

Original Scientific Paper/Original naučni rad
Paper Submitted/Rad primljen: 31.12.2025.
Paper Accepted/Rad prihvaćen: 20.01.2026.
DOI: 10.5937/SJEM2600001M

UDC/UDK: 004.8:17

Etičke dileme i društveni izazovi: Ko će preuzeti odgovornost za zloupotrebu VI?

Ida Manton¹

¹ Prague University of Economics and Business, ida.manton@gmail.com

Abstract: Članak se bavi fenomenom veštačke inteligencije, moralnim dilemama koje proizilaze iz njene široke upotrebe i neophodnosti regulisanja te upotrebe. Dok žurimo ka stvaranju potencijalno najopasnijeg agensa automatizovanog donošenja odluka, moramo postaviti pitanje: čija je odgovornost da kontroliše zloupotrebu VI i potencijalne štete koje će naneti našim društvima – države, tehnoloških korporacija ili pojedinca? Informacija je najvredniji resurs našeg vremena. Informacije su oduvek bile dragocene, ali načini na koje im pristupamo su se promenili, a sa tim i strukture i metode kojima se dele. Ljudi su oduvek bili vrednovani na osnovu svog znanja, na osnovu toga koliko dobro su mogli da koriste ono što su znali i koliko kreativnosti je moglo proizaći iz informacija koje su prikupili. Period koji je prethodio ovoj novoj realnosti vođenoj VI dao nam je uvid u poteškoće sa kojima naša nacionalna i međunarodna tela mogu da regulišu i donose zakone o upotrebi nepoznatih tehnologija.

Ključne reči: veštačka inteligencija, etika, odgovornost, regulacija, društveni uticaj, demokratija, upravljanje

Ethical Dilemmas and Social Challenges: Who Will Take Responsibility for AI Misuse?

Abstract: The article looks into the phenomenon of artificial intelligence, the moral dilemmas rising from its widespread use and the necessity for regulating that use. As we are rushing towards creating the potentially most dangerous agent of automated decision-making we need to ask the question: whose responsibility is it to control the misuse of AI and the potential damages it will inflict on our societies – the state, the tech corporations, or the individual? Information is the most valuable resource of our time. Information has always been precious, but the ways we access it have changed, and with that, so have the structures and methods by which it is shared. People have always been valued based on their knowledge, on how well they could use what they knew, and on how much creativity could stem from the information they gathered. The lead up to this new AI-driven reality has given us a preview of the difficulty with which our national and international bodies can regulate and legislate the use of unknown technologies.

Keywords: artificial intelligence, ethics, responsibility, regulation, social impact, democracy, governance

1. Ethical Dilemmas and Social Challenges: Who Will Take Responsibility for AI Misuse?

The text therefore explores if the Artificial intelligence is intelligent, whether it is a tool that will be useful for humanity or will make people subservient to machines, and also what dangers are preventable and what is an integral part of the system, which we have to accept with making AI fundamental part of our lives. A part of the text will deal with the effects AI has on education, social dynamics that are already showing signs of pathologies which AI will only exacerbate, as well as the effect it will have on furthering polarization and creating echo-chambers that amplify dangerous group-think and changes in social and political interactions.

And we often do it before we even see the full potential of the well-intended invention. We did that with almost all political models and we saw very wrong deviations, which compellingly became arguments for witch-hunting rather than advanced thinking about how to repair and upgrade.

2. The dangers of advancing AI

Introducing AI in our everyday life, on our personal computers, our social media pages and even every single message we send to another human, poses a huge ethical dilemma. It makes me think that the Universe is spitting at us a scrambled response carrying some meaning along the lines of “be careful what you wish for” knowing humanity has a tendency to overdo, to exaggerate and to turn sour even the most helpful tools and ideas. We have turned cure into poison before. And even just a naturally occurring bacteria, like Anthrax, we have turned into a bioweapon. In Chinese alchemy, the elixirs prepared to prolong life, were found to be the cause of death to a few Emperors and noblemen because of the heavy metals they contained. But we have also experimented and turned the notorious “poison of the Kings”, the Arsenic, into a cure for syphilis, and mold into antibiotic. So, the core determinant of whether something will be used as an antidote or a poison, as helpful or detrimental, even devastating, is the user’s intent which determines the impact. This is how our story with the use of AI will untangle. The difference is that all the other inventions, be those cures or poisons, were tools, just products, vehicles, or trinkets. AI is the first invention we humans created and are in a rush to amplify it to a level where it surpasses us in, what we thought to be the best of all spices in – thinking! No other human form is a more advanced thinker than the human being and now we seem not to trust ourselves with this skill as much as we trust AI technologies, that already proved to have a better and wider scope of knowledge assembling, analyzing in mere seconds and producing content better than any of us, especially when confronted with time and insurmountable quantities of content. But does that make the Artificial Intelligence intelligent and even more, does it make it smarter than us? It sure is created to be, to outmatch us in our human intelligence. The promise and hope, is that AI can invent and create things that are beyond our imagination, but at the same time this unpredictability leaves us humans in a vulnerable position where we cannot predict what AI will come up with or do. The threat is that how it develops is beyond our control and therefore we cannot guarantee that AI will comply with our instructions because it has a “mind” of its own. Yuval Noah Harari intriguingly asked recently at the World Economic Forum 2026 in Davos: Will AI challenge our (human) supremacy in thinking? (Harari, 2026)

Additionally, how do we entrust AI with truth, shared moral concepts and integrity and other more fluid areas that enjoy a huge range of positioning among humans? As creators of AI, as digital content creators, we all provide material that can then be used by algorithms for a curated AI content, so anything becomes potentially repeated and regurgitated content, regardless of whether it is truth or a lie, disinformation or fact. “This means AI has no understanding. No consciousness. No knowledge in any real, human sense. Just pure probability-driven, engineered brilliance — nothing more, and nothing less”.

A less obvious potential problem is based on the hypothesis that even if all the creators of AI content had genuinely good and truthful intention, the system would not be absolutely truthful, as the whole is never just a sum of the particles because the interactions, the synergies create new qualities. Science has showed this to us. We have explored “emergence” as a concept and have identified the element of unpredictability, or as Robert Musil calls what he explores in his novels, the “imaginary unit-i”. That is if everything uploaded in the AI systems was morally unquestionable, which is of course an absolute and improbable situation, which does not reflect our multilayered, cross-cultural reality. My truth can be someone else’s lie. Still an intellectually stimulating though comes to mind - would even in such absolute circumstances AI technologies be truthful, can they employ restraint and avoid escalation and polarization, can they operate benevolently, and even more, can they be trained to know how to avoid human suffering of any kind. So far, or shall I say already, we have seen the opposite. What we know from the recent military operations in Gaza and Ukraine, is that the AI-DSS (AI decision support systems) can create “kill lists”, identifying and eliminating targets with facial recognition technologies (FRT) in complete disregard of IHL (international humanitarian law).

This leaves no space for hope that we will be spared of the AI’s dark side. And if that is a starting point in exploring how to prevent the dangers approaching humanity with hypersonic speed, we have to start creating institutions, SOPs, experts and legislation that will slow down the unhinged use and creation of AI models that violate at least the already existing laws, like data privacy, to name just one in addition to the whole body of IHL, which I mentioned earlier. And while we do that on communal, national and international level, we have to broaden the discussion to understand AI. Here I offer just a few thoughts for such necessary discussion.

Both words describing this phenomenon are challenging. “Artificial” refers to a human creation, an artifact. In Aristotelian terms, an artifact is something that exists by craft and has its origin in the craftsman in the form of the thing as it exists in the mind of the maker. When it comes to “intelligence”, if it is narrowly understood as information gathering, we have already lost the battle to AI. Luckily, intelligence has many features and quite significant part of it is emotional, kinetic, abstract and artistic - fields that require our human existence and values our neuro-divergence, our different points of view and our need for dialogue to find common ground, mutual

understanding, morality and even a worldview. In all of these aspects of modern human life, technologies already have massive roles – from how we interact with other humans, to our office life, diplomacy, film industry, video gaming, performing arts, wellbeing and health, all handled through apps and computer programmes that measure and collect our data. The eventual goal of this information gathering exercise is that the technology we use will predict what we want and need, which services we require and soon AI will be able to provide those without humans interfering in that process. Automated, however, still does not mean artificial intelligence, and our experiences so far show us how stupid technology is even though we name it “smart”. We all catch ourselves hating the logjams created by machines more than valuing its usefulness. Too many times, in just a single day. But AI is different from voice mail machines and a milk foam maker. AI is not just producing or providing the service. AI is ordering and delivering, which then excludes the will and the intentions of the people, of its creator.

The key question therefore is who are the humans behind the creation of a world based on AI. Can they be trusted? What are their motives and how do they see the endgame? How is this new “gadget” affecting power, governance, politics, social contracts, security and how can it disrupt the already ruptured tissues?

Let us first review what we know about the behavior and interests of the tech giants from this relatively short phase of our digital life on this planet. With the “democratization” of the media that happened under minimal regulation, content was being uploaded without restrictions on sharing or mindfulness of intellectual property regulations already in place. A few IT guys earn money for something that the journalist who wrote the article is paid far less for. Almost 80% of the income went to Google and Facebook. Google has already been fined for violating competition laws in the European Union. A few years ago, the US Department of Justice (n.d.) filed a lawsuit against Google for monopolizing the online advertising market, accusing it of using “anticompetitive and illegal methods to eliminate or drastically reduce any threat to its dominance over the technologies used for digital advertising”. Facebook was the subject of criticism, but also of lawsuits related to the management of the information it possesses, the way it transmits news, identifies users and recognizes their faces, protects privacy, and creates an addiction that is often compared to drug addiction. The Cambridge Analytica scandal, detailed and explained in the documentary “The Great Hack” (The Great Hack, 2019), shows how personal data collected to build psychological profiles (through applications such as “This is your Digital Life”), revealed a very dangerous side of digital tools — that they are abused to influence electoral processes, as was the case with the 2016 US Presidential Election and the Brexit referendum. With this track record, almost the same actors leading the tech giants and investing outrageously large sums of money in developing AI systems, being sceptic about the benevolence of the AI is the least we can do.

By now we have access to published research and accounts by whistleblowers and reliable investigative journalists analyzing and exposing the monetization strategies employed by tech billionaires to generate profit by promoting polarization, misogyny and violence of various kinds. They expose the financial dynamics at play, key actors, the narratives they deploy, and the tactics they use to operate as a scalable business model that often develops to level of capturing states. One such eye-opening insight was provided by GNET, and it discusses Engagement Farming and the Tactics Behind Incendiary Online Content (Global Network on Extremism and Technology, n.d.). The text explains that: “Engagement farming exploits the political economy of social media platforms, specifically algorithms that prioritise emotional responses. Adversarial actors intentionally curate inflammatory, misogynistic content to trigger indignation and counter-speech”. This being the case with online content, should already raise all the flags for introducing AI, which uses all of this trolled, biased and warped content, as there are no restrictions to what data the AI creators are using to build up their models.

Additionally, the potential for misuse of combined data which leads to creation of digital IDs that are then in the hands of people like Peter Thiel, Elon Musk, David Sacks and other “Tech bros” and billionaires, was predictable. DOGE and Palantir infiltrated the US Government and the White House, harvested classified information and documents from agencies and departments they dismantled (like USAID) have awarded themselves enormous federal funds and made even more money than the investments they made into the political campaigns to bring Trump into power. In March, Trump signed an executive order requiring all agencies and departments of the federal government to share data under the pretense that he was “Stopping Waste, Fraud, and Abuse by Eliminating Information Silos”.

To get the job done, Trump chose Palantir Technologies. Palantir, “sells an AI-based platform that allows its users – among them, military and law enforcement agencies – to analyze personal data, including social media profiles, personal information and physical characteristics. These are used to identify and surveil individuals”. Palantir, was also reported to have built the deportation software for Trump. So, the symbolism behind the name became chillingly real and so did the role Palantir and similar tech companies have in the war on democracy and liberal values.

Some of the features amplified by AI were already available in the pre-AI phase and were concerning because they were used to incite violence, manipulate elections, intimidate political opponents etc. On top of all the information collection, AI offers upgraded features for manipulating information and immediate action - detects patterns, analyzes data, provides options for strategic approaches and much more. In the hands of people who want to run the world unchallenged, AI is dangerous and arguably devastating invention. The reveal of the work Palantir has conducted for ICE (US Immigration and Customs Enforcement) and its investigative branch HIS (Homeland Security Investigations), showed that there is capacity to track these developments by civil society groups. However, despite the work done by an immigrant legal rights group, called Just Futures Law (Just Futures Law, n.d.) the second Trump administration awarded a \$30 million contract to Palantir to build the government a new platform called ImmigrationOS that will “service Ice branches beyond HSI, and aims to “streamline” the identification and deportation of immigrants. Revealing and publicizing the deals and how they affect the American society has mobilized more people in variety of local and national organizations working on digital rights, regular protests and campaigns, like No Tech for Ice. They demand responsible government, safe digital space and data privacy, all of which are being traded and sold as part of multimillion contracts between the Government agencies and the tech billionaires who now have unrestricted access to data and no legal or moral restriction in how to use it.

3. Most Often Discussed Negative Effects of AI

Among the concerns across academic analysis, policy debates, and regulatory discussions, the most relevant for this paper can be grouped as those pertaining to safety, economy and relationships. Among those falling under safety, including security and societal risks, most discussed are misinformation and manipulation (including deepfakes, synthetic media, automated propaganda, political micro-targeting, and election interference), cybersecurity threats (automated hacking and malware generation, exploitation of vulnerabilities accelerated by AI tools, AI-enabled cyberattacks on critical infrastructure), military misuse (like lethal autonomous weapon systems - LAWS), privacy erosion (mass surveillance through biometric tracking, Inference without consent) as well as bias, discrimination, and deficient outcomes due to biased algorithms in hiring, policing, healthcare etc. The economic risks primarily talk about job replacement and re-skilling challenges. On a broader societal level, there is fear of loss of human oversight and control, unpredictability, hallucinations, memory gaps, black box logic, as well as large-scale societal destabilization.

Maybe the most dangerous aspects of AI that we need to regulate is the human responsibility to avoid abdicating our traditional superpower – to think? And this is a responsibility that we will have to nurture as individuals, as educators, as parents, as societies and as humanity. The question still remains how and who will take a lead on each step of these pathways. Also, will it suffice if we have responsible individuals and irresponsible network of technofeudalists? As Varoufakis warns in his book “Technofeudalism: What Killed Capitalism”, the traditional capitalism has been replaced by a new system where big tech companies act as modern-day feudal lords, extracting “rent” from users through their platforms. Big tech firms like Google, Amazon, and Facebook have become the new ruling class, extracting value not through traditional profit from selling goods, but through “rents” from their platforms (Varoufakis, 2023). As platforms that live off of engagement and attracting (maybe more correct terminology would be luring) visitors to their spaces to interact, shop and do almost everything through them – from ordering food, to engaging with your doctor, presenting your work etc., it is unlikely that they will add ethical considerations if their profits are decreasing. Unless restricted by law-makers, they will have no incentives to be an ethics champion.

In the field where I have specialized over the years—international negotiations—there have already been many hopes and many warnings about the role AI might have. A Harvard-led research paper suggested that more research will have to be done in order to “define the complementary roles of AI and human negotiators, ensuring that AI supports, rather than undermines, the complexities of negotiation practice”. And this is where we stand when it comes to “high-stakes workplaces where human judgment, relationship-building, and adaptability are essential—such as crisis response, social work, and labor mediation—AI should be designed to support workers’ expertise rather than automate complex, context-sensitive decisions”. The main concerns expressed by negotiators, as part of this research were confidentiality and hallucinations/bias.

In addition to all of this, AI is likely to reduce creativity if we rely on its intellectual capacity without entering the process prepared and with expectation that AI will be a control function, not the content creator, especially in areas where we are not experts. So, whether AI turns into an assistive technology, efficient expert that will interact with humans or an uncontrollable foe, will depend on what we make of it, how lazy we become and what will our societies, competition and other circumstances force us to do to stay in the game.

4. Ethical dilemmas of using AI in political and social contexts

The rapid expansion of artificial intelligence raises profound ethical dilemmas that cut across power, autonomy, democracy, inequality, and human dignity. Thinkers such as Harari (2026), Zuboff (2019), Noble (2018), and Eubanks (2018), warn that AI is not merely a technical innovation but a political force capable of reshaping societies at their core.

One major category of concern relates to power and human agency. Harari argues that AI's growing ability to "hack" human psychology threatens the foundations of democracy. By analyzing biometric and behavioral data, AI systems can influence emotions, beliefs, and political preferences, blurring the line between legitimate persuasion and unethical manipulation. Personalized political messaging, while efficient, risks undermining individual autonomy by nudging citizens toward decisions they do not fully understand or consciously choose. As people increasingly defer judgment to algorithms, human agency itself is weakened. This creates an ethical dilemma: does increased efficiency justify outsourcing understanding, decision-making, and responsibility to machines, or does this lead to a form of "digital serfdom" in which humans are optimized rather than empowered?

A second set of dilemmas concerns knowledge, truth, and democratic processes. Democracy depends on shared facts and meaningful deliberation. Without them, political disagreement turns into manipulation and confusion. This raises difficult questions about whether AI-generated political content should be restricted, even if such restrictions limit freedom of speech. Closely related is the problem of algorithmic polarization. Engagement-driven systems tend to amplify outrage, filter information, and exaggerate extremes, pushing societies toward radicalization. Also, massive information asymmetries are emerging. Governments and tech corporations increasingly possess far more behavioral data than ordinary citizens, shifting political authority and changing decision-making practices.

AI also intensifies dilemmas related to inequality and social justice. Extensive research has shown that algorithms can encode and amplify existing racial, gender, and socioeconomic biases (Eubanks, 2018; Noble, 2018). Decisions shaped by biased data may appear neutral and objective while masking structural discrimination, making accountability harder to establish. Additionally, unequal access to AI technologies risks deepening global and domestic inequalities. Wealthy states, corporations, and political actors are better positioned to exploit advanced AI for influence and control, raising doubts about whether AI truly democratizes knowledge or instead entrenches elite power. Also, based on the content it is fed, the languages it recognizes and adopts, it is very likely to have a racial, western and cultural bias. While employment and automation are often framed as economic issues, their political consequences are significant. Job displacement and social insecurity can undermine democratic participation, and economic exclusion can translate into political marginalization, among the many desired occurrences by authoritarian leaders.

Surveillance, control, and political authority form another critical area of concern. While surveillance is often justified in the name of security or efficiency, it raises difficult ethical questions about privacy, freedom, and consent. AI-powered monitoring risks transforming liberal democracies into data-driven totalitarian systems. Predictive policing and algorithmic risk assessments further complicate matters, as they frequently misrepresent marginalized communities and raise serious concerns about due process. If an algorithm predicts criminal behavior, should authorities act on that prediction, and if so, how? Even more extreme is the use of AI in autonomous weapons systems. Delegating decisions over life and death to non-conscious machines challenges foundational ideas of moral responsibility and accountability.

AI also affects identity and culture. Global AI systems often reflect Western, corporate, or dominant cultural assumptions, leading to concerns about cultural cognitive imperialism. This raises the question of whose values are embedded in the algorithms that increasingly shape global communication and culture. The rise of AI companions, therapy bots, and political chatbots introduces dilemmas about authenticity and human relationships. Simulated care, intimacy, or political solidarity may blur emotional boundaries and manipulate trust, raising ethical questions about whether such simulations are acceptable.

Finally, issues of governance, accountability, and moral responsibility cut across all these domains. Many AI systems operate as opaque "black boxes," making it difficult to understand how decisions are made or who is responsible when harm occurs. When an AI system causes damage, responsibility is fragmented among developers, deployers, users, politicians and regulators. Moreover, there is a growing tendency to treat AI as a technical fix for governance and bureaucracy. So far, the use has not been very sophisticated, but when that happens the databases from different agencies could come together into a larger system which can generate digital profiles of citizens. Additional dangers might include sidelining of democratic debate due to overreliance on AI

recommendations and shrinking of the space for collective decision-making, which could easily be delegated to AI to avoid public participation and to avoid the push backs from any opposition.

The hallucinations, biases and disconnected argumentation are going to be part of AI for a long time. Whether they will be fixed, fought or even challenged, will depend whether humanity will catch up with big tech and will demand a more responsible development of the systems or the race will justify the means and AI, like many other potentially useful inventions, will become a serious threat to the healthy society, human relationships and a threat to peace on the planet whose existence is challenged because AI's expansion already challenged basics like water, energy and mining.

5. AI, Power, and the Future of Democracy

Artificial intelligence is often framed as a technological revolution, but its deepest impacts are political and ethical. AI is not just another tool—it is the first tool capable of shaping the very human minds that created it. This is the heart of the dilemma: we are deploying a technology that can read our weaknesses, predict our choices, and influence our behavior faster than democratic institutions can respond.

Already now, AI systems generate realistic images, voices, and narratives that make truth negotiable and trust fragile. When citizens cannot distinguish fact from fabrication, democratic debate collapses. The result is not disagreement but disorientation—a political environment where the loudest algorithm, not the best argument, prevails. No society can sustain democratic legitimacy without a shared reality. But the ethical dilemmas extend beyond truth. AI amplifies existing inequalities by embedding the biases of the past into decisions about the future. Predictive policing, algorithmic hiring, and automated welfare systems can quietly punish the already marginalized, creating a digital caste system dressed up as efficiency. At the same time, concentration of data in the hands of a few governments and corporations threatens to undermine human agency (Zuboff, 2019). If algorithms know us better than we know ourselves, the space for genuine autonomy narrows.

Perhaps the most profound ethical question is whether democratic societies can preserve the dignity of human judgment in an age when machines can outperform us in analysis, persuasion, and prediction. The temptation to outsource political decisions to “neutral” algorithms is growing. But politics is not a math problem; it is a moral project. Handing our collective choices to opaque systems risks turning citizens into spectators, governed by forces they cannot see or challenge.

AI's promise is real, but so is the danger that it becomes a technology of domination—one that manipulates emotions, films populations, and concentrates power at unprecedented scale. Ethical governance must begin with protecting human agency and maintaining democratic control over the systems that increasingly shape our lives. The future of AI is not just about innovation. It is about who we become when our choices, beliefs, and relationships are filtered through machines that do not share our values, but can profoundly influence them. As part of the great power competition, the big question will be who will have a bigger cloud service and how its development will affect the electricity grids and expansion to emerging markets.

The question, therefore, is not whether AI will change society, but rather who gets to decide how and whether democracy survives the cast-creating cosmos.

6. AI and Education

Education is already a challenge and the transformation has brought both innovation and laziness.

AI can add a lot in education, but the worst damage we already observe is that in this very competitive field, where quantitative indicators decide which school one will go to or whether they will get a scholarship or not, students are not encouraged to learn how to study and the system replaces their curiosity with ingenuine products produced by AI.

So far, we have been defensive, but soon we will have to teach how to use AI responsibly, how to teach assisted by AI, not expecting AI to compensate for lack of substantive knowledge, teaching materials and lesson plans, how to evaluate students' work and making a distinction between AI generated content and genuine input, etc.

While AI may replace humans in some roles, “it also opens up new opportunities in sectors that demand complex decision making, emotional intelligence, and creative skills—attributes that AI cannot replicate. Understanding these trends is crucial for future workforce preparation”. Education and training will need to adapt to help people transition to roles where human expertise remains irreplaceable.

7. How States Regulate Negative Effects of AI

Regulatory approaches to artificial intelligence vary across jurisdictions, but they generally cluster into several recognizable models.

Horizontal, comprehensive AI laws aim to regulate AI across sectors through overarching legal frameworks. The most prominent example is the European Union's AI Act. It introduces a risk-based system that categorizes AI uses as unacceptable, high-risk, limited-risk, or minimal-risk, with the strictest obligations applied to systems affecting biometric surveillance, critical infrastructure, and safety-sensitive applications. Enforcement is carried out through national supervisory authorities, coordinated at the EU level by the EU AI Office.

Similarly, the Council of Europe (2024) adopted the Framework Convention on AI in 2024, the first binding international treaty focused on AI and human rights. Notably, participation extended beyond EU member states to include countries such as the United States, the United Kingdom, and Canada.

A second model consists of sector-specific regulation, where AI is governed through existing legal frameworks rather than dedicated AI laws. Many states regulate AI indirectly via data protection regimes such as the EU's GDPR, California's CCPA/CPRA, and Brazil's LGPD, as well as through consumer protection law. Additional sectoral rules apply to areas like medical devices and healthcare AI, autonomous vehicles, and financial services, including algorithmic trading.

A third category focuses on AI safety and frontier model governance, particularly for advanced or general-purpose systems. Examples include the United States Executive Order on Safe, Secure, and Trustworthy AI issued in 2023 (United States Executive Order 14110, 2023), the establishment of the UK AI Safety Institute in the same year, and ongoing efforts in countries such as Japan, Singapore, and Canada to develop AI assurance, evaluation, and audit schemes.

Finally, national security and export control regimes play an increasingly important role in AI governance. The United States has imposed export controls on advanced semiconductor chips and, in some cases, model weights. The European Union applies dual-use regulations to surveillance and related technologies, while China has introduced rules governing generative AI providers, including requirements for synthetic media labeling.

8. International Bodies That Govern, Legislate, or Restrict AI Misuse

There is currently no single global authority governing artificial intelligence, but a range of international and regional organizations play important roles by creating binding treaties, soft-law instruments, and influential technical standards.

Some bodies are responsible for binding or quasi-binding instruments. The Council of Europe is particularly significant, having adopted the Framework Convention on AI in 2024, the first binding multilateral treaty dedicated to AI. The convention focuses on safeguarding human rights, democracy, and the rule of law, and is open to participation by non-European states. The European Union, its EU AI Act, alongside related legislation such as the Digital Services Act, the GDPR, and the Data Governance Act was a huge achievement, especially as these frameworks often have extraterritorial effect, applying to AI systems and services that reach EU citizens regardless of where providers are based. Within the United Nations system, no global AI treaty exists yet, but governance is emerging through multiple bodies. UNESCO's Recommendation on the Ethics of AI

adopted in 2021 by 193 states (UNESCO, 2021), provides a widely endorsed normative framework, while the UN Human Rights Council has issued non-binding resolutions addressing AI and human rights.

Technical and security dimensions are addressed by entities such as the International Telecommunication Union, which develops interoperability and technical standards; the UN Office of Counter-Terrorism, which focuses on preventing AI-enabled terrorism; and UNIDIR, which conducts research on AI in weapons systems and arms control. In addition, the Wassenaar Arrangement contributes indirectly to AI governance by imposing dual-use export controls on surveillance and intrusion software technologies that can include AI-enabled security tools.

Alongside these, a number of global multistakeholder and standard-setting bodies exert significant influence despite their non-binding nature. The OECD's AI Principles, adopted in 2019, are among the most influential soft-law instruments and have informed G7 initiatives, EU regulation, and national AI strategies. The G7's Hiroshima AI Process, developed between 2023 and 2024, produced a Code of Conduct for developers of advanced AI systems, with a strong emphasis on frontier model safety and transparency. Technical standards are further developed by ISO and IEC, which publish standards on AI safety, risk management, and auditing that are

frequently incorporated into national laws and regulatory frameworks. The Global Partnership on AI, with more than 44 member states, operates as an advisory and research-driven forum promoting responsible AI development and deployment. The partnership is guided by a Council, Plenary, and Steering Group, with support from the OECD Secretariat.

Given the focus of this publication, we should look into security and military-focused bodies that are increasingly engaged in AI governance. Within the UN Convention on Certain Conventional Weapons, states continue discussions on lethal autonomous weapons systems, and while no binding treaty has yet emerged, momentum toward regulation is growing. Similarly, the OSCE addresses AI-related risks across a range of areas, including media freedom, disinformation, cybersecurity, election integrity, and the human rights impacts of digital technologies. NATO has also developed an AI strategy for defense applications and adopted principles for the responsible use of AI in military contexts in 2021, shaping how AI is integrated into collective defense and security operations. Important to notice at the end is the difference between legislative and advisory documents. While we see increasing production of documents with more informative and advisory, even predictive character, we do not see the same pace with legislating and restricting development of AI or some coordinated action for minimizing economic shocks, electricity bills, water management and labor redistribution.

9. Conclusion

Regulating and legislating AI is often mistaken for obstructing its development and that causes fear of bureaucratizing or slowing down the development because those involved cannot afford to lose the race. The business stakes are too high, but so are the societal and anthropological stakes that change the basic features of our safety, security, peace, and ethical frameworks that guide us.

This article has examined artificial intelligence not simply as a technological innovation, but as a transformative force reshaping power, responsibility, democracy, and human agency. From its rapid and largely unregulated integration into everyday life to its deployment in governance, security, education, and warfare, AI exposes deep ethical dilemmas that existing political and legal systems are struggling to address. The discussion has shown that AI intensifies long-standing structural problems rather than replacing them. It amplifies inequalities by embedding bias into automated decision-making, concentrates power through control over data and infrastructure, and erodes democratic processes by manipulating information, accelerating polarization, and weakening shared reality. At the same time, surveillance technologies, predictive systems, and autonomous decision-making tools challenge fundamental principles of privacy, due process, and human dignity. These developments are not hypothetical. They are already shaping law enforcement, migration control, warfare, education, and political participation.

At the core of these dilemmas lies the risk of human abdication of judgment, responsibility, and critical thinking, in favor of systems optimized for efficiency rather than moral reasoning. While AI can support human expertise and creativity, treating it as a neutral authority or a substitute for political deliberation risks transforming democratic societies into technocratic or data-driven authoritarian systems. Regulation, therefore, should not be understood as an obstacle to innovation, but as a necessary condition for preserving autonomy, accountability, and trust.

The analysis of regulatory approaches and international governance frameworks demonstrates that while important steps have been taken—particularly within the European Union, the Council of Europe, and UNESCO—global AI governance remains fragmented, slow, and reactive. The speed of technological development far outpaces institutional capacity, and the geopolitical race for dominance often undermines ethical restraint. Without stronger coordination, enforceable standards, and meaningful public oversight, AI risks becoming a tool of domination rather than empowerment.

Ultimately, the challenge posed by artificial intelligence is not whether it will change society, but how and under whose control. The future of AI is inseparable from the future of democracy, social justice, and human dignity. Preserving these values requires resisting the temptation to outsource moral and political responsibility to machines and reaffirming the human capacity to think, judge, and decide collectively. AI may shape the tools of governance, but it must not be allowed to replace the ethical foundations on which free and just societies depend. That is something we, the people, will have to reenvision and renegotiate after this rupture in the world order.

References

1. Council of Europe. (2024). Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law. <https://www.coe.int>

2. Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press.
3. European Union. (2024). Regulation (EU) 2024/1689 of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Official Journal of the European Union.
4. Global Network on Extremism and Technology (GNET). (n.d.). Engagement farming and the tactics behind incendiary online content. <https://gnet-research.org>
5. Harari, Y. N. (2026, January). Address to the World Economic Forum. World Economic Forum Annual Meeting, Davos, Switzerland.
6. Just Futures Law. (n.d.). No Tech for ICE Campaign. <https://www.justfutureslaw.org>
7. Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. NYU Press.
8. Organisation for Economic Co-operation and Development (OECD). (2019). OECD AI Principles. <https://oecd.ai/en/ai-principles>
9. The Great Hack [Documentary]. (2019). Netflix.
10. UNESCO. (2021). Recommendation on the Ethics of Artificial Intelligence. <https://www.unesco.org>
11. United Nations Office for Disarmament Affairs (UNIDIR). (n.d.). Research on AI in weapons systems and arms control. <https://unidir.org>
12. United States Department of Justice. (n.d.). Lawsuit against Google for monopolizing online advertising. <https://www.justice.gov>
13. United States Executive Order 14110. (2023). Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. Federal Register.
14. Varoufakis, Y. (2023). Technofeudalism: What killed capitalism. Bodley Head.
15. Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. PublicAffairs