EAC-PM Working Paper Series
EAC-PM/WP/26/2023

A Complex Adaptive System Framework to Regulate AI

December 2023

Sanjeev Sanyal, Pranav Sharma and Chirag Dudani

Contents

Sanjeev Sanyal is Member, Economic Advisory Council to the Prime Minister (EAC-PM), Pranav Sharma is Consultant (Science Diplomacy), Indian National Science Academy, and Chirag Dudani is an Assistant Consultant, Economic Advisory Council to the Prime Minister (EAC-PM). The contents of the paper including facts and opinions expressed are sole responsibility of the authors. EAC-PM or the Government of India does not endorse the accuracy of the facts, figures or opinions expressed therein.

A Complex Adaptive System Framework to Regulate AI

*****

Abstract

Artificial Intelligence (AI) has been making significant strides, with advancements in computer vision, language processing, robotics, and more. This rapid development, however, has raised substantial concerns about the risks associated with uncontrolled AI propagation. The increasing capabilities of AI systems bring potential dangers, such as the prospect of "runaway AI" where systems could recursively self-improve beyond human control and misalign with human welfare. This has led to a growing demand for regulatory frameworks that ensure AI development and deployment are conducted safely, ethically, and transparently, safeguarding against risks like systemic breakdowns, loss of privacy, national security threats, and compromised critical infrastructure.

The current regulatory approaches, which typically rely on ex-ante impact analysis and risk assessment, face challenges in effectively governing AI. Traditional methods fall short due to AI's non-linear, unpredictable nature. AI systems are akin to Complex Adaptive Systems (CAS) where components interact and evolve in unpredictable ways. This complexity results in butterfly effects, where minor changes can lead to significant, unforeseen consequences. Consequently, standard regulatory methods that work for static, linear systems are inadequate for AI, necessitating a different approach that acknowledges and accommodates its complex and adaptive nature.

Regulating artificial intelligence technology and applications is becoming increasingly challenging. Current landscape of policy frameworks has a broad spectrum of control each with its own associated issues. This includes a 'hands-off' & self-regulation approach (USA) and pro-innovation laissez-faire approach (UK) which has significant associated dangers of regulation without teeth given the far-reaching and unpredictable consequences of AI. The risk-based approach (EU) inherently assumes that a classification of risks in a non-deterministic system is possible. While the complete state bureaucratic control (China) can fail and have cascading effects as demonstrated by likely lab-leak origin of Covid-19.

Aforementioned approaches do not fully address the CAS characteristics of AI, underlining the need for regulatory frameworks that set boundary conditions, enable real-time monitoring, and guide the evolution of AI systems. Our analysis suggests a complex adaptive system (CAS) approach to AI regulation.

To effectively regulate AI (algorithm, training data sets, models, and applications), a novel framework based on CAS thinking is proposed, consisting of five key principles:

- **Establishing Guardrails and Partitions**: Implement clear boundary conditions to limit undesirable AI behaviors. This includes creating "partition walls" between distinct and within deep learning AI systems to prevent systemic failures, similar to firebreaks in forests.
- **Mandating Manual 'Overrides' and 'Authorization Chokepoints'**: Critical infrastructure should include human control mechanisms at key stages to intervene when necessary, emphasizing the need for specialized skills and dedicated attention without limiting automation of systems. Manual overrides empower humans to intervene when AI systems behave erratically or create pathways to cross embedded partitions. Meanwhile, multi-factor authentication authorization protocols provide robust checks before executing high-risk actions, requiring consensus from multiple credentialed humans.

- **Ensuring Transparency and Explainability**: Open licensing of core algorithms for external audits, AI factsheets, and continuous monitoring of AI systems are crucial for accountability. There should be periodic mandatory audits of explainability.
- **Defining Clear Lines of AI Accountability**: Mandate standardized incident reporting protocols to document any system aberrations or failures. Establish predefined liability protocols to ensure that entities or individuals are held accountable for AI-related malfunctions or unintended outcomes. This proactive stance inserts an ex-ante "Skin in the Game," ensuring that system developers and operators remain deeply invested in and accountable for AI outcomes.
- **Creating a Specialist Regulator**: Traditional regulatory mechanisms often lag the rapid pace of AI evolution. A dedicated, agile, and expert regulatory body with a broad mandate and the ability to respond swiftly is pivotal to bridge this gap, ensuring that governance remains proactive and effective.

The regulation of financial markets offers valuable insights for AI regulation. Financial markets exemplify CAS with emergent behaviors arising from diverse interacting agents. Individual traders adapt strategies responding to price signals, news and each other's actions. This feeds back in nonlinear ways, resulting in volatile, unpredictable market dynamics. Yet proactive governance demonstrates feasible regulation approaches.

Establishing dedicated oversight through bodies like SEBI in India and SEC in the United States of America (USA) provides specialized expertise attuned to nuances. Requirements like financial statements, auditing and mandated incident reporting enhance transparency and traceability - akin to explainability standards and auditing for AI algorithms. Circuit breakers act as control chokepoints, halting trading when indices cross threshold triggers to prevent cascading crashes. Margin requirements, settlement periods and other boundaries constrain destabilizing behaviours. Mechanisms like bankruptcy and insolvency provide ultimate overrides when entities become non-viable. Fixing liability through personal accountability and 'skin in the game' holds individuals responsible for corporate actions. While not a perfect parallel, insights from governing chaotic markets can inform AI regulation. By applying similar strategies, such as partitions, transparency standards, control points, and accountability measures, AI's development can be steered responsibly, ensuring safety and ethical deployment akin to the orderly functioning of financial markets.

This proposed approach aims to navigate the complexities of AI regulation by learning from other CAS and adapting to the unique challenges posed by AI's evolving nature. Implementing prudent measures today can help ensure that AI develops in a way that is beneficial and safe for society.

A Complex Adaptive System Framework to Regulate AI

*****

1. Introduction

Artificial Intelligence (AI) has transitioned from academic and technological experimentation to widespread application in the past few years where advanced AI frameworks have been deployed exponentially with multilayered application modalities. Governments across the world have been continuously doing efforts to support and regulate AI technology and applications[1].

Key to the regulatory discourse is the 'black box problem' in AI models. They are end products of an algorithm (set of procedures) that has been trained on a large set of examples (data). To keep the intellectual property intact, generally models are hidden or are put in a 'black box'[2]. Consumption oriented opaqueness where if the model is able to give an output to the user have increasingly become challenging which require adequate policy attention. In addition, even developers are facing difficulty explaining these models, especially the ones that use emergent deep learning algorithms.

One of the key policy challenges is understanding the difficulty of explaining and understanding the complex decisions made by AI systems throughout they evolutionary and application cycle. In the context of superintelligent/supercognitive AI, *Finitum Non Capax Infiniti* (the finite cannot contain the infinite) becomes contextually resonant, contrasting the wide range of differences between human and machine capabilities and consequent challenges in comprehensibility. The lack of clear tangible explanations can severely hinder our ability to evaluate, debug, and improve AI systems, especially when failures occur. The current state of AI poses significant risks and concerted efforts are needed to make these systems more transparent and their decisions more interpretable to those who rely on them[3,4], one such effort is explainable artificial intelligence (XAI)[5] initiative.

As AI's capability grows, calls for regulation have increased to ensure safety, accountability and transparency. Unfortunately, standard regulatory approaches centred on ex-ante impact analysis and risk assessment seem unlikely to sustain. For effective regulation, it is essential to recognize that AI constitutes a Complex Adaptive System (CAS) that does not follow a predictable and fundamentally deterministic path. AI systems combine elements of CAS - where multiple components interact and evolve in nonlinear ways. Much like chaos theory and the butterfly effect, small initial changes in complex AI systems can cascade into disproportionate impacts potentially going out of control. The evolutionary trajectory of AI and

---

[1] Regulate AI? Here's What That Might Mean in the US. (2023). Bloomberg.com. [online] 30 Oct. Available at: https://www.bloomberg.com/news/articles/2023-10-30/ai-regulation-what-biden-s-new-rules-might-mean-in-the-us#xj4y7vzkg.

[2] Bagchi, S. (2023) Why we need to see inside AI's Black Box, Scientific American. Available at: https://www.scientificamerican.com/article/why-we-need-to-see-inside-ais-black-box/ (Accessed: 06 November 2023).

[3] Yampolskiy, R. (07 2020) 'Unexplainability and Incomprehensibility of AI', Journal of Artificial Intelligence and Consciousness, 07, pp. 1–15. doi: 10.1142/S2705078520500150.

[4] sdgs.un.org. (n.d.). Beyond a black-box approach to artificial intelligence policy – a simple guide to definitions, functions and technology types, by Richard A Roehrl (DESA) | Department of Economic and Social Affairs. [online] Available at: https://sdgs.un.org/documents/beyond-black-box-approach-artificial-intelligence-policy-simple-guide-definitions [Accessed 6 Nov. 2023].

[5] Johal, A. (2023) From black box to glass box: The evolution of explainable AI, Medium. Available at: https://medium.com/geekculture/from-black-box-to-glass-box-the-evolution-of-explainable-ai-c4b932c9fe94 (Accessed: 06 November 2023).

the unintended consequences it may engender cannot be understood or fully anticipated a priori through a reductionist approach, which is why most of the reactionary and conservative regulations keep spiralling in a revision loop.

As AI rapidly approaches and potentially surpasses human-level capabilities, an array of profound danger arises. One concern is the loss of control over recursively self-improving AI that eclipses human comprehension, also called runaway AI [6,7,8]. Additionally, interconnected critical infrastructure could face disruption through hacking of systems such as power grids. The integration of AI into defence, space, and biotechnology could also escalate cyber, outer space, nuclear, and biowarfare risks through engineered pathogens, autonomous weapons, etc. AI-powered mass surveillance[9,10] risk the erosion of privacy, free discourse disruption, and manipulation of information.

However, more subtle dangers also arise from advanced AI's ability to manipulate perceptions and simulated realities[11]. Through a combination of surveillance, persuasive messaging and synthetic media generation, malevolent AI could increasingly control information ecosystems and even fabricate customized deceitful realities to coerce human behavior. As realities increasingly blur between virtual and physical, developing oversight and security protocols will be critical to preserving human agency and autonomy.[12] Indeed, given the unintended consequences of self-generation, the AI need not be malevolent, but in fact attempting to be benevolent (eradication of cancer from the world could mean eradication of the disease or the patients).

This working paper undertakes a comparative analysis across several AI regulations and distils key learnings from global practices and ethical charters. The analysis demonstrates significant gaps vis-a-vis the parameters above in existing regulatory regimes. Better understanding of AI regulation and ability to regulate would also help us effectively regulate quantum computing before it blows out of proportions[13].

We propose a tailored forward-looking regulatory framework with a new lens of AI as a CAS. It recommends an adaptive approach cantered on setting ethical boundaries, mandated human oversight for high-risk applications, partitioning AI systems to limit contagion, and governing AI as a living technology that evolves in ways hard to fully anticipate today. The aim is not to exactly forecast and meticulously regulate AI's developmental arc over decades; instead,

---

[6] Ebhardt, J. D. (2018). The Threat of Artificial Superintelligence. Exigence, 2 (1). Retrieved from https://commons.vccs.edu/exigence/vol2/iss1/1

[7] MIT Technology Review. (n.d.). Yes, We Are Worried About the Existential Risk of Artificial Intelligence. [online] Available at: https://www.technologyreview.com/2016/11/02/156285/yes-we-are-worried-about-the-existential-risk-of-artificial-intelligence/.

[8] Davidson, Tom. "The Danger of Runaway AI." Journal of Democracy, vol. 34 no. 4, 2023, p. 132-140. Project MUSE, https://doi.org/10.1353/jod.2023.a907694.

[9] Feldstein, S. (2019). The Global Expansion of AI Surveillance. [online] Carnegie Endowment for International Peace. Available at: https://carnegieendowment.org/2019/09/17/global-expansion-of-ai-surveillance-pub-79847.

[10]Baptista, E. (2022). China uses AI software to improve its surveillance capabilities. Reuters. [online] 8 Apr. Available at: https://www.reuters.com/world/china/china-uses-ai-software-improve-its-surveillance-capabilities-2022-04-08/.

[11] Mustak, M. et al. (2023) 'Deepfakes: Deceptions, mitigations, and opportunities', Journal of Business Research, 154, p. 113368. doi: 10.1016/j.jbusres.2022.113368.

[12] Emsley, R. ChatGPT: these are not hallucinations – they're fabrications and falsifications. Schizophr 9, 52 (2023). https://doi.org/10.1038/s41537-023-00379-4

[13] Kop, V.W., Mauritz (n.d.). Why Quantum Computing Is Even More Dangerous Than Artificial Intelligence. [online] Foreign Policy. Available at: https://foreignpolicy.com/2022/08/21/quantum-computing-artificial-intelligence-ai-technology-regulation/.

it is to institute ethical guardrails today for foreseeable risks, mechanisms to dynamically adapt safeguards as needed, and sufficient human oversight throughout AI life cycles.

2. Policy Analysis

In the broadest sense of AI regulation, including data protection and privacy laws that indirectly affect AI development and deployment, one could argue that the earliest forms of AI-related regulation date back to the data protection laws of the 1970s and 1980s. For instance, the OECD's Guidelines on the Protection of Privacy and Transborder Flows of Personal Data, established in 1980, set out principles that are fundamental to the use of AI, such as data minimization and purpose specification.[14]

Specific to AI regulation, which is designed with the express purpose of governing AI systems, is a much more recent development, largely emerging in the 21st century as AI technology began to proliferate modern living and global development. For example, in 2019, the High-Level Expert Group on Artificial Intelligence, which was appointed by the European Commission, released Ethics Guidelines for Trustworthy AI[15], which outlines a framework for achieving ethical AI that is lawful, ethical, and robust. The European Union has been particularly proactive in this area, proposing the Artificial Intelligence Act in April 2021[16], which is often cited as one of the first comprehensive legislative proposals to regulate AI specifically.

It's important to note that various countries and regions are at different stages of developing and implementing AI regulation, and many are still in the process of adapting existing laws or creating new ones to address the unique challenges posed by AI technologies. These policies range from combinations of sporadic frameworks coupled with self-regulation to conservative and stringent fundamental rights-based regulation, and regulations that give the State complete control.

United States

The 2023 Update on The National Artificial Intelligence R&D Strategic Plan[17] emphasises the role of the federal government in ensuring responsible development of AI technologies and making variety of thoughtful investments across diverse AI R&D and deployments for public good. It reaffirms key eight strategies of the 2016 draft[18] and 2019 update[19].

The US strategy emphasizes long-term investments in responsible AI research to advance foundational capabilities while mitigating risks from generative systems. It also focuses on effective human-AI collaboration through understanding complementarity, developing competencies, and measuring team performance. Additional priorities include expanding access to high-quality training data and environments, establishing multi-dimensional evaluation frameworks for AI systems standards and benchmarks from Administration's Blueprint for an

---

[14] Michael Kirby, The history, achievement and future of the 1980 OECD guidelines on privacy, International Data Privacy Law, Volume 1, Issue 1, February 2011, Pages 6–14, https://doi.org/10.1093/idpl/ipq002

[15] European Commission (2021). Ethics guidelines for trustworthy AI | Shaping Europe's digital future. [online] digital-strategy.ec.europa.eu. Available at: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai.

[16] European Parliament (2023). *EU AI Act: First Regulation on Artificial Intelligence | News | European Parliament*. [online] www.europarl.europa.eu. Available at: https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence.

[17]https://www.whitehouse.gov/wp-content/uploads/2023/05/National-Artificial-Intelligence-Research-and-Development-Strategic-Plan-2023-Update.pdf

[18] https://www.nitrd.gov/pubs/national_ai_rd_strategic_plan.pdf

[19] https://www.nitrd.gov/pubs/National-AI-RD-Strategy-2019.pdf

AI Bill of Rights[20] and AI Risk Management Framework (RMF)[21], building an AI-ready workforce through education, and promoting public-private partnerships for sustained R&D investment and practical transition of advances. The overarching aim is supporting innovation in beneficial AI while strengthening oversight to minimize harms.

The US approach to regulation has evolved from a complete 'hands-off' one initially to the recent shift towards 'self -regulation' and 'voluntary commitments'. The US has a plethora of regulations on AI (Federal, Congressional, State, Industry Self-Regulation)[22] which has made their regulatory stance laissez-faire for a while given that Federal administration did not want to contain the development of technology, They have asserted that that US administration will *"maintaining American leadership in AI requires a concerted effort to promote advancements in technology and innovation, while protecting American technology, economic and national security, civil liberties, privacy, and American values and enhancing international and industry collaboration with foreign partners and allies*"[23]. Following which US Administration issued a Memorandum in 2019 for the Heads of Executive Departments and Agencies issuing Guidance for Regulation of Artificial Intelligence Applications[24] that stated:

> "*As stated in Executive Order 13859, "the policy of the United States Government [is] to sustain and enhance the scientific, technological, and economic leadership position of the United States in AI." The deployment of AI holds the promise to improve safety, fairness, welfare, transparency, and other social goals, and America's maintenance of its status as a global leader in AI development is vital to preserving our economic and national security. The importance of developing and deploying AI requires a regulatory approach that fosters innovation, growth, and engenders trust, while protecting core American values, through both regulatory and nonregulatory actions and reducing unnecessary barriers to the development and deployment of AI.*
>
> ***To that end, Federal agencies must avoid regulatory or non-regulatory actions that needlessly hamper AI innovation and growth. Where permitted by law, when deciding whether and how to regulate in an area that may affect AI applications, agencies should assess the effect of the potential regulation on AI innovation and growth. Agencies must avoid a precautionary approach that holds AI systems to such an impossibly high standard that society cannot enjoy their benefits.***"

The aforementioned memorandum stated that in cases where existing regulations suffice or the costs of new regulations outweigh their benefits for AI applications, agencies may opt for non-regulatory approaches. This emphasizes avoiding hampering innovation and assessing regulatory impact on AI growth. This aligns with laissez-faire principles prioritizing technological advancement over precaution. Additionally, agencies are encouraged to prefer voluntary consensus standards developed by the private sector or independent organizations for managing AI risks, considering these standards' robustness before proposing new regulations or compliance programs.

---

[20] https://www.whitehouse.gov/wp-content/uploads/2022/10/Blueprint-for-an-AI-Bill-of-Rights.pdf
[21] https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf
[22] www.morganlewis.com. (n.d.). The United States' Approach to AI Regulation: Key Considerations for Companies. [online] Available at: https://www.morganlewis.com/pubs/2023/05/the-united-states-approach-to-ai-regulation-key-considerations-for-companies.
[23] Executive Order 13859 of February 11, 2019; Federal Register. (2019). Maintaining American Leadership in Artificial Intelligence. [online] Available at: https://www.federalregister.gov/documents/2019/02/14/2019-02544/maintaining-american-leadership-in-artificial-intelligence.
[24] The White House, Vought, R. T.(2019). MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES FROM. [online] Available at: https://www.whitehouse.gov/wp-content/uploads/2020/11/M-21-06.pdf

The approach marked a shift in July 2023 when Biden-Harris Administration secured Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI[25]. While the *"voluntary commitments"* and self-regulation from companies[26] gesticulates a positive intention toward public safety and ethical considerations but its relying solely on these pledges presents a significant risk in itself[27] and being argued as pro-business[28,29]. The commitments to pre-release testing, information sharing, and investment in security are steps in the right direction, but without a binding regulatory framework, these promises lack enforceability and accountability. History has shown that self-regulation in rapidly advancing technological domains can lead to gaps in safety and ethics, as corporate interests might overshadow public welfare. The very nature of AI as an evolving and complex field makes it difficult to anticipate all potential risks, and voluntary measures are inadequate to cover unforeseen vulnerabilities. The US's regulatory approach to AI, with its emphasis on voluntary industry commitments, carries significant risks and may be inadequate to address the full range of potential issues posed by the development and deployment of AI technologies.

This self-regulatory approach was further enhanced with proposed limited oversight in the latest OMB Memorandum[30] that was sent out for public review in October 2023. It reflects a regulatory stance focused on promoting responsible AI innovation while rigorously managing its risks through "*specific minimum risk management practices for uses of AI that impact the rights and safety of the public, especially those affecting public rights and safety*". It mandates specific governance structures, strategies for AI implementation, and comprehensive risk management practices, emphasizing accountability, transparency, and public engagement. However, it does not indicate any regulatory guidance for the agencies as indicated in the 2019 memorandum, only suggests better governance. Since it's not a superseding document and still open to public review, the previous guidelines do remain effective. The industry self-regulation has been seen to promote industry's interest at the expense of the public[31] and its failures have been noted across disciplines[32,33].

---

[25] The White House (2023). FACT SHEET: Biden-Harris Administration Secures Voluntary Commitments from Leading Artificial Intelligence Companies to Manage the Risks Posed by AI. [online] The White House. Available at: https://www.whitehouse.gov/briefing-room/statements-releases/2023/07/21/fact-sheet-biden-harris-administration-secures-voluntary-commitments-from-leading-artificial-intelligence-companies-to-manage-the-risks-posed-by-ai/

[26] Zakrzewski, C. and Tiku, N. (2023). AI companies form new safety body, while Congress plays catch-up. Washington Post. [online] 27 Jul. Available at: https://www.washingtonpost.com/technology/2023/07/26/ai-regulation-created-google-openai-microsoft/.

[27] Alikhani, K. (n.d.). Council Post: Why It's Dangerous For AI To Regulate Itself. [online] Forbes. Available at: https://www.forbes.com/sites/forbestechcouncil/2019/03/22/why-its-dangerous-for-ai-to-regulate-itself/?sh=3d79c20d7e54 [Accessed 9 Nov. 2023].

[28] VentureBeat. (2023). Why self-regulation of AI is a smart business move. [online] Available at: https://venturebeat.com/ai/why-self-regulation-of-ai-is-a-smart-business-move/ [Accessed 9 Nov. 2023].

[29] Adonis Hoffman, opinion contributor (2023). Why self-regulation is best for artificial intelligence. [online] The Hill. Available at: https://thehill.com/opinion/4300288-why-self-regulation-is-best-for-artificial-intelligence/ [Accessed 9 Nov. 2023].

[30] The White House, Young, S. (2023). PROPOSED MEMORANDUM FOR THE HEADS OF EXECUTIVE DEPARTMENTS AND AGENCIES FROM. [online] Available at: https://www.whitehouse.gov/wp-content/uploads/2023/11/AI-in-Government-Memo-draft-for-public-review.pdf [Accessed 16 Nov. 2023].

[31] Edwards, Benjamin P., "The Dark Side of Self-Regulation" (2017). Scholarly Works. 1117.https://scholars.law.unlv.edu/facpub/1117

[32] UK Parliament, EDM (Early Day Motion)1345: tabled on 08 February 1993, FAILURE OF SELF-REGULATION OF FINANCIAL SERVICES, Available at: https://edm.parliament.uk/early-day-motion/6448/failure-of-selfregulation-of-financial-services

[33] Amit Narang, O.C. (2019). Corporate self-regulation is failing. [online] The Hill. Available at: https://thehill.com/blogs/congress-blog/the-administration/436328-corporate-self-regulation-is-failing/.

Recently, the US Government took another step towards regulation through the Presidential Executive Order of 30 October 2023[34]. The order presents a plan to address the risks of AI, requiring companies to rigorously test new products like ChatGPT and share findings with U.S. officials, a move seen as a significant regulatory step. In terms of regulation, it calls for a '*potential voluntary, regulatory, and international mechanisms to manage the risks and maximize the benefits*' of AI in its emergent regulatory framework while maintaining that "*Administration will support programs to provide Americans the skills they need for the age of AI and attract the world's AI talent to our shores — not just to study, but to stay — so that the companies and technologies of the future are made in America"* making regulatory sector specific regulatory authorities independent allowing them, "*…as they deem appropriate, to consider whether to mandate guidance through regulatory action in their areas of authority and responsibility*". The order includes provisions like AI-enabled product watermarks to guard against deepfakes and mandates biotech precautions, but these are advisory rather than mandatory. The order also instructs U.S. agencies to adopt AI practices, setting an example for the private sector, with specific departments addressing AI's threat to critical infrastructure. Highlighting AI threats to critical infrastructure acknowledges interconnected systemic risks. Aligned with complex systems perspective. One key challenge that it highlights is the risk of AI induced biowarfare resounding the concerns of a 2017 study by John T. O'Brien and Cassidy Nelson[35]. The Executive Order moves towards more oversight by requiring testing and information sharing. This a somewhat departure from pure laissez-faire, but still remains advisory rather than mandatory regulation. The Blueprint for an AI Bill of Rights[36] highlights the development of responsible automated systems and creates barriers such as the ones we also propose. Overall, recent gestures indicate movement away from pure laissez-faire to self-regulation and limited oversight. But binding governance is yet to be seen.

United Kingdom

The United Kingdom adopted a cautious yet pro-innovation stance towards AI governance, aimed at supporting growth while managing risks. The 2017 Industrial Strategy White Paper[37], now retracted, initially stimulated business environments, skills and infrastructure. It was transitioned to the 2021 Plan for Growth, which does not explicitly mention AI[38]. The 2018 AI Sector Deal committed £1 billion to advance AI across sectors, emphasizing research, innovation, infrastructure and skills development[39]. This aligned with the 2022 National AI Strategy's[40] core focuses on sustaining UK AI leadership. The UK administration advocates an application-centric framework, with regulations tailored to specific use contexts rather than one-size-fits-all governance. Overall, the UK strategy rejects

[34] The White House (2023). Executive Order on the Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence. [online] The White House. Available at: https://www.whitehouse.gov/briefing-room/presidential-actions/2023/10/30/executive-order-on-the-safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence/.

[35] O'Brien JT, Nelson C. Assessing the Risks Posed by the Convergence of Artificial Intelligence and Biotechnology. Health Secur. 2020 May/Jun;18(3):219-227. doi: 10.1089/hs.2019.0122. PMID: 32559154; PMCID: PMC7310294.

[36] The White House (2022). Blueprint for an AI Bill of Rights. [online] The White House. Available at: https://www.whitehouse.gov/ostp/ai-bill-of-rights/.

[37] Department for Business, Energy & Industrial Strategy (2017). Industrial Strategy: building a Britain fit for the future. [online] GOV.UK. Available at: https://www.gov.uk/government/publications/industrial-strategy-building-a-britain-fit-for-the-future.

[38] GOV.UK. (2021). Build Back Better: our plan for growth (HTML). [online] Available at: https://www.gov.uk/government/publications/build-back-better-our-plan-for-growth/build-back-better-our-plan-for-growth-html.

[39] GOV.UK. (2022). National AI Strategy - HTML version. [online] Available at: https://www.gov.uk/government/publications/national-ai-strategy/national-ai-strategy-html-version.

[40] GOV.UK. (2022). National AI Strategy - HTML version. [online] Available at: https://www.gov.uk/government/publications/national-ai-strategy/national-ai-strategy-html-version.

precautionary restraints on AI advancement focusing on innovation, instead preferring selective governance guided by ethical principles and aligned with national interests. [41] This is largely antithetical to our principle of 'Guardrails' advocated later in the paper.

It centres on targeted support for technology growth, while instituting sector specific regulations to address foreseeable sectoral risks. This approach enables sector-specific regulators to customize AI governance based on contextual needs, capitalizing on their specialized knowledge and experience. This is unlike the specialized common regulator, given the pervasive nature and the risks of super connected AI systems, that we advocate.

Principally, UK articulates a quintet of foundational principles designed to scaffold the suggested regulatory apparatus[42], principles which are applicable across a broad spectrum of sectors such as safety, appropriate transparency and explainability, equity, and mandatory accountability and governance. It also advances the proposition that, where deemed pertinent, aggrieved parties and entities engaged in the AI lifecycle should have the agency to challenge AI-generated decisions or outcomes that are detrimental or pose a significant risk. While these principles, especially transparency and explainability find resonance in the approach we propose, the principles-based approach to AI regulation shared the UK is undermined by the ambiguity of the principles and the challenge of converting them into actionable standards, further complicated by potential conflicts with established legal frameworks. The minimal legislative efforts do not equip regulators to enforce these principles effectively.

A significant deficiency in the proposal is the lack of a detailed regulatory model, particularly in defining the operational structure and mechanisms of the central regulatory function, leaving questions about standardization, resources, accountability, and implementation.

The ultimate configuration of this proposed regulatory framework, as outlined in the White Paper, may yet undergo evolution. Several updates have already emerged.

The UK's pro-innovation agenda in AI regulation is concerning for its potential inadequacy to meet the complex challenges posed by AI[43]. The focus on positioning the UK as a prime location for AI companies seems to favour industrial growth at the expense of public welfare in the face of AI's risks. They state that:

> "*While we should capitalise on the benefits of these technologies, we should also not overlook the new risks that may arise from their use, nor the unease that the complexity of AI technologies can produce in the wider public. We already know that some uses of AI could damage our physical and mental health, infringe on the privacy of individuals and undermine human rights.*
>
> *Public trust in AI will be undermined unless these risks, and wider concerns about the potential for bias and discrimination, are addressed. By building trust, we can accelerate the adoption of AI across the UK to maximise the economic and social benefits that the technology can deliver, while attracting investment and stimulating the creation of high-skilled AI jobs. In order to maintain the UK's position as a global AI leader, we need to ensure that the public continues to see how the benefits of AI can outweigh the risks.*"[44]

---

[41] GOV.UK (2023). A pro-innovation approach to AI regulation. [online] GOV.UK. Available at: https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper.
[42] Ada Lovelace Institute. Regulating AI in the UK. [online] Available at: https://www.adalovelaceinstitute.org/report/regulating-ai-in-the-uk/.
[43] Charlesworth, Andrew and Fotheringham, Kit and Gavaghan, Colin and Sanchez-Graells, Albert and Torrible, Clare, Response to the UK's March 2023 White Paper "A pro-innovation approach to AI regulation" (June 19, 2023). Available at SSRN: https://ssrn.com/abstract=4477368 or http://dx.doi.org/10.2139/ssrn.4477368
[44] GOV.UK (2023). A pro-innovation approach to AI regulation. [online] GOV.UK. Available at: https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper.

It goes on further to state that:

> "*New rigid and onerous legislative requirements on businesses could hold back AI innovation and reduce our ability to respond quickly and in a proportionate way to future technological advances. Instead, the principles will be issued on a non-statutory basis and implemented by existing regulators. This approach makes use of regulators' domain-specific expertise to tailor the implementation of the principles to the specific context in which AI is used. During the initial period of implementation, we will continue to collaborate with regulators to identify any barriers to the proportionate application of the principles and evaluate whether the non-statutory framework is having the desired effect.*"[45]

The purportedly agile regulatory framework may be a pretext for possible deregulation, advocating for the removal of "barriers" without justifying the abandonment of existing safeguards amidst heightened AI risks. This could compromise the autonomy and efficacy of regulatory bodies essential for overseeing AI's multifaceted impact.

Similar to US's 'hands-off' approach, UK's pro-innovation approach where principles governing AI regulation in the UK are issued on a non-statutory basis. This approach allows existing domain specific regulators to tailor AI governance to specific contexts, leveraging their domain-specific expertise. This method underscores a laissez-faire attitude by preferring guidance and principles over hard legislation.

European Union

In contrast to the UK and the US, the European Union (EU) has taken a risk-based approach in recognizing the need to balance AI innovation with safeguarding fundamental rights and addressing potential risks[46]. The EU's approach to AI regulation is centred around the European Commission's proposal for the Regulation on Artificial Intelligence[47]. This legislative proposal is a component of a more comprehensive Artificial Intelligence strategy, inclusive of the refreshed Coordinated Plan on AI[48].

The cornerstone of the EU's AI regulations is the EU's AI Act[49], which aims to create a comprehensive framework for AI governance. The Act emphasizes a risk-based approach, categorizing AI systems into different levels of risk and imposing specific obligations accordingly. It prohibits AI systems that are considered "unacceptable" under EU law. In this approach, it classifies AI systems into four categories: Unacceptable Risk, High Risk, Limited Risk, and Minimal Risk. Unacceptable Risk AI systems, such as those used for social scoring or manipulating human behaviour, are outright banned. High-risk AI systems, such as those employed in critical infrastructure, healthcare, or law enforcement, are subject to stringent requirements, including conformity assessments, documentation, and human oversight. The Act also imposes specific obligations on high-risk AI systems to ensure transparency, accountability,

---

[45] ibid

[46] Feingold, S. (2023). *'On artificial intelligence, trust is a must, not a nice to have,' one lawmaker said. #AI*. [online] World Economic Forum. Available at: https://www.weforum.org/agenda/2023/06/european-union-ai-act-explained/.

[47] digital-strategy.ec.europa.eu. (2023). *Regulatory framework proposal on artificial intelligence | Shaping Europe's digital future*. [online] Available at: https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai#:~:text=The%20Commission%20is%20proposing%20the.

[48] digital-strategy.ec.europa.eu. (n.d.). Coordinated Plan on Artificial Intelligence | Shaping Europe's digital future. [online] Available at: https://digital-strategy.ec.europa.eu/en/policies/plan-ai.

[49] European Commission (2021). EUR-Lex - 52021PC0206 - EN - EUR-Lex. [online] Europa.eu. Available at: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A52021PC0206.

and robust governance. High-risk AI systems also must undergo conformity assessment procedures before being placed on the market or deployed. [50,51]

EU after successfully deploying GDPR and GDR passed the revised draft AI Act in European Parliament earlier this year and its now with European Council for negation with a goal to get it passed before end of 2023.[52] This revised draft took a tough stance on AI risks and prohibited AI's use for biometric surveillance.

The critical loophole in the European regulation is its regulatory localization in 'European Values' which are ambiguous for the continent. This has been raising a variety of concerns on interpretation of the term and then eventually the law.

The designating risk system is completely dependent on the designation of the risk. However, in a complex system such as AI, it cannot be deterministically designated. What might be perceived as low or no risk can lead to cascading high risk outcomes given the unpredictability of the algorithm and perpetuation of processing logic through deep learning processes even on a heterogeneous data set.

There is a difference of opinion between European Parliament & European Council and European Commission. The upcoming "trialogue" discussions are crucial for finalizing a cohesive AI regulatory framework, with pressing issues like AI definitions, high-risk categories, and biometric identification on the table. The proposals differ in their approach to enforcement— European Parliament favours a single national surveillance authority per member state for stronger, more centralized oversight, diverging from the multiple authorities model proposed by the European Council and Commission. Fundamental challenges that are of concern in EU's AI regulation are[53]:

(i) **Regulation of AI Systems by Risk Category**

a) Risk-Based Approach: The AI Act categorizes AI technologies based on their risk to safety, health, or fundamental rights, sorting them into unacceptable, high, and limited or minimal risk levels.

b) Prohibited AI Systems: AI systems considered unacceptable are banned outright under Article 5. These include systems that manipulate behavior, exploit vulnerabilities, or perform social scoring that results in harm or discrimination.

c) High-Risk AI Systems: High-risk AI systems are allowed but come with stringent obligations, such as informing human operators of risks like automation bias and conducting fundamental rights impact assessments.

d) Limited or Minimal-Risk AI Systems: AI systems with lower risks have minimal obligations, primarily around transparency to enable informed user decisions.

---

[50] Title II & III of the EU AI Act

[51] Chee, F.Y., Mukherjee, S., Chee, F.Y. and Mukherjee, S. (2023). EU lawmakers vote for tougher AI rules as draft moves to final stage. Reuters. [online] 14 Jun. Available at: https://www.reuters.com/technology/eu-lawmakers-agree-changes-draft-artificial-intelligence-rules-2023-06-14/.

[52] www.europarl.europa.eu. (2023). EU AI Act: first regulation on artificial intelligence | News | European Parliament. [online]. Available at: https://www.europarl.europa.eu/news/en/headlines/society/20230601STO93804/eu-ai-act-first-regulation-on-artificial-intelligence#:~:text=On%2014%20June%202023%2C%20MEPs [Accessed 6 Nov. 2023].

[53] Müge Fazlioglu, (2023). Contentious areas in the EU AI Act trilogues. [online] Available at: https://iapp.org/news/a/contentious-areas-in-the-eu-ai-act-trilogues/.

e) Contention Over Risk Categories: Debate continues over which systems belong to which risk category, especially regarding biometric surveillance. The European Parliament has pushed for a broader list of prohibitions, including real-time remote biometric identification, despite attempts for derogations. Conversely, the European Council advocates for a narrower scope of prohibited systems. Upcoming trilogue discussions are expected to address these disputes and potentially alter the list of prohibitions and exceptions within Article 5.

(ii) **High-Risk AI System Obligations and Categories**

a) Article 6 and Annex III Requirements:
- Eight primary high-risk AI system categories defined.
- Categories include critical infrastructure, biometric identification, education, employment, essential services, law enforcement, border control, and administration of justice/democracy.
- Additional sub-categories and expansions by the European Parliament include social media platforms and election influence.

b) High-Risk Classification Ambiguity:
- Challenges in determining high-risk status for AI systems.
- European Commission's list revised by the Council and Parliament.
- Risk classification studies show unclear risk levels for many AI systems.

While we agree with the setup of a separate regulator proposed in the legislation, the risk-based approach to AI regulation, which categorizes AI applications based on their potential threat to safety, health, or fundamental rights, presents several challenges. There is no consensus on what constitutes 'high risk' in AI. While it may be obvious in a few instances (e.g., nuclear weapons), the problem is with ex-ante risk categorisation. AI technologies evolve rapidly, and a system deemed low-risk today may become high-risk tomorrow due to changes in usage, context, or capabilities. The challenge is that this approach only works for static, linear systems with predictable risks. AI combines qualities of complex adaptive systems (CAS) where components interact and evolve in nonlinear ways. This can lead to butterfly effects where small changes can cascade disproportionately through AI systems. Similarly, its evolutionary trajectory cannot be predicted through reductionist thinking.

China

China was one of the earliest actors to have an AI strategic plan. They came up with New Generation Artificial Intelligence Development Plan (2015-2030)[54]. The document delineates a strategic blueprint for China's advancement in AI. Initially, by the terminus of 2020, the ambition was to narrow the innovation chasm separating China from global frontrunners in technology, elevating its AI sector to a status of international prominence. Subsequently, the intermediate milestone of 2025 targets seminal triumphs in the core research domains of AI and aims to establish the foundational structure for Next Generation Artificial Intelligence (NGAI). The culminating objective for 2030 envisions China's emergence as a global vanguard of AI innovation, complete with a comprehensive suite of regulatory, legal, and ethical frameworks to guide the responsible evolution of AI. Since China started very early in understanding, harnessing, and regulating AI, it is in a better position to regulate more complex applications than any other country[55].

[54] Webster, G., Creemers, R., Kania, E. and Triolo, P. (2017). Full translation: China's 'new generation artificial intelligence development plan' (2017). [online] DigiChina. Available at: https://digichina.stanford.edu/work/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/.

[55] Sheehan, M. (n.d.). China's AI Regulations and How They Get Made. [online] Carnegie Endowment for International Peace. Available at: https://carnegieendowment.org/2023/07/10/china-s-ai-regulations-and-how-they-get-made-pub-90117.

In addressing the evolving landscape of algorithmic governance, China has instituted a trio of seminal and influential regulatory frameworks targeting artificial intelligence (AI) technologies. These include the 2021 mandate on recommendation algorithms[56], the 2022 provisions governing deep synthetic media[57], and the prospective 2023 guidelines concerning generative AI systems[58].

The core objective shared by these regulatory actions is to fortify information oversight, yet their scope extends beyond this focal point, incorporating various significant clauses. Specifically, the 2021 regulation on recommendation algorithms introduces measures to curb pronounced price differentiation practices and safeguards the welfare of personnel influenced by algorithmically managed work schedules. The 2022 rule targeting deep synthetic content mandates that such material be clearly marked as artificially generated, thereby ensuring transparency.

The proposed 2023 draft rules on generative AI pose a particularly stringent criterion, mandating the veracity and precision of both the input data and the outputs produced by AI, presenting a challenging standard for AI chatbot services to meet. Moreover, each set of regulations compels developers to formally register their algorithms with a novel state-operated algorithm registry. This repository is designed to catalogue detailed information regarding the training methodologies of algorithms. Additionally, developers are obliged to conduct and submit a security self-assessment, attesting to the integrity and safety of their algorithmic systems.

The Wuhan experience has shown the potential pitfalls of an all-powerful centralised bureaucratic system with the potential lab-leak origin of Covid-19[59,60,61] It is only imperative to suggest that complete bureaucratic control over algorithms, training sets, and models with a goal to maximize interests would lead to catastrophic outcomes and expecting the Chinse state to follow transparent and ethical standards would not be a wise expectation. If the proactive regulation in China continues to go at the pace, it might be in a position of crating global standards much before the international community is able to build consensus on fundamental issues regarding the technology and its applications[62]. It is also attempting to balance multiple goals that may not be mutually conducive.[63]

---

[56] www.cac.gov.cn. (n.d.). **互**联网信息服务算法推荐管理规定-**中共中央网**络安全和信息化委员会办公室. [online] Available at: http://www.cac.gov.cn/2022-01/04/c_1642894606364259.htm.

[57] www.cac.gov.cn. (n.d.). **国家互**联网信息办公室等三部门发布《互联网信息服务深度合成管理规定》-**中共中央网**络安全和信息化委员会办公室. Available at: http://www.cac.gov.cn/2022-12/11/c_1672221949318230.htm

[58] www.cac.gov.cn. (n.d.). **生成式人工智能服**务管理暂行办**法_中央网**络安全和信息化委员会办公室. [online] Available at: http://www.cac.gov.cn/2023-07/13/c_1690898327029107.htm.

[59] Knight D. COVID-19 Pandemic Origins: Bioweapons and the History of Laboratory Leaks. South Med J. 2021 Aug;114(8):465-467. doi: 10.14423/SMJ.0000000000001283. PMID: 34345925; PMCID: PMC8300139.

[60] Ciuriak, Dan, Why Wuhan? What the Circumstantial Evidence Says About the Origins of COVID-19 (June 6, 2021). Available at SSRN: https://ssrn.com/abstract=3763542 or http://dx.doi.org/10.2139/ssrn.3763542

[61] Cohen, J. (2022). Do three new studies add up to proof of COVID-19's origin in a Wuhan animal market? [online] www.science.org. Available at: https://www.science.org/content/article/do-three-new-studies-add-proof-covid-19-s-origin-wuhan-animal-market.

[62] China Briefing News. (2022). China to Regulate Deep Synthesis (Deepfake) Technology from 2023. [online] Available at: https://www.china-briefing.com/news/china-to-regulate-deep-synthesis-deep-fake-technology-starting-january-2023/.

[63] S. Zheng, J. Zhang, Time. (2023). China Tries to Balance State Control and State Support of AI. [online] Available at: https://time.com/6304831/china-ai-regulations/.

China is aggressively advancing in the Internet of Things (IoT), aiming to become a global leader and posing significant challenges to non-Chinese economic and security interests. With its expansive growth, IoT is set to revolutionize numerous sectors with billions of connected devices. Yet, issues of operation, safety, and data security remain unresolved. China's strategy encompasses leveraging national resources and its large market for IoT promotion, resulting in a commanding presence in global markets. China's IoT strategy is self-serving, prioritizing Chinese interests and showing little regard for the economic and national security concerns of non-Chinese entities[64]. While, we agree with the human hierarchical control, a super connected AI system without any 'Guardrails' and 'Partitions' makes it almost impossible to govern and is rampant with risks of a limited system failure infecting the entire ecosystem.

The United Nations

It should be highlighted that currently, there is no comprehensive international regulatory framework governing AI that is legally binding, neither under the auspices of the UN nor within any regional intergovernmental organizations.

Following which High-level Panel on Digital Cooperation which they outline Roadmap for Digital Cooperation[65] and resulted in the establishment of the Office of the Secretary-General's Envoy on Technology (OSET), which spearheads the initiatives of the UN Secretary-General and the UN system in advancing new technologies, encompassing AI. In addition to the UN Secretary-General's Roadmap for Digital Cooperation, the Common Agenda, and the Global Digital Compact—which features a principal action on fostering worldwide cooperation on AI and are all driven by the UN Secretariat—various UN entities have introduced their distinct reports and initiatives. These are aimed at tackling AI and related emerging technologies, along with the specific challenges they pose to their individual mandates and stakeholders.[66,67,68,69,70].

UNESCO's AI Ethics global standards that recommend a series of principles that align with the advancement and safeguarding of human rights. It promotes fundamental tenets including transparency, accountability, and adherence to the rule of law in the digital realm have also been adopted.[71] These tenets resonate with the CAS principle of making AI systems transparent and their operations understandable, which is essential for accountability and ethical considerations.

In UNSC remarks, Secretary General also pointed out that, "*the interaction between AI and nuclear weapons, biotechnology, neurotechnology, and robotics is deeply alarming. Generative AI has enormous potential for good and evil at scale. Its creators themselves have warned that much bigger, potentially catastrophic*

[64] Henk H.F. Smid, www.thespacereview.com. (2023). The Space Review: Internet of Things: the China perspective. [online] Available at: https://www.thespacereview.com/article/4566/1.
[65] www.un.org. (n.d.). Secretary-General's Roadmap for Digital Cooperation. [online] Available at: https://www.un.org/en/content/digital-cooperation-roadmap/.
[66] www.itu.int. (n.d.). ITU Partner2Connect Digital Coalition. [online] Available at: https://www.itu.int/itu-d/sites/partner2connect/.
[67] www.unicef.org. (n.d.). Giga. [online] Available at: https://www.unicef.org/innovation/giga.
[68] sdgailab.org. (n.d.). SDG AI Lab. [online] Available at: https://sdgailab.org/ [Accessed 8 Nov. 2023].
[69] WHO Guidance (2021). Ethics and governance of artificial intelligence for health. [online] www.who.int. Available at: https://www.who.int/publications/i/item/9789240029200.
[70] www.icaiea.com. (n.d.). 2024 5th International Conference on Artificial Intelligence and Electromechanical Automation (AIEA 2024). [online] Available at: http://www.icaiea.com/ [Accessed 8 Nov. 2023].
[71] UNESCO (2022). Ethics of artificial intelligence | UNESCO. [online] www.unesco.org. Available at: https://www.unesco.org/en/artificial-intelligence/recommendation-ethics.

and existential risks lie ahead.[72]" The United Nations' approach to AI regulation, while aspirational and aligned with global cooperation, falls short of the specificity and enforceability required by the principles proposed for regulating AI as a Complex Adaptive System. The broad strokes of international collaboration, ethical standards, and human-centric AI development lack the rigorous, enforceable guardrails and partitions necessary to prevent systemic risks. As of now, the absence of a comprehensive, legally binding international regulatory framework for AI means that these principles and initiatives serve as guidelines or recommendations rather than enforceable rules. This underscores the emerging nature of global AI governance and the need for continued development towards binding international agreements.

3. AI: A Complex Adaptive System

Complex adaptive systems (CAS) describe a broad class of intricate, dynamic networks found throughout nature and society. Diverse agents interact and adapt in CAS, exhibiting emergent collective behaviours that are nonlinear and cannot be predicted by analysing individual components. Examples range from ant colonies to brains to cities. These decentralized, self-organizing systems evolve through feedback loops, phase transitions and sensitivity to initial conditions. Even small perturbations can cascade in chaotic, unpredictable ways. Understanding CAS requires embracing their inherent complexity rather than oversimplifying analyses. Conceptualizing artificial intelligence as sharing key qualities with CAS provides a nuanced perspective on governance to match its profound technological dynamism.[73]

From the lens of AI as a complex system, the emergent behaviour from simple rules has been an intriguing aspect of study across various disciplines. AI, particularly in its function as a complex adaptive system, epitomizes this phenomenon. Like the brain's neurons that modify connections based on environmental stimuli, or the financial market reflecting the collective will of its participants, AI systems learn and adapt, showcasing the power of agent-based models. In this realm, AI is akin to a self-organizing entity, where individual algorithms—acting as agents—follow basic protocols yet collectively manifest sophisticated behaviour. This adaptation is crucial for AI systems, enabling them to refine their predictive capabilities and responses to dynamic environments. Conventional scientific and mathematical models often fall short in explaining these adaptive processes, prompting a shift toward computational simulations and other innovative methodologies. The complexity in AI arises as these agents adjust not only their actions but also the fundamental rules guiding these actions in response to external changes. This self-modifying characteristic of AI agents is what underpins the field's complexity and challenges traditional analytical approaches, underscoring the need for new scientific paradigms to understand and harness the potential of AI as a complex adaptive system.[74]

Artificial intelligence exemplifies a "complex adaptive system" - where multiple components dynamically interact and evolve in nonlinear, unpredictable ways. AI systems combine diverse technologies like machine learning, neural networks, evolutionary algorithms and reinforcement learning. As these elements interact, they produce emergent behaviours that cannot be deduced by analysing individual components alone.

---

[72] www.un.org. (n.d.). Secretary-General's remarks to the Security Council on Artificial Intelligence | United Nations Secretary-General. [online] Available at: https://www.un.org/sg/en/content/sg/speeches/2023-07-18/secretary-generals-remarks-the-security-council-artificial-intelligence#:~:text=And%20the%20interaction%20between%20AI [Accessed 8 Nov. 2023].

[73]Chan, S. (2001). Complex Adaptive Systems. [online] Available at: https://web.mit.edu/esd.83/www/notebook/Complex%20Adaptive%20Systems.pdf.

[74] Booker, Lashon, and Stephanie Forrest, 'Adaptation, Evolution, and Intelligence', in Lashon Booker, and others (eds), Perspectives on Adaptation in Natural and Artificial Systems (New York, 2005; online edn, Oxford Academic, 12 Nov. 2020), https://doi.org/10.1093/oso/9780195162929.003.0004, accessed 9 Nov. 2023.

Like chaos theory's butterfly effect, small initial changes in AI systems can cascade and compound rapidly into disproportionate impacts if left unchecked. The intricacy of combinatorial explosions from multiple technology facets interacting makes modelling the long-term trajectory intractable. AI development resembles an evolutionary process, with continual interplay between exploration, selection, and propagation of technical variants.

This nonlinear complexity stems from the autonomy, interdependencies, and nesting of AI subsystems. For instance, a facial recognition algorithm may iterate through billions of parameter tweaks chosen by a separate neural architecture search algorithm. If the search algorithm evolves in unexpected ways, it dramatically alters the final facial recognition system.

These cascading effects challenge traditional reductionist thinking centred on prediction and control. AI's dynamic complexity requires a new regulatory perspective focused on instituting real-time guardrails based on ethical principles[75], rather than ex-ante restrictions attempting to forecast outcomes[76]. Setting initial boundary conditions provides a scaffolding to guide development, while retaining flexibility to course correct.

Consider the difficulty predicting future risks from today's AI research, which could evolve unrecognizably over decades. The combinatorial complexity makes modelling long-term capabilities speculative. This uncertainty highlights why adaptive governance is imperative for rapidly advancing technologies like AI. Rather than prescribe rigid ex-ante rules, regulatory approaches must themselves evolve as capabilities grow. Prudent oversight mechanisms will be critical to steering inevitable AI progress in beneficial directions for humanity.

4. Contours of AI Regulation: A Complex Adaptive System Approach

As we traverse the dynamic landscape of Artificial Intelligence (AI), it becomes increasingly evident that traditional regulatory frameworks might falter under the weight of this burgeoning technology[77]. AI, when viewed through the lens of Complex Adaptive System (CAS) thinking, offers new insights into how it can be effectively governed. CAS, characterized by its emergent behaviours and non-linear interactions, provides a suitable analogy for the intricate, evolving nature of AI. Building on this perspective, we propose a regulatory approach anchored in five core principles:

Instituting Guardrails and Partitions to 'Prevent Wildfire'

Drawing inspiration from CAS theory, it's imperative to establish robust 'Guardrails' and 'Partitions' around AI systems. These Guardrails in the form of boundary conditions delineate the operational space, ensuring AI technologies neither exceed their intended functions nor encroach into potentially hazardous territories, such as nuclear armament decision-making. But setting boundaries is only one part of the solution. Secondly, to avoid the domino effect where a malfunction in one system cascades into a larger systemic failure, introducing 'partitions' becomes essential. Much like firebreaks in forests, these partitions act as barriers, preventing the spread of dysfunction. Importantly, it's crucial to underline that these partitions should be universally applied, regardless of the perceived risk of a system. By stringently segregating AI and automated/Internet of Things (IoT) systems, we insulate them from one another, thereby dramatically minimizing the contagion risk.

---

[75] ibid

[76] ibid

[77] Candelon, F., Carlo, R.C. di, Bondt, M.D. and Evgeniou, T. (2021). AI Regulation Is Coming. [online] Harvard Business Review. Available at: https://hbr.org/2021/09/ai-regulation-is-coming.

While 'Internet of Things' may exist, the partitions mean that the 'Internet of Everything' within statutory and geographical limits should be discouraged as a state policy. Modern AI systems, akin to Complex Adaptive Systems (CAS), can exhibit unpredictable emergent behaviours. These require robust boundary mechanisms to regulate their operation. Guardrails provide a foundational layer, setting the rules of engagement, while partitions act as insulators, preventing systemic risks from spiralling out of control. The intention here is to prohibit a super connected AI System or one that is either source-to-delivery or globally integrated.

- **Boundary Setting**:
  - Establish clear, technically defined thresholds within which an AI system operates.
  - Regularly reassess and adjust these thresholds based on system performance and external environmental changes.
- **Partitioning Strategy:**
  - Implement strict separation protocols for distinct AI systems, regardless of their perceived risk levels.
  - Use containerization or virtualization techniques to encapsulate AI processes, ensuring controlled and isolated execution environments.

Ensuring Human Control through Manual 'Overrides' and 'Authorization Chokepoints'

Human agency should never be entirely overshadowed by automation, especially in domains of critical infrastructure.[78] Manual overrides empower humans as agents of intervention when AI systems behave erratically[79]. Meanwhile, multi-factor authentication authorization protocols provide robust checks before executing high-risk actions, requiring consensus from multiple credentialed humans.

Mandating manual overrides ensures that humans can intervene when AI systems begin to behave unpredictably[80]. Implementing multi-factor authentication authorization protocols, especially for high-risk actions, provides a robust check and balance system. Multiple human validators, with the requisite credentials, must give their consensus before an action is executed. This layered approach, supported by hierarchical governance, ensures that there are multiple stages and opportunities for human intervention. However, it is pivotal to note that this strategy requires equipping humans with specialized skills to effectively interface with and oversee AI. Retaining human oversight in the AI-driven world ensures that machines serve humanity's best interests. Manual overrides and chokepoints ensure that when AI systems approach ambiguity or potential harm, human judgment intervenes.

- **Manual Overrides:**
  - Identify key technical junctures within the AI processing chain where human intervention can be most effective.
  - Design and integrate manual override mechanisms at these junctures, enabling halt or redirection of AI processes.
- **Multi-Factor Authentication:**
  - Design authentication systems that require multiple human validators for high-stake decisions.
  - Ensure redundancy in validation to eliminate single points of failure.

[78] McKendrick, J. and Thurai, A. (2022). AI Isn't Ready to Make Unsupervised Decisions. [online] Harvard Business Review. Available at: https://hbr.org/2022/09/ai-isnt-ready-to-make-unsupervised-decisions.
[79] Firlej, M. and Taeihagh, A. (2021), Regulating human control over autonomous systems. Regulation & Governance, 15: 1071-1091. https://doi.org/10.1111/rego.12344

- **Hierarchical Governance:**
  - Establish a multi-tiered review process, with human decision-makers at each level equipped to assess AI decisions.
  - Provide specialized training to these decision-makers, ensuring they understand AI system behaviours and the nuances of their respective oversight roles.

Transparency and Explainability

The elusive nature of AI decisions can often make it a "black box." To mitigate this, emphasizing transparency and explainability is critical[81]. Adopting open licensing for core algorithms fosters an environment where external experts can conduct audits, enhancing system accountability. National open algorithms and data training sets would also go far in this initiative. Complementing this, "AI factsheets" can offer insights into the AI's upbringing: the training data, metrics used, potential uncertainties, and other critical parameters. This offers users a clear snapshot of the system's capabilities and limitations. Moreover, continuously monitoring and debugging black-box systems provides a real-time pulse on their operations.

Transparency is the bedrock of trust in AI systems. By demystifying AI's operational black box, we not only foster trust but also ensure accountability at all stages.

- **Open Licensing of Algorithms:**
  - Promote the use of open-source licenses for core AI algorithms to enable scrutiny by external experts.
  - Encourage industry-wide adoption of open standards to facilitate audits and cross-system comparisons.
- **AI Factsheets:**
  - Mandate the creation of standardized documentation detailing AI system upbringing: data sources, training procedures, performance metrics, and known limitations.
  - Implement a uniform format for these factsheets to ensure consistency and ease of interpretation.
- **Mandated Dynamic Monitoring:**
  - There should be periodic mandatory audits of explainability.
  - Utilize AI debugging and monitoring tools to trace system decisions in real-time.
  - Establish protocols for regular audits, ensuring continuous oversight and system accountability.

Distinct Accountability

While AI's influence permeates various sectors, it's imperative to have clear lines of responsibility[82]. In a world where technology often outpaces legal frameworks, having predefined liability protocols can bridge this gap. Whether it's a malfunction, unintended consequence, or any aberration, there should always be a clear trail leading to an entity or individual held responsible. This ensures that accountability isn't an afterthought but rather an integral part of AI deployment. This proactive stance inserts an ex-ante "Skin in the Game," ensuring that system developers and operators remain deeply invested in and accountable for AI outcomes.

---

[81] ibid

[82] Novelli, C., Taddeo, M. & Floridi, L. Accountability in artificial intelligence: what it is and how it works. AI & Soc (2023). https://doi.org/10.1007/s00146-023-01635-y

AI's pervasiveness necessitates clear liability pathways. Ensuring that responsibility is attributed correctly is fundamental to maintain public trust and ensure ethical AI deployment.

- **Predefined Liability Protocols:**
  - Draft clear legal frameworks detailing liability in case of AI system malfunctions or unintended outcomes.
  - Clearly delineate responsibilities among developers, operators, and end-users of AI systems.

- **Incident Reporting:**
  - Mandate standardized incident reporting protocols to document any system aberrations or failures.
  - Establish investigation mechanisms to delve into these incidents, ensuring lessons are learned and system robustness is improved.

Specialized, Agile Regulatory Bodies

Given AI's rapid progression, traditional regulatory methods, often mired in bureaucratic delays, might prove inadequate. Herein lies the necessity for a specialized AI regulator which has been central to global conversation on AI regulation[83,84]. This dedicated body, conceptualized as a dynamic task force, is not only equipped with the expertise but also endowed with a broad mandate. With the agility to swiftly adapt to the ever-evolving technological landscape, such a regulator can proactively identify potential pitfalls and recalibrate regulations in real-time, ensuring that governance remains both relevant and robust.

Traditional regulatory mechanisms often lag the rapid pace of AI evolution. A dedicated, agile, and expert regulatory body is pivotal to bridge this gap, ensuring that governance remains proactive and effective.

- **Broad Mandate:**
  - Endow the regulatory body with a wide-ranging mandate, allowing it to swiftly address emerging challenges without bureaucratic delays.
- **Dynamic Adaptation:**
  - Equip the regulator with tools and methodologies to continuously scan the AI horizon, identifying potential pitfalls or areas requiring regulatory attention.
  - Encourage a feedback-driven approach, where the regulator engages with industry and academia to refine its directives continually.
- **Continuous Monitoring:**
  - Employ real-time monitoring tools to track AI system behaviours against the set boundaries.
  - Integrate automated alert systems to notify of potential boundary breaches or aberrant behaviours.

As AI continues to evolve and compound in its capability, ensuring it operates within a well-defined, adaptable regulatory framework becomes paramount. By embracing the principles of CAS, we not only recognize AI's intricate, evolving nature but also craft a governance framework that is resilient, responsive, and responsible.

---

[83] Levin, B. and Downes, L. (2023). Who Is Going to Regulate AI? [online] Harvard Business Review. Available at: https://hbr.org/2023/05/who-is-going-to-regulate-ai.
[84] Finocchiaro, G. The regulation of artificial intelligence. AI & Soc (2023). https://doi.org/10.1007/s00146-023-01650-z

5. Financial Markets as a Complex Adaptive System

Financial markets exemplify intricate complex adaptive systems[85,86], with diverse human and algorithmic traders constantly interacting and reacting to emerging signals. As traders adapt strategies in response to price fluctuations, news developments and each other's actions, it creates nonlinear feedback loops and unpredictable turbulence. Emergent behaviours arising from diverse interacting agents[87,88]. Yet despite this inherent chaos, proactive governance approaches demonstrate that regulating complex market dynamics is feasible.

Dedicated regulatory bodies like SEBI (Securities and Exchange Board of India) and SEC (Securities and Exchange Commission) in the United States provide specialized oversight honed over decades of financial evolution. Equipped with broad mandates and continually engaging with market complexities, they develop tailored interventions balancing stability and progress.

Strict protocols govern automated algorithmic trading to contain risks, while still permitting innovation. Requirements of transparency through regular reporting of systemic changes, quarterly financial audits, and disclosure standards enhance traceability. This is akin to mandating algorithmic explainability and audits for artificial intelligence systems.

When volatility escalates beyond thresholds, automatic circuit breakers act as control chokepoints to temporarily halt trading and regain equilibrium. Constraints like margin requirements, settlement periods and exposure limits curb excessive speculation and leverage. Graded warnings, fines and penalties deter violations. Insolvency and bankruptcy laws provide ultimate overrides when entities become non-viable.

Fixing liability is crucial, regulations hold individuals and corporate boards personally accountable for corporate actions through codes of conduct and controls like audit committees. Imposing 'skin in the game' for market participants aligns risks and rewards. Licensing and surveillance regimes weed out misconduct.

Of course, differences exist between AI systems and financial markets – technologies like recursive self-improvement have no parallel in capital markets. However, regulating complex, emergent behaviours arising from webs of interdependencies does share useful principles. Leveraging expertise, transparency, control points and accountability. Regulation centred on expertise, transparency, control points and accountability helps manage market complexities essential for economic functioning, and has proven effective for market oversight across decades of turmoil. Adapting such insights to guide AI governance can similarly help steward progress responsibly. Building guardrails suited to AI's specific intricacies will be vital as capabilities scale exponentially.

---

[85] James, Paulin., Anisoara, Calinescu., Michael, Wooldridge. (2018). Agent-Based Modeling for Complex Financial Systems. IEEE Intelligent Systems, 33(2), 74-82. Available from: 10.1109/MIS.2018.022441352

[86] M., Flood., Alexander, Lipton., D., Quintana., Antoaneta, Serguieva. (2018). Guest Editorial Special Issue on Complex Systems in Finance and Economics. IEEE Systems Journal, 12(2), 1087-1089. Available from: 10.1109/JSYST.2018.2817978

[87] Sheri M. Markose, Computability and Evolutionary Complexity: Markets as Complex Adaptive Systems (CAS), The Economic Journal, Volume 115, Issue 504, June 2005, Pages F159–F192, https://doi.org/10.1111/j.1468-0297.2005.01000.x

[88] Polacek, G.A., Gianetto, D.A., Khashanah, K. and Verma, D. (2012), On principles and rules in complex adaptive systems: A financial system case study. Syst. Engin., 15: 433-447. https://doi.org/10.1002/sys.21213

6. Conclusion

Artificial intelligence (AI) can be considered a complex adaptive system (CAS) with unpredictable emergent behaviors arising from the interactions of multiple components like machine learning and neural networks. Just as small changes can cascade in unpredictable ways in CAS, small tweaks to AI systems can compound into disproportionate impacts that are hard to model long-term. AI development is an evolutionary process with continual interplay between variations in technical components.

Effective regulation of AI as a CAS involves instituting real-time guardrails rather than attempting to predict outcomes. Boundaries and partitions can contain risks, preventing malfunctions from spreading across systems. Retaining human oversight through chokepoints and overrides ensures human judgment intervenes when AI systems approach uncertainty or potential harm. Multi-factor authentication and hierarchical governance provide robust check and balance systems.

Emphasizing transparency and explainability of AI systems builds accountability and trust. Open licensing, national algorithms, AI factsheets, monitoring and audits by external experts demystify the "black box" nature of AI. Clear lines of responsibility and liability protocols bridge legal gaps, ensuring accountability is not an afterthought in AI deployment. Dedicated, agile regulatory bodies with specialized expertise can adapt swiftly to AI's rapid changes. A dynamic, informed regulator can scan the horizon to recalibrate regulations in real-time.

Insights from governing chaotic systems like financial markets demonstrate feasible regulation approaches for complex technologies. Dedicated oversight bodies provide specialized expertise. Requirements like audited financial statements enhance transparency, similar to explainability standards for AI. Circuit breakers act as control chokepoints, while liability and skin in the game ensure accountability.

This approach has many implications. For instance, we may never allow a super-connected internet of everything. Those creating AI tools will not be let off easily for supposed unintended consequences – thereby inserting an ex-ante "Skin in the Game". Humans will retain override and authorization powers. Regular mandated audits will have to enforce explainability.