials.

This assumption has recently been reflected in legislative developments in the Czech Republic, as in the course of discussions on the modernization of the Code of Administrative Procedure, a proposal was introduced to explicitly regulate the possibility of automated administrative acts, allowing certain procedural acts to be performed automatically without the participation of an official, while explicitly excluding the use of automation in situations where administrative discretion must be exercised^[2]. This model is, of course, not

unique to the Czech Republic, as other jurisdictions have adopted similar approaches to the automation of administrative decision-making. Such is, for instance, the case of the German VwVfG, stipulating that administrative acts may only be issued automatically provided that neither discretion nor room for interpretation is present in the case (although VwVfG does not prevent the legislator from adopting special legislation (*lex specialis*) as well as subsequently enacted legislation (*lex posterior*) modifying this setup)^[3].

At first sight, such a limitation may appear intuitive. Administrative discretion is often perceived as a domain that necessarily requires human judgment, sensitivity to context, and the ability to weigh competing interests. From this perspective, excluding discretionary decisions from automated processing seems to be a natural safeguard against the risks associated with algorithmic decision-making^[4]. A closer examination, however, reveals a significant conceptual flaw underlying this approach: modern administrative law in continental European legal systems is structured in such a way that elements of discretion are present in a large number of administrative decisions, including those that are otherwise highly routine and standardized. Legislative provisions frequently authorize administrative authorities to impose sanctions, grant benefits, or determine obligations within a range of possible outcomes^[5]. Formally speaking, such provisions grant the administrative authority a wide range of possible choices. If the existence of such discretion were interpreted as an absolute barrier to automation, the practical result would be that even the most routine administrative decisions could not be processed automatically.

This observation exposes a deeper problem in the way administrative discretion is often conceptualized, especially in the context of automated decision-making. Much of the current debate implicitly assumes that discretion operates in a binary manner: either a decision involves discretion (in which case automation should be excluded), or it does not (in which case automation may be permissible). In the present article, I intend to challenge this assumption. I argue that administrative discretion is not a uniform or binary phenomenon, but rather a linear concept that manifests itself in different degrees and modalities. While some forms of discretion indeed involve complex value judgments that are difficult to formalize, other forms are already highly structured and constrained in administrative practice.

The gap between the theoretical conception of discretion and its practical application plays a central role in this discussion. This article proceeds from the hypothesis that in practice, administrative authorities rarely exercise discretion in a completely open-ended manner, which manifests itself in the fact that discretionary powers are often structured through internal guidelines, established administrative practice, and methodological frameworks that seek to ensure consistency and equality in decision-making. As a result, many discretionary decisions are in reality guided by predetermined criteria that significantly limit the range of acceptable outcomes. Recognizing this reality has important implications for the debate on automated decision-making in public administration. If certain forms of discretion are already exercised within a structured framework of criteria and established practice, the categorical exclusion of automated systems from all discretionary decisions becomes difficult to justify. Rather than treating discretion as an absolute barrier to automation, it may be more appropriate to examine the different forms that discretion can take and to consider whether some of them are compatible with carefully designed forms of automated decision-making.

The aim of this article is therefore to reconsider the relationship between administrative discretion and artificial intelligence, or automation of administrative decision-making in general. This objective can be translated into the research question of whether the presence of administrative discretion necessarily precludes the use of automated decision-making in public administration, or whether this incompatibility is based on an overly simplified binary understanding of discretion. I seek to demonstrate that the prevailing binary understanding of discretion obscures important distinctions within the concept itself and leads to an overly restrictive approach to automation in public administration. Thus, by examining the doctrinal foundations of administrative discretion, its practical application, and the different modalities in which it operates, this article aims to develop a framework for assessing the compatibility of artificial intelligence with certain forms of discretionary decision-making.

The analysis, based upon doctrinal methodology, proceeds in several steps. I first examine the concept of administrative discretion from a doctrinal perspective and explore the reasons why legal systems rely on discretionary powers in the first place. The research focus then turns to the practical operation of discretion in

administrative decision-making, highlighting the extent to which discretionary powers are structured and constrained in practice. Building on these observations, the subsequent section develops a more nuanced understanding of discretion by distinguishing between different degrees and modalities of discretionary decision-making. Against this background, I finally examine the potential role of artificial intelligence in replicating certain forms of structured discretion and explore alternative models for integrating automated systems into administrative decision-making without abandoning the safeguards traditionally associated with discretionary power.

2. The concept of administrative discretion and its *raison d'être*

It would probably be fair to call administrative discretion one of the central concepts of modern administrative law. In the simplest terms, the concept of administrative discretion (also known as administrative consideration or discretionary power)^[6] refers to situations in which the law authorizes an administrative authority to choose among several legally permissible solutions when deciding a particular case^[7]. Rather than dictating a single predetermined outcome, legislation often leaves decision-makers with a range of permissible options within which they must exercise judgment. As a result, administrative authorities must interpret statutory provisions, evaluate factual circumstances, and select an appropriate course of action within legally defined boundaries^[8]. Hence, unlike strictly bound decision-making, where the legal norm prescribes a single outcome once the relevant facts are established, discretionary powers allow the administrative authority to determine which of several possible outcomes is most appropriate under the circumstances of the case^[9].

In doctrinal terms, administrative discretion typically arises when the legal norm is formulated in a way that does not fully predetermine the content of the administrative decision. This may occur, for example, where the law authorizes an authority to impose a sanction “up to” a specified amount^[10], to grant or refuse a permit depending on an assessment of certain public interests, or to select one of several possible regulatory measures. In such situations, the law establishes the legal framework within which the authority must operate, while

leaving a margin of choice regarding the final decision. The exercise of discretion therefore presupposes that several outcomes are legally permissible and that the authority must determine which of them should be applied in the specific case.^[11] At this point, it is necessary to clarify that the concept of administrative discretion is not understood uniformly across legal systems. For instance, in German administrative law, a distinction is traditionally drawn between “Rechtsfolgenermessen”, referring to the authority’s discretion in selecting among several legally permissible legal consequences, and “Beurteilungsspielräume”, which concern the administration’s margin of appreciation in assessing whether the legal requirements for administrative action are fulfilled. While the former constitutes the purer form of administrative discretion, the latter occupies a more contested position. German doctrine generally proceeds from the assumption that vague legal terms do not, as a rule, confer autonomous margins of appreciation upon the administration but remain fully subject to judicial interpretation, save for certain exceptional categories recognized by legislation and case law. By contrast, many other legal systems treat both the determination of legal consequences and the assessment of open-ended legal concepts as manifestations of administrative discretion. The present article is primarily concerned with the former category, namely discretionary powers relating to the choice among legally permissible outcomes. Nevertheless, the broader comparative perspective illustrates that the concept of discretion itself admits different doctrinal configurations, reinforcing rather than weakening the argument that discretion should not be regarded as a single, homogeneous phenomenon^[12].

The existence of discretionary powers reflects certain structural characteristics of our administrative governance. Legislators are rarely able to foresee all possible factual situations that may arise in practice. If legal norms attempted to casuistically regulate every individual case in advance, they would either become excessively rigid or require an unrealistic level of detail. Discretion therefore functions (at least in theory) as an instrument that allows administrative authorities to adapt the application of the law to the particular circumstances of individual cases. By leaving a certain margin of choice to administrative authorities, the legal system acknowledges the limits of legislative foresight and enables the administration to respond flexibly to diverse and evolving situations.

Another important function of administrative discretion lies in the pursuit of fairness and proportionality in administrative decision-making. Strictly rule-based systems may produce outcomes that are formally correct but substantively inappropriate in particular circumstances^[13]. Discretion allows the administrative authority to take into account relevant contextual factors and to balance competing interests in a manner that better reflects the underlying purpose of the law. In this sense, discretion serves as a mechanism that helps to mitigate the rigidity that might otherwise result from the application of general legal rules to highly diverse factual situations^[14].

At the same time, the exercise of discretion is not unlimited. Even where the law grants a margin of choice, the administrative authority remains bound by the legal order and must exercise its discretion within the limits established by law^[15]. Administrative law doctrine therefore traditionally distinguishes between the existence of discretion and its lawful exercise. While the authority may choose among several possible outcomes, this choice must respect the purpose of the legal norm, the general principles of administrative law, and the requirements of equality and proportionality. Discretionary powers must not be exercised arbitrarily or in a manner that is inconsistent with the objectives pursued by the legislature^[16]. These limitations are reflected in the doctrinal concept often described as the prohibition of misuse of discretion. Administrative authorities must not base their decisions on irrelevant considerations, must take into account all relevant circumstances of the case, and must not exceed the limits of the discretion granted by law. In addition, discretionary decisions are subject to judicial review, although courts typically refrain from substituting their own assessment for that of the administrative authority and instead examine whether the authority has remained within the legal boundaries of its discretionary power^[17].

Although the concept of administrative discretion is described in doctrine by the term “power”, which might suggest that the administrative authority is free to decide whether it will exercise discretion at all, such an interpretation would be misleading. The fact that the law grants a discretionary margin does not mean that the authority may arbitrarily choose whether or not to use it. If the legal norm requires that certain factors be assessed through discretionary evaluation, the administrative authority is not entitled, but rather obliged to carry out that

evaluation and to base its decision on it. The Czech Supreme Administrative Court has observed that *«the counterpart of the discretionary power of an administrative authority is the duty to exercise such discretion, that is, to consider all factors that the law establishes as premises of such reasoning, to obtain the necessary evidence for that purpose, to derive factual and legal findings from that evidence, and subsequently, through discretionary reasoning, to reach an appropriate decision»*^[18].

In other words, discretion is not merely a privilege granted to administrative authorities but also a procedural and substantive obligation. If an authority fails to exercise discretion in situations where the law requires it, its decision would be unlawful.

From a doctrinal perspective, administrative discretion is therefore commonly understood as a legally structured margin of choice that enables administrative authorities to adapt the application of the law to the specific circumstances of individual cases. It reflects the recognition that legal regulation cannot fully anticipate the complexity of social reality and that administrative decision-making requires a certain degree of flexibility in order to achieve fair and proportionate outcomes. This traditional understanding of administrative discretion plays an important role in contemporary debates on the automation of administrative decision-making. Discretion is frequently perceived as the domain of human judgment and as a safeguard against overly mechanical forms of governance. For this reason, many legislative proposals and scholarly contributions assume that discretionary decisions should remain outside the scope of automated processing. Before accepting such conclusions, however, it is necessary to examine more closely how discretionary powers are actually exercised in administrative practice. As the following section will demonstrate, the practical operation of discretion often differs significantly from the theoretical model that is commonly described in administrative law doctrine.

3. Administrative discretion in practice

The doctrinal account outlined above portrays administrative discretion as a legally structured margin of choice that enables administrative authorities to adapt the application of legal norms to the specific circumstances of individual

cases. While this theoretical model accurately captures the legal framework within which discretion operates, the evidence suggests that the way discretionary powers are exercised in administrative practice often differs from this idealized model. While “discretion in the books” is based upon the abstract ideal in which every case is individually decided upon while carefully considering all the evidence and all the specific facts and circumstances of each case, “discretion in action” appears to be much more mundane. In reality, discretionary decision-making rarely takes the form of an entirely open-ended evaluation in which the administrative authority freely selects one of many possible outcomes. Instead, it appears that the exercise of discretion is typically shaped and constrained by a variety of practical mechanisms that aim to ensure consistency, predictability, and equal treatment in the best-case scenario, or realistically to simply increase efficiency^[19].

One important reason for this development lies in the institutional nature of administrative decision-making. Administrative authorities do not operate in isolation, but rather as organizations responsible for deciding large numbers of similar cases. In such a context, purely individualized and unconstrained exercises of discretion would inevitably lead to significant variations in outcomes. If different officials were to decide comparable cases in substantially different ways, the resulting disparities would raise concerns regarding equality before the law and the legitimacy of administrative action. For this reason, administrative bodies often develop internal mechanisms that guide the exercise of discretion and help to align individual decisions with a broader institutional practice^[20].

The tendency to structure discretionary decision-making is not merely a practical response of administrative bodies but is also reflected in broader European administrative law standards. Guidance on the exercise of discretionary powers has long emphasized that discretion must be exercised within a framework of objective and consistent criteria. This approach is reflected, for example, in Recommendation No. R (80) 2 of the Committee of Ministers of the Council of Europe concerning the exercise of discretionary powers by administrative authorities. The Recommendation establishes a number of principles designed to prevent arbitrariness and ensure consistent decision-making. Among other requirements, administrative authorities exercising discretion must observe objectivity and impartiality, take into account only relevant factors, respect the

principle of equality before the law, and apply any general administrative guidelines in a consistent manner while still considering the particular circumstances of each case. These principles illustrate that discretion has long been understood in European administrative law not as an unfettered freedom of decision but as a legally structured form of decision-making that operates within a set of normative constraints^[21].

These mechanisms take a variety of forms. In many areas of administrative law, discretionary decisions are guided by internal guidelines, methodological instructions, established administrative practice^[22], but also constrained by the principle of proportionality, which operates as a fundamental limit on the exercise of discretionary powers and simultaneously structures judicial review^[23]. Such instruments typically specify criteria that should be taken into account when exercising discretion, identify relevant aggravating or mitigating circumstances, and sometimes even indicate approximate ranges of appropriate outcomes in typical cases. Although these instruments are typically only internally binding and thus neither have the same external effect as statutory provisions nor limit the legal obligation of the administrative authority to differentiate between specific cases and individualize its decisions, they nevertheless play a significant role in shaping the day-to-day operation of administrative decision-making and have significant *de facto* power over officials who are in principle obligated to adhere to internal instruments^[24].

A clear illustration of this phenomenon can be found in the field of administrative punishment. Legal provisions frequently authorize administrative authorities to impose sanctions within a relatively broad range. A statutory provision may, for example, allow the imposition of a fine “up to” a certain maximum amount. From a purely formal perspective, such a provision should be considered the highest form of discretion, since it grants the authority a wide range of possible outcomes. In practice, however, administrative authorities rarely treat each possible amount within this range as an equally plausible option. Instead, sanctioning practices tend to develop around relatively stable patterns, often supported by internal guidelines or methodological frameworks that link the severity of the sanction to specific factors such as the seriousness of the offence, the circumstances of its commission, or the conduct of the offender^[25]. These forms of structured decision-making reflect an effort by administrative

authorities to reconcile the flexibility inherent in discretionary powers with the need for consistency and predictability. By articulating common criteria for the exercise of discretion, administrative bodies seek to reduce the risk of arbitrary differences between similar cases and to provide a degree of transparency regarding the reasoning underlying their decisions. In this way, discretionary powers that appear broad at the level of statutory formulation may in practice be exercised within a significantly narrower and more structured framework^[26].

Another factor that contributes to the structuring of discretion is judicial review. Courts reviewing discretionary decisions typically refrain from substituting their own assessment for that of the administrative authority, but they nevertheless examine whether the authority has exercised its discretion in a lawful and rational manner. Requirements such as the prohibition of arbitrariness, the obligation to consider relevant circumstances, and the principle of proportionality encourage administrative authorities to articulate and apply coherent criteria when exercising discretionary powers. Over time, these expectations may further consolidate administrative practices that promote consistency across cases^[27].

Taken together, these developments reveal an important feature of modern administrative governance: while discretion remains formally embedded in many legal norms, its practical exercise is often mediated by a variety of institutional and methodological constraints. Discretionary decision-making therefore frequently operates within a structured environment that limits the range of acceptable outcomes and channels the reasoning of administrative authorities through predefined criteria and established practices. This observation plays a crucial role in the context of automated decision-making. If discretionary powers are already exercised through structured frameworks designed to ensure consistency, the question arises whether certain aspects of such structured discretion might, at least in principle, be replicated or supported by automated systems. The following section therefore turns to a closer examination of the different degrees and modalities of discretion that can be identified within administrative decision-making.

4. Modalities of administrative discretion

The discussion in the preceding sections reveals an important tension between

the doctrinal understanding of administrative discretion and the way in which discretionary powers are actually exercised in administrative practice. While doctrine often describes discretion as a margin of choice between several legally permissible outcomes, the practical operation of discretion frequently takes place within structured frameworks that guide the decision-making process and limit the range of acceptable outcomes. This observation raises a broader conceptual question. If discretion is routinely exercised within such constraints, is it appropriate to treat it as a uniform and indivisible category, or should it instead be understood as a phenomenon that manifests itself in different degrees and modalities?

Traditional doctrinal discussions tend to approach discretion as a binary concept. Either a legal norm leaves the administrative authority with a margin of choice, in which case the decision is discretionary, or the norm determines the outcome with sufficient precision, in which case the decision is bound^[28]. This binary distinction undoubtedly serves an important analytical function: it allows legal doctrine to distinguish between situations in which the administrative authority merely applies the law and those in which it must also evaluate and select among different permissible outcomes. At the same time, however, this distinction may obscure significant differences between various forms of discretionary decision-making.

A review of different instances of discretion across legislation suggests that not all discretionary powers are alike. Some involve complex value judgments, the balancing of competing interests, and the evaluation of open-ended legal concepts. In such cases, the administrative authority must consider a wide range of contextual factors and determine which solution best reflects the purpose of the legal norm in the circumstances of the particular case. Decisions of this kind often involve a high degree of responsibility and normative evaluation, as the authority must effectively determine how the legal framework should be concretely applied to a specific situation^[29].

Other forms of discretion, however, operate in a markedly different manner. For instance, in many areas of administrative law, discretionary powers take the form of selecting a value or outcome within a numerically defined range. The law may authorize the administrative authority to impose a fine “up to” a certain amount, to determine the duration of a measure within specified limits, or to select one of

several regulatory options that differ primarily in degree rather than in kind. Although such provisions formally leave a wide margin of choice, the practical decision-making process often resembles the application of a structured methodology rather than an open-ended evaluative judgment^[30].

A further form of discretion can be observed in situations where the law formally confers a discretionary power on the administrative authority, yet the practical outcome of the decision depends almost entirely on whether the applicant fulfils a set of predetermined criteria. This situation frequently arises in licensing or authorization procedures^[31]. In many jurisdictions and in certain categories of licensing procedures, legal provisions may state that the authority “may” grant a license if certain conditions are met, thereby formally framing the decision as discretionary. In practice, however, the decision-making process often revolves around the verification of whether the applicant satisfies the statutory requirements, such as professional qualifications, technical standards, or other objective conditions laid down by law. Once these criteria are fulfilled, the authority will normally grant the license, whereas their absence typically leads to refusal. Although such decisions may formally fall within the category of discretionary powers, the actual scope for evaluative judgment is often very limited. The task of the administrative authority primarily consists in verifying compliance with predefined conditions rather than weighing a broad range of policy considerations or selecting among qualitatively different outcomes. In this sense, the decision-making process resembles the application of a structured checklist more than the exercise of open-ended discretionary reasoning. The authority’s role is not so much to determine which of several possible outcomes would be the most appropriate, but rather to ascertain whether the legally prescribed criteria have been satisfied.

The difference between these forms of discretion becomes particularly visible when considering the structure of the choices available to the administrative authority. In some cases, the authority is effectively faced with a limited number of qualitatively distinct options. A decision whether to grant or refuse a particular status or authorization may involve only two possible outcomes, each carrying significant legal and social consequences. In other situations, however, the authority may formally be able to choose from a large number of possible outcomes that differ only incrementally. For example, where the law authorizes

the imposition of a fine of up to a specified maximum amount, the administrative authority may theoretically select from a wide spectrum of possible amounts. Yet this multiplicity of potential outcomes does not necessarily translate into a correspondingly broad exercise of evaluative judgment.

Paradoxically, the existence of a large number of formally available options may actually encourage the development of structured decision-making frameworks that reduce the effective scope of discretion. Faced with the possibility of choosing among hundreds or thousands of potential outcomes, administrative authorities often seek to ensure consistency by establishing methodological criteria that link particular factual circumstances to typical ranges of sanctions or measures. In such cases, the practical exercise of discretion increasingly resembles the application of a structured decision-making model rather than the exercise of entirely open-ended judgment, thus resembling what might be described as the “paradox of choice” of a kind^[32].

These observations suggest that administrative discretion is better understood as a spectrum rather than as a binary category. At one end of this spectrum lie forms of discretion that require complex normative evaluation and the balancing of competing considerations. At the other end are forms of discretion that are primarily technical or quantitative in nature and that are frequently exercised through structured methodologies designed to ensure consistency and predictability. Between these poles lies a variety of intermediate situations in which discretion is present but significantly shaped by institutional practices, guidelines, and methodological frameworks^[33].

Recognizing these different modalities of discretion has important implications for the debate on automated decision-making in public administration. If discretion is treated as a single, undifferentiated concept, it may appear natural to exclude all discretionary decisions from automated processing. If, however, discretion is understood as a spectrum encompassing different forms of decision-making, the relationship between discretion and automation becomes more nuanced. Certain forms of discretion that involve complex normative judgments may indeed require the involvement of human decision-makers. Other forms, particularly those that are already exercised through structured criteria and methodological frameworks, may be more compatible with carefully designed forms of automated decision support or decision-making. The purpose of

distinguishing between different degrees and modalities of discretion is therefore not to deny the importance of discretionary powers in administrative law, but rather to refine the conceptual framework within which discretion is understood. By recognizing that discretion operates in different ways across different types of decisions, it becomes possible to move beyond the simplistic assumption that the mere presence of discretion necessarily precludes the use of automated systems. This refined understanding provides the conceptual foundation for the analysis examining whether and to what extent artificial intelligence may be capable of replicating certain forms of structured discretion in administrative decision-making.

5. Artificial intelligence and the replication of structured discretion

Much of the existing scholarly debate on artificial intelligence and administrative decision-making is framed around the question of whether automated systems should be allowed to exercise discretionary powers^[34], or prohibited from doing so^[35]. This formulation of the problem, however, may itself be based on a questionable premise. From my perspective, the crucial issue is not simply whether artificial intelligence should or should not be allowed to use discretion, but rather what is meant by discretion in the first place. If discretion is understood as the process of deciding cases that require genuinely individualized assessment, complex contextual judgment, and the balancing of competing values, the argument that such decisions should remain within the domain of human officials is both intuitive and convincing. In such situations, the exercise of discretion indeed presupposes forms of reasoning and responsibility that are difficult to translate into purely technical processes and should be kept in the human domain. If, however, discretion refers to the application of structured and standardized decision-making frameworks that are designed to navigate the wide range of options formally provided by legislation, the picture appears considerably different.

Professor Juli Ponce Solé has famously remarked^[36] that artificial intelligence is “psychopathic”, in the sense that it lacks the capacity for empathy and moral sensitivity that may be required when exercising administrative discretion. From

this perspective, automated systems appear inherently unsuitable for decisions that require individualized assessment and the weighing of competing human interests. Modern administrative systems are not designed to encourage purely subjective or emotionally driven decisions. On the contrary, administrative law traditionally seeks to limit the influence of personal preferences and individual sensibilities by requiring officials to base their decisions on legally relevant criteria, to apply general guidelines consistently, and to treat comparable cases alike^[37]. In this sense, administrative institutions deliberately discipline the decision-making process in order to reduce variability between individual officials, thus rendering administrative decision-making similarly detached from individual emotional judgment. The exercise of discretion is therefore often structured in a way that minimizes the role of purely personal judgment and channels decision-making through predefined methodological frameworks. When discretion operates within such structured systems, the decision-making process may already resemble the application of standardized rules rather than the expression of individual moral intuition.

In many areas of administrative practice, the exercise of discretion already involves the systematic application of predefined criteria, internal guidelines, and methodological frameworks that substantially limit the freedom of the individual decision-maker. In this sense, the institutional design of discretionary decision-making often seeks to reduce variability and align individual decisions with established patterns of administrative practice. Where discretion operates in this highly structured manner, it becomes difficult to maintain that such decision-making is intrinsically human in nature. Rather, it may increasingly resemble the application of complex but formalizable decision rules, which raises the possibility that at least some forms of such structured discretion could, in principle, be replicated by technical means.

A categorical exclusion of discretionary decisions from automated processing would have far-reaching consequences for the practical deployment of artificial intelligence in public administration. As noted above, elements of discretion are present in a large number of administrative decisions, including those that are otherwise highly routine and standardized. Legal provisions frequently authorize administrative authorities to determine outcomes within a certain range, to assess compliance with various criteria, or to select among several permissible measures.

If the mere presence of such discretionary elements were interpreted as an absolute barrier to automation, the result would be that artificial intelligence could be used only in a very limited number of administrative tasks. Many areas of decision-making that are today considered among the most suitable candidates for automation would remain outside the scope of automated processing^[38].

This consequence appears difficult to reconcile with the practical realities of modern public administration. Administrative authorities are often required to process large volumes of relatively similar cases in a consistent and timely manner. In many such situations, the exercise of discretion already takes place within structured frameworks that guide the decision-making process. Internal guidelines, methodological instructions, and established administrative practices frequently provide detailed criteria that link specific factual circumstances to particular types or ranges of outcomes. As discussed in the previous sections, the practical exercise of discretion in these contexts often resembles the application of a structured methodology rather than the exercise of entirely open-ended evaluative judgment.

I believe that the debate on artificial intelligence and administrative discretion should not be framed as a choice between two extreme positions. On the one hand, it would be difficult to justify granting automated systems unrestricted authority to exercise discretionary powers in the same manner as human officials^[39]. Decisions involving complex value judgments, the balancing of competing interests, or the assessment of open-ended legal concepts raise significant concerns regarding accountability, transparency, and the protection of fundamental rights. On the other hand, a categorical exclusion of all discretionary decisions from automated processing ignores the structured and methodologically guided nature of many forms of discretionary decision-making in contemporary administrative practice. A more nuanced approach would therefore recognize that certain forms of discretion may be compatible with carefully designed forms of automated decision-making.

In particular, artificial intelligence may be capable of replicating forms of structured discretion that are already exercised through predefined criteria, methodological frameworks, or established administrative practice. In such situations, the role of the automated system would not be to invent new standards of decision-making but rather to apply existing criteria in a consistent

and predictable manner. One possible model involves the replication of structured sanctioning practices^[40]. In areas such as administrative punishment, internal guidelines often link specific types of violations and aggravating or mitigating circumstances to particular ranges of sanctions. An automated system could, in principle, apply these criteria in a manner similar to that used by human officials, thereby ensuring consistency in the treatment of comparable cases. Rather than replacing the underlying legal framework, such a system would operate within the same methodological structure that already guides discretionary decision-making in practice^[41].

A second model concerns situations in which discretionary decisions are effectively determined by the fulfilment of predefined criteria. As discussed above, licensing procedures frequently involve the verification of whether the applicant satisfies a set of statutory conditions. Although the legal norm may formally grant the administrative authority a discretionary power to grant or refuse the license, the practical outcome of the decision often depends primarily on the objective assessment of whether these conditions have been met. In such cases, automated systems may assist or even perform the decision-making process by verifying compliance with the relevant criteria^[42].

Another possible approach involves the restructuring of administrative procedures in a manner that separates standardized and discretionary elements of decision-making. In such a model, the initial stage of the procedure could involve the automated application of objective criteria and structured methodologies^[43]. Human decision-makers would then retain the possibility of intervening in cases where the circumstances of the individual case justify a departure from the standard outcome. Discretion would thus function as a corrective mechanism rather than as the primary mode of decision-making. Instead of being exercised in every individual case, discretionary judgment would be reserved for situations in which the standardized decision does not adequately reflect the particular circumstances of the case or where exceptional factors call for a different outcome. In this sense, the role of the human official would shift from routinely determining outcomes in every case to supervising the functioning of the standardized system and intervening when necessary to ensure that the resulting decisions remain fair and proportionate.

Such a layered structure may offer a way to reconcile two objectives that are often

perceived as being in tension. On the one hand, administrative systems require consistency, predictability, and the ability to process large volumes of cases efficiently. On the other hand, the legal order seeks to preserve the possibility of individualized consideration and to avoid overly rigid decision-making that fails to account for the particularities of individual cases^[44]. By combining automated standardization with the possibility of discretionary correction, administrative procedures could preserve the advantages of both approaches. Human discretion would be exercised primarily in atypical situations, where the circumstances of the case justify deviating from the standard approach. In this respect, the introduction of automated decision-making would not necessarily represent a radical transformation of administrative reasoning, but rather a technological extension of an existing institutional logic^[45].

This layered approach reflects an important insight into the function of discretion in administrative law. Discretion does not necessarily have to operate as an unrestricted margin of choice at every stage of the decision-making process. In some contexts, it may instead serve as a safeguard that allows administrative authorities to depart from standardized solutions when the particular circumstances of a case so require. By combining structured decision-making with the possibility of discretionary intervention, administrative systems may be able to reconcile the benefits of consistency and efficiency with the need to ensure fairness and individual consideration.

The proposed two-stage model naturally raises a further question: who should determine whether a particular case is sufficiently standardized to be decided automatically or whether it requires human assessment? This preliminary classification should not itself become an unconstrained algorithmic decision. As Timo Rademacher has convincingly observed^[46], contemporary AI systems operate primarily by identifying statistical patterns and may fail to recognize legally relevant circumstances that fall outside those patterns or have not been incorporated into the model's informational framework. They may therefore lack the ability to identify their own informational blind spots or to appreciate that a particular case calls for an individualized assessment rather than the application of an established pattern. Rather than entrusting this gateway decision exclusively to the automated system, a hybrid allocation mechanism appears preferable. The automated system could initially process only those cases that

satisfy predefined objective criteria indicating a sufficiently standardized factual and legal situation, while referring cases presenting predefined risk indicators, inconsistencies, missing information or low confidence levels to a human decision-maker. At the same time, the affected party should retain the procedural right (or even responsibility) to contest the automated classification and request human review where it considers that the case involves circumstances insufficiently reflected in the automated assessment.

Such an arrangement recognizes that the individual concerned will frequently be in the best position to identify factual or contextual features that distinguish the case from the ordinary pattern on which the automated decision relies. Instead of requiring the administration or the AI system to anticipate every possible exceptional circumstance, the procedural burden of identifying relevant atypical features could, to a limited extent, shift to the party to the proceedings. Where the party substantiates that the case involves legally relevant circumstances requiring individualized consideration, the matter would proceed to a human official exercising the full scope of administrative discretion. The automated decision would thus operate as the default outcome only for cases that remain uncontested and genuinely correspond to the standardized factual pattern for which the automated procedure was designed. From this perspective, the proposed model does not eliminate individualized administrative decision-making but reallocates responsibility for identifying the need for individualization. Routine cases may be resolved through automated procedures, whereas individualized assessment becomes available whenever either the administration or the affected party demonstrates that the predefined standardized framework is insufficient.

It must also be acknowledged that the integration of artificial intelligence into discretionary decision-making raises a number of important legal and institutional questions. Issues such as the transparency of algorithmic decision-making, the risk of reinforcing problematic administrative practices, the potential influence of automated outputs on human reviewers, and the broader implications for procedural fairness require careful examination^[47]. These questions extend beyond the scope of the present article and will therefore be addressed in a separate study focusing specifically on the risks and institutional challenges associated with the automation of administrative decision-making.

The analysis presented in this section nevertheless demonstrates that the relationship between artificial intelligence and administrative discretion is more complex than is often assumed. Once discretion is understood as a spectrum encompassing different modalities of decision-making, it becomes possible to identify situations in which automated systems may operate within existing frameworks of structured discretion without undermining the fundamental principles that govern the exercise of public authority.

6. Conclusions

The debate on the use of artificial intelligence in public administration is often framed around the question of whether automated systems can or cannot be used in situations involving administrative discretion (and predominantly on the assumption of the latter). I argue that such a question per se rests on a problematic premise, and that the perceived incompatibility between discretion and automation is largely a consequence of treating discretion as a binary concept. Once a decision is classified as discretionary, automation is assumed to be excluded, and vice versa. The analysis presented in this article suggests that this premise might be conceptually flawed.

I believe that administrative discretion in the age of automation should not be perceived as a uniform phenomenon that operates in an all-or-nothing manner. As the preceding sections have shown, discretionary powers manifest themselves in different degrees and modalities. While some forms of discretion involve complex value judgments, individualized assessments, and the balancing of competing interests, others operate within highly structured frameworks defined by predetermined criteria, internal guidelines, and established administrative practice. In many areas of administrative decision-making, discretion already functions through methodological frameworks designed to ensure consistency and predictability across large numbers of similar cases.

This phenomenon even gives rise to a somewhat paradoxical situation in which the broader the formal margin of choice granted to administrative authorities by legislation, the stronger the institutional incentives to limit that freedom through standardization, internal guidelines, and consistent administrative practice. Faced with a wide spectrum of formally permissible outcomes, administrative bodies

frequently seek to reduce variability by developing structured decision-making methodologies. As a result, some forms of discretion operate in practice as what may be described as pseudo-discretion: although the legal framework formally leaves several options open, the decision-making process is effectively guided by predetermined criteria that significantly narrow the range of outcomes that can realistically be chosen.

This observation bears important implications for the relationship between administrative discretion and automated decision-making. If discretion is understood as a spectrum rather than a binary category, the question of automation cannot be answered by simply asking whether discretion is present. The more relevant issue concerns the nature and structure of the discretionary decision-making involved in a particular case. Decisions requiring genuine individualized judgment and the balancing of complex human considerations may remain inherently human tasks. By contrast, forms of discretion that primarily involve the application of structured criteria or standardized methodologies may in principle be capable of replication through technical means. From this perspective, the introduction of artificial intelligence into administrative decision-making does not necessarily require abandoning the concept of discretion. Instead, it invites a reconsideration of the role that discretion plays within administrative procedures. In many areas of administrative practice, discretion currently functions as a permanent element of decision-making, meaning that officials are expected to exercise evaluative judgment in every individual case.

The analysis developed in this article suggests that an alternative institutional design may be possible. Rather than relying on discretion at every stage of the decision-making process, administrative procedures could combine standardized decision-making with the possibility of discretionary correction. In such a model, automated systems could apply predefined criteria and established administrative methodologies in routine cases, while human officials would retain the authority to depart from the standardized outcome where the particular circumstances of a case justify such intervention. Discretion would thus operate primarily as a corrective mechanism rather than as the default mode of decision-making. This approach could preserve the flexibility traditionally associated with discretionary power while simultaneously allowing administrative systems to benefit from the

consistency and efficiency of structured or automated decision-making.

The central challenge in the age of artificial intelligence is therefore not whether discretion should exist in administrative law, but how different forms of discretion should be understood and organized within increasingly digital systems of governance. Moving beyond the simplistic assumption that discretion and automation are inherently incompatible may open the way for more nuanced approaches to the integration of artificial intelligence into public administration.

1. This article was prepared within the framework of the research project No. MA/25/F3/0/017 titled *“Europe: a Laboratory of a Digital State”* supported by the 4 EU+ European University Alliance, and builds on the author’s presentation delivered at the international conference *“Swan song of the anthropocentric administration in Europe”* held in Prague on 27-28 November 2025. The manuscript was prepared with the assistance of artificial intelligence tools, which were used in full alignment with current standards for academic writing and applicable ethical guidelines. Specifically, GPT-5.2 (OpenAI) was used for language refinement, proofreading and citation, whereas Gemini (Google) and GPT Deep Research Mode were used to facilitate research and literature review. I would like to thank Jakub Handrlica for his valuable support and guidance throughout the entire project. I also find it fit to gratefully acknowledge the intellectual contribution of Juli Ponce Solé, whose insightful discussions with me helped me refine my own ideas and approach the presented research from new perspectives. Last but not least, I would like to also thank Alessia Monica and Leonardo Scuto for their help with the publication process and their patience.
2. See Parliamentary Print No. 845/0: Proposal by Members of Parliament Tomáš Dubský, Milada Voborská, Martina Ochodnická, Jiří Havránek, and Jiří Carbol for the enactment of a law amending Act No. 128/2000 Coll., on Municipalities (Municipal Establishment), as amended, and other laws in connection with supporting cooperation among municipalities, proposing the introduction of a new Section 15a of the Administrative Procedure Code.
The wording of the proposed Section 15a (1) reads as follows (translated by the author): *«Where the nature of the matter under consideration, the protection of the rights of the persons concerned, or the protection of the public interest does not require an act in the proceedings to be performed by an authorized official, the act may be performed automatically without the participation of an authorized official. An act may not be performed in this manner, in particular, where it requires the exercise of administrative discretion or concerns a decision on a remedy»*.
3. See Section 35a of the German Verwaltungsverfahrensgesetz (VwVfG) (Administrative Procedures Act), which reads as follows (translated by the author): *«An administrative act*

may be issued entirely by automated means, provided that this is allowed by law and that neither administrative discretion nor any margin of assessment is involved».

4. See e.g. M. Martini, D. Nink, *Subsumtionsautomaten ante portas? – Zu den Grenzen der Automatisierung in verwaltungsrechtlichen (Rechtsbehelfs-) Verfahren*, in *Deutsches Verwaltungsblatt*, 133, 17, 2018, pp. 1128-1137.
5. A typical example can be found in the field of administrative punishment, where the law may authorize the imposition of a fine “up to” a certain amount, but also in many other cases. See further.
6. Many continental European systems share a similar doctrinal concept developed mainly in German administrative law, which is represented by different terms of the same normative meaning in different European languages, e.g. *správní uvážení* in Czech, *Ermessen* or *Verwaltungsermessen* in German, *pouvoir discrétionnaire* in French or *discrezionalità amministrativa* in Italian.
7. See e.g. H. Maurer, C. Waldhoff, *Allgemeines Verwaltungsrecht (General Administrative Law)*, XXI Ed., C.H. Beck, München, 2024, pp. 147 et seq. Similarly e.g. in D. Hendrych et al., *Správní právo. Obecná část (Administrative Law. General Part)*, IX Ed., C.H. Beck, Prague, 2016, p. 55, G. Cartier, *Administrative Discretion: Between Exercising Power and Conducting Dialogue*, in C.M. Flood, L. Sossin, *Administrative Law in Context*, II Ed., Emond Montgomery Publications, Toronto, 2013, pp. 381 et seq., P. Craig, *EU Administrative Law*, II Ed., Oxford University Press, Oxford, 2012, pp. 401 et seq., J.H. Grey, *Discretion in Administrative Law*, in *Osgoode Hall Law Journal*, 17, 1, 1979, pp. 107-132, etc.
8. K.F. Warren, *Administrative Discretion*, in E. Berman (ed.), *Encyclopedia of Public Administration and Public Policy*, II Ed., Routledge, New York, 2007, pp. 58-62.
9. Ibidem.
10. While examples of this type of discretion can be found in any administrative sentencing, the discretionary aspects are put together nicely e.g. in M.N. Lintvedt, *Putting a price on data protection infringement*, in *International Data Privacy Law*, 12, 1, 2022, pp. 1-15.
11. See also F.A. Hessick, C.B. Hessick, *Facts, Policy, and Discretion*, in *University of California, Davis Law Review*, 59, 2026, pp. 1581-1633.
12. See e.g. H. Maurer, C. Waldhoff, *Allgemeines Verwaltungsrecht*, 21. ed., C. H. Beck, 2024, pp. 150-172, or also C. Hufen, *Ermessen und unbestimmter Rechtsbegriff*, in *Zeitschrift für das Juristische Studium*, 2010, pp. 603–607.
13. See J.P. Solé, *Artificial intelligence and interaction between machines and humans: discretionary powers, human reserve in the public sector and the EU Regulation of 2024*, in *CERIDAP*, 1, 2026.
14. See K.C. Davis, *Discretionary Justice: A Preliminary Inquiry*, LSU Press, Baton Rouge, 1969.
15. This idea is pointed out in the case law of administrative judiciary repeatedly, see e.g. Supreme Administrative Court of the Czech Republic, decision of 23 March 2005, No. 6 A 25/2002-42.

16. See J.H. Grey, *Discretion in Administrative Law*, in *Osgoode Hall Law Journal*, 17, 1, 1979, pp. 107-132.
17. Ibidem.
18. See Supreme Administrative Court of the Czech Republic, judgment of 13 March 2008, No. 5 As 51/2007-110, or also Supreme Administrative Court of the Czech Republic, judgment of 22 May 2008, No. 5 As 9/2008-106.
19. See M. Lipsky, *Street-Level Bureaucracy: Dilemmas of the Individual in Public Services*, Russell Sage Foundation, New York, 1980, pp. 81 et seq.
20. See e.g. K.C. Davis, *Discretionary Justice: A Preliminary Inquiry*, LSU Press, Baton Rouge, 1969, p. 126. See also A. McHarg, *Administrative Discretion, Administrative Rule-Making and Judicial Review*, SSRN, 2017.
21. See Committee of Ministers of the Council of Europe, Recommendation No. R (80) 2 of the Committee of Ministers to Member States concerning the Exercise of Discretionary Powers by Administrative Authorities, 11 March 1980.
22. See G.E. Metzger, K.M. Stack, *Internal Administrative Law*, in *Michigan Law Review*, 115, 2017, pp. 1239-1307.
23. See N. Marsch, V.M. Tünsmeier, *The Principle of Proportionality in German Administrative Law*, in S. Ranchordás, B. de Waard (eds.), *The Judge and the Proportionate Use of Discretion: A Comparative Administrative Law Study*, Routledge, 2015, pp. 13-42.
24. Ibidem. See also W.F. West, *Structuring Administrative Discretion: The Pursuit of Rationality and Responsiveness*, in *American Journal of Political Science*, 28, 2, 1984, pp. 340-360. This structuring effect should not be understood as permitting administrative guidelines to replace or exhaust the discretion conferred by legislation. In German administrative law, such guidelines are generally binding only within the administrative hierarchy and do not alter the statutory obligation to exercise discretion in accordance with the purpose of the empowering provision and the circumstances of the individual case. Although consistent application of a guideline may produce an indirect external effect through the principle of equality and the administration's self-binding practice (*Selbstbindung der Verwaltung*), the authority must remain capable of departing from the general approach where relevant particularities render the case atypical or otherwise require differentiated treatment. Administrative guidelines may therefore channel and standardize discretion, but they cannot lawfully convert it into an inflexible rule that excludes individualized assessment.
25. See S. Easton, C. Piper, *Sentencing and Punishment*, V Ed., Oxford University Press, Oxford, 2022, pp. 41-72. Also, numerous instances of various fining guidelines can be found, such as European Data Protection Board, *Guidelines 04/2022 on the Calculation of Administrative Fines under the GDPR*, Version 2.1, 24 May 2023, or European Commission, *Guidelines on the method of setting fines imposed pursuant to Article 23(2)(a) of Regulation No. 1/2003*, OJ C 210, 1 September 2006.
26. See S. Easton, C. Piper, *Sentencing and Punishment*, V Ed., Oxford University Press,

- Oxford, 2022, pp. 41-72.
27. See G.E. Metzger, K.M. Stack, *Internal Administrative Law*, in *Michigan Law Review*, 115, 2017, pp. 1239-1307.
 28. Such is, after all, also the logic behind Section 35a of the German VwVfG, or proposed Section 15a of the Czech Code of Administrative Procedure.
 29. Such cases might include competition regulators deciding whether to approve corporate mergers by applying broad economic concepts and complex assessment, municipal planning authorities exercising broad discretion when deciding whether to allow construction projects, environmental permits for industrial projects, or some types of politically, socially or otherwise sensitive decisions such as expropriation etc.
 30. See G.E. Metzger, K.M. Stack, *Internal Administrative Law*, in *Michigan Law Review*, 115, 2017, pp. 1239-1307.
 31. This observation, of course, is not necessarily universally applicable. For example, German administrative law traditionally distinguishes between licensing regimes that constitute bound administrative decisions (“gebundene Genehmigungen”), such as many building permits or permits under the “Bundes-Immissionsschutzgesetz”, and genuinely discretionary licensing regimes (“Ermessensgenehmigungen”). In the former category, once the statutory requirements are fulfilled, the authority is generally obliged to issue the license, meaning that § 35a VwVfG does not preclude the use of fully automated administrative acts. By contrast, where licensing powers are genuinely discretionary, the discretion often remains substantively relevant in practice, for example with regard to the imposition of ancillary provisions (“Nebenbestimmungen”) attached to the license. The present example is therefore intended to illustrate a phenomenon observable in a number of legal systems rather than to describe a universal feature of licensing procedures.
 32. See e.g. B. Schwartz, *The Paradox of Choice*, in S. Joseph (ed.), *Positive Psychology in Practice*, Wiley-Blackwell, Chichester, 2015, pp. 121-138.
 33. G.E. Metzger, K.M. Stack, *Internal Administrative Law*, in *Michigan Law Review*, 115, 2017, pp. 1239-1307.
 34. See J.C. Covilla, *Artificial Intelligence and Administrative Discretion: Exploring Adaptations and Boundaries*, in *European Journal of Risk Regulation*, 16, 1, 2025, pp. 36–50.
 35. See e.g. J.P. Solé, *Artificial intelligence and interaction between machines and humans: discretionary powers, human reserve in the public sector and the EU Regulation of 2024*, in *CERIDAP*, 1, 2026, or M. Martini, D. Nink, *Subsumtionsautomaten ante portas? – Zu den Grenzen der Automatisierung in verwaltungsrechtlichen (Rechtsbeheb-) Verfahren*, in *Deutsches Verwaltungsblatt*, 133, 17, 2018, pp. 1128-1137.
 36. Both during our conversation, but also in his publications, see e.g. J.P. Solé, *Artificial intelligence and interaction between machines and humans: discretionary powers, human reserve in the public sector and the EU Regulation of 2024*, in *CERIDAP*, 1, 2026.
 37. See also C. Harlow, *Global Administrative Law: The Quest for Principles and Values*, in *European Journal of International Law*, 17, 1, 2006, pp. 187-214.

38. Such is the case of the Czech Republic, but also other European administrative systems traditionally relying on administrative discretion in almost every aspect of decision-making.
39. See also O. Butler, *Algorithmic Decision-Making, Delegation and the Modern Machinery of Government*, in *Oxford Journal of Legal Studies*, 45, 3, 2025, pp. 727-752.
40. See also A. Lengyel, *Automated Decision-Making in Administrative Penal Matters*, in *Studia Iuridica Cassoviensia*, 13, 2025, pp. 107-119.
41. See further Y. Li, *The Applicability of Automated Administrative Penalties: Legal Challenges and Regulatory Responses*, in *Journal of Humanities, Arts and Social Science*, 8, 8, 2024, pp. 1960-1964.
42. In the light of preceding references, it has to be noted that Sec. 35a VwVfG mentioned earlier is only applicable to fully automated decision-making, meaning that discussion should strictly speaking distinguish between fully automated administrative acts and decision-support systems, where the final administrative decision remains attributable to a human official. This article deliberately remains technologically neutral, whereas additional questions fall outside the scope of this article.
43. E.g. in a form of a simplified decision such as payment orders, summary judgments etc.
44. As is pointed out and argued by Timo Rademacher, who explains that machine-learning systems operate through statistical patterns and therefore face a problem of *Einzelfalladäquanz*, meaning adequacy to the specific case. See T. Rademacher, *Recht ohne Zufall? Algorithmische Rechtsverwirklichung in der freiheitlichen Demokratie*, Nomos Verlag, 2025, pp. 291-294.
45. At the same time, such a model would require careful institutional design. Mechanisms would need to ensure that human review remains meaningful and that officials retain the authority and willingness to depart from automated outcomes when justified by the circumstances of the case. Without such safeguards, there is a risk that automated decisions might become de facto final in practice, with human oversight remaining merely formal. These concerns, however, relate primarily to the institutional design and governance of automated decision-making systems and will therefore be addressed in greater detail in a separate study focusing specifically on the risks associated with automation in public administration.
46. See T. Rademacher, *Recht ohne Zufall? Algorithmische Rechtsverwirklichung in der freiheitlichen Demokratie*, Nomos Verlag, 2025, pp. 291-323.
47. In this regard, see e.g. M. Suksi, *Administrative Due Process when Using Automated Decision-Making in Public Administration: Some Notes from a Finnish Perspective*, in *Artificial Intelligence and Law*, 29, 2021, pp. 87-110, U.B.U. Roehl, *Automated Decision-Making and Good Administration: Views from Inside the Government Machinery*, in *Government Information Quarterly*, 40, 4, 2023, or J.C. Covilla, *Artificial Intelligence and Administrative Discretion: Exploring Adaptations and Boundaries*, in *European Journal of Risk Regulation*, 16, 1, 2025, pp. 36-50.

CERIDAP

Legal Framework for Artificial Intelligence in Outer Space

Kristýna Šopková

DOI: 10.13130/2723-9195/2026-2-69

Il presente contributo esamina l'intersezione tra intelligenza artificiale e spazio extra-atmosferico, mettendo in evidenza le peculiari sfide regolamentari poste da sistemi autonomi ad alto rischio che operano in un ambiente orbitale fragile. Si sostiene che, nonostante le critiche di obsolescenza, il diritto internazionale dello spazio – attraverso l'attribuzione di responsabilità agli Stati – offre un quadro strutturale per una governance ex ante. Avvalendosi dei regimi nazionali di autorizzazione, del controllo preliminare al lancio, del monitoraggio continuo e degli obblighi di condivisione dei dati, gli Stati possono mitigare i rischi sistemici, garantire la responsabilità e gestire il dispiegamento di sistemi spaziali abilitati all'IA. L'articolo considera altresì i compromessi implicati, tra cui il potenziale rallentamento dell'innovazione e la necessità di bilanciare il principio di precauzione con il progresso tecnologico, sottolineando l'importanza di una calibrazione attenta nella progettazione normativa.

This article examines the intersection of artificial intelligence and outer space, highlighting the unique regulatory challenges posed by high-risk, autonomous systems operating in a fragile orbital environment. It argues that, despite criticisms of being outdated, international space law – through the allocation of responsibility and liability to States – provides a structural framework for ex ante governance. By leveraging national authorisation regimes, pre-launch scrutiny, ongoing monitoring, and data-sharing obligations, States can mitigate systemic risks, ensure accountability, and manage the deployment of AI-enabled space systems. The paper also considers trade-offs, including the potential slowing of innovation and the need to balance precaution with technological advancement, highlighting the importance of careful calibration in regulatory design.

Summary: 1. Introduction.- 2. Foundations of International Space Law.- 3. From Obsolete to Essential: Leveraging International Space Law for AI Governance.- 3.1 Orbital Access as a Structural Bottleneck.- 3.2 Ex Ante Oversight Through Domestic Authorization.- 3.3 Orbital Fragility and the Incentive for Strict Supervision.- 4. Challenges: Assessing the Limits of the Proposed Framework.- 4.1 Expertise Gaps and Institutional Capacity.- 4.2 Growth in Orbital Actors and the Durability of Ex Ante Control.- 4.3 Innovation Costs and the Risk of Regulatory Overreach.- 4.4 Interim Assessment.- 5. Conclusion.

1. Introduction

Outer space is an inherently hazardous environment. Extreme temperatures, radiation exposure, orbital debris, and the immense speed at which objects travel mean that even minor technical failures may lead to catastrophic consequences. The margin for error is exceptionally narrow, and incidents rarely remain confined to a single actor. Damage caused by one satellite may cascade across orbital regimes, affecting multiple operators and potentially even activities on Earth^[1]. In such an environment, the structural risks of space significantly heighten the regulatory importance of prevention and oversight.

At the same time, the legal governance of outer space is being tested by profound technological and commercial change. Rapid innovation, declining launch costs, and the rise of private operators have transformed the space sector from a predominantly state-led domain into a far more complex ecosystem of public and private activity^[2]. Private mega-constellations, in-orbit servicing missions, robotic satellite platforms, and space resource initiatives are no longer speculative concepts, but operational realities. This transformation has increased both the scale of space activity and the difficulty of supervising it effectively^[3].

A further source of regulatory complexity is the growing integration of artificial intelligence into space systems. AI technologies are increasingly used for autonomous navigation, collision avoidance, mission optimization, data processing, and other operational functions. While these systems offer clear benefits in terms of efficiency, resilience, and responsiveness, they also introduce new forms of opacity, complexity, and unpredictability. As operational control is

delegated to adaptive or partially autonomous systems, traditional assumptions about human oversight, traceability, and foreseeability of harm become more difficult to sustain.

These developments acquire particular significance in light of broader normative concerns surrounding artificial intelligence. Across legal and policy debates, questions of accountability, responsibility, and transparency have emerged as central challenges of AI governance. When decision-making processes are partially autonomous and technically non-transparent, it becomes increasingly difficult to determine who should be answerable for harmful consequences, how responsibility should be allocated, and whether algorithmic behaviour can be meaningfully scrutinized or contested. In the context of outer space – where harmful consequences may be transboundary, irreversible, and system-wide – these concerns become especially acute ^{[4][5]}.

Taken together, these developments converge on a common regulatory problem. The orbital environment demands robust *ex ante* governance, yet the systems increasingly deployed within it are autonomous, complex, and difficult to supervise through traditional means. Unlike many terrestrial domains, harm in outer space is not easily contained or remedied. A single malfunctioning satellite, one defective autonomous manoeuvre, or one compromised AI-enabled object may generate debris or disruption that affects all operators, regardless of nationality. In such circumstances, regulatory fragmentation itself becomes a source of systemic risk ^[6].

This, in turn, raises a broader legal question: whether the existing framework of international space law is capable of responding to these emerging governance challenges. International space law is often criticized as outdated and insufficiently responsive to contemporary commercialization and technological transformation ^[7]. The core treaties were drafted in an era dominated by state-led space activity, long before the rise of private mega-constellations and AI-enabled autonomous systems ^[8]. That critique is not unfounded. Yet it may overlook an important structural feature of the existing legal architecture: the allocation of responsibility and liability to states for national activities in outer space.

This article argues that, despite widespread criticism of international space law as outdated, its state-centric allocation of responsibility and liability may still offer a meaningful structural foundation for the *ex ante* governance of artificial

intelligence in outer space. In particular, Article VI of the Outer Space Treaty requires states to authorize and continuously supervise the activities of non-governmental entities, thereby embedding *ex ante* control and ongoing oversight into the legal architecture of space governance. Complementing this framework, Article VII attributes international liability for damage caused by space objects to the launching state. Together, these provisions create a dual mechanism of responsibility and liability: states must not only supervise private actors, but also internalize the legal and financial consequences of potential harm.

Far from being merely an obsolete feature of an earlier era, this state-centric design may offer certain regulatory advantages in the context of AI-enabled space activities. Because access to orbit depends on identifiable launch infrastructure, licensing procedures, and national authorization mechanisms, the space sector remains more institutionally mediated than many terrestrial domains in which AI is deployed. Moreover, the catastrophic and potentially cascading consequences of orbital incidents create strong incentives for preventive regulation. The combination of physical bottlenecks, systemic fragility, and treaty-based state responsibility may therefore make international space law more adaptable to AI-related risks than is often assumed.

This article therefore examines whether the existing international legal framework governing outer space is capable of addressing the risks associated with artificial intelligence and autonomous systems as applied within the space domain/industry. It argues that, although the current regime was not designed with AI in mind, its emphasis on authorization, continuing supervision, and state responsibility provides an important basis for *ex ante* governance. Rather than treating international space law solely as an outdated framework in need of replacement, the article suggests that it also contains underappreciated structural resources for regulating emerging technologies in an exceptionally hazardous environment.

2. Foundations of International Space Law

International space law is founded on five key multilateral treaties adopted under the auspices of the United Nations through the Committee on the Peaceful Uses of Outer Space (COPUOS). Concluded between 1967 and 1984, these treaties

established the basic legal architecture that continues to shape both international space governance and national space legislation. Together, they articulate fundamental principles governing the exploration and use of outer space, allocate rights and obligations among states, and provide core rules on matters such as responsibility, liability, rescue, and registration^[9].

The five core treaties are the Treaty on Principles Governing the Activities of States in the Exploration and Use of Outer Space, including the Moon and Other Celestial Bodies (1967, “Outer Space Treaty”), the Agreement on the Rescue of Astronauts, the Return of Astronauts, and the Return of Objects Launched into Outer Space (1968, “Rescue Agreement”), the Convention on International Liability for Damage Caused by Space Objects (1971, “Liability Convention”), the Convention on Registration of Objects Launched into Outer Space (1976, “Registration Convention”), and the Agreement Governing the Activities of States on the Moon and Other Celestial Bodies (1984, “Moon Agreement”). While all five treaties together form the historical and normative basis of international space law, the first four continue to carry the greatest practical significance in contemporary space governance.

At the same time, the foundational treaties were drafted in a very different technological and institutional context. They were designed primarily for a state-centric and largely non-commercial era of space activity, long before the rise of private mega-constellations, rapid innovation cycles, and AI-enabled autonomous systems. For this reason, scholars and practitioners have often criticized the treaty framework as insufficiently responsive to contemporary regulatory challenges, particularly in relation to private actors and complex liability scenarios^[10].

For the purposes of this article, particular attention must be paid to the distinction between state responsibility and state liability within the legal framework established by the Outer Space Treaty and the Liability Convention. Although these concepts are frequently translated into some languages by a single term, they represent distinct legal categories serving different functions within international law^[11]. Article VI of the Outer Space Treaty is based on the concept of responsibility and requires States to ensure that national activities in outer space, including those conducted by non-governmental entities, comply with international law. In practice, this obligation is implemented through the

requirement to authorize and continuously supervise private actors. By contrast, Article VII adopts the concept of liability for damage caused by space objects and establishes that the launching State bears international liability for damage caused by such objects irrespective of whether the relevant activity was carried out by governmental or private actors^[12].

Questions of international responsibility in general, and of responsibility and liability in space law in particular, have been the subject of considerable scholarly debate, and any analysis of Article VI of the Outer Space Treaty must engage with that literature rather than treat the provision's meaning as settled. The prevailing view construes Article VI as imposing a due diligence obligation on States: States are required to adopt reasonable measures to authorise, monitor, prevent and, where appropriate, sanction unlawful conduct by private actors, without incurring automatic responsibility for each such act as such^[13]. This interpretation is consistent with the general law of state responsibility, which does not ordinarily attribute the conduct of private actors to the State but instead imposes obligations relating to prevention, supervision and control. A broader – and contested – interpretation, advanced by Bin Cheng, suggests that the conduct of private entities engaged in outer space activities may be regarded as attributable to the State, such that a breach of international law by a non-governmental operator may engage state responsibility in a manner approaching direct state conduct^[14].

The doctrinal uncertainty surrounding the interpretation of Articles VI and VII of the Outer Space Treaty, particularly regarding the scope of state obligations and the attribution of responsibility and liability for resulting harm, presents difficulties even in the context of conventional space activities. Divergent interpretations concerning the extent of supervisory duties under Article VI and the operation of liability mechanisms under Article VII already generate uncertainty in relation to traditional space activities involving private actors and a substantial degree of human decision-making and control. These challenges become even more pronounced in the context of artificial intelligence. The deployment of autonomous and adaptive systems in outer space may complicate the assessment of conduct, causation and oversight, particularly where the operation of such systems involves limited, indirect or evolving forms of human intervention. Existing legal uncertainty concerning the attribution of conduct

and responsibility may therefore be further amplified where AI systems generate outcomes that are difficult to predict, explain or control *ex ante*.

The purpose of this article is not to resolve the broader doctrinal debate concerning the interpretation of Articles VI and VII of the Outer Space Treaty or to advocate for a particular understanding of state responsibility and liability in space law. Rather, this article proceeds from the premise that the existence of competing interpretations itself constitutes a relevant legal risk that must be taken into account when addressing the governance of AI systems in outer space. Irrespective of which interpretation ultimately prevails, the absence of a settled legal understanding may significantly affect regulatory design, supervisory expectations and the allocation of legal responsibility for harm caused by AI-enabled space activities.

In principle, it would be best if the treaty regime could be modernized or supplemented by new binding international agreements better tailored to contemporary space activities. In practice, however, the negotiation of new multilateral treaties in this field has proven politically difficult^[15]. As a result, much recent normative development has taken place through non-binding instruments, guidelines, and other forms of soft law adopted within the UN framework, including the COPUOS Space Debris Mitigation Guidelines. Although such instruments do not carry the force of law, they play an important role in shaping expectations of responsible conduct, particularly in areas such as orbital safety and debris mitigation. At the same time, the regulatory center of gravity has increasingly shifted toward national legislation, as states develop domestic licensing and supervisory frameworks for commercial space activities. This trend reflects a pragmatic response to the limits of treaty reform and highlights the continuing importance of state-based governance in the contemporary regulation of space activities^[16].

3. From Obsolete to Essential: Leveraging International Space Law for AI Governance

Before examining how the structural features of space governance may be leveraged for AI oversight, it is necessary to clarify the relationship between the international and domestic legal levels. International space law – most

importantly the Outer Space Treaty and the Liability Convention – does not itself establish detailed rules. Rather, it creates a main framework within which states regulate national space activities through domestic law (i.e. allocation of responsibility and liability)⁽¹⁷⁾. It is therefore at the domestic level that authorization procedures are designed, technical and safety standards are specified, supervisory mechanisms are created, and conditions for the deployment of autonomous systems are defined. This distinction is especially important in the AI context, because autonomous systems challenge traditional assumptions about foreseeability, controllability, and human oversight – assumptions that are relevant both to the international allocation of responsibility and liability and to the domestic operationalization of authorization and supervision. This dual-level structure is sometimes underappreciated in critiques of international space law.

A common objection in the literature is that the space treaties, having been drafted in the second half of the twentieth century, are poorly suited to contemporary commercialization and AI-driven operations. That criticism is partly justified: the treaties were not designed with machine learning, autonomous decision-making, or private mega-constellations in mind. Yet such critiques may overstate the consequences of that mismatch if they assume that the international framework must itself contain the full substantive content of AI regulation. Its primary function is instead structural: to allocate responsibility, impose supervision obligations, and create incentives for states to build domestic regulatory systems. In this respect, the question is not simply whether international space law contains AI-specific rules, but whether its institutional design is capable of supporting AI governance through domestic implementation. It is this latter question that is examined here.

The argument advanced in this section is that the state-centric architecture of international space law may provide more regulatory value than is often acknowledged. This is not because the treaties offer a complete or technologically tailored response to AI. They clearly do not. Rather, their continuing significance lies in the way they channel regulatory authority through states and link that authority to legal and financial consequences. In particular, Article VI of the Outer Space Treaty obliges states to authorize and continuously supervise the activities of non-governmental entities, while the liability regime gives states

strong incentives to prevent harm before it occurs. For the purposes of AI governance, this structure matters because the deployment of autonomous systems in orbit remains dependent on highly concentrated physical infrastructure and on national authorization processes capable, at least in principle, of imposing *ex ante* conditions. The discussion below develops this claim in three steps: first, it identifies the structural bottleneck created by orbital access; second, it explains how this bottleneck enables *ex ante* oversight; and third, it considers how the fragility of the orbital environment reinforces incentives for strict supervision.

3.1 Orbital Access as a Structural Bottleneck

Although technological progress has significantly reduced some of the costs associated with space activities, especially in satellite manufacturing and standardized components, access to orbit remains subject to substantial technological, financial, and infrastructural barriers. The relevant comparison with terrestrial AI is therefore instructive. Most terrestrial AI systems can be developed, modified, and deployed with little or no governmental awareness, and often without passing through any centralized institutional checkpoint. By contrast, an AI-enabled space object cannot enter orbit without reliance on launch services, launch infrastructure, licensing, and state authorization. In this respect, outer space remains a highly mediated environment rather than an open-access domain.

Launch capabilities continue to be concentrated among a relatively limited number of actors worldwide, and the number of geographically suitable and operational launch sites is inherently constrained. Effective orbital launches require particular geographic, safety, and logistical conditions, which not every territory can provide. This concentration has direct regulatory significance. Any operator seeking to place an AI-enabled satellite into orbit must depend on a launch vehicle and launch site functioning under some form of state authority or state-approved regulatory framework. Moreover, under international law, the concept of the “launching state” is interpreted broadly enough to extend legal accountability beyond the single state from whose territory a launch physically takes place. States that procure, authorize, or conduct launches may all become

relevant actors within the responsibility and liability framework.

This concentration of access creates a structural bottleneck, and that bottleneck may be normatively valuable. Much of the literature presents the state-centric structure of space law as a residue of an earlier era, when space activity was dominated by governments rather than private actors. Yet in the context of AI governance, that same structure may serve as a regulatory advantage. Because access to orbit is not freely scalable, states retain a meaningful opportunity to intervene before deployment occurs. In practical terms, autonomous systems cannot simply be “released” into orbit in the same manner as terrestrial software can be deployed online. They must pass through a state-regulated gateway.

3.2 Ex Ante Oversight Through Domestic Authorization

The existence of this bottleneck does not by itself guarantee effective governance. Its importance lies in the regulatory opportunities it creates at the domestic level. Because states remain the legal focal point of responsibility and liability under international law, they have both the authority and the incentive to condition launch approvals on compliance with specific authorization criteria. Those criteria may include safety testing, fail-safe requirements, cybersecurity protections, software verification, audit documentation, or disclosure obligations tailored to AI-enabled systems. The concrete design of such measures is a matter of domestic law, but the incentive to create them is reinforced by the international framework.

This is where the interaction between international and domestic law becomes most important. International law does not prescribe the full content of AI governance, but it does shape the behavior of states by exposing them to reputational, legal, and potentially financial consequences if inadequately supervised national activities cause harm. To that extent, the treaties may be understood not as self-sufficient regulatory codes, but as framework instruments that externalize pressure onto domestic authorization systems. This is especially significant in the AI context, where ex post responses may be insufficient. Once an autonomous system is deployed in orbit, errors may be difficult to predict, difficult to interrupt, and impossible to reverse in practical terms. For high-risk orbital activities, the preventive dimension of governance therefore becomes far

more important than in many terrestrial settings.

A possible objection is that this argument assumes too much regulatory capacity on the part of states. Not all states possess equivalent levels of technical expertise, administrative resources, or institutional experience in assessing AI-enabled systems. That concern is well founded and should not be minimized. However, it does not negate the structural point being made here. The claim is not that all states currently exercise robust and effective supervision, but that the architecture of space law creates identifiable control points through which such supervision can in principle be organized. In other words, the framework may be institutionally imperfect and unevenly implemented, yet still structurally better positioned for *ex ante* AI oversight than many terrestrial legal settings in which no comparable bottleneck exists.

This point also bears on the scholarly debate over whether the current regime is functionally obsolete. Critics are right to observe that the treaties do not directly regulate autonomous decision-making, machine learning models, or AI validation procedures. But it does not necessarily follow that they are irrelevant. If the central problem of AI governance in space is how to ensure that high-risk systems are scrutinized before deployment, then a legal framework that channels deployment through state authorization and continuing supervision may retain significant regulatory value even without AI-specific treaty language. The present argument therefore departs from stronger obsolescence narratives by suggesting that the existing system may still provide an adaptable institutional foundation for domestic AI governance.

3.3 Orbital Fragility and the Incentive for Strict Supervision

The plausibility of strong *ex ante* oversight is further reinforced by the physical characteristics of the orbital environment itself. Outer space is not merely dangerous in the ordinary sense; it is structurally fragile. Harm is cumulative, difficult to contain, and potentially irreversible. The phenomenon commonly described as the Kessler syndrome illustrates this logic particularly clearly: a single collision or technical malfunction may generate debris that triggers cascading collisions, degrading or even rendering unusable particular orbital regimes. In

such a context, the risks associated with AI-enabled systems are not limited to isolated operator error. Autonomous decision-making, software faults, cybersecurity failures, or poorly supervised updates may all contribute to events with systemic and transboundary consequences.

This feature of orbital activity strengthens the case for demanding standards of authorization and supervision. In legal terms, it supports the view that even if state responsibility under Article VI is interpreted through a due diligence lens, the content of that due diligence may have to be especially exacting in the context of AI-enabled space activities. Put differently, the significance of the responsibility/liability debate is not merely conceptual. It bears directly on the level of oversight that should be expected where the consequences of failure may affect the entire orbital commons. In this respect, the orbital environment itself provides an important reason to read the obligations of authorization and supervision as substantively robust rather than purely formal.

Comparable patterns can be observed in other highly regulated sectors characterized by low-probability but high-impact failures. Nuclear energy and systemic financial regulation offer familiar examples. In both contexts, isolated technical or institutional failures have produced cascading, transboundary, and long-term harm, leading states to adopt dense frameworks of licensing, preventive oversight, and continuous monitoring. These sectors are often criticized as heavily regulated, yet such regulatory intensity is widely understood as a proportionate response to systemic risk. Although orbital incidents are less frequent than nuclear accidents or financial collapses, their potentially irreversible consequences justify a similarly preventive orientation.

None of this means that the state-centric model is free from difficulty. It depends on administrative expertise, international coordination, and the willingness of states to enforce demanding authorization standards even where commercial pressures point in the opposite direction. Nor does it eliminate the challenge that AI systems may evolve post-launch or be updated remotely in ways not fully captured by initial review. Nevertheless, these limitations do not undermine the central point. Far from being merely obsolete, the state-centered architecture of international space law may provide a concentrated and practically meaningful framework for the *ex ante* governance of AI in orbit. Its value lies not in offering a complete substantive code for artificial intelligence, but in creating legal

incentives, institutional checkpoints, and regulatory leverage at the points where access to orbit can still be controlled.

While this model is not without institutional challenges, it demonstrates that the state-centric structure of space law does not necessarily represent regulatory obsolescence. On the contrary, it provides a concentrated regulatory checkpoint that, if effectively utilized, may function as a pragmatic mechanism for governing autonomous systems in outer space.

An additional structural factor reinforcing the plausibility of strong state oversight is the extreme fragility of the orbital environment. Unlike most terrestrial domains, outer space is characterized by cumulative and potentially irreversible externalities^[18]. The phenomenon commonly referred to as the Kessler syndrome illustrates how even a single collision or technical malfunction may trigger a cascading chain reaction of debris generation, ultimately rendering certain orbital regimes unusable^[19]. Given the autonomous decision-making of AI systems, errors may trigger or accelerate debris generation predicted by the Kessler model. In such a scenario, the consequences are not confined to a single operator or state; rather, they affect the entire international community. Access to orbit is therefore structurally interdependent: either it remains viable for all actors, or it may become effectively inaccessible to all. This “all-or-nothing” characteristic significantly increases the systemic risk associated with autonomous systems operating at scale.

Comparable risk structures can be observed in other highly regulated sectors where small failures may produce disproportionate systemic consequences. Nuclear energy provides a paradigmatic example: the accidents at Chernobyl Nuclear Power Plant and Fukushima Daiichi Nuclear Power Plant demonstrated that localized technical failures can have transboundary and long-term effects^[20]. Similarly, in the financial sector, the collapse of Lehman Brothers during the 2008 crisis revealed how interconnected systems amplify localized risk into global instability^[21]. In both sectors, states responded by establishing comprehensive frameworks of *ex ante* supervision, licensing, and continuous monitoring. These industries are frequently described as “overregulated,” yet such regulatory intensity is widely recognized as necessary and proportionate to the magnitude of potential harm. While incidents in orbital space occur less frequently than nuclear or financial crises, their systemic and potentially irreversible

consequences justify a similarly intensive regulatory approach.

4. Challenges: Assessing the Limits of the Proposed Framework

The preceding analysis has argued that the state-centric architecture of international space law, reinforced by the physical bottlenecks of orbital access, may provide a meaningful foundation for the *ex ante* governance of AI-enabled space systems. That argument, however, should not be overstated. A governance model may appear coherent at the structural level yet still encounter serious difficulties in practice. It is therefore necessary to assess not only the strengths of the proposed framework, but also its institutional limits, potential points of erosion, and broader policy trade-offs^[22].

This section addresses three principal challenges. The first concerns whether states possess the technical expertise and administrative capacity necessary to evaluate increasingly complex AI systems within authorization regimes originally designed for more conventional space technologies. The second asks whether the proposed model may weaken over time if the number of orbital actors continues to expand and commercialization further accelerates. The third concerns the danger that a precautionary, authorization-based approach may impede innovation by imposing excessive regulatory burdens on emerging technologies. Each of these objections deserves careful consideration. None is fatal to the proposed framework, but each reveals important tensions that any realistic model of AI governance in outer space must confront.

4.1 Expertise Gaps and Institutional Capacity

A central institutional weakness of the proposed *ex ante* model lies in the limited technical expertise available within many national administrations. AI systems – particularly those deployed in safety-critical orbital environments – are complex, adaptive, and often opaque. Their meaningful assessment may require interdisciplinary competence in machine learning, software assurance, cybersecurity, orbital mechanics, and systems reliability. Yet many national authorities responsible for space authorization were not originally designed to evaluate high-dimensional AI models or autonomous decision architectures^[23].

They were created primarily to assess more conventional questions of launch safety, registration, licensing, and technical compliance.

The resulting risk is that the existence of an authorization framework may create only the appearance of regulatory control. Without sufficient internal expertise, or without access to independent technical review, states may lack the capacity to scrutinize AI-enabled systems in a substantively meaningful way prior to launch. In that scenario, authorization becomes formal rather than effective. This is a serious objection, and it underscores that the viability of *ex ante* governance depends not only on legal structure, but also on institutional capability.

At the same time, the weakness should not be assessed in isolation from the incentive structure created by international space law. Under the liability regime applicable to space activities, launching states may face significant legal and financial consequences if inadequately supervised space objects cause damage. Although liability rules differ depending on the location and nature of the damage, the broader point remains: orbital accidents, debris-generating collisions, and loss of high-value infrastructure may produce exceptionally costly consequences^[24]. From this perspective, investment in *ex ante* expertise may be economically rational. Even costly mechanisms such as expert review panels, third-party certification, technical audits, or specialized AI assessment units are likely to be less expensive than the consequences of catastrophic orbital harm.

A further challenge arises from disparities among states. Not all launching states possess comparable scientific infrastructure, administrative capacity, or access to expert communities. This asymmetry may produce uneven supervisory standards and create weak links in the overall system^[25]. Here, too, the objection is real. The state-centric model does not ensure uniform quality of oversight across jurisdictions. Yet the shared and interdependent nature of the orbital environment also creates strong incentives for cooperation. Because the costs of a poorly supervised launch are not confined to the authorizing state, orbital safety functions as a collective good. This suggests that technical expertise, model standards, and supervisory practices may need to be shared across jurisdictions if *ex ante* governance is to remain credible in practice. The challenge of expertise therefore qualifies the proposed model, but does not necessarily defeat it; rather, it indicates that effective state-based governance may require institutional support mechanisms beyond the state acting alone^[26].

4.2 Growth in Orbital Actors and the Durability of Ex Ante Control

A second objection is more structural. Even if state-based authorization currently provides a workable checkpoint, this may not remain true as access to space becomes cheaper and the number of private operators grows. If orbital activities continue to expand rapidly, states may find it increasingly difficult to exercise meaningful oversight over all AI-enabled systems deployed by their national operators. Regulatory capacity is finite, while the complexity of machine-learning-based systems may place increasing pressure on already limited administrative resources. On this view, the bottleneck that currently supports *ex ante* control may gradually weaken as commercialization accelerates.

This concern deserves serious attention, but its empirical basis should not be overstated. Some parts of the space sector – most notably small-satellite manufacturing and standardized components – have indeed become more affordable^[27]. Yet launch capability has not undergone a comparable degree of democratization. The development, certification, and operation of orbital launch vehicles remain technologically sophisticated, capital-intensive, and heavily regulated. Access to orbit therefore continues to depend on a relatively small number of launch providers and operational launch sites. Even if the number of payload developers and satellite operators increases substantially, this does not necessarily imply a corresponding fragmentation of the launch segment^[28].

Indeed, market dynamics may point in the opposite direction. In infrastructure-heavy sectors, declining costs in one segment often produce concentration rather than unlimited entry in another. Launch services may increasingly resemble a centralized transport function, in which a limited number of providers offer capacity to a growing number of users. In this respect, commercial developments such as “rideshare” or “launch-as-a-service” models may actually reinforce, rather than undermine, the structural concentration through which regulation can operate. Operators investing substantial sums in AI-enabled payloads are also likely to prefer established launch providers with demonstrated reliability, rather than untested entrants. This preference further supports concentration in the launch market^[29].

The significance of this point is regulatory rather than merely economic. Even if

the number of actors designing satellites increases, access to orbit may remain funneled through a relatively small and identifiable set of gateways. If so, the state does not need to supervise every possible technological actor in the abstract; it needs to ensure that no payload reaches orbit without passing through authorization and launch control processes. This creates a structural asymmetry that distinguishes the space context from most terrestrial AI applications, where systems may be developed and deployed with minimal governmental visibility.

That said, the objection cannot be dismissed entirely. If launch markets were to become substantially more decentralized, or if new forms of access to orbit were to erode the present concentration of infrastructure, the effectiveness of the bottleneck model could diminish over time. The claim advanced here is therefore contingent rather than absolute. It is not that the bottleneck will necessarily remain intact indefinitely, but that under current technological and market conditions it remains sufficiently concentrated to support meaningful *ex ante* governance.

Where such concentration persists, domestic authorization frameworks can operate as practical tools of control. States may require payload operators to obtain authorization before seeking launch services, to demonstrate compliance with safety, cybersecurity, and technical reliability requirements, and, where necessary, to submit documentation, testing results, or audit materials relevant to AI-enabled functions. Launch providers, in turn, may be required to verify the existence of valid authorization before carrying payloads to orbit. In this way, the authorization process can function not merely as a formal licensing step, but as an upstream checkpoint through which deployment is conditioned on prior scrutiny.

Ex ante review, however, does not eliminate all risk. AI systems may evolve post-launch, may be remotely updated, or may behave in unforeseen ways despite pre-deployment assessment. For that reason, authorization should not be understood as a complete solution. Its importance lies rather in reducing the probability that inadequately assessed high-risk systems reach orbit in the first place, while also creating the legal and procedural basis for continuing supervision after deployment. In this respect, upstream control remains valuable even where downstream uncertainty cannot be fully removed.

4.3 Innovation Costs and the Risk of Regulatory Overreach

A third challenge concerns the costs of precaution itself. An authorization-based governance model may reduce the likelihood of harmful deployment, but it may also slow the pace of innovation. Space activities have developed rapidly in recent decades, particularly as private actors have entered the market and technical development cycles have accelerated. More demanding authorization procedures for AI-enabled systems could lengthen deployment timelines, increase compliance costs, and reduce flexibility for operators. These effects may be especially significant for smaller or newer market entrants^[30].

A related concern is that states, motivated by the prospect of liability and by the catastrophic potential of orbital accidents, may adopt overly conservative authorization criteria. If regulatory institutions become excessively risk-averse, they may block or delay the deployment of technologies whose risks are uncertain but manageable. In this sense, the very features that make *ex ante* control attractive from a safety perspective may also produce a chilling effect on innovation. This is a genuine policy tension and should not be underestimated^[31].

At the same time, the existence of such trade-offs does not mean that precautionary governance is misguided. In sectors characterized by high-impact, potentially irreversible harms, slower deployment may sometimes be a justified cost of responsible regulation. The relevant question is not whether oversight imposes burdens – it inevitably does – but whether those burdens are proportionate to the magnitude of the risks at stake. In the orbital environment, where a single failure may generate long-term debris or compromise access for multiple actors, some degree of regulatory friction may be not only acceptable, but necessary.

The challenge, then, is one of calibration rather than rejection. If authorization mechanisms are to remain both legitimate and effective, they should be designed to distinguish between unacceptable risk and manageable uncertainty. This may require proportionate review procedures, staged approvals, experimental licensing, expert advisory bodies, and mechanisms for iterative testing under controlled conditions. Such approaches would allow states to preserve the preventive logic of *ex ante* governance while reducing the danger that authorization becomes an instrument of excessive conservatism^[32].

The broader implication is that the proposed framework should not be understood as advocating rigid or static control. Its value lies in establishing legal and institutional checkpoints through which high-risk autonomous systems can be evaluated before deployment, while still leaving room for differentiated and adaptive regulatory techniques. The challenge of innovation, therefore, does not refute the *ex ante* model. Rather, it highlights the need to design it carefully.

4.4 Interim Assessment

Taken together, these three challenges show that the proposed framework is neither automatic nor cost-free. Its effectiveness depends on institutional expertise, continued concentration in access to orbit, and the careful calibration of precautionary oversight. These are meaningful constraints, and they should temper any overly optimistic account of state-centered AI governance in outer space. Yet they do not alter the central conclusion of this article. Even acknowledging its limits, the state-centric structure of international space law remains more adaptable to *ex ante* AI governance than critiques of obsolescence often suggest. Its value lies not in eliminating uncertainty, but in providing legal incentives, regulatory leverage, and institutional control points through which uncertainty can be managed before harm occurs.

5. Conclusion

In sum, the regulatory landscape of outer space reflects a complex interplay between extraordinary environmental hazards, rapid technological advancement, and evolving legal frameworks. The combination of high systemic risk, AI integration, and increasing private sector participation underscores the urgent need for proactive governance measures. *Ex ante* authorization, embedded within the state-centric structure of international space law, provides a unique and enforceable mechanism to manage these risks before they materialize, leveraging states' legal responsibility, control over launch providers, and economic incentives to ensure compliance.

While international treaties are often critiqued as outdated, their allocation of responsibility and liability to states creates strong legal and economic incentives for rigorous *ex ante* oversight. When states implement rigorous pre-launch

scrutiny, ongoing monitoring, and data-sharing obligations, they establish a coherent framework capable of mitigating systemic risks, including those arising from autonomous AI-enabled systems.

Nevertheless, this governance model is not without trade-offs. The precautionary nature of *ex ante* review may slow the pace of innovation or introduce conservative biases in deployment decisions. Balancing the need for safety against the benefits of technological advancement remains a central challenge. Effective implementation therefore requires not only robust legal and procedural structures but also the careful calibration of oversight thresholds and the integration of technical expertise.

Ultimately, the state-centric approach illustrates that the apparent obsolescence of international space law may be overstated. Instead, it provides a foundational platform upon which contemporary governance of AI-enabled space activities can be built – leveraging historical legal principles, concentrated control points, and interdependent incentives to address the novel risks of the twenty-first-century orbital domain.

1. L. Hall, *The History of Space Debris*, Space Traffic Management Conference, Embry-Riddle Aeronautical University, 6, Listopadu, 2014.
2. W. W. Baber & A. Ojala, *New Space Era: Characteristics of the New Space Industry Landscape*, in *Space Business: Emerging Theory and Practice*, pp. 3–26, DOI: 10.1007/978-981-97-3430-6_1, Springer, Singapore, 2024.
3. G. M. Danilenko, *International law-making for outer space*, in *Space Policy*, vol. 37, no. 3, pp. 179–183, DOI: 10.1016/j.spacepol.2016.12.002, Elsevier, Amsterdam, 2016.
4. A. Dey & J. Jagadanandan, *Unveiling the realm of AI governance in outer space and its importance in national space policy*, in *Acta Astronautica*, vol. 228, pp. 253–264, DOI: 10.1016/j.actaastro.2024.11.022, Elsevier, Amsterdam, 2025.
5. T. Graham, K. Thangavel & A.-S. Martin, *Navigating AI-lien Terrain: Legal Liability for Artificial Intelligence in Outer Space*, in *Acta Astronautica*, vol. 217, pp. 197–207, DOI: 10.1016/j.actaastro.2024.01.039, Elsevier, Amsterdam, 2024.
6. L. Hall, *The History of Space Debris*, in *Space Traffic Management Conference*, Embry-Riddle Aeronautical University, 6, Listopadu, 2014.
7. G. M. Danilenko, *International law-making for outer space*, in *Space Policy*, vol. 37, pp. 179–183, DOI: 10.1016/j.spacepol.2016.12.002, Elsevier, Amsterdam, 2016.
8. F. G. Von der Dunk, *The Origins of Authorisation: Article VI of the Outer Space Treaty and International Space Law*, in F. G. Von der Dunk (ed.), *National Space Legislation in Europe: Issues of Implementation*, Springer, Dordrecht, 2009.

9. F. G. Von der Dunk, *Advanced Introduction to Space Law*, in *Elgar Advanced Introductions series*, Edward Elgar Publishing, Cheltenham, 2020.
10. F. G. Von der Dunk, *The Origins of Authorisation: Article VI of the Outer Space Treaty and International Space Law*, in F. G. Von der Dunk (ed.), *National Space Legislation in Europe, Studies in Space Law*, vol. 6, Nijhoff, Leiden, pp. 1–2, 2011; D. Linden, *The Impact of National Space Legislation on Private Space Undertakings: Regulatory Competition vs. Harmonization*, in *Journal of Science Policy & Governance*, vol. 8, no. 1, pp. 1–14, 2016.
11. *Ibid.*
12. C. Stotler, *Interpretation as Creation: Article VI of the Outer Space Treaty*, in *Chicago Journal of International Law*, vol. 26, no. 1, pp. 133–146, 2025.
13. F. G. von der Dunk, *The Origins of Authorisation: Article VI of the Outer Space Treaty and International Space Law*, in F. G. von der Dunk (ed.), *National Space Legislation in Europe*, Martinus Nijhoff, Leiden, pp. 3–28, 2011; S. E. Bartz, *Reframe: A New Interpretive Framework for Non-Governmental Entity Responsibility in Outer Space*, in *Journal of Air Law and Commerce*, vol. 90, no. 2, pp. 154–173, 2025.
14. B. Cheng, *Article VI of the 1967 Space Treaty Revisited: "International Responsibility", "National Activities", and "The Appropriate State"*, in *Journal of Space Law*, vol. 26, no. 1, pp. 7–33, 1998.
15. F. G. Von der Dunk, *The Origins of Authorisation: Article VI of the Outer Space Treaty and International Space Law*, in F. G. Von der Dunk (ed.), *National Space Legislation in Europe, Studies in Space Law*, vol. 6, 2011.
16. J. Handrlica, *Kosmické právo správní – skica moderního právního odvětví*, in *Studia Iuridica Cassoviensia*, roč. 13, č. 1, pp. 4–19, DOI: 10.33542/SIC2025-1-02, 2025.
17. Hobe, Schmidt-Tedd and Schrogl, *The Cologne Commentary on Space Law*, pp. 418–419, Carl Heymanns Verlag, Cologne, 2017.
18. G. Rausser, E. Choi & A. Bayen, *Public-Private Partnerships in Fostering Outer Space Innovations*, in *Proceedings of the National Academy of Sciences of the United States of America*, vol. 120, no. 43, National Academy of Sciences, USA, e222013120, DOI: 10.1073/pnas.222013120, 2023.
19. I. Bratu, *Mission Impossible? Artificial Intelligence, Space Debris, and the Legal Implications for Space Sustainability*, SSRN Electronic Journal, VU University Amsterdam, 28 pp., DOI: 10.2139/ssrn.5112687, 30 Aug 2024.
20. P. D. Cameron, *Nuclear Safety After Chernobyl: The Role of International Law*, in *Leiden Journal of International Law*, vol. 1, no. 2, Leiden, pp. 121–135, DOI: 10.1017/S0922156500000820, 1988.
21. A. M. Hashmall, *After the Fall: A New Framework to Regulate "Too Big to Fail" Non-Bank Financial Institutions*, in *New York University Law Review*, vol. 85, no. 3, 2010.
22. R. L. Spencer Jr, *State Supervision of Space Activity*, in *Air Force Law Review*, roč. 63, str. 75, 127. ISSN 0094-8381, pp. 75–82, 2009.

23. D. O. Cajueiro & V. R. R. Celestino, *A Comprehensive Review of Artificial Intelligence Regulation: Weighing Ethical Principles and Innovation*, in *Journal of Economy and Technology*, vol. 4, pp. 77–91, DOI: 10.1016/j.ject.2025.07.001, Elsevier, Amsterdam, 2026.
24. K. Schmalenbach, *Convention on International Liability for Damage Caused by Space Objects*, in P. Gailhofer et al., *Corporate Liability for Transboundary Environmental Harm*, Springer, Cham, https://doi.org/10.1007/978-3-031-13264-3_11, pp. 525–530, 2023.
25. J. Hendricks, *South Africa's Place in Space: Towards a New Space Regulatory Framework*, in *Space Fostering African Societies*, pp. 1–24, DOI: 10.1007/978-3-030-89887-8_1, Springer, Cham, 2022.
26. G. Rausser, E. Choi & A. Bayen, *Public-private partnerships in fostering outer space innovations*, in *Proceedings of the National Academy of Sciences of the United States of America*, vol. 120, no. 43, National Academy of Sciences, USA, e2222013120, DOI: 10.1073/pnas.2222013120, 2023.
27. S. Rementeria, *Power Dynamics in the Age of Space Commercialisation*, in *Space Policy*, vol. 60, Art. 101472, DOI: 10.1016/j.spacepol.2021.101472, Elsevier, Amsterdam, 2022.
28. *Ibid.*
29. *Ibid.*
30. D. O. Cajueiro & V. R. R. Celestino, *A Comprehensive Review of Artificial Intelligence Regulation: Weighing Ethical Principles and Innovation*, in *Journal of Economy and Technology*, vol. 4, pp. 77–91, DOI: 10.1016/j.ject.2025.07.001, Elsevier, Amsterdam, 2026.
31. K. J. D. Chan, G. Papyshv & M. Yarime, *Balancing the Tradeoff Between Regulation and Innovation for Artificial Intelligence: An Analysis of Top-Down Command and Control and Bottom-Up Self-Regulatory Approaches*, in *Technology in Society*, vol. 79, Art. 102747, DOI: 10.1016/j.techsoc.2024.102747, Elsevier, Amsterdam, 2024.
32. *Ibid.*

From Human to Robot Interpreters: Legal Reflections on the Draft Amendment to the Foreign Nationals Residence Act

Kamila Balounová

DOI: 10.13130/2723-9195/2026-2-30

L'interpretazione e la traduzione robotica potrebbero in futuro offrire vantaggi in termini di efficienza, sotto il profilo dei costi e della rapidità nei procedimenti dinanzi alle autorità pubbliche. Tuttavia, le tecnologie attuali non dispongono della qualità e dell'affidabilità necessarie per garantire il diritto all'interpretariato e ad un equo processo. L'articolo si esprime a favore di un modello ibrido che combini l'intelligenza artificiale con la supervisione umana e auspica una futura regolamentazione integrata direttamente nel diritto processuale.

Robotic interpreting and translation may in the future offer efficiency gains in terms of cost and speed in proceedings before public authorities. However, current technologies lack the quality and reliability required to safeguard the right to interpretation and a fair trial. The article argues for a hybrid model combining artificial intelligence with human oversight and calls for future regulation to be embedded directly in procedural law.

Summary: 1. Introduction.- 2. The Role of Court Interpreting and Translation.- 3. Machine Interpreting and Machine Translation.- 4. Machine interpretation in the Foreigners Residence Act.- 5. Conclusion.

1. Introduction

The accelerated development of artificial intelligence in recent years has fundamentally transformed the field of interpreting and translation. While until recently professional discussion focused primarily on computer-assisted translation and interpreting as support tools for human work, today the question of their complete replacement by automated systems is increasingly being raised. Moreover, this debate concerns not only conference or community interpreting and translation but is increasingly affecting the field of legal interpreting and translation as well.

Legal interpreters and translators represent a specific group of professionals whose activity is directly connected with the protection of fundamental rights of participants in judicial and administrative proceedings, particularly the right to a fair trial, the right to defence, and the right to be heard. The quality of interpreting and translation in this context is not merely a matter of professional standards or procedural efficiency but has a direct constitutional law dimension. Poor quality or inaccurate interpreting or translation may lead to interference with the procedural rights of participants in proceedings and, in extreme cases, to an unjust decision.

The article therefore focuses on the issue of using automatic interpreting and translation in the judicial environment, at the intersection of administrative and constitutional law, linguistics, and translation studies. The first part devotes attention to the fundamental principles of legal interpreting and translation, and the current development of technologies based on artificial intelligence. The second part then concentrates on the specific legislative proposal contained in the amendment to the Act on the Residence of Foreign Nationals, which envisages the use of automated systems instead of legal interpreters and translators, and evaluates it from the perspective of fundamental rights protection and procedural guarantees.

2. The Role of Court Interpreting and Translation

Whereas interpreting is the oral transfer of a message from one language to another occurring in real time (i.e., interpreting the spoken word)^[1], translation is, in simplified terms, the written transfer of a message from one language to

another (i.e., translation of the written word)^[2]. Both interpreting and translation serve to facilitate intercultural communication and overcome language barriers between people who do not speak the same language. Court interpreting and court translation are specific disciplines. They are performed in a “legal” setting (in court, at an administrative authority, at the police station, at a notary’s office). Many countries have decided to introduce legislation regulating the activities of court interpreters and make the performance of this activity conditional upon obtaining special authorization in order to ensure a certain quality of their performance^[3]. In the Czech Republic, too, court interpreting and translation are conditional upon obtaining authorization, pursuant to Act No. 354/2019 Coll., on Court Interpreters and Translators, which forms the basis of the legal framework for the activities of court interpreters and translators.

The right to interpretation is enshrined in many international instruments and is often linked to the right to a fair trial. The International Covenant on Civil and Political Rights enshrines it in Article 14(3)(f)^[4]. The European Convention on Human Rights provides for it in Article 5(2) and Article 6(3)(e)^[5]. Directive 2010/64/EU of the European Parliament and of the Council of 20 October 2010 on the right to interpretation and translation in criminal proceedings specifies the right of accused persons and suspects to interpretation and translation in criminal proceedings in EU Member States. In the Czech legal order, the right to an interpreter is enshrined not only at the constitutional level (in Article 37(4) of the Charter of Fundamental Rights and Freedoms), but also in procedural regulations.

The right to interpretation means that a person who does not understand or speak the language of the proceedings has the right to free assistance of a court interpreter and translator in judicial proceedings or proceedings before administrative authorities. It also applies to deaf and deaf-mute persons, who are provided with a sign language interpreter. The right to interpretation, its content and scope, is further defined and specified by the international instruments themselves or by Czech procedural regulations and case law.

The right to interpretation is considered part of the right to a fair trial, as a so-called ancillary right to the principle of equality of the parties in proceedings^[6]. The right to a fair trial is one of the fundamental pillars of the rule of law and democratic society^[7]. It ensures that every person may assert their rights through

established procedures before an independent and impartial court and, in specified cases, before another body^[8]. According to the Constitutional Court of the Czech Republic, it is not a one-dimensional right, but rather a complex and structured right comprising multiple independent subjective fundamental rights^[9].

The right to an interpreter is an essential component of the principle of equality of the parties. William Hewitt emphasizes that «*interpreters help ensure that such persons may enjoy equal access to justice, and that court proceedings and court support services function efficiently and effectively*»^[10]. Ensuring equality of the parties in all types of proceedings is a necessary prerequisite for achieving a fair trial. In order to achieve true equality of the parties in proceedings, it is essential to ensure the right to use one's mother tongue before public authorities, from which follows the right to an interpreter, which guarantees that no party will be disadvantaged due to a language barrier^[11]. The right to an interpreter (together with the right to defence) represents a significant guarantee of properly conducted and fair proceedings before public authorities^[12].

Failure to ensure the right to an interpreter, despite a party to the proceedings having requested it or where the need for a court interpreter is evident, results in a failure to ensure equality of the parties^[13] and a substantial procedural defect, as it significantly curtails their rights^[14]. Violation of the right to an interpreter impermissibly infringes upon human dignity under Article 1 of the Charter of Fundamental Rights and Freedoms and simultaneously violates the principles of a fair trial, thereby constituting a violation of the right to a fair trial^[15].

Without the ability to understand the ongoing proceedings and without the ability to express oneself adequately, a fair trial would remain a mere formality and would be practically unattainable for those who do not understand the language of the proceedings. This right deserves all the more attention in connection with the discussion on the possibilities of using technology and artificial intelligence in the field of interpreting and translation in judicial and administrative proceedings.

3. Machine Interpreting and Machine Translation

Automatic speech translation, also referred to as speech-to-speech translation or

spoken language translation, represents an innovative field of artificial intelligence aimed at creating machines capable of translating spoken discourse from one language to another in real time and for immediate practical use. In recent years, this technology has experienced significant performance improvements due to the development of modern machine learning algorithms, particularly neural networks, and the availability of extensive language data (for certain language combinations). There are two fundamental approaches to machine interpreting: the cascade approach and the end-to-end approach^[16].

In the cascade approach, the process of oral translation is divided into several successive sub-steps: automatic speech recognition (ASR), which converts spoken discourse into text form; machine translation (MT), which translates written text from the source language into the target language; and text-to-speech synthesis (TTS), which generates the spoken form of the translated text^[17]. Cascade models outperform end-to-end models in accuracy, particularly for high-resource languages and complex tasks, due to their modularity and better error management^[18].

The end-to-end approach is based on the use of a unified component that performs direct conversion of the input audio signal into an output audio signal in the target language without generating intermediate textual representation, i.e., transcription^[19]. This approach enables the translation of spoken speech from one language directly into speech in another language without relying on intermediary text processing^[20]. End-to-end models excel in latency and deployment simplicity.

Machine interpreting systems are capable of facilitating spoken communication to a certain extent. They also achieve real-time processing of input and provide voice output^[21], but they still rely on machine translators developed primarily for text translation. These systems cannot fully analyse and interpret the content of messages, which is essential for proper mediation of their meaning. They are unable to process meanings conceptually in a manner that corresponds to human understanding, which limits their ability to accurately capture and convey the sense of communication, particularly in culturally specific and more complex situations^[22]. Machine interpreting systems are incapable of understanding typical aspects of verbal communication, such as inference from situational context, interpretation of prosodic features, correction of imperfect utterances,

etc.^[23]. When processing and converting the source message, an interpreter utilizes not only linguistic skills but also their presence in the communicative situation and ability to adapt to the context of communication^[24]. This temporal and spatial immediacy of interaction, which encompasses not only the spoken word but also non-verbal expressions, makes human interpreting unique and irreplaceable by machines^[25].

MT is the automatic translation from one language to another using computer technology without human intervention^[26]. Currently, neural machine translation (NMT) is most frequently utilized, based on the use of deep neural networks and employing deep learning^[27]. The Transformer is a deep neural network composed of many encoder and decoder layers that utilize a self-attention mechanism for efficient analysis of relationships between all words in a sequence simultaneously^[28]. The most efficient type of NMT consists of models based on transformer architectures, primarily the Transformer model, which has surpassed earlier methods such as recurrent neural networks (RNN) and convolutional neural networks (CNN)^[29].

Another approach to machine translation is hybrid machine translation, combining different approaches to achieve higher accuracy and flexibility^[30]. Most commonly, they combine neural machine translation methods with older statistical or rule-based models^[31]. The functioning of a hybrid model consists in the individual methods complementing each other. While neural networks enable better understanding of context and more fluent translations, statistical methods or rule-based systems can assist with consistency of terminology and handling of specific linguistic structures.

Among the main advantages of machine translation is the elimination of typographical errors, grammatical and punctuation mistakes, which ensures formally correct output. Today, thanks to progress, NMT models are better able to adapt to linguistic variations and produce more fluent and natural translations, preserve context, more accurately interpret idiomatic expressions, and more effectively convey subtle linguistic nuances^[32].

However, despite significant progress, machine translation still faces challenges, particularly in the areas of contextual accuracy, stylistic adaptability, and specialized terminology^[33]. Problems remain with literalness, insufficient sensitivity to dialects and cultural specifics, and difficult processing of idioms and

ambiguous expressions. Translation systems often inaccurately interpret specialized jargon and cannot fully preserve the stylistic nuances of different genres^[34]. Further progress therefore requires more sophisticated language models, expansion of training data, and refinement of adaptation to context and specific communicative purposes^[35].

In the case of court interpreting and translation, the demands are even higher—here, not only linguistic accuracy is required, but also deep understanding of legal concepts and their specific meaning in different legal cultures. For these reasons, current machine translation and interpreting technologies cannot be considered capable of fully replacing court interpreters or translators. Their use in proceedings before public authorities is, at this time, inappropriate from both a legal and practical perspective.

4. Machine interpretation in the Foreigners Residence Act

In 2024, the Ministry of the Interior submitted a government draft of a new Act on the Entry and Residence of Foreign Nationals in the Territory of the Czech Republic, which is to replace the currently applicable Act on the Residence of Foreign Nationals. This draft, which has not yet passed through the legislative process, contains Section 492, which was intended to enable interpretation by means of a technical device: *«for the interpretation of an act in proceedings, an administrative authority may use a technical device instead of an interpreter registered in the list of court interpreters and court translators»*^[36]. For the time being, this would be the only legal regulation (both substantive and procedural) in the Czech Republic that would permit machine interpreting or translation.

According to the explanatory memorandum, the main reasons for introducing the possibility of interpretation by means of a technical device are the effort to make the administrative process more efficient and faster, as automatic interpretation will enable faster communication and eliminate time delays associated with organizing interpretation with a court interpreter; cost savings, as the use of technology will lead to savings in expenditure on fees for court interpreters; elimination of errors in interpretation, as the Ministry of the Interior assumes that technology should be capable of eliminating errors arising from “human” interpretation, such as omissions or unintentional or intentional

omission of information; and the development and support of digitalization of state administration.

The draft presupposes a technical device enabling automatic interpretation; however, it does not take into account the possibility of a technical device for automatic translation, despite the fact that greater progress has been achieved precisely in the field of machine translation. Section 492 provides the administrative authority with the option to use a technical device instead of a court interpreter, but without clear specification of the conditions under which this could occur. The text of the Act does not make clear whether prior informed consent of the party to the proceedings would be required^[37]. The text of the Act does not even envisage the issuance of implementing regulations that would specify the conditions for the use of this technical device.

The provision in question is conceived in such a way that the technical device cannot be regarded merely as a supportive tool intended for basic assistance in communication (e.g., in the form of a chatbot), but as an autonomous means capable of generating interpreting outputs at all stages of the proceedings. The legal regulation does not establish any limitations regarding the type of acts for which the device may be used. No distinction is thus made between situations where technology serves only for informal communication and cases where its outputs may have a direct impact on the content of decisions in the matter and the procedural rights of parties to the proceedings.

The deployment of a technical device for machine interpreting (and translation) into live operation without prior thorough testing of its reliability, accuracy, and security can be considered highly problematic. Such an approach could jeopardize the fundamental procedural rights of parties to the proceedings, particularly the right to a fair trial and the right to understand the content of the proceedings. Before any practical deployment, it is therefore essential to conduct pilot verification in a controlled environment that will enable identification of potential system failures and verification that the device meets the required standards comparable to the performance of a court interpreter or translator.

Paragraph 2 sets out the basic and only requirements for the technical device itself: *«the technical device according to paragraph 1 must enable automatic interpretation of spoken foreign language into Czech language and back into the foreign language in real time. The accuracy and speed of interpretation must be*

comparable to an interpreting act performed by an interpreter registered in the list of court interpreters and court translators»^[38].

It is precisely in the requirement set out in paragraph 2 for accuracy and speed comparable to an interpreting act performed by a court interpreter that one of the fundamental deficiencies of the legislative proposal can be seen. It is in no way clear who and how should assess this comparability and what the tolerated deviation is. It can be assumed that in real time this is not even possible when the administrative authority and the party to the proceedings each command only one of the languages used. If this is a condition that must be met before the deployment of the technical device into “live operation”, it is not clear through what criteria the performance of court interpreters will be assessed and who will evaluate their performance in order to subsequently establish parameters for evaluating the performance of the technical device.

It is also necessary to emphasize that machine interpreting technology is dependent on the availability of large amounts of quality language data from which it learns. However, for less widespread languages (so-called minor languages), there is a shortage of such data, which makes adequate training of systems impossible. The outputs of machine interpreting could therefore be not only inaccurate but also completely misleading, thereby infringing upon the right of parties to the proceedings to a fair trial.

In order to ensure effective control of machine interpreting outputs, it would be essential to systematically make recordings of interpreted acts and simultaneously establish clear rules for their retention and subsequent review. In cases of doubt, it would be necessary for the recording to be evaluated by a court interpreter, which would, however, bring not only additional costs but also possible prolongation of the proceedings.

Furthermore, it remains unclear who and in what manner would be authorized to subsequently review the outputs of machine interpreting, or whether the burden of proof would rest on the party to the proceedings, who would have to prove the discrepancy between the interpreted content and the actual meaning of the statement, as well as the impact of such an error on the course or outcome of the proceedings and on their procedural rights. It is also not clear whether the party would have access to a recording of the machine interpreting, its transcript, or whether they would not be granted access to any materials.

Another question that needs to be resolved before deploying such a device is protection against cyberattacks and data security. Interpreting, like translation, often involves working with highly sensitive information, which may include personal data, information relating to ongoing proceedings, etc.^[39] Data processed by the technical device may be stored or transmitted through cloud services, which increases the risk of their theft or unauthorized access.^[40] A leak of such information could have serious consequences, including jeopardizing ongoing proceedings, harming the parties, or undermining trust in public administration.

A fundamentally crucial legal question is liability for the functioning of the technical device, particularly in connection with possible errors in automatic interpreting or translation, leakage of sensitive data, or manipulation of the algorithm. Unlike a court interpreter or translator, who is legally liable under Czech law for their errors (cf. Section 21, Section 37(1)(c) or (2)(c) of Act No. 354/2019 Coll., on Court Interpreters and Translators) and for breach of the duty of confidentiality (cf. Section 37(1)(h) or (2)(h) of Act No. 354/2019 Coll., on Court Interpreters and Translators), the technical device bears no legal liability.

The Czech legal order does not address the liability of a technical device, thus creating a legal vacuum, as it is not clear who should bear liability for potential errors—whether the administrative authority that used the device, the software manufacturer, or the technology operator^[41]. The state would likely bear liability for any damage caused by this device. This approach is based in the Czech legal order on strict liability without the need for fault on the part of the obligated person (the so-called no-fault liability model). The state, which implemented the possibility of using machine interpreting or translation into the legal order, would also be liable for any damage, cf. Section 1(1) of Act No. 82/1998 Coll., on Liability for Damage Caused in the Exercise of Public Authority by a Decision or by Incorrect Official Procedure. The purpose of Act No. 82/1998 Coll. is to enable compensation for all cases in which damage was caused by an unlawful decision or incorrect official procedure^[42].

Act No. 82/1998 Coll. allows, in certain cases, recourse claims to be made against official persons, but does not allow them to be made against other entities, such as software manufacturers. The state is thus presented with the option of using

the private law route, most likely liability for damage caused by a defective product under Section 2939 et seq. of Act No. 89/2012 Coll., the Civil Code^[43]. Application of this provision is conceivable in situations where during the production process, typically when programming artificial intelligence algorithms, a bug occurs that later causes deviation from the standard and required behavior of artificial intelligence^[44]. For fully autonomous devices, however, the detectability and provability of model failure appears problematic, as the manner in which it arrives at individual outputs remains non-transparent and difficult or even inaccessible for humans to understand^[45]. This insufficient explainability of algorithms leads to the so-called black box phenomenon^[46], where even a technical expert is unable to understand its internal functioning (so-called internal opacity)^[47], retrospectively identify the specific cause of the error, and consequently is unable to determine the liable entity^[48]. In such a case, there would be no entity liable for the damage that occurred^[49] and it would be an example referred to in the literature as a responsibility gap^[50].

5. Conclusion

A robotic interpreter or translator could in the future represent an efficient alternative to a human court interpreter and translator, particularly in terms of economic efficiency, speed, and objectivity of interpreting and translation processes.

From an economic perspective, these technologies could bring significant savings in operational costs, as after the initial investment in acquiring the system and training personnel, the costs of their long-term operation would be lower than the fees paid to court interpreters and translators.

Furthermore, the use of a robotic interpreter or translator would lead to considerable acceleration of administrative processes, as delays associated with organizing interpretation or waiting for the production of a translation by a court translator would be eliminated.

Another significant advantage would be the objectivity of interpretation. It would be free from subjective elements, as automated systems do not work with individual interpretations or subjective views of the content of the message, which could reduce the risk of distortion of the meaning of interpretation,

thereby contributing to greater neutrality.

At present, however there is no such system that could equal the performance of a “human” interpreter^[51], and contemporary technologies of machine interpreting or translation cannot be considered suitable for use in proceedings before public authorities. The reasons are primarily their insufficient quality and reliability, which can lead to significant errors in outputs, distortion of statements by parties to the proceedings, or incomprehensibility of translated text. Such deficiencies would directly threaten the right of parties to the proceedings to interpretation, and thereby their right to a fair trial.

In the field of translation, however, it can be reasonably anticipated that technologies enabling automated text processing will achieve a high level of accuracy in the near future. Particularly for repetitive acts, such as the translation of standardized documents like birth certificates, the effective use of machine translation can be considered given their template-like structure. In these cases, automated systems could achieve a high degree of reliability, as their outputs would be based on predefined patterns and established terminological databases. Alongside technical limitations, however, there also exists a legal vacuum. Current legislation does not contain any framework regulating liability for incorrect outputs of machine translation or interpreting, nor does it address mechanisms by which it would be possible to ensure defense against potential deficiencies in machine-generated outputs. It is not yet clear how one would proceed, for example, in the case of objections by a party to the proceedings against the incorrectness of translation or interpreting—whether the output would be reprocessed by machine, possibly by different software, or whether it would be necessary to call in a court interpreter or translator for verification. The absence of these legal solutions currently represents one of the obstacles to the deployment of these technologies in practice.

On the other hand, it is undeniable that legislation will sooner or later have to adapt to rapidly developing technologies, particularly in the field of machine translation. Section 492 of the draft Act on the Residence of Foreign Nationals can be perceived in this regard as a first step toward integrating these innovations into the area of public administration. Given the current challenges and unresolved questions, however, it would be advisable to approach this possibility with prudence.

One possible way to integrate machine outputs into proceedings before public authorities could be a pilot project^[52], within which a machine interpreting or translation system would operate alongside a human interpreter or translator, who would simultaneously perform a control role or post-editing of outputs. It would be necessary to test the device for each language under consideration separately, as the device learns from different datasets for each language pair. Such experimental deployment would enable not only evaluation of the actual quality and benefit of the technologies in real conditions, but also identification of weaknesses and preparation of appropriate legislative measures.

The introduction of machine interpreting or translation into administrative proceedings has the potential to significantly affect the rights of parties to the proceedings, particularly their right to interpretation, and thereby the right to a fair trial. In this context, a legitimate question arises as to whether the primary motive for introducing machine interpreting is not primarily economic savings, at the expense of the quality of interpretation and protection of the rights of parties to the proceedings. Particularly concerning is also the risk of discrimination against persons who speak less widespread languages, for which sufficiently high-quality language models do not exist. It is evident that the draft completely disregards the fundamental role of court interpreting as a procedural safeguard, which serves not only to overcome language barriers, but primarily to protect the rights of parties and to ensure fairness of proceedings.

At this moment, the most ideal approach appears to be a hybrid model that combines the use of artificial intelligence with the human factor. Such models promise to accelerate and simplify the work of court interpreters and translators while maintaining the required level of accuracy. In the field of translation, these are so-called CAT tools (computer-assisted translation)^[53] and in the field of interpreting, computer-assisted interpreting^[54], which involves the use of digital technologies to increase the efficiency and accuracy of human interpreter work^[55].

From the perspective of the systematics of legal regulation, it can be recommended *de lege ferenda* that in the event of achieving such technological progress that would allow machine interpreting and translation to be considered equivalent to the performance of a court interpreter and translator, the legal regulation of their use should be enshrined directly in procedural regulations.

This would ensure uniform and coherent application of these technologies in all types of proceedings, instead of their partial and unsystematic regulation in individual sectoral acts.

1. I. Čeňková, *Úvod do teorie tlumočení*, Česká komora tlumočnicků znakového jazyka, Prague, 2008, p. 11.
2. D. Knittlová et al., *Překlad a překládání*, Univerzita Palackého v Olomouci, Filozofická fakulta, Olomouc, 2010, pp. 221-225.
3. C. Wadensjö, *Interpreting as Interaction*, Addison Wesley Longman, New York, 1998, p. 74.
4. «*In the determination of any criminal charge against him, everyone shall be entitled to the following minimum guarantees, in full equality: [...] To have the free assistance of an interpreter if he cannot understand or speak the language used in court*».
5. «*Everyone who is arrested shall be informed promptly, in a language which he understands, of the reasons for his arrest and of any charge against him.*» a «*Everyone charged with a criminal offence has the following minimum rights: [...] to have the free assistance of an interpreter if he cannot understand or speak the language used in court.*».
6. J. Stavínohová, *Zásada rovnosti účastníků*, in D. Hendrych et al., *Právníký slovník*, III Ed., C. H. Beck, Prague, 2009.
7. J. Blahož, K. Klíma, J. Skála, *Ústavní právo Evropské unie*, Aleš Čeněk, Dobrá Voda, 2003, p. 590.
8. Art. 36(1) Charter of Fundamental Rights and Freedoms of the Czech Republic.
9. Constitutional court, judgement, 22 January 2008, n. Pl. ÚS 54/05.
10. W. Hewitt, in H. Mikkelsen, *Evolving Views of the Court Interpreter's Role: Between Scylla and Charybdis*, in C. Valero-Garcés, A. Martin (eds.), *Crossing Borders in Community Interpreting: Definition and Dilemmas*, John Benjamins, Amsterdam, 2008, p. 81.
11. A. Winterová, *Civilní právo procesní*, III Ed., Linde, Prague, 2004, pp. 66-67.
12. Constitutional court, judgement, 12 March 2019, n. III. ÚS 3464/17.
13. Constitutional court, judgement, 8 August 2005, n. II. ÚS 186/05.
14. Supreme Court, resolution, 12 July 2001, n. 6 To 550/2001.
15. Constitutional court, judgement, 7 July 2021, n. II. ÚS 482/21.
16. T. Etchegoyhen et al., *Cascade or Direct Speech Translation? A Case Study*, in *Applied Sciences*, 22, 3, 2022, pp. 1019-1121.
17. J. Iranzo-Sánchez et al., *Streaming cascade-based speech translation leveraged by a direct segmentation model*, in *Neural Networks*, 142, 2021, pp. 303-315.
18. M. Labied et al., *A Multicriteria Comparison of End-to-End and Cascade Speech-to-Text Translation Models*, in *Bulletin of Electrical Engineering and Informatics*, 14, 4, pp. 2837-2848.
19. H. Wang et al., *Progress in Machine Translation*, in *Engineering*, 18, 11, 2022, p. 148.
20. C. Fantinuoli, *The Emergence of Machine Interpreting*, in *EST Newsletter*, 62, 2023, p. 1.

21. Ibid, p. 12.
22. W. Xiaoman, C. Fantinuoli, *Exploring the Correlation between Human and Machine Evaluation of Simultaneous Speech Translation*, in *Proceedings of the 25th Annual Conference of the European Association for Machine Translation (volume 1)*, Sheffield: European Association for Machine Translation, 2024, pp. 329-330.
23. N. P. Russo, *Faut-il craindre l'interprétation automatique ? De la fiction à la réalité, le point sur les «traducteurs électroniques»*, in *Traduire*, 246, 2022, p. 102.
24. Y. Cao, *Comparative Analysis of Machine Interpreting and Human Interpreting: Insights into Consecutive Interpreting Teaching*, in *Advances in Social Science, Education and Humanities Research*, 914, 851, 2024, pp. 730-741.
25. F. Pöschhacker, *Is Machine Interpreting Interpreting?*, in *Translation Spaces*, 2024, p. 13.
26. W. J. Hutchins, H. Somers, *An Introduction to Machine Translation*, Academic Press, London, 1992, p. 150; cf. P. Naveen, P. Trojovský, *Overview and Challenges of Machine Translation for Contextually Appropriate Translations*, in *iScience*, 27, 10, 2024, p. 2.
27. Deep learning is a specific method of machine learning that uses deep neural networks. Each layer analyzes the inputs it receives from the previous layer, and its output is further processed by the next layer.
28. Y. Zhao, J. Zhang, C. Zong, *Transformer: A General Framework from Machine Translation to Others*, in *Machine Intelligence Research*, 20, 4, 2023, pp. 515-519.
29. S. Lankford, H. Afli, A. Way, *Human Evaluation of English-Irish Transformer-Based NMT*, in *Information*, 13, 7, 2022, pp. 1-19.
30. A. Anugu, G. Ramesh, *A Survey on Hybrid Machine Translation*, in *E3S Web of Conferences*, 2020, p. 3.
31. R. K. Dwivedi, P. Nand, O. Pal, *Hybrid NMT Model and Comparison with Existing Machine Translation Approaches*, in *Multidisciplinary Science Journal*, 7, 4, 2024, p. 1.
32. A. Sugiyama, N. Yoshinaga, *Data Augmentation Using Back-Translation for Context-Aware Neural Machine Translation*, in *Proceedings of the Fourth Workshop on Discourse in Machine Translation (DiscoMT 2019)*, Association for Computational Linguistics, Hong Kong, 2019, pp. 35-44.
33. Chamber of Court Interpreters and Translators of the Czech Republic, *AI – umělá inteligence v překladatelské praxi?*, in *Soudní tlumočník*, 2, 2023, p. 8.
34. P. Naveen, P. Trojovský, *Overview and Challenges of Machine Translation for Contextually Appropriate Translations*, in *iScience*, 27, 10, 2024, pp. 8-9.
35. For less widely spoken languages, an additional challenge arises in the form of limited training data. To improve the performance of neural machine translation models, back-translation may be employed; however, its main drawback lies in the generation of so-called pseudo-sentences, i.e. low-quality translations that may negatively affect model performance (S. Ranathunga et al., *Neural Machine Translation for Low-Resource Languages: A Survey*, in *ACM Computing Surveys*, 55, 11, 2023, pp. 1-37).
36. Translation from Czech; original wording in Czech.
37. According to the explanatory memorandum to the draft law on the residence of

foreigners, Section 492, it is up to the administrative authority to decide whether technical equipment will be used.

38. Translation from Czech; original wording in Czech.
39. I. Horváth, *AI in Interpreting: Ethical Considerations*, in *Across Languages and Cultures*, 23, 1, 2022, pp. 8-9. Cf. J. Moorkens, D. Kenny, F. De Carmo, *Fair MT: Towards Ethical, Sustainable Machine Translation*, in *Translation Spaces*, 9, 1 (special issue), 2020, pp. 1-11.
40. C. Canfora, A. Ottmann, *Risks in Neural Machine Translation*, in *Translation Spaces*, 9, 1, 2020, pp. 58-77. Cf. K. Siau, W. Wang, *Artificial Intelligence (AI) Ethics: Ethics of AI and Ethical AI*, in *Journal of Database Management*, 31, 2, 2020, pp. 74-87. P. Yang et al., *Data Security and Privacy Protection for Cloud Storage: A Survey*, in *IEEE Access*, 8, 2020, p. 2.
41. Uncertainty in this area could negatively affect the exercise of the rights of parties to proceedings, particularly with regard to claims for damages, which could undermine public confidence in procedural fairness.
42. The situation described above can be compared to liability for damage caused by particularly dangerous operations under Section 2925(1) of the Civil Code, where the operator is always the person responsible for a particularly dangerous plant. According to this provision, anyone who operates a particularly dangerous plant or other facility shall compensate for damage caused by the source of increased danger; operation is particularly dangerous if the possibility of serious damage cannot be reasonably excluded in advance, even with the exercise of due care. Case law has concluded that things whose use, under certain circumstances, has secondary harmful effects that are not fully controllable by humans and which are associated with a high degree of probability of causing harm (such as a frame saw located in unsuitable geological conditions) should also be considered a source of increased danger (Supreme Court of the Czechoslovak Socialist Republic, judgment 31 May 1983, n. 1 Cz 13/83 [R 24/1986]). Artificial intelligence, which is capable of reaching its own conclusions based on collected and analyzed data and subsequently behaving in a manner that may be unpredictable, can be considered a source of increased danger.
43. The responsible person pursuant to Sections 2939 and 2940 of the Civil Code may be (i) the manufacturer of the product, (ii) the quasi-manufacturer (a person who labels the product with their name, trademark, or other symbol that leads the customer to believe that such person is the manufacturer), (iii) the importer who imported the product as part of their business for the purpose of placing it on the market, or (iv) if the manufacturer or importer cannot be identified, the supplier is the responsible person (M. Hulmák et al., *Commentary on Sections 2055–3014*, in M. Hulmák (ed.), *Civil Code: Commentary*, vol. VI, Obligations Law – Special Part, C. H. Beck, Prague, 2014, p. 1652.).
44. In the development of artificial intelligence systems, it is also relevant to consider the concept of shared responsibility among entities involved in design, programming and production (D. C. Vladeck, *Machines without Principles: Liability Rules and Artificial*

- Intelligence*, in *Washington Law Review*, 89, 2014, p. 149.).
45. T. Svoboda, *Chatboty ve veřejné správě – stručný nástin problematiky*, in *Správní právo*, 6-7, 2024, p. 513.
 46. J. Nešpor, *Automated Administrative Decision-Making: What Is the Black Box Hiding?*, in *Acta Universitatis Carolinae – Iuridica*, 2, 2024, p. 69 ff.
 47. J. Strakoš, *Právní aspekty automatizace ve správním trestání v kontextu algoritmů strojového učení*, in *Správní právo*, 6-7, 2024, p. 486.
 48. M. Koubová, *Umělá inteligence ve zdravotnictví: Kdo ponese odpovědnost v případě škody?*, *Ekonomický deník*, online, 16 August 2019.
 49. L. Kolaříková, *Odpovědnost (za) robota aneb právo umělé inteligence*, in *Bulletin advokacie*, 3, 2018, p. 15.
 50. D. G. Johnson, *Technology with No Human Responsibility?*, in *Journal of Business Ethics*, 127, 4, 2015, pp. 707-715 or A. Matthias, *The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata*, in *Ethics and Information Technology*, 6, 2004, pp. 175-183.
 51. Explanatory Memorandum to the Draft Act on the Entry and Residence of Foreign Nationals, Section 492, p. 434.
 52. A pilot project can be likened to a regulatory sandbox within the meaning of the Artificial Intelligence Regulation (AI Act), which provides businesses with the opportunity to experiment with new products and services with limited application of certain regulatory requirements, thereby facilitating innovation and investment in the field of artificial intelligence.
 53. CAT tools are software programs that help translators streamline their work by storing previously translated segments, managing terminology, and ensuring consistency and accuracy in translation (T. Svoboda, *Kapitoly z překladatelské praxe: odborný překlad mezi němčinou a češtinou*, Charles University, Faculty of Arts, Prague, 2012, p. 80.).
 54. CAI tools include, for example, speech-to-text tools (automatic transcription of spoken words on screen) or remote interpreting tools that connect interpreters and participants remotely.
 55. B. Prandi, *Computer-Assisted Simultaneous Interpreting: A Cognitive-Experimental Study on Terminology*, Language Science Press, Berlin, 2023, pp. 35-38.

AI as a tool for ensuring the housing rights of internally displaced persons in Ukraine

Nataliya Isayeva, Liliia Serhiichuk

DOI: 10.13130/2723-9195/2026-3-7

Il contributo analizza le sfide connesse alla garanzia dell'accesso a un'abitazione adeguata per gli sfollati interni in Ucraina e valuta il potenziale dell'intelligenza artificiale (IA) nel migliorare l'assegnazione degli alloggi e il funzionamento dei registri degli sfollati interni. Il fenomeno dello sfollamento interno su larga scala, iniziato nel 2014 e intensificatosi a seguito dell'invasione su vasta scala del 2022, ha messo in luce la frammentazione dei registri statali, l'assenza di un coordinamento digitale unitario e la limitata capacità istituzionale di soddisfare il fabbisogno abitativo. Sulla base di richieste di accesso alle informazioni presentate alle autorità centrali e regionali, lo studio individua significative lacune nei programmi abitativi esistenti, l'assenza di una risorsa nazionale centralizzata e incoerenze nelle prassi amministrative.

L'articolo sostiene che l'integrazione dell'IA nel nuovo Sistema informativo e analitico dei beni immobili potrebbe migliorare l'accuratezza dei dati, automatizzare le verifiche, individuare frodi, supportare analisi predittive e migliorare i processi decisionali. Gli strumenti basati sull'IA potrebbero inoltre semplificare l'interazione con gli utenti, accelerare l'esame delle domande e ridurre i ritardi burocratici. Al contempo, lo studio evidenzia i rischi giuridici ed etici connessi alla protezione dei dati, alla trasparenza e ai bias algoritmici, sottolineando la necessità di armonizzare la legislazione ucraina con l'AI Act dell'Unione europea e con il GDPR.

This article analyses the challenges of ensuring access to adequate housing for internally displaced persons (IDPs) in Ukraine and evaluates the potential of artificial intelligence (AI) to improve housing allocation and the functioning of IDP registries. Large-scale internal displacement since 2014, intensified after the 2022 full-scale invasion, has exposed fragmentation in state registries, a lack of unified digital coordination, and limited institutional capacity to meet housing needs. Based

on information requests submitted to central and regional authorities, the study identifies significant gaps in existing housing programmes, the absence of a centralised national resource, and inconsistencies in administrative practices. The article argues that integrating AI into the newly established Information and Analytical System of Real Estate Assets could enhance data accuracy, automate verification, detect fraud, support predictive analytics, and improve decision-making. AI-based tools may also simplify user interaction, accelerate application processing, and reduce bureaucratic delays. At the same time, the study highlights legal and ethical risks related to data protection, transparency, and algorithmic bias, emphasising the need to harmonise Ukrainian legislation with the EU AI Act and GDPR.

Summary: 1. Introduction.- 2. Constitutional and legal framework of housing rights of IDPs.- 3. Gaps of the existing housing support system for IDPs.- 4. AI as a tool for improving housing allocation and IDP registries.- 5. Legal safeguards and compliance with EU AI standards.- 6. Conclusions.

1. Introduction

The large-scale internal displacement in Ukraine, significantly intensified since 2014 and further escalated following the full-scale invasion in 2022, has created a persistent and complex challenge in ensuring access to adequate housing for IDPs^[1].

The number of IDPs in Ukraine has changed significantly from 2014 to 2024. While the figure remained around 1.5 million in 2021, in 2022, more than 7.5 million IDPs were officially registered. This large-scale increase has significantly affected the state's ability to ensure these citizens' constitutional rights, in turn affecting the level of social protection available to them.

The analysis begins with a brief overview of macro-level indicators that demonstrate the scale, significance, and ongoing relevance of the issue under consideration. This dynamic is clearly reflected in the changing number of IDPs, as illustrated in the diagram below^[2].

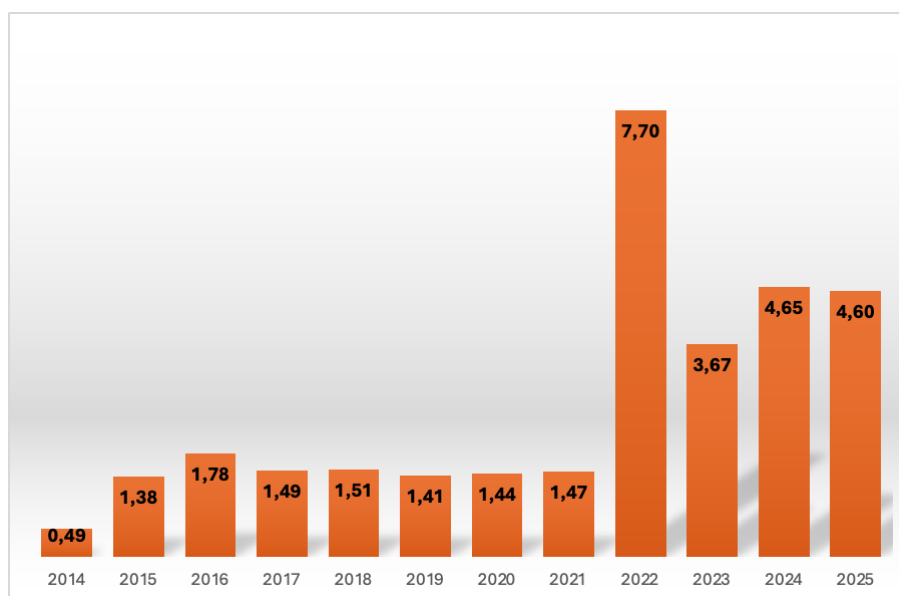
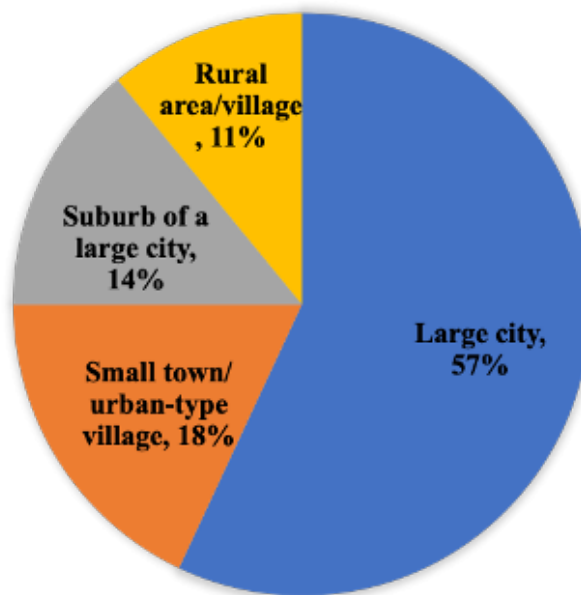


Figure 1

In addition to the overall number of IDPs, it is also necessary to consider their uneven territorial distribution across the country, including differences by type of settlement. In particular, the chart «*Share of IDPs by Type of Settlement*» presented below clearly illustrates the concentration of displaced persons in large urban areas. This pattern, in turn, may lead to localised pressures on the housing market, the labour market, and other public sectors.

Figure 2^[3]

One of the key obstacles in this context is the fragmentation and inefficiency of state registries and administrative procedures related to housing provision for IDPs. Dispersed databases, limited interoperability between public authorities, and a lack of unified digital infrastructure hinder timely and transparent decision-making. As a result, many IDPs continue to face difficulties in accessing reliable information about available housing programs and in effectively exercising their constitutional right to housing.

In response to these challenges, increasing attention has been given to the potential use of AI and data-driven technologies in public administration. AI-based tools may significantly improve the management of IDP-related data by enhancing data accuracy, enabling predictive analytics, reducing administrative burdens, and supporting more efficient allocation of housing resources. However, the integration of AI into such sensitive domains raises important legal and ethical concerns, particularly regarding data protection, transparency, accountability, and the prevention of algorithmic bias.

At the same time, the legal framework governing the use of AI in Ukraine remains fragmented and underdeveloped. While certain aspects are regulated through legislation on digital transformation, personal data protection, and

electronic governance, Ukraine has not yet adopted a comprehensive legal act addressing the development and use of AI systems. In this regard, the European Union's Artificial Intelligence Act (AI Act) represents a relevant reference model for shaping national regulatory approaches.

The purpose of this study is to comprehensively assess access to adequate housing for IDPs in Ukraine and to analyze the realization of their constitutional right to housing under martial law, as well as to substantiate the feasibility of integrating AI into existing state registers in order to enhance the effectiveness of implementing this constitutional right, optimize the work of public officials, minimize excessive bureaucratic procedures, and accelerate the provision of public services for IDPs. In addition, the study emphasises the need to establish a National Register of State, International, and Local Housing Credit Support Programs for IDPs (the Unified State Register of Preferential Credit Programs for IDPs) and to integrate AI tools into this register.

This study involved the systematic collection of information from all relevant public authorities at both the regional and national levels through the submission of formal information requests. The data obtained are subsequently used as primary empirical material, enabling the authors to reconstruct the current de facto governance model in matters related to the redistribution of resources aimed at safeguarding the rights and freedoms of IDPs. This approach also allows for the identification of structural shortcomings, gaps in the coherence of institutional practices, and key areas for improving governance processes.

2. Constitutional and legal framework of housing rights of IDPs

According to Article 47, 48 of the Constitution of Ukraine, everyone shall have the right to housing. The State shall create conditions enabling every citizen to build, purchase, or rent housing. Citizens in need of social protection shall be provided with housing by the government authorities and local government, free of charge or at a price affordable for them in accordance with law. No one shall be forcibly deprived of housing other than based on the law pursuant to a court's decision. Everyone shall have the right to a standard of living sufficient for themselves and their families including adequate nutrition, clothing, and

housing^[4].

In addition to the Constitution of Ukraine, the housing rights of IDPs are regulated by the Law of Ukraine «*On Ensuring the Rights and Freedoms of Internally Displaced Persons*», as well as by a number of subordinate regulatory acts, including: «*Certain Measures for the Formation of Housing Funds Intended for the Temporary Accommodation of Internally Displaced Persons*» (Resolution of the Cabinet of Ministers of Ukraine of April 29, 2022, No. 495); «*On Approval of the Procedure for the Inspection of Real Estate Objects for Housing Internally Displaced Persons and the Standard Form of the Real Estate Inspection Act*» (Resolution of the Cabinet of Ministers of Ukraine of April 29, 2025, No. 489); «*Certain Issues of Organizing the Inspection of Real Estate Objects for Housing Internally Displaced Persons*» (Resolution of the Cabinet of Ministers of Ukraine of April 29, 2025, No. 493); «*Certain Issues Concerning the Establishment and Functioning of the Information and Analytical System of Real Estate Assets for Ensuring Housing for Internally Displaced Persons and Its Interaction with Other Information and Communication Systems*» (Resolution of the Cabinet of Ministers of Ukraine of July 16, 2025, No. 894), as well as other regulatory acts.

According to Article 41 of the Law of Ukraine «*On Ensuring the Rights and Freedoms of Internally Displaced Persons*», the Unified Information Database on Internally Displaced Persons is established for the purpose of registering such persons. The central executive authority responsible for the formulation and implementation of state policy in the areas of social protection, social services for the population, volunteer activity, family and children's affairs, children's health improvement and recreation, as well as the protection of the rights of persons deported on the basis of nationality who have returned to Ukraine, is entrusted with ensuring the establishment and maintenance of the Unified Information Database on Internally Displaced Persons^[5].

The Unified Information Database on IDPs is owned by the State, represented by the Ministry of Social Policy of Ukraine, which also holds the exclusive proprietary intellectual property rights to its software. The database contains information on the housing and other essential needs of IDPs and the level of their satisfaction.

For assessing housing needs, relevant data from the database may be provided to

the Ministry for Communities and Territories Development of Ukraine upon request. Accordingly, the Ministry of Social Policy serves as the central executive authority possessing up-to-date information on the housing needs of IDPs.

Pursuant to the Resolution of the Cabinet of Ministers of Ukraine of 21 July 2025 on the optimization of central executive authorities, the Ministry of Social Policy was reorganized into the Ministry of Social Policy, Family and Unity of Ukraine, which became the legal successor to the rights and obligations of the Ministry of National Unity. Consequently, the Ministry of Social Policy, Family and Unity of Ukraine is currently the central executive authority responsible for ensuring the rights of IDPs.

3. Gaps of the existing housing support system for IDPs

In order to examine the state of housing provision for IDPs in Ukraine, the authors submitted information requests to the Ministry of Social Policy of Ukraine, the Ministry of National Unity of Ukraine, the Ministry for Communities and Territories Development of Ukraine, the Ministry of Finance of Ukraine, the State Specialized Financial Institution «*State Fund for Youth Housing Construction Assistance*», and all Regional State Administrations regarding the following issues: housing programs for IDPs that were implemented or remain in force in Ukraine during the period from 24 February 2022 to 15 July 2025; the number of IDPs registered nationwide as persons in need of housing or improvement of housing conditions during the period from 24 February 2022 to 15 July 2025, broken down by year and by region; where information on available housing programs for IDPs in Ukraine is published; whether there is a centralized online resource in Ukraine (website, register, etc.) that contains comprehensive information on all housing programs available to IDPs nationwide, including those involving credit financing, and, if so, a request to provide a link to such a resource.

However, in its letter dated 24 July 2025 No. 420/0/98-25/314, the Ministry of Social Policy of Ukraine stated that it does not possess the requested information and forwarded the authors' inquiry to the Ministry for Communities and Territories Development of Ukraine.

The Ministry of National Unity (which until 3 December 2024 operated under

the name Ministry for Reintegration of the Temporarily Occupied Territories of Ukraine), in its letter dated 22 July 2025 No. 22/4.1-2639-25, reported that, according to the information available, during the period from 2021 to 2024 the Ministry for Reintegration of the Temporarily Occupied Territories of Ukraine, together with its partners, implemented international technical assistance projects (hereinafter referred to as ITA projects), namely: «*Support to the Development of Social Infrastructure (UFSI VI)*», «*Support to the Development of Social Infrastructure (UFSI VIII)*», «*Support to the Development of Social Infrastructure (UFSI IX)*», «*Improving Housing Conditions for Internally Displaced Persons in Eastern Ukraine*», and «*Housing Premises for Internally Displaced Persons*». The Ministry for Community Development provided the following information, as summarised below:

Table 1

Nº	Program Name	Responsible Executor	Program Implementation Period	Allocated Funds for Program Implementation	Funding Sources (state/local/budget/etc.)	Number of IDP Families Receiving Funds per Year	Number of IDP Families Provided with Housing per Year
1	«Provision of Preferential Mortgage Loans to IDPs» (CMU Resolution No. 451 of 28/04/2021)	SSFU «State Youth Housing»	2021–2028	For 2025 – UAH 215.67 million from the state budget	State budget, KfW reconstruction grant	2021 – 241 families 2022 – 332 families 2023 – 94 families 2024 – 322 families 2025 – 64 families	2021 – 241 families 2022 – 332 families 2023 – 94 families 2024 – 322 families 2025 – 64 families
2	Provision of Preferential Mortgage Loans to IDPs using loans from the Council of Europe Development Bank	SSFU «State Youth Housing»	Grant agreement of 2025 between Ukraine and CEB	Planned funding 2026–2028	State budget, grant funds from the Council of Europe Development Bank	–	–

CERIDAP

3	Budget program for providing long-term preferential loans to IDPs, ATO participants and/or Joint Forces Operation participants for housing (CMU Resolution No. 980 of 17/11/2019)	SSFU «State Youth Housing»	2019–2028	For 2025 – UAH 22.00 million from the state budget	State budget	2022 – 7 families2023 – 6 families2024 – 9 families2025 – 3 families	2022 – 7 families2023 – 6 families2024 – 9 families2025 – 3 families
	State preferential lending to individual rural developers for construction (reconstruction) and purchase of housing (CMU Resolution No. 1547 of 05/10/1998)	Regional Support Funds	Accounting of citizens with IDP status has been conducted since 2022; funding planned until 2028	For 2025 – UAH 31.56 million from the state budget	State budget	2022 – 37 families2023 – 54 families2024 – 21 families	2022 – 37 families2023 – 54 families2024 – 21 families
4							

This information pertains exclusively to housing acquisition or mortgage lending. Furthermore, in the aforementioned response, the Ministry of Communities and Territories Development of Ukraine recommended seeking more detailed information from the State Specialized Financial Institution «*State Fund for Youth Housing Assistance SSFU*» «*State Youth Housing*» (hereinafter referred to as the Fund), which is indicated as the responsible executor of the housing provision programs for IDPs listed in the table. However, in its response to authors dated 21 July 2025, No. 1436/005, the Fund stated that it does not possess information regarding any such programs, including where the information is located or who holds it.

Consequently, the organization designated as responsible for the implementation of state programs, to which the Ministry refers, does not have any information regarding the existence of such programs. As a result, internally displaced persons who approach this institution to obtain information about the programs in which they may participate do not receive such information. This situation constitutes a violation not only of their constitutional right to access information but also of their right to housing, leading to a failure of the State to fulfill its obligations toward individuals and to guarantee the protection of human rights.

Additionally, the Ministry of Communities and Territories Development of Ukraine provided a response stating that there is no single centralised online resource containing information on all housing programs available to IDPs across the country.

Regarding the number of IDPs registered nationwide as requiring housing or improvements to their living conditions for the period from 24 February 2022 to 15 July 2025, the Ministry provided the information presented below.

Table 2

№	Administrative-Territorial Unit	Social Housing Register / Temporary Housing Fund Register			
		2022	2023	2024	2025 (I half)
1	Vinnitska	67/137	129/353	129/353	214/548
2	Volynska	37/24	97/230	114/391	113/471
3	Dnipropetrovska	203/586	204/752	204/752	130/485
4	Donetska	29/622	41/789	44/792	43/774
5	Zhytomyrska	8/104	35/223	51/1402	54/1455
6	Zakarpatska	3/703	3/267	3/267	3/267
7	Zaporizka	1/39	0/39	0/103	8/119
8	Ivano-Frankivska	32/76	35/57	33/708	71/905
9	Kyivska	0/0	0/0	211/773	211/773
10	Kirovohradska	0/0	0/715	0/1007	0/2459
11	Luhanska	0/0	0/0	0/0	0/0

CERIDAP

12 Lvivska	0/430	0/510	0/743	0/1581
13 Mykolaivska	5/35	0/112	0/112	0/112
14 Odeska	5/109	5/356	10/373	7/0
15 Poltavska	32/300	816/817	38/1028	38/1180
16 Rivnenska	0/477	0/704	5/704	1/1796
17 Sumska	43/69	2/114	3/196	3/335
18 Ternopil'ska	2/14	10/390	12/1247	13/1423
19 Kharkiv'ska	46/112	71/165	57/1906	114/446
20 Kherson'ska	0/0	5/36	5/384	5/384
21 Khmelnytska	23/374	23/378	23/1169	23/1176
22 Cherkaska	30/144	30/226	48/454	52/1076
23 Chernivetska	0/0	0/982	0/2272	0/2662
24 Chernihiv'ska	2/404	0/571	7/333	4/327
25 Kyiv City	38/1418	76/1346	131/2657	54/3069

606/6177	1582/10132	1128/20126	1161/23814
(186	(352	(399	(434
families	families	families	families
provided	provided	provided	provided
with	with social	with social	with social
social	housing /	housing /	housing /
housing /	2868 IDP	3950 IDP	4205 IDP
2472 IDP	families	families	families
families	provided	provided	provided
provided	with	with	with
with	temporary	temporary	temporary
temporary	housing)	housing)	housing)
housing)			

A comparative analysis of the number of IDPs over the years (Figure 1) and the Ministry's data on the total number of IDPs nationwide (Table 2) who need housing demonstrates that IDPs are largely compelled to address their housing needs independently, primarily by renting accommodations at their own expense. This observation is further corroborated by the data in Figure 3 bellow.

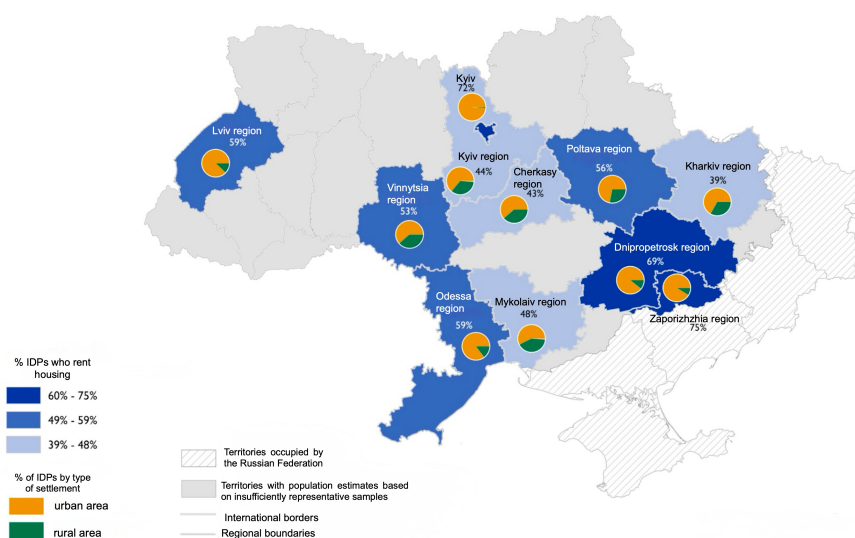


Figure 3^[6]

In 2021, Isayeva N. proposed the creation and management of a Unified State Register of state, communal, and private real estate that could be made available for IDPs, including the introduction of relevant amendments to the Law of Ukraine «*On Ensuring the Rights and Freedoms of Internally Displaced Persons*»^[7]. In November 2024, the relevant amendments were adopted, and an Information and Analytical System of Real Estate Objects was introduced to provide housing for internally displaced persons^[8]. On July 16, 2025, by Resolution No. 894 of the Cabinet of Ministers of Ukraine, the Procedure for the Functioning of the Information and Analytical System of Real Estate Objects for Providing Housing to Internally Displaced Persons and Its Interaction with Other Information and Communication Systems (hereinafter – the Information and Analytical System, IAS) was approved^[9].

The Resolution also establishes the procedure for the development and implementation of the Information and Analytical System, ensuring access for IDPs and integration with other state registers. Since this IAS is currently being established, the Verkhovna Rada of Ukraine and the Cabinet of Ministers of Ukraine have time to legislatively approve and implement AI tools within the system.

4. AI as a tool for improving housing allocation and IDP registries.

The legal basis for the provision of electronic administrative services in Ukraine is established by the Law of Ukraine «*On Administrative Services*», which provides for the functioning of the Unified State Web Portal of Electronic Services as the official source of information on administrative services in Ukraine^[10]. In addition, the Concept for the Development of Digital Competencies, approved by the Cabinet of Ministers of Ukraine, forms part of the broader national strategy for digital transformation and the integration of modern technologies into public administration^[11].

Despite progress in digital governance, the current system of IDP registration and housing support remains fragmented and administratively burdensome. Traditional information systems have proven insufficiently flexible to process rapidly changing data on displacement, housing needs, and social support. In

particular, the lack of unified digital coordination among state institutions complicates the effective allocation of housing resources and slows decision-making processes.

In this context, AI technologies may become an effective tool for improving housing governance and the functioning of IDP registries. AI systems based on machine learning, natural language processing, and intelligent data analysis can significantly improve the processing of large volumes of information and support evidence-based administrative decision-making.

Therefore, while the IAS is still in the stage of development and integration with other state registers, it is necessary to develop appropriate procedures and define algorithms for processing information, including confidential data, as well as for protecting the personal data of individuals whose information will be included in the IAS.

The authors believe that integrating AI into the IAS will simplify the work of the responsible commissions, as it will allow:

- rapid verification and updating of IDP data;
- detection of duplicate applications and prevention of fraud;
- forecasting housing and social support needs;
- optimisation of housing allocation procedures;
- analytical support for state authorities responsible for resource distribution.
- automatic assessment of the technical condition of real estate objects (based on photos/videos);
- intelligent classification of housing by parameters (suitability, location, energy efficiency) with forecasting of financial needs and potential resources for housing stock renovation;
- automatic generation of reports, text recognition, and digitisation of information;
- obtaining recommendation algorithms for housing allocation.

For IDPs, the use of AI will significantly simplify interaction with the IAS, as it would enable: the implementation of a chatbot for user interaction and application submission; intelligent housing selection taking into account family

size, age, health condition, place of work or study, and accessibility of infrastructure; faster data processing and minimization of the time between application submission and receiving a decision; and elimination of the human factor in considering all parameters necessary for making a well-founded decision. Furthermore, a centralised, uniform approach with real-time data synchronisation could enable the redistribution of workload across the country based on available resources and facilitate evidence-based analysis to help individuals choose a specific settlement before relocation. This could also accelerate the resolution of the sensitive issue of individuals refusing to leave dangerous areas due to uncertainty and fear of the unknown. Awareness that housing is available, coupled with access to detailed information and real photos, can significantly speed up the decision to leave conflict zones, thereby reducing civilian casualties and easing the work of security forces.

For the state as a whole, the use of AI in this IAS will enable: forecasting housing needs for IDPs by integrating with the Unified Information Database on IDPs; mitigating corruption risks; detecting fraud; identifying duplicates, errors, and risks in the database; reducing the time required to manage the IAS; ensuring transparency in housing allocation and fund usage; automating the application submission process; monitoring efficiency through analysis of system use and housing allocation speed; and quickly obtaining statistical data for strategic decision-making in housing policy and fulfilling the state's obligations to provide housing.

The Ministry of Digital Transformation of Ukraine, together with the Ministry of Social Policy, has already initiated projects to integrate AI components into digital public services, particularly through the Diia ecosystem. system. In particular, digital services for internally displaced persons were introduced to automate procedures for social assistance and housing-related payments^[12]. These developments demonstrate the growing role of AI and data-driven technologies in modernising public administration and improving service delivery to vulnerable groups.

At the same time, the integration of AI into housing allocation systems and IDP registries requires appropriate legal safeguards. The use of intelligent systems in such sensitive areas must comply with the principles of transparency, accountability, data protection, and non-discrimination. Therefore, further

development of AI-based administrative systems in Ukraine should be accompanied by the gradual harmonisation of national legislation with relevant European standards, including the principles reflected in the EU AI Act and data protection regulations.

5. Legal safeguards and compliance with EU AI standards

The integration of artificial intelligence into housing allocation systems and registries for internally displaced persons requires an appropriate legal framework that ensures both administrative efficiency and the protection of fundamental rights. Although Ukraine has developed a regulatory framework for digitalisation, electronic governance, administrative services, and personal data protection, it still lacks comprehensive legislation governing the development and use of AI systems^{[13] [14] [15] [16] [17] [18]}.

These challenges are particularly relevant in the context of housing support for IDPs, where AI systems may process large volumes of personal and socially sensitive data. Therefore, the use of AI in housing allocation and social support mechanisms must be accompanied by safeguards that ensure transparency, accountability, non-discrimination, and the protection of personal data^{[19] [20]}.

In this regard, the European Union's regulatory approach provides an important reference point. Regulation (EU) 2024/1689 (EU AI Act) establishes a risk-based framework for AI governance and introduces requirements particularly relevant to public administration systems, including transparency, explainability, human oversight, cybersecurity, and accountability. These principles are closely linked to the standards of the General Data Protection Regulation (GDPR), which emphasises lawful processing of personal data, transparency, purpose limitation, and protection against fully automated decisions that affect individuals' rights^[21]
^[22].

As a candidate country for EU membership, Ukraine is gradually aligning its digital policy with European standards^{[23] [24] [25]}. Such harmonisation is especially important for AI-based systems intended for housing allocation and management of IDP registries. The introduction of AI tools into the Information and Analytical System of Real Estate Assets should therefore be accompanied by legal mechanisms that ensure algorithmic transparency, protect

personal data, provide effective human oversight, and safeguard against discriminatory outcomes.

Consequently, the development of AI-supported housing governance for IDPs should be based on the principles established by the EU AI Act and GDPR. This would enhance public trust, strengthen legal certainty, and ensure that technological innovation contributes to the effective realisation of the constitutional right to housing.

6. Conclusions

This study demonstrates that the current system of housing support for IDPs in Ukraine remains fragmented and insufficiently coordinated. The analysis of responses received from central executive authorities and regional administrations revealed significant institutional gaps, including the absence of a unified information resource containing comprehensive information on housing programmes available to IDPs, inconsistencies in administrative practices, and limited access to reliable information for potential beneficiaries.

The research confirms that, despite the existence of several state and international housing support programmes, most internally displaced persons continue to address their housing needs independently. The available mechanisms do not provide effective coordination of information, transparent allocation of resources, or comprehensive assessment of housing demand at the national level.

A significant step towards improving this situation was the establishment of the Information and Analytical System of Real Estate Assets for Ensuring Housing for Internally Displaced Persons. However, the effectiveness of this system will largely depend on its ability to integrate information from different state registers and provide timely, accurate, and transparent decision-making.

The study argues that AI may become an important instrument for strengthening the implementation of the constitutional right to housing for IDPs. AI technologies can improve verification procedures, identify duplicate applications, detect fraud risks, support predictive analysis of housing needs, and facilitate evidence-based allocation of available housing resources. In addition, AI-powered tools may simplify communication between IDPs and public authorities, reduce bureaucratic barriers, and accelerate access to housing support

programmes.

Particular attention should be paid to the creation of a Unified State Register of State, International, and Local Housing Support Programmes for IDPs. The establishment of such a register, combined with AI-based analytical and recommendation tools, would improve access to information, increase transparency of housing policies, and enhance coordination among public authorities responsible for housing provision.

At the same time, the integration of AI into housing allocation systems must be accompanied by appropriate legal safeguards. The use of AI in this sensitive area should comply with the principles of transparency, human oversight, accountability, non-discrimination, and personal data protection. In this regard, the standards established by the EU AI Act and GDPR provide an appropriate regulatory framework for the future development of AI-based housing governance in Ukraine.

The study concludes that AI should not be viewed as a substitute for public policy but rather as a tool that can increase the effectiveness of existing housing support mechanisms. Its responsible implementation within housing registries and allocation systems may contribute to the more effective realisation of the constitutional right to housing, strengthen public trust in state institutions, and support Ukraine's broader recovery and reconstruction efforts.

1. «*AI will be either the best or the worst thing ever to happen to humanity*», S. Hawking, *AI will be either best or worst thing for humanity*, in *The Guardian*, 19 October 2016, <https://www.theguardian.com/science/2016/oct/19/stephen-hawking-ai-best-or-worst-thing-for-humanity-cambridge>.
2. N. Isayeva, *Socio-economic Rights of IDPs in Ukraine: Challenges and Solutions Amidst the War*, in *Deutsch-Polnische Juristen-Zeitschrift*, No. 1–2, 2025, pp. 29–35, <https://www.dpjz.de/wp-content/uploads/2025/04/DPJZ-1-2.pdf>.
3. International Organization for Migration (IOM), in *DTM Ukraine — Internal Displacement Report — General Population Survey Round 17* (August 2024), 4 September 2024, I O M, U k r a i n e, <https://dtm.iom.int/reports/ukraine-internal-displacement-report-general-population-survey-round-17-august-2024>.
4. No. 254K/96-BP (*Constitution of Ukraine*), 1996, <https://zakon.rada.gov.ua/laws/show/en/254%D0%BA/96-%D0%B2%D1%80>.
5. No. 1706-VII (*On ensuring the rights and freedoms of internally displaced persons*), 2014, <https://zakon.rada.gov.ua/laws/show/1706-18#Text>.

6. International Organization for Migration (IOM), in *DTM Ukraine – Protracted Displacement and Housing Insecurity* – October 2025, 12 December 2025, IOM, Ukraine, <https://dtm.iom.int/reports/ukraina-trivale-peremischennya-ta-nestabilne-zabezpechennya-zhitlom-zhovten-2025-roku?close=true>.
7. N. S. Isayeva, *Socio-Economic and Cultural Rights and Freedoms of Internally Displaced Persons and Guarantees of Their Implementation: Theoretical and Legal Characteristics (Dissertation)*, 2021, <https://dduvs.edu.ua/wp-content/uploads/files/Structure/science/rada/df023/d.pdf>.
8. Ibid. 8. See also source - No. 1706-VII (*On ensuring the rights and freedoms of internally displaced persons*), 2014, <https://zakon.rada.gov.ua/laws/show/1706-18#Text>.
9. No. 894-2025-II (*On certain issues of creating and operating an Information and Analytical System of real estate objects to provide housing for internally displaced persons – Order of the Cabinet of Ministers of Ukraine*), 2025, <https://zakon.rada.gov.ua/laws/show/894-2025-%D0%BF#Text>.
10. No. 5203-VI (*About Administrative Services*), 2017, <https://zakon.rada.gov.ua/laws/show/5203-17>.
11. No. 167-p/2001 (*On approval of the Concept for the development of digital competencies and approval of the action plan for its implementation – Order of the Cabinet of Ministers of Ukraine*), 2001, <https://zakon.rada.gov.ua/laws/show/167-2021-%D1%80#Text>.
12. *Ukraine launches new e-service for internally displaced persons*, in *UNDP Ukraine*, 22 April 2022, <https://www.undp.org/ukraine/press-releases/ukraine-launches-new-e-service-internally-displaced-persons>
13. No. 537-V (*On the basic principles of developing the information society in Ukraine for 2007–2015*), 2007, <https://zakon.rada.gov.ua/laws/show/537-16#Text>.
14. No. 1089-IX (*About electronic communications*), 2020, <https://zakon.rada.gov.ua/laws/show/1089-20#Text>
15. No. 2155-VIII (*On electronic identification and electronic trust services*), 2017, <https://zakon.rada.gov.ua/laws/show/2155-19#Text>.
16. No. 2297-VI (*About personal data protection*), 2010, <https://zakon.rada.gov.ua/laws/show/2297-17>.
17. Ibid. 13. See also source - No. 5203-VI (*About administrative services*), 2017, <https://zakon.rada.gov.ua/laws/show/5203-17>.
18. Ibid. 14. See also source 14 - No. 167-p/2001 (*On approval of the Concept for the development of digital competencies and approval of the action plan for its implementation – Order of the Cabinet of Ministers of Ukraine*), 2001, <https://zakon.rada.gov.ua/laws/show/167-2021-%D1%80#Text>.
19. O. Obolenskyi, V. Kosytska, A. Rvach, *Artificial intelligence in public administration: requirements, problems and risks*, in *Scientific Notes*, No. 33(4), 2023, pp. 121-137, <https://ir.kneu.edu.ua/server/api/core/bitstreams/8f740b46-5a48-41d8-b5e0-5e2f24eff3d2/content>.

20. Ibid 19. See also source 19 - No. 2297-VI (*About personal data protection*), 2010, <https://zakon.rada.gov.ua/laws/show/2297-17>.
21. European Parliament and Council of the European Union, (2016), *Regulation (EU) 2016/679 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)*, <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679>.
22. European Parliament and Council, (2024), *Regulation (EU) 2024/1689 on Artificial Intelligence (AI Act)*, in *Official Journal of the European Union*, L 1689/1.
23. Ibid. See also source 24 - European Parliament and Council of the European Union, (2016). *Regulation (EU) 2016/679 of the European Parliament and of the Council on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)*, <https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32016R0679>.
24. European Commission, Digital Europe Programme, *Aligning AI Policies with the EU Legal Framework. Guidelines for Associated Countries*, Brussels, 2023, <https://digital-strategy.ec.europa.eu/en/activities/digital-programme>.
25. European Commission, *AI4Government: European approach to artificial intelligence*, Brussels, <https://digital-strategy.ec.europa.eu/en/policies/european-approach-artificial-intelligence>.

AI Administration in Construction: Vienna's Innovative Solution

Radka Semancová

DOI: 10.13130/2723-9195/2026-2-57

Il contributo analizza il progetto "BRISE Vienna" quale modello di automazione algoritmica dei procedimenti di rilascio dei titoli edilizi, valutandone la conformità ai principi del diritto amministrativo e al quadro normativo europeo in materia di intelligenza artificiale. Contestualmente, esso esamina il grado di adeguatezza dell'ordinamento della Repubblica Slovacca rispetto a un'analoga trasformazione.

The paper examines the "BRISE Vienna project" as a paradigm for the algorithmic transformation of building permit proceedings and evaluates its adherence to the principles of administrative law and the European regulatory framework on artificial intelligence. Concurrently, it assesses the Slovak Republic's normative preparedness for a transformation of a similar nature.

Summary: 1. From an anthropocentric model to technology-supported decision-making.- 2. Building permit proceedings as a complex administrative process.- 3. The procedural model of BRISE Vienna.- 3.1 Preparation phase and preliminary automated review.- 3.2 Formal submission and expert assessment.- 3.3 Transparency, participation and augmented reality.- 3.4 Pilot phase and empirical results.- 4. Formalisation of legal norms and its limits.- 5. The procedural model of BRISE and its normative significance.- 6. Normative implications for the Slovak Republic.- 7. Conclusion.

1. From an anthropocentric model to technology-supported

decision-making

Administrative law has traditionally been conceived as an anthropocentric system of decision-making, in which the administrative authority is the bearer of discretionary power, evaluative judgment and individual assessment of the case (“in concreto”)^[1]. However, the digitalisation of public administration is gradually transforming this model. The question therefore arises to what extent procedural and normative structures may be “translated” into algorithmic language without weakening the principles of the rule of law.

Between 2018 and 2023, the City of Vienna implemented the BRISE Vienna project^[2] as a research and development initiative aimed at the digitalization and partial algorithmicizing of building permit proceedings. The project was co-funded by a European urban innovation program with approximately EUR 4,8 million and was implemented through interdisciplinary cooperation between the City of Vienna, the Vienna University of Technology, experts in digital engineering and the Chamber of Civil Engineers. It was tested on real construction cases within the administrative practice of the city. The BRISE Vienna project thus represents a significant experiment in the field of the algorithmicising of building permit proceedings. It is a system that integrates digital building modelling with automated verification of project compliance with legal norms. It does not constitute fully automated decision-making; rather, it represents a model of supportive automation that fundamentally alters the procedural dynamics of the proceedings.

2. Building permit proceedings as a complex administrative process

Building permit proceedings rank among the most complex forms of administrative procedure. They affect property rights, environmental protection, urban development and the public interest in regulated land use, while at the same time significantly influencing the level of building culture as an integral part of the cultural and social environment^[3].

The traditional model is based on documentation manually reviewed by the administrative authority. This system is time-consuming, administratively

burdensome and prone to formal deficiencies.

In the conditions of a rapidly growing Vienna, where approximately thirteen thousand building applications are submitted annually, the classical model proved to be institutionally unsustainable. The response was the BRISE project, which represents a transformation of the very procedural architecture of building permit proceedings.

3. The procedural model of BRISE Vienna

Unlike partial digitalisation initiatives, BRISE represents a comprehensively conceived model of public administration transformation that affects the very essence of permitting processes in the field of construction. The project was not designed merely as a technical solution for the electronic submission of applications, but as a research and development concept of intelligent, efficient and transparent public administration. Through its implementation, the City of Vienna has positioned itself among European leaders in digital administration.

BRISE (Building Regulations Information for Submission Evolvment) was implemented as an interdisciplinary project in cooperation between the City of Vienna, the Vienna University of Technology (civil engineers and computer scientists), experts in digital engineering, the Chamber of Civil Engineers, and other partners. The project received approximately EUR 4.8 million from a European initiative supporting innovative urban development and was successful in competition with 175 projects submitted by applicants from 23 Member States. The very selection of the project demonstrates its systemic and transferable character.

3.1 Preparation phase and preliminary automated review

The process begins with the creation of a digital three-dimensional building model using the Building Information Modelling (BIM) methodology. BIM represents a network-based planning method integrating architectural, structural and technical data into a coordinated model. Within the BRISE project, the internationally recognised software-neutral IFC standard is used, enabling interoperability and comparability of models.

The digital model contains precise data concerning building height, volume,

built-up area, setback distances, functional use and other parameters of the structure. The model is uploaded into the system, where a series of automated checks is carried out. These include verification of compliance with height limits, control of setbacks from neighbouring plots, verification of volumetric regulations and building density, comparison with binding elements of land-use planning documentation, and verification of formal requirements (such as signatures and completeness of documentation).

The system employs a combination of algorithmic checks, collision tests and intelligent searches of legal provisions. Artificial intelligence is applied primarily in the form of machine learning and semantic search, enabling the rapid assignment of relevant legal sources to a specific model.

The result is immediate feedback to the designer already at the preliminary review stage. This phase significantly increases planning certainty and reduces the number of procedural requests for supplementary documentation. It represents a fundamental shift from *ex post* review to preventive *ex ante* validation.

3.2 Formal submission and expert assessment

After the identified inconsistencies have been remedied, the application is submitted through the digital platform. The administrative authority has access to the complete digital model together with the outputs of the automated checks. All stakeholders may monitor the status of the proceedings in real time.

The administrative burden shifts from manual measurement and verification of basic parameters to substantive assessment of the project. Officials may focus on urbanistic, social, architectural and environmental aspects that require qualitative evaluation and the exercise of discretionary power.

BRISE therefore does not alter the subject of decision-making but rationalises its performance. Automation assumes precise, quantifiable tasks, while evaluative and interpretative elements remain within human competence.

3.3 Transparency, participation and augmented reality

A distinctive feature of the system is the use of Augmented Reality. Digital models may be placed directly into real space and displayed via mobile devices. The public may thus view the planned building in its future form even before

construction commences.

This tool strengthens transparency of the proceedings and promotes democratic participation. In this respect, BRISE exceeds the framework of purely administrative innovation and establishes a new model of communication between building authorities, the professional community and citizens.

3.4 Pilot phase and empirical results

The project was tested on thirteen real construction projects. Empirical results confirmed a significant reduction in the duration of proceedings – digital building permits were issued up to 50% faster than under the traditional paper-based model. At the same time, administrative burden, risk of errors and costs associated with repeated supplementation of documentation were reduced.

The project thus not only optimised processes but also created know-how transferable to other areas of public administration. BRISE is therefore understood as a pilot model of intelligent administration whose principles may be applied beyond construction law^[9].

4. Formalisation of legal norms and its limits

At the core of the system lies the formalisation of certain legal norms into algorithmic form (“norma in codice”). Such an approach enhances uniformity in the application of law and reduces human error. At the same time, it raises questions concerning the permissible scope of algorithmicising.

Legal norms frequently contain indeterminate legal concepts and evaluative elements. Their complete formalisation could lead to excessive rigidity and weaken discretionary power. The principle of individual assessment of the case must therefore remain preserved.

The regulatory framework of the European Union, represented by the Artificial Intelligence Act (Regulation (EU) 2024/1689), requires human oversight, transparency and accountability in the case of high-risk systems. The BRISE model, which preserves decision-making authority with the administrative body, complies with the requirement of human oversight^[9].

The Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law likewise emphasises the

protection of fundamental rights, proportionality and reviewability of decisions. Algorithmic assistance must therefore not result in the creation of a non-transparent “black box”^[6].

5. The procedural model of BRISE and its normative significance

BRISE transforms the very procedural architecture of building permit proceedings. Instead of traditional documentation, the foundation of the procedure becomes a digital three-dimensional building model containing precisely quantifiable data concerning its parameters. This model is subjected to automated review on the basis of formalised rules derived from the building regulations of the City of Vienna.

The process may be divided into three principal phases:

1. Preliminary automated review – the system identifies non-compliance with height, volume or setback regulations prior to the formal submission of the application.
2. Expert assessment by the administrative authority – the authority focuses on qualitative and evaluative aspects of the project.
3. Transparent communication with the public – digital visualisations enable better understanding of the impact of the construction on its surroundings.

Empirical results of the pilot phase indicated a significant acceleration of proceedings and a reduction in administrative burden. From a normative perspective, however, it is essential that decision-making authority remains vested in the administrative body.

The principle of legality (“*principium legalitatis*”) constitutes one of the fundamental principles of administrative law. Public authorities may act only on the basis of law and within its limits. The algorithmicising of part of building regulations into program code (“*norma in codice*”) represents a specific form of application of law.

A key question is whether the translation of a legal norm into an algorithm

preserves its content and purpose (“ratio legis”). Formalisation of a legal norm presupposes its unequivocal interpretation, whereas in practice interpretation is often the subject of professional debate. If an algorithm were to fix only a single interpretative alternative, this could weaken the principle of legality in its substantive sense. BRISE addresses this issue by automating only those parts of regulation that are objectively quantifiable. Evaluative and interpretative elements remain within the competence of the administrative authority^[7].

The principle of proportionality requires that an intervention by public authority be suitable, necessary and proportionate to the pursued objective. Automated systems operate deterministically, which may lead to excessive rigidity. Building permit proceedings frequently involve borderline situations where it is necessary to consider the specific circumstances of the case. Algorithmic application of rules without the possibility of individual correction could result in a breach of the principle of proportionality. In this respect, the BRISE model is a compromise. The algorithm performs precise checks, yet the final decision remains in human hands. This preserves the possibility of taking exceptional circumstances into account^[8].

The principle of equality before the law requires that similar cases be treated alike (“similia similibus”). Algorithmicising may paradoxically strengthen this principle, as it reduces the risk of subjective discrepancies between individual officials. On the other hand, if input data or system settings were to contain systematic biases, discriminatory effects could be reproduced. It is therefore essential to ensure transparency of algorithms and the possibility of their scrutiny^[9].

The principle of protection of the rights of participants in proceedings and the right to a fair process require that a decision be reviewable and reasoned. In the context of algorithmicising, the question of explainability of system outputs arises. The AI Act places emphasis on ensuring human oversight in the case of high-risk systems used in public administration. The Council of Europe Framework Convention on Artificial Intelligence likewise stresses the protection of fundamental rights, transparency and reviewability. The BRISE model, which preserves final decision-making authority within the administrative body, is compatible with these requirements. The algorithm provides supporting inputs, not the final decision.

The algorithm operates deterministically, whereas discretionary power enables consideration of the specific circumstances of the case. BRISE represents a model of complementarity, in which technology assists but does not replace human decision-making. The future of administrative law therefore does not lie in complete automation, but in seeking a balance between “efficientia” and “iustitia”.

6. Normative implications for the Slovak Republic

The BRISE Vienna project represents significant inspiration for the Slovak environment, but at the same time constitutes a test of the normative preparedness of the Slovak legal order. The introduction of a comparable system could contribute to the acceleration of building proceedings, reduction of formal deficiencies in submissions and increased predictability of decision-making. The question remains whether Slovak legal, technological and institutional conditions allow such a qualitative shift without weakening the fundamental principles of administrative law.

The new Slovak Building Act (Act No. 25/2025 Coll., Building Act and Amendments to Certain Acts) redefines the concept of a building and the process of its permitting. It defines a building as a construction created by building works from construction products, permanently connected to the ground or requiring ground modification, while emphasising its functional and spatial character. This definition is significant from the perspective of digitalisation, as it creates a unified concept of an object that could in the future be modelled and assessed in digital form.

The Slovak Republic has already undertaken the electronic submission of building applications through the Construction Portal. Applications and documentation are submitted electronically, primarily in PDF or DWG formats. However, this represents digitalisation of the information carrier rather than transformation of the procedural logic. Although documentation is submitted electronically, its assessment remains manual. The system does not include algorithmic verification of normative compliance nor automated application of regulatory provisions through artificial intelligence. Digitalisation in the Slovak context therefore currently represents formal modernisation rather than genuine

algorithmicising of decision-making^[10].

Algorithmicising of part of building proceedings would require a higher degree of normative determinacy. Rules suitable for automated control must be precise, quantifiable and unambiguous. Slovak construction law, however, contains numerous indeterminate legal concepts and evaluative criteria enabling individual assessment of the case (“in concreto”). Their formalisation would require methodological standardisation and clear separation of provisions suitable for algorithmic application from those that must remain within the discretionary power of the administrative authority. Otherwise, there would be a risk that the algorithm fixes a single interpretative alternative and indirectly modifies the content of the legal norm, thereby weakening the material understanding of the principle of legality (“*principium legalitatis*”).

At the same time, full protection of procedural rights must be ensured. If a supportive algorithmic system were introduced, its transparency and explainability would have to be guaranteed. The participant in the proceedings must be able to understand the reasons why the system identified non-compliance and must have access to an effective remedy. Under the European regulatory framework, such a system would likely qualify as high-risk, requiring human oversight, risk management and documentation of decision-making processes.

A model preserving final decision-making authority with the administrative body therefore appears normatively acceptable^[11].

Particular attention must also be paid to the principle of proportionality. Automated systems operate deterministically, which may result in excessive rigidity. Building proceedings often involve borderline situations requiring consideration of specific circumstances^[12].

If the algorithm were to replace individual assessment without the possibility of correction, proportionality could be violated and discretionary power restricted. Digitalisation must therefore not be perceived as an end in itself but as a tool for enhancing the quality, transparency and legal certainty of decision-making.

The Slovak Republic thus possesses the basic prerequisites for gradual transformation of building proceedings, yet the current state represents a phase of digitalisation rather than algorithmisation. Transition to a model comparable to BRISE would require systematic normative revision, methodological

standardisation and technological infrastructure enabling machine-readable models instead of static documents. Technological efficiency (“efficientia”) may constitute a legitimate objective of public administration, provided it is not achieved at the expense of substantive justice (“iustitia”) and the protection of procedural rights.

7. Conclusion

The BRISE Vienna project represents a significant example of how technological tools may be integrated into administrative proceedings without negating their anthropocentric foundation. It demonstrates that algorithmicising need not replace human decision-making but may constitute structured support enhancing efficiency, reducing administrative burdens and strengthening uniform application of law.

Analysis of the procedural model confirms that the key condition of normative acceptability lies in preserving discretionary power, ensuring transparency of algorithmic processes and respecting the principles of legality, proportionality, equality and protection of procedural rights. Formalisation of legal norms into “norma in codice” is permissible only insofar as it does not undermine the material understanding of legality nor create a non-transparent decision-making environment.

In the Slovak context, the current stage is characterised primarily by digitalisation of submission rather than transformation of procedural logic. The new Building Act establishes a more unified normative framework, including a definition of a building as a technical and functional entity, yet algorithmic assessment through artificial intelligence remains absent. The Slovak Republic therefore stands at the threshold of potential transformation, which will require careful normative preparation, methodological standardisation and institutional capacity-building. The future of administrative law does not lie in complete automation of decision-making, but in achieving equilibrium between “efficientia” and “iustitia”, between technological rationality and substantive justice. Only if algorithmicising remains subordinate to the principles of the rule of law and the protection of fundamental rights can it serve as a legitimate instrument of public administration. It is within this tension between

innovation and normative continuity that the further development of building permit proceedings in the era of artificial intelligence will unfold.

1. This article was produced within the framework of the research project APVV-24-0063 Protection of Architectural Heritage: Innovative Approaches and Solutions.
2. See BRISE-VIENNA - Planning and living together - Digitales Wien.
3. F. Branický, *Dialogue Between Tradition and Innovation in the Rural Context*, in. *EUROSTAV*, 5, 2025 and B. Cepek, M. Maslen, *Substantive Administrative Law*, Wolters Kluwer, Bratislava, 2018.
4. BRISE-VIENNA - Planning and living together - Digitales Wien.
5. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act).
6. Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law dated Vilnius, 5.IX.2024.
7. B. Cepek, B. Maslen, *Substantive Administrative Law*, *cit.*
8. J. Boguszak, *Theory of law*, ASPI Publishing, 2004, 2. Edition, 2004.
9. J. Tekeli, *Substantive Administrative Law – Special Part*, šafárikPress, Košice, 2022.
10. Act No. 25/2025 Coll., Building Act and Amendments to Certain Acts.
11. M. Mesarčík, Z. Gyurász, *Law and AI*, Faculty of Law Comenius University in Bratislava, Bratislava, 1. Edition, 2024.
12. S. Košičiarová, *Substantive Administrative Law*, Vydavatelství a nakladatelství Aleš Čeněk, Plzeň, 2022.