The Imperative of Proactive AI Regulation: Balancing Innovation and Security in the Age of Superintelligence

November 30, 2022, the day when OpenAI released ChatGPT, is a significant moment that "overthrew" our current understanding of technology-enhanced society. In less than three years, AI's rapid evolution has already begun to influence industries, governance, and public policy. However, while the AI industry brought substantial benefits to the prospect of technological development by rapidly learning patterns and analyzing data, its applications also start to lurk behind potential severe social crises.

Historically, emerging technologies have brought significant improvement in social development as well as severe crises. From the emerging mechanization in the 19th century to the development of nuclear weapons during World War II and the sweeping addiction to social media, all have proven and foreshadowed the potential revolution and subsequent technological crisis brought by Artificial Intelligence. Regulations are therefore needed to navigate the modern challenges of AI.

By feeding large amounts of data into the machine-learning model, the AI engineers have already accelerated the update of their program to ensure a higher level of self-progress. It should not be denied that AI is approaching human-level intelligence. In the wake of such rapid technological transformation, issues of job displacement, cybersecurity, over-reliance, industry opacity, and even warfare could be redefined by the application of AI.
Considering AI's self-improving potential, which distinguishes it from previous technologies, early, preemptive regulation is preferable to ex post facto regulation. Even Elon Musk claimed that "AI is a rare case where I think we need to be proactive in regulation than be reactive". The quote is reasonable in its foresight: regulation should be proactive to address the consequences AI might bring in the future, particularly in military applications and cybersecurity. Given AI's self-improving nature and the difficulties of enacting traditional regulations, soft law, or quasi-legal instruments could be a viable solution to compromise between safety and innovation.

To begin, the application of AI to military systems has altered modern warfare at the fundamental level of decision-making. Unlike traditional warfare, where human decision-making serves as a buffer against violence, current AI-operated systems are still highly unpredictable. In AI-enhanced applications —— such as autonomous drones, robotic soldiers, and missile guidance systems, which all have significantly increased the precision and lethality of military operations —— the Lethal Autonomous Weapons System (LAWS) operates. By definition, autonomous weapons systems require "autonomy" to perform their functions, and Artificial intelligence is not a prerequisite for the functioning of autonomous weapons systems, but, when incorporated into them, AI could further enable such systems by making evaluations without

required human participation. (*Lethal Autonomous Weapon Systems (LAWS) – UNODA*, n.d.) In fact, the development of LAWS has raised debates worldwide. Critics warn that the possibility of an AI-led arms race will echo the twentieth century's nuclear arms race, condemning humanity to face an even greater existential crisis than before.

Aided Target Recognition (AiTR), for instance, assists in decision-making, target recognition, and casualty care in the field. The most challenging task is identifying the target's intent. Whereas, in the past, the detection of a targeted weapon or aggressive behavior may have been sufficient, today the soldier must also determine the intent of the human behind the behavior to avoid unnecessary attacks. (Ratches, 2011) Therefore, considering the possibility of inadvertent casualties to civilians, the US Department of Defense enacted ethical principles for AI in 2020, which include "Responsible", "Equitable", "Traceable", "Reliable", and "Governance" to establish fundamental ethical rules for AI's military application. (*DOD Adopts Ethical Principles for Artificial Intelligence*, n.d.) In the case of AiTR, proactive policies are essential due to the severe consequences of military decision-making and to prevent the intolerable war disaster caused by autonomous weapons.

In addition, the US military's growing reliance on AI in cybersecurity highlights the urgent need for preemptive regulation. On one hand, AI has increased threat detection, helping the US government and other organizations to quickly identify and resolve malicious activities. However, non-governmental criminals are adapting the same technology, posing unprecedented risks. AI-powered hacking tools can autonomously identify computer system vulnerabilities, generate convincing phishing scams, and adapt in real-time to bypass security measures. Large Language Models, for example, have been used for phishing attacks. Advanced LLMs are capable of generating human-like language to create personalized spear-phishing messages for as little as a few cents. (Hazell, 2023) By using the open-source LLM and manipulating human emotions of fear, greed, and anxiety, malicious phishing emails spread at a low cost. According to a new report by cybersecurity firm SlashNext, since the fourth quarter of 2022, there's been a 1,265% increase in malicious phishing emails and a 967% rise in credential phishing. (Violino, 2023)

It should be surprising to observe that AI models themselves become the targets of criminals. Data poisoning, in which attackers manipulate data inputs to the machine learning model, can corrupt AI systems and lead to widespread misinformation. Some may even lead to inestimable costs in cybersecurity or military decisions. Hence, the corrupt AI system will be a unique threat to social stability. Without proactive regulation, the cybersecurity landscape could become increasingly volatile, with AI accelerating nefarious actors 'defensive and offensive capabilities.

Because of the extensive risks posed by AI, proactive regulation is essential to resolve the problem of "pace regulation", or the challenge of regulatory policies attempting to keep pace with emerging technologies. However, AI's self-improving capabilities and the large share of the private sector already invested in AI technology make traditional legal frameworks difficult to apply. In this context, soft law could be used as a flexible, practical approach. Compared to complex law, which consists of absolute legislation and strict enforcement, soft law includes

voluntary guidelines, principles, and codes of conduct established by governments and other organizations. This adaptable framework is particularly suited for AI regulation because it allows for rapid updates in regulatory oversight in response to technological advancements, preventing policies from becoming outdated. Furthermore, soft law promotes international cooperation by encouraging shared ethical standards without falling into the complex negotiations required for legally binding treaties.

While overly rigid legal frameworks could stifle AI development, soft law strikes a balance. It also offers flexibility, allowing regulators to respond to emerging risks without impeding technological progress. The OECD AI Principles, for example, adopted in 2019, set a standard to promote trustworthy AI practices. It provides five major value-based principles: inclusive growth, human rights, transparency, robustness, and accountability. (AI Principles, n.d.) Contrary to traditional regulations and laws, the OECD AI Principles are not legally binding; instead, they serve as a global benchmark adopted by governments and companies. The advantages of soft laws such as OECD AI principles include: the non-binding benchmark allows immediate updates, which helps following rapid technological improvement. Also, the soft-law approach effectively encourages innovation by avoiding over-regulation. Nations and governments voluntarily accept the law for both cybersecurity and reputation.

On top of these benefits, investor pressure and credit ratings can also play roles in promoting these soft-law principles. For example, Goldman Sachs and JPMorgan evaluate AI risk using the OESD AI principles. The overall legal and theoretical structure of soft law thus provides adaptable frameworks that promote safety, transparency, and accountability while fostering innovation. Also, soft law encourages corporate responsibility by motivating companies to adopt ethical practices, such as algorithmic fairness and data privacy protections. Therefore, the soft laws currently could serve as the optimal choice for implementing proactive regulation.

The potential risks of Artificial Intelligence, from military applications to cybersecurity threats, demonstrate the necessity of proactive regulation. Previously, cybersecurity issues such as the Cambridge Analytica Scandal already demonstrated how AI and data could be used as weapons to manipulate public opinion and how important proactive standards or regulations are to prevent misinformation and cybersecurity problems. Similarly, the development and the arms race of nuclear weapons during WWII also serve as a historical precedent for the threat of technological revolution. While the nuclear age required international treaties to prevent catastrophe, it is even more urgent for Artificial Intelligence to adopt regulation, especially proactive rules, for its ability to self-improve and the widespread AI production in the private sector. Soft laws, such as the OECD AI Principles, provide a balanced approach to set a standard for AI without stifling innovation. As AI continues to permeate global commerce and national security, flexible governance will be essential to prevent the recurrence of technological crises while maximizing the benefits and accountability of AI.

References:

AI principles. (n.d.). OECD. Retrieved April 2, 2025, from https://www.oecd.org/en/topics/ai-principles.html

DOD Adopts Ethical Principles for Artificial Intelligence. (n.d.). U.S. Department of Defense. Retrieved April 2, 2025, from https://www.defense.gov/News/Releases/release/article/2091996/dod-adopts-ethical-principles-for-artificial-intelligence/https%3A%2F%2Fwww.defense.gov%2FNews%2FReleases%2FRelease%2FArticle%2F2091996%2Fdod-adopts-ethical-principles-for-artificial-intelligence%2F

Hazell, J. (2023). Spear Phishing With Large Language Models (No. arXiv:2305.06972). arXiv. https://doi.org/10.48550/arXiv.2305.06972

Lethal Autonomous Weapon Systems (LAWS) – UNODA. (n.d.). Retrieved April 2, 2025, from https://disarmament.unoda.org/the-convention-on-certain-conventional-weapons/background-on-laws-in-the-ccw/

Ratches, J. A. (2011). Review of current aided/automatic target acquisition technology for military target acquisition tasks. Optical Engineering, 50(7), 072001. https://doi.org/10.1117/1.3601879

Violino, B. (2023, November 28). AI tools such as ChatGPT are generating a mammoth increase in malicious phishing emails. CNBC. https://www.cnbc.com/2023/11/28/ai-like-chatgpt-is-creating-huge-increase-in-malicious-phishing-email.html