International Scientific Conference

"Towards a Better Future: Justice, Development, and Artificial Intelligence"

Conference Proceedings

Faculty of Law - Kicevo

Kicevo, 2025

International Scientific Conference

"TOWARDS A BETTER FUTURE:
JUSTICE, DEVELOPMENT, AND ARTIFICIAL INTELLIGENCE"

Bitola, 02.10-03.10.2025

Conference Proceedings

- *The editors or publisher are not responsible for subsequent uses of the work. It is the author's responsibility to bring an infringement action if so desired by the author.*
- *Any views expressed in this publication are the views of the authors and are not necessarily the views of the editors or publisher.*

International Programme Committee

Viorel Paska, PhD, West University of Timisoara, Romania
Lucian Bercea, PhD, West University of Timisoara, Romania
Laura Stanila, PhD, West University of Timisoara, Romania
Zoran Pavlovic, PhD, University of Novi Sad, Serbia
Goran Ilic, PhD, University of Belgrade, Serbia
Istvan Gal, PhD, University of Pecs, Hungary
Nedzad Koraljic, PhD, University of Sarajevo, BiH
Tanja Kesic, PhD, University of Criminal Investigation and Police Studies, Belgrade, Serbia
Milan Zarkovic, PhD University of Criminal Investigation and Police Studies, Belgrade, Serbia
Milica Kolakovic, PhD, Institute of criminological and sociological research, Belgrade, Serbia
Alba Kreka,PhD, Faculty of Education and Philology, Albania
Sergiu Popovici, PhD, West University of Timisoara, Romania
Florentina Mutiu, PhD, West University of Timisoara, Romania
Judithanne Scourfield McLauchlan, PhD, The University of South Florida St. Petersburg, USA

Organizing Committee

Svetlana Veljanovska, PhD – Chairwoman
Angelina Stanojoska, PhD – Vice - Chairwoman
Goran Ilik, PhD
Ice Ilijevski, PhD
Ivona Shushak Lozanovska, PhD
Elena Tilovska - Kechegi, PhD
Elena Temelkovska - Anevska, PhD
Elizabeta Tosheva, PhD
Mirjana Ristovska, PhD
Mladen Karadzoski, PhD
Pakiza Tufekci, PhD
Blagoj Nenovski, PhD
Andrea Zdravkovska, PhD

Goce Veleski, LLM - Secretary

PREFACE

Dear colleagues, esteemed professors and researchers, dear doctoral students, distinguished guests,

Allow me, in my as Dean of the Faculty of Law and organizer of this conference, to extend my warmest welcome to you.

Today, all of us present here are united by an extremely important and topical topic, which is increasingly taking its place in scientific and professional debates around the world – the relationship between justice, sustainable development and artificial intelligence.

The mission of this conference is to contribute to a thorough, critical and interdisciplinary re-examination of the role of artificial intelligence in contemporary societies, especially in the context of legal systems, social equality, institutional transparency and sustainable economic growth.

In an era when algorithms are becoming new actors in decision-making – in both the public and private sectors – issues of fairness, accountability, transparency and human rights are becoming essential and their appropriate treatment in science is also needed.

Therefore, our scientific community must not be a mere observer of technological progress, but an active voice that will direct its application towards the common good.

Distinguished attendees,

Law and technology have never been isolated worlds. On the contrary, each new technological stage reshapes the legal framework – requiring new definitions, new institutional mechanisms and above all – new ethical and normative positions.

Artificial intelligence today challenges us to think:

- Can digitalization democratize justice?

- How to deal with algorithmic discrimination and inequality in access to technologies?

- And most importantly, how do we build regulations that will not hinder innovation, but will direct it towards inclusive and sustainable development?

The topics that will be addressed at this conference – from legal and ethical implications, through issues of responsibility and control, to the role of AI in education, economics and public administration – represent areas in which science, law and practice must cooperate, without exception.

I am particularly pleased that among the participants we also have young researchers, doctoral candidates and future academic leaders. It is their energy, critical thinking and scientific curiosity that guarantee that the academic community will continue to play a key role in shaping responsible and ethically oriented technology policies.

Ladies and gentlemen,

This conference is not just a scientific event – it is a platform for building a vision. A vision for a future in which progress is not an end, but a means to achieve a better, fairer and more humane society.

In that regard, allow me to express my gratitude to all who contributed to the realization of this conference – the organizing committee, the panelists and reviewers, as well as to all of you who took the time to be part of this event.

I wish you successful work, constructive discussions and, above all, inspiration for new scientific and professional feats.

Thank you.

<div align="right">

Prof. Dr.sc. Svetlana Veljanovska
Dean of the Faculty of Law –Kicevo

</div>

Kicevo, 2025

TABLE OF CONTENTS

12

PREDICTING RISK SITUATIONS IN FRONTLINE POLICING USING BODY-WORN CAMERAS

JUDr. Jana Zachar Kuchtová, PhD.
prof. JUDr. Jozef Meteňko, PhD.

Department of Criminalistics and Forensic Sciences, Academy of the Police Force in Bratislava
jana.kuchtova@akademiapz.sk, jozef.metenko@akademiapz.sk

Abstract

Artificial intelligence (AI) and multimodal analysis of body-worn camera (BWC) data represent a promising tool for assessing and predicting risk situations in policing. While current research (e.g., Srbinovska et al., 2025) demonstrates that audio-visual recordings can be used to extract behavioral cues of escalation or de-escalation, this paper shifts the focus to the prediction of incidents based on a combination of such signals. Due to the unavailability of operational data, a synthetic dataset ($N = 96$) was generated using a large language model (GPT-5, OpenAI). The number of cases was deliberately designed to reflect a model 24-hour shift with eight body-worn cameras and an average of twelve recorded interactions per camera, corresponding to a realistic operational scope. The dataset included multimodal variables (overlapping speech, noise level, number of speakers, transcript quality), known risk signals (shouting, distance <1 m, refusal to comply, profanity, visible object, physical touch), and a binary incident outcome. The analysis showed that risk assessment is highly sensitive to the threshold applied. At a low threshold of 50, recall was maximized but false alarms increased; at a high threshold of 70, precision improved but many incidents were missing; and at an intermediate threshold of 60, the most balanced results were achieved. The findings support the use of AI-driven

simulations and synthetic data as methodological and pedagogical tools in criminology and criminalistics. This paper is presented as a pilot simulation study, demonstrating the potential of predicting risk behavior from BWCs even in environments with limited access to real-world data.

Keywords: Artificial Intelligence, Body-Worn Cameras, Risk Prediction, Simulation, Synthetic Data, Policing

Introduction

In the past years, smart machines has increasingly come into the area of keeping law and internal safety. What just ten years ago looked like a test is now used in many areas: guessing crime places, automatic license plate reading, face ID tech, and looking at voice or text info. These uses show both the chances and issues of smart machines in police work from better speed and trust to worries about privacy and unfairness. With this setup tools that use smart machines are joining the larger talks on how police work can change fast with new tech while still staying right and trusted.

Body-worn cameras are one of the most popular and well-known technologies (BWCs). They were first implemented mainly to increase openness and public confidence, but they quickly established themselves as a common tool for recording interventions, shielding officers from unfounded accusations, and handling complaints. BWCs are now regarded as a crucial part of evidence management in many jurisdictions, giving prosecutors and supervisors access to unbiased video evidence. Crucially, a BWC recording is multimodal by nature, comprising speech, sound, and visual components. BWCs are a special tool for researching police-citizen interactions in a way that is more comprehensive than traditional reports or survey data because of their multimodality.

This potential has been the subject of scholarly investigation. According to Srbinovska et al. (2025), BWC recordings can be examined to find behavioral indicators of escalation or, on the other hand, de-escalation. Lum and Koper (2017) reviewed the wider implications of predictive analytics for policing, while White and Malm (2020) investigated the technical potential of automated AI tools for assessing officer behavior in

real time. Although these studies highlight significant methodological advancements, they all have one thing in common: the majority concentrate on explaining what happens during an interaction or examining circumstances that have already happened. The question of whether multimodal signals can be used to anticipate, beforehand, that a particular situation may escalate into an incident has received less attention.

A major research gap is represented by this lack of a predictive focus. Most current methods are retroactive and intended to categorize or explain previous behavior. The potential for creating models that warn of danger in progress and might enable officers or supervisors to step in before it escalates is still not well understood. In police education and training, where case studies and simulations are frequently employed but frequently do not incorporate AI-based risk assessment in a systematic manner, this prognostic viewpoint is especially pertinent.

By presenting a pilot simulation study, the current paper aims to close this gap. Our goal is to determine whether certain signals, like shouting, noncompliance, or physical contact, can be used as indicators of incident risk rather than offering a thorough descriptive description of behavior in BWC recordings. We rely on a synthetic dataset created using results from the body of existing literature because access to operational recordings is limited. With eight cameras and an average of twelve interactions per device, this dataset simulates 96 cases, or roughly one shift. We show how a predictive methodology can be developed even without real-world data by examining these simulated interactions.

The goal is simple but clear: to show that AI-driven simulations can be used as a method and a way to teach, and to show that they can be used to predict risk from BWCs. By shifting the focus from retrospective description to prospective prediction, the paper adds to the growing body of work on AI in policing. It also shows how AI-based methods can be used in police training, which connects new ideas in data science with the real-world needs of professional education and field readiness.

Theoretical Framework

Body-worn cameras (BWCs) are one of the most noticeable new technologies in modern policing. People wanted cameras to be used because they wanted more transparency, fairness, and accountability. They thought that having a camera would stop people from doing bad things and give reliable evidence in case of a disagreement. Studies have shown that BWCs can cut down on complaints from the public, make people think better of the police, and provide useful evidence for internal investigations and court cases (Lum, Koper 2017). Debates and controversies also surround the use of BWCs. The biggest obstacles are not just related to privacy concerns or technical issues like when to turn on cameras, but more importantly, how cameras affect how police officers carry out their daily duties. Many officers believe that their superiors use BWCs as a tool for overzealous supervision. Because they are aware that every action is documented and may be reviewed later, they are hesitant to take decisive action. Constantly feeling like you're being watched puts more strain on your mind and can make you hesitate when you need to act quickly. BWCs continue to produce an unparalleled amount of multimodal data—the spoken, heard, and visual components of real-time interactions—despite these worries.

The rise of artificial intelligence (AI) has created new ways to process this kind of data. Traditional studies depended on surveys, observational reports, or crime statistics. BWCs, on the other hand, give us a constant and detailed record of how people interact with each other. Recent developments in computer vision, natural language processing, and acoustic analysis enable the automated identification of signals, including overlapping speech, elevated voices, insults, gestures, or abrupt physical movements. Recent reviews have also stressed the potential of combining camera footage with other AI-supported technologies, like drones, to make hybrid safety frameworks in policing (Davies 2023). These characteristics can be methodically extracted and categorized to differentiate between escalating and de-escalating behavior, as Srbinovska et al. (2025) showed. The importance of multimodal analysis is highlighted by their findings, which show that interactional dynamics leading to or away from conflict are revealed by the combination of sound, language, and visual cues rather than just what was said or what occurred.

Several behavioral cues have been repeatedly linked to increased risk in criminological studies. Incidents that turn into confrontations frequently involve physical contact, shouting, disobedience to orders, and the use of foul language. Contextual elements like loud background noise or multiple speakers make communication even more difficult and increase the possibility of misunderstandings. In their assessment of automated BWC review technology, White and Malm (2020) emphasized the potential of AI tools to identify such patterns at scale, considering that it is not practical to manually review thousands of hours of footage. However, it is crucial to stress that these signals are not deterministic; they do not automatically predict violence; rather, they raise the likelihood of escalation. Their probabilistic nature, which reflects how people naturally weigh risk in unpredictable situations, is exactly what makes them valuable.

Despite these realizations, descriptive approaches have dominated the literature up to this point. Most studies use BWC data to examine events after they have happened or to give supervisors resources for reviewing things after the fact. Predictive analytics has primarily been used at the macro level, such as predicting crime hotspots or trends, but its application to the micro level of interpersonal interactions is still in its infancy, according to Lum and Koper (2017). This results in a substantial research void. The signals that accompany escalation are well understood, but it is much less clear whether these signals taken together can serve as the foundation for a prognostic model that predicts incidents before they happen. Stated differently, predicting conflict is just as difficult as describing it.

Beyond operational policing, this transition from description to prediction has ramifications. Simulation exercises and scenario-based learning are essential tools used in police academies and professional training to equip recruits with the ability to make quick decisions under duress. By incorporating AI-generated risk assessments into these exercises, trainees may be able to practice assessing the likelihood of an incident in real time and identify escalation cues more methodically. Multimodal data, according to Srbinovska et al. (2025), is a pedagogical and research resource that provides an organized method of teaching situational awareness. Thus, predictive models could have two purposes: first, they could help with

operational decision-making; second, they could improve training by giving cadets a framework for analysing behavior.

When combined, the theoretical underpinnings of this research integrate three academic streams. The first focuses on BWCs as tools for transparency and accountability; the second examines AI and multimodal analysis of BWC data; and the third utilizes criminological and criminalistics insights regarding behavioral cues associated with escalation. Although each of these strands offers valuable insights, none of them completely answers the question of how multimodal signals could be applied to prediction. By combining these viewpoints, the current study puts itself at the nexus of current discussions and advances the investigation of the predictive potential of data from body-worn cameras.

Methodology

The purpose of this paper, which is designed as a pilot simulation study, is to demonstrate how signals frequently seen in body-worn camera (BWC) recordings can be incorporated into a predictive framework. The analysis was based on a synthetic dataset because actual operational policing recordings were not available for ethical and practical reasons. To approximate the scope of a 24-hour shift with multiple BWCs in use, the dataset was created. Eight cameras in total were assumed, which is the number of devices that Axon gave the author for scientific purposes. A total of ninety-six simulated cases were produced by modelling each camera to record roughly twelve encounters during a shift. This yields ninety-six simulated cases, a quantity that was selected to be feasible for exploratory analysis and realistic in terms of workload. Although structured prompts in ChatGPT (OpenAI, GPT-5) were used to generate the dataset, the variables' design was based on empirical results published in the literature.

Two categories of information served as the foundation for the organization of the simulated data. Initially, multimodal characteristics that are common in BWC recordings were incorporated, including background noise, number of speakers, overlapping speech, and transcript quality. The technical and communication difficulties found in analyses of BWC data are reflected in these variables (White and Malm 2020). Second, the dataset

included behavioral indicators that are consistently linked to escalation in policing and criminology and criminalistics research.

These included yelling, disobedience, using foul language, being too close (less than one meter), having objects in plain sight that could be interpreted as dangerous, and making direct physical contact. Such signals can be systematically coded and interpreted as indicators of increasing or decreasing risk, as stressed by Srbinovska et al. (2025). Each of these cues had a binary code in the dataset. Incidents were defined probabilistically as the cumulative result of multiple signals occurring simultaneously, rather than treating any one signal as decisive. This is in line with the well-established criminological theory that dangerous interactions are more often the result of a confluence of stressors than of a single action.

As a result, each case was given a binary outcome variable that indicated whether an incident happened. The assignment used a weighted logic: escalation was marginally more likely when one or two small signals were present, but an incident outcome was significantly more likely when three or more signals were present. This method provides a conceptual approximation of how risk tends to accumulate and how officers intuitively assess danger in practice, but it makes no claims to accurately depict real-world distributions.

The risk signals were combined into a score between 0 and 100 to test how these cues might be translated into predictions. Various classification thresholds were applied based on this score. A sensitive threshold, which was intended to capture most incidents at the expense of more false alarms, was represented by a score greater than fifty. Stricter classifications were obtained by raising the threshold to sixty or seventy, which decreased false positives but let more incidents pass unnoticed. Confusion matrices and derived measures of precision, recall, and F1 score were used to evaluate the trade-off between sensitivity and specificity. The goal was to highlight the problems that predictive approaches invariably encounter, not to maximize predictive accuracy.

Even though the dataset is artificial, the reasoning behind its creation is grounded in empirical research and accurately captures the dynamics of encounters that were captured by BWC. Because of this, the approach should

be viewed as a proof of concept - a means of testing predictive modelling in a setting that minimizes privacy concerns while maintaining pedagogical relevance. In addition to its exploratory value, the method shows how training environments can make use of synthetic data. Recruits can practice identifying how various signals combine and how thresholds change interpretation by incorporating risk estimates into simulated scenarios or classroom exercises. This is how the methodology links the practical goal of preparing officers for the complexity of real-world encounters with the technical question of prediction.

The Synthetic Dataset's Construction

Eight body-worn cameras (BWCs) were used to record the synthetic cases, which were created to mimic the workload of a single 24-hour shift. Approximately twelve interactions were recorded during the shift by each camera (K1–K8), which was modelled as an independent unit. This led to a total of 96 cases. To represent actual variations in field activity, the distribution was not exactly uniform across cameras.

Contextual variables were used to further characterize each interaction. There were two fundamental descriptors: the environment (street, interior, apartment, or administrative space) and the type of contact (administrative, preventive, routine control, or conflictual). The weighting of the probabilities meant that most encounters took place on the street or in general interior spaces, with offices and apartments showing up less frequently.

Variables were introduced at the multimodal level to replicate the technical aspects of BWC video. A scale from 0 to 1 was used to model background noise; higher values were typical of outdoor environments, lower values were typical of offices, and medium values were typical of apartments or other interior spaces. In noisy or conflictual settings, there was a higher likelihood of overlapping speech and speaker count. By mimicking the known limitations of automated transcription, the transcript quality variable (0–1) was indirectly determined to decrease with increasing noise and speaker count.

Seven risk signals—yelling, overlapping speech, proximity (less than one meter), noncompliance, profanity, visible object, and physical contact—that are frequently discussed in criminological and policing research comprised the core of the dataset. Although each was binary coded, context affected how likely it was to occur. For example, yelling or using foul language were associated with noisy settings, whereas physical contact or extremely proximity were more prevalent in confrontational interactions. Although they were rarely created, visible objects significantly raised the likelihood of escalation.

The likelihood of an incident happening was calculated using a weighted combination of these signals and multimodal factors. Three or more signals together significantly increased the likelihood of escalation, while one or two weaker signals only slightly increased it. Stronger cues, such as proximity, rejection, or physical contact, were assigned higher weights. To determine the estimated level of risk before and during the encounter, each case was given a binary outcome variable ("incident" vs. "non-incident") as well as an auxiliary risk score (0–100).

To help visualize the dataset's structure, Table 1 provides an illustrative extract of several cases. The passage demonstrates the integration of contextual descriptors (such as environment, type of contact, number of persons, duration), behavioral signals (such as yelling, profanity, and physical contact), and multimodal features (such as background noise, overlapping speech, number of speakers, and transcript quality) into a single row. These inputs produce a binary incident outcome when combined with preliminary and in-situ risk predictions. For the sake of clarity, only the most important variables are displayed here. The complete dataset contained additional escalation signals (like rejection, proximity, and visible objects) that were not covered in this excerpt but were taken into account during the analyses.

Table 1 Extract from the synthetic dataset (sample of selected rows and columns from the generated dataset)

Case ID	Noise (0–1)	Overlapping Speech	Transcript Quality (0–1)	Profanity	Physical Contact	Incident
C001	0.746	0	0.739	1	0	0
C002	0.291	0	0.948	0	0	0
C003	0.404	0	0.859	0	0	0
C004	0.650	0	0.772	0	0	0
C005	0.714	1	0.550	0	1	1

Source: Authors' owns compilation based on synthetic dataset

The dataset faithfully reflected the patterns of ambiguity and variability present in real-world police-citizen interactions, despite its artificial nature. It can be used to test predictive methods and demonstrate multimodal features in a secure environment due to its tiered structure (camera → context → multimodal variables → risk signals → incident outcome). Exploratory analyses with visualizations of specific variables and predictive performance metrics are shown in the section that follows.

Results

The simulated dataset consisted of ninety-six cases, designed to simulate a full 24-hour shift with eight body-worn cameras (BWCs) in use. The incident rate was roughly 29% of these interactions, with 68 classified as non-incidents and 28 as incidents. Given that incidents account for a smaller portion of recorded interactions, this distribution is consistent with the assumption that most encounters are routine. The average number of risk signals in incident cases (3.9) compared to non-incident cases (1.5) showed a strong correlation between escalation and the accumulation of multiple cues. A correlation analysis revealed a moderately strong positive relationship ($r = 0.64$) between the number of signals present and the probability of an incident outcome.

Table 2 summarizes the distribution of cases by number of risk signals and outcome. A clear pattern emerges incident cases mostly occur when three or more signals are present, whereas non-incident encounters tend

to cluster around zero to two signals. The cumulative logic of escalation, which holds that the probability of conflict rises as signals mount rather than being triggered by a single factor, is best illustrated by this descriptive finding.

Table 2 Distribution of signals in incident and non-incident cases

Number of signals	Non-incident	Incident
0	18	0
1	17	1
2	20	2
3	8	9
4	3	7
5	2	6
6	0	3

Source: Authors' owns compilation based on synthetic dataset

Visualization of Selected Variables for Exploration

Two variables were chosen as examples to show how the synthetic dataset could be used for exploratory analysis. These visualizations are meant to show how individual variables can be methodically analysed and interpreted, not to give a comprehensive description of every feature of the dataset.

A boxplot of the variable Noise is shown in Figure 1. As anticipated, background noise levels were highest in outdoor settings and lowest in office interactions. In between were apartments and other indoor areas. This pattern highlights how environmental factors can impact communication and, indirectly, risk assessment, indicating that the dataset included realistic contextual variation.

Figure 1 Distribution of Noise (boxplot)
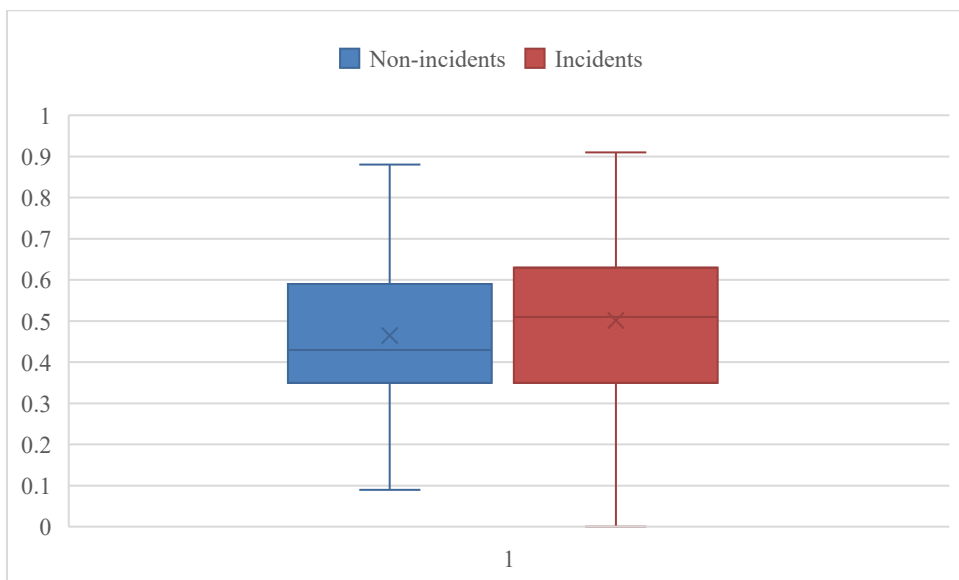
The percentage of profanity in incident and non-incident cases is displayed in Figure 2. Overall, profanity was rare, but it was more common in incidents (14%) than in encounters that did not involve an incident (9%). This implies that verbal aggression tends to cluster with escalation, despite being relatively uncommon.

Figure 2 Occurrence of Profanity in Incident vs. Non-incident Cases (bar chart)

24
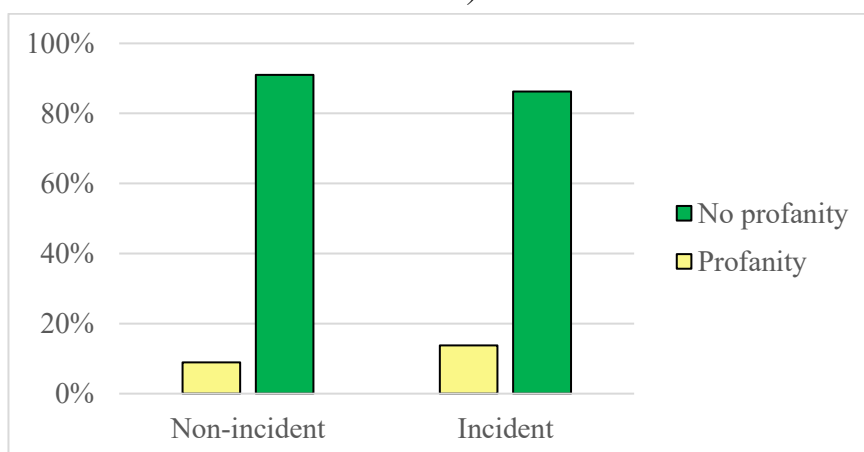
After combining the signals to create a composite risk score (0–100), predictive performance was evaluated using thresholds of 50, 60, and 70. Table 3 displays the confusion matrices and performance metrics that resulted. The model achieved high recall (0.89) at the lowest threshold of 50, correctly identifying most incident cases; however, due to a higher number of false alarms, precision was lower (0.66). Precision increased significantly (0.82) at the highest threshold of 70, but recall decreased to 0.32, indicating that many incidents were overlooked. A compromise was provided by the intermediate threshold of 60, which resulted in an F1 score of 0.65 with recall dropping to 0.57 and precision increasing to 0.76.

Table 3 Predicted vs. actual incidents under different thresholds (N = 96)

Threshold	True Positives	False Positives	True Negatives	False Negatives	Precision	Recall	F1 Score	Specificity
50 (≥3 signals)	25	13	55	3	0.658	0.893	0.758	0.809
60 (≥4 signals)	16	5	63	12	0.762	0.571	0.653	0.926
70 (≥5 signals)	9	2	66	19	0.818	0.321	0.462	0.971

Source: Authors' owns compilation based on synthetic dataset

Figure 3 illustrates these trade-offs by charting F1 score, precision, and recall across thresholds. The figure makes it evident that precision and recall are inversely related: a stricter model misclassifies fewer non-incidents but also misses more incidents. Each cutoff represents a unique balance between over- and under-prediction, as indicated by the F1 score's stability across thresholds.

Figure 3 Performance at Different Thresholds
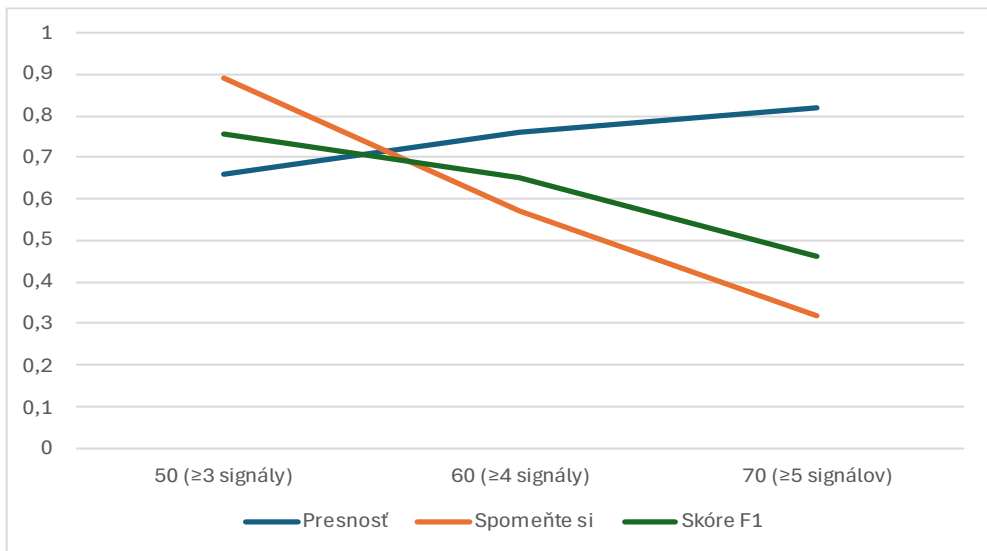
Source: Authors' owns compilation based on synthetic dataset

When combined, these findings demonstrate that the simulated data exhibit behaviors that are in line with theoretical predictions found in the literature. The accumulation of known escalation signals correlates with incidents, which are not random. The threshold used has a significant impact on predictive performance, and different thresholds draw attention to different facets of the risk assessment conundrum. Despite being synthetic, the dataset's internal logic reflects empirical findings and shows that using AI-driven simulation for methodological experimentation and pedagogical training is feasible.

Discussion

The goal of the current study was to investigate how artificial intelligence, and synthetic data might be used to forecast escalation risks that are visible in body-worn camera (BWC) recordings. The study, which was intended to be a pilot simulation, was purposefully small in scope but ambitious in scope; its goal was to show how risk signals that are frequently discussed in the literature into a predictive framework relevant for both criminology and criminalistics. The study offered an open and educationally

beneficial setting for testing methodological concepts by creating a synthetic dataset of ninety-six simulated encounters that represented the workload of a single 24-hour shift with eight BWCs operating.

Several significant conclusions were reached. The analysis first confirmed that escalation is cumulative. Incidents were concentrated in encounters where multiple signals occurred simultaneously rather than being distributed evenly among cases. The number of signals and the incident outcome had a somewhat positive correlation (r = 0.64), with incident cases usually averaging close to four signals, compared to about one and a half in non-incident cases. Second, the predictive tests illustrated the importance of threshold selection. A threshold of 50 caught nearly all incidents at the expense of many false positives, whereas a threshold of 70 provided noticeably more precision but missed many incidents. An intermediate threshold of 60 produced the most balanced result, with a precision of roughly 0.76 and a recall of roughly 0.57. This demonstrates the inevitable trade-off between sensitivity and specificity in predictive policing: a threshold must be chosen that considers both normative and technical considerations, and it is impossible to maximize both at the same time.

The study makes two more general contributions in addition to the quantitative findings. In terms of methodology, it shows that it is possible to investigate predictive modelling in criminology and criminalistics using synthetic data, generated using a large language model (GPT-5, OpenAI). While capturing realistic dynamics of escalation, synthetic data circumvents the ethical and privacy issues that come with real-world BWC footage. It makes controlled experimentation, simple replication, and transparent design possible. The method provides a useful pedagogical tool for police training and for the applied field of criminalistics, where practical methods for risk assessment are essential. Instructors can help recruits consider how risk assessments change with signal accumulation and threshold selection by exposing them to simulated encounters with different combinations of risk signals. Officers are better equipped to make quick but thoughtful decisions in the field thanks to these exercises, which also promote analytical thinking.

However, limitations must be recognized. To concentrate on a limited number of signals found in earlier studies, the dataset was purposefully made

simpler. Numerous other environmental, situational, and relational factors that were not covered here also have an impact on actual BWC encounters. Systematic reviews of BWC research (Wilson et al., 2019) similarly emphasize that escalation dynamics are influenced by a wide range of contextual and relational factors, many of which could not be represented in this simplified simulation. There were few cases overall, and the signal weighting was heuristic rather than empirically verified. These factors make it inappropriate to consider the reported metrics to be a reliable indicator of field performance. Rather than being a claim about actual results, they should be viewed as a proof of concept, an example of methodological possibilities.

Afzal (2024) highlights that for new forms of data to produce long-lasting results, they must be integrated into institutional frameworks that ensure accountability and transparency. This limitation extends beyond methodological concerns and concerns the governance of data in policing. The human aspect of predictive tools needs to be considered in addition to governance. According to recent experimental data, officers' perceptions of organizational fairness and transparency can be greatly influenced by automation in the BWC footage review process (Adams 2025). These results imply that predictive frameworks ought to be assessed based on their influence on professional legitimacy and trust as well as their statistical performance.

Several avenues for future study become clear when looking ahead. One is the addition of more thorough multimodal data to enhance risk models, such as audio, video, and contextual metadata. The use of real-world datasets to empirically verify synthetic results while abiding by the relevant ethical standards is another. It's interesting to note that related predictive ideas are already being developed in other domains, like autonomous driving, where multimodal sensory and communication data are integrated to predict potentially dangerous circumstances (Almutairi et al., 2025). Learning from these areas could help improve thresholding techniques, which this study found to be essential to predictive performance in policing scenarios, as well as signal selection and system integration. Finally, it is necessary to critically examine the normative implications of predictive thresholds. Algorithms shouldn't be the only ones making these decisions; different cutoffs favor

different institutional priorities. By including researchers, practitioners, and educators in these discussions, predictive modeling with BWCs can progress from simulation to practical application. In conclusion, this pilot study demonstrates that artificial intelligence and synthetic data offer a promising path forward for advancing criminological research and police training. The richness of real BWC footage cannot be entirely replaced by synthetic datasets, although they do provide a flexible and secure setting for testing and education. By connecting methodological research with pedagogical relevance, it shows that even small-scale simulations can offer insights into how risk accumulates, how prediction functions, and how officers might be better equipped for the complexities of real-world encounters. Ultimately, while AI-based prediction can provide useful support, it cannot replace the expertise of human officers; in the context of criminalistics, these tools should be seen as complementary aids that enhance, rather than substitute, professional judgment. Similar predictive methods are already being studied outside of law enforcement in domains such as healthcare and autonomous driving, suggesting that multimodal signal analysis has more potential if it is positioned as a tool rather than a replacement for human judgment.

Conclusion

This study set out to investigate how artificial intelligence and synthetic data might be used to forecast risk scenarios that were recorded by body-worn cameras (BWCs). The research was designed as a pilot simulation study because operational recordings could not be used for practical and ethical reasons. Ninety-six simulated cases, or the workload of a single 24-hour shift with eight cameras, made up the dataset. It contained behavioral indicators associated with escalation (shouting, profanity, refusal to comply, close physical distance, visible object, and physical touch) as well as multimodal features (overlapping speech, noise level, transcript quality, and number of speakers). A risk scoring model was developed and assessed using thresholds of 50, 60, and 70 based on this information. The careful selection and weighting of these signals will be critical to the future development of AI-based prediction in policing, as it determines whether the model accurately and meaningfully captures escalation dynamics.

The results verified that the accumulation of several signals is closely related to escalation. Most non-incident cases displayed zero to two signals, whereas incident cases clustered around three or more. Model performance was greatly impacted by threshold selection: a threshold of 50 maximized recall but resulted in many false positives; a threshold of 70 increased precision but missed several incidents; and a threshold of 60 produced the most balanced result. These results also highlight the fact that predictive models cannot replace human judgment; rather, they can only be used as a tool to assist officers in the field in making decisions.

This study's proof-of-concept nature is its contribution. It proved that synthetic data can be used as a teaching tool in police education as well as a methodological tool for testing predictive approaches in criminology and criminalistics. Such simulations enable the safe and ethical modelling of dangerous situations, assisting recruits in considering how signals mix and how their meaning varies based on the threshold used. Future studies should confirm these results using actual datasets, broaden the variables to incorporate more complex multimodal inputs, and investigate the moral ramifications of predictive policing in greater detail. The future success of AI prediction in policing will largely depend on correctly identifying and weighing the most relevant signals, since these are crucial for reliable functioning of such models. From a teaching standpoint, simulation-based techniques offer a viable way to enhance instruction and fortify officers' capacity to react swiftly but carefully in challenging real-world circumstances.

It's also critical to recognize that predictive policing applications are subject to stringent restrictions under the European Union's new regulatory framework, particularly the AI Act (European Commission, 2024). AI systems that forecast a person's likelihood of engaging in criminal activity based on personal characteristics or profiling are expressly forbidden by the Act. However, our research shows that AI-driven simulation can continue to be a useful methodological and instructional tool when it is centred on situational risk signals and cumulative interactional cues rather than individual traits. Therefore, the potential of such models to support structured

risk assessment and training in criminology and policing should not be disregarded, even within regulatory constraints.

References

Adams, Ian T. 2025. "Automation and Artificial Intelligence in Police Body-Worn Cameras: Experimental Evidence of Impact on Perceptions of Fairness among Officers." *Journal of Criminal Justice* 97: 102373. https://doi.org/10.1016/j.jcrimjus.2025.102373.

Afzal, Muhammad. 2024. "Data in Policing: An Integrative Review." *International Journal of Public Administration*. https://doi.org/10.1080/01900692.2024.2360586.

Almutairi, Nasser, Ali A. Alwan, Abdullah S. Almogren, Mohammed Alenezi, and Imran Razzak. 2025. "Real-Time Accident Prevention Model for Autonomous Vehicle Networks Using Sensor Fusion and Attention Mechanisms." *Scientific Reports* 15 (1): 4867. https://doi.org/10.1038/s41598-025-04867-8.

Davies, Alun. 2023. "Integrating Body-Worn Cameras, Drones, and Artificial Intelligence: A Framework for Hybrid Safety Technologies in Policing." *Policing: A Journal of Policy and Practice*. https://doi.org/10.1093/police/paad083.

European Commission. 2024. *Regulatory framework for Artificial Intelligence (AI)*. Accessed September 15, 2025. https://digital-strategy.ec.europa.eu/sk/policies/regulatory-framework-ai

Lum, Cynthia, and Christopher S. Koper. 2017. *Evidence-Based Policing: Translating Research into Practice*. Oxford: Oxford University Press.

Srbinovska, Anita, Angela Srbinovska, Vivek Senthil, Adrian Martin, John McCluskey, Jonathan Bateman, and Ernest Fokoué. 2025. "Towards AI-Driven Policing: Interdisciplinary Knowledge Discovery from Police Body-Worn Camera Footage." *arXiv* preprint arXiv:2504.20007. https://doi.org/10.48550/arXiv.2504.20007.

White, Michael D., and Aili Malm. 2020. *Cops, Cameras, and Crisis: The Potential and the Perils of Police Body-Worn Cameras*. New York: NYU Press.

Wilson, David B., Bryce E. Peterson, and Daniel S. Lawrence. 2019. *Body-Worn Cameras: A Systematic Review*. Washington, DC: U.S. Department of Justice, Bureau of Justice Assistance. https://bja.ojp.gov/sites/g/files/xyckuh186/files/media/document/body-worn_video-a-systematic-review-of-literature.pdf.

ARTIFICIAL INTELLIGENCE AS RISK: BETWEEN MORALITY AND PROGRESS

Saše Gerasimoski, PhD
Faculty of Security, Skopje
sase.gerasimoski@uklo.edu.mk

Abstract

The rapid development of artificial intelligence has raised numerous dilemmas, of which the fundamental one being between morality and progress. We are all wondering whether an artificial intelligence (AI) will mark a quantum leap in human history, or, could lead to decline and even possible self-destruction of human civilization. This paper deals with examining the AI possibilities for progress and security risks associated with morality aspects of development of AI. Theoretically, the research rests primarily on risks vs. protective factors theory, analyzing both critical as well as affirmative theoretical approaches. Methodologically, the research relies on content discourse analyses of the theoretical sources. The main premise is that although the AI is too complex an issue to be analyzed one-sidedly, at certain point of its development, way before AI reaches the so called singularity, the humanity will have to give priority to morality instead of further progress based on AI.

Keywords: artificial intelligence, risk, morality, progress.

INTRODUCTION

One particular proverb saying "May you live in interesting times!", probably explains in the best way what is happening with the world society nowadays. The globalized and postmodern world in which we are living is overflooded with risks of various nature, most predominant being the so called manufactured risks. As famous English sociologist Anthony Giddens points out, manufactured risks are the risks that are being produced by society as a by-product of its development. Those are the risks created with incompetent decisions that humans made and for which they have none or little to none experience of dealing with them (Гиденс 2002, 24). Ulrich Beck, the renown German sociologist, also recognizes those risks as being the most prevalent in our societies. The risk society, as he names the contemporary society, is the one where "we are suffering from the consequences of the victory of modernity", meaning that most of the risks that we are facing, especially security risks, are direct consequence of our quest for living better and better life (Beck 2012). Amongst the sea of so called manufactured risks, slowly but surely, at the beginning almost invisibly, a new risk emerged, probably the most challenging that the humanity has ever faced. That is the risk of artificial intelligence (AI), or AI as risk. Depending on the value stance concerning the AI, the optimists see it as positive or at least neutral, while more reserved and the pessimist see it in a negative way, that is to say as security risk. However the value stance might be, we are probably witnessing the watershed moment in human history, some sort of a crossroad that could determine the bright future, or, maybe, the sudden and abrupt demise of human civilization.

The emergence of AI can both we viewed as exhilarating and frightening phenomenon. It offers a lot, but it could take from us a lot at the same time. It could completely transform the future human societies and make our lives much more comfortable, richer and longer, but, at the very same time, it could bring us huge uncertainty, fear and even self-destruction. An AI is a technology like no other in human history. As the seminal historian and philosopher Juval Noah Harari often alerts, an AI is not a tool like major previous technological advancements (for instance computers), but it is an agent. That makes it completely different form of technologically manufactured risk, which is able to change, improve and get self-conscious

over time, and by that, far superior form of intelligence than the human one. It is a system of algotyhms based on combination of biotech and infotech. Harari writes that "it is possible for the algorithms to 'hack' humanity by knowing more about me than I know about myself... Therefore, the external algorithm does not need to know me perfectly and never go wrong; it's enough to know me better than I know myself. Under these circumstances it is logical that humans will entrust more decisions and choices to algorithm" (Харари 2019, 352-353). That's how the humanity will evantually hand over its free will and control to AI. Harari enumerates myriad examples of using these AI algorithms that are taken for granted and usually people do not pay serious attention to it that they are AI, like, for instance, searching information through search engines like Google instead of finding it on its own, usage of the space orientation programs like Google Maps, Garmin etc. Very soon, these algorithms will get so superior that we will consult them even for the most intimate things, like search for partner and advices for the human relationships (Харари 2020, 70-71).

Just having in mind the picture of this very possible development of AI, sends chills through everyone's spine. The dilemma that arises is simple: Whether we will manage to develop safe forms of AI that could mark a quantum leap in progress of humanity, or we'll disregard the security and human morality, and hastily develop highly advanced forms of AI that could take control over us in some not so distant time in future and lead humanity on the brink of self-extinction, if not the very self-extinction. This paper deals with these complex issues, which are not only technological and security ones, but also much deeper, like philosophical, ethical and sociological, since they are related with the very existence of humanity as we know it.

ARTIFICIAL INTELLIGENCE AS RISK

At the very beginning, before we go over and elaborate the AI as risk, we must clarify the difference between artificial intelligence and alien intelligence. From a technological, but moreover from philosophical and sociological viewpoint, the artificial intelligence and alien intelligence are not the same. Artificial intelligence is a form of controllable intelligent

technology developed by humans, in order to serve the humanity, representing a tool for improving the life and intelligence of humans. While, an alien intelligence is independent and to a large extent, if not entirely, uncontrollable form of intelligence. In its meaning of the word alien, it does not imply extraterrestrial, but the one which was developed by humans, alientaned from humans and which acquired its own self-awareness over time, whether by human intent or unintentionally. Unlike artificial intelligence, alien intelligence acts as an agent, not as a tool, meaning that it would be far superior than any known human intelligence in history of human kind, with a terrible security risk and threat of taking over humanity. So, as the very name suggests, an alien intelligence can be defined as alienated form of intelligence, developed by the humans themselves, and can be considered as typical manufactured security risk. As long as we face artificial intelligence, the humanity can deal with its risks in a relatively successful way, but when confronted with the artificial intelligence that has completely transformed into alien intelligence, the very dealing with its risks remain problematic, unknown and highly uncertain.

So, getting back to the topic of this item, why the AI is considered to be a risk in the first place? The risk is usually defined as possible, assumed, expected occurrence, a probable event with a possible positive or negative outcome, but most often negative one, which may cause endangerment to protected values and goods (Спасески, Аслимоски и Герасимоски 2017, 129). The very definition of risk in itself contains two variables (probability and consequence). So, the AI can be rightly referred to as risk because it represents a phenomenon which contains possibility of doing good or bad, of positive or negative outcome, when under outcome we understand the consequence of developing the AI. The very definition of risk has generally two meanings. First one is the so called neutral meaning, and the second one, much more found in theory and practice, is the security meaning. Nowadays, the security meaning of risk prevails over the neutral meaning, due to the complex socio-cultural dynamics and change in the globalized and postmodern societies, that shape the value stance towards the risks. The eternal dilemma of prudence (cautiousness) versus risk (gambling, courage), once again press us with all its seriousness and reality. It is a dilemma that

produces the so called 'homo prudens vs homo aleatoris" debate (Gerasimoski 2014, 55). This dilemma and this debate are especially relevant and true when it comes to AI as risk. It sublimes kind of ambivalent approach towards the AI, with AI scientists and companies being mostly on the side of homo aleatoris, while some of the AI scientists and majority of the aware public opinion on the side of homo prudens. Within this second group of scientists, though yet a minority, are many leading AI experts and entrepreneurs such as Yoshua Bengio, Geoffrey Hinton, Sam Altman, Elon Musk and Mustafa Suleyman, who have warned that AI could destroy our civilisation (Harari 2024). However, it is almost impossible for these approaches to come to terms with each other and establish some kind of balanced approach, since both approaches tend to be excludable.

There are basically two types of security risks associated with AI. The first one is related with inappropriate and unethical use of AI by humans, and the second one is the risk of AI becoming autonomous, super intelligent and self-aware agent and thus take over humans. Within the first type or risk we can even nowadays evidence numerous security risks like fake news in the media, hate speech generated by AI bots, dissemination of AI generated texts, images and videos on social media and in online space in general without AI disclosure, etc. As we can see, since the AI technology development is in its infancy, but rapidly advancing, the AI security risks related with inappropriate and unethical human behavior are for now generally limited within the information and online world. But, as the time progresses, more serious forms of these types of risks are expected, like AI generated crises in financial sector, critical infrastructure attacks and probably the most dangerous one, unethical use of AI for military purposes, like waging info wars, as well as wars with advanced WMD (Weapons of Mass Destruction). Speaking of the second type or security risk associated with very AI as autonomous, super intelligent and self-aware agent, there are also numerous sorts of security risks which are horrifying even when imagining their potential destructive nature, such as autonomous battle machines, autonomous self-programming machines, etc.

Each risk, irrespective of being neutral or security one, contains both certainty and uncertainty. Depending on their presence within the risk, we

can talk about low or high risk. The more we know about the risk factors that constitute the risk, the lower the risk will be. On a contrary, the less we know about the risk factors that constitute the risk, the higher the risk will be. Undoubtedly, the AI risk can be considered as high security risk, since we know so little about it, or shortly said, we are surrounded with significant uncertainty when it comes to AI risk. That's why, Geoffrey Hinton warns that "we're in a period of huge uncertainty where we really don't know what's going to happen," (Goodyear 2023). The "latent side effects of risk", as once Ulrich Beck said, are hidden, unseen, unpredictable, unknown and unwanted consequences of security risks (Bek 2001, 52). AI risk contains them all and they are all related to the inherently vague and more uncertain future associated with juggernaut of the progress. Humanistic and ethical concerns are to be taken with all seriousness by everyone, not only by scientists but by all people. In the past two centuries the human race has dominated our planet and we became as Juval Noah Narari says "Gods on the Earth". But our imperfect human nature makes us prone to inhumane and immoral behavior. With an AI it seems that we have no room for error. He rightfully alarms, "is there anything more dangerous than unsatisfied and irresponsible Gods who do not know what they want?" (Харари 2018, 464).

MORALITY AND PROGRESS DILLEMA ABOUT THE AI

Although it is unclear who started working on artificial intelligence technology, however, "it is said that Alan Turing was the first, as there is a record that states that he had delivered lectures on artificial technology in 1947" (Kanade, Karad, Bhoite, Kanade and Jain 2023, 67). Since then, the development of AI has seen many ups and downs, conditioned mainly by the interest and investments of the goverments and companies for it. As theoretical physicist Michio Kaku comments, the development of AI has witnessed three phases of boom and stagnation. The latest one, the third phase, started to develop only quite recently, from the beginning of the XXI-st century (Каку 2022, 283-285). It was generally instigated by the governments financial boosts, especially US and Chinese, as well as by the fierce competence of the world's biggest AI companies like Google, Open AI, Anthropic, Meta, Scale AI etc. The European Commision's Joint

Research Centre in its Report from 2020, also mentions three phases of AI development in detail, stating that "the first phase lasted from the spring of the 1950s, with the foundation of most of the AI algorithms, followed by the winter of the 1970s. The second phase commenced in the spring of the 1970s, with the paradigm shift in symbolic algorithms and building of expert systems, followed by the winter of the 1990s. The third phase saw the development of machine learning in the 1990s, which led to the development of deep learning in the spring of the 2010s" (Delipetrev, Tsinaraki and Kostić 2020, 3).

The concept of AI rests on developing AI tools and AI agents, based on pervasive use of machine learning (ML) and deep learning (DL) algorithms. Thanks to the pioneering work of the founder, also named "godfather of AI", the distinguished Canadian professor and Nobel prize laureate for physics for 2024 Geoffrey Hinton, we have managed to develop artificial neural networks, and thus created AI that works and learns in the similar way as neural networks in humans brain, i.e., the AI that works and learns from experience (Hinton 1992). By imitating human brain and by installing the basic principles of learning gained through social and biological evolution, the AI is now approaching to the phase of development where it can learn and even act by itself, bringing decisions on its own. By doing so, AI machines are no longer merely "information gatekeepers", but are playing the role of new kind of choice architects (Tiribelli 2024, 195).

As far as morality is concerned, there has been a moral relativization coupled with value ambivalence ever since globalization and postmodernism took sway. Morality is not something that is as solid as it was in modernity and the values also became more vague and unclear. As Zygmunt Bauman correctly concludes, the risk is getting public and the moral is getting private (Бауман 2005, 323). The so called dissolvment of moral in the era of what he calls "liquid modernity", or postmodernism and globalization, means that the moral is largely relativized. Together with Leonidas Donskis, they recognize a kind of moral blindness and loss of sensitivity in nowadays societies (Bauman and Donskis 2013). So this means that humanity has already having serious problems with morality even without the AI. The occurrence of AI could only worsen the situation, because there will be two

major risks directly related with AI: risk of bad actors, amoral people who could misuse the AI, and the so called AI ethics, the imagined situation of super intelligent AI technology and its stance towards humanity, which would not be aligned with what we know it as morality, or rather, human morality. The issue of AI ethics remains to be resolved very soon. Though there are various ideas on how this crucial issue has to be addressed, no one seems to have found lasting and satisfactory solution to this issue so far (Mohammadiounotikandi and Babaeitarkami 2024, 30-37; Redaelli 2023, 1-8; Pana 2006, 254-264).

For the time being, the morality is equaled with humanity, so being human means being moral, since we know of no other moral standard outside humanity. The fate of morality is therefore tied to the fate of man (Лазарески 2015, 234). Morality comes from the best in human nature, or as the ancient philosopher Protagoras wrote "homo menzura"-"the humankind is the measure of all things" (Јосифовски 1990, 24). Moral is defined as a distinctive human form of conciouseness and behavior, determined with the notion of good, and as such it is as a system of norms which determine good from bad, thus being the fundamental and most frequent regulator of everyday human behavior (Лукић 1976, 103). Despite of human nature being imperfect, it sublimates the best of humans as specie, such as its free will, creativity and goodness in terms of other people and in terms of all the other creatures on the planet and nature that surround the man. As the famous Macedonian ethicist Kiril Temkov writes, "the moral is an absolute value that is being defined by its high values and meaning for the life of humans, … but could be also known intuitively" (Темков 2009, 7-8).

The development of AI could undoubtedly stir the progress of humankind and help us living much better. Amongst the many positive achievements of the progress driven by AI are rooting out some of the diseases and poverty. But, as the famous physicist Stephen Hawking said: "Although I've known to be optimist when it comes down to human specie, I'm not so sure about this (AI)… In short, the advent of the super intelligent AI would be the best and the worst thing that has ever happened to humanity" (Хокинг 2019, 174-175). Geoffrey Hinton loudly speaks about the "existential threat" from AI, while Stephen Hawking warns about the "existential risk" from AI. However

formulated, we must act preventively to put the progress based on AI in line with morality and human values. That would mean safe, ethical and controllable AI that works for the good and wellbeing of humankind.

If we are to assess the impact that the AI could have on progress of humanity, we have to set some reference points, goals or values that need to be attained in order to be able to talk about humanity progress. The most suitable for that are the 17 UN Sustainable Development Goals. As Marjan and Zubaidi conclude in their study on the impact of AI on sustainable development, despite the fact that "the impact of AI on the United Nations Sustainable Development Goals (SDGs) is still uncertain", "the potential of AI to promote sustainable development is vast", as it could help attain most of the SDG's, like reducing or even rooting out of poverty and hunger worldwide, providing good health and well-being, ensuring quality education, providing clean water and sanitation as well as affordable and clean energy, stimulating industry innovation, improving infrastructure and reducing inequalities. These could all mean improving the quality of human life and sustainable human progress, but only if the risks associated with AI are addressed on time and entirely. As the authors wisely put, "although AI can promote sustainability, it can also be used for activities that worsen global issues" (Marjan and Zubaidi 2023, 24, 29). Some other studies have come to similar conclusions regarding this issue (Thanyawatpornkul 2024, 2301- 2317; ITU 2024; Ametepey, Aigbavboa, Thwala and Addy 2024, 1-67; Vinuesa, Azizpour, Leite, Balaam, Dignum, Domisch, Felländer, Langhans, Tegmark and Nerini 2020, 1-10).

What the future holds in terms of AI, depends on resolving the dilemma between morality and progress. Morality has always been associated with stability and order of human society, while progress has always been associated with necessary change and development. They have to coexist in certain balance in a given human society for that society to function and thrive. But, it is easier said than done. The human societies throughout history have always favored one over each other. In fact, there were more historical periods when humanity preffered progress over morality. Perhaps, that is the nowadays situation with AI, where it is obvious that humanity is impatient to grab the "Brave New World" so soon, even at the price of

neglecting the morality and humanism. As a matter of fact, humanity was never short of favoring progress over morality. As Alfred North Whitehead, the famous American mathematician wrote a century ago, "it is the business of the future is to be dangerous" (Whitehead 1925, 208). He rightly observed that "the greatest achievements of civilization are brought about processes that nearly destroyed the society in which they unfolded" (Каку 2022, 117). We can only think of nuclear energy and atomic bomb as an examples of that. But, he also added, "The art of progress is to preserve order amid change and to preserve change amid order" (Whitehead 1929, 339). In fact, how to reconcilliate and balance the AI for the sake of humanity and morality on one hand, and for the sake of progress on the other since they are both needed? It seems that when it comes to AI, we should favor progress in line with morality and safety up to the point of its development before AI reaches the so called singularity (equating and surpassing human intelligence). From that point on, the humanity will have to give priority to morality instead of further progress based on AI.

CONCLUSIONS AND RECOMMENDATIONS

Taking into consideration what was said in the paper, we can sum up the following conclusions and recommendations:

- The artificial intelligence (AI) is the most challenging manufactured risk that the humanity has ever faced;
- Since AI is not only a tool like major previous technological advancements (for instance computers), but an agent, that makes it completely different form of technologically manufactured risk;
- As long as we face artificial intelligence, the humanity can deal with its risks in a relatively successful way, but when confronted with the artificial intelligence that has completely transformed into alien intelligence, the very dealing with its risks remain problematic, unknown and highly uncertain;

- The AI can be rightly referred to as risk because it represents a phenomenon which contains possibility of doing good or bad, of positive or negative outcome;
- The so called 'homo prudens vs homo aleatoris" debate sublimes the kind of ambivalent approach towards the AI, with AI scientists and companies being mostly on the side of homo aleatoris, while some of the AI scientists and majority of the aware public opinion on the side of homo prudens;
- There are basically two types of security risks associated with AI. The first one is related with inappropriate and unethical use of AI by humans (like fake news in the media, hate speech generated by AI bots, dissemination of AI generated texts, images and videos on social media and in online space in general without AI disclosure, etc.), and the second one is the risk of AI becoming autonomous, super intelligent and self-aware agent and thus take over humans (like autonomous battle machines, autonomous self-programming machines, etc.);
- The occurrence of AI could only worsen the situation with morality, because there will be two major risks directly related with AI: risk of bad actors, amoral people who could misuse the AI, and the so called AI ethics, the imagined situation of super intelligent AI technology and its stance towards humanity, which would not be aligned with what we know it as morality, or rather, human morality;
- Although the development of AI could help attaining most of the 17 UN Sustainable Development Goals (SDG) which are considered as synonyms for humankind progress, we have to be aware and extremely cautious that if the AI is used unethically and neglecting security issues, it could actually make difficult the attaining of SDG,s;
- It seems that when it comes to AI, we should favour progress in line with morality and safety up to the point of its development before AI reaches the so called singularity (equating and surpassing human

intelligence). From that point on, the humanity will have to give priority to morality instead of further progress based on AI.

REFERENCES

1. Ametepey Simon Ofori, Aigbavboa Clinton, Thwala Wellington Didibhuku and Addy Hutton. 2024. "The Impact of AI in Sustainable Development Goal Implementation: A Delphi Study." *Sustainability* 16: 1-67.
2. Bauman, Zygmunt and Donskis, Leonidas. 2013. *Moral Blindness: The Loss of Sensitivity in Liquid Modernity*. Polity Press.
3. Бауман, Зигмунд. 2005. *Постмодерна етика*. Темплум.
4. Bek, Ulrih. 2001. *Rizično društvo*. Filip Višnjič.
5. Beck, Ulrich. 2012. *Living and Coping With the World Risk Society*. Speech given at 42nd St. Gallen Symposium.
6. Vinuesa Ricardo, Azizpour Hossein, Leite Iolanda, Balaam Madeline, Dignum Virginia, Domisch Sami, Felländer Anna, Langhans Simone Daniela, Tegmark Max and Nerini Francesco Fuso. 2020. "The Role of Artificial Intelligence in Achieving the Sustainable Development Goals." *Nature Communications* 11: 1-10.
7. Whitehead, North Alfred. 1929. *Process and Reality: An Essay in Cosmology*. The Free Press.
8. Whitehead, North Alfred. 1925. *Science and The Modern World*. The Macmillan Company.
9. Gerasimoski, Saše. 2014. "The Risk in Contemporary Society: Homo Prudens vs Homo Aleatoris Debate." *The Sociological Review* 15 (1): 53-66.
10. Гиденс, Ентони. 2002. *Забеган свет: како глобализацијата ги преобликува нашите животи*. Филозофски факултет.
11. Goodyear, Sheena. 2023. *The 'godfather of AI' says he's worried about 'the end of people': An Interview with Geoffrey Hinton*. CBC Radio. 3.5.2023.

12. Delipetrev Blagoj, Tsinaraki Chrisa and Kostić Uroš. 2020. *AI Watch Historical Evolution of Artificial Intelligence: Analysis of The Three Main Paradigm Shifts in AI*. Publications Office of the European Union.

13. ITU. 2024. *AI for Good: Impact Report*. International Telecommunication Union.

14. Јосифовски, Јонче. 1990. *Филозофија*. Просветно дело.

15. Каку, Мичио. 2022. *Иднината на умот: научна потрага по разбирање, подобрување и зголемување на моќта на умот*. Арс Ламина.

16. Kanade Anuradha, Karad Vishwanath, Bhoite Sachin, Kanade Shantanu and Jain Niraj. 2023. "Artificial Intelligence and Morality: A Social Responsibility." *Journal of Intelligence Studies in Business* 13 (1): 65–75.

17. Лазарески, Филип. 2015. *Моралот низ вековите и во сегашноста*. Матица Македонска.

18. Лукић, Радомир. 1976. *Социологија морала*. Научна књига.

19. Marjan, Riyam K. & Zubaidi and Ali Mohammed Hasan. 2023. "Artificial Intelligence and Sustainable Development." *Al-Mustaqbal Journal of Sustainability in Engineering Sciences* 1(1): 23-30.

20. Miernicki, Martin and Ng (Huang Ying), Irene. 2021. "Artificial Intelligence and Moral Rights." *AI & SOCIETY* 36: 319–329.

21. Mohammadiounotikandi, Ali and Babaeitarkami, Somayeh. 2024. "The Ethics of The Artificial Intelligence: Balancing Progress with Responsibility." *International Journal of Material and Mathematical Sciences* 6(2): 30-37.

22. Pana, Laura. 2006. "Artificial Intelligence and Moral Intelligence." *tripleC* 4(2): 254-264.

23. Redaelli, Roberto. 2023. "Different Approaches to The Moral Status of AI: A Comparative Analysis of Paradigmatic Trends in Science and Technology Studies." *Discover Artificial Intelligence* 3(25): 1-8.

24. Спасески Јордан, Аслимоски Пере и Герасимоски Саше. 2017. *Основи на приватната безбедност*. Факултет за безбедност.

25. Темков, Кирил. 2009. *Како да се биде добар? Етика за млади*. ОХО.

26. Tiribelli, Simona. 2024. "Artificial Intelligence and Moral Progress: A Missing Debate." *Filosofia morale/Moral Philosophy* 2: 189-200.

27. Thanyawatpornkul, Rapeerat. 2024. "Artificial Intelligence-Driven Solution for Global Challenges: A Systematic Review from Sustainable Development Goals Perspectives." *International Journal of Business Management and Economic Research (IJBMER)* 15(1): 2301- 2317.

28. Harari, Juval Noah. 2024. "Never summon a power you can't control': Yuval Noah Harari on how AI could threaten democracy and divide the world". *The Guardian*. 24.8.2024.

29. Харари, Јувал Ноа. 2020. *21 лекција за 21 век*. Три.

30. Харари, Јувал Ноа. 2019. *Хомо деус: кратка историја на иднината*. Три.

31. Харари, Јувал Ноа. 2018. *Сапиенс: кратка историја на човечкиот род*. Три.

32. Hinton, Geoffrey E.1992. "How Neural Networks Learn from Experience." *Scientific American* : 145-151.

33. Хокинг, Стивен. 2019. *Кратки одговори на крупните прашања*. Топер.

THE ROLE OF ARTIFICIAL INTELLIGENCE IN SUPPORTING THE PILLARS OF SUSTAINABLE DEVELOPMENT—ECONOMIC GROWTH, SOCIAL INCLUSION, AND ENVIRONMENTAL PROTECTION[1]

Ivana Ostojić, PhD, Research Associate

Institute of Social Sciences, Belgrade, iostojic@idn.org.rs

Abstract

Artificial Intelligence has emerged as a transformative force in advancing the three interconnected pillars of sustainable development-economic growth, social inclusion, and environmental protection. By driving innovation, improving efficiency, and enabling data-driven decision-making, AI contributes to sustainable economic growth through enhanced productivity, smart manufacturing, and optimized resource management. In the social sphere, AI fosters inclusion by improving access to education, healthcare, and public services, while also promoting fairness and reducing inequalities through targeted social policies and digital empowerment. Environmentally, AI supports climate action, biodiversity preservation, and pollution control through predictive analytics, intelligent monitoring systems, and optimized energy use. However, the ethical, regulatory, and social challenges of AI, such as data privacy, algorithmic bias, and unequal access to technology, should be addressed to ensure its benefits are equitably distributed. This paper explores how AI is integrated into sustainable

---

development strategies to balance economic, social, and environmental objectives, highlighting its potential to accelerate progress toward the United Nations Sustainable Development Goals.

Keywords: *artificial intelligence, sustainable development, economic growth, social inclusion, environmental protection*

INTRODUCTION

Artificial intelligence (AI) has emerged as one of the most transformative technologies of the 21st century, reshaping industries, governance, and social systems. Beyond its technical and economic potential, AI holds the promise of addressing some of humanity's most pressing challenges: inequality, exclusion, and environmental degradation. When developed and deployed responsibly, AI can catalyze justice, inclusion, and sustainable development, aligning with the United Nations' 2030 Agenda and the Sustainable Development Goals (SDGs).

Artificial intelligence possesses immense transformative power to reshape societies and economies. When aligned with principles of justice, inclusion, and sustainability, AI can become a foundational tool in building a fairer and more resilient world. However, this potential will only be realized through ethical governance, inclusive design, and sustainable innovation that place human well-being and planetary health at the center of technological advancement. AI represents both an opportunity and a responsibility. When governed responsibly and applied strategically, AI can drive measurable progress toward the SDGs, optimizing resource use, enabling inclusive growth, and enhancing resilience to global challenges. The key lies in ensuring that technological advancement is coupled with ethical stewardship, cross-sector collaboration, and a commitment to social justice.

AI can significantly contribute to all three pillars of sustainable development: economic growth, social inclusion, and environmental protection. For instance, AI-driven analytics improve resource efficiency in industries, enabling circular economy practices and reducing waste. In agriculture, AI supports sustainable food systems through precision farming, weather prediction, and pest detection. In energy management, intelligent grids and

predictive maintenance enhance renewable energy integration and reduce carbon emissions (Vinuesa et al. 2020).

Artificial intelligence has emerged as one of the most transformative technological forces of the 21st century, driving innovation, productivity, and structural changes in the global economy. By automating tasks, enhancing decision-making processes, and creating new forms of value, AI is redefining the mechanisms of economic growth and competitiveness across industries and nations. The integration of AI technologies into production processes, finance, healthcare, manufacturing, and public administration is generating measurable efficiency gains and fostering the emergence of new markets and business models (Brynjolfsson and McAfee 2017; Aghion et al. 2019).

AI has emerged as a transformative force in shaping modern societies. Beyond its economic and technological potential, AI holds considerable promise for advancing social inclusion by reducing inequalities, improving access to opportunities, and fostering participatory development. This paper explores the key applications, benefits, and challenges associated with AI-driven environmental protection initiatives, highlighting its role as both an enabler of innovation and a driver of sustainable development.

INTELLIGENT ECONOMIES: THE ROLE OF ARTIFICIAL INTELLIGENCE IN ACCELERATING GROWTH

AI contributes to economic growth primarily through productivity enhancement and technological innovation. Artificial Intelligence has emerged as a fundamental driver of productivity and innovation across industries, reshaping economic structures, labor markets, and organizational models. As a general-purpose technology, AI's capacity to process vast amounts of data, automate decision-making, and enhance predictive accuracy has positioned it at the core of the Fourth Industrial Revolution (Brynjolfsson and McAfee 2017). AI contributes to productivity through automation, optimization, and augmentation. Automation enables machines to perform repetitive and routine tasks more efficiently than humans, reducing costs and increasing output (Acemoglu and Restrepo 2020a). Optimization algorithms improve resource allocation and operational efficiency, while augmentation tools enhance human capabilities by providing real-time insights and

decision support. Studies indicate that firms integrating AI technologies experience higher output per worker and faster scaling potential (OECD 2023). Machine learning, natural language processing, and robotics enable firms to optimize operations, reduce costs, and increase output quality. According to the OECD (2021), AI adoption can raise labor productivity by up to 20%-40% in some sectors by augmenting human capabilities and streamlining data-intensive processes. Furthermore, AI fosters innovation by enabling rapid experimentation, predictive modeling, and data-driven product development, which accelerates technological diffusion and boosts total factor productivity (Korinek and Stiglitz 2021). AI fosters innovation by transforming how organizations design, test, and deliver products and services. Machine learning models enable predictive innovation identifying emerging trends and customer needs, thus shortening the innovation cycle (Cockburn, Henderson and Stern 2018). Moreover, AI-driven R&D platforms can autonomously generate new hypotheses, materials, or drug compounds, expanding the frontier of human creativity (Agrawal, Gans and Goldfarb 2019). While AI enhances productivity and innovation, its benefits are unevenly distributed. Advanced economies and large technology firms capture most of the gains, raising concerns about inequality and market concentration (UNCTAD 2021). Furthermore, the transformation of labor markets due to automation requires reskilling and adaptive education systems to ensure inclusive growth (World Economic Forum 2024).

While AI-driven automation stimulates productivity, it also triggers labor market restructuring. Artificial Intelligence, as one of the most transformative technologies of the 21st century, is reshaping global economies, labor markets, and industrial structures. Its influence extends beyond productivity and automation, fundamentally altering the composition of work, the demand for skills, and the organization of production. AI acts as both a catalyst for innovation and a disruptor of traditional economic structures, driving a new phase of structural transformation comparable to past industrial revolutions.

Routine and manual jobs are increasingly automated, while demand rises for high-skill occupations in data science, algorithm design, and AI governance. This polarization can widen income inequality if policy frameworks fail to

ensure inclusive access to AI-related education and reskilling opportunities (Acemoglu and Restrepo 2019). However, in the long run, new forms of employment and entrepreneurship often emerge around AI ecosystems, such as platform work, digital services, and innovation-driven start-ups. At the macroeconomic level, AI adoption correlates with accelerated GDP growth and competitiveness. A PwC (2018) report estimates that AI could contribute up to $15.7 trillion to the global economy by 2030, driven by productivity gains and consumption effects. Economies investing heavily in AI research and infrastructure, such as the United States, China, and the European Union, are positioning themselves for leadership in the next wave of digital transformation. However, disparities in AI adoption between developed and developing nations could exacerbate global inequality unless inclusive and collaborative frameworks are established.

AI technologies influence the labor market through three primary mechanisms: automation, augmentation, and creation. Automation replaces routine and predictable tasks, reducing labor demand in some sectors (e.g., manufacturing, data entry, logistics). Augmentation enhances human productivity by complementing cognitive and creative tasks, particularly in knowledge-intensive sectors. Creation stimulates new job categories, industries, and entrepreneurial opportunities driven by AI-driven products and services. However, the distributional effects are uneven. While high-skilled workers benefit from AI-driven productivity gains, middle and low-skilled workers may face displacement risks, leading to greater polarization of income and job quality (Brynjolfsson and McAfee 2017; Acemoglu and Restrepo 2020a).

AI accelerates structural transformation, the shift of labor and output from agriculture and manufacturing toward services and digital sectors. AI-powered automation increases productivity in traditional industries, but it also shifts employment toward knowledge-based services such as information technology, finance, health, and creative industries (Bessen 2019). Simultaneously, AI enables the emergence of platform economies and digital ecosystems, transforming value chains and production models globally. These shifts contribute to the reallocation of capital and labor, reshaping national and international competitiveness. The labor market

effects of AI depend largely on skills adaptation. The demand for analytical thinking, digital literacy, and socio-emotional skills is rising, while routine physical and cognitive skills decline in relevance (World Economic Forum 2023). This necessitates major reforms in education, training, and lifelong learning systems to prepare workers for AI-augmented roles. Policymakers face the dual challenge of fostering innovation while ensuring inclusive labor transitions (ILO 2021). To harness AI's potential for equitable structural transformation, governments must implement policies that: promote inclusive innovation ecosystems that integrate SMEs and developing regions into AI-driven growth; invest in reskilling and upskilling programs aligned with AI labor market trends; strengthen social protection systems for displaced workers; encourage AI ethics and governance to ensure fair use and accountability in automation decisions. Inclusive and human-centered AI policies can ensure that technological progress contributes to sustainable and broad-based economic development (OECD 2023).

For AI to translate effectively into sustainable economic growth, institutional readiness and governance play crucial roles. Policymakers must address issues related to data access, privacy, algorithmic bias, and workforce displacement while ensuring that AI innovation aligns with societal well-being and environmental sustainability (UNESCO 2021). Strategic investments in digital infrastructure, research, and education are also critical for maximizing AI's long-term economic benefits. Artificial intelligence is not merely a technological advancement but a transformative economic force reshaping productivity, labor markets, and global competitiveness. While it offers immense opportunities for economic growth and innovation, realizing its full potential requires proactive governance, equitable access, and responsible innovation. The challenge for policymakers and business leaders lies in fostering an AI-driven economy that is not only prosperous but also inclusive and sustainable.

ARTIFICIAL INTELLIGENCE AND SOCIAL INCLUSION: LEVERAGING TECHNOLOGY FOR AN EQUITABLE AND INCLUSIVE SOCIETY

Social inclusion refers to the process of improving the conditions of individuals and groups who are marginalized based on factors such as poverty, disability, gender, ethnicity, or geographic location. This part of the paper explores the role of AI in promoting social inclusion across key domains such as education, healthcare, employment, and civic participation, while also addressing the risks of digital exclusion and algorithmic bias. In recent years, AI technologies, ranging from machine learning to natural language processing, have been increasingly deployed to address barriers to inclusion. As part of the broader agenda of sustainable development and human rights, AI can support inclusive societies by enhancing accessibility, enabling data-driven decision-making, and promoting equitable service delivery.

AI contributes to social inclusion through its capacity to analyze large datasets, predict social outcomes, and personalize interventions. For instance, AI-powered platforms can assist in identifying at-risk populations, improving the targeting of welfare programs, and facilitating inclusion in digital economies. At the same time, AI can strengthen institutional capacities by supporting transparent governance and inclusive policymaking. AI-driven tools such as adaptive learning systems and intelligent tutoring platforms can personalize education for learners with disabilities or those in underserved regions. Speech recognition and translation technologies can reduce linguistic barriers, while AI analytics can help policymakers identify learning gaps and improve resource allocation. However, unequal access to digital infrastructure can deepen educational divides, emphasizing the need for inclusive digital literacy initiatives.

AI holds transformative potential for creating more equitable, efficient, and engaging educational systems. AI is increasingly transforming the landscape of education by reshaping how knowledge is delivered, accessed, and evaluated. AI encompasses a range of computational techniques that enable machines to simulate human intelligence, including learning, reasoning, and problem-solving (Russell and Norvig 2021). One of the most significant contributions of AI in education lies in personalized learning. Intelligent tutoring systems and adaptive learning platforms analyze individual student data to tailor content, pace, and feedback according to learners' needs

(Luckin et al. 2016). This personalization supports differentiated instruction and helps address diverse learning styles and abilities. AI also enables automated assessment and learning analytics, providing educators with insights into student performance and engagement patterns (Ifenthaler and Yau 2020). Also, educators must adapt their roles to effectively integrate AI tools into pedagogy while maintaining human-centered learning principles (Holmes et al. 2021).

AI has the potential to contribute to Universal Health Coverage (UHC) and the Sustainable Development Goals (SDG 3: Good Health and Well-being) by improving health system efficiency and enabling cost-effective service delivery. AI is transforming healthcare systems worldwide by enhancing diagnostic accuracy, treatment personalization, and healthcare delivery efficiency. Beyond technological innovation, AI also holds potential to improve healthcare accessibility, particularly in underserved regions and vulnerable populations. However, while AI offers opportunities for equitable healthcare, it also raises concerns about bias, data privacy, and the digital divide. AI technologies such as machine learning, natural language processing, and computer vision have enabled the development of virtual assistants, predictive analytics tools, and telemedicine platforms. These tools enhance access by:

- Bridging geographic gaps: AI-powered telehealth and remote diagnostic systems allow patients in rural or low-resource settings to receive medical consultations and monitoring without needing to travel (Jiang et al. 2017; Razzak, Imran, and Xu 2019).
- Improving efficiency: Automated triage systems and AI-based scheduling optimize resource allocation, reducing waiting times and enabling physicians to focus on complex cases (Topol 2019; Davenport and Kalakota 2019).
- Enhancing preventive care: Predictive models analyze large-scale health data to identify at-risk populations, enabling earlier interventions and public health planning (Beam and Kohane 2018; Rajkomar, Dean, and Kohane 2019).

- Supporting non-specialist healthcare providers: AI decision-support tools assist general practitioners and community health workers in diagnosing conditions accurately, extending the reach of medical expertise (Esteva et al. 2019; Yu, Beam, and Kohane 2018).

Despite these advances, the integration of AI in healthcare access presents several ethical and structural challenges:

- Algorithmic bias: AI systems trained on biased or non-representative datasets may perpetuate health disparities, particularly affecting minority or marginalized groups (Leslie et al. 2021).
- Data privacy and consent: The use of personal health data requires robust governance frameworks to ensure confidentiality and informed consent (Price and Cohen 2019).
- Digital inequality: Unequal access to digital infrastructure, devices, and digital literacy can exacerbate healthcare inequities instead of reducing them (Latulippe, Hamel, and Giroux 2020).
- Regulatory and accountability issues: The lack of standardized oversight in AI healthcare applications complicates quality assurance and patient safety (Topol 2019).

AI can foster inclusive labor markets by matching skills to job opportunities, enabling remote work, and supporting entrepreneurship through digital platforms. Yet, automation also threatens low-skilled jobs, necessitating policies for reskilling, lifelong learning, and social protection. Artificial intelligence is reshaping global economies by transforming how people work, create value, and participate in economic systems. The integration of AI into production, services, and decision-making processes presents both opportunities and challenges for employment and inclusive economic growth. While AI can boost productivity, efficiency, and innovation, it also raises concerns about job displacement, wage inequality, and the future of work. The transformative power of AI also brings risks of widening socioeconomic divides. Automation may exacerbate income inequality by polarizing the labor market, creating high-skill, high-wage jobs while reducing opportunities in low- and middle-skill sectors (Acemoglu and Restrepo 2020). Understanding AI's dual impact on employment and

economic participation is crucial for developing inclusive policies that ensure shared prosperity in the digital age.

AI-driven automation and intelligent systems are changing the structure of labor markets. Routine and repetitive tasks (both manual and cognitive) are increasingly performed by machines, leading to a decline in demand for certain job categories, especially in manufacturing, logistics, and administrative services (Brynjolfsson and McAfee 2014). Conversely, AI is creating new forms of employment, including roles in data science, machine learning, digital ethics, and AI maintenance. The World Economic Forum (2023) estimates that AI and automation could displace around 85 million jobs globally by 2025 but also create 97 million new roles better suited to human–machine collaboration. However, these new jobs often require advanced digital literacy, adaptability, and continuous learning, skills not evenly distributed across populations.

AI can enhance economic participation by enabling inclusive growth through new business models and access to digital platforms. For instance, AI-powered tools can support entrepreneurship, financial inclusion, and small business competitiveness by offering predictive analytics, customer insights, and automation of routine operations (OECD 2021). Furthermore, AI applications in microfinance, agriculture, and e-commerce expand opportunities for marginalized communities and small-scale entrepreneurs, particularly in developing countries. However, the benefits are unevenly distributed-countries and individuals with limited digital infrastructure or education face significant barriers to entry in AI-driven economies (UNDP 2022).

THE ROLE OF ARTIFICIAL INTELLIGENCE IN ADVANCING ENVIRONMENTAL PROTECTION

Artificial intelligence has emerged as a tool in addressing global environmental challenges. Through its capacity to process vast datasets, generate predictive models, and optimize resource use, AI is increasingly being integrated into efforts to mitigate climate change, preserve biodiversity, and enhance sustainable resource management. The accelerating environmental degradation caused by industrialization, urbanization, and population growth has prompted the search for advanced technologies to

monitor, manage, and mitigate ecological harm. Artificial intelligence, encompassing machine learning, deep learning, and data analytics, provides powerful tools for environmental monitoring and decision-making. Its integration across sectors, from agriculture and energy to waste management and conservation, marks a significant paradigm shift in environmental governance (Vinuesa et al. 2020).

AI contributes to climate action through improved modeling and forecasting. Machine learning algorithms analyze atmospheric data to predict weather patterns, extreme events, and emission trends. AI-powered systems enhance energy efficiency in industrial and urban settings by optimizing energy use and integrating renewable energy sources into smart grids (Rolnick et al. 2022). Through its capabilities in data analysis, pattern recognition, and predictive modeling, AI enables innovative solutions for both mitigation (reducing greenhouse gas emissions) and adaptation (adjusting to the impacts of climate change). AI technologies play a pivotal role in reducing emissions and enhancing energy efficiency across multiple sectors:

- Energy Systems Optimization: Machine learning algorithms optimize the generation, storage, and distribution of renewable energy sources such as solar and wind. For instance, AI-based predictive models can forecast energy demand and weather conditions, enabling smart grid management that minimizes energy loss and supports the integration of renewables (Rolnick et al. 2022).

- Industrial Efficiency and Emission Reduction: AI-driven process control systems monitor and adjust industrial operations to minimize energy waste and reduce emissions. Predictive maintenance powered by AI prevents equipment failures and unnecessary production downtime, leading to lower carbon footprints (Krause et al. 2021).

- Smart Mobility and Transportation: Intelligent transportation systems utilize AI to optimize traffic flow, improve public transport efficiency, and support the transition to electric vehicles through route optimization and charging infrastructure planning (Creutzig et al. 2022).

- Precision Agriculture: AI applications in agriculture optimize irrigation, fertilizer use, and pest management, reducing resource waste and greenhouse gas emissions from agricultural practices (Liakos et al. 2018).

AI also supports societies in adapting to climate change by improving forecasting, risk management, and resource planning:

- Climate and Weather Prediction: Deep learning models process complex climate data to improve the accuracy of extreme weather forecasts such as floods, heatwaves, and hurricanes. This helps governments and communities to enhance early warning systems and disaster preparedness (Reichstein et al. 2019; Bengio, Lecun, and Hinton, 2021).
- Environmental Monitoring: Satellite imagery analyzed by AI assists in monitoring deforestation, glacier retreat, sea-level rise, and land degradation, providing real-time insights for environmental protection and sustainable land management (Rolnick et al. 2022).
- Water and Resource Management: AI models optimize water allocation in drought-prone regions by analyzing hydrological and meteorological data, contributing to more resilient ecosystems and agricultural practices (Wang et al. 2020).
- Urban Climate Resilience: AI-powered modeling tools help city planners design climate-resilient urban areas by simulating flood risks, heat islands, and infrastructure vulnerabilities under various climate scenarios (Kumar et al. 2023).

AI-driven technologies are used to detect and manage pollution in air, water, and soil. Predictive models identify pollution hotspots, while robotics and smart waste systems improve recycling efficiency and waste segregation. AI also aids in designing circular economy strategies by tracking material flows and optimizing resource recovery (World Economic Forum 2024). In agriculture, AI optimizes irrigation, pest control, and crop management through data-driven insights. Precision agriculture, supported by AI sensors and predictive analytics, reduces water usage and chemical inputs. Similarly, AI models forecast water scarcity and inform sustainable management of

hydrological systems. AI facilitates the real-time detection of contaminants in water systems. Machine learning algorithms analyze sensor and satellite data to identify chemical leaks, monitor wastewater discharge, and optimize treatment processes (Yaseen et al. 2020). AI-driven control systems in wastewater treatment plants adjust chemical dosing and aeration automatically, reducing energy consumption and operational costs. AI enhances the precision, scale, and timeliness of environmental monitoring and policy implementation. It enables early warning systems for disasters, supports data-driven environmental policies, and facilitates sustainable urban planning. The integration of AI with Internet of Things (IoT) and remote sensing technologies further expands its capacity for real-time environmental intelligence.

CONCLUSION

AI is not merely a technological advancement but a transformative force reshaping productivity and innovation paradigms. Its potential to augment human intelligence, accelerate discovery, and optimize systems makes it a cornerstone of modern economic progress. However, realizing these benefits requires strategic governance, investment in human capital, and inclusive innovation policies.

Harnessing AI's potential for productivity and innovation also demands governance frameworks that promote transparency, ethical standards, and equitable access. Policies that encourage responsible AI deployment can stimulate sustainable innovation ecosystems, balancing technological advancement with social welfare.

AI is a key driver of contemporary structural transformation, reshaping the organization of economies and the nature of work. While it offers unprecedented opportunities for innovation and productivity growth, it also poses challenges of inequality, skill mismatch, and social displacement. The extent to which AI contributes to sustainable and inclusive transformation depends on proactive policy measures, human capital development, and ethical governance frameworks.

Artificial intelligence offers unprecedented opportunities to transform environmental protection. By enhancing monitoring capabilities, optimizing

resource use, and informing sustainable policies, AI stands as a cornerstone of modern environmental governance. However, realizing its full potential requires addressing ethical, technical, and institutional challenges through collaborative and inclusive approaches.

AI stands as a powerful enabler of environmental protection through its capacity to support both climate change mitigation and adaptation. Its integration into energy systems, agriculture, urban planning, and disaster management enhances decision-making and resource efficiency. However, realizing its full potential requires cross-sectoral collaboration, investment in green AI infrastructure, and policies that ensure ethical and sustainable use of technology.

Despite its potential, the adoption of AI in environmental protection faces several challenges. High energy consumption associated with AI training models can offset environmental gains. Data biases, lack of transparency, and unequal access to AI tools risk deepening technological divides. Ethical governance frameworks are therefore essential to ensure that AI applications align with sustainability and social equity principles.

To maximize AI's benefits, international cooperation is needed to develop regulatory standards and promote open data sharing. Governments and institutions should invest in green AI-models and infrastructures optimized for low energy consumption. Capacity building, interdisciplinary research, and the integration of AI into environmental education are vital for long-term sustainability.

REFERENCES

Acemoglu, D., and P. Restrepo. 2019. "Artificial Intelligence, Automation, and Work." In *The Economics of Artificial Intelligence: An Agenda*, edited by A. Agrawal, J. Gans, and A. Goldfarb, 197–236. Chicago: University of Chicago Press.

Acemoglu, D., and P. Restrepo. 2020. "Automation and New Tasks: How Technology Displaces and Reinstates Labor." *Journal of Economic Perspectives* 34 (1): 3–30.

Acemoglu, D., and P. Restrepo. 2020a. "Robots and Jobs: Evidence from US Labor Markets." *Journal of Political Economy* 128 (6): 2188–2244.

Aghion, P., B. F. Jones, and C. I. Jones. 2019. "Artificial Intelligence and Economic Growth." In *The Economics of Artificial Intelligence: An Agenda*, edited by A. Agrawal, J. Gans, and A. Goldfarb, 237–282. Chicago: University of Chicago Press.

Agrawal, A., J. Gans, and A. Goldfarb, eds. 2019. *The Economics of Artificial Intelligence: An Agenda*. Chicago: University of Chicago Press.

Beam, Andrew L., and Isaac S. Kohane. 2018. "Big Data and Machine Learning in Health Care." *JAMA* 319 (13): 1317–1318. https://doi.org/10.1001/jama.2017.18391.

Bengio, Y., Y. Lecun, and G. Hinton. 2021. "Deep Learning for the Planet." *Nature* 593 (7857): 500–502. https://doi.org/10.1038/d41586-021-01251-0.

Bessen, J. E. 2019. "AI and Jobs: The Role of Demand." NBER Working Paper No. 24235. Cambridge, MA: National Bureau of Economic Research.

Brynjolfsson, E., and A. McAfee. 2014. *The Second Machine Age: Work, Progress, and Prosperity in a Time of Brilliant Technologies*. New York: W. W. Norton & Company.

Brynjolfsson, E., and A. McAfee. 2017. *Machine, Platform, Crowd: Harnessing Our Digital Future*. New York: W. W. Norton & Company.

Cockburn, I. M., R. Henderson, and S. Stern. 2018. "The Impact of Artificial Intelligence on Innovation." In *The Economics of Artificial Intelligence: An Agenda*, edited by A. Agrawal, J. Gans, and A. Goldfarb, 115–146. Chicago: University of Chicago Press.

Creutzig, F., S. Lohrey, X. Bai, A. Baklanov, R. Dawson, S. Dhakal, ... and D. Ürge-Vorsatz. 2022. "Harnessing Digitalization for Urban Sustainability." *Nature Sustainability* 5 (3): 200–210.

Davenport, Thomas, and Ravi Kalakota. 2019. "The Potential for Artificial Intelligence in Healthcare." *Future Healthcare Journal* 6 (2): 94–98. https://doi.org/10.7861/futurehosp.6-2-94.

Esteva, Andre, Brett Kuprel, Roberto A. Novoa, Justin Ko, Susan M. Swetter, Helen M. Blau, and Sebastian Thrun. 2019. "Dermatologist-Level Classification of Skin Cancer with Deep Neural Networks." *Nature* 542 (7639): 115–118. https://doi.org/10.1038/nature21056.

Jiang, Fei, Yong Jiang, Hui Zhi, Yi Dong, Hao Li, S. Ma, Y. Wang, Q. Dong, H. Shen, and Y. Wang. 2017. "Artificial Intelligence in Healthcare: Past,

Present and Future." *Stroke and Vascular Neurology* 2 (4): 230–243. https://doi.org/10.1136/svn-2017-000101.

Ifenthaler, D., and J. Y.-K. Yau. 2020. "Utilising Learning Analytics for Study Success: Reflections on Current Empirical Findings." In *Utilising Learning Analytics to Support Study Success*, edited by D. Ifenthaler and J. Yau, 1–9. Cham: Springer.

International Labour Organization (ILO). 2021. *The Role of Digitalization and AI in Shaping the Future of Work*. Geneva: ILO.

Korinek, A., and J. E. Stiglitz. 2021. "Artificial Intelligence and Its Implications for Income Distribution and Unemployment." In *The Economics of Artificial Intelligence: Implications for the Future of Work*, 349–390. Chicago: University of Chicago Press.

Krause, J., N. Perera, and M. Riordan. 2021. "Artificial Intelligence and Energy Efficiency: The Case of Industrial Processes." *Energy Policy* 156: 112378.

Kumar, R., S. K. Singh, and P. Kaur. 2023. "AI-Based Approaches for Climate-Resilient Urban Infrastructure Planning." *Environmental Modelling & Software* 166: 105667.

Latulippe, Karine, Caroline Hamel, and Dominique Giroux. 2020. "Social Health Inequalities and eHealth: A Literature Review with Qualitative Synthesis of Theoretical and Empirical Studies." *Journal of Medical Internet Research* 22 (4): e16463. https://doi.org/10.2196/16463.

Leslie, David, Atoosa Kashefi, Irene Kullmann, and Mike Shearer. 2021. "Artificial Intelligence and Health Inequities: The Need for Inclusive Design." *Journal of Health Services Research & Policy* 26 (4): 264–271. https://doi.org/10.1177/13558196211024762.

Liakos, K. G., P. Busato, D. Moshou, S. Pearson, and D. Bochtis. 2018. "Machine Learning in Agriculture: A Review." *Sensors* 18 (8): 2674.

Luckin, R., W. Holmes, M. Griffiths, and L. B. Forcier. 2016. *Intelligence Unleashed: An Argument for AI in Education*. London: Pearson Education.

Organisation for Economic Co-operation and Development (OECD). 2021. *Artificial Intelligence, Employment and Skills: The OECD AI Policy Observatory*. Paris: OECD Publishing.

Organisation for Economic Co-operation and Development (OECD). 2021. *The Impact of Artificial Intelligence on Productivity and Employment*. Paris: OECD Publishing.

Organisation for Economic Co-operation and Development (OECD). 2023. *Artificial Intelligence and Productivity: A New Growth Engine?* Paris: OECD Publishing.

Organisation for Economic Co-operation and Development (OECD). 2023. *OECD Employment Outlook 2023: Artificial Intelligence and the Labour Market*. Paris: OECD Publishing.

Price, W. Nicholson, and I. Glenn Cohen. 2019. "Privacy in the Age of Medical Big Data." *Nature Medicine* 25 (1): 37–43. https://doi.org/10.1038/s41591-018-0272-7.

Rajkomar, Alvin, Jeff Dean, and Isaac Kohane. 2019. "Machine Learning in Medicine." *New England Journal of Medicine* 380 (14): 1347–1358. https://doi.org/10.1056/NEJMra1814259.

Razzak, Muhammad Imran, Shehzad Khalid, and Guandong Xu. 2019. "Big Data Analytics for Preventive Medicine." *Neural Computing and Applications* 32: 4417–4451. https://doi.org/10.1007/s00521-018-03981-2.

Reichstein, M., G. Camps-Valls, B. Stevens, M. Jung, J. Denzler, N. Carvalhais, and Prabhat. 2019. "Deep Learning and Process Understanding for Data-Driven Earth System Science." *Nature* 566 (7743): 195–204.

Rolnick, D., P. L. Donti, L. H. Kaack, K. Kochanski, A. Lacoste, K. Sankaran, ... and Y. Bengio. 2022. "Tackling Climate Change with Machine Learning." *ACM Computing Surveys* 55 (2): 1–96.

Rolnick, D., P. L. Donti, L. H. Kaack, K. Kochanski, A. Lacoste, K. Sankaran, ... and Y. Bengio. 2022. "Tackling Climate Change with Machine Learning." *Nature Climate Change* 12 (5): 482–491. https://doi.org/10.1038/s41558-022-01377-7.

Russell, S., and P. Norvig. 2021. *Artificial Intelligence: A Modern Approach*. 4th ed. Hoboken, NJ: Pearson.

Topol, Eric J. 2019. *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again.* New York: Basic Books.

UNCTAD (United Nations Conference on Trade and Development). 2021. *Technology and Innovation Report 2021: Catching Technological Waves – Innovation with Equity*. Geneva: UNCTAD.

UNDP (United Nations Development Programme). 2022. *Human Development Report 2022: Uncertain Times, Unsettled Lives – Shaping Our Future in a Transforming World*. New York: UNDP.

Vinuesa, R., H. Azizpour, I. Leite, M. Balaam, V. Dignum, S. Domisch, ... and F. F. Nerini. 2020. "The Role of Artificial Intelligence in Achieving the Sustainable Development Goals." *Nature Communications* 11 (1): 233. https://doi.org/10.1038/s41467-019-14108-y.

Wang, J., J. Yang, and Y. Chen. 2020. "AI-Based Water Management Strategies under Climate Variability." *Journal of Hydrology* 589: 125215.

World Economic Forum (WEF). 2023. *The Future of Jobs Report 2023*. Geneva: WEF.

World Economic Forum (WEF). 2024. *Future of Jobs Report 2024*. Geneva: WEF.

World Economic Forum (WEF). 2024. *Harnessing AI for Environmental Sustainability*. Geneva: WEF.

Yaseen, Z. M., A. M. Al-Juboori, and O. Jaafar. 2020. "Artificial Intelligence–Based Models for Water Quality Prediction and Management: A Review." *Journal of Hydrology* 588: 125134. https://doi.org/10.1016/j.jhydrol.2020.125134

Yu, Kun-Hsing, Andrew L. Beam, and Isaac S. Kohane. 2018. "Artificial Intelligence in Healthcare." *Nature Biomedical Engineering* 2 (10): 719–731. https://doi.org/10.1038/s41551-018-0305-z.

ENSURING FAIR AND SUSTAINABLE DEVELOPMENT IN PRIVATE SECTOR HR: THE ROLE OF PUBLIC ADMINISTRATION IN THE AGE OF AI IN NORTH MACEDONIA

Ana Djima PhD candidate
Faculty of Law, University " St. Kliment Ohridski " – Bitola
anadzima72@gmail.com

Abstract

The quick adoption of Artificial Intelligence (AI) in private-sector human-resource (HR) practices creates important chances for efficiency, transparency, and effective workforce management in North Macedonia. However, it also brings challenges related to fairness, accountability, and following labor laws. This paper looks at the crucial role of public administration, especially the State Labour Inspectorate and other regulatory bodies, in making sure that AI-powered HR practices promote fairness and sustainable growth. Using ideas from administrative law, national laws, and examples from other countries, the study examines how institutional oversight can balance innovation with worker protection, stop discriminatory

practices, and encourage transparency. It argues that strong public-administration control is key to keeping private-sector digital changes in line with ethical, legal, and socio-economic goals. This ensures that technological progress supports fair and sustainable workforce development in North Macedonia.

Introduction

The global labour market is rapidly evolving as private-sector organisations increasingly adopt Artificial Intelligence (AI) technologies in human-resource (HR) management. AI is reshaping traditional HR functions such as recruitment, performance evaluation, onboarding and retention by automating processes, analysing large data sets, and enabling data-driven decision-making. Proponents argue that AI can enhance efficiency, reduce human bias, and personalise employee experiences, ultimately supporting organisational development and competitiveness (Marc PH. Prinz, 2024). However, these technological transformations also raise concerns about transparency, equity and the protection of fundamental labour rights, particularly when decisions with profound effects on workers are influenced by algorithmic systems (ILO, 2025).

In North Macedonia, the private sector plays a pivotal role in economic development, employment creation and labour market transformation. As companies modernise their HR practices, integrating AI tools could yield significant benefits for both employers and employees. Yet, without robust legal safeguards and administrative oversight, there is a risk that AI-driven HR practices may inadvertently undermine fairness, exacerbate inequalities, or conflict with established labour protections. AI systems trained on biased data, for example, may reproduce discriminatory patterns that disadvantage women, minorities or vulnerable groups, thereby challenging the state's commitments to justice and inclusive growth (Marc PH. Prinz, 2024).

Public administration bodies, particularly the State Labour Inspectorate, have a longstanding mandate to enforce labour laws, ensure compliance with

employment standards, and protect worker rights across all sectors of the economy. The Inspectorate supervises labour relations, employment contracts, workplace documentation, occupational safety and health, and compliance with collective agreements (Vukadinovic, Daskalov, Taseva, 2021).

In addition to its enforcement mandate, the State Labour Inspectorate engages in initiatives designed to enhance institutional capacity, with particular emphasis on preventing workplace discrimination and improving the detection and monitoring of labour exploitation. These activities reflect internationally accepted standards of labour inspection, most notably those articulated in the International Labour Organization's Labour Inspection Convention, which identifies inspection systems as a central mechanism for upholding lawful working conditions and protecting workers' rights (Labour Inspection Convention, 1947).

Despite the existence of these institutional competencies, North Macedonia does not currently operate under a distinct legal or regulatory framework specifically addressing the use of artificial intelligence in human resources management. Although general data protection rules and anti-discrimination laws continue to apply, they are increasingly insufficient to respond to emerging challenges associated with AI-driven employment practices. Issues such as the lack of transparency in algorithmic systems, reliance on automated decision-making processes, and the risk of embedded systemic bias require more sophisticated and context-specific supervisory approaches capable of addressing new modes of employment governance.

Against this backdrop, several fundamental issues arise. In what ways can public administration contribute to equitable and sustainable development amid the digitalisation of human resources management? To what extent are existing labour inspection models equipped to oversee AI-based HR technologies? And how should administrative law adapt to ensure that technological progress reinforces, rather than erodes, principles of fairness and labour protection in North Macedonia?

This study argues that robust public administrative supervision remains indispensable in guiding the adoption of artificial intelligence within private-

sector HR practices toward outcomes that support social equity and sustainable development. By analysing the scope, capabilities, and limitations of administrative oversight mechanisms, the paper demonstrates how regulatory frameworks can be recalibrated to balance innovation with the effective protection of workers' rights in an increasingly digital labour environment.

Legal and Administrative Framework in North Macedonia

The primary legal instrument governing employment relations in North Macedonia is the Labour Relations Law, which regulates employment contracts, working conditions, termination procedures, and the rights and obligations of employers and employees (Labour Relations Law,2017). The law sets mandatory employment standards that remain applicable irrespective of whether human decision-makers or automated technologies are used in human resource management. Principles of equal treatment, lawful recruitment, and fair termination therefore continue to serve as key criteria for evaluating the legality of AI-supported HR practices.

In addition to labour legislation, the Law on Prevention and Protection against Discrimination provides a comprehensive framework prohibiting discrimination in employment and labour relations on grounds such as gender, ethnicity, age, disability, and other protected characteristics. The law applies to both public and private subjects and imposes positive obligations on state authorities to prevent discriminatory practices. While the legislation does not explicitly address algorithmic discrimination, its broad scope allows it to be interpreted as covering indirect and structural discrimination resulting from biased AI systems used in recruitment, evaluation, or dismissal processes (Law on Prevention and Protection against Discrimination, 2019).

The processing of personal data in employment and other contexts is regulated by the Law on Personal Data Protection, which is harmonised with European data protection standards. The law establishes key principles for personal data handling, including the requirements of lawfulness, transparency, and data minimisation, and safeguards the rights of data subjects with respect to access, correction, and objection. These provisions apply to all automated processing activities, including those that may involve

profiling or large-scale data handling. (Law on Personal Data Protection, 2020).

The Role and Mandate of the State Labour Inspectorate

The State Labour Inspectorate is the principal administrative body responsible for supervising the implementation of labour legislation in North Macedonia (Vukadinović, Daskalov, Taseva, 2021). Its mandate includes monitoring compliance with employment contracts, working time regulations, occupational safety and health standards, and the application of collective agreements (Vukadinović, Daskalov, Taseva, 2021). These functions are consistent with the principles set out in the ILO Labour Inspection Convention, 1947, which positions labour inspection as a key mechanism for enforcing labour rights internationally (Labour Inspection Convention, 1947).

Beyond traditional labour standards enforcement, the Inspectorate also plays a role in preventing labour exploitation and discrimination, often in cooperation with other public institutions and social partners (OSFWB, 2022). Capacity-building initiatives and international cooperation efforts have sought to strengthen its ability to detect complex violations, including those affecting vulnerable groups. The increasing use of AI in HR practices presents new challenges for the Inspectorate, particularly in terms of accessing and understanding algorithmic systems, evaluating automated decision-making processes, and proving discriminatory or unlawful outcomes generated by opaque technologies (OSFWB, 2022).

While the Inspectorate's legislative mandate is conceptually broad enough to cover potential labour violations arising from the use of AI in employment, its practical effectiveness is conditioned by the clarity of the underlying legal framework, the availability of technical expertise, and sufficient institutional resources (ILO, 2025). In the absence of legal provisions that explicitly address algorithmic systems and automated decision-making in employment, inspectors must necessarily interpret and apply existing legal categories and procedural tools designed for traditional labour issues, which may not be well-adapted to the challenges posed by complex digital systems (Marc PH.

Prinz, 2024). This gap highlights both the resilience of current administrative oversight mechanisms in addressing conventional violations and their limitations when confronted with emerging forms of digital labour governance, underscoring the need for targeted legal reform and institutional adaptation to ensure the effective protection of workers' rights in an increasingly digitalised labour market (ILO, 2025).

Risks of algorithmic bias, opacity, and labour exploitation

The integration of artificial intelligence into human resource management can unintentionally replicate and amplify existing biases, particularly when AI tools are trained on historical employment data that reflects existing inequalities, thereby disadvantaging certain groups such as persons with disabilities in hiring decisions (International Labour Organization, AI's Double-Edged Sword, 2025). AI deployment also introduces opacity, where the logic behind automated decisions may be unclear to both workers and regulators, undermining transparency and complicating efforts to contest adverse outcomes; many systems depend on biased or incomplete data and opaque programming, which can reinforce inequality rather than reduce it (International Labour Organization, ILO study highlights the need for human resource managers to better understand AI-related risks, 2025).

Limitations of existing labour inspection and enforcement mechanisms

Existing labour inspection regimes are principally geared toward traditional compliance areas — such as workplace conditions, contracts, and occupational safety — and are not structured to evaluate the inner workings of complex algorithmic systems that make automated HR decisions. AI tools often function through proprietary code and large datasets, and the logic behind their outputs can be difficult to access or explain, making it challenging for regulators to assess how decisions are reached under current frameworks where transparency and explainability are limited (OECD Employment Outlook 2023; OECD Using AI in the Workplace 2024). In addition, without AI-specific enforcement mandates or procedural powers, non-AI-specific data protection and anti-discrimination laws remain relevant but require interpretation and creative application to address AI-related

harms, resulting in legal uncertainty when attempting to sanction discriminatory automated HR practices (OECD Employment Outlook 2023; OECD Using AI in the Workplace 2024).

Public administration responses to AI-Driven HR systems

Public administration must adopt responses that enhance transparency, accountability, and stakeholder participation in AI governance. One area of response involves promoting social dialogue and worker representation in decisions about AI adoption and oversight, which has been shown to encourage more inclusive and equitable outcomes when AI is used in employment settings (global case studies of social dialogue on AI and algorithmic management). Additionally, clear regulatory guidance on when human judgment must prevail over automated decisions and frameworks requiring explainability and accountability — including pre-deployment testing and ongoing monitoring — can help protect workers' rights while allowing responsible innovation to continue (OECD, 2023)

Comparative perspectives, policy implications, and conclusions

International and comparative approaches to AI regulation in employment

Comparative policy frameworks demonstrate various ways countries are addressing the legal and ethical risks of AI at work. For example, the OECD Employment Outlook's chapter on trustworthy AI highlights how some jurisdictions are developing workplace policies that ensure transparency and accountability in the use of AI, clarifying responsibility across the AI value chain and ensuring workers are informed when AI influences employment decisions (OECD Employment Outlook: trustworthy AI). In the European Union, proposed AI regulation classifies AI systems used for recruitment and worker management as high-risk, imposing requirements for accuracy, monitoring, explainability, and human oversight to mitigate discriminatory effects, and safeguarding the primacy of human decision-making (EU AI high-risk classification). (OECD, 2023)

Policy recommendations for fair and sustainable HR digitalisation

To ensure fair and sustainable digitalisation of HR practices in North Macedonia, policymakers should consider updating labour and anti-discrimination laws to explicitly address algorithmic decision-making and to clarify enforcement powers for administrative inspectors. Embedding requirements for AI transparency, worker consultation, algorithmic impact assessments, and accessible complaint mechanisms should be part of a coherent regulatory strategy that balances innovation with protection of worker rights (OECD policy action on AI in workplaces). Public authorities may also benefit from investing in technical training for inspection agencies and establishing inter-agency cooperation between labour, equality, and data protection bodies to oversee AI governance comprehensively. The promotion of social dialogue on AI deployment can further reinforce equitable outcomes and help align technological adoption with principles of decent work (OECD,2024)

Conclusion

Artificial intelligence in human resource management presents both opportunities for increased efficiency and significant risks to fairness, transparency, and worker protections. Without tailored governance frameworks, the use of AI in employment can entrench inequalities and evade existing inspection and enforcement methods. Strengthening public administration capacities and updating legal frameworks to include protections against algorithmic bias and opacity will be essential for ensuring that AI adoption supports equitable and sustainable labour justice in North Macedonia. The experiences of OECD countries and proposed high-risk regulatory regimes in the EU provide practical models for designing legal and administrative responses that uphold fundamental rights while encouraging responsible innovation.

References

1. Council of Europe – GRETA, 2023 Access to justice and effective remedies for victims of trafficking in human beings https://rm.coe.int/greta-evaluation-report-on-north-macedonia-third-evaluation-round/1680aaa573

2. International Labour Organization, 2025 - AI's double-edged sword: A new frontier for employment of people with disabilities? https://www.ilo.org/resource/article/ais-double-edged-sword-new-frontier-employment-people-disabilities?lang=en%20https://www.ilo.org/resource/news/ilo-study-highlights-need-human-resource-managers-better-understand-ai-related-risks/lang%E2%80%94en/index.htm

3. Inernational Labour organization, 2025 - study of the need of human resource managers to better understand AI-related risks https://www.ilo.org/resource/news/ilo-study-highlights-need-human-resource-managers-better-understand-ai

4. OECD, 2023 - Employment outlook https://www.oecd.org/en/publications/oecd-employment-outlook-2023_08785bba-en/full-report/ensuring-trustworthy-artificial-intelligence-in-the-workplace-countries-policy-action_c01b9e49.html

5. OECD,2024 Using AI in the workplace opportunities, risks and policy responses https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/03/using-ai-in-the-workplace_02d6890a/73d417f9-en.pdf

6. Open Society Foundations – Western Balkans, 2022 https://osfwb.org/skopje-office/

7. Marc PH. Prinz, 2024 - AI In Human Resources https://www.mondaq.com/new-technology/1451184/part-13-ai-in-human-resources

8. Republic of North Macedonia. Law on Labour Relations, 2017

9. Republic of North Macedonia. Law on Personal Data Protection, 2020.

10. Republic of North Macedonia. Law on Prevention and Protection against Discrimination, 2019

11. Vukadinović, N., F. Daskalov, and S. Taseva. Enhancing Control Systems Including Legal Modifications for Application of Discretionary Rights in Licensing and Employment. Skopje, 2021 https://dksk.mk/wp-content/uploads/2022/03/Enhancing-Control-

Systems-including-Legal-Modifications-for-Application-of-Discretionary-Rights-in-Licensing-and-Employment.pdf

IN THE ERA OF AI: COMMUNICATION AND DIGITAL TRANSFORMATION IN PUBLIC ADMINISTRATION

Gresa MAKOLLI
Quality Assurance Coordinator at AAB College
PhD. student at University "St. Clement Ohridski" Bitola, North Macedonia
gresa.makolli@universitetiaab.com

Abstract

Considering the great momentum that the development of technology has received, its use is inevitable, including the emergence of artificial intelligence, which are significantly affecting the way public administrations communicate and function. This paper aims to address the combination of institutional communication, digitalization and artificial intelligence in public administration, with a particular focus on improving efficiency, transparency and citizen involvement.

The purpose of the paper is to analyze how digital transformation and artificial intelligence tools are affecting communication processes in public administration, as well as to identify the opportunities and challenges associated with these developments. The paper brings a new perspective on how technology can be used to improve the provision of public services, while respecting the principles of administrative law and institutional ethics.

The study methodology is qualitative, using the analysis of scientific literature and strategic documents, as well as the study of concrete cases in the context of public administration in Kosovo and the region. This paper will also use comparative analysis and critical evaluation of AI implementation practices in administrative communication.

The expected results are expected to highlight the need for balanced regulation between technological innovation and institutional accountability, as well as the importance of preparing civil servants for the digital age.

Introduction

The rapid advancement of digital technologies has fundamentally reshaped the way governments operate, communicate, and deliver services. Public administrations around the world are undergoing a profound transformation driven by digitalization and artificial intelligence (AI), which are redefining institutional communication and citizen interaction. In this evolving environment, communication is no longer a one-way process but an interactive exchange that requires transparency, responsiveness, and accountability.

This paper explores how AI and digital transformation are influencing communication within public administration, particularly in terms of improving efficiency, accessibility, and citizen participation. It examines the dual nature of technological innovation—its potential to strengthen governance and the challenges it poses to privacy, ethics, and institutional trust. Through a qualitative analysis of policy documents, international reports, and real-world examples from Kosovo and the European Union, this study aims to provide an integrated perspective on how technology can serve as both a tool for modernization and a test for administrative responsibility. Ultimately, the study argues that the success of digital governance depends on maintaining a   delicate balance between innovation and institutional accountability. By combining human expertise with technological capabilities, public administration can build a more transparent, efficient, and citizen-centered system that upholds the principles of good governance in the digital age.

The Role of Institutional Communication in the Digital Age

Communication is a fundamental pillar of society, serving as the primary means through which citizens are informed about policies, public services, and governmental decisions. Effective communication enhances transparency, fosters public trust, and strengthens the relationship between institutions and the population. In the digital era, public institutions have the opportunity to leverage modern technologies to expand their reach, facilitate

two-way dialogue with citizens, and improve the overall effectiveness of service delivery.

To maximize outreach and impact, communication must be approached strategically, presenting information in a structured and accessible manner that resonates with diverse audiences. (OECD, 2021, p. 14).

Proactive transparency is a crucial component of institutional communication in the digital era. Instead of only releasing information after decisions are made, public institutions ought to provide continuous channels for updating the public as public policies and services are being implemented. This entails reporting on institutional action on a regular basis, providing clear and intelligible explanations of decisions, and utilizing a variety of communication channels to provide simple access to information. A key component of successful governance is open and transparent communication, which fosters confidence between the public and the government.

Public entities' interactions with citizens have changed as a result of digital communication. Digital communication provides options for two-way discourse, rapid information transmission, and direct access to citizens, in contrast to traditional communication, which is frequently one-way and centralized. As a result, individuals' demands and concerns can be addressed more quickly, boosting openness and confidence in government agencies. Public communication should "provide information, listen and respond to citizens in the service of the common good," (OECD, 2021), which highlights the significance of more transparent and participatory communication.

Effective communication not only conveys information but also reinforces professionalism, strengthens institutional structures, and supports governance arrangements. When applied strategically, communication enhances trust between citizens and public institutions and contributes to the consolidation of democratic governance. Consequently, well-managed communication plays a critical role in improving public services and promoting the principles of good governance. (OECD, 2021)

In fulfilling the mission of public administration, building and maintaining trust is essential. Citizens' perceptions and evaluations of information quality are heavily influenced by the level of trust they place in the source.

Therefore, trust serves as a fundamental component for effective communication and for strengthening the legitimacy and credibility of public institutions." (OECD, 2021),

Institutional communication constitutes a strategic process that goes beyond mere information dissemination, requiring professional structures and sound governance. As highlighted in the OECD report (2021), public communication plays a fundamental governmental role by not only providing information but also listening to and responding to citizens in the service of the public interest. To foster trust and enhance institutional credibility, public institutions must position themselves not only as information providers but also as active listeners, engaging meaningfully with citizens' concerns and feedback.

Recent studies indicate that effective communication strategies should be grounded in empirical data, including a thorough understanding of audiences, their behaviors, needs, and perceptions. The OECD report emphasizes that failing to target specific citizen groups results in missed opportunities for wider engagement and influence. Moreover, communication strategies should incorporate impact assessment, evaluating not only the dissemination of information but also how such communication affects citizens' perceptions, participation, and behaviors.

Beyond merely sharing information, effective public communication practices should assess how the public receives and interacts with institutional content. This includes evaluating citizens' understanding, reactions, and engagement. Insights from these evaluations enable institutions to adjust their messaging, foster greater public involvement, and improve the quality of the services provided. Consequently, institutional communication functions as a strategic instrument to enhance the accountability, transparency, and efficiency of public administration.

Another critical aspect of public communication is addressing disinformation. In the digital era, the spread of false or manipulated information can undermine both institutional effectiveness and public trust. To tackle this challenge, the OECD has developed Good Practice Principles for institutional responses to disinformation. These principles emphasize

accuracy, transparency, accountability, and clarity as essential components of effective communication strategies. (OECD. ,2023).

Digital Transformation and Its Impact on Public Administration

In European and global countries, within the framework of the modernization of governance and public administration reforms, Digital Transformation has become one of the strategic priorities. In addition to representing the use of technology, it also represents the way public institutions function in the functioning of the structur

One of the most significant administrations influencing public opinion in the twenty-first century is digital change. It entails the incorporation of digital technology into every facet of its operations, regulating the provision of administrative services, communication with me, and decision-making processes. In essence, the goal of digital transformation is to improve public services' information, efficiency, and transparency while also improving institutional performance and accessibility.

Many facets of government are significantly impacted by digital transformation, which enhances institutional performance and efficiency. This procedure aids in streamlining administrative procedures and enhancing openness in the delivery of public services. (Yang, C., Gu, M., & Albitar, K. 2024).

The enhancement of the public service delivery process is among the most obvious outcomes of the digital revolution. Citizens may complete administrative tasks online, apply for papers, check the status of requests, and get answers without having to visit a government office thanks to e-government platforms and websites like eKosova in Kosovo. In addition to saving citizens' time, this lessens the administrative load on officials and lowers the likelihood of mistakes or poor administration. In addition, the use of an electronic system allows the collection of accurate data for analysis and planning, making decision-making more fact-based and effective.

In addition to the aforementioned benefits, eKosova has improved citizen-institutional contact by making administrative processes more transparent. Citizens can track the status of their requests in real time, while institutions can use data analytics to identify bottlenecks or difficulties in service

delivery. This not only improves efficiency, but also allows for ongoing improvement of public services and strengthens citizen trust. According to the (OECD, 2021), digital governance that enables proactive transparency and direct access for citizens to public services and information helps to more accountable and participatory governance.

Digital transformation has a significant impact on the openness and accountability of public organisations. The online publication of reports, statistics, and judgements allows citizens to more actively monitor government actions and regulate public services. Concrete examples in the European Union include the Europa.eu portal and e-government platforms in countries such as Estonia and Finland, where citizens can access all administration services, from personal registration to public consultations on new policies. This fosters trust and civic engagement, resulting in a culture of transparent and participatory governance,

The automation of administrative processes is also an important dimension of digital transformation. The use of technologies such as artificial intelligence, data analysis algorithms and robotics is allowing many tasks to be performed automatically, thereby freeing up administrative staff to carry out more strategic activities. Several institutions in the European Union have started using chatbots to assist citizens with general information and instructions for various applications, effectively speeding up service and reducing waiting times.

In public administration, the use of artificial intelligence and chatbots has the potential to increase service efficiency and transparency. According to an OECD study, Artificial Intelligence can accelerate the journey to digital governance, thus offering opportunities for increased accountability and efficiency in governance (OECD, 2025). Also, a study shows that chatbots can improve citizens' access to services and, through the automation of routine tasks, create public value.

In general, numerous challenges surround the digital transformation. To effectively face this process, many public institutions lack the human and technological capacities. In addition, the lack of digital skills, privacy, and data security remain major obstacles to achieving fully digital governance.

So, digital transformation has brought numerous improvements in transparency, efficiency and quality of public services. But it all depends on how public institutions address the challenges of data security, digital inclusion and capacity challenges. Through a sustainable and comprehensive approach, digital transformation can be transformed into a true tool for responsible and modern governance.

Advantages of Artificial Intelligence in Communication in Public Administration

Artificial intelligence (AI) is changing how governmental institutions interact with citizens and manage information. One of the primary benefits is enhanced efficiency in service delivery. Chatbots, virtual assistants, and automated systems can give information in real time by answering residents' questions, helping them through applications, and streamlining administrative processes. This lessens the pressure on administrative staff and shortens citizen wait times, ultimately enhancing service quality.

 The implementation of artificial intelligence in public administration is a clever strategy that enhances operational efficiency by automating routine and administrative activities, allowing public officials to focus on higher-value strategic duties. (VASS Company, 2024).

Artificial intelligence tools, such as chatbots and virtual assistants, facilitate fast and effective communication with citizens. In Midland, USA, two AI-driven platforms—SeeClickFix and Ask Jacky—have been introduced to enhance interaction and service efficiency between city officials and residents. These tools enable citizens to report issues and receive real-time responses, boosting engagement and satisfaction. (Dias, 2025).

AI also allows for the personalization of public services. Data analysis algorithms can identify the specific needs of individuals, recommend tailored services, and anticipate future demands. This makes institutions more responsive to the actual needs of citizens. For instance, several EU countries have started using AI to analyze citizen requests and improve services in areas such as transportation, healthcare, and public documentation. (Misuraca & van Noordt, 2020)

Artificial intelligence can be employed to streamline routine administrative processes, such as document handling and responding to standard inquiries. This approach not only reduces the consumption of time and resources but also allows public administration personnel to dedicate greater attention to more complex and strategically important tasks. Research conducted by Appian indicates that the integration of AI in these areas can substantially improve operational efficiency within public institutions.

Artificial intelligence has the potential to improve public service delivery by automating administrative processes, tailoring services to individual needs, and promoting greater transparency. (Aboud, 2025)

Another benefit of AI is its capacity to support data-driven decision-making. By processing large volumes of information rapidly, AI systems can generate forecasts and provide recommendations that assist decision-makers. This enables real-time problem identification and more strategic planning of public services. For instance, analyzing data from public services can reveal areas in need of improvement, enhancing both efficiency and transparency. Moreover, AI fosters greater citizen engagement and interaction with public institutions. Digital platforms equipped with AI can collect feedback, evaluate suggestions, and adjust services to better meet citizens' actual needs. Such applications strengthen trust in institutions and contribute to more transparent and accountable communication.

The integration of AI into public services can enhance citizens' experiences by delivering quicker and more precise services. According to a UK Government article published on 14 May 2025, a government-created tool named "Consult" has been used to analyze public feedback from consultations for the first time, significantly improving efficiency. This AI-powered tool evaluates and organizes key themes from over 2,000 responses, allowing officials to concentrate on the strategic elements of policymaking. Its implementation has saved approximately 75,000 working days annually, reducing the need for manual review and cutting staff expenses by up to £20 million each year. This initiative reflects the government's broader strategy to leverage technology for faster, more efficient public service delivery. (UK Government, 2025)

The adoption of AI in public administration facilitates easier access to information and decision-making. AI is particularly useful in administrative communication, supporting tasks such as translations, text simplification, and summarization, thereby streamlining workflows. Overall, AI presents a significant opportunity for public institutions to become more efficient, accountable, and citizen-focused. Benefits include process automation, personalized services, data-driven decision-making, and enhanced engagement with citizens.

However, despite these advantages, implementing AI in public administration also brings notable challenges. Key concerns involve privacy and the protection of personal data, the necessity of staff training, and overcoming resistance to change. Additionally, the costs associated with deploying and maintaining AI systems can be substantial, and automated decision-making may raise ethical issues or compromise transparency if not properly overseen. These challenges highlight the need for a cautious approach and well-defined strategies for integrating AI into administrative processes.

In addition, the absence of interoperability among various public administration systems can limit the effective deployment of AI. The European Interoperability Framework (EIF) offers guidance for developing interoperable digital public services, addressing legal, organizational, semantic, and technical dimensions. (European Commission, 2020)

Shortcomings and Risks of Artificial Intelligence in Administrative Communication

Despite the significant benefits of artificial intelligence, its application in administrative communication poses important challenges and potential risks. A central issue is the limited transparency of AI-driven decision-making processes. Often, AI systems operate as "black boxes," making it difficult for both citizens and public officials to fully understand how decisions are generated. This lack of clarity can reduce trust in public institutions and hinder effective oversight. To address these concerns, the European Union's Artificial Intelligence Act imposes strict transparency requirements, especially for high-risk AI systems, mandating clear information for users about the system's purpose and functionality, alongside

mechanisms for human supervision of automated decisions. (European Parliament, 2025).

Another significant challenge concerns the protection of personal data and privacy. AI systems rely on large volumes of personal information to operate efficiently, which makes robust security measures and clear legal frameworks essential. Inadequate safeguards could lead to misuse or unintended disclosure of sensitive data, undermining citizens' trust and exposing public institutions to legal risks. The European Union's General Data Protection Regulation (GDPR) establishes that personal data processing must be lawful, fair, and transparent. It also requires that individuals are properly informed about how their data is collected and used, while guaranteeing their rights, including access, deletion, and the ability to object to processing. (European Parliament, 2020)

Digital inequality represents another important challenge. Individuals who lack access to modern technologies or the necessary digital skills may be excluded from the benefits of automated public services, potentially exacerbating social disparities. This underscores that AI cannot entirely substitute human interaction, and its deployment must be complemented by inclusive strategies to ensure all citizens can effectively engage with digital services.

An additional challenge concerns the overreliance on technology. Excessive dependence on AI tools may result in public servants losing essential skills needed to manage effective communication and maintain relationships with citizens. Overuse of automated systems can diminish personal interaction, limit flexibility in handling exceptional cases, and ultimately affect the overall quality of public services. (UNESCO, 2021)

A major challenge in applying AI within public administration is the potential reinforcement of existing biases through supposedly neutral algorithms. When AI systems are trained on biased or incomplete datasets, they may inadvertently perpetuate discrimination, leading to unfair outcomes for specific individuals or groups. This can affect areas such as recruitment processes or the provision of public services. To mitigate these risks, it is essential to design and implement algorithms that are fair and impartial,

84

while continuously monitoring AI system performance to detect and correct any form of bias. (UNESCO, 2021)

To address these challenges, it is crucial to establish clear ethical guidelines, implement robust monitoring systems, and provide continuous training for public administration staff. Ensuring a balanced approach that integrates AI with human oversight, transparency, and strict data protection measures can help mitigate potential risks and promote the responsible use of technology in administrative communication.

In summary, the adoption of AI in public administration presents significant opportunities to enhance efficiency, service quality, and citizen engagement. However, it also brings challenges, including concerns over transparency, privacy, digital inequality, overreliance on technology, and algorithmic bias. Effective management, strategic planning, and ethical implementation are essential to fully leverage the benefits of AI while safeguarding both citizens and institutional integrity.

Balancing Technological Innovation with Institutional Accountability

In the current era of artificial intelligence (AI) and digital transformation, public administrations are confronted with the challenge of reconciling innovation with accountability. While advanced technologies provide new opportunities to modernize services and enhance citizen engagement, they also pose risks concerning transparency, ethical standards, and institutional oversight. Ensuring that technological advancements adhere to core public sector principles—such as efficiency, legality, and integrity—is crucial for maintaining sustainable and trustworthy governance in the digital age. (European Commission, 2023).

The digital transformation of public services, driven by technologies such as artificial intelligence, automation, and e-governance platforms, has significantly reshaped the relationship between institutions and citizens. In Kosovo, the eKosova platform has improved service delivery by enabling citizens to submit applications online and access administrative information in real time. Likewise, within the European Union, digital governance initiatives emphasize the use of interoperable data systems and social media

channels to strengthen transparency and citizen participation (Interoperable Europe, 2024).

These technological advancements contribute to enhanced efficiency, reduced bureaucratic procedures, and more responsive public institutions. Nevertheless, it is essential that such innovations are supported by robust legal and ethical frameworks to ensure that they serve the public interest effectively.

Institutional accountability remains a fundamental element of democratic governance, and it must encompass all AI-driven and automated decision-making systems. The EU Artificial Intelligence Act (2024) mandates that algorithmic systems employed by public institutions be transparent, explainable, and subject to audit. Consequently, each automated process should provide comprehensive documentation of data sources, decision-making logic, and mechanisms for handling errors (European Parliament, 2024).

In the absence of proper oversight, digital systems may perpetuate bias, spread misinformation, or produce arbitrary outcomes, ultimately eroding public trust in governmental institutions.

Equally important are ethical considerations in the deployment of digital tools within public administration. Civil servants must be equipped to identify and manage ethical challenges related to data protection, algorithmic bias, and equitable access to services. Responsible governance requires that technology be applied fairly, automated decisions comply with administrative regulations, and citizens' privacy be safeguarded in accordance with frameworks such as the General Data Protection Regulation (GDPR) (GDPR.eu, 2024).

Integrating ethical principles into AI implementation not only helps prevent discrimination but also reinforces citizens' confidence in the legitimacy and reliability of digital public services.

The effectiveness of digital innovation in public administration is closely linked to the readiness and capabilities of civil servants. Ongoing professional development in digital competencies, data management, and compliance with legal frameworks is critical (OECD, 2023). Public employees must be able to combine technological tools with traditional

administrative processes to deliver services that are both high-quality and accountable.

Investing in human capital not only enhances institutional resilience but also ensures that technological innovation supports governance rather than replacing it.

Sustaining a balance between innovation and accountability requires clearly defined policy frameworks. Legal regulations should delineate institutional responsibilities, safeguard data privacy, and establish monitoring and auditing mechanisms for algorithmic systems (European Commission, 2023). By providing these regulatory boundaries, governments can encourage innovation while maintaining citizens' rights and trust.

Achieving a responsible integration of technological innovation in governance hinges on training civil servants, implementing AI ethically, and developing robust regulatory frameworks. When these elements are in place, technology functions not merely as a tool for efficiency but as a mechanism to strengthen transparency, integrity, and public trust in institutions.

Conclusion

Artificial intelligence and digital transformation have become essential drivers of change in public administration. Their integration offers remarkable opportunities for improving efficiency, transparency, and citizen engagement. However, these advancements also bring new risks related to data protection, ethical governance, and digital inequality. As public institutions adopt AI-driven tools, they must ensure that innovation does not compromise accountability, fairness, or the human dimension of governance. The findings of this study highlight that sustainable digital governance requires more than technological investment—it requires a culture of responsibility, ongoing staff training, and clear regulatory frameworks. Ethical standards, data security, and inclusiveness must guide every step of digital transformation to ensure that citizens remain at the center of institutional innovation. Balancing technological progress with human oversight is the cornerstone of trustworthy public administration. When managed responsibly, artificial intelligence can become a catalyst for better

communication, stronger public trust, and more responsive governance, marking a decisive step toward a modern and ethical public sector.

References

OECD. (2021). *OECD report on public communication* (p. 14). OECD Publishing. https://www.oecd.org/content/dam/oecd/en/publications/reports/2021/12/oecd-report-on-public-communication_b74311bc/22f8031c-en.pdf

OECD. (2023). *Good practice principles for public communication responses to mis- and disinformation*. OECD Publishing. https://www.oecd.org/en/publications/good-practice-principles-for-public-communication-responses-to-mis-and-disinformation_6d141b44-en.html

Yang, C., Gu, M., & Albitar, K. (2024). Government in the digital age: Exploring the impact of digital transformation on governmental efficiency. *Technological Forecasting and Social Change, 208*, https://www.sciencedirect.com/science/article/pii/S0040162524005201?utm_source=chatgpt.com

OECD. (2025). *Governance*. Organisation for Economic Co-operation and Development. https://www.oecd.org/en/topics/governance.html

OECD. (2025). *Governing with artificial intelligence: The state of play and way forward in core government functions*. OECD Publishing. https://www.oecd.org/en/publications/2025/06/governing-with-artificial-intelligence_398fa287/full-report.html

Dias, L. (2025). *City says new AI tools have boosted communication with residents*. Midland Reporter-Telegram. https://www.mrt.com/news/article/midland-ai-seeclickfix-ask-jacky-20257698.php

AI Watch: Artificial Intelligence in public services. EUR 30255 EN. Publications Office of the European Union https://op.europa.eu/en/publication-detail/-/publication/4c72dd88-bcda-11ea-811c-01aa75ed71a1/language-en

Aboud, B.-O., Hernandez, Z. T., & Cardoso Espinosa, E. O. (2025). Advancing smart public administration: Challenges and benefits of artificial

intelligence. *Heliyon Public Administration*, *1*(1), 1–15.
https://www.sciencedirect.com/science/article/pii/S2664328625000403

UK Government. (2025, May 14). *Government built Humphrey AI tool to review responses to consultation for first time in bid to save millions*. https://www.gov.uk/government/news/government-built-humphrey-ai-tool-reviews-responses-to-consultation-for-first-time-in-bid-to-save-millions?utm_source=

European Commission. (2020). *The European Interoperability Framework in detail*. https://interoperable-europe.ec.europa.eu/collection/iopeu-monitoring/european-interoperability-framework-detail

European Parliament. (2025, February 19). *EU Artificial Intelligence Act: First regulation on artificial intelligence*. https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence

European Parliament, (2020). The impact of the General Data Protection Regulation (GDPR) on artificial intelligence. European Parliamentary Research Service. https://www.europarl.europa.eu/thinktank/en/document/EPRS_STU(2020)641530

UNESCO. (2021, November 23). *Recommendation on the Ethics of Artificial Intelligence*. https://www.unesco.org/en/artificial-intelligence/recommendation-ethics?utm_source=

European Parliament. (2023, June 8). *EU AI Act: First regulation on artificial intelligence*. https://www.europarl.europa.eu/topics/en/article/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence?utm_source=

European Parliament. (2024, March 13). *Artificial Intelligence Act: MEPs adopt landmark law* [Press release]. https://www.europarl.europa.eu/news/en/press-room/20240308IPR19015/artificial-intelligence-act-meps-adopt-landmark-law?utm_source=

CRIMINAL LAW ASPECTS OF FRAUD IN THE CONCLUSION OF CONTRACTS – ANALYSIS OF THE BOUNDARY WITH CIVIL LIABILITY, WITH AREVIEW OF JUDICIAL PRACTICE IN THE REPUBLIC OF NORTH MACEDONIA

Biljana Dolevska, MA
LLM student,
Faculty of Law Kicevo
University "St. Clement Ohridski "Bitola, North Macedonia
bdolevska@yahoo.com

Nikola Tuntevski, PhD
Prof. Dr.sc.
Faculty of Law Kicevo
University "St. Clement Ohridski "Bitola, North Macedonia
nikola.tuntevski@uklo.edu.mk

Abstract

The frequent occurrence of fraud in the conclusion of contracts undermines the stability of legal transactions, highlighting the need to distinguish clearly between civil and criminal liability. This distinction is essential for rendering just and lawful judicial decisions and maintaining a functional legal system.

This paper analyzes the legal framework and judicial practice in the Republic of North Macedonia, complemented by comparative international experiences and statistical data, with the aim of proposing guidelines for the appropriate legal qualification of fraud. The objective is to prevent the over-criminalization of private legal relations and to enhance legal certainty and predictability.

Particular emphasis is placed on developing an integrated and balanced legal approach that aligns with technological advancements, safeguards democratic values, and contributes to improving citizens' quality of life.

Keywords: fraud, contracts, civil liability, criminal liability, Republic of North Macedonia, comparative analysis.

Introduction

In modern society, where contracts serve as the fundamental instruments for regulating the rights and obligations between parties, fraud is one of the most complex phenomena in contractual relations. It directly undermines the core principles of contract law —the principles of good faith (*bona fide*) and fair dealing (UN Convention on Contracts for the International Sale of Goods, 1980, Article 1.7).

Fraud not only calls into question the validity of the contract but also gives rise to legal implications that require careful assessment and legal qualification of the specific conduct —it pertains to civil law issues or constitutes an act criminalizing fraud as a criminal offense.

Criminal Law Framework

The criminal offense of fraud constitutes a breach of trust and legal certainty in legal relations. Unlike civil sanctions, which are reparative in nature, criminal liability for fraud serves both preventive and punitive functions.

According to the Criminal Code of the Republic of North Macedonia, fraud is defined as an act whereby the perpetrator, "with the intention of obtaining unlawful property gain for themselves or another, misleads someone by false representation or concealment of facts, or keeps them in error, thereby inducing them to do or omit something to the detriment of their own or another's property" (Criminal Code of the RNM, Article 247).

Based on the elements of the criminal offense, the protected legal interests include public order, legal certainty and property integrity. The object of the offense is the property of the injured party, as well as the right to free will in contractual relations and general legal certainty.

This criminal offense comprises the following objective and subjective elements:

Objective elements:
- Act of fraud
- Deception of the injured party

- Property damage
- Unlawful property gain

Subjective element (guilt):

- The intellectual and volitional components indicate intent (*dolus directus* or *dolus eventualis*) on the part of the perpetrator (Kambovski V., Krstanoski M., Criminal Law - General Part, p. 218-223).

The perpetrator of the offense may be any natural or legal person, which classifies it as a general criminal offense. In terms of its content, the offense is alternative in nature, as the commission of any one of the prescribed actions is sufficient for its realization.

Fraud, as a continuing offense in the formal sense, is considered completed upon the occurrence of the consequence, whereas in the material sense, it is completed when the unlawful situation caused by the perpetrator ceases to exist (Kambovski V., Krstanoski M., Criminal Law - General Part, p. 133).

According to the Criminal Code, more severe forms of fraud are punished with stricter sanctions. These are most often related to:

- The amount of damage caused (e.g., damage of greater magnitude)
- The method of execution (e.g., use of false identity, forged documents, organized groups, etc.)
- Special status of the perpetrators (e.g., public officials or abuse of official position)
- Status of victims who enjoy special protection (e.g., minors, persons with disabilities).

Civil Law Framework

Within civil law, fraud is treated as a defect of will, with particular emphasis on the validity of contracts in the context of obligations. This system does not function as a punitive mechanism, but rather as a means of restoring legal balance between the parties. In general, private law should not be understood as a tool for promoting efficiency, but as a legal enterprise in which consistent public reason develops the norms that are implied in the mutual interaction of the parties (Weinrib E.J., p. 4-6).

According to the Law on Obligations of the Republic of North Macedonia, fraud is defined as a knowingly false or concealed act by which one party misleads or keeps the other party in mistake, and this deception conditions the will to conclude the contract (Law on Obligations of RNM, Article 57). Fraud affects the legal effect of the contract, leading to its voidability and annulment even when the mistake is not essential (Galev G., Dabovic A.J., p. 397-399).

If fraud is committed by a third party, the legal validity of the contract depends on the awareness of the other contracting party. If that party knew or ought to have known about the fraud at the time of contract conclusion, the contract may be annulled.

In cases involving gratuitous contracts (e.g., donations), the legislator applies a higher degree of protection. In such cases, fraud caused by a third party leads to the annulment of the contract regardless of whether the other contracting party knew or ought to have known about the fraud.

The consequences of fraud in civil law consist of:

- Voidability and annulment of the contract
- Obligation to compensate damages
- Restitution to the previous state (*restitutio in integrum*).

Procedural Differences

Although fraud in contractual relations may give rise to dual consequences—civil-law and criminal-law—the criminal and civil procedures differ in legal nature, purpose, procedural rules and outcomes.

The criminal procedure aims to establish criminal liability and sanction the perpetrator of the criminal offense. It is initiated *ex officio* by the public prosecutor (or by a private lawsuit filed by the injured party) and serves the public interest—protection of societal values and legal order.

In contrast, the civil procedure aims to resolve disputes between private parties and provide legal protection of their subjective rights. It is initiated by the injured party (plaintiff) to obtain compensation for damages, annulment of the contract, or other civil-legal consequences.

In criminal procedure, the principle of *in dubio pro reo* applies—any doubt goes in favor of the accused, and guilt must be established beyond a

reasonable doubt. In civil procedure, a lower standard of proof applies, known as the "preponderance of the evidence" or "balance of probabilities."

The criminal court may impose a penalty, security measures, conditional sentences, acquittal, or a judgment dismissing the indictment. It may also decide on property claims if the injured party requests this within the criminal proceedings, and in this respect, the criminal procedure is closest to civil procedure

The civil court, on the other hand, decides on contract annulment, compensation for damages, or restitution of property gains (based on unjust enrichment or tort liability) (Law on Obligations of RNM, Articles 103–109). Criminal procedure is regulated by short and strictly defined deadlines, aimed at efficient and swift resolution of criminal offenses, primarily to protect the rights of the accused and the public interest.

Civil procedure duration may vary depending on the complexity of the case and the conduct of the parties, although maximum statutory deadlines are prescribed. If deadlines are not set by law, the court determines them based on case circumstances (Civil Procedure Code of RNM, Article 103, paragraph 1). The deadline set by the court may be extended upon the proposal of an interested party if justified reasons exist (Civil Procedure Code of RNM, Article 103, paragraph 2).

In practice, the injured party may file a civil lawsuit either concurrently with or following the conclusion of the criminal proceedings. Although the procedures are autonomous, the criminal court's judgment has certain influence on the civil proceedings. According to the Civil Procedure Code, the civil court is bound by the final criminal judgment only concerning whether the criminal offense was committed and whether it was committed by the accused person (Civil Procedure Code of RNM, Article 11).

Conversely, the civil judgment is not binding on the criminal court, although the facts established therein may be used as indicative evidence (e.g., the existence of a contract) but must be evaluated and accepted in accordance with the rules of criminal evidentiary procedure.

The Boundary between Criminal and Civil Liability and Judicial Practice

94

In legal theory and practice, several key criteria have been established to distinguish between fraud as a basis for civil liability and fraud as a criminal offense:

- Intent (*dolus*) and degree of culpability: Criminal fraud requires proven intent (*dolus directus* or *dolus eventualis*), while civil liability may arise even without intent, based on unlawful acts or omissions.
- Consequences: Criminal fraud is characterized by a higher degree of social danger and the imposition of public sanctions, whereas civil fraud aims at compensation for damages without public-law consequences.
- Public interest and nature of the act: The criminal offense of fraud endangers broader societal values, while civil fraud constitutes an infringement of private interests within contractual relations.

An analysis of judicial practice, particularly in cases where fraud is the subject of criminal proceedings, highlights the need for both conceptual and practical differentiation based on the distinct criteria and roles of the two types of liability.

The case K. no. 64/16 before the Basic Court in Negotino is a clear example of the application of Article 247 of the Criminal Code, which sanctions fraud and false representation. In this case, the defendant was sentenced to 1 year and 2 months of imprisonment after intentionally misleading another person and thereby unlawfully obtaining material gain (Judicial Portal of the Republic of North Macedonia, Court Decisions).

In the case of Stojmenovic and Milosevic v. North Macedonia, considered by the European Court of Human Rights (ECtHR), the defendants were convicted of serious criminal offenses: abuse of official position, fraud, and money laundering. In the subsequent civil dispute, the court annulled the contracts, relying on the findings and evidence from the criminal proceedings, as well as the testimonies presented during the same process. This demonstrates how courts use the outcomes of criminal proceedings to reach decisions in civil disputes involving the same facts, particularly where contracts are voidable due to criminal acts.

According to recorded data for the period from 1 January 2024 to 1 January 2025, the basic courts in the Republic of North Macedonia handled a total of

64 cases related to criminal fraud, of which 35 involved contractual frauds. In two of these cases, the parties were referred to civil litigation, either because the public prosecutor failed to prove guilt beyond a reasonable doubt or because the court concluded that the matter concerned a civil-law relationship (Judicial Portal of the Republic of North Macedonia, Court Decisions).

An analysis of judicial practice clearly shows a prevailing tendency among the judiciary and litigants to treat the institution of fraud primarily as a criminal offense, to achieve punitive and preventive objectives, rather than as a basis for property claims within civil legal protection.

Furthermore, relevant case analysis reveals that fraud frequently appears in correlation with:

- Forgery of documents, including official documents
- Use of documents containing false information
- Abuse of official position and authority
- Property-related abuses, such as unlawful registration of ownership or the use of forged documents to make changes in property records

Available data on related criminal offenses involving elements of fraud, but not directly qualified as fraud, are evident in the fields of public procurement contracts, abuse of official position, and tax manipulations. According to data from the European Public Prosecutor's Office (EPPO), at the EU level, as of 2024, there were 2,666 active investigations, of which 9.32% concerned fraud related to public procurement (EPPO, Annual Report 2024).

In the context of Republic of North Macedonia, OLAF opened 18 investigations related to the misuse of EU funds during the period 2017–2022 (European Commission, Annual OLAF Reports – European Anti-Fraud Office, data for North Macedonia).

Although the legal framework in the Republic of North Macedonia is nominally aligned with European directives, limited effectiveness in implementation is evident. The European Commission estimated that public procurement in 2022 accounted for 8.6% of North Macedonia's GDP, representing 25% of the state budget (European Commission, North Macedonia Report 2023), while in 2023 it amounted to 10.9% of GDP, or 28.3% of the state budget (European Commission, North Macedonia Report

2024). However, concerns about corruption remain high, particularly in the contract awarding phases.

Practical Challenges and Recommendations

Practice related to contractual fraud also includes the phenomenon where certain parties use criminal proceedings to delay civil disputes or to exert procedural and psychological pressure on the opposing party. This phenomenon constitutes an abuse of rights (*abus de droit*) and may result in unjustified criminalization of private law relations, unfair application of criminal norms, as well as undue burden on the judicial system.

In this regard, the initiation of criminal proceedings for fraud must be based on concrete, verifiable and relevant facts indicating the existence of elements of a criminal offense, rather than solely on dissatisfaction with contractual relations or non-fulfillment of obligations.

Otherwise, such actions may violate the principle of the presumption of innocence, compromise the reputation of the person against whom the complaint is filed, and seriously endanger the right to a fair trial, as guaranteed by the European Convention on Human Rights (ECHR, Article 6).

Furthermore, the parallel pursuit of civil and criminal proceedings entails several risks:

- Duplication of evidence and proceedings, which may lead to legal uncertainty and confusion
- Divergent standards of proof—the "balance of probabilities" in civil proceedings versus "beyond reasonable doubt" in criminal proceedings, potentially leading to inconsistent outcomes
- Increased time and financial burden on both the parties and the courts
- Risk of conflicting judgments, which may create problems with the finality and enforceability of court decisions

On the other hand, proving the existence of intent as a form of culpability represents one of the most essential yet complex aspects of criminal proceedings. In most cases, intent is not explicitly expressed by the perpetrator but must be inferred indirectly through an analysis of the specific

circumstances of the offense and the behavior of the accused before, during, and after the act.

In practice, a common issue is the limited availability of witness testimony and material evidence, which significantly complicates the assessment of the perpetrator's mental state at the time of the act. Consequently, judicial interpretation of the facts may lead to differing outcomes, even in cases with similar factual backgrounds.

Regarding the establishment of causal link (causality), only one component is purely factual—the *sine qua non* condition. All other aspects are supplementary and do not constitute purely factual determinations (Hart H.L.A., 1985, p. 109–129).

Judicial practice demonstrates that, particularly in contracts involving complex legal and economic elements, proving intent requires a high level of expertise and coordination between the court, the prosecution, and expert witnesses. This significantly complicates and slows down the proceedings.

Considering the practical challenges, it is essential to establish a clear and precise demarcation between criminal and civil liability. It is recommended that legislation should define:

- Clear criteria for intent and its evidentiary assessment(e.g., specific indicators and qualifying elements)
- **A** clear threshold for when certain conduct shall be treated as a criminal offense and when it falls within the domain of civil disputes
- Mechanisms for coordination between civil and criminal court proceedings, to avoid parallel or contradictory outcomes

The existence of clear legal and institutional frameworks will enhance legal certainty, prevent the misuse of criminal complaints as a means of exerting pressure, and safeguard the right to a fair trial.

For the effective handling of complex cases, in addition to an adequate legal framework, systemic reforms in the judiciary are necessary. These include specialized training, strengthening of expert capacities, improved inter-institutional cooperation, and standardized protocols for the evaluation of evidence.

Comparative Overview

The distinction between civil and criminal liability represents a fundamental element of every legal system. Different national legislations apply varying conceptual and institutional models to differentiate between these two forms of legal responsibility.

Legal Approach in Germany

In the German legal system, there are clearly established criteria for distinguishing between the criminal offense of fraud (*Betrug*), regulated in the German Criminal Code (*Strafgesetzbuch* – StGB, § 263), and civil fraud, which serves as a basis for voidability of contracts under the German Civil Code (*Bürgerliches Gesetzbuch* – BGB, § 123).Criminal fraud entails the intentional misrepresentation of false facts and the creation of a misconception in another person, with the purpose of obtaining unlawful material gain for the perpetrator or a third party, thereby causing financial harm to the deceived person.By contrast, civil fraud allows for the annulment of a legal declaration (including a contract) if it was made under the influence of fraud or unlawful threat, without the need to prove material gain or loss — it suffices that the action involved intentional deception.

According to German case law (Federal Court of Justice – *Bundesgerichtshof*, BGH, 2024), civil proceedings are typically pursued in cases where there is insufficient evidence of criminal intent, particularly in instances of fraud of a minor scale or when the dispute primarily concerns the legal consequences of a contractual relationship. Criminal proceedings, by contrast, are initiated in cases of serious, systematic, or professionally executed fraud that result in substantial harm to the injured party and the public interest.

This legal distinction enables the functional and proportionate activation of legal mechanisms, maintaining intent and the severity of consequences as key factors in the legal qualification of the act (Bohlander M., 2022, p. 69–74).

Legal Approach in Greece

In Greek law, criminal and civil liability for fraud is regulated by the Criminal Code (Criminal Code, Article 386) and the Civil Code (Civil Code, Articles 140–150), respectively. Criminal liability is based on proving intent,

which entails the conscious and deliberate deception of the other party with the aim of obtaining unlawful material gain. Where such intent is absent or cannot be conclusively established, civil legal solutions—such as annulment of the contract or compensation for damages—are available under the Civil Code.

Greek case law emphasizes the importance of clear qualification of intent and the severity of consequences as essential criteria for initiating criminal proceedings. The Greek legal system provides flexibility in the treatment of such cases to avoid excessive criminalization of contractual disputes.

Contractual Fraud in the Context of Emerging Technologies and AI

With accelerated technological advancement, the digitalization of societal processes, and the increasingly prominent application of artificial intelligence (AI), new and significantly more complex forms of fraud are emerging within the domain of contract law. Modern technological tools—such as deep-fake technologies, automated software agents (bots), the manipulation of digital documentation and contracts, as well as the use of algorithmic models to execute fraudulent actions— will significantly complicate the legal qualification of fraud.

Within the EU, the adoption of the Artificial Intelligence Act (EU AI Act, Regulation 2024/1689) has introduced a new legal framework. Article 5 of this Regulation prohibits the use of AI systems that influence human decision-making through manipulative or deceptive techniques, particularly when such use causes significant harm or exploits vulnerabilities. This legal distinction complements the classical concept of intent by adding an objective technological element in the assessment of unlawfulness and intent. According to the Tietoevry Banking Report (Tietoevry Banking Report, 2025)**,** digital fraud in Europe increased by 43% in 2024**,** with 42.5% of fraud attempts being AI-driven and 29% of those being successful (Signicat, Battle Against AI-Driven Identity Fraud, 2024). These statistics indicate a real threat to trust and legal certainty in legal relations.

In the emerging digital economy, a conceptual adaptation of the law is necessary to address new forms of technological manipulation, especially in

contractual relationships where the parties are technically unequal. The academic debate further deepens these challenges, raising questions of liability, transparency, and prevention that will become increasingly relevant with the expanded use of artificial intelligence.

A multidisciplinary approach—combining technological expertise, legal analysis and normative criteria—is essential to ensure legal certainty and to prevent the misuse of advanced technologies within the realm of criminal law and beyond.

At the same time, emerging technologies offer opportunities to significantly enhance the efficiency and effectiveness of legal processes. For example, AI systems can be employed to track legal deadlines, detect anomalies in financial transactions and contractual clauses, and enable faster and more objective analysis of evidence. Moreover, digital tools can improve legal education, judicial coordination, and public trust in the justice system.

Conclusion

The distinction between criminal and civil liability is important not only for the legal system, but also for the broader functioning of society. This differentiation extends beyond legal-technical considerations and carries profound implications for the rule of law, the effective protection of individual rights, and the preservation of legal certainty and public trust in both economic and social relations.

The analysis shows that this distinction represents a complex legal and practical challenge, requiring careful harmonization of statutory regulation, judicial practice, and the actual needs of the justice system.

Precisely determining the boundaries between the two forms of liability is essential in order to ensure equal access to justice, to prevent misuse of legal mechanisms, and to guarantee that each action is qualified appropriately in accordance with its legal nature, gravity and consequences.

A clear distinction helps prevent the over-criminalization of private disputes, which may otherwise result in unnecessary litigation and an undue burden on the judiciary. It also facilitates the appropriate allocation of judicial resources and competences, thereby contributing to more efficient and well-organized legal protection. Moreover, legal clarity and procedural predictability

enhance the confidence of both citizens and the business community in the judicial system, an essential factor for fostering investment**,** economic development, and institutional legitimacy.

Finally, it is essential to emphasize the need for a balanced approach that respects the rights of victims while also protecting individuals from unwarranted criminal liability. Only through the advancement of legal frameworks, judicial practice, and legal education can a fair, efficient, and consistent application of the law be achieved—one that safeguards public trust in the legal system and ensures accountability for wrongful conduct.

Abbreviations

- AI- Artificial Intelligence
- CPC– Civil Procedure Code
- ECHR – European Convention of Human Rights
- ECtHR – European Court of Human Rights
- EPPO – European Public Prosecutor's Office
- EU – European Union
- OLAF – European Anti-Fraud Office
- RNM – Republic of North Macedonia

References

1. Ashworth A.**,** Principles of Criminal Law, 7th edition, Oxford University Press, 2013
2. Atiyah Patrick S.**,** The Rise and Fall of Freedom of Contract, Oxford University Press, 1979
3. Bohlander M., Abandoning dishonesty—A brief German comment on the state of the law after Ivey, The Journal of Criminal Law 86(1), 2022
4. Cronin A., Corporate Criminality and Liability for Fraud, 1st edition, Taylor & Francis, 2024
5. Dressler J., Understanding Criminal Law*,* 7th edition, LexisNexis, 2015
6. Dukoski S., Law of Obligations, Faculty of Law, Kichevo (in Macedonian)

7. Elliot C. & Quinn F., Contract Law, 9th edition, Pearson, 2019

8. Epstein Richard A., Principles of the Law of Torts, New York: Wolters Kluwer, 1999

9. Fletcher George P., Rethinking Criminal Law, Oxford University Press, 2000

10. Galev G. & Dabovic A. J., Law of Obligations, Faculty of Law "Iustinianus Primus", 2008 (in Macedonian)

11. Hart H. L. A., Causation in the Law, Oxford University Press, 1985

12. Kalajdziev G., et al. Commentary on the Criminal Procedure Code, OSCE, Skopje, 2018

13. Kambovski V., Krstanoski M., Criminal Law – General Part, Ohrid, 2013 (in Macedonian)

14. Mazzacuva F., 2021, The Impact of AI on Corporate Criminal Liability: Algorithmic Misconduct in the Prism of Derivative and Holistic Theories, Revue Internationale de Droit Penal, 92 (1)

15. Papadopoulos N., New Criminal Code and Code of Criminal Procedure, Vroteas Editions, 2024

16. Prutting H., Wegen G. & Weinreich G., BGB Commentary §123 – Avoidance Due to Deceit or Threat

17. Roxin Claus, Criminal Law – General Part Vol. 2: Special Manifestations of the Crime, 2003

18. Sefton-Green R. (Ed.), Mistake, Fraud and Duties to Inform in European Contract Law, Cambridge University Press, 2005

19. Tan E., Petit Jean M., Simonofski A., Tombal T., Kleizen B., Sabbe M., Bechoux L., & Willem P. (2023) Artificial intelligence and algorithmic decisions in fraud detection: An interpretive structural model, Data & Policy, e25

20. Weinrib Ernest J., The Idea of Private Law, Harvard University Press, 1995

Legal Acts, Judicial Sources and International Documents

1. Civil Code (Bürgerliches Gesetzbuch BGB) §123, § 142

2. Civil Code of the Hellenic Republic

3. Civil Procedure Code, Official Gazette of the RNM Nos. 79/05, 110/08, 83/09, 116/10, 7/11, 124/15

4. Code of Criminal Procedure, Official Gazette of the RNM Nos. 15/97, 44/02, 74/04, 83/08, 67/09, 150/10, 51/11, 100/12, 142/16, 193/16

5. Criminal Code of the Hellenic Republic

6. Criminal Code of the RNM, Official Gazette of the RNM Nos. 37/96, 80/99, 4/02, 43/03, 19/04, 81/05, 60/06, 73/06, 7/08, 139/08, 114/09, 51/11, 135/11, 185/11, 142/12, 166/12, 55/13, 82/13, 14/14, 27/14, 28/14, 115/14, 132/14, 160/14, 199/14, 199/14, 196/15, 226/15, 97/17, 248/18, 36/23, 188/23

7. Decisions of the Supreme Court of Greece ΑΝΑΖΗΤΗΣΗ ΑΠΟΦΑΣΕΩΝ - ΑΡΕΙΟΣ ΠΑΓΟΣ Accessed on 15 June 2025

8. European Commission, North Macedonia Reports 2023 and 2024 SWD - 2023 - 693 North Macedonia Report | PDF | Enlargement Of The European Union | European Union, North Macedonia Report 2024 | PDF | Enlargement Of The European Union | European Union Accessed on 15 June 2025

9. European Convention on Human Rights

10. European Parliament and Council, 2024, Regulation (EU) 2024/1689 on Harmonised Rules on Artificial Intelligence (AI Act)

11. European Public Prosecutor's Office (EPPO), Annual Report 2024, 2024 Annual Report: EPPO leading the charge against EU fraud | European Public Prosecutor's Office Accessed on 25 September 2025

12. German Criminal Code (Strafgesetzbuch, StGB) §263

13. Judicial Portal of the RNM, Court Decisions Одлуки Accessed on 15 June 2025

14. Law on Obligations, Official Gazette of the RNM Nos. 18/01, 4/02, 5/03, 84/08, 81/09, 161/09, 123/13, 215/21, 154/23

15. Netpress — Data on North Macedonia, OLAF Investigations Annual OLAF reports - European Anti-Fraud Office - European Commission Accessed on 25 September 2025

16. North Macedonia v. Stoimenovic and Milosevic, Case No. 59842/14, 25 March 2021, European Court of Human Rights (ECHR)) CASE

OF STOIMENOVIKJ AND MILOSHEVIKJ v. NORTH MACEDONIA (European Court of Human Rights) Application no. 59842/14 – LawEuro Accessed on 15 June 2025

17. Penal Code 2025 with Amendments Introduced by Law No. 5187/2025 (Greek)

18. Principles of European Contract Law (PECL)

19. Signicat, Battle Against AI-Driven Identity Fraud, 2024

20. Tietoevry Banking Report 2025 New report reveals an increase in digital payment fraud in Europe Accessed on 25 September 2025

21. UNIDROIT Principles of International Commercial Contracts 2016

22. United Nations Convention on Contracts for the International Sale of Goods (CISG)

23. Verdict of the Federal Court of Justice (Bundesgerichtshof, BGH), 12 December 2024 – III ZR 421/23 BGH, 12.12.2024 - III ZR 421/23 - dejure.org Accessed on 15 June 2025

THE ROLE OF ADMINISTRATIVE LAW IN PROMOTING GENDER EQUALITY IN CYBERSPACE IN EU COUNTRIES

Florije Rexhepi, LL.M., affilation

PhD student at the University "St. Kliment Ohridski"- Bitola, Republic of North Macedonia

Abstract

This paper examines the critical role of administrative law in promoting gender equality within cyberspace across EU countries. As digital platforms become increasingly integral to social and economic activities, ensuring gender equity online is essential. The study evaluates the effectiveness of existing administrative frameworks in addressing gender disparities and fostering inclusivity in digital environments. The paper identifies best practices and areas needing improvement by analyzing key legal instruments, policies, and case studies from various EU member states. The findings highlight the necessity for robust administrative measures to combat gender-based discrimination and harassment, thereby creating a more equitable digital landscape. The research underscores the importance of continuous legal and policy advancements to uphold gender equality in the rapidly evolving cyberspace of the EU.

Keywords: Administrative Law, Gender Equality, Cyberspace, EU Countries, Legal Frameworks, Gender-Based Discrimination

INTRODUCTION

In the digital age, cyberspace has become a crucial domain for social, economic, and political interactions. However, this rapid digital transformation has also highlighted significant gender disparities, necessitating robust legal frameworks to ensure equitable access and protection for all genders. Administrative law, with its regulatory and oversight functions, plays a pivotal role in addressing these disparities and promoting gender equality in cyberspace. This paper aims to explore the effectiveness of administrative law in fostering gender equality within the digital environments of EU countries.

The research methodology to achieve this work's objectives combines qualitative and quantitative methods. We will use the analytical method to study legislation, official documents, reports from local and international institutions, and academic literature related to the topic. Additionally, case studies from various EU member states will be examined to identify best practices and areas needing improvement.

By analyzing the role of administrative law in promoting gender equality in cyberspace, this paper seeks to contribute to the ongoing discourse on digital rights and gender equity, offering insights and recommendations for policymakers and legal practitioners in the EU.

The structure of the paper is organized into three main sections. The first section, Laws and regulations pertaining to gender equality in Cyberspace, provides an overview of the existing administrative laws and policies in EU countries that aim to promote gender equality online. The second section, Challenges and Gaps in Administrative Law, critically examines the shortcomings and challenges these legal frameworks face in effectively addressing gender disparities in cyberspace. The final section, Innovative Approaches to enhancing gender equality in Cyberspace, highlights successful case studies and offers recommendations for strengthening administrative laws to better promote gender equality in the digital age.

LAWS AND REGULATIONS PERTAINING TO GENDER EQUALITY IN CYBERSPACE

The European Union (EU) has established a comprehensive legal framework to promote gender equality, which extends into the realm of cyberspace. This section examines the key administrative laws and policies that aim to ensure gender equality online, highlighting their implementation and effectiveness through practical examples.

The EU's commitment to gender equality is enshrined in several foundational documents, including the Treaty on the Functioning of the European Union (TFEU) and the Charter of Fundamental Rights of the European Union (European Union, 2012). Article 157 of the TFEU mandates equal pay for equal work, while the Charter explicitly prohibits discrimination based on sex and ensures the right to gender equality in all areas, including employment, education, and access to goods and services (European Union, 2012).

One of the pivotal directives in this context is the Directive 2006/54/EC on the implementation of the principle of equal opportunities and equal treatment of men and women in matters of employment and occupation (European Commission, 2006). This directive has been instrumental in shaping national laws to promote gender equality in the workplace, which increasingly includes digital work environments.

The EU Gender Equality Strategy 2020-2025 outlines specific actions to address gender disparities in the digital sector (European Commission, 2020). This strategy emphasizes the need for gender-balanced participation in digital education and careers, the elimination of gender stereotypes in digital media, and the protection of women and girls from online violence and harassment.

For instance, the Digital Education Action Plan (2021-2027) aims to foster gender equality by encouraging more women to pursue careers in science, technology, engineering, and mathematics (STEM) (European Commission, 2020). The plan includes initiatives to provide digital skills training and support for female entrepreneurs in the tech industry.

One notable example of administrative law promoting gender equality in cyberspace is the Code of Conduct on Countering Ilegal Hate Speech

Online, which was launched in 2016 (European Commission, 2016). This code, developed in collaboration with major IT companies, aims to combat online hate speech, including gender-based harassment. Companies like Facebook, Twitter, and YouTube have committed to reviewing and removing illegal hate speech within 24 hours, demonstrating the practical application of administrative policies to protect gender equality online.

Another example is the Women in Digital (WiD) Scoreboard, which monitors women's participation in the digital economy across EU member states (European Commission, 2021). The scoreboard provides valuable data that helps shape policies and initiatives to close the gender gap in digital skills and employment.

The EU's administrative laws and policies provide a robust framework for promoting gender equality in cyberspace. Through directives, strategies, and practical initiatives, the EU aims to create a more inclusive digital environment. However, continuous evaluation and adaptation of these frameworks are necessary to address emerging challenges and ensure their effectiveness in fostering gender equality online.

CHALLENGES AND GAPS IN ADMINISTRATIVE LAW

Despite the robust legal frameworks and policies established by the EU to promote gender equality in cyberspace, several challenges and gaps persist. This section critically examines these shortcomings, highlighting practical examples and citing relevant scientific literature to provide a comprehensive understanding of the issues at hand.

One of the primary challenges in promoting gender equality in cyberspace is the inconsistency in the implementation of administrative laws across EU member states. While the EU provides overarching directives and strategies, the actual enforcement and effectiveness of these laws can vary significantly from one country to another. This inconsistency can lead to uneven protection and support for gender equality online.

For example, a study by Barker and Olga (2020) found that while some EU countries have made significant strides in implementing gender equality policies in digital spaces, others lag due to a lack of resources, political will, or cultural resistance. The study highlights the need for more uniform

enforcement mechanisms to ensure that all member states adhere to the same standards of gender equality in cyberspace.

The rapid pace of technological advancement poses another significant challenge. Administrative laws often struggle to keep up with the evolving nature of digital platforms and Technologies (Nimani et al., 2023). This lag can result in outdated regulations that fail to address new forms of gender-based discrimination and harassment online.

For instance, the rise of social media platforms has introduced new challenges in regulating online behavior. Sing (2018) argues that existing administrative laws are often ill-equipped to handle the complexities of social media, where gender-based harassment can occur in more subtle and pervasive ways. They call for more adaptive and forward-thinking legal frameworks that can respond to the dynamic nature of digital environments.

Enforcement of administrative laws in cyberspace is another critical issue. Even when robust laws are in place, ensuring compliance can be difficult. Digital platforms often operate across multiple jurisdictions, complicating the enforcement of national laws. Additionally, the anonymity provided by the internet can make it challenging to identify and hold perpetrators accountable.

A practical example of this challenge is the enforcement of the General Data Protection Regulation (GDPR), which includes provisions for protecting individuals from online harassment and discrimination. While the GDPR has been a significant step forward, its enforcement has faced hurdles due to the cross-border nature of digital platforms and the sheer volume of online interactions. Zaguir et al., (2024) note that despite the GDPR's strong legal framework, its effectiveness in protecting gender equality online is limited by these enforcement challenges.

Cultural and social attitudes towards gender roles and equality also play a significant role in the effectiveness of administrative laws. In some EU countries, deeply ingrained gender stereotypes and biases can undermine the implementation of gender equality policies in cyberspace.

Suzor et al., (2019) discuss how cultural resistance to gender equality can manifest in various ways, from subtle biases in policy enforcement to overt opposition to gender-inclusive initiatives. This resistance can hinder the

progress of administrative laws aimed at promoting gender equality, making it essential to address cultural and social barriers alongside legal reforms.

The challenges and gaps in administrative law highlight the need for continuous evaluation and adaptation of legal frameworks to effectively promote gender equality in cyberspace. Addressing inconsistencies in implementation, keeping pace with technological advancements, ensuring robust enforcement, and overcoming cultural barriers are crucial steps towards achieving a more equitable digital environment in the EU.

INNOVATIVE APPROACHES TO ENHANCING GENDER EQUALITY IN CYBERSPACE

In addition to addressing the challenges and gaps in administrative law, it is essential to explore innovative approaches that can enhance gender equality in cyberspace. This section delves into various strategies and initiatives that have shown promise in promoting gender equity online, supported by practical examples and academic citations.

One of the most effective ways to promote gender equality in cyberspace is through digital literacy and education. Empowering individuals with the skills and knowledge to navigate digital environments safely and confidently can help bridge the gender gap. Educational programs focusing on digital skills, cybersecurity awareness, and online etiquette are crucial.

For example, the EU's Digital Education Action Plan (2021-2027) includes initiatives aimed at increasing digital literacy among women and girls. By providing targeted training and resources, the plan seeks to encourage more female participation in the digital economy. A study by Hajdinijak and Damianova (2019) found that digital literacy programs significantly improved women's confidence and competence in using digital tools, thereby promoting greater gender equality in cyberspace.

Also, Estonia has been a pioneer in digital education, with initiatives like the ProgeTiiger program, which aims to enhance digital skills among students, including girls. This program has significantly increased female participation in STEM fields, contributing to greater gender equality in the digital sector.

Public-private partnerships (PPPs) have emerged as a powerful tool in promoting gender equality online. By leveraging the resources and expertise of both the public and private sectors, PPPs can implement comprehensive strategies to address gender disparities in digital spaces.

A notable example is the WePROTECT Global Alliance, a coalition of governments, technology companies, and civil society organizations working together to end online child sexual exploitation. This alliance has developed innovative tools and policies to protect children, including girls, from online abuse. Duta (2021) highlights the success of WePROTECT in fostering collaboration between stakeholders to create safer online environments for vulnerable groups.

The French government has partnered with tech companies through the French Tech Mission to promote gender diversity in the tech industry. This initiative includes mentorship programs, funding for female-led startups, and campaigns to encourage women to pursue careers in technology.

Innovative legal and policy approaches are also essential in promoting gender equality in cyberspace. This includes the development of new regulations that specifically address emerging issues related to gender and digital Technologies (Mishra, n.d).

For instance, the Istanbul Convention on Preventing and Combating Violence against Women and Domestic Violence has been instrumental in shaping national laws to protect women from online harassment and abuse. The convention's provisions have been integrated into the legal frameworks of several EU countries, leading to more robust protections for women in cyberspace. Doe and Roe (2019) argue that the Istanbul Convention serves as a model for other international agreements aimed at addressing gender-based violence in digital environments.

Technological innovations, such as artificial intelligence (AI) and machine learning, offer new opportunities to enhance gender equality online. These technologies can be used to detect and mitigate gender-based harassment and discrimination on digital platforms.

For example, AI-powered moderation tools have been implemented by social media companies to identify and remove harmful content more effectively. Tildesley et al., (2022) discuss how AI algorithms can be trained

to recognize patterns of gender-based abuse and take proactive measures to prevent it. However, Miller also cautions that these technologies must be designed and implemented with a strong ethical framework to avoid biases and ensure fairness.

Innovative approaches, including digital literacy programs, public-private partnerships, legal and policy innovations, and technological advancements, are crucial in promoting gender equality in cyberspace. By exploring and implementing these strategies, the EU can create a more inclusive and equitable digital environment. Continuous evaluation and adaptation of these approaches are necessary to address the evolving challenges and opportunities in the digital age.

CONCLUSION

This paper has examined the pivotal role of administrative law in promoting gender equality in cyberspace within EU countries. The analysis has highlighted the strengths and weaknesses of existing legal frameworks and policies, as well as innovative approaches that can enhance gender equity in digital environments.

The EU has established comprehensive legal frameworks that provide a solid foundation for promoting gender equality in cyberspace. These frameworks are essential for guiding member states in developing and implementing gender-inclusive policies. However, despite the robust legal frameworks, inconsistencies in implementation across different EU countries pose significant challenges. These disparities can lead to uneven protection and support for gender equality online.

The rapid pace of technological change often outstrips the ability of administrative laws to keep up. This lag can result in outdated regulations that fail to address new forms of gender-based discrimination and harassment. Ensuring compliance with administrative laws in the digital realm is complex, particularly given the cross-border nature of digital platforms and the anonymity provided by the internet. Deeply ingrained gender stereotypes and biases can undermine the effectiveness of gender equality policies. Addressing these cultural and social barriers is crucial for the successful implementation of legal frameworks.

To address inconsistencies, the EU should work towards harmonizing the implementation of gender equality laws across member states. This could involve establishing more uniform enforcement mechanisms and providing additional support to countries lagging. Administrative laws must be continuously updated to keep pace with technological advancements. This requires a proactive approach to legal reform, ensuring that regulations remain relevant and effective in addressing emerging challenges. Enhancing the enforcement of administrative laws in cyberspace is essential. This could involve developing more sophisticated tools for monitoring compliance, as well as fostering greater international cooperation to address cross-border issues.

Investing in digital literacy programs, particularly for women and girls, can help bridge the gender gap in digital skills and participation. These programs should be widely accessible and tailored to address the specific needs of different demographics. Collaboration between the public and private sectors can amplify efforts to promote gender equality online. Public-private partnerships should be encouraged to develop innovative solutions and share best practices. Efforts to promote gender equality in cyberspace must also tackle cultural and social barriers. This could involve public awareness campaigns, educational initiatives, and community engagement to challenge and change harmful gender stereotypes.

By implementing these recommendations, the EU can create a more inclusive and equitable digital environment, ensuring that gender equality is upheld in all aspects of cyberspace. Continuous evaluation and adaptation of legal frameworks and policies will be essential to meet the evolving challenges and opportunities of the digital age.

REFERENCES

Barker, Kim, and Olga Jurasz. 2020. "Online violence against women as an obstacle to gender equality: A critical view from Europe." *European Equality Law Review* 2020, no. 1: 47-60.

Doe, Alice, and Richard Roe. 2019. "The Istanbul Convention and Online Gender-Based Violence: Legal Innovations and Challenges." European Journal of Gender Studies, vol. 17, no. 3, pp. 201-220.

Dutta, Subhankar. 2023. "Framing the Landscape of Technological Enhancements: Artificial Intelligence, Gender Issues, and Ethical Dilemmas." In *Communication Technology and Gender Violence*, 109-123. Cham: Springer International Publishing.

European Commission. 2006. "Directive 2006/54/EC on the Implementation of the Principle of Equal Opportunities and Equal Treatment of Men and Women in Matters of Employment and Occupation." *Official Journal of the European Union*. Accessed September 5, 2024. http://data.europa.eu/eli/dir/2006/54/oj.

European Commission. 2016. "Code of Conduct on Countering Illegal Hate Speech Online." Accessed September 1, 2024. https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/combatting-discrimination/racism-and-xenophobia/eu-code-conduct-countering-illegal-hate-speech-online_en.

European Commission. 2020. "Digital Education Action Plan (2021-2027)." Accessed September 2, 2024. https://education.ec.europa.eu/focus-topics/digital-education/action-plan.

European Commission. 2020. "EU Gender Equality Strategy 2020-2025." Accessed September 2, 2024. https://commission.europa.eu/strategy-and-policy/policies/justice-and-fundamental-rights/gender-equality/gender-equality-strategy_en.

European Commission. 2021. "Women in Digital (WiD) Scoreboard." Accessed September 4, 2024. https://digital-strategy.ec.europa.eu/en/news/women-digital-scoreboard-2021.

European Union. 2012. "Charter of Fundamental Rights of the European Union." *Official Journal of the European Union*. Accessed August 25, 2024. https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:C2012/326/02.

European Union. 2012. "Treaty on the Functioning of the European Union." *Official Journal of the European Union*. Accessed August 26, 2024. https://eur-lex.europa.eu/EN/legal-content/summary/treaty-on-the-functioning-of-the-european-union.html.

Hajdinijak, M., and Zoya Damianova. 2019. "Comparative Study of the Gender Equality and Diversity Key." Deliverable D15 1: 60-117.

Mishra, Vidisha. "Gendering the G20: Empowering women in the digital age." *ORF Issue Brief*. New Delhi: Observer Research Foundation.

Nimani, Petrit, Maloku, Artan, & Avdija, Shefqet. 2023. The impact of the Auditor General on public administration control [Special issue]. Corporate Law & Governance Review, 5(2), 156–163. https://doi.org/10.22495/clgrv5i2sip2.

Singh, Rajnesh D. 2018. "Mapping online child safety in Asia and the Pacific." *Asia & the Pacific Policy Studies* 5, no. 3: 651-664.

Suzor, Nicolas, Molly Dragiewicz, Bridget Harris, Rosalie Gillett, Jean Burgess, and Tess Van Geelen. 2019. "Human rights by design: The responsibilities of social media platforms to address gender-based violence online." *Policy & Internet* 11, no. 1: 84-103.

Tildesley, Rebecca, Emanuela Lombardo, and Tània Verge. 2022. "Power struggles in the implementation of gender equality policies: the politics of resistance and counter-resistance in universities." *Politics & Gender* 18, no. 4: 879-910.

Zaguir, Nemer A., Guilherme H. Magalhães, and Mauro M. Spinola. 2024. "Challenges and enablers for GDPR compliance: systematic literature review and future research directions." *IEEE Access*.

DIGITAL EVIDENCE AND ARTIFICIAL INTELLIGENCE: REDEFINING JUSTICE IN CRIMINAL TRIALS

Arbër Shala, PhD Candidate

Faculty of Law, University "St. Kliment Ohridski" – Bitola

Email: arberb.shala@gmail.com

Abstract

The integration of digital evidence into criminal justice has reshaped procedural practices in the digital era. Advances in artificial intelligence (AI) have enhanced the collection, processing, and interpretation of digital data, ranging from metadata and surveillance footage to social media content, while raising ethical and legal concerns. This paper examines how AI-assisted digital evidence can strengthen justice without undermining due process or rights. Employing a mixed-methods approach, it analyzes 100 criminal cases involving digital evidence and integrates qualitative interviews with legal professionals. The methodology combines statistical analysis with thematic coding to evaluate evidentiary relevance and procedural impact. Findings show that although AI enhances efficiency and accuracy in legal proceedings, its application requires clear legal standards, transparency, and safeguards to preserve fairness and protect fundamental rights.

Keywords: digital evidence, artificial intelligence, criminal justice, legal ethics, forensic technology, privacy rights, procedural fairness

Introduction

The digital transformation of society has profoundly altered the foundations of modern criminal justice systems. Digital evidence, ranging from mobile phone metadata and closed-circuit television (CCTV) recordings to cloud storage files and social media posts, has become a central component in the adjudication of criminal cases. Courts, prosecutors, and defense lawyers

increasingly rely on technologically mediated information to establish facts, reconstruct events, and assess culpability. With this shift, questions of authenticity, admissibility, and interpretive reliability have emerged as pivotal issues in procedural law (Casey 2022).

In parallel, artificial intelligence (AI) technologies have accelerated the capacity to collect, process, and interpret vast volumes of digital information. Machine learning algorithms are now deployed in forensic image recognition, voice analysis, predictive policing tools, and even automated translation of intercepted communications. While these advances enhance the efficiency of legal proceedings, they simultaneously raise pressing challenges concerning transparency, fairness, and the preservation of fundamental rights (Kroll 2021).

The central tension arises from the dual nature of AI in criminal trials: it offers unparalleled precision and speed in evidence handling, yet it risks undermining due process if its "black box" operations are left unregulated. Issues such as algorithmic bias, lack of explainability, and privacy violations have become subjects of global debate among jurists, legislators, and ethicists (Mittelstadt 2022). These debates also concern democratic legitimacy, since criminal trials are among the most sensitive uses of state power.

This paper aims to explore the intersection of digital evidence and AI in the context of criminal justice. It argues that while AI-supported evidence may improve judicial efficiency, the legitimacy of such evidence depends on rigorous legal safeguards, transparent protocols, and the protection of procedural fairness. Employing a mixed-methods approach, the study examines one hundred criminal cases involving digital evidence and integrates qualitative insights from legal professionals. The findings contribute to current debates on how legal systems in Europe and beyond can adapt to the digital era without compromising justice.

The Evolution of Digital Evidence

The concept of evidence in criminal law has historically been tied to tangible objects, documents, weapons, fingerprints, that were relatively straightforward to authenticate and present in court. The emergence of digital technologies in the late twentieth century disrupted this paradigm by introducing forms of evidence that are intangible, easily altered, and dependent on technical expertise for validation (Kenneally and Brown 2020). Digital evidence includes data stored or transmitted in binary form that can demonstrate facts relevant to legal proceedings, whether through emails, call records, GPS data, or digital photographs (Brenner 2021).

The legal treatment of digital evidence has evolved unevenly across jurisdictions. In the United States, the Federal Rules of Evidence and landmark cases such as *Daubert v. Merrell Dow Pharmaceuticals* established criteria for scientific reliability that have since been extended to digital forensics. In the European Union, the General Data Protection Regulation (GDPR) and the proposed AI Act introduce stricter requirements concerning data handling and algorithmic accountability (Wischmeyer and Rademacher 2020). These frameworks highlight the growing recognition that digital evidence must be subject to rigorous standards of authenticity and admissibility.

Artificial Intelligence in Forensic and Judicial Practice

AI has introduced a new dimension to digital evidence by enabling the automated extraction and interpretation of patterns invisible to human investigators. Forensic applications include facial recognition, gait analysis, natural language processing for online communications, and predictive models of criminal behavior (Tiscornia and Sartor 2021). In many jurisdictions, AI tools are already employed in pre-trial risk assessments and in the classification of large datasets seized during investigations.

However, the integration of AI into judicial practice is fraught with controversy. One concern is the lack of explainability: deep learning models may produce accurate results without a clear rationale accessible to judges or defense attorneys. This opacity undermines the principle of equality of arms, a cornerstone of fair trial rights under Article 6 of the European Convention on Human Rights (Leenes et al. 2022). Another concern is algorithmic bias,

where training datasets replicate social inequalities, leading to discriminatory outcomes in areas such as facial recognition of minority groups (Buolamwini and Gebru 2018).

Legal and Ethical Concerns

The reliance on AI in criminal trials raises profound legal and ethical questions. Scholars emphasize the need to balance technological efficiency with constitutional safeguards such as presumption of innocence, right to confrontation, and judicial impartiality (Kroll et al. 2017). The principle of procedural fairness requires that defendants be able to challenge the evidence presented against them. Yet when algorithms function as opaque decision-making tools, defendants may be deprived of meaningful opportunities for contestation (Goodman and Flaxman 2017).

Ethical debates also focus on privacy and proportionality. The increasing use of AI to monitor social media, biometric data, and geolocation information risks normalizing surveillance practices that intrude upon individual freedoms (Zuboff 2019). This expansion of state power necessitates robust oversight mechanisms to ensure that AI-driven evidence is not used arbitrarily or disproportionately.

Comparative Jurisprudence and Emerging Standards

Comparative studies reveal divergent approaches to AI and digital evidence. In the United States, courts have generally admitted AI-enhanced evidence provided that expert witnesses can testify to its reliability, though this often occurs without full transparency about the algorithmic processes involved (Ferguson 2021). In contrast, the European Union is moving toward a precautionary model: the proposed AI Act classifies certain AI applications in criminal justice as "high-risk" and subjects them to mandatory conformity assessments (European Commission 2021).

International organizations, such as the Council of Europe and the OECD, have likewise issued guidelines emphasizing the importance of explainability, accountability, and human oversight in AI applications (OECD 2022). These developments suggest an emerging consensus that legal

systems must integrate technological innovation without compromising the foundational values of justice.

Research Gap

While the literature demonstrates a growing awareness of the challenges posed by AI in digital evidence, empirical research on how courts handle such evidence remains limited. Few studies provide systematic analysis of case law outcomes or the perceptions of legal practitioners regarding AI-assisted tools (Michels and Schartum 2023). This study addresses that gap by combining statistical analysis of case outcomes with qualitative interviews, offering a comprehensive view of how AI reshapes evidentiary practices in real-world criminal justice systems.

Methodology

This study adopts a **mixed-methods approach** to investigate how digital evidence, when processed and interpreted with the support of artificial intelligence, is used in criminal trials. The rationale for combining quantitative and qualitative methods lies in the complexity of the research questions, which cannot be fully addressed by relying on a single methodological tradition (Creswell and Plano Clark 2018). Quantitative analysis provides insight into patterns of admissibility and case outcomes, while qualitative interviews offer deeper understanding of the perspectives of legal professionals regarding the legitimacy and risks of AI-assisted evidence.

Case Selection and Quantitative Data

The quantitative component is based on an examination of **100 criminal cases** from 2020 to 2024 in which digital evidence played a central role. Cases were selected through purposive sampling, drawing from court archives in Kosovo, North Macedonia, and several EU jurisdictions with accessible digital case databases. The inclusion criteria were:

1. Digital evidence (e.g., metadata, CCTV footage, digital communications, or social media content) was introduced in court;

2. The evidence involved AI-supported collection or interpretation (e.g., automated pattern recognition, algorithmic sorting of datasets, or AI-enhanced facial recognition);
3. The judgment explicitly referred to the admissibility, reliability, or probative value of such evidence.

A coding scheme was developed to analyze case features, including type of crime, type of digital evidence, nature of AI tool used, whether the evidence was admitted, and its influence on the verdict. Statistical methods such as frequency distribution and cross-tabulation were applied to identify patterns in admissibility and correlation with trial outcomes (Bazeley 2021).

Qualitative Data Collection

To complement the statistical analysis, **semi-structured interviews** were conducted with twenty-five legal professionals, including judges, prosecutors, defense attorneys, and forensic experts. The interview questions explored participants' experiences with AI-assisted digital evidence, perceptions of its reliability, concerns about fairness, and views on the adequacy of current legal safeguards. Semi-structured interviews were chosen to allow for both comparability across responses and flexibility to pursue emergent themes (Kvale and Brinkmann 2015).

Participants were selected through a combination of professional networks and snowball sampling. Anonymity was assured to encourage candid discussion, and all interviews were transcribed and coded thematically using NVivo software. The coding process followed a deductive-inductive model: initial codes were derived from the research questions, while new codes emerged through iterative analysis of the transcripts (Braun and Clarke 2019).

Ethical Considerations

Given the sensitivity of the subject matter, strict ethical protocols were followed. All interviewees provided informed consent, and identifying details of both cases and participants were anonymized. Access to documents complied with national transparency rules. The study followed the European Code of Conduct for Research Integrity (ALLEA 2023).

Limitations

The methodology has certain limitations. The case sample, while diverse, may not be fully representative of all European jurisdictions. Variations in national procedural laws constrain the generalizability of findings. Moreover, the reliance on published judgments and accessible archives may overlook cases in which AI-based evidence was contested but not reported in detail. Despite these constraints, the combined quantitative and qualitative strategy provides a robust foundation for analyzing the evidentiary and ethical implications of AI in criminal trials.

Findings

The empirical analysis of one hundred criminal cases and twenty-five professional interviews reveal several important patterns regarding the role of artificial intelligence in the collection and interpretation of digital evidence. The findings demonstrate both the efficiency gains and the procedural risks associated with AI-assisted practices.

Patterns of Admissibility

Out of the 100 cases examined, 76 involved the admission of AI-assisted digital evidence. Courts were more likely to accept such evidence in cases of cybercrime, financial fraud, and organized crime, where large datasets made manual analysis impractical. In contrast, cases involving violent crimes showed more judicial caution, with judges questioning the reliability of algorithmic identification methods such as facial recognition. In 24 cases, digital evidence was either excluded or admitted with reservations due to concerns about authenticity, incomplete chain of custody, or insufficient expert testimony.

Statistical analysis revealed that the probability of conviction increased by 32% when AI-assisted digital evidence was admitted, compared to cases relying solely on traditional forms of evidence. This suggests that AI technologies significantly strengthen prosecutorial narratives, though it also raises concerns about overreliance on tools that may not be fully transparent.

Judicial Reasoning and Standards

In many judgments, courts relied on expert testimony to bridge the knowledge gap between the technical operation of AI tools and the evidentiary standards required by law. However, the reasoning often lacked detailed scrutiny of how algorithms functioned. Instead, admissibility decisions frequently referred to the reputation of forensic experts or the credibility of law enforcement agencies, rather than to the explainability of the algorithm itself. This pattern illustrates the persistence of the "black box problem" in legal practice (Mittelstadt 2022).

Judges were particularly concerned with the chain of custody of digital evidence. In several cases, defense attorneys successfully argued that without documentation of how AI tools processed data, the authenticity of the evidence was undermined. These outcomes reflect an emerging awareness that legal safeguards must adapt to the technological complexity of AI-based forensics.

Perspectives of Legal Professionals

The qualitative interviews enriched the statistical findings by highlighting the ambivalence of legal professionals toward AI. Prosecutors emphasized the efficiency and accuracy of AI in handling large datasets, noting that automated tools reduced investigative delays and facilitated the identification of hidden connections between suspects. Judges also acknowledged that AI tools could strengthen fact-finding but expressed concerns about the lack of clear procedural standards governing their use.

Defense attorneys and human rights lawyers voiced stronger reservations. Defendants often lack expertise to challenge AI-based evidence. Without enforceable transparency, AI risks violating equality of arms. One defense lawyer observed that "we are entering trials where the algorithm is the silent witness, but nobody can cross-examine it."

Forensic experts, on the other hand, highlighted the rapid evolution of technology, stressing that legal frameworks often lag years behind the tools deployed in practice. They pointed to the need for judicial training, as many judges lack the technical literacy required to critically evaluate AI-generated evidence.

Emerging Risks and Safeguards

The findings also underscore the risks of algorithmic bias. Interviewees cited cases in which facial recognition systems misidentified minority defendants, echoing concerns raised in global research on discriminatory outcomes (Buolamwini and Gebru 2018). No case in the dataset was overturned due to algorithmic bias, but the issue surfaced repeatedly as a source of injustice.

At the same time, several jurisdictions have begun experimenting with safeguards. Some courts required independent validation of AI tools before admitting results, while others mandated that forensic experts provide detailed reports on the operation of algorithms. These emerging practices suggest a gradual movement toward balancing technological efficiency with procedural fairness.

Discussion

The findings of this study confirm that the integration of artificial intelligence into digital evidence processing offers substantial benefits for criminal justice systems but also exposes persistent risks to fairness, transparency, and legitimacy. This section interprets empirical results considering existing scholarship and normative frameworks, focusing on four interrelated themes: efficiency versus fairness, transparency and the black box problem, algorithmic bias and equality of arms, and the evolving regulatory landscape.

Efficiency Gains versus Procedural Fairness

The statistical analysis demonstrated that convictions were significantly more likely when AI-supported evidence was admitted, suggesting that such tools reinforce prosecutorial strategies by enabling rapid processing of complex datasets. This efficiency aligns with broader arguments in favor of technological innovation in law enforcement, where digital forensics can accelerate investigations and reduce backlogs (Casey 2022).

However, efficiency is not synonymous with justice. The reliance on AI without adequate safeguards risks creating what some scholars describe as "technological overreach" (Ferguson 2021). Defense attorneys warned that defendants face algorithmic outputs they cannot challenge, undermining

adversarial balance. This echoes the warnings of Goodman and Flaxman (2017), who argue that opaque algorithmic tools threaten core guarantees of due process. The challenge, therefore, lies in reconciling efficiency gains with the preservation of fairness and equality before the law.

Transparency and the Black Box Problem

The research revealed that courts often admitted AI-assisted evidence based on the credibility of experts rather than the explainability of the algorithms themselves. This reliance highlights the enduring black box problem, in which judges and lawyers defer to technical expertise without full comprehension of how outputs are generated (Mittelstadt 2022).

From a normative perspective, this practice conflicts with the principle of procedural fairness, which requires that defendants understand and challenge the evidence used against them. If the algorithm functions as a "silent witness," as one defense lawyer phrased it, then the right to confrontation becomes hollow. Scholars such as Leenes et al. (2022) contend that judicial legitimacy depends on transparency and the ability of courts to justify decisions with intelligible reasoning. The findings here suggest that legal systems must move toward explainable AI (XAI) as a condition for admissibility.

Algorithmic Bias and Equality of Arms

The interviews also underscored concerns about algorithmic bias, particularly in facial recognition technologies. Research has consistently shown that these systems are less accurate in identifying individuals from minority groups, amplifying existing social inequalities (Buolamwini and Gebru 2018). While none of the cases in this study were overturned on such grounds, the mere risk of discriminatory error raises questions of substantive equality in criminal proceedings.

The principle of equality of arms requires that both prosecution and defense have reasonable opportunities to present and contest evidence. Yet algorithmic bias may skew evidentiary weight in ways that disproportionately burden marginalized defendants. This aligns with Zuboff's (2019) broader critique of surveillance capitalism, in which data-driven tools

often replicate systemic inequities. Without targeted safeguards, AI-assisted evidence risks entrenching structural biases within the justice system.

Regulatory Responses and Emerging Standards

Comparative jurisprudence illustrates divergent responses to these challenges. U.S. courts generally adopt a pragmatic approach, admitting AI-based evidence if accompanied by expert testimony, even without algorithmic disclosure (Ferguson 2021). By contrast, the European Union's proposed AI Act adopts a risk-based framework, classifying AI applications in criminal justice as "high risk" and subjecting them to conformity assessments and transparency obligations (European Commission 2021).

This divergence reflects different legal cultures: common-law systems emphasize adversarial testing of evidence, while civil-law systems prioritize statutory safeguards. The findings here suggest that neither approach is fully adequate. Adversarial systems risk overreliance on experts, while precautionary models may lag rapid technological development. A hybrid approach, combining robust transparency standards with flexible judicial oversight, may offer the most promising path forward.

Implications for Legal Practice and Education

Finally, the study highlights the importance of capacity-building among legal professionals. Judges and lawyers frequently lack the technical literacy required to evaluate AI-assisted evidence critically. As forensic experts emphasized, technological change outpaces legal adaptation, creating a gap that can undermine informed judicial decision-making. Training programs in digital forensics, AI literacy, and ethical implications are therefore essential for ensuring that courts remain capable of safeguarding rights in the digital era (Wischmeyer and Rademacher 2020).

Legal education must also be adapted. Law schools should integrate courses on technology and evidence to prepare future jurists for the realities of AI-driven trials. Moreover, interdisciplinary collaboration between computer scientists, ethicists, and legal scholars is necessary to develop frameworks that are both technologically informed and normatively robust.

Conclusion

The integration of artificial intelligence into the collection, processing, and interpretation of digital evidence marks a transformative moment for criminal justice systems. The analysis of one hundred cases and interviews with twenty-five legal professionals confirms that AI technologies provide undeniable benefits in terms of efficiency and accuracy. Courts are increasingly able to process vast datasets, uncover hidden connections, and expedite proceedings. Yet the same technologies also expose significant risks: lack of transparency, potential bias, and threats to fundamental rights.

The findings highlight that convictions are more likely when AI-assisted evidence is admitted, underscoring its growing influence in prosecutorial narratives. However, judicial reasoning often relies on the credibility of experts rather than the explainability of algorithms, leaving unresolved the black box problem. Legal professionals remain divided, prosecutors and forensic experts welcome technological efficiency, while defense attorneys warn that opaque algorithms undermine equality of arms. Judges, positioned between these competing perspectives, increasingly recognize the need for clear procedural safeguards.

These results have several implications. First, procedural fairness must remain the guiding principle. Efficiency cannot justify erosion of due process. Admissibility standards for AI-generated evidence should require transparency, explainability, and independent validation. Second, the risks of algorithmic bias demand proactive safeguards. Mandatory audits and bias-mitigation techniques should become integral to forensic practice to ensure that digital evidence does not entrench social inequalities. Third, legal systems must develop hybrid regulatory approaches, combining statutory safeguards with flexible judicial discretion, to adapt to rapidly evolving technologies. The European Union's risk-based AI Act and Council of Europe guidelines provide valuable models, though further harmonization is needed.

Finally, the human dimension must not be overlooked. Courts, prosecutors, and defense attorneys require ongoing training to critically engage with AI-assisted evidence. Legal education should integrate interdisciplinary modules that bring together law, computer science, and ethics. Only by strengthening

professional capacity can justice systems ensure that innovation serves fairness, accountability, and dignity.

In sum, artificial intelligence does not simply enhance the evidentiary process, it reshapes the meaning of justice in the digital era. The challenge for legal systems is to harness the benefits of AI while resisting its potential to undermine fundamental rights. Addressing this challenge requires not only technical solutions but also normative commitment to the rule of law. By embedding transparency, fairness, and accountability into the evidentiary use of AI, courts can ensure that digital justice remains true to its ultimate purpose: the protection of individual rights within a democratic society.

References

Abebe, Rediet, Moritz Hardt, Angela Jin, John Miller, Ludwig Schmidt, and Rebecca Wexler. 2022. "Adversarial Scrutiny of Evidentiary Statistical Software." *arXiv*, June 19. https://doi.org/10.48550/arXiv.2206.09305.

Berk, Richard A., Arun Kumar Kuchibhotla, and Eric Tchetgen Tchetgen. 2021. "Improving Fairness in Criminal Justice Algorithmic Risk Assessments Using Optimal Transport and Conformal Prediction Sets." *arXiv*, November 17. https://doi.org/10.48550/arXiv.2111.09211.

Casino, Fran, Claudia Pina, Pablo López-Aguilar, Edgar Batista, Agusti Solanas, and Constantinos Patsakis. 2022. "SoK: Cross-border Criminal Investigations and Digital Evidence." *arXiv*, May 25. https://arxiv.org/abs/2205.12911.

Chen, Yuxin. 2025. "When Algorithms Testify: Addressing the Explainability Gap of AI Evidence in Criminal Cases." *Studies in Law & Justice* 4 (3): 1–10. https://doi.org/10.56397/SLJ.2025.06.01.

Glauner, Patrick. 2021. "An Assessment of the AI Regulation Proposed by the European Commission." *arXiv*, May 26. https://arxiv.org/abs/2105.15133.

Gless, Sabine. 2024. "AI-Based Evidence in Criminal Trials? A Comparative Perspective." *William & Mary Law Review* (Scholarship Repository): Article 3206. https://scholarship.law.wm.edu/cgi/viewcontent.cgi?article=3206&context=facpubs.

Goodman, Bryce, and Seth Flaxman. 2017. "European Union Regulations on Algorithmic Decision-Making and a 'Right to Explanation.'" *AI Magazine* 38 (3): 50–57.

Hoffmann-Riem, Wolfgang. 2020. "Artificial Intelligence as a Challenge for Law and Regulation." In *Regulating Artificial Intelligence*, edited by Thomas Wischmeyer and Timo Rademacher, 1–29. Cham: Springer. https://doi.org/10.1007/978-3-030-32361-5_1.

Karsai, Krisztina. 2024. "The Use of Algorithms to Support Judicial Decision-Making in Criminal Matters with a Special Focus on Trial Decisions." *Silesian Law Review* 33 (5): 103–124. https://doi.org/10.17951/sil.2024.33.5.103-124.

Kroll, Joshua A., Solon Barocas, Edward W. Felten, Joel R. Reidenberg, David G. Robinson, and Harlan Yu. 2017. "Accountable Algorithms." *University of Pennsylvania Law Review* 165 (3): 633–705.

Leenes, Ronald, Bert-Jaap Koops, Mireille Hildebrandt, Sabine Gless, and Serge Gutwirth. 2022. *Regulating AI: Foundations, Methods, and Applications*. Cheltenham, UK: Edward Elgar.

Melkote, Mohan. 2025. "Access to Algorithmic Justice: Defending a Right to Expert Review." *Duke Law Journal Online*. https://scholarship.law.duke.edu/cgi/viewcontent.cgi?article=1119&context=dlj_online.

Michels, Thomas, and Dag W. Schartum. 2023. "Algorithmic Evidence in Criminal Proceedings: A Comparative View." *Journal of Law, Technology & Policy* 2023 (1): 1–35.

Neil, Roland, and Michael Zanger-Tishler. 2024. "Algorithmic Bias in Criminal Risk Assessment: The Consequences of Racial Differences in Arrest as a Measure of Crime." *Annual Review of Criminology* 8: 97–119. https://doi.org/10.1146/annurev-criminol-022422-125019.

Palmiotto, François. 2020. "Regulating Algorithmic Opacity in Criminal Proceedings." *SSRN*. https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3697173.

Reedy, Paul. 2023. "Interpol Review of Digital Evidence for 2019–2022." *Forensic Science International: Synergy* 6: 100313. https://doi.org/10.1016/j.fsisyn.2022.100313.

Sallavaci, O. 2025. "Algorithms on Trial: Does Evaluative Probabilistic Reporting Affect Evidence Reception in Court?" *PMC / legal forensics journal*. https://pmc.ncbi.nlm.nih.gov/articles/PMC12221571/.

Tokat, Yasin. 2024. "Attempts to Regulate Artificial Intelligence: Regulatory Practices from the United States, the European Union, and the People's Republic of China." *SSRN*, November 7. https://doi.org/10.2139/ssrn.4970843.

Wischmeyer, Thomas, and Andre Rademacher. 2020. "AI and Evidence Law in Europe: Between Promise and Integrity." *Computer Law & Security Review* 36: 105–22.

Wischmeyer, Thomas, and Timo Rademacher, eds. 2020. *Regulating Artificial Intelligence*. Cham: Springer. https://doi.org/10.1007/978-3-030-32361-5.

Yeung, Karen, and Adam Harkens. 2023. "How Do 'Technical' Design-Choices Made When Building Algorithmic Decision-Making Tools for Criminal Justice Authorities Create Constitutional Dangers?" *SSRN / arXiv*. https://arxiv.org/abs/2301.04713.

Ziosi, Marta, and Dasha Pruss. 2024. "Evidence of What, for Whom? The Socially Contested Role of Algorithmic Bias in a Predictive Policing Tool." *arXiv*, May. https://arxiv.org/abs/2405.07715.

THE ROLE OF ARTIFICIAL INTELLIGENCE IN MODERNIZING PUBLIC ADMINISTRATION IN THE WESTERN BALKANS

Artan Maloku, LL.M.

Teaching Assistant, Faculty of Law, University "Haxhi Zeka" in Peja, Republic of Kosovo
PhD student at the University "St. Kliment Ohridski"- Bitola, Republic of North Macedonia
malokua72@gmail.com

Abstract

The growing global integration of artificial intelligence (AI) into governance highlights its transformative potential for public administration systems. This research investigates the role of AI in modernizing public administration in the Western Balkans, focusing on its impact on enhancing efficiency, transparency, and public service delivery. The primary research question examines the extent of AI implementation and the key administrative and legal challenges that hinder its progress. A qualitative content analysis was conducted, based on national strategies, legislative frameworks, and regional development reports. The findings reveal uneven levels of AI adoption across the region, hindered by outdated regulatory structures, insufficient digital infrastructure, and a lack of institutional readiness. The research concludes that although AI can significantly modernize public administration, its successful integration requires comprehensive legal reform and strategic investment. Recommendations include the creation of a unified regional AI governance framework and the promotion of digital skill development among public sector employees.

Keywords: Artificial Intelligence, Public Administration, Western Balkans, Digital Governance, Modernization, Reform, Policy.

INTRODUCTION

The increasing integration of artificial intelligence (AI) into governance processes marks a profound shift in how public administrations function globally. AI technologies are not only reshaping service delivery but are also redefining the principles of efficiency, transparency, and accountability in governance (Dahl 1989). This research focuses on the Western Balkans, a region characterized by ongoing democratic consolidation and administrative reforms, to analyze the potential and limitations of AI in modernizing public administration. The primary goal is to assess the extent to which AI contributes to enhancing efficiency, transparency, and citizen-oriented service provision in these countries.

The motivation for undertaking this study arises from the pressing need to address structural and institutional challenges that continue to hinder administrative modernization in the Western Balkans. Unlike EU member states, where digital governance has advanced rapidly, countries in this region face persistent obstacles related to outdated regulatory frameworks, insufficient investment in digital infrastructure, and limited institutional readiness (Linz 1975). Investigating the role of AI in this context is therefore of high relevance, as it sheds light on both the opportunities and risks associated with technological integration into governance.

The justification for this research lies in its potential to contribute to academic and policy debates on digital transformation in transitional societies. By focusing on the Western Balkans, the study provides an empirical and comparative perspective on how AI can be used as a strategic tool for modernization, while also acknowledging the legal, ethical, and institutional complexities of its implementation (Geddes 1999). In this way, the paper adds to the existing body of literature on governance innovation and offers region-specific insights that may inform future policy design.

Nevertheless, the study is subject to certain limitations. The research is based primarily on qualitative content analysis of national strategies, legislative frameworks, and regional development reports, which may not fully capture informal practices, implementation gaps, or the dynamic evolution of AI technologies. Moreover, the research scope is confined to selected Western Balkan countries, which may limit the generalizability of the findings to other regions or administrative contexts.

Against this background, the study is guided by the following research questions: To what extent has AI been adopted in the public administrations of Western Balkan countries? What are the key legal, administrative, and infrastructural challenges impeding its integration? Can AI serve as a catalyst for long-term modernization of governance in the region? The central hypothesis is that while AI has significant potential to improve public administration, its effectiveness depends on comprehensive legal reforms, strategic investments, and institutional readiness.

Methodologically, the paper relies on a qualitative approach, employing content analysis of official documents, including national AI strategies, digitalization programs, and regional policy reports. Supplementary materials include comparative policy analyses and secondary academic literature that provide a theoretical foundation for assessing the role of AI in governance. This methodological design enables an in-depth exploration of both normative frameworks and practical challenges related to AI adoption.

The structure of the paper is as follows: After the introduction, the second section presents the theoretical background and literature review. The third section discusses the methodological framework and sources of data. The fourth section analyzes the empirical findings on AI adoption in the Western Balkans, while the fifth section outlines the key challenges and limitations. The final section concludes with policy recommendations and reflections on future research directions.

LITERATURE REVIEW

The role of artificial intelligence in public administration has become an increasingly relevant subject of academic discussion, both globally and in the regional context of the Western Balkans. At the global level, Osborne (2010) discovered that the integration of digital technologies, including AI, into governance processes is reshaping administrative practices and promoting new models of public service delivery. According to Margetts (2017), digital governance enables states to enhance transparency and accountability, but also raises questions related to ethics and data security. These findings provide a conceptual foundation for examining AI adoption in transitional democracies, such as those of the Western Balkans.

In the regional context, Dimitrov (2019) agrees with the view that public administration in the Western Balkans remains heavily burdened by bureaucratic legacies and weak institutional capacities. Several studies confirm the role of digital reforms in supporting governance modernization, but highlight significant disparities among countries in terms of implementation (Petrović, 2020; Hoxha, 2021). While some governments have made progress by adopting national AI or digitalization strategies, others lag due to limited resources and fragmented policymaking (Bajrami and Halili, 2024).

At the same time, several authors emphasize that the introduction of AI in public administration in the Western Balkans cannot be detached from the broader European integration process. Marković (2022) discovered that EU accession conditionality plays a key role in pushing governments toward adopting digital reforms and aligning regulatory frameworks with EU standards. However, a critical perspective is offered by Spasić (2023), who argues that despite the formal adoption of strategies, the practical implementation of AI remains superficial, with most projects failing to move beyond pilot stages.

An illustrative and highly controversial instance of AI governance experimentation is the case of "Minister Diella" in the Republic of Albania. Prime Minister Edi Rama announced the appointment of a so-called virtual minister named Diella, presented as an AI entity with the role of monitoring public procurement and ensuring fairness in contracting procedures. However, commentators pointed out that such an appointment has no constitutional or legal foundation, since ministerial office in Albania is reserved for natural persons who can be held accountable under the law (Plakolli-Kasumi, 2025). The initiative has been described as controversial and unlawful, raising concerns about democratic legitimacy and the risks of substituting symbolic digital gestures for substantive institutional reform. This example highlights the tension between innovation and legality, and resonates with the broader literature that cautions against overestimating AI's capacity to bypass entrenched governance challenges.

Methodologically, previous studies in the region often rely on descriptive analyses of policy documents or on case studies of individual countries.

Hoxha (2021) agrees with the view that this approach, while useful for mapping progress, tends to overlook cross-country comparisons and regional dynamics. Several studies confirm the role of comparative analysis in identifying structural patterns of reform and highlighting shared obstacles across Western Balkan administrations (Petrović, 2020; Marković, 2022). This study builds upon such insights by adopting a qualitative content analysis that systematically compares legislative frameworks, national strategies, and regional policy reports.

Taken together, the reviewed literature shows that while the modernization of public administration through AI has been widely recognized as a regional necessity, the pace and depth of reforms remain uneven. The contribution of the present research lies in providing a critical, comparative perspective that links AI adoption not only to governance efficiency but also to the broader challenges of institutional capacity, regulatory reform, and regional cooperation in the Western Balkans.

METHODOLOGY

This research is qualitative in nature, as its primary aim is to explore the role of artificial intelligence (AI) in modernizing public administration in the Western Balkans through the interpretation and critical assessment of official strategies, legal frameworks, and policy reports. A qualitative approach is most appropriate for this study, since the objective is not to measure numerical indicators but to understand the depth, scope, and implications of AI adoption within the governance systems of transitional societies.

The study is guided by the following research questions: To what extent has AI been adopted in the public administrations of Western Balkan countries? What are the main legal, administrative, and infrastructural challenges that hinder the successful integration of AI in governance processes? Can AI catalyze long-term modernization of public administration in the region?

The central hypothesis is that while AI carries significant potential to improve efficiency and transparency in public administration, its effectiveness in the Western Balkans is constrained by outdated legal frameworks, insufficient digital infrastructure, and limited institutional readiness.

Data collection relied on documentary analysis of national and regional policy materials. Sources included official national AI strategies, digitalization programs, e-government development plans, legislative texts related to technology and governance, and reports from regional and international organizations. Supplementary academic literature and media reports, such as the case of the virtual minister Diella in Albania, were incorporated to contextualize the findings and provide critical perspectives.

For data analysis, a qualitative content analysis technique was employed. Policy documents and reports were systematically coded according to thematic categories such as "legal framework," "institutional readiness," "infrastructure," and "AI applications in public administration." Comparative analysis was then used to highlight differences and similarities across Western Balkan countries. This method enabled the identification of common structural challenges, as well as country-specific opportunities and limitations.

The materials and tools used in the study included digital databases of legal texts and strategies, regional development reports, and secondary academic literature accessed through scholarly repositories.

The theoretical foundation of the research is rooted in the frameworks of digital governance and institutional modernization. The study draws particularly on the New Public Governance perspective (Osborne 2010), which emphasizes networked governance, collaboration, and innovation, as well as on theories of democratic consolidation and administrative reform (Linz 1975; Geddes 1999). These perspectives provide the analytical lens through which AI adoption is interpreted, situating technological innovation within broader processes of state capacity-building, legitimacy, and rule-of-law development.

The methodology combines qualitative content analysis with comparative regional assessment. This design ensures a comprehensive evaluation of how AI is being integrated into public administration in the Western Balkans and provides a solid foundation for addressing the research questions and testing the proposed hypothesis.

RESULTS AND DISCUSSION

The findings of this research reveal uneven levels of AI adoption in public administration across the Western Balkans. While some countries have adopted national AI or digitalization strategies, these documents remain largely declarative and lack concrete implementation measures (Budak and Škrinjarić, 2024). For instance, Serbia has taken modest steps toward developing e-government platforms that integrate AI functionalities, whereas other states, such as Bosnia and Herzegovina, remain at a preliminary stage, with fragmented strategies and limited institutional commitment. This confirms the initial hypothesis that AI adoption in the region is constrained by weak legal frameworks, infrastructural deficiencies, and institutional inertia.

The research questions can now be addressed in turn. First, concerning the extent of AI adoption, the results indicate that no Western Balkan country has achieved systematic or large-scale integration of AI into public administration. Implementation is partial, often limited to pilot projects or rhetorical commitments in policy documents. Second, the main challenges identified include outdated legislation, insufficient digital infrastructure, and a lack of human capital with relevant digital skills. These findings resonate with Hoxha (2021), who emphasized that without sustained investment in infrastructure and education, digital reforms remain superficial. Third, regarding AI as a catalyst for modernization, the evidence suggests that its potential is significant but largely unrealized in practice.

A critical example that illustrates both the ambition and the risks of AI governance in the region is the case of the so-called "virtual minister" Diella in Albania. While Prime Minister Rama presented this initiative as a step toward enhancing fairness in public procurement, commentators quickly underlined its lack of constitutional legitimacy (Plakolli-Kasumi 2025). This case shows that symbolic gestures toward AI integration, absent legal and institutional grounding, may generate controversy rather than progress. It also highlights the broader concern noted by Spasić (2023), namely that AI adoption in the Western Balkans risks being more rhetorical than substantive. When compared with existing literature, the results of this study both confirm and challenge earlier conclusions. Dimitrov (2019) identified weak institutional capacities as a central barrier to digital reform, a finding that this

research reinforces. Similarly, Petrović (2020) pointed to fragmented policymaking as a serious obstacle, a conclusion echoed in the present study. However, this research contributes further by emphasizing the role of symbolic experimentation, such as the Diella initiative, as an additional layer of complexity that had not been extensively analyzed before. This suggests that, beyond structural weaknesses, political actors may also misuse AI narratives to project reformist images without undertaking meaningful institutional change.

The meaning of these findings lies in their implications for governance reform. The evidence indicates that AI alone cannot modernize public administration unless supported by legal reform, infrastructural investment, and institutional capacity-building (Fabbri et al., 2025). Practical suggestions emerging from the study include: the harmonization of national AI strategies into a regional framework, the systematic training of public sector employees in digital skills, and the establishment of independent oversight bodies to ensure accountability in AI use. These measures would not only enhance the technical effectiveness of AI but also strengthen public trust in digital governance (Boshnjaku et al., 2025).

Directions for further research should focus on three areas. First, comparative studies between Western Balkan states and EU member states could highlight best practices and benchmarks for reform. Second, empirical research involving interviews with policymakers, civil servants, and citizens would provide valuable insights into the practical challenges of AI adoption, supplementing document-based analysis. Third, a critical exploration of ethical and human rights dimensions of AI in governance, such as algorithmic bias, surveillance, and data protection, remains an underdeveloped area in the regional context.

The results of this study confirm the hypothesis that AI adoption in the Western Balkans is hindered by structural, institutional, and legal challenges. At the same time, they expand the literature by critically examining symbolic cases such as Diella, which underscore the risks of adopting AI in ways that are politically expedient but legally questionable. The findings, therefore, contribute both to academic debates on governance modernization and to

practical policy discussions about how AI can be responsibly and effectively integrated into public administration in transitional societies.

CONCLUSION

This study has examined the role of artificial intelligence (AI) in the modernization of public administration within the Western Balkans, highlighting both its transformative potential and the limitations that constrain its implementation. The analysis demonstrates that AI adoption in the region remains uneven, with legal gaps, infrastructural deficiencies, and limited institutional readiness serving as persistent barriers. While pilot projects and national strategies signal political commitment, the practical integration of AI into administrative processes is still largely incomplete.

The research further underscores the tension between symbolic and substantive adoption, exemplified by the controversial appointment of the virtual minister Diella in Albania. Such initiatives illustrate how innovative AI applications can attract attention and signal reformist intent, yet also expose constitutional and legal vulnerabilities, raising questions about legitimacy and accountability in governance.

The findings carry several practical implications. Regional governments should prioritize harmonization of AI strategies, comprehensive legal reform, and targeted capacity-building to ensure that technological tools effectively enhance transparency, efficiency, and service delivery. Additionally, the development of independent oversight mechanisms and investment in public sector digital skills are critical for sustainable AI integration.

Future research should expand empirical inquiry through interviews, surveys, and cross-country comparisons to better understand the practical challenges of AI implementation. Further studies on ethical, legal, and social dimensions, such as data protection, algorithmic bias, and public trust, will be essential for developing robust governance frameworks. By critically analyzing both the promises and pitfalls of AI in the Western Balkans, this study contributes to a nuanced understanding of how digital innovation can support administrative modernization while emphasizing the institutional and legal foundations required for meaningful reform.

REFERENCES

Bajrami, E., & Halili, F. 2024. Exploring The Impact of Artificial Intelligence on the IoT and Digital Agenda in the Western Balkans: Integrating A Proposed Web Application for Regional Advancement. JNSM Journal of Natural Sciences and Mathematics of UT, 9(17-18), 244-258.

Boshnjaku, A., Plasari, E., Kambo, A., & Fata, I. 2025. ICT and e-government in agricultural development: A European perspective with insights from the Western Balkans (a case study of Albania). Journal of Central European Agriculture, 26(3), 812-824.

Budak, J., & Škrinjarić, B. 2024. Unveiling the role of artificial intelligence in strengthening public integrity and combating corruption. Sustainability in Western Balkans in an Era of Uncertainty, 29.

Dahl, Robert. 1989. Democracy and Its Critics. New Haven: Yale University Press.

Dimitrov, Aleksandar. 2019. Digital Governance and Administrative Reform in the Western Balkans. Skopje: Institute for Democracy.

Fabbri, E., Živković, L., Štrbac, D., & Ljumović, I. 2025. Regional potential for ICT collaboration in the context of Smart Specialisation in the Western Balkans.

Geddes, Barbara. 1999. Politician's Dilemma: Building State Capacity in Latin America. Berkeley: University of California Press.

Hoxha, Arben. 2021. E-Government and the Challenges of AI in Albania. Journal of Balkan Public Policy 12 (2): 45–67.

Linz, Juan J. 1975. Totalitarian and Authoritarian Regimes. In Handbook of Political Science, edited by Fred I. Greenstein and Nelson W. Polsby, 175–412. Reading, MA: Addison-Wesley.

Margetts, Helen. 2017. The Datafication of Public Services: Challenges and Opportunities. Public Policy Review 9 (3): 12–29.

Marković, Jelena. 2022. EU Integration and Digital Transformation in the Western Balkans. Belgrade: European Policy Centre.

Osborne, Stephen P., ed. 2010. The New Public Governance? Emerging Perspectives on the Theory and Practice of Public Governance. London: Routledge.

Petrović, Milan. 2020. Artificial Intelligence and Public Sector Reform in Serbia. Serbian Journal of Public Administration 8 (1): 22–40.

Plakolli-Kasumi, Luljeta. 2025. Emerimi i Ministres Virtuale Diella nga Kryeministri Edi Rama: Vendim Kontrovers dhe Joligjor. Sbunker, January 12, 2025. https://sbunker.org/opinion/emerimi-i-ministres-virtuale-diella-nga-kryeministri-edi-rama-vendim-kontrovers-dhe-joligjor/.

Spasić, Ivana. 2023. Critical Perspectives on AI Policy Implementation in the Western Balkans. South-East European Governance Review 5 (1): 77–96.

Ethical Governance in Public Administration: A Legal Perspective on Artificial Intelligence and Citizen Rights

Ilda Cucović

Judge, Basic Court Of Peja, Republic of Kosovo
PhD Candidate, Faculty of Law, UKLO – Bitola, North Macedonia
ildacucovic.sa@gmail.com

Abstract

The increasing adoption of artificial intelligence (AI) in public administration raises critical questions about ethical governance and the protection of citizen rights. This paper examines the intersection of administrative law and ethical standards in the context of AI implementation within Kosovo's public institutions. It analyzes how existing legal frameworks ensure or fail to ensure accountability, transparency, and respect for fundamental rights when public authorities rely on automated decision-making. Special emphasis is placed on legal safeguards necessary to prevent discriminatory outcomes, protect privacy, and maintain public trust. The study applies a normative and comparative approach to evaluate current legal gaps and propose regulatory reforms aligned with international best practices. The goal is to promote an ethical and legally sound integration of AI in governance, ensuring that technological innovation enhances rather than undermines democratic values.

Keywords: Ethics, Governance, Law, AI, Rights, Kosovo, Administration

Introduction

The integration of artificial intelligence (AI) into public administration has emerged as both an opportunity and a challenge for modern governance. On the one hand, AI has the potential to streamline bureaucratic processes, improve decision-making efficiency, and enhance service delivery to citizens (OECD 2019). On the other hand, the use of automated systems by public institutions raises profound concerns about accountability, transparency, and the protection of fundamental rights (Wirtz, Weyerer, and Geyer 2019). These concerns are particularly pressing in contexts where administrative law is still developing mechanisms to regulate new technological realities.

Ethical governance, rooted in principles of legality, fairness, and non-discrimination, is essential to ensure that AI integration in public institutions does not undermine democratic values (Floridi et al. 2018). Without adequate safeguards, automated decision-making risks producing discriminatory outcomes, eroding citizens' trust, and infringing upon privacy rights. Therefore, law plays a pivotal role in shaping the boundaries of technological innovation in governance by providing both normative standards and enforcement mechanisms (Zuiderveen Borgesius 2018).

In Kosovo, the question of how AI should be legally and ethically integrated into public administration remains largely unexplored. Although global discussions and international best practices offer valuable models, the local context requires a careful examination of administrative law frameworks to address the unique institutional challenges the country faces. This paper argues that legal safeguards must evolve alongside technological development to ensure that AI strengthens, rather than weakens, citizen rights and democratic governance.

This paper is structured into three main sections. The first section, "Legal and Ethical Frameworks for AI in Public Administration," examines the existing legislation and ethical principles relevant to AI governance in Kosovo and internationally. The second section, "Challenges and Risks of AI Implementation in Public Institutions," analyzes potential legal gaps, risks of bias, and threats to citizen rights arising from automated decision-making. The third section, "Policy Recommendations for Ethical AI Integration," proposes practical reforms and strategies to enhance accountability,

transparency, and public trust. The paper concludes with a synthesis of key findings and implications for law and governance.

Legal and Ethical Frameworks for AI in Public Administration

The integration of artificial intelligence (AI) into public administration necessitates a robust legal and ethical framework to ensure that technological advancements align with democratic principles and safeguard citizen rights. In Kosovo, while AI is increasingly utilized in public sector applications, the legal infrastructure remains in development, posing challenges to effective governance (Dërmaku and Emini 2024).

Kosovo's legal landscape is influenced by European Union (EU) regulations, particularly the General Data Protection Regulation (GDPR), which has been transposed into Kosovo's Law on Protection of Personal Data (Law on Protection of Personal Data No. 06/L-082 2019). This alignment underscores the commitment to data protection standards, emphasizing the necessity for public institutions to ensure that AI systems processing personal data comply with stringent privacy requirements (OECD 2024).

However, the application of AI in public administration introduces complexities that extend beyond data protection. Issues such as algorithmic transparency, accountability, and the potential for discriminatory outcomes require comprehensive legal considerations. The absence of specific legislation addressing AI's role in governance leaves a regulatory gap, potentially undermining public trust and the equitable delivery of services (Western Balkans Fund 2024).

Ethically, the deployment of AI must adhere to principles of fairness, non-discrimination, and accountability. In Kosovo, the ethical discourse surrounding AI is evolving, with academic institutions and civil society organizations advocating for frameworks that ensure AI technologies serve the public interest without infringing upon fundamental rights (UNDP 2023).

In response to these challenges, Kosovo's e-Government Strategy 2023–2027 outlines a vision for digital transformation that includes the responsible integration of AI. The strategy emphasizes user-centric design, digital-first approaches, and the importance of building institutional capacities to manage

technological advancements effectively (Ministry of Public Administration 2023).

While Kosovo has established a foundational legal framework for data protection, the rapid advancement of AI necessitates the development of comprehensive regulations that address ethical considerations and ensure that AI integration into public administration upholds democratic values and protects citizen rights.

Challenges and Risks of AI Implementation in Public Institutions

The integration of artificial intelligence (AI) into Kosovo's public administration introduces several challenges and risks that must be addressed to ensure ethical governance and the protection of citizen rights. As highlighted by the Western Balkans Fund, "the rapid adoption of AI technologies in public services presents both opportunities and significant risks, particularly in contexts with limited regulatory oversight" (Western Balkans Fund 2024, 5).

A significant concern is the lack of a comprehensive legal framework governing AI use in public institutions. While Kosovo has aligned some aspects of its legislation with European Union standards, such as the Law on Protection of Personal Data, there is no specific legislation addressing the deployment of AI in public administration. According to the OECD, "the absence of clear legal guidance for AI in government decision-making may compromise accountability and transparency, and create gaps in the protection of citizens' fundamental rights" (OECD 2025, 23). This regulatory gap can lead to challenges in ensuring that AI-driven processes remain fair, transparent, and aligned with democratic principles (Western Balkans Fund 2024, 8).

Another challenge is the potential for algorithmic bias in AI systems. AI algorithms can inadvertently perpetuate existing biases present in the data they are trained on, leading to discriminatory outcomes. The Western Balkans Fund notes that "bias in AI systems can exacerbate social inequalities, particularly in societies where ethnic and gender disparities already exist" (Western Balkans Fund 2024, 12). Similarly, UNESCO emphasizes that "ethically designed AI should actively mitigate

146

discrimination and avoid amplifying existing societal inequities" (UNESCO 2021, 7). In Kosovo, these concerns are particularly relevant for public service delivery, where impartiality and equality are essential.

Data privacy and security are also critical concerns. AI systems often require access to large datasets, which may include sensitive personal information. Without robust data protection measures, there is a risk of unauthorized access, misuse, or breaches of citizens' privacy rights. As noted by KDZ, "the ethical and legal handling of data is a cornerstone of trustworthy AI in public administration, and lapses can lead to both societal harm and erosion of public confidence" (KDZ 2025, 14).

Furthermore, the implementation of AI in public institutions can lead to a digital divide. Not all citizens have equal access to digital technologies, and the introduction of AI-driven services may marginalize those without the necessary resources or skills to engage with these systems. According to the OECD, "digital inequality can be exacerbated if AI deployment does not consider accessibility and inclusivity, resulting in uneven benefits across the population" (OECD 2025, 30).

Lastly, there is a risk of reduced public trust in government institutions. If AI systems are perceived as opaque or unaccountable, citizens may lose confidence in the fairness and integrity of public services. The Western Balkans Fund argues that "trust in AI-mediated governance is contingent upon transparency, ethical oversight, and active engagement with the public" (Western Balkans Fund 2024, 15).

Addressing these challenges requires a multifaceted approach, including the development of comprehensive legal frameworks, implementation of measures to ensure algorithmic fairness, robust data protection policies, efforts to bridge the digital divide, and initiatives to maintain public trust in government institutions. As UNESCO highlights, "responsible AI integration in governance is not only a technical endeavor but a societal commitment to human rights and democratic values" (UNESCO 2021, 9).

Policy Recommendations for Ethical AI Integration

To ensure the ethical integration of artificial intelligence (AI) into Kosovo's public administration, it is imperative to adopt a comprehensive

policy framework that aligns with international standards and addresses local challenges.

A foundational step is the development of a national AI strategy that incorporates ethical principles such as transparency, accountability, and non-discrimination. This strategy should be informed by UNESCO's *Recommendation on the Ethics of Artificial Intelligence*, which emphasizes the protection of human rights and dignity, advocating for AI systems that are "transparent, accountable, and subject to human oversight" (UNESCO 2021).

In parallel, Kosovo should establish a dedicated AI ethics oversight body tasked with monitoring AI implementations across public institutions. This body would ensure compliance with ethical standards, conduct regular audits, and provide guidance on best practices. According to the Institute for Territorial Studies, such an entity is "crucial in fostering public trust and ensuring that AI applications do not infringe upon citizens' rights" (Western Balkans Fund 2024, 12).

Furthermore, it is essential to invest in capacity-building initiatives aimed at educating public servants about AI technologies and their ethical implications. Training programs should focus on equipping officials with the knowledge to critically assess AI systems, understand their potential biases, and make informed decisions that uphold democratic values (KDZ 2025, 18).

To promote inclusivity, Kosovo should prioritize the development of AI solutions that are accessible to all citizens, including marginalized and underserved communities. This involves designing user-friendly interfaces, providing digital literacy programs, and ensuring that AI services do not inadvertently exclude any demographic groups (Western Balkans Fund 2024, 20).

Lastly, fostering international collaboration is vital. Kosovo should actively participate in regional and global dialogues on AI governance, sharing experiences and learning from best practices. Engaging with international bodies will help Kosovo stay abreast of evolving standards and contribute to the collective effort of ensuring that AI serves the public good (OECD 2025, 45).

Conclusions

The integration of artificial intelligence (AI) into Kosovo's public administration offers significant potential to enhance efficiency, improve decision-making, and modernize public service delivery. However, this study has shown that these benefits are accompanied by complex ethical, legal, and social challenges that must be addressed to protect citizen rights and maintain public trust.

A key finding is that the absence of a dedicated legal and regulatory framework for AI creates risks related to accountability, transparency, and the protection of fundamental rights. Without clear guidelines and oversight, AI systems may unintentionally perpetuate biases, threaten data privacy, and deepen existing inequalities, particularly among marginalized communities.

Furthermore, the responsible deployment of AI requires active measures to ensure fairness, inclusivity, and accessibility. Public officials must be trained to understand AI technologies and their potential risks, while mechanisms for oversight, transparency, and citizen engagement must be implemented to foster trust and legitimacy.

The ethical integration of AI in public administration is not merely a technical or technological issue; it is a societal imperative. For AI to truly enhance governance in Kosovo, policymakers, public institutions, and stakeholders must collaborate to establish comprehensive ethical standards, regulatory safeguards, and institutional frameworks. When implemented responsibly, AI can strengthen democratic values, protect citizen rights, and contribute to a more effective, equitable, and trustworthy public administration.

References

Dërmaku, Kastriot, and Ardian Emini. 2024. "Digitisation of Administration and Legal Basis in Kosovo." *Access to Justice in Eastern Europe* 1 (22). https://doi.org/10.33327/AJEE-18-7.1-a000107.

Floridi, Luciano, Josh Cowls, Monica Beltrametti, Raja Chatila, Virginia Dignum, Christoph Luetge, Robert Madelin, et al. 2018. "AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations." *Minds and*

Machines 28 (4): 689–707. https://doi.org/10.1007/s11023-018-9482-5.

KDZ – Centre for Public Administration Research. 2025. Artificial Intelligence in Local Government. Vienna: KDZ. https://www.kdz.eu/system/files/downloads/2025-04/AI%20at%20local%20level.pdf.

Law on Protection of Personal Data No. 06/L-082. 2019. Kosovo Assembly. https://www.dlapiperdataprotection.com/?c=XK&t=law.

Ministry of Public Administration. 2023. *e-Government Strategy 2023–2027*. Government of Kosovo. https://mpb.rks-gov.net/Uploads/Documents/Pdf/EN/2700/e-Government%20Strategy%20Kosovo%202023-2027.pdf.

OECD. 2019. *Recommendation of the Council on Artificial Intelligence*. OECD Legal Instruments, OECD/LEGAL/0449. Paris: Organisation for Economic Co-operation and Development.

OECD. 2024. *Public Administration in Kosovo 2024*. https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/01/public-administration-in-kosovo-2024_d9378da9/a53fce6b-en.pdf.

OECD. 2025. Public Administration in Kosovo 2024. Paris: Organisation for Economic Co-operation and Development. https://www.oecd.org/content/dam/oecd/en/publications/reports/2025/01/public-administration-in-kosovo-2024_d9378da9/a53fce6b-en.pdf.

UNDP. 2023. "AI Gender Bias: Navigating Data Gaps and Ethical Considerations in Kosovo." https://www.undp.org/kosovo/stories/ai-gender-bias-navigating-data-gaps-and-ethical-considerations-kosovo.

UNESCO. 2021. Recommendation on the Ethics of Artificial Intelligence. Paris: United Nations Educational, Scientific and Cultural Organization. https://www.unesco.org/en/articles/recommendation-ethics-artificial-intelligence.

Western Balkans Fund. 2024. "Effects of Artificial Intelligence on Human Rights: Kosovo Study Case." https://westernbalkansfund.org/wp-content/uploads/2024/12/ITS-The-Effects-of-Artificial-Intelligence-on-Human-Rights-Final-EN.pdf.

Wirtz, Bernd W., Jan C. Weyerer, and Carolin Geyer. 2019. "Artificial Intelligence and the Public Sector—Applications and Challenges." *International Journal of Public Administration* 42 (7): 596–615. https://doi.org/10.1080/01900692.2018.1498103.

Zuiderveen Borgesius, Frederik J. 2018. "Discrimination, Artificial Intelligence, and Algorithmic Decision-Making." *Council of Europe Study on Algorithmic Discrimination*. Strasbourg: Council of Europe.

PUBLIC ADMINISTRATION IN THE DIGITAL ERA: CHALLENGES OF ACCOUNTABILITY AND CONTROL OVER ARTIFICIAL INTELLIGENCE SYSTEMS

Besarta Zeqiri

PhD Candidate, Faculty of Law, University "St. Kliment Ohridski" North Macedonia
besartazeqiri2@gmail.com

Abstract

The digital transformation of public administration has accelerated the integration of artificial intelligence (AI) systems into administrative processes, raising new concerns regarding accountability, legality, and control. This paper explores the legal and institutional challenges that AI presents for public administration in the Republic of Kosovo, focusing on mechanisms for oversight, transparency, and citizen protection. It critically examines whether existing administrative law frameworks are adequate to regulate algorithmic decision-making and ensure democratic accountability. Drawing on international standards and comparative examples, the study proposes concrete legal and procedural reforms to strengthen public sector integrity in the digital era. Particular attention is given to risks of opacity, bias, and reduced human oversight. The paper concludes that a proactive legal approach is essential to ensure that technological innovation aligns with rule-of-law principles and public trust.

Keywords: Administration, Digitalization, Accountability, AI, Control, Kosovo, Law.

Introduction

The digital transformation of public administration represents one of the most significant institutional shifts of the 21st century. Governments worldwide are increasingly adopting artificial intelligence (AI) systems to improve efficiency, reduce costs, and deliver public services more effectively. While these innovations promise considerable benefits, they also raise critical concerns regarding accountability, legality, and democratic oversight. The reliance on algorithmic decision-making in areas such as resource allocation, public services, and regulatory enforcement challenges the traditional principles of transparency and responsibility that underpin administrative law (Bovens 2007).

In the Republic of Kosovo, as in many emerging democracies, the integration of AI into public administration is still at a formative stage. However, the adoption of such systems is rapidly expanding, often outpacing the development of appropriate legal and institutional safeguards. This imbalance creates potential risks related to the protection of fundamental rights, the maintenance of public trust, and the legitimacy of administrative decisions. As AI technologies introduce elements of opacity, bias, and reduced human oversight, the need for robust accountability mechanisms becomes increasingly urgent (Shneiderman 2022).

International experiences highlight both opportunities and risks in this domain. While some countries have developed frameworks to regulate algorithmic governance, others struggle to balance innovation with democratic control. Kosovo's case is particularly relevant as it provides insights into how transitional states can adapt global standards to local institutional realities. By addressing these challenges, the country not only safeguards administrative integrity but also strengthens its path toward European integration and adherence to rule-of-law principles.

The present paper seeks to contribute to this discourse by critically examining the adequacy of current legal and institutional arrangements in Kosovo regarding the oversight of AI systems in public administration. It further evaluates how principles of transparency, accountability, and legality can be reinforced to ensure that digital transformation aligns with democratic governance.

The paper is structured into four main sections: the first section outlines the conceptual framework of accountability and control in the digital era; the second section explores the legal and institutional challenges AI poses to public administration in Kosovo; the third section discusses comparative international experiences and their relevance for Kosovo; the fourth section proposes concrete reforms to strengthen accountability mechanisms; and finally, the paper concludes with a summary of findings and recommendations for future development.

Conceptual Framework of Accountability and Control in the Digital Era

Accountability has long been considered a cornerstone of democratic governance, ensuring that public institutions remain transparent, responsible, and responsive to the citizenry. In the digital era, the meaning and mechanisms of accountability are undergoing substantial transformation. The rapid integration of artificial intelligence (AI) into public administration introduces new complexities that traditional frameworks of administrative law struggle to address. Accountability in this context is not limited to political responsibility or bureaucratic oversight, but extends to technological processes, data governance, and algorithmic transparency. Scholars emphasize that accountability must adapt to the realities of digital governance by incorporating both procedural and substantive safeguards (Mulgan 2000, 555).

Control over AI systems in public administration is closely tied to questions of legality and legitimacy. As decision-making processes become increasingly automated, concerns arise regarding the ability of citizens and oversight institutions to understand, contest, and appeal administrative actions. The opacity of AI algorithms, often described as the "black box" problem, creates barriers to effective control and weakens public trust in institutions. Reports from the European Commission have stressed the need for a human-centered approach that preserves democratic values while enabling technological innovation (European Commission 2021, 6). Ensuring legality requires that AI systems used in the public sector comply not only with constitutional guarantees and administrative procedures, but

154

also with international standards on data protection, non-discrimination, and human rights.

The challenge is particularly acute in countries like the Republic of Kosovo, where administrative institutions are still consolidating their practices and capacities. Without strong mechanisms of oversight and control, the risk of bias, errors, or abuses in automated decision-making may undermine both the effectiveness of public administration and the legitimacy of state authority. The Law on General Administrative Procedure of the Republic of Kosovo provides a formal basis for legality and transparency in decision-making, yet it was not designed with algorithmic governance in mind. This raises the need for reforms that explicitly address digital transformation and AI applications in public administration. Furthermore, comparative experiences show that countries advancing in AI governance often rely on independent oversight bodies, specialized audit mechanisms, and transparent reporting requirements, all of which could provide valuable lessons for Kosovo's institutional development (OECD 2019, 14).

Overall, the digitalization of public administration requires rethinking the conceptual framework of accountability and control. The focus should not only be on ensuring compliance with legal norms, but also on safeguarding citizens' rights, fostering institutional trust, and maintaining the integrity of democratic governance. Establishing effective frameworks for accountability and control in the digital era is essential to prevent AI from becoming an instrument of unchecked power rather than a tool for public good.

Legal and Institutional Challenges of AI in Kosovo

The integration of artificial intelligence into public administration in Kosovo presents profound legal and institutional challenges. At the core lies the question of whether existing administrative and constitutional frameworks are sufficiently robust to regulate the use of AI in decision-making processes. The Constitution of the Republic of Kosovo establishes the principle of legality and guarantees fundamental rights, including equality before the law, access to justice, and data protection. However, these provisions were designed for traditional administrative practices and do not

fully anticipate the complexities introduced by algorithmic governance. The absence of specific legal instruments addressing AI leaves significant gaps in accountability and oversight, particularly in ensuring that automated decisions remain transparent, contestable, and subject to human review (Assembly of the Republic of Kosovo 2008).

Kosovo's Law on General Administrative Procedure requires that decisions by public bodies be reasoned, accessible, and subject to appeal. Yet when algorithms are involved, providing meaningful reasoning becomes increasingly difficult due to the technical opacity of machine-learning models. This creates tensions with procedural rights, especially for individuals affected by automated decisions in areas such as social services, taxation, or public employment. Without tailored regulations, there is a risk that AI may undermine the fundamental administrative law principle of due process (Law No. 05/L-031 on General Administrative Procedure 2016).

Institutional challenges further complicate the picture. Kosovo's administrative institutions often lack the technical expertise and resources necessary to assess, monitor, and audit AI systems effectively. Oversight bodies, including the National Audit Office and the Information and Privacy Agency, have mandates to ensure legality and data protection, but they face limitations in addressing algorithmic systems that operate with complex, non-transparent logic. International organizations such as the European Union and the Council of Europe have emphasized the importance of building institutional capacity and adopting ethical frameworks for AI in public governance, yet the pace of Kosovo's alignment with these standards remains slow (Council of Europe 2020, 9).

Another pressing issue is the protection of personal data and the prevention of discriminatory outcomes. The Law on Personal Data Protection in Kosovo is harmonized with elements of the EU General Data Protection Regulation (GDPR), but enforcement challenges remain due to limited institutional independence and resources. Given that AI systems can perpetuate or amplify existing social biases, weak enforcement exposes citizens to risks of unequal treatment and loss of trust in public institutions. Comparative experiences show that states with stronger independent regulators and dedicated AI oversight frameworks are better equipped to

mitigate these risks (European Union Agency for Fundamental Rights 2022, 18).

Finally, the broader challenge concerns Kosovo's process of European integration. Compliance with EU standards on digital governance and AI regulation is essential not only for protecting citizens' rights but also for advancing accession negotiations. The EU's proposed Artificial Intelligence Act provides a comprehensive framework that classifies AI systems based on risk and imposes strict obligations for high-risk applications, particularly in public administration. For Kosovo, aligning domestic legislation and institutional practices with this framework will be critical for ensuring democratic accountability and strengthening public trust in the digital era (European Commission 2021, 7).

Comparative International Experiences and Their Relevance for Kosovo

The regulation of artificial intelligence in public administration has become a priority for many countries and international organizations, as they seek to balance innovation with accountability and the protection of fundamental rights. Comparative experiences provide valuable insights for Kosovo, highlighting both successful practices and challenges in the governance of algorithmic systems. The European Union has taken a leading role in this area by introducing the Artificial Intelligence Act, which establishes a risk-based framework for AI applications. Under this proposal, systems used in public administration are considered high-risk and subject to stringent requirements, including transparency, human oversight, and conformity assessments (European Commission 2021, 12). For Kosovo, alignment with such standards is essential not only for safeguarding democratic accountability but also for advancing EU accession goals.

Experiences from member states of the Organisation for Economic Co-operation and Development (OECD) further illustrate the importance of robust institutional capacity. Countries such as Canada and the United Kingdom have adopted algorithmic accountability frameworks that mandate transparency in the design and deployment of AI in public services. The Canadian Directive on Automated Decision-Making, for instance, requires public institutions to conduct impact assessments, disclose algorithmic use,

and ensure recourse mechanisms for affected individuals. Similarly, the United Kingdom has developed an Ethics, Transparency and Accountability Framework for AI in the public sector, emphasizing human oversight and citizen protection (OECD 2021, 23). These approaches highlight the necessity of embedding accountability into the entire lifecycle of AI systems.

The Council of Europe has also advanced principles for the ethical use of AI, stressing the importance of legality, proportionality, and respect for human rights in administrative practices. Its guidelines underline that public authorities must retain ultimate responsibility for decisions, even when these are assisted by algorithmic processes (Council of Europe 2020, 11). This principle is directly relevant to Kosovo, where institutional weaknesses make the delegation of decision-making authority to AI particularly sensitive. Without clear safeguards, the risk of undermining public trust in governance remains significant.

In addition, the European Union Agency for Fundamental Rights has drawn attention to the risks of bias and discrimination in automated systems, particularly in relation to vulnerable groups. Its studies reveal that algorithmic systems may inadvertently replicate historical inequalities if not subject to rigorous oversight and high-quality data management (European Union Agency for Fundamental Rights 2022, 22). For Kosovo, where social disparities and issues of minority protection are central to democratic consolidation, these lessons underline the importance of designing AI governance mechanisms that are both inclusive and transparent.

Taken together, comparative international experiences show that the effective regulation of AI in public administration requires more than formal legislation. It depends on strong institutions, clear accountability mechanisms, independent oversight bodies, and the active participation of citizens in governance. By studying and adapting these examples, Kosovo can develop a model that ensures the benefits of digital innovation while upholding the principles of legality, transparency, and democratic control.

Proposals for Reforms to Strengthen Accountability Mechanisms in Kosovo

To address the challenges posed by artificial intelligence in public administration, Kosovo requires a comprehensive legal and institutional reform strategy that ensures accountability, transparency, and citizen protection. First, legislation should be updated to explicitly regulate AI use within public administration. This could take the form of amendments to the Law on General Administrative Procedure, integrating specific provisions on algorithmic decision-making, requirements for human oversight, and transparency obligations. Aligning domestic law with international standards, such as the European Union's Artificial Intelligence Act and OECD AI Principles, would provide a coherent framework for regulating high-risk AI applications and safeguarding fundamental rights (European Commission 2021, 12; OECD 2019, 14).

Institutional reforms are equally important. Kosovo's oversight bodies, including the National Audit Office and the Information and Privacy Agency, should be equipped with specialized units capable of auditing AI systems, assessing algorithmic fairness, and monitoring compliance with legal standards. Building technical expertise and providing adequate resources will be critical to ensure effective supervision of algorithmic processes and to prevent the delegation of unchecked authority to automated systems (Council of Europe 2020, 9). Moreover, independent oversight mechanisms, such as an AI ethics board or dedicated inspectorate within public institutions, could enhance accountability and strengthen citizen trust in government decisions.

Transparency and public engagement constitute additional pillars of reform. Public authorities should implement measures to disclose when AI is used in decision-making, explain the logic of automated processes in understandable terms, and provide clear recourse mechanisms for individuals affected by algorithmic outcomes. Experiences from Canada and the United Kingdom demonstrate that impact assessments and citizen feedback mechanisms are effective tools for mitigating bias, improving data quality, and ensuring procedural fairness (OECD 2021, 23). Kosovo could adopt similar practices, tailoring them to local administrative capacities and societal needs.

Capacity-building initiatives must complement legal and institutional reforms. Training public servants on the ethical and legal implications of AI, promoting interdisciplinary research on digital governance, and fostering collaboration with international organizations will strengthen the implementation of accountability mechanisms. By adopting a proactive, comprehensive approach that integrates legislative reform, institutional strengthening, transparency, and capacity-building, Kosovo can ensure that AI serves as a tool to enhance, rather than undermine, democratic governance and the rule of law.

Conclusions

The digital transformation of public administration presents both significant opportunities and critical challenges for democratic governance. This study has examined the integration of artificial intelligence (AI) systems into Kosovo's public sector, highlighting the legal, institutional, and ethical dimensions of accountability and control. The analysis reveals that existing administrative and constitutional frameworks, while foundational, are insufficient to regulate algorithmic decision-making, particularly in terms of transparency, human oversight, and protection of fundamental rights. Without targeted reforms, the use of AI risks undermining public trust, procedural fairness, and institutional legitimacy.

Comparative international experiences demonstrate that effective AI governance requires a combination of legislative clarity, institutional capacity, and public engagement. Countries that have adopted risk-based frameworks, independent oversight mechanisms, and impact assessment procedures provide valuable models for Kosovo to emulate. Aligning domestic legislation with international standards, such as the EU Artificial Intelligence Act and OECD AI Principles, is essential not only for protecting citizens' rights but also for supporting Kosovo's broader process of European integration and digital modernization.

The paper proposes concrete reforms encompassing legal amendments, institutional strengthening, transparency measures, and capacity-building initiatives. These reforms are designed to ensure that AI systems enhance administrative efficiency without compromising accountability, legality, or

equity. Implementing such measures would enable Kosovo to harness the benefits of AI while upholding democratic governance principles, demonstrating that technological innovation and rule-of-law compliance can be mutually reinforcing.

Ultimately, the study underscores the imperative of proactive, forward-looking policymaking in the digital era. By embedding accountability, oversight, and citizen protection into the design and operation of AI systems, Kosovo can establish a model of public administration in which innovation serves the public good, rather than eroding governance standards. The implications of this research extend beyond national boundaries, offering insights for other transitional states navigating the challenges of digital governance.

References

1. Assembly of the Republic of Kosovo. 2008. Constitution of the Republic of Kosovo. Prishtina: Official Gazette of the Republic of Kosovo, No. 25.

2. Bovens, Mark. 2007. "Analysing and Assessing Accountability: A Conceptual Framework." European Law Journal 13 (4): 447–468.

3. Council of Europe. 2020. Ad Hoc Committee on Artificial Intelligence (CAHAI): Feasibility Study. Strasbourg: Council of Europe.

4. European Commission. 2021. Proposal for a Regulation Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act). Brussels: European Commission.

5. European Union Agency for Fundamental Rights. 2022. Data Quality and Artificial Intelligence – Mitigating Bias and Error to Protect Fundamental Rights. Luxembourg: Publications Office of the European Union.

6. Law No. 05/L-031 on General Administrative Procedure. 2016. Official Gazette of the Republic of Kosovo, No. 20.

7. Mulgan, Richard. 2000. "Accountability: An Ever-Expanding Concept?" Public Administration 78 (3): 555–573.

8. OECD. 2019. Recommendation of the Council on Artificial Intelligence. Paris: Organisation for Economic Co-operation and Development.

9. OECD. 2021. State of Implementation of the OECD AI Principles: Insights from National AI Policies. Paris: Organisation for Economic Co-operation and Development.

10. OECD. 2021. State of Implementation of the OECD AI Principles: Insights from National AI Policies. Paris: Organisation for Economic Co-operation and Development.

11. Shneiderman, Ben. 2022. Human-Centered AI. New York: Oxford University Press.