Ethics and bias in artificial intelligence

Challenges and opportunities for the Republic of North Macedonia and the region

Authors:

Suad Seferi - Founder and President of "VI SEGA"
prof. Dr. **Sead Jigal** - University Professor
Prof. Dr. **Sanja Adzaip Velichkovski** - University Professor

Publisher
VI SEGA - Association for Artificial Intelligence
Skopje, 2025

Sorzhina

Introduction by the authors

This publication is the result of the joint work of three authors with different expertise, united around a common goal: to raise the issue of ethics and bias in artificial intelligence in a way that is appropriate, applicable, and scientifically based for the Macedonian and broader Balkan context.

The initiative was born from **Suad Seferi,** a researcher, author of the first book on artificial intelligence **“The AI Journey”** in the Balkans and founder of **“AI NOW”** – the first association for artificial intelligence in the Republic of North Macedonia. His mission is to connect technological progress with the public interest, through information, advocacy and regional networking.

Two prominent academic experts were also involved in the preparation of the document: **Prof. Dr. Sead Dzigal,** university professor and researcher in the field of communication and media studies, and **Prof. Dr. Sanja Adzaip Velichkovski,** Dean of the Faculty of Humanities and Social Sciences at the International Balkan University, whose many years of experience in education was crucial for the analysis of the ethical challenges associated with the use of artificial intelligence in the teaching process.

This report is not a definitive answer, but a call for shared reflection and action. It aims to encourage informed policies, responsible research and practices that will position the Republic of North Macedonia as a digitally innovative,
but also an ethically conscious society.

Ethics and bias in artificial intelligence – state and prospects in the Republic of North Macedonia and the region

Summary

The Republic of North Macedonia and neighboring countries face the challenge of reconciling the rapid expansion of artificial intelligence *(AI)* with ethical standards and the need to prevent bias. Global and European initiatives *(such as the new EU Artificial Intelligence Act, AI-Act)* are introducing strict rules for the responsible use of AI, but there is mixed progress in the region.

Serbia is leading the way with a national strategy and adopted ethical guidelines for trustworthy and accountable AI, while the Republic of Macedonia and other countries are still developing similar frameworks and lag behind their neighbors that have already adopted guidelines or laws. Key challenges include the emergence of bias in algorithms *(due to unbalanced data or design)* and dilemmas around transparency, privacy and accountability. This white paper analyses the current state of play, including global trends and national efforts, and highlights examples of good practices.

Finally, practical recommendations are proposed – from aligning regulations with European standards and establishing ethical guidelines, to strengthening education and regional cooperation – to ensure trustworthy and fair AI systems at the service of society. Implementing these recommendations will help the region fully exploit the potential of AI, without sacrificing human rights and ethical values.

Introduction

Artificial intelligence is becoming an integral part of modern society, with applications in healthcare and education, to public services and business. However, alongside the benefits, significant questions about ethics and bias in AI systems are also emerging. Ethical use of AI involves respecting human rights, non-discrimination, privacy and transparency in the design and implementation of algorithms. Bias, on the other hand, occurs when algorithms unknowingly favor or discriminate against certain groups due to inadequate data or assumptions in the model. These phenomena can have real-world consequences – for example, a biased hiring system may single out candidates based on gender or ethnicity, or a facial recognition model may be less accurate for certain demographic groups.

The importance of ethics and preventing bias in AI is particularly emphasized in the European Union, where trustworthy AI *is* a key goal in digital policies. For the countries of the Western Balkans, including the Republic of North Macedonia and its neighbors, this topic is also important, as they strive for European integration and digital transformation. In this paper, the emphasis is placed on the context of the Republic of Macedonia and the region *(Albania, Serbia, Bulgaria, Greece and Kosovo).* Through an analysis of global trends and local policies, a clear picture will be obtained of where the region stands in terms of ethical development of AI, what the main challenges are and what steps are needed next.

Situation analysis

Global and European policies and trends for ethical AI

Globally, several key initiatives to set standards for the ethical use of artificial intelligence have emerged in recent years. One of the first was the 2019 **OECD** AI Recommendation – the first international agreement calling for innovative yet trustworthy AI that respects human rights and democratic values. This document established five principles for responsible governance of AI, including inclusiveness, human values, transparency, robustness and accountability. Many countries *(including some in the region)* have adopted these principles as the basis for their own

policies.

UNESCO also played a significant role by adopting the Recommendation on the Ethics of Artificial Intelligence in November 2021, supported by 193 countries. This global recommendation establishes a set of values *(such as human rights, human dignity, diversity and sustainability)* and principles *(e.g. transparency, non-discrimination, fairness, accountability)* that should be followed by states in the development of AI. Serbia, for example, actively participated in the drafting of this document and subsequently aligned its national ethical guidelines with it. These international efforts emphasize the need for AI systems to be human-centric and safe, meaning that technological progress must be accompanied by the protection of fundamental rights.

In the European Union, the topic of ethics and bias in AI has been given concrete form through **the European Artificial Intelligence Act** *(AI-Act).* This regulatory act, the first of its kind in the world, was agreed in 2024 and entered into force on 1 August 2024. The Act aims to establish harmonised rules for the development and use of AI within the EU, encouraging innovation but also ensuring the responsible and safe use of AI-

technologies.

The Act introduces a risk-based classification of AI systems: from prohibited practices *(e.g., "social scoring" systems or mass surveillance)* to high-risk systems *(e.g., AI in health, transport, education)* that will be subject to strict conditions for transparency, human oversight and risk assessment. In addition, the Act requires member states to establish national oversight bodies and AI literacy strategies . Although only Bulgaria and Greece among the analyzed countries are EU members *(and directly bound by the Act),* the remaining candidate countries (Repu[illegible]ia, Albania) show intentions to comply with these rules as part of the EU accession proces[illegible]ciples that they should [illegible]actively develop their policies for trustworthy, fair and human-centric AI in line [illegible]uropean standards.

Additionally, since 2019, the EU has also iss[illegible] the Ethi[illegible]uidance on Trustw[illegible]y AI – a set [illegible]uidelines de[illegible]ed by an expert group, which define 7 main re[illegible]ements: [illegible]n oversight, technic[illegible]ustness, pr[illegible]y and data management, transparency, non-discriminatio[illegible]d diversit[illegible]cial and environme[illegible]enefits and [illegible]untability. [illegible]e guidelines were voluntary, but laid the foundat[illegible]on which[illegible]I Act now estab[illegible]inding norm[illegible]

The trend in Europe is clear – ethics in AI is no longer just an option, but a necessity, with regulations like the AI Act needing to provide a common framework for all.

This also has an impact on neighboring countries: Bulgaria and Greece must implement this Act, and other countries in the region are launching initiatives to keep up. The following is an overview of national strategies, initiatives and practices in the Republic of Macedonia and in neighboring countries, with an assessment of their progress in relation to these global and European trends.

Regional situation: national strategies, initiatives and regulations in neighboring countries

Within the SEE region, approaches to AI vary from country to country. Some, like Serbia, already have formulated strategies and concrete steps, while others, like Albania and Kosovo, are still in the early stages of policy design. This chapter provides a brief overview of the situation in each country – the Republic of Macedonia, Albania, Serbia, Bulgaria, Greece and Kosovo – in terms of national AI strategy, regulatory framework, institutional support and practices.

Republic of North Macedonia

The Republic of Macedonia has not yet adopted a specific national strategy or legislation on artificial intelligence and lags behind some neighboring countries in this field. However, there are initiatives that signal change. In 2021, through **the Fund for Innovation and Technological Development** *(FITR),* the Government established a working group to develop a National Strategy for AI. This step was also supported by international organizations such as **the World Bank** and **UNDP,** in order to align the strategy with EU standards. Unfortunately, progress has been slow – so far *(2025)* the strategy has not been finalized, and according to reports, there have been very few meetings of the working group since its establishment. This points to institutional problems, such as a lack of capacity and lack of coordination.

Although a formal framework is lacking, some AI-related activities are being implemented in the country. FITR finances domestic startups with an artificial intelligence component – at least 6 AI companies have been supported so far with around 441,000 euros. Additionally, in 2022, the Government adopted the Growth Acceleration Plan 2022 – 2026, which provides a fund of 27 million euros to support startups and innovative companies, including companies developing advanced technologies such as AI. This plan is part of a broader strategy for economic development and digital transformation. In the absence of a specific law on AI, the regulations that apply are general ones – e.g., **the Law on Personal Data Protection** *(compliant with the GDPR – General Data Protection Regulation)* which is relevant for AI systems that process personal information, as well as cybersecurity strategies that mention the need for regulation of new technologies.

On the positive side for the Republic of Macedonia is the ra elopment of the IT and S cosystem, which is one the most dynamic in the region. According to **Startup M onia,** there are around 400 startups e country, with the majority targeting foreign markets and operating in n sectors such as AI, agritech and biotec gy. This entre urial spirit means that there is a pool of talent and ideas can fuel th AI. romine uccessful ex is the Macedonian startup **Pixyle AI,** which uses a l intelliger object recogn automa n online fas – software that identifies fashion items in images generate se attributes for p s. Pixyle w ounded in by a researcher with a PhD in computer vision, a 2022 it s d an investment of on euros to and its bu These successes show that even without a fo strategy, vations in AI are har ng, but expe warn that th of a national framework carries risks: citizens an ciety can xposed to unintend nsequences AI *(e.g., data se, discrimination)* without safeguards. Therefore, Republic acedonia is in a to urgenth lize and imple its strategy to steer this growth of AI in a safe and cal direct

Albania

Albania currently does not have a specific national strategy for artificial intelligence. Instead, aspects of digital policy are covered in broader documents. In 2021, the Government adopted the Digital Agenda 2022-2026, which sets out the guidelines for ICT development. However, it does not mention specific measures for the development or regulation of artificial intelligence. This shows that the topic of AI has not yet received independent attention in national planning, but is treated indirectly through digitalization in general. Also, the National Cybersecurity Strategy 2020-2025 emphasizes the need for alignment with EU legislation and improving the regulation of new technologies, which implies that AI may be included in the future, but there are no concrete steps yet.

Despite the lack of a strategy, AI-related initiatives are emerging in Albania from academic institutions and the civil society. The Albanian Institute for Artificial Intelligence is a non-governmental organization established as the first national AI platform in the country. Its mission is to raise awareness, train young professionals and foster the development of the AI ecosystem. Institutionally, a key role is played by **the National Agency for Information Society** *(NAIS),* which is responsible for the digitalization of government services.

NAIS focuses on e-government, but it is possible that in the future it will also grow into a carrier of AI initiatives in the public sector.

Albania, like other EU candidate countries, is closely following European AI policies. Although it does not have its own AI Act, it is logical that Albania will have to implement the EU principles for ethical AI as part of the accession process. According to expert analyses, an opportunity will be missed if Balkan governments stand aside for too long and do not take proactive steps in the field of AI. In Albania's case, this means building on the digital agenda with specific AI activities – such as developing guidelines for the responsible use of AI, pilot projects in the public sector *(e.g., smart cities with AI, but with respect for privacy),* as well as initiatives to create human resources *(courses, training in machine learning at universities).*

Currently, Albania relies more on regional cooperation and learning than others: it participates in regional events on digital transformation and within the framework of the Open Balkans and other forums, AI is discussed. An example of this is that at the 2022 Western Balkans Digital Fair, in Pristina, Albanian repres ies.

In summary, Albania has yet to approach AI strat ily: the current situation is at the awaren tage, but without a concrete policy, which leaves room for rapid progr there is poli ort.

Serbia

Serbia is positioning itself as a regional leader in the development and regulation of artificial intelligence. Back in 2019, the Serbian government adopted the Strategy for the Development of AI 2020 – 2025, placing Serbia among the first European countries with a national AI strategy. This strategy was the result of the early realization that AI would have a transformative impact on the economy and society, and Serbia wanted to be

ready.

The strategy covers several pillars: development of AI infrastructure, support for science and innovation, application of AI in the public sector and education of personnel. For its implementation, **the Institute of Artificial Intelligence of Serbia was established in March 2021,** the first such state institute in the Balkans. The government stated that the goal of the Institute is to contribute to innovation and entrepreneurship, to create top experts and to attract investments in the AI-based economy. The Institute is already working on projects and cooperation with universities and industry, positioning Serbia as a regional center for AI research.

On the regulatory front, Serbia is taking steps to align with European and global ethical standards. In March 2023, the Serbian government officially adopted the Ethical Guidelines for the Development, Application and Use of Trustworthy and Responsible AI. These guidelines draw heavily on the 2021 **UNESCO** Recommendation on Ethics in AI and are a voluntary code for all those developing or using AI in Serbia. They aim to prevent AI systems that could endanger people or marginalize certain groups, and to ensure freedom of decision and non-discrimination.

The guidelines establish basic principles: explainability and verifiability *(transparency of models),* human dignity *(AI should not subjugate* people), non-harmfulness *(safety and protection from abuse) ,* and fairness. A self-evaluation questionnaire for developers is also prescribed, as well as recommendations for improving systems in accordance with the principles. With this, Serbia is taking steps towards establishing regulatory readiness before adopting binding acts *(such as the EU AI Act, which it would still implement upon accession).*

Practical support for AI in Serbia is significant. The government is investing in infrastructure – in 2020, **a State Data Center in Kragujevac** was opened with high-performance data hosting and supercomputers. The project to establish a national AI platform that infrastructure through the Office for IT and e-Government, in partnership with **NVIDIA,** Serbia is working to get one of the fastest AI supercomputers in Southeast Europe. NVIDIA will also help with training and joint research for domestic startups and talents, which is a huge boost for the local ecosystem. In parallel, Serbia is attracting foreign investments oriented to AI – an example is the opening of the **Huawei Innovation Center** in Belgrade, where the Chinese company, with the support of the Serbian government, is working on projects for smart cities, digitalization of education and the introduction of AI solutions.

However, this aspect has also brought criticism: digital rights organizations warn of the risks of China's technology in surveillance and privacy, pointing to the example of the installation of **Huawei** facial recognition cameras in Belgrade and the **"Safe City" project,** which could lead to mass surveillance. The European Parliament has also followed the issue, concerned about **"China's penetration of Europe"** through such technologies.

Serbia, therefore, is balancing between the desire to be a leader in AI and the pressure to ensure that that development is in line with Western ethical standards and values.

The results so far show that Serbia has established solid foundations: a strategy, a research institute, a code of ethics and investments. The challenge remains effective implementation – to realize the goals of the strategy *(which ends in 2025)* and potentially to adopt the first special law on AI.

According to some announcements, Serbia is also working on the first draft law on AI that would establish oversight bodies and rules, following the model of eureviewofailaw.com. At this pace, Serbia is positioning itself as an example for other Balkan countries for a proactive approach to AI, combining innovation with ethics.

Bulgaria

Bulgaria, as an EU member state, is following the European agenda on artificial intelligence and ethics. It has already developed a Concept for the Development of AI *(2020 – 2030),* a document that serves as a national strategy for AI. The main goal of this concept is the integration of AI into national priorities *(industry, science, e-government, agriculture, healthcare)* while building a trustworthy and human-centric AI.

Bulgaria defines six strategic goals:
(1) Modern infrastructure for AI – investments in communication and supercomputing infrastructure;

(2) Education and skills – reforms in education to create AI talent;
(3) Research and innovation – supporting centers of excellence and connecting science with industry;

(4) Data and digital resources – improving access to quality data and open AI databases;

(5) Integration of AI in key sectors – encouraging the application of AI solutions in the economy and public services;
(6) Trustworthy AI and regulation – creating a legal and ethical framework for the safe development and use of AI.

It is particularly emphasized that Bulgaria will develop regulations and standards aligned with EU principles to ensure that AI systems are human-centric, safe and respect rights. This includes the development of ethical guidelines for trustworthy AI, data protection, cybersecurity measures and oversight mechanisms.

In practice, Bulgaria participates in pan-European AI initiatives. It is part of the European Network of Digital Innovation Hubs and has a supercomputer **"Discoverer"** in Sofia, as part of the **EuroHPC network,** which is available to researchers for AI projects. The government has also announced the establishment of AI laboratories and increased funding for research, following the guidelines of the concept. In terms of legislation, as an EU member, Bulgaria will have to implement the AI Act – which means that by 2026 it will have to establish a national AI authority, adopt lists of high-risk systems and align local laws *(e.g. on consumer protection, product safety, etc.)* with the provisions of the Act.

One challenge for Bulgaria, identified in the concept is the brain drain. Many young IT professionals and scientists are moving to Western Europe, the United States, which means that investment is needed to retain and return experts. To this end, the plan envisages incentives for researchers, cooperation with the Bulgarian diaspora as well as specialized training programs.

From an ethical perspective, Bulgaria does not yet have a specific national code of ethics for AI *(like in Serbia),* but relies on EU principles. However, there is awareness of risks and risks. For example, within the framework of the Open Government Partnership, Bulgaria recognizes the need for a law that would regulate AI and guarantee human rights when using algorithms. This shows that policymakers recognize the need for legal certainty against risks and consequences *(discrimination, violation of privacy).*

In the next few years, Bulgaria will likely update its strategies: some reports state that 4 out of 5 Nordic countries are updating their AI strategies in 2024, so Bulgaria could also revise its concept in line with new trends.

Summary for Bulgaria: it has a strategic vision for AI until 2030, which also includes ethics, has technological infrastructure and access to EU resources, but needs to move from concept to implementation – to adopt specific ethical guidelines, national regulation under the AI Act and invest in human capital.

Greece

Greece has made significant steps in recent years towards developing a comprehensive policy on artificial intelligence, emphasizing ethical aspects as a central part of that policy. In November 2024, the document **"Blueprint for Greece** *'s AI Transformation" was published.* This roadmap, prepared by **the National Advisory Committee on AI and the Special Secretariat for Foresight ,** represents the finalized national AI strategy of Greece. The strategy has a structured vision for using AI for economic growth, better public services and ethical, responsible development of technology. This means that ethical considerations, transparency and multi-sectoral cooperation have been embedded as principles from the very beginning of the strategy.

Key initiatives of the Greek strategy include: establishing a national center of excellence for AI, establishing robust data governance frameworks *(especially since quality data is the basis for neutral and accurate models),* and promoting AI literacy and education through the education system. Greece is also investing in infrastructure – an example is the **“Pharos” project,** a supercomputer-based AI factory, which uses the capacities of the **“DAEDALUS”** supercomputer to provide powerful computational resources for AI research and innovation. This project, part of the European **EuroHPC initiative,** is intended to advance research in health, culture, sustainability, etc., while ensuring that the results are compliant with EU regulations *(e.g. on data protection).*

As an EU member state, Greece is bound by the AI Act, and hence its strategy already anticipates those obligations: it prioritizes an ethical and responsible approach at all stages – from development to implementation and oversight. The national document emphasizes values such as human dignity and transparency and sets out six **"flagship"** AI programs, including: ethical data management, cutting-edge research and educational infrastructure, broad AI literacy, ecosystem development, cultural and creative applications of AI, etc. In other words, Greece wants not only to introduce AI into various sectors but to do so in a way that will be an example of combining innovation with ethics.

There are already examples of AI applications in Greece with ethical oversight: in healthcare, AI tools are used for diagnostics, but under strict privacy rules; in public administration, chatbots for services are being piloted, but human rights impact assessments are being carried out before their use. Greece is also part of European AI Auditing net[illegible] encouraging companies to develop the capacity to check their AI systems for bias a[illegible]compliance wi[illegible]ndards.

Overall, Greece has an ambitious goa[illegible]f bec[illegible]ng a regi[illegible]leade[illegible]n the ethical an[illegible]nnovative use of AI, relying on a solid strategy, sign[illegible]cant i[illegible]stments *(th[illegible]flux o[illegible]ata center inve[illegible]ents from the US and Europe is increasing data p[illegible]essin[illegible]apacities)* a[illegible]active [illegible]rticipation in s[illegible]ing EU policies. The main challenge remain[illegible]mple[illegible]ntation: it is[illegible]cessar[illegible]o consistently i[illegible]ement the roadmap, monitor the results and up[illegible]te the[illegible]easures[illegible]eeded[illegible]ut, at least on pa[illegible], Greece is currently an example of a compreher[illegible]ve app[illegible]at co[illegible]th the technologica[illegible]nd ethical aspects of AI.

Kosovo

Kosovo, as the youngest country in the region and without **UN/UNESCO** membership status , faces specific circumstances in the development of its AI policies. It has not yet adopted a national strategy for artificial intelligence, but there are announcements and plans for its development. Within the framework of the government document **Digital Agenda of Kosovo 2030,** it is stated that the development of a Kosovo Strategy for AI and the development of a legal framework for an intelligent ecosystem are foreseen. This indicates that digital transformation is a priority and that AI is recognized as part of that agenda, although for now without a specific document.

In practice, the development of legislation and strategies for digital technologies in Kosovo is assessed as low or slow. The government has not yet adopted comprehensive policies for the introduction of advanced technologies such as AI, either in the public or private sector. However, it is positive that civil society and the IT community in Kosovo are very active. There are several organizations and initiatives that promote AI. For example, **INET Kosova** and **the IPKO Foundation** organize events, trainings and hackathons on the topic of AI, in order to familiarize the public and young people with this technology.

Additionally, Kosovo has a young and rapidly growing startup community: AI startups have attracted millions in investments in recent years. This rapid entrepreneurial development is happening without a state strategy – which speaks to the initiative of the private sector and the diaspora.

Kosovo has also hosted important conferences related to AI and digitalization, such as the aforementioned **Western Balkans Digital Summit 2022** *(WB6 Digital Summit)* in Pristina. Such events engage the country in regional dialogue and allow it to follow the trends and experiences of its neighbors. One interesting development is that political parties in Kosovo have started to include AI and ICT in their election platforms. For example, the ruling Vetevendosje party announced plans for a national AI program with a budget of 15 million euros, the creation of an AI strategy, and the establishment of an AI research center. This indicates a growing political awareness of the importance of AI and a promise of concrete steps, although these are still on paper.

One specific challenge for Kosovo in the context of AI ethics is related to geopolitics: in the northern part of Kosovo *(dominated by the Serb population),* Chinese-made *(Huawei, Dahua)* video surveillance and facial recognition systems funded by Serbia were installed. The government in Pristina opposes the presence of such technologies, considering them uncontrolled and potentially abused for citizen surveillance.

This example highlights the risk: without a comprehensive strategy and legal framework, advanced AI technologies can be introduced spontaneously or by external actors, bringing threats to privacy and security. On a positive note, the Kosovo Press Council has even updated its codes of ethics for journalists to cover the use of AI, in order to prevent the spread of disinformation and promote responsible use of algorithms in the media. This is a small step, but an important one in building awareness.

In conclusion, Kosovo is in the early stages of shaping an AI policy: it has a vision *(part of the 2030 Digital Agenda),* an active civil society and startups, but institutions need to be systematically involved.

A priority for the coming period would be: drafting and adopting a National Strategy for AI, establishing a working group or agency for AI, and harmonizing with international recommendations *(e.g., following the example of the UNESCO recommendation, although Kosovo is not a member, it can apply the same principles).* Given the young population and IT talents, Kosovo has the opportunity to "jump" directly into modern trends, if these initiatives are implemented appropriately.

Countries' progress in ethical and responsible AI development

Земја	Национална Стратегија	Етички упатства	Институционална поддршка	Образовни иницијативи	Закони/Регулативи
Северна Македонија	X (во изработка)	X	(ФИТР, стартап фондови)	(стартапи, курсеви)	X (општи закони)
Албанија	X (дел од дигитална агенда)	X	(NAIS, AI Institute NGO)	(невладини обуки)	X
Србија	(2020-2025)	(2023)	(AI Институт)	(AI курикулум)	(во подготовка)
Бугарија	(AI Concept 2030)	X (следи ЕУ норми)	(Discoverer SuperPC)	(Центри за иновации)	(AI Act имплементација)
Грција	(Blueprint 2024)	Да	(National AI Center)	(AI писменост)	(AI Act)
Косово	X (планира)	X	X (Civil Society)	(IPKO, хакатони)	X

X = No, or in progress

For better clarity, the following table compares the six countries *(Republic of North Macedonia, Albania, Serbia, Bulgaria, Greece, Kosovo)* according to key aspects: national AI strategy, legal/regulatory framework, ethical guidelines, institutional support, and educational initiatives. The table reflects the current state and progress of each country in these categories.

Comparison of the state of ethical and impartial AI in six countries

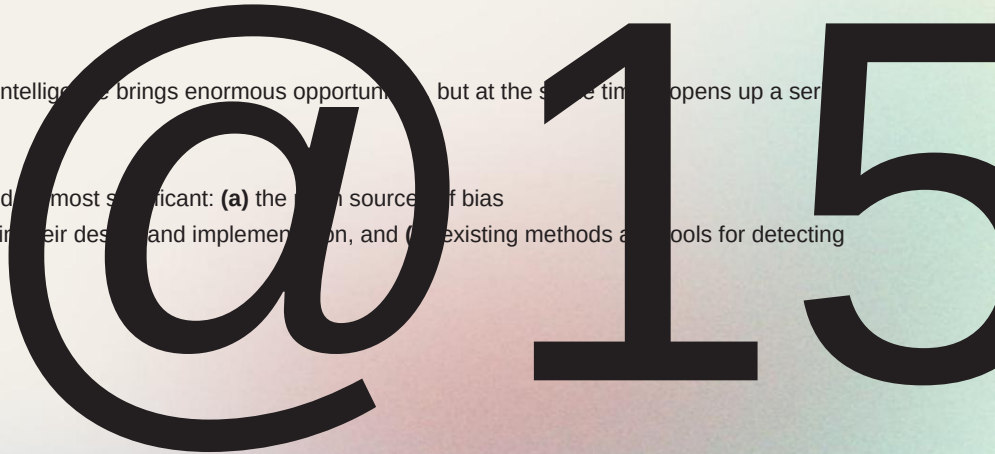
Земја	Етички принципи	Непристрасност и инклузивност	Алгоритмиска транспарентност	Граѓанска вклученост	Регулаторна рамка
Северна Македонија	Нема усвоени	Лингвистички празнини	Ограничена видливост	Слаба јавна дебата	Без конкретна регулатива
Албанија	Во најава	Ниски ресурси за различност	Ниско ниво на транспарентност	Активна НВО сцена	Без специфични прописи
Србија	Развиени водичи (2023)	Фокус на социјална инклузија	Повеќе пилот-проекти	Универзитетски панели	Во процес на донесување
Бугарија	Следи ЕУ препораки	Работи на јазичен AI	Вклученост преку унии	Проекти по универзитети	Има AI план
Грција	Подготвен етички кодекс	Вклучува маргинализирани групи	Отворени податоци	Платформи за учество	Применува AI Act
Косово	Нема	Слаба застапеност на малцинства	Слаба правна транспарентност	Невладин ангажман	Нема регулативи

The table shows that Serbia, Bulgaria and Greece have made the most progress in establishing a national AI framework *(strategies, institutions),* with Serbia and Greece standing out with explicit ethical documents and/or principles. The Republic of Macedonia, Albania and Kosovo are in the initial or preparatory stages – they recognize the need, but have not yet finalized or adopted formal strategies and guidelines. However, despite these differences, in all countries there is a trend of increasing interest in AI and gradual alignment with global standards, which is driven by both regional cooperation and the European perspective.

Key challenges

The technological development of artificial intelligence brings enormous opportunities, but at the same time opens up a series of ethical challenges that societies face.

This section discusses the two aspects cited as most significant: **(a)** the main sources of bias in AI systems and typical ethical dilemmas in their design and implementation, and **(b)** existing methods and tools for detecting and reducing bias, which are crucial for building fairer AI systems.



Main sources of bias and ethical dilemmas in AI

AI bias refers to a systematic error or disproportion in the output of an AI system, whereby a certain group of people is favored or discriminated against.

This bias most often stems from two sources:
(1) training data and
(2) the design of the algorithms themselves.

The first source, data bias, occurs when the data set on which a model is trained is unbalanced or has built-in historical biases. For example, if an employment model is trained on data that historically has predominantly employed men in certain positions, the model may "learn" a biased pattern that favors men in future predictions.

Algorithmic bias, on the other hand, can arise from the assumptions and decisions made by programmers when building the model – for example, how they define the objective function *or* which variables to include. If the team unknowingly introduces their own views or uses appropriate metrics, the algorithm can favor certain outcomes that are not fair.

The problem is that AI systems can amplify these biases on a large scale. If an algorithm “learns” a bias from data, it will repeat it and potentially amplify it in every decision it makes. Even small initial distortions in the data can lead to broadly discriminatory outcomes when applied across automated systems with millions of users. That’s why one of the main ethical imperatives is to identify bias before AI systems are applied to situations that affect people *(employment, credit, health diagnoses, etc.).*

In practice, numerous ethical dilemmas and cases of bias in artificial intelligence applications across various domains have already been documented:

Bias in HR : CV filtering or candidate ranking algorithms can discriminate against the weaker sex or minority groups. There is a well-known example where a recruitment tool favored words associated with men and downgraded female candidates based on historical data. The ethical dilemma here is how to combine efficiency *(automating selection)* with fairness and equal opportunities.

Credit scoring and banking: AI credit approval systems may disproportionately reject applicants from poorer regions or certain ethnic groups if historical data shows higher risk in those groups. This leads to deepening socioeconomic disparities. The ethical question is whether and how to use factors that correlate with race/gender/social class.

Healthcare: If a diagnostic algorithm is trained primarily on data from one ethnic group, it can produce inaccurate results for patients from other groups. This has happened before – skin cancer detection systems have been less accurate for patients with darker skin tones. The dilemma here is how to ensure equal care through AI without putting anyone’s life at risk due to biased algorithms.

Predictive policing: Tools that predict crime hotspots often target increased surveillance in minority neighborhoods because they rely on historical police data *(which may be the result of selective law enforcement).* This creates a self-reinforcing cycle – more police in an area means more recorded crime – which AI interprets as “higher crime”. **The ethical dilemma: safety vs. fair law enforcement.**

Facial recognition: Many facial recognition algorithms show higher error rates for women and people with darker skin. This has led to cases of wrongful arrests in the US due to incorrect identification of African Americans. The question is whether such technology should even be used by police until high accuracy is achieved for everyone and the built-in biases are removed.

Generative AI and content: Models like GPT run the risk of generating text with stereotypes or offensive statements, if such patterns exist in their database. Also, image generation sometimes reproduces racial or gender stereotypes *(e.g., if you ask for an image of an "engineer," the model returns mostly images of men).* This raises questions about the responsibility of creators to train models to avoid these pitfalls.

Transparency and explainability: Many AI systems are “black boxes.” For example, if an AI decides who gets a loan, it must ethically provide an explanation to rejected customers. The lack of transparency is a dilemma – how to balance the complexity of the models with understandability for the people affected.

Privacy: Applications such as facial recognition in public spaces *(the cameras in Northern Kosovo, or smart cities in Serbia with Chinese technology)* raise questions about privacy violations and potential misuse of surveillance. The ethical dilemma here is public **safety vs. privacy,** and the risk of a slippery slope towards Orwellian surveillance.

As can be seen, ethical dilemmas often arise from the clash of traditional values *(justice, equality, privacy)* with automated AI decisions.

The main principles that should be respected to address these dilemmas – according to international standards – are: non-discrimination, human control, transparency, accountability and privacy. The challenge is how to specifically apply these principles in different contexts. For example, non-discrimination means that AI does not discriminate on the basis of protected characteristics *(race, gender, etc.),* but this may also require active measures such as positive discrimination in data to correct past injustices – which is sometimes controversial.

In the region of RS Macedonia and its neighbors, an additional challenge is awareness and understanding of these issues among decision-makers and the public. The United Nations Development Programme *(UNDP)* research highlights that despite progress, continued discussion about the benefits and risks of AI is needed to adopt appropriate policies. It is important for societies to address these dilemmas preventively, and not only after harmful consequences have occurred. This leads us to the second aspect – what tools and methodologies are available to detect and reduce bias before AI systems are released “into the wild”.

Methods and tools for detecting and mitigating bias

Faced with the risk of biased and unfair AI systems, researchers and industry have developed multiple approaches to detect *and* mitigate *bias .* These strategies operate at different stages of an AI model's life cycle – before training, during training, or after training – in order to ensure that the final decisions are as fair as possible.

Here are the key methods:

Data pre *-processing:* These techniques are applied before the algorithm learns from the data. They include data cleaning, balancing, and transformation to reduce the influence of discriminatory factors. For example, if we have a data set of job applicants in which women are significantly underrepresented, we might oversample *the* records for women or undersample *the* records for men to achieve balance. Or, sensitive fields *(such as race)* might be removed or encrypted so that the model cannot "see" them. Such approaches ensure that biases in the data are mitigated before the model can process them.

Fairness- **aware algorithms :** This means that the models themselves or their learning algorithms are designed with built-in constraints or goals related to fairness. For example, in machine learning, in addition to optimizing accuracy, one can add conditions that the model must have a similar rate of positive decisions for different groups *(men/ women, minorities/ majorities).* There are algorithmic techniques that penalize bias in the training process – e.g., regularization that minimizes the difference in performance between groups. Research in the field of **fair ML** is developing new methods where fairness is a metric as important as accuracy.

Post- **processing :** These methods are applied after the model has been trained, directly to its outputs, to adjust them and make them more accurate. An example is the so-called recalibration – if a credit system has a threshold after which it approves credit, and it is determined that a disproportionate number of women are below the threshold, the threshold for women can be lowered or a "coefficient" can be added that corrects the output results.

Another example: In generative language models, a filter can be added that detects hate speech or gender-biased language and removes it before rendering. Post-processing is like a [illegible] catches potential injustices after the model has produced an output, and adjusts it before final use.

Bias testing and validation: Before being de[illegible]yed, an[illegible] system shou[illegible]ndergo t[illegible]rough testing with [illegible]ous scenarios and data to see if it produces biased results. [illegible]s is oft[illegible]done by creati[illegible]est data[illegible]s where expectati[illegible]for fairness are known in advance.

For example, we create a group with the same [illegible] twice, o[illegible]ce it is a [illegible] name, and the s[illegible] is a female nam[illegible] – if the model evaluates them differently, we kno[illegible]at there is gender bias. Such auditing can be internal *(by the developers)* and external *(by independent bodies or experts).*

There are also automated auditing tools – for example, **IBM AI Fairness 360 Toolkit** or **Google What-If Tool –** that allow for model analysis according to several fairness metrics.

Auditing goes hand in hand with transparency: developers need to provide insight into how the model makes decisions so that the audit can be meaningful. Some regulations *(such as the EU AI Act)* will likely require documentation of such assessments for high-risk
systems.

Human oversight and hybrid systems: One practical way to prevent major harm from biased AI is to keep a human in the decision-making process *(human-in-the-loop).* This means that the AI makes a recommendation, but the final decision is made by a human, who can assess whether there is any injustice. This principle is part of the EU's ethical requirements for human oversight, especially for critical applications. For example, an algorithm may preliminarily screen job candidates, but an HR officer reviews the list with a sense of diversity and makes corrections, if necessary.

Similarly in the judiciary: tools that predict the risk of reoffending *(as used in the US)* should be used only as advice, and the judge should have the final say, aware of possible biases.

Diversity in development teams: Interestingly, some solutions are not technical but organizational. One of them is ensuring diverse teams that create AI systems. Research shows that a team composed of people of different backgrounds, genders, and ethnicities is better able to identify potential biases, as each member brings a different perspective. Companies are encouraged to hire more women in IT, minority representatives, social science experts, along with engineers, etc. In this way, ethical questions are raised at the design stage that might otherwise have been overlooked.

Corporate AI Governance Policies: For the private sector in particular, it is advisable to introduce internal AI policies and ethics boards. These bodies would set guidelines, oversee projects, and approve models before they are released. Bringing together legal and service departments with IT teams is an important strategy for ensuring that products comply with legal and ethical standards. For example, stipulating that each model must pass a non-discrimination and reputational risk review, with both legal and engineering teams reviewing it.

In addition to these, new trends in fair AI are emerging and gaining momentum. One of them is Explainable AI *(XAI)* – the development of techniques that make models' decisions [illegible] humans.

This indirectly helps detect bias: if you can e[illegible]ain why [illegible]model m[illegible]ertain [illegible]cision *(for example, which 5 characteristics had the most influence),* it will [illegible]easier t[illegible]e if an inapp[illegible]ate cha[illegible]teristic *(such as e[illegible]city)* has a hidd[illegible] influence. Another trend is user-centric desig[illegible]involvi[illegible]end users in th[illegible]evelopm[illegible] process, especia[illegible]ose from vulnerable groups, to provide feedback and e[illegible]ure tha[illegible]e system take[illegible]eir need[illegible] to account and do[illegible]ot harm them.

Artificial data is also used to mitigate bias: if t[illegible]are not enough exampl[illegible] for a group, artificial data can be generated, which increases diversity without violating anyone's pri[illegible] the region, adopting these methods is part of building capacity for responsible AI.

Some initial steps are visible: for example, in Serbia, part of the ethical guidelines foresee a self-assessment questionnaire for developers of AI systems, which essentially forces them to think about bias and other risks before releasing the system.

In Greece, the AI Center of Excellence plan includes an AI Audit and Standardization segment, which would provide companies and public institutions with tools to validate their models. In the Republic of Macedonia and other countries, although such formalized mechanisms do not yet exist, existing global resources *(e.g., open tools such as the aforementioned IBM Toolkit, or consulting assistance from international projects)* could be used to evaluate emerging pilots.

In summary, bias in AI is a solvable problem if approached systematically: with good data preparation, properly designed algorithms, post-training control, and human oversight, injustices can be significantly reduced. This is not only ethically correct, but also increases public trust in AI, which is critical for its successful integration into everyday life.

Examples and case studies

This chapter presents concrete examples that illustrate efforts towards ethical AI – both regionally and internationally. Through real-world case studies, we can see how the principles discussed so far are being applied in practice, which approaches have proven successful, and what can be learned from those experiences. We will start with regional examples from the Republic of Macedonia and surrounding countries, and then we will also list international examples that can serve as models for adaptation in our context.

Regional examples of ethical AI initiatives and startups

"Pixyle AI" – Macedonia: Pixyle, already mentioned as a successful startup, is an interesting example from the aspect of the ethical use of AI in business. Its product – an algorithm for visual recognition of fashion products – does not directly involve sensitive personal data or make decisions that could discriminate against people, which means that ethical risks are minimal. Additionally, the company was founded by a female AI expert, **Svetlana Kordumova,** and is often highlighted in the media as a positive example of female leadership in technology. This contributes to gender inclusion in the IT sector. Pixyle collaborates with large retailers to enable better product search – indirectly, helping shoppers find what they want faster, without being led by biased recommendations *(e.g., the algorithm does not recommend clothes according to gender stereotypes, but exclusively according to visual characteristics).*

In this way, Pixyle shows how AI can be used for a positive user experience without ethical controversies. The success *(funded by EU funds and regional investors)* proves that innovations from a small market can compete globally, which should encourage other local startups.

"ADA" – a digital assistant in the public administration of North Macedonia: As part of the [illegible] Innovation Fund supported the development of the first Macedonian AI chatbot for citizen services called **"ADA".** This assistant is designed to answer questions from citizens and help them navigate administrative procedures.

What is ethically significant: ADA is trained on official information, passes accuracy checks, and is not intended to replace human officials but to supplement them. Through ADA, the Republic of Macedonia gets a chance to see how AI works in the public sector and establish guidelines for transparency *(the chatbot should clearly identify itself as a machine)* and accountability *(tracking responses, correcting if it gives incorrect advice).*

This project is also an example of government-startup collaboration – a small team of innovators working on a solution to a public problem, with government support. If ADA is successfully integrated and shows results, it will open the door to more AI projects in the administration, with experiences with ADA shaping future ethical rules *(e.g., how to store conversation data, how to ensure equal treatment of all citizens in responses ,* **etc.).**

Ethical Guidelines and Commissions – Serbia: The adoption of the national Ethical Guidelines for AI in Serbia *(2023)* is a regional precedent that could serve as an example for its neighbors. Serbia has not only adopted the guidelines but also established a mechanism *(in the pipeline)* for their implementation – most likely through a Commission or body that will monitor their implementation in public institutions. For example, it is being considered how ministries could create a catalogue of AI systems they use and conduct a self-assessment against the guidelines' questionnaire. **This is a practical step towards responsible government use of AI and a good learning case.** The Republic of Macedonia could, while developing its strategy, adapt and adopt similar ethical principles and questionnaires for its domestic needs. Regional cooperation here could directly benefit – the experts who developed the Serbian guidelines could share knowledge with

Macedonian colleagues.

The startup ecosystem in Serbia and the Balkans: In addition to Pixyle, there are other AI startups in the region that deserve attention. In Serbia, for example, the Institute for Artificial Intelligence operates, which has an incubator component and has already helped create several startups in the domain of medical AI solutions and agrotechnical AI platforms. In Kosovo, the startup **"Joma"** *(a fictitious name for example)* received investments for a platform that uses machine learning to analyze market trends - which shows that there is room for AI entrepreneurship in less developed markets as well. It is crucial that many of these startups understand that for global success they must also pay attention to their reputation in terms of ethics. A product with the label **"ethical AI"** can more easily enter Western markets. That is why we see that these startups *(especially those led by young founders with foreign education)* apply **privacy by design,** conduct diversity testing of their models, and the like. This trend should be encouraged - perhaps through competitions and awards for **"AI for social good"** in the region.

NGO Initiatives – Albania and Kosovo: In the abs[ence] of govern[ment] action, th[e] non-go[vern]mental sector is filling the gap. In Albania and Kosovo. For example, in Albania, **Op[en] Data Alb[ania]** is organiz[ing] workshops [on] fair algorithms in the public s[ector], trying to build the capacity of journalists and acti[vists] to dete[ct] bias in decisions *[including algorithms for public services).* In Kosovo, **the IPKO Foundation** implemented an educati[onal] progra[m] for young girls in t[echnology], w[here] elements of machine [learning] were taught along with a discussion of how stereotyp[es] can be [trans]lated into code. These exam[ples] are important bec[ause] they demonstrate societal awareness and commitm[ent] to ethic[s], regardless o[f] state. The[se] initiatives are sm[all], but they are bringing a new generation of IT professionals w[ho] will think [about] the grou[nd] up [about] th[e] impact of technol[ogy] on society.

Bias and ethical challenges in education: the role of AI

The integration of artificial intelligence *(AI)* into education represents a transformative shift with significant potential to improve learning across diverse student populations. However, this advancement also poses serious ethical challenges. As AI becomes increasingly integrated into teaching practices and personalized learning systems, ethical issues related to access, data privacy, and algorithmic bias require special attention. These aspects highlight the particularly important moral responsibility of educators, education policymakers, and policymakers in ensuring equal learning opportunities.

Accessibility is often cited as one of the greatest benefits of AI in education. AI-based technologies can personalize learning for students with disabilities or those from underrepresented communities. Educational tools such as various adaptive learning platforms and voice recognition software enable customized instruction and support, helping students who do not speak their native language and those from socially vulnerable backgrounds *(Khreisat et al., 2024).* However, without a planned design that takes into account inclusiveness, these technologies risk reinforcing existing inequalities. As Eden et al. *(2024)* emphasize , ethical frameworks for AI must prioritize accessibility, otherwise AI systems may inadvertently exclude certain groups.

Data privacy is another key aspect. AI in education relies on collecting data about students, from academic scores to behavioral patterns. While this enables more effective personalized learning, it also raises questions about consent, data ownership, and information security. Institutions must address these issues by implementing rigorous privacy and

transparent data handling practices. Eden et al. *(2024)* point out that data protection is crucial not only for legal compliance, but also for maintaining trust in education systems. Inappropriate data handling can cause harm and erode trust in school environments that use AI.

Algorithmic bias poses an additional ethical challenge. While AI can tailor learning to individual needs, it can also accentuate social inequality if developed on biased or incomplete data. For example, an AI system trained predominantly on data from economically more affluent backgrounds may not take into account the experiences of students from rural or socially disadvantaged backgrounds. This can result in the disproportionate treatment of certain marginalized groups. To address this, policymakers must ensure that algorithms are trained on diverse data and are continuously tested for bias and inclusiveness.

While AI has the potential to democratize education, Khreisat et al. *(2024)* warn that unequal access to digital technology, such as devices with high-speed internet connections, can further deepen inequality. Students from resource-poor backgrounds may be excluded from the opportunities offered by AI, thereby undermining the inclusive nature of the technology. Providing access to the necessary tools and digital connectivity is key to realizing the full developmental potential of AI.

The role of educators is another ethically critical area affected by the integration of AI in education. As AI systems take on more and more educational responsibilities, concerns are growing about the possible devaluation of traditional educational roles and the potential displacement of educators *(Donatus et al., 2024).*

With the introduction of artificial intelligence *(AI)* into classrooms, from automated assessment tools to intelligent instructional systems, there is growing concern that teachers may be seen as redundant. However, rather than replacing teachers, AI should be used to enhance and enhance their capabilities. Teachers have the potential to move from a role as transmitters of knowledge to a role as facilitators of tailored interactive learning. This requires a redefinition of teacher training and education, which should actively include digital and AI literacy, ethical use of educational technologies, and critical pedagogy that empowers teachers to question and shape the way AI is implemented.

The need for changes in educational methodology and the promotion of modernized educational practices that are human-centered and that prioritize ethics and critical thinking at the center of the educational process should also be emphasized. Schiff *(2022)* highlights the importance of preserving a human-centered educational philosophy. In the context of the Republic of Macedonia and the Balkans, this means finding alternatives to the adoption of foreign models of artificial intelligence without local adaptation. Imported educational technologies often do not reflect the linguistic, cultural, and socio-economic realities of students in the region. For example, AI systems that function well in English-speaking or high-income contexts may face difficulties using the Macedonian language or lack sufficient contextual understanding to serve students from rural areas or minority communities.

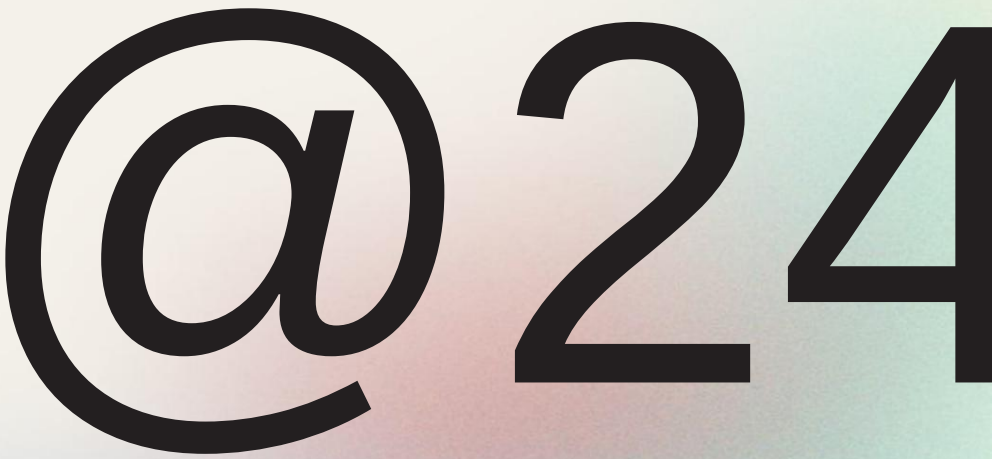
In addition, ethical principles governing the use of AI in education must be clearly defined by decision-makers, educators, and technologists. Researchers *(Nguyen et al., 2023)* propose the development of an ethical code that responds to the unique challenges posed by AI, with a focus on respect, responsibility, and the promotion of the common good. The need for education in AI ethics is essential; developing ethical awareness among future educators and managers will play a significant role in the integration of AI into the education system.

Developing ethical standards for the use of AI in education must be a collaborative effort between educators, policymakers, and technology leaders in the region. This could include partnerships between the Ministry of Education and Science, universities, civil society organizations, and international experts to jointly create guidelines that address ethical issues such as data privacy, algorithmic bias, and equitable access to AI.

The region could benefit from a joint initiative on ethical AI in education, modeled on the European Commission's AI guidelines. Such collaboration would allow countries to pool resources and share best practices, ensuring that AI-based solutions are tailored to the region's educational needs and respect local values. Establishing clear ethical guidelines, supported by transparent practices, is essential to maintaining public trust in educational reforms involving AI.

To maximize the benefits of AI and minimize the harms, educational institutions must adopt a comprehensive ethical approach. This includes creating inclusive technologies, protecting student data, and combating bias in algorithmic systems. Institutions need to critically evaluate the infrastructures and databases on which AI models are based, ensuring that they reflect the diversity of all students. The promotion of ethical AI must include the voices of marginalized communities, ensuring that technology is a bridge, not a barrier, to educational equity.

In conclusion, AI brings great opportunities for improving education, its implementation must be guided by ethical principles. In doing so, accessibility, data privacy, and algorithmic balance must be prioritized, thus creating an innovative and efficient educational model.



International examples applicable in the region

The Estonian model for AI in the public sector: Estonia, a small country with a highly developed digital administration, has implemented several AI projects *(e.g., a judicial AI assistant, a risk prediction system in social services).* Each project is accompanied by an ethical assessment and public debate. In addition, Estonia has issued guidelines for public sector employees on the use of AI, which is a practical step in building institutional capacity. For countries like the Republic of Macedonia, adopting and adapting such guidelines in the local context would be of great benefit – it would help officials understand what AI is, where it can be applied, and what risks they should be aware of *(e.g., using an "AI Check-list" before purchasing an AI software tool to check whether the supplier meets ethical standards).*

Finnish course "Elements of AI": Finland, in collaboration with a university, created a free online course on the basics of AI for the general public, available in multiple languages. The goal was to **educate 1% of citizens** on the basic concepts of AI and to close the gap in understanding. The course also includes an ethics module. This course has already been translated into Serbian and other languages in the region, so it would be useful to provide it in Macedonian/Albanian and promote it. This way, the public would be better informed and less susceptible to unfounded fears or excessive expectations – both extremes that can be harmful. Educational tools are an international example that is relatively easy to implement with a small investment.

Nordic Ethical AI Strategy: In 2024, the Nordic countries came up with joint recommendations for ethical AI, including five strategic actions: creating a joint strategy, establishing a Nordic Center for Ethical AI, promoting language models for Nordic languages, developing skills, and increasing transparency and sustainability of AI. Although the Nordic countries are highly developed, their model of regional cooperation is inspiring. Countries from our region could consider a similar approach – for example, establishing a Western Balkans Working Group or Center for Responsible AI, where experts from different countries would collaborate on common standards, language technologies *(for Macedonian, Albanian, Serbian, etc.)* and exchange knowledge. This would allow for more efficient use of limited resources and strengthen the regional position in negotiations with the EU on the AI Act and similar initiatives.

Bias checking tools (example: IBM AI Fairness 360): Internationally, major technology companies and universities have developed open tools for detecting and correcting bias in al[illegible] released the AI Fairness [illegible] Google offer[illegible] the What-If Tool, and there are also academic librari[illegible]*airLearn, AIF360)* that[illegible]elop metrics f[illegible]ness po[illegible] These tools are used worldwide by organizations that wa[illegible]ndependent verification of the[illegible]models.

Applicability: Institutions in the region *(e.g. [illegible]versity [illegible]oratories or IT [illegible]mpanies)* [illegible]n adopt and use[illegible]e tools when developing their own AI systems. In this way[illegible]d with[illegible]waiting for n[illegible]al regula[illegible]n, self-regulati[illegible]nd verification can already be practiced.

Example: if the Ministry of Health of the Republic of [illegible]nia ord[illegible]system for predicting hospital workload, it may ask the contractor to submit a report on a bias test with one of these tools.

This is a direct lesson from international practice on how to incorporate ethics into the technical development process.

Canada – Algorithmic Transparency in Government: The Government of Canada has introduced a directive requiring all federal institutions using automated decision-making systems to conduct an Algorithmic Impact Assessment *and* publish information about the system *(purpose, logic, anti-bias measures).* This is an excellent example of transparency in the public sector. A similar form or procedure could be adapted in countries in our region – first on a voluntary basis, and then on a mandatory basis.

For example, if the Employment Agency were to use AI in the future to recommend training to unemployed people, it would fill out such a questionnaire and publish it on its website so the public would know: **"We use model X, trained on data Y, tested for gender and age bias, results showed <5% difference between groups, measures foreseen...".**

All of these examples – both regional and international – show that ethics in AI is not an abstract principle, but a set of concrete practices that can be implemented.

Success stories are a true guide, as they show that innovative solutions based on artificial intelligence can be developed while preserving basic human values. The next logical step is to translate these insights into recommendations relevant to the Republic of Macedonia and the region.

Based on the detailed analysis, several strategic directions emerge, the implementation of which will enable the development of artificial intelligence in an ethical, impartial and consistent manner with the best global practices. These recommendations apply to all relevant stakeholders – state institutions, the civil sector, educational institutions and private industry, as only a coordinated approach can bring lasting results.

First, it is necessary for the Republic of Macedonia to finalize and adopt a national strategy for artificial intelligence as a matter of priority, which must explicitly include a chapter on ethical principles and bias management, following the example of countries such as Greece and Bulgaria. This strategy should foresee the establishment of an Ethics Council for Artificial Intelligence, which will develop guidelines and monitor their implementation in the public sector. In addition, the strategy should set clear goals, such as training a significant number of civil servants in trustworthy artificial intelligence or integrating subjects related to this area into study programs.

Furthermore, harmonisation with European regulations, in particular the European Union Artificial Intelligence Act, as well as international standards, is essential. Governments in the region, in particular the Republic of Macedonia, Albania and Serbia, should start aligning themselves with the future European legal framework in a timely manner, which includes identifying a national supervisory body for artificial intelligence, mapping potentially high-risk systems and creating a registry. Adopting the principles of the UNESCO Ethics Recommendation as well as joining international initiatives will contribute to a unified approach and exchange of experiences.

It is also necessary to develop specific ethical guidelines for different sectors, such as health, justice, and security.

These guides should be prepared by experts and practitioners in the relevant fields, respecting general principles, but also addressing the specific dilemmas arising from each domain. For example, the guide on artificial intelligence in human resources should prohibit the use of algorithms that directly or indirectly select for protected characteristics, such as gender or religion, provide for a requirement for transparency towards candidates, and provide for the right to object.

Such guidelines may initially be voluntary, but over time, best practices should be incorporated into legislation. Institutions have an obligation to strengthen their capacities and skills by establishing internal teams or contact points for artificial intelligence. Each ministry should have an advisor trained in data science and ethics, who will provide expert support in the procurement or development of tools. In parallel, it is necessary to organize training for regulatory bodies, such as competition or financial market commissions, so that they can timely recognize cases of bias or violation of rights.

The judiciary, i.e. judges and lawyers, should also be familiar with the basics of algorithmic accountability, as such cases will become increasingly common. The government can collaborate with universities and NGOs to provide such programs.

Universities in the Republic of Macedonia and the region should integrate courses on ethics in artificial intelligence within computer science, but also as electives for law, philosophy or sociology students, thus creating a new cadre that understands both the technology and the social context. In addition, research and projects aimed at fair algorithms and local biases should be encouraged, as well as funding for such initiatives through state and donor funds.

In the private sector, it is advisable for larger IT companies and startups to develop internal policies for responsible AI, through publicly published ethics statements or the establishment of an AI ethics officer position. This proactive approach will provide them with a competitive advantage in markets where such issues are highly valued.

Chambers of Commerce and ICT should organize workshops and develop a code of good practice for the industry, with guidelines on transparency, security and mechanisms for reporting problems or bias. Regional cooperation and resource sharing are key for further progress. Western Balkan countries should establish a regional network for ethical AI through existing initiatives, with the aim of exchanging data, tools [illegible] will enable the develo[illegible]tter natural language processing models, as well as a joint appr[illegible] when applying for Eur[illegible]n funds.

Finally, it is necessary to raise public awaren[illegible]s of the [illegible]tential and ris[illegible] artificia[illegible]telligence throu[illegible]blic debates, forums and media campaigns on ethics. Org[illegible]zing ar[illegible]l conferenc[illegible]d enco[illegible]ging investiga[illegible]journalism will contribute to a better informed public, thus reducing the [illegible] of unf[illegible]d f[illegible]or [illegible]se [illegible]e technolog[illegible]

All of these recommendations are mutually reinforcing and aim to create an ecosystem in which AI will develop in a way that increases well-being and competitiveness, while minimizing harm and injustice. The key is to be proactive – to act now, in parallel with technological development, rather than reactively after problems arise.

Conclusion and next steps

Artificial intelligence represents an opportunity for transformative development of societies – from improving public services and economic growth, to solving complex challenges. The Republic of Macedonia and neighboring countries, although small in terms of market and resources, have a real chance to catch up with global trends, harnessing the benefits of AI while mitigating the risks. To realize this vision, it is necessary to embed ethics and impartiality into the foundations of digital development strategies.

The analysis showed that the region is not starting from scratch: there are already initiatives, strategies in the pipeline, success stories and local talents. However, it is clear that increased efforts and greater coordination are needed. In conclusion, several key points can be highlighted. The vision for the future implies that the region develops "trusted and fair AI" as its recognizable brand – that is, the technology is used for social well-being *(smart cities, better healthcare, agro-innovations),* in a way that is transparent, under control and safe for citizens. The future could include integrated AI systems between countries *(data exchange for disaster prevention using AI, common translation tools for Balkan languages),* which will encourage regional development.

In terms of future regulation, it is expected that in the next 3 to 5 years the Republic of Macedonia will adopt a legal act or rules on AI, especially under the influence of the EU AI Act. In this process, it is important to include local specificities – such as the protection of minorities in language-based algorithms *(which is especially important for multilingual environments),* or the establishment of an oversight mechanism that will also have an advisory role towards the private sector, and not only punished.

Specific activities as next steps

Organizing a series of workshops and trainings under the auspices of the Government and chambers of commerce, dedicated to familiarizing with the European AI Act, using fair AI tools and sharing good practices *(with the possibility of bringing experts from Serbia, Slovenia, Finland for transfer of experiences)*

Establishing an inter-sectoral working group that will unite the ministries of information society, justice, education, together with representatives from the academic community and the IT industry, with the task of working on the implementation of the recommendations and translating them into an action plan. For example, developing a plan for implementing pilot projects for "ethical AI" in several ministries, with clearly defined deadlines.

Providing a fund, through FITR or donors, for projects related to **"AI for Good"** The criterion for success will be both social impact and ethical sustainability. This will encourage startups and researchers to develop solutions that explicitly target social challenges *(e.g., an AI tool for people with disabilities, a system for recognizing disinformation)* with integrated ethical considerations.

Organizing a broad public debate before adopting any legal framework, using digital consultation tools, where citizens, NGOs and businesses can express their views on what kind of AI future they want and what protections they expect. This will result not only in better rules, but also in increased confidence that the process is inclusive.

By implementing these activities, the Republic of Macedonia and its neighbors will take a significant step towards responsible digital transformation. This will mean that the region will not be just a passive user of other people's AI solutions, but also an active creator and controller of technology, in line with its own values.

Finally, it is worth emphasizing that ethics and bias in AI are not **a "one-time task,"** but a continuous process. Technology will continue to change, and with it, our strategies and guidelines will need to be renewed and adapted. Therefore, the most important **"next action"** is to establish a culture of ongoing dialogue and learning between the technical community, policymakers, and the public. Only in this way will we ensure that AI systems serve people, not the other way around.

SWOT analysis: Republic of Macedonia and ethically impartial AI

To assess the readiness and position of the Republic of Macedonia in the domain of ethical and unbiased development of artificial intelligence, it is useful to conduct a SWOT analysis ***(Strengths, Weaknesses, Opportunities, Threats).*** This analysis summarizes some of the previously presented points and provides a picture of where to capitalize and where to be careful in the future.

Strengths

Developed IT community and startup ecosystem: The Republic of Macedonia has a dynamic IT sector, numerous startups *(around 400)* and talented engineers. This foundation enables rapid adoption of new technologies and has already generated successful AI products, indicating a capacity for innovation and awareness of the latest trends.

Support from international partners: The process of creating a national AI strategy is supported by the World Bank and UNDP, and digitalization in general is driven by EU projects. This provides access to expertise and resources that can help embed ethical standards from the start.

Regulatory alignment with the EU: The country already has GDPR-compliant legislation *and* is guided by European regulations in telecommunications and related areas. This acquis approach means that the Republic of Macedonia is well-positioned to quickly adopt future EU AI rules, rather than starting from scratch.

Multicultural society – awareness of non-discrimination: As a country with different ethnic communities, the Republic of Macedonia has developed mechanisms and sensitivity for non-discrimination *(e.g. Ombudsman, anti-discrimination commissions).* This can be built upon in the context of AI – there is an understanding that systems should be fair to all groups.

Weaknesses

Lack of a formal AI strategy and regulation: While there is an initiative, no national strategy, law or regulation specific to AI has been adopted to date. This means that there is currently no clear vision and guidance at the state level, which carries the risk of uncoordinated or unethical implementations.

Limited human resources and "brain drain": The lack of sufficient domestic experts in AI and ethics is a serious problem – many young IT professionals have emigrated. A small number of specialists are tasked to cover many spheres *(academia, industry, state administration),* which makes parallel progress difficult.

Low level of public awareness: The general population and some decision-makers are not sufficiently informed about what AI is, so ethics in AI seem like a distant topic for them. This can result in apathy *(and thus important issues being ignored)* or panic when the first controversies arise. More work is needed on education and promotion.

Fragmented approach: Current AI-related activities are fragmented across different institutions *(FITR, Ministry of Information Society, academia)* without a strong coordination mechanism. This can lead to duplication of efforts or omissions – for example, developing AI solutions without consulting ethics experts, due to a lack of communication between sectors.

Opportunities

Regional leadership: Given that other countries such as Albania and Kosovo do not yet have concrete policies, the Republic of Macedonia has the opportunity to position itself as a leader in ethical AI in the Western Balkans if it quickly adopts the strategy and guidelines. This would bring a reputational advantage and the opportunity to attract projects and investments, such as AI test centers that need to be aligned with the EU.

Inclusion in European programs: Funds for research and capacity building in AI are available through IPA and Horizon Europe. There is a possibility to provide funds for projects such as **an "AI regulatory sandbox",** where Macedonian startups will be able to test solutions under the supervision of regulators, or for educational programs. This would accelerate learning and create their own models of best practice.

Diaspora and return of experts: Many Macedonian professionals work in technology companies abroad *(USA,* EU), specifically in AI and data science. If a favorable scientific and business climate were created in the country *(e.g., an AI research center, startup incentives),* some of them would collaborate or return, bringing valuable knowledge, especially about the ethical standards applied by large companies.

Intersectoral Innovations: The Republic of Macedonia has needs in various areas that could be addressed with AI – such as precision agriculture, pollution prevention, personalized tourism, etc. These are opportunities for the development of AI solutions with a positive societal impact, which, if designed ethically from the start, will serve as a framework and increase public trust.

Threats

Rapid import of foreign technologies without control: The market may be flooded with offers for AI software *(for video surveillance, education, businesses)* from global companies. Th[illegible]hat institutions or companies will adopt tools that are not aligned with our values or legal framework. An e[illegible]ple is the discusse[illegible]allation of Chines[illegible]eras – such solutions could bring about a massive violati[illegible] privacy if not controlled.

Abuses and incidents: If a significant incide[illegible]ccurs – [illegible]xample, an AI[illegible]em in a [illegible]k discriminating agai[illegible]articular ethnic group in loans, or deepfake disinforma[illegible] during[illegible]election – it cou[illegible]eriously u[illegible]ermine trust in AI[illegible] panicky, inappropriate reactions *(outright ban[illegible]stagna[illegible]n technology a[illegible]ion).* This[illegible]a threat if oversight mechan[illegible] and rapid correction are not established befo[illegible]and.

Slow adoption of regulations: If a country delays too long in establishing a framework, it may face pressure to adopt other solutions without adapting them *(for example, the EU VI Act when it becomes binding for members – the Republic of Macedonia would have to quickly implement many provisions, which could be overwhelming for the administration and businesses).* This would create a regulatory shock rather than gradual adaptation.

Public distrust and resistance: If processes are not transparent, there is a risk that citizens will develop distrust and resistance to AI. This can be fueled by conspiracy theories or real-life cases of abuse. Distrust can lead to the failure of useful projects – for example, citizens refusing to share their data even for legitimate purposes, such as better healthcare through AI, for fear of abuse.

How to exploit or mitigate these factors?

The SWOT analysis suggests that the Republic of Macedonia should use its strengths and opportunities – especially human potential and international support – to quickly build the necessary framework and infrastructure for ethical AI. At the same time, it should work on its weaknesses *(accelerate the strategic process, invest in talent retention)* and protect itself from threats through preventive measures *(regulation, educational efforts, transparency).*

In practice, this means investing in people *(training, preventing attrition),* creating clear policies before the mass adoption of commercial AI solutions, and actively involving the public to build trust. If this path is followed, the Republic of Macedonia can position itself as an example of a small country successfully implementing big principles – showing that market size is not an obstacle to establishing high standards in artificial intelligence. This will ensure that technology will truly be at the service of citizens and development, and not a source of new injustices or divisions.

If the recommended next steps are taken in a timely manner, the vision outlined in this white paper – of an innovative, yet ethical and impartial digital Balkans – is feasible and achievable. With the joint efforts of all societal actors, artificial intelligence can become a tool that will contribute to the prosperity of the region, while strengthening the values of fairness, transparency and respect for the human being as the center of technological progress.

Sources used:

- *Serbia prepares for the AI revolution* | ***United Nations Development Program*** - https://www.undp.org/serbia/blog/serbia-prepares-ai-revolution
- *Serbia adopts ethics guidelines for artificial intelligence* - https://www.schoenherr.eu/content/serbia-adopts-ethics-guidelines-for-artificial-intelligence
- The status and future prospects of AI regulation and development in North Macedonia- https://www.schoenherr.eu/content/the-status-and-future-prospects-of-ai-regulation-and- development-in-north-macedonia
- AI principles – **OECD** - https://www.oecd.org/en/topics/sub-issues/ai-principles.html

ÿ Background - **OECD** - https://oecd.ai/en/about/background

- **AI Act** | Shaping Europe's digital future - **European Union** - https://digital- strategy.ec.europa.eu/en/policies/regulatory-framework-ai
- AI Act enters into force - **European Commission** - https://commission.europa.eu/news/ai-act- enters-force-2024-08-01_en

ÿ Using artificial intelligence without a strategy exposes citizens to risk, the state has not yet started a procedure – **Metamorphosis** - https://metamorphosis.org.mk/blog/koristenjeto-na-veshtachkata-inteligencija-bez-strategija-gi-izlozuva-graganite-na-rizik-drzavata-se-ushte-nema-pochnato-procedura-meta-mk-novinska-agencija/

ÿ Macedonia with a national AI strategy - https://fitr.mk/%D0%BD%D0%B0%D1%86%D0%B8%D0%BE%D0%BD%D0%B0%D0%BB%D0%BD%D0%B0-ai-%D1%81%D1%82%D1%80%D0%B0%D1%82%D0%B5%D0%B3%D0%B8%D1%98%D0%B0/

- Serbia Introduces First National Law on Artificial Intelligence (Serbia) - https://www.reviewofailaw.com/Tool/Evidenza/Single/view_html?id_evidenza=3142
- **Ethical guidelines** for development, use of artificial intelligence - https://op.europa.eu/en/publication-detail/-/publication/d81a0d54-5348-11ed-92ed-01aa75ed71a1/language-en

ÿ Bulgaria's AI Development Concept (2020–2030) | **Digital Watch Observatory** Development of Standards in the Use of Artificial Intelligence - https://dig.watch/resource/bulgarias-ai-development-concept-20202030

- PRESS RELEASE: Nordic AI experts launch five recommendations for ethical AI | **Nordic Innovation** - https://www.nordicinnovation.org/news/press-release-nordic-ai-experts-launch-five-recommendations-ethical-ai
- Greece Information Technology **National AI Strategy** - https://www.trade.gov/market-intelligence/greece-information-technology-national-ai-strategy
- Greece Unveils Ambitious **Blueprint** to Lead AI Transforma https://babl.ai/greece-un mbitious-blueprin d-ai-tran ion/
- A Blueprint for Greece's AI Transformation | **Digital ch Obser** https n/reso a-blueprint-for-greeces-ai-transformation

ÿ Digital Agenda of Kosovo 2030 - https://me.rks- net/wp-content/uploads/2023/07/Digital-Agenda-of-K ovo-20 ocx

- Kosovo parties plans for AI: Vetëvendosje - _data
- Kosovo Press Council Updates Code of Ethi Comba ps: nt.com/2024/02/20/kosovo-press-council-updates-code ethics-to-combat-ai-risks/
- **OSCE** launches new Policy Briefs on Media and Informa ://www.osce.org/mission-in-kosovo/587894

ÿ **Adams, C.,** Pente, P., Lemermeyer, G., & Rockwell, G. (2023). Ethical principles for artificial intelligence in K-12 education. *Computers and Education: Artificial Intelligence,* 4, 100131. https://www.sciencedirect.com/science/article/pii/S2666920X23000103

ÿ **Akgun, S.,** & Greenhow, C. (2022). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *AI and Ethics,* 2(3), 431-440. https://link.springer.com/article/10.1007/s43681-021-00096-7

ÿ **Alahmed, Y.,** Abadla, R., Ameen, N., & Shteiwi, A. (2023). Bridging the gap between ethical AI implementations. *International journal of membrane science and technology,* 10(3), 3034-3046. https://www.researchgate.net/profile/Yazan-Al-Alahmed/publication/375091161_Bridging_the_Gap_Between_Ethical_AI_Implementations/links/6540e88bf7d021785f24e88c/Bridging-the-Gap-Between-Ethical-AI-Implementations.pdf

ÿ **Eden, CA,** Chisom, ON, & Adeniyi, IS (2024). Integrating AI in education: Opportunities, challenges, and ethical considerations. *Magna Scientia Advanced Research and Reviews,* 10(2), 006-013. https://text2fa.ir/wp-content/uploads/Text2fa.ir-Integrating-AI-in-education-Opportunities-challenges-and.pdf

ÿ **Khreisat, MN,** Khilani, D., Rusho, MA, Karkkulainen, EA, Tabuena, AC, & Uberas, AD (2024). Ethical implications of AI integration in educational decision making: Systematic review. *Educational Administration: Theory and Practice,* 30(5), 8521-8527. https://www.researchgate.net/profile/Maher-Rusho2/publication/380855490_Ethical_Implications_Of_AI_Integration_In_Educational_Decision_Making_Systematic_Review/links/66519a2422a7f16b4f4ab627/Ethical-Implications-Of-AI-Integration-In-Educational-Decision-Making-Systematic-Review.pdf

ÿ **Leta, FM,** & Vancea, DPC (2023). Ethics in education: Exploring the ethical implications of artificial intelligence implementation. *Ovidius University Annals, Economic Sciences Series,* 23(1), 413-421. https://ibn.idsi.md/sites/default/files/j_nr_file/Full%20Vol.%20XXIII%20Issue%201%202023.pdf#page=427

ÿ **Nguyen, A.,** Ngo, HN, Hong, Y., Dang, B., & Nguyen, BPT (2023). Ethical principles for artificial intelligence in education. *Education and Information Technologies,* 28(4), 4221-4241. https://link.springer.com/article/10.1007/s10639-022-11316-w

ÿ **Nur, N.,** Goh, SJ, Patel, J., & Mizrahi, M. (2024). Navigating the ethical landscape of AI integration in educational settings. In *INTED2024 Proceedings* (pp. 7654-7663). IATED. https://library.iated.org/view/NUR2024NAV

ÿ **Reiss, MJ** (2021). The Use of AI in Education: Practicalities and Ethical Considerations. *London Review of Education,* 19(1), n1. https://eric.ed.gov/?id=EJ1297682

ÿ **Tang, L.,** & Su, YS (2024). Ethical implications and principles of using artificial intelligence models in the classroom: A systematic literature review. https://reunir.unir.net/handle/123456789/16210

Footnotes

- **UNESCO** Recommendation on the Ethics of Artificial Intelligence (2021) – https://unesdoc.unesco.org/ark:/48223/pf0000381137
- The **OECD AI Principles** promote the use of AI that is innovative and trustworthy and that respects human rights and democratic values. Adopted in May 2019, they set standards for AI that are practical and flexible enough to stand the test of time. – https://oecd.ai/en/ai-principles
- **EU Artificial Intelligence Act.** The AI Act is the first-ever legal framework on AI, which addresses the risks of AI and positions Europe to play a leading role globally. – https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-AI
- **IEEE Ethically Aligned Design.** New technology comes with unknown obstacles and unintended risks requiring accountable design and lifecycle planning to ensure responsible innovation. – https://ethicsinaction.ieee.org/

ÿ **National AI Strategy – Serbia (2020–2025)**
https://nitra.gov.rs/cir/ministarstvo/vesti/usvojena-strategija-za-razvoj-vestacke-inteligencije-u-republici-srbiji-za-period-od-2025-do-2030-godine

- **AI4Media** project Strengthen Europe's Excellence on AI for Media, Society, and Democracy and ensure the development of ethical & trustworthy AI! (Horizon 2020) – https://ai4media.eu/
- Reports from **UNDP, ITU, GIZ, Digital Europe**

ÿ Examples: **Pixyle.ai, EmbedSocial, FINKI research**

- **World Economic Forum, Brookings Institution**

Final note from the authors

Ethics and bias in artificial intelligence are not just technical challenges, but fundamental issues of trust, justice, and social responsibility. At a time when new technologies are becoming more deeply embedded in our everyday lives, the need for critical awareness, transparency, and ethical balance is greater than ever.

This document is intended for all stakeholders: government institutions and policymakers, educational and research institutions, civil society organizations, private companies, startups, technology developers, media, as well as all individuals participating in or affected by the digital transformation. It aims to be a basis for institutional dialogue, interdisciplinary partnerships and thoughtful policymaking.

We believe that only through cooperation between the state, academia, civil society, and industry can we build a digital future that is fair, sustainable, and truly human-centered.

We encourage everyone to use this paper as a tool for analysis, education, and action toward a fair, safe, and socially responsible application of artificial intelligence.

AI Now - Association for Artificial Intelligence
AINOW - Association for Artificial Intelligence
Skopje, 2025
contact@ainow.mk | www.ainow.mk