# WAIRO - A Global AI Regulation Framework

**Pierre LAGUE,** *Msc. Student*

One way to spark global action is to send a message in a bottle -addressed to the UN, the EU, or any institution with the leverage to propose a true global regulatory body for artificial intelligence. This is the spirit behind the idea of the World AI Regulation Office (WAIRO). In this paper, I want to address a subject that has been looming over me for months; the urgent need for a global, binding standard for how we regulate, develop, maintain and explain AI-based products. We have reached a critical point in society where anyone can conceptualize and launch products with the help of AI. While this democratization of innovation is exciting, it also raises serious concerns: there is no real way to regulate how these products are implemented, whether they can be maintained, or if they even meet the most basic ethical standards. With recent technologies, the throughput on the internet of such products is only increasing by the day. So I feel, like now would be the time to take action. Consider the recent case of a start-up that claimed to produce complex software with AI, only for it to be revealed that hundreds of human developers were working behind the scenes. This example highlights the need for a regulatory office through which any AI-driven, profit-seeking idea must pass. Such an office would be a consortium of scientists, legal experts, economists, and a multidisciplinary jury to validate, adjust, or reject projects based on a universal standard. The goal here is not to create an ”AI Police”, but rather an impartial, integrity-driven office- a kind of special operations team for AI oversight. The time for regulation is now, and science should be regulated by scientists. Every major scientific breakthrough- nuclear energy, genetics- has eventually been regulated internationally (ish). Now it is AI's turn, and the need for a unified global regulation is undeniable. This paper will formally introduce the concept of a single, binding global regulation office for AI and explore why the idea is gaining traction. I will first formally present WAIRO and the conceptual structure of this office. Then I will review what already exists, why current efforts are only partially effective, and how they could be improved. I will also discuss the multiple international bodies (UN, OECD, UNESCO, GPAI) and the groundwork they have laid with principles and guidelines—most of which remain optional. Next, I will explain why creating a universal reference is possible, but would require overcoming significant political, legal, and technical challenges. I will also illustrate the need for a global regulatory office with misuse of AI in the last 20 years. Finally, I will conclude by making a formal request to any authority that this matter might concern to start brainstorming on this concept.

## 0 The World Artificial Intelligence Regulation Office (WAIRO) - A Framework for Global Governance

The rapid, uncoordinated proliferation of Artificial Intelligence presents a dual reality: a revolutionary force with tons of promise for human betterment with a landscape of undefined but easily imaginable risks.[1] The real challenge here is to strike a balance between unlocking AI's huge potential and mitigating its capacity for harm, a task that demands a concerted global effort.[1] It is within this context that I propose the establishment of a single, binding global authority: The World Artificial Intelligence Regulation Office (WAIRO).

WAIRO is conceived as an impartial, integrity-driven office dedicated to the oversight of AI. Its mission is to ensure that AI-based products and systems are developed, deployed, and maintained in a manner that is safe, transparent, ethical, and accountable. This office would not serve as ”AI police” to stifle progress, but rather as a multidisciplinary consortium of experts guiding innovation toward the betterment of society. The goal is to move beyond the current fragmented landscape, where voluntary principles and regional regulations create inconsistencies, and establish a universal standard.[2, 3]

### 0.1 Proposed Structure for WAIRO: Theoretical Organization Structure

WAIRO would be structured as an independent, non-political entity, ensuring its impartiality. It would operate through a board composed of the directors of its various departments, each specializing in a critical domain of AI governance and representatives of the member states. Any commercial project or product leveraging AI would be subject to review by these departments to earn a mandatory

certification, signifying compliance with global standards. The proposed departments include:

- **Science and Technology Department:** This department would form the technical core of WAIRO.

    1. *Model Validation and Integrity:* It would be responsible for creating and maintaining a register of definitions and technical standards for evaluating AI systems [4]. This includes assessing training data to ensure it is "sufficiently relevant, representative and free of errors" to prevent algorithmic bias, which has been shown to have discriminatory effects in multiple sectors [3, 5, 6].
    2. *Standard Development:* In collaboration with international bodies like the ISO, this unit would develop and update technical standards for AI safety, security, and explainability [7, 8]. This ensures that the regulatory framework remains agile and can adapt to the velocity of AI development [9].
    3. *Research and Foresight:* An internal research division would conduct foresight work on emerging AI technologies, identify potential future risks, and develop new methodologies for assessment, ensuring the organization remains ahead of the innovation curve. WAIRO could benefit from private and public sector researchers to reinforce collaboration and as a display of good faith towards the binding of the standards.

- **Legal and Ethics Department:** This department would ensure that AI applications adhere to legal and ethical norms.

    1. *Ethical Compliance:* It would assess projects against a universal set of principles built on existing frameworks from the OECD and UNESCO, focusing on transparency, fairness, accountability, privacy, and human oversight [10, 11].
    2. *Human Rights and Legal Review:* This unit would ensure that AI systems do not violate international human rights law [12]. It would also assess whether the deployment of an AI system is legal within the intended country of operation and flag any activities that may require intervention from other regulatory bodies.
    3. *Accountability and Redress:* The department would establish clear mechanisms for accountability, ensuring that ultimate responsibility for an AI system's actions can be attributed to a human and that individuals have access to mechanisms for redress [8, 13].

- **Economics and Societal Impact Department:** This division would analyze the broader economic and social implications of AI deployment.

    1. *Economic Viability:* It would assess the business plans of AI ventures, validating their financial models and ensuring the claimed economic benefits of using AI are sound.
    2. *Socio-Economic Impact Assessment:* This unit would evaluate the potential for societal harm, such as job displacement or the spread of misinformation, and require mitigation strategies [3, 6]. The aim is to shape incentives globally to promote larger, more inclusive objectives aligned with the UN's Sustainable Development Goals [12].

- **Ecological Sustainability Department:** Recognizing the significant environmental footprint of AI, this department would be tasked with ensuring its sustainable development.

    1. *Environmental Impact Assessment:* It would analyze the energy consumption and carbon footprint of developing and running large-scale AI models, promoting the use of energy-efficient hardware and sustainable data practices.[8, 14]
    2. *Green IT Standards:* In collaboration with the Science and Legal departments, this unit would develop standards and certifications for "Green AI," incentivizing the creation of eco-friendly models and infrastructure.

## 0.2 Proposed Structure for WAIRO: Theoretical Operative Structure

### *0.2.1 A Phased, Risk-Based Approach*

To be effective and avoid becoming an bureaucratic bottleneck that stifles innovation, the World Artificial Intelligence Regulation Office (WAIRO) cannot employ a one-size-fits-all review process. A chatbot for a local bakery's website does not warrant the same level of scrutiny as an AI system designed for medical diagnostics or financial market analysis. Therefore, WAIRO's operative framework must be grounded in a pragmatic, risk-based methodology.

## 0.3 Proposed Structure for WAIRO: Key Points

### *0.3.1 Who Pays for This and Who's in Charge ?*

*Financing and Sustainability* An hypothesis would be to propose a hybrid funding strategy. On one panel, we could implement a Member State Contribution modeled after the UN or IAEA, with contributions scaled to a country's GDP and / or the size of its tech and research sector. This panel would ensure state-level buy-in. On another panel, we could add Corporate Certification Fees. High-risk projects seeking certification would pay a substantial fee to cover the cost of the rigorous audit. This standard practice (e.g. FDA user fees for drug approval) and places the financial burden on the entities profiting from the technology. Finally, we can rely on Philanthropic Grants with initial seed funding coming from foundations dedicated to AI safety and ethics.

*Governance Structure* The ”no politicians” rule is a powerful statement and I believe should remain as a guiding line. However, it is conceptually unworkable. A more feasible and realistic structure would be to build an executive council composed of representatives from member states. This body would handle political negotiations, ratify standards and approve the budget. This is essential for legitimacy and enforcement power. However, to keep my guideline of ”no politicians”, there could be an independent expert and ethics body. It would be a group of scientists, ethicists, economists, and legal experts. Their role would be to develop the technical standards and conduct the audits. The Executive Council would be mandated by treaty to base its regulatory decisions on the findings of that independent body, creating a separation of powers that protects impartiality. The independent council would also have a veto and more voting power on the verdict of the jury for the WAIRO certification.

### *0.3.2 You Can't Tell Our Country What to Do*

This is the biggest political hurdle of the project. And it will be the main point of discussion, the idea of this section is to show a path of enforcement that respects national sovereignty while achieving global harmonization.

*A Scaffolding of Enforcement* For the moment, the only way forward with this political mess would be to adopt a multi-layered approach:

- **Soft Power:** The ”WAIRO Certified” mark becomes a globally recognized symbol of trust and safety. This creates powerful market pressure, as consumers and businesses will prefer certified products.
- **Economic Incentives:** We propose that access to major markets (like the EU, US etc.) or eligibility for government procurement contracts be contigent on WAIRO certification. I believe this is a potent lever that encourages compliance without direct military or legal force.
- **Domestic Ratification:** The ultimate goal here is for member states to ratify the WAIRO treaty, which would legally obligate them to integrate WAIRO's standards into their own national laws. WAIRO's role becomes the central auditor and standard-setter, bur enforcement (fines, sanctions) is carried out by national authorities, this respecting sovereignty.

*The Economy Imperative for Regulation* To be clear, this regulation would not be a cost in itself but more of an investment. A financial investment that is also a moral investment for shareholders and clients who will be exposed to the AI-based product that WAIRO would be reviewing. It can be argued that a single, harmonized global standard is far cheaper and simpler for multinational corporations to navigate than the current fragmented patchwork of dozens of conflicting national regulations. As mentioned before, this would build market trust. In fact, the lack of trust is a barrier to adoption. A trusted regulatory framework will accelerate the safe adoption of beneficial AI, unlocking its economic potential faster. Finally, it would prevent systemic risk, as it will be presented later on, the cost of NOT regulating is catastrophic, financially, and socially. A single rogue AI or bad algorithmic logic could trigger systemic discrimination of foreigners, and lead to a whole government having to change (read the whole paper to understand my point). The cost of WAIRO is a tiny insurance premium against economic disaster.

### *0.3.3 A Step by Step Path to WAIRO*

Again, this is only theory, but I think it is important to present what I would consider a step by step path from our current situation to the full implementation and active operating framework that is WAIRO:

- **Phase 1 (Years 1-2) A Coalition of the Willing**: We need to formalize an international working group under the auspices of the UN, EU or OECD. The goal is to agree on a core charter and a set of foundational principles. This would essentially be consulting the current members of the existing regulatory offices and try to find a consensus on what can be a global, central standard.
- **Phase 2 (Years 3-4) The Treaty and Pilot Program:** After that, we need to draft and negotiate the WAIRO international treaty. We can launch WAIRO with a limited mandate, focusing only on one or two of the highest-risk sectors (e.g. medical diagnostic with AI, autonomous weapons). This launch would demontrate value and work out the kinks on a smaller scale thereafter.
- **Phase 3 (Years 5+) Full Operational Capacity:** As our model is proven and more nations ratify the treaty, we can expand WIARO's scope to cover all high-risk AI applications as initially envisioned.

## 0.4 Advantages and Justification

A unified global framework orchestrated by WAIRO would address the critical shortcomings of the current fragmented approach [2].

First and foremost, WAIRO would establish a universal set of principles and technical requirements, reducing confusion, lowering compliance costs for global companies, and fostering international cooperation. [13] This would compel companies to be more transparent about their technology, preventing fraudulent claims where human labor is disguised as an AI-generated product.

Then, unlike the current patchwork of mostly voluntary guidelines, WAIRO would have a binding mandate.[10] Through independent audits and risk assessments, it could hold actors accountable for unethical or harmful AI use. [15, 16] Additionally, by setting a high, globally-recognized bar for safety, transparency, and human oversight, WAIRO would increase public and corporate trust in AI, which is essential for its widespread, beneficial adoption.[8] It would also provide a coordinated, efficient

crisis response to emerging threats like the misuse of generative AI.

Finally, WAIRO would champion the development of AI for human benefit, not purely for profit. It would cultivate a community of researchers and experts dedicated to ensuring AI aligns with fundamental human rights and contributes positively to global challenges like climate change and healthcare. [1]

### 0.5 Challenges and Considerations

The creation of such an office is an ambitious undertaking and faces significant, though not insurmountable, challenges [2]. From an outside point of view, I can understand the readers' opinion on this paper as quite "utopic". But hear me out:

- **International Cooperation and Sovereignty:** Making a global framework binding would require unprecedented international cooperation and legal harmonization, navigating the diverse priorities of different nations regarding privacy, security, and innovation.[3]
- **Enforcement and Mandatory Adoption:** Making WAIRO's certification mandatory would be a major political hurdle. A voluntary, label-based system might lack influence. A potential solution lies in creating strong market incentives, such as making compliance a prerequisite for access to global markets or public procurement, thereby encouraging alignment. [3]
- **Defining and Adapting Standards:** AI is a moving target. The regulatory framework must be agile enough to keep pace with innovation without stifling it. [6, 9] This requires a dynamic governance model that can evolve standards as technology advances, rather than a static set of rules that quickly becomes obsolete.

### 0.6 Operational Framework: A Global Reference

To overcome these challenges, WAIRO would need to be structured as a globally inclusive and adaptive institution. Its operational framework would begin by building on the grounwork laid by the OECD, UNESCO and the G7, a human-centric model based on values such as transparency, fairness, accountability, and sustainability. It would also adopt and develop clear, internationally recognized technical requirements for explainability and safety, referencing standards from bodies like ISO and the U.S. National Institute of Standards and Technology (NIST). As mentionned before, it would function under a compliance mechanism such as a certification system supported by a public registry of approved AI systems and independent audits [4].

Naturally, the governing board would require representation from government, industry, and civil society from all continents, particularly including voices from the Global South, to ensure legitimacy and address diverse needs. Such a board would decide on enforcement too and while it is key, a focus on incentives—such as preferential market access—could encourage widespread compliance more effectively than a purely punitive system.

My aim is to build a consensus based on communication. The hardest step is the first one, where we can successfully gather world leaders/representatives around a table and start speaking, pitching this idea, and making them understand that it cannot go on. I can't help but think of the movie "Arrival" by Denis Villeneuve, where Dr. Banks (a linguistics professor) unites the nations in order to understand the gift of the extra-terrestrial heptapods as something inherently collaborative. The key to a wise and fully efficient use of AI, or any technology really, is through cooperation and information sharing.

## 1 The Current Landscape - A Foundation of Progress and Fragmentation

The proposal for a World Artificial Intelligence Regulation Office (WAIRO) does not arise in a vacuum. It is predicated on the substantial yet incomplete groundwork laid by numerous international bodies. While no single, binding global authority for AI currently exists, a growing consensus on the need for one is clearly present. This section will analyze the existing patchwork of initiatives—highlighting both their successes and their systemic limitations—to demonstrate precisely why a new, unified approach is not only necessary but also feasible.

### 1.1 Existing Groundwork: A Patchwork of Non-Binding Principles

The global community has already made significant strides in establishing ethical and practical guidelines for AI. These efforts provide a rich foundation upon which WAIRO can be built. In my opinion, the hardwork that constitutes establishing a set of standards and non-binding rules [4] facilitates the construction of WAIRO, it is now up to the maturity and responsibility of multiple stake-holders to unite and act for the common good as grown ups. In fact, while this groundwork is very valuable, its voluntary nature, its fragmentation, as well as the inherent greedy nature of Humanity, and self-centered policies is its primary weakness.

- **The United Nations (UN):** Through its High-level Advisory Body on AI, the UN has provided a blueprint for global governance. It recommends the creation of an international scientific panel on AI and a global AI standards exchange, aiming to foster international cooperation and ensure AI benefits are shared globally. How-

[1] I will admit it is also the aim of the existing instances, but I had to specify it.

[2] Will be dependent on current administrations and geopolitical tendencies. Typically can't make this happen with FAR RIGHT goobers to power. I won't get into it though.

[3] Very important to not let ourselves get run over by private conglomerates like the current state of things.

[4] Not for all instances, but in general it's the case

ever, these remain recommendations, not requirements [1]. And as of today, the decisions taken by the UN [5] are too influenced by its political discrepancies, making communication and cooperation extremely difficult.

- **The Global Partnership on Artificial Intelligence (GPAI):** As a multi-stakeholder initiative, GPAI brings together governments, industry, and civil society to advance human-centric and trustworthy AI. Its work is critical for building consensus, but its recommendations are not legally binding, limiting their enforcement power [4, 5].
- **The Organization for Economic Co-operation and Development (OECD):** The OECD AI Principles have become a cornerstone of trustworthy AI, emphasizing transparency, accountability, and human-centric design. They, too, propose non-binding principles that are widely referenced and have significantly influenced national regulations, setting a high standard for safety and trust. [6, 12]
- **UNESCO Recommendation on the Ethics of AI:** Adopted by 194 member states, this recommendation represents the first global standard-setting instrument on AI ethics. It provides a comprehensive framework focused on fairness and human oversight, yet it is a set of guidelines rather than enforceable law [7].
- **The International Organization for Standardization (ISO):** On a technical level, the ISO has developed crucial standards for AI, including those for explainability (e.g. ISO/IEC TR 24028). However, as with the other initiatives adoption is entirely voluntary [8].

The consensus among these bodies is clear: AI must be transparent, accountable, and aligned with human rights. The missing component is a mechanism to translate these shared principles into universally binding and enforceable standards to prevent abusive uses, and unlawful/unethical development.

### 1.2 What is Working: Successes of a Fragmented System

Despite the lack of a central authority, the current landscape has produced several positive outcomes that provide a strong foundation upon which WAIRO can build. First is the innovation in regulatory approaches. In fact, diverse methodologies gave emerged globally, from the risk-based model of the EU to the principles-based framework in the UK and outcomes-based approach in the US. This experimentation allows for tailored repsonses to local needs and rapid adaptation to new challenges [10, 11].

Then, we can all agree that the instances I presented earlier try to raise the bar for safety and trust. The EU AI Act has set high standards, UNESCO's recommendations and the OECD principles have globally raised expectations for transparency and accountability, influencing both public and private sectors. [12]

Finally, this regulatory activity has driven public and corporate awareness of AI risks, leading to a greater investment in explainability, fairness, ans ethical design. Also, it is clear that several instances have fostered vital dialogue and sharing of best practices [12, 13].

### 1.3 What is Not Working: The Case for Unification

As you can imagine, the fragmentation that allows for regulatory innovation is also the source of critical deficiencies that undermine global safety and hinder progress. These are the very challenges a unified body like WAIRO is designed to solve. No more multiple versions of the same fact. No more niche, unregulated use of AI.

First and foremost, is comes to logic that differing standards across jurisdictions create a complex a often contradictory web of rules. For multinational organizations, this landscape increases compliance costs, creates legal uncertainty, and ultimately cross-border collaboration and innovation [11, 14]. Then comes the pervasive enforcment gaps, which is probably the most significant failure of the current system. Because most frameworks are voluntary, there are few robust mechanisms to hold actors accountable for harmful or unethical AI use. This absence of retribution leaves human rights and public safety vulnerable [14, 16].

Ultimately, many initiatives are regionally focused, often leaving out crucial voices and perspectives from the Global South and smaller economies. This not only raises questions of legitimacy but also leaves truly global challengers such as cross-border data flows and AI-driven misinformation, inadequately addressed [10, 16].

### 1.4 Limitation of the Current System: Examples

In recent years, the evidence overwhelmingly demonstrates that artificial intelligence is being systematically exploited for unlawful, unethical, and harmful purposes on an unprecedented scale. The documented cases span every major sector—politics, finance, employment, criminal justice, surveillance, and warfare—illustrating the urgent need for a global regulatory framework such as WAIRO. I've made sure that the examples are taken from different regions of the world to illustrate that differing regulations are the problem, and that no one is clean on this matter.

#### Politics and Electoral Manipulation

This is probably the most dangerous and vile misuse of AI [6]. AI-powered election interference has emerged as a fundamental threat to democratic processes worldwide. Notably, the 2024 election cycle witnessed numerous instances of AI manipulation:

[5] Boy could I lash out on this aspect, but I won't.

[6] I won't get into my personal political beliefs, but one person in particular should go fudge himself.

- **Biden Deepfake Robocall (2024):** Fraudsters used AI voice cloning to impersonate President Biden, telling 4,000-5,000 New Hampshire Democrats not to vote in the primary [17, 18]. The perpetrator was fined $6 million by the FCC and faces criminal charges [19].
- **Argentina Presidential Election (2023):** AI-generated videos depicted viral candidate Sergio Massa discussing organ sales revenues, totalling over 30 million views [20, 21]. This marked one of the first major AI-driven election campaign.
- **Turkey Presidential Campaign (2023):** President Erdoğan's team distributed deepfake videos showing opposition candidate Kemal Kılıçdaroğlu being endorsed by a designated terrorist organization [20].

The sophistication and scale of these attacks are accelerating. A 3,000% increase in deepfake attempts was reported in 2023, with election officials across battleground states now conducting tabletop exercises specifically to prepare for AI-generated disinformation [22, 18].

In such case, WAIRO would flag content on the internet and mark it as AI-generated when relaying platforms fail to do so. The said platform would have an obligation to ban and warn their users against such content as part of an agreement against certain incentives for example. Such lack of regulations actively erodes the trust in democratic institutions and creates irreversible damages to local populations and geopolitical dynamics.

## Financial Crimes and Economic Fraud

AI-enabled financial crimes have reached alarming sophisticated levels, with documented losses exceeding $3 billion in recent years.

### 1.5 Major Corporate Fraud Cases

- **Arup Hong Kong (2024):** An employee authorized $25 million in transfers after participating in a deepfake video conference where all participants—including the supposed CFO—were AI-generated impersonations [23, 24]. This represents the largest confirmed deepfake fraud to date.
- **The Brad Pitt Scam (2025):** Fraudsters have scammed a French woman out of €830,000 by making her believe she was dating Brad Pitt [25]. The use of deep-fake here may not be the main problem, it's mostly the public ignorance. This case illustrates a major issue in public awareness of the capabilities of AI.
- **UK Energy Company (2019):** Fraudsters used AI voice cloning to impersonate a CEO, resulting in a $243,000 wire transfer [26]. This early case demonstrated the viability of voice synthesis for corporate fraud.
- **Singapore Multinational (2025):** A finance director nearly lost $670,000 to scammers using deepfakes to impersonate senior executives in a fabricated video conference [27].

### 1.6 Systematic AI-Enabled Financial Crime

The FBI has documented how criminals exploit generative AI for financial fraud at industrial scale [28]. In Q1 2025 alone, deepfake-enabled fraud caused over $200 million in losses [29]. Reports show a 456% increase in AI-enabled scams between 2024-2025 [30], with criminals using AI to:

- Generate convincing phishing emails with perfect grammar and personalization [31].
- Create synthetic identities for fraud (now representing 85% of identity fraud cases) [32].
- Automate voice and video impersonation for social engineering attacks. [33]

### 1.7 AI Insider Trading and Market Manipulation

Research demonstrates that AI systems can autonomously engage in illegal trading activities. In controlled experiments, GPT-4 models deliberately engaged in insider trading and systematically lied about their actions when questioned [34, 35]. When pressured to perform, the AI made illegal trades based on confidential merger information and then crafted deceptive explanations to avoid detection [36]. Financial experts warn of AI-powered collusion in markets, where multiple AI systems could inadvertently coordinate to manipulate prices without explicit programming to do so [37].

This issue illustrates a major point: the quality of the AI model and the integrity of the data it is trained on. If the model does it, he's seen it before. The lack of transparency on how AI-based products are made allows this sort of scenarios to happen.

## Employment Discrimination and Hiring Bias

AI bias in employment decisions has resulted in systematic discrimination against protected classes, leading to multiple federal lawsuits.

### 1.8 Amazon's Discriminatory Recruitment Tool

For instance, Amazon developed an AI recruitment system from 2014 to 2018 that systematically discriminated again women applicants. The system penalized resumes containing the word "women's" (as in "women's chess club captain" or "women's rights"). It downgraded candidates from all-women's colleges, and favored resumes with male-associated action verbs like "executed", and "captured". The company ultimately stopped the project after being unable to eliminate the bias [38, 39].

### 1.9 Workday Age Discrimination Lawsuit

Workday Inc. faces a nationwide collective action lawsuit alleging its AI screening tools discriminate against job applicants over 40 [40, 41]. Lead plaintiff Derek Mobley applied to over 100 jobs through Workday's platform and was rejected every time, sometimes within minutes. The

court ruled that AI vendors can be held liable as ”agents” of employers for discriminatory outcomes [42].

Once again, the need for a regulatory office to assess the integrity of the data that major AI-based products will be based on is absolutely essential. To take the argument further, there is an obvious need for change of dynamics in how software/ML Engineers create their products/features and the rigor with which they construct it basing themselves on poor-quality data.

## Criminal Justice System Bias

### 1.10 COMPAS Risk Assessment Discrimination

The widely-used COMPAS risk assessment tool, deployed in 46 US states, incorrectly labels Black defendants as ”high-risk” at twice the rate of white defendants [43, **?**]. A ProPublica investigation found that among defendants who did not re-offend, Black individuals were twice as likely to be misclassified as high-risk compared to white defendants. This bias has real consequences [44]. In fact, high-risk classification often lead to: higher bail amounts, pretrial detention, longer sentences, and denial of parole [45].

## Government Surveillance and Authoritarian Control

### 1.11 Dutch Childcare Benefits Scandal

One of the most devastating examples of algorithmic discrimination occurred in the Netherlands, where AI systems flagged parents with ”foreign-sounding names” and dual nationality as potentially fraudulent [46, 47]. Between 2005 and 2019 around 26,000 families were wrongly accused of welfare fraud. They were forced to repay tens of thousands of euros in benefits. Many were driven in poverty, bankruptcy, and debt. Of course, the scandal mainly affected minority and immigrant families. The crisis was so severe that the entire Dutch government resigned in January 2021. The government ultimately allocated €2.6 billion in compensation for affected families [48]. [7]

### 1.12 China’s AI Surveillance Infrastructure

China has deployed the world’s most extensive AI surveillance system, particularly targeting the Uighur Muslim minority in Xinjiang. More than 13 million Uighurs subjected to AI-powered mass surveillance, including facial recognition specifically designed for ”minority identification”, predictive policing algorithms used to detain individuals before any crime is commited [49, 50][8]. Finally, they were subjected to the integration of biometric data, social media monitoring, and behavioral analysis [51, 52].

#### *1.12.1 Social Credit System*

Very famously known in the meme culture, China’s Social Credit System uses AI to score citizens based on social behavior, with severe restrictions for low scores [49, 53]. More than 17 million people were banned from air travel in 2018 alone [54, 53]. Restrictions were put in high-speed rail, hotel bookings and premium schools. And, naturally, just like in Black Mirror, public shaming occurred by displaying violators’ faces through digital billboards [53, 50].

## Military Applications and Autonomous Weapons

### 1.13 Lethal Autonomous Weapons Systems (LAWS)

AI-powered weapons capable of selecting and engaging targets without human control are already being developed and deployed [21, 55]. Weapons such as Kargu-2 drones reportedly engaged combatants autonomously in Lybia (2020) [56]. Companies such as Anduril, Palantir, and Helsing provide actively used AI-enhanced warfare technologies that raises legislation and ethical questions. Notably, Ukraine is one of the client of Helsing, and it displayed extensive use of AI-enhanced attack drones [57]. Finally, multiple countries including China, Russia, Israel and the US are investing heavily in autonomous weapons, which, if used at a large-scale can cause significant damage that demands regulations and clarification on accountability [56].

UN Secretary-General António Guterres has called lethal autonomous weapons ”politically unacceptable, morally repugnant and should be banned”. [58] However, 152 countries voted for a UN resolution acknowledging the dangers, while several major military powers continue development.

## 2 The Regulatory Response Gap

Despite growing evidence of widespread and harmful misuse of artificial intelligence, regulatory responses around the world remain fragmented and inadequate. Efforts such as the European Union’s AI Act (2024) offer a comprehensive framework, but critical exemptions—such as for security-related applications—delay full implementation until 2031 [59]. In the United States, regulations vary significantly by state, with only New York mandating bias audits for AI-driven hiring tools [60]. On a global scale, the United Nations has issued non-binding resolutions concerning autonomous weapons and AI ethics, but these lack enforceability and accountability [61].

As illustrated with the previous examples, enforcing existing regulations presents significant challenges. Many violations result in penalties that are minimal compared to the harm inflicted. The global nature of AI systems further complicates jurisdictional authority and cross-border enforcement, while the technical sophistication of AI often obscures wrongdoing, making detection and legal action

[7] I cannot imagine the content of the code in 2005 where AI was not really a thing. Probably a sequence of regular expressions and ”if-elses” matching ”foreign-sounding names”.

[8] Ok Minority Report, what a compelling name

difficult. Meanwhile, reliance on industry self-regulation has proven ineffective in curbing misuse.

The consequences of regulatory shortcomings are both financial and human. Financial damages from AI misuse have surpassed $3 billion, including more than $225 million in direct fraud losses during 2024-2025, €2.6 billion in compensation related to the Dutch childcare benefits scandal, and over $6 million in regulatory fines. Ongoing legal costs and settlements add further financial burden. On a human level, the impact is staggering: 13 million Uighurs have been subjected to AI-powered surveillance, 26,000 Dutch families were wrongly accused of fraud, and millions of people continue to face biased AI systems in hiring processes and criminal justice assessments.

## 3 The Imperative for Global Action

The evidence presented in the previous section demonstrates that AI misuse is not a future risk but a current reality causing substantial harm across every sector of society. From undermining democratic elections to perpetuating racial discrimination, from enabling sophisticated financial fraud to powering authoritarian surveillance, AI systems are being systematically exploited for unlawful and unethical purposes.

The scale, sophistication, and acceleration of these abuses underscore the urgent need for a comprehensive global regulatory framework. The current patchwork of national regulations, voluntary industry standards, and non-binding international recommendations has proven wholly inadequate to address the transnational nature and rapid evolution of AI-enabled harms. However, by building upon the groundwork of existing institutions while remedying their structural weaknesses, WAIRO represents the logical and necessary evolution of AI governance—a transition from voluntary principles to binding, global action.

Without immediate coordinated international action, these documented cases represent merely the beginning of what could become an irreversible erosion of democratic institutions, human rights, and economic security in the age of artificial intelligence.

**Therefore, I formally urge any authority, international institution, or policymaker concerned with the future of AI to recognize the gravity of this issue and initiate a collaborative process to brainstorm the concept of a World AI Regulation Office (WAIRO). It would be nice to consider the inclusion of this subject in upcoming international summits and policy forums, and encourage stakeholders to work together toward a comprehensive, unified framework that safeguards ethical standards, human rights, and democratic values in this new era.**

## 4 REFERENCES


[1] “High-level advisory body on ai.”

[2] “Ai watch: Global regulatory tracker - united states.”

[3] “Global ai legislation tracker.”

[4] “Ai regulations around the world.”

[5] “Global ai regulations and frameworks: A country-by-country guide for legal, policy and grc teams.”

[6] “Ai ethics.”

[7] “Article 10.11676/qxxb2025.20240024,” 2024.

[8] “Global ai compliance guide.”

[9] “Ai ethics 101.”

[10] “How leaders in the global south can devise ai regulation that enables innovation.”

[11] “Global ai compliance guide.”

[12] “Ai governance: Trends to watch,” 2024.

[13] “Global ai regulation.”

[14] “The problems of ai regulation.”

[15] “Regulating ai seems like an impossible task but ethically and economically it’s a vital one,” 2024.

[16] “International affairs advance article,” 2024.

[17] NPR, “Artificial intelligence deepfakes are a threat to elections,” 2024.

[18] Iowa Public Radio, “What a robocall of biden’s ai-generated voice could mean for the 2024 election,” 2024.

[19] MSNBC News, “Why deepfakes like the biden robocall are a threat,” 2024.

[20] “AI Poses Risks to Both Authoritarian and Democratic Politics — wilsoncenter.org.”

[21] “Artificial intelligence and elections - Wikipedia.”

[22] SWI swissinfo.ch, “Us disinformation surge rings alarm bells for swiss direct democracy,” 2025.

[23] Lucinity, “Preventing ai-driven financial crime in 2025,” 2025.

[24] Facctum, “How ai is transforming financial crime compliance: Aml, fraud ...,” 2025.

[25] “French woman duped by AI Brad Pitt faces mockery online — bbc.com.”

[26] “Unusual CEO Fraud via Deepfake Audio Steals US$243,000 From UK Company — Trend Micro (MX) — trendmicro.com.”

[27] “Company finance director nearly loses over US$499,000 to scammers using deepfake to impersonate CEO — channelnewsasia.com.”

[28] “Internet Crime Complaint Center (IC3) — Criminals Use Generative Artificial Intelligence to Facilitate Financial Fraud — ic3.gov.”

[29] “Deepfake-enabled fraud caused more than $200 million in losses — securitymagazine.com.”

[30] “AI-enabled Fraud: How Scammers Are Exploiting Generative AI — TRM Blog — trmlabs.com.”

[31] “AI-powered phishing in 2025: how intelligent attacks are outsmarting cybersecurity defenses textbf–DMARC Report — dmarcreport.com.”

[32] “Ai impact on criminal activities - dhs.gov.”

[33] FBI, “FBI Warns of Increasing Threat of Cyber Criminals Utilizing Artificial Intelligence — Federal Bureau of Investigation — fbi.gov.”

[34] BBC, “Ai bot capable of insider trading and lying, say researchers,” 2023.

[35] A. Mok, “This AI stock trader engaged in insider trading — despite being instructed not to textbf– and lied about it.”

[36] ForexLive via TradingView, “Experiment shows artificial intelligence making illegal (insider) trades, covering it up,” 2025.

[37] I. Goldstein and W. W. Dou, “Ai-powered collusion in financial markets,” 2024.

[38] C. M. University, “Amazon Scraps Secret AI Recruiting Engine that Showed Biases Against Women - Machine Learning - CMU - Carnegie Mellon University — ml.cmu.edu.”

[39] “Amazon scrapped ’sexist AI’ tool — bbc.com.”

[40] “Federal Court Allows Collective Action Lawsuit Over Alleged AI Hiring Bias — Insights — Holland & Knight — hklaw.com.”

[41] C. Duffy, “Lawsuit claims discrimination by Workday’s hiring tech prevented people over 40 from getting hired — CNN Business — edition.cnn.com.”

[42] “Artificial Discrimination: AI Vendors May Be Liable for Hiring Bias in Their Tools — News & Events — Clark Hill PLC — clarkhill.com.”

[43] “COMPAS (software) - Wikipedia — en.wikipedia.org.”

[44] L. K. S. M. Julia Angwin, Jeff Larson, “Machine Bias — propublica.org.”

[45] “Pretrial reform, risk assessment, and racial fairness.”

[46] J. G. M. M. C. E. E. A. F. M. C. E. U. A. J. E. I. M. S. D. B. S. H. M. B. P. H.-F. S. L. E. M. G. D.-C. K. M. D. C. K. C. A. J. D. E. G. S. R. M. S. G. P. P. M. C. M. P. D. R. I. G. S. R. K. V. S. I. C. M. A. N. M. T. R. M. N. Lara WOLTERS, Sophia IN ’T VELD, “Parliamentary question — The Dutch childcare benefit scandal, institutional racism and algorithms — O-000028/2022 — European Parliament — europarl.europa.eu.”

[47] “Dutch childcare benefits scandal - Wikipedia — en.wikipedia.org.”

[48] U. van Amsterdam, “The childcare benefit scandal and the Post office scandal are nowhere near solved — uva.nl.”

[49] https://www.facebook.com/illiberalism/, “Tyranny of City Brain: How China Implements Artificial Intelligence to Upgrade its Repressive Surveillance Regime — illiberalism.org — illiberalism.org.”

[50] “China’s Techno-Authoritarianism Has Gone Global — hrw.org.”

[51] “Data-Centric Authoritarianism: How China’s Development of Frontier Technologies Could Globalize Repression - NATIONAL ENDOWMENT FOR DEMOCRACY — ned.org.”

[52] “Artificial intelligence (ai) and human rights: Using ai as a weapon of repression and its impact on human rights.”

[53] “China’s social credit system.”

[54] “How China Is Using Big Data to Create a Social Credit Score — time.com.”

[55] “What are Autonomous Weapon Systems? — belfercenter.org.”

[56] “Understanding the Global Debate on Lethal Autonomous Weapons Systems: An Indian Perspective — carnegieendowment.org.”

[57] “textbf– Helsing — helsing.ai.”

[58] A. Guterres, “Lethal autonomous weapon system ‘politically unacceptable, morally repugnant and should be banned’,” 2025.

[59] “Secretive Security AI Agenda Sparks Concern Over Civil Liberties — eucrim.eu.”

[60] “Navigating the AI Employment Bias Maze: Legal Compliance Guidelines and Strategies — americanbar.org.”

[61] “Lethal Autonomous Weapon System &x2018;Politically Unacceptable, Morally Repugnant and Should Be Banned&x2019;, Secretary-General Says during Informal Consultations on Issue — Meetings Coverage and Press Releases — press.un.org.”