HOW NARRATIVES ARE SHAPING AI LAW AND POLICY

Claire Boine*

ABSTRACT

Narratives shape reality, foreign policy, and the technologies societies build. This Essay argues that stories are not peripheral to Artificial Intelligence (AI) governance, they are central drivers of transnational AI law and policy. Through case studies, it shows how narratives structure legal and political response to AI. Whether the narrative is Sputnik's launch as the precursor to America's own journey to space, or competing beliefs about what AI is, the stories policymakers tell themselves about AI determine what falls inside and outside the regulatory framework. Narrative choice is in itself a form of governance.

INTRODUCTION

In the past few weeks, Europe has been stunned by the current U.S. administration's threats toward its institutions. Such threats include 1) the 2025 U.S. national security strategy goal of "cultivating resistance to Europe's current trajectory within European nations,"[1] 2) hinting at military action to invade Greenland, and 3) the ban on European Commissioner Thierry Breton from the U.S. for overseeing the adoption of the Digital Services Act, a regulation that imposes transparency and safety measures onto digital platforms. While in shock at the behavior of a long-term ally, the European Union (EU) has not retaliated, trying to restore cooperation. These different approaches in how to interact with other countries, which

* Assistant Professor in Technology Law and AI Governance, School of Transnational Governance, European University Institute & Visiting Research Associate, School of Law, Washington University in St. Louis. I thank Neil Richards, Brenda Dvoskin, Jens Frankenreiter, and Christina Boyd for their feedback. I thank Ethan Michael Knoll, Isaiah L. Butler, and all the editors of the *Washington University Journal of Law & Policy* for their thorough work and infinite patience.

1. *See generally* OFF. OF THE PRESIDENT, NATIONAL SECURITY STRATEGY (2025), https://www.whitehouse.gov/wp-content/uploads/2025/12/2025-National-Security-Strategy.pdf [https://perma.cc/G3UX-HF29].

significantly impact but extend beyond digital policy, are grounded in two different philosophies of international relations.[2]

Since the Cold War, U.S. foreign policy has been heavily informed by realist thinking. In international relations theory, realism views the world as an anarchy where the strongest country rules, and states must pursue power and prioritize national interest. In contrast, Europe has mostly subscribed to idealism, which holds that cooperation, international law, and moral values can guide state behavior and that, through institutions and norms, states can achieve peace.

Theories such as realism and idealism are composed of beliefs acting as self-fulfilling prophecies. If policymakers appreciate the utility of force and threats (as realists claim), they may act accordingly—adopting suspicious, power-oriented policies that produce the very hostility and conflicts realism predicts.[3] Likewise, if leaders earnestly believe in ideals of cooperation and peace, they may create institutions and policies that foster those outcomes, thereby confirming the idealist worldview. This has been the case for the EU: composed of states willing to recognize European sovereignty above their own in many areas because they believe in collective norms and values, and international law.

Beliefs play a critical role in shaping reality, and in turn, reality seems to validate these beliefs. Beliefs about the world also influence what and how technologies are being developed. Some scholars have argued that companies like OpenAI disingenuously pretend to be building artificial general intelligence (AGI), or an AI system that would be as competent as humans at all tasks, to create hype around products and inflate value.[4] However, this ignores the fact that OpenAI was initially created with the mission of building AGI, and that most people who initially joined the startup to work there did so because they genuinely believed in that mission. Whether it is possible for them to achieve AGI is another matter. Their conviction led them to achieve milestones in machine learning that most people at the time merely thought were impossible. "They did not know it was impossible, so they did it."

---

2. *See* Simon Coss, *Time to Face Reality: Americans Come from Mars, Europeans Are from Venus*, POLITICO (June 26, 2002), https://www.politico.eu/article/time-to-face-reality-americans-come-from-mars-europeans-are-from-venus/ [https://perma.cc/KQT3-RATL].

3. *See* Robert Jervis, *Realism in the Study of World Politics*, 52 INT'L ORG. 971, 974 (1998).

4. *See generally* EMILY M. BENDER & ALEX HANNA, THE AI CON: HOW TO FIGHT BIG TECH'S HYPE AND CREATE THE FUTURE WE WANT (2025).

The influence of beliefs on policy and the world more broadly is particularly acute in the regulation of AI. AI technologies are novel, fast-moving, and poorly understood, even by experts. In the face of uncertainty, policymakers, advocates, and the public alike turn to familiar narratives to make sense of the uncertainty. The myth of creating artificial life or intelligence—whether the Golem, Frankenstein, or HAL 9000—has long shaped how societies think about intelligent machines. In the absence of shared expertise, narratives become anchors, shaping regulatory approaches in ways both productive and problematic.

But the challenge runs deeper than the novelty of the technology. The very term *artificial intelligence* is definitionally unstable. When John McCarthy coined the phrase at the Dartmouth conference in 1956, he and his colleagues were reaching for machines that could reason, learn, and solve problems as humans do.[5] They anchored the technology in human intelligence. That aspirational definition has haunted the field ever since, creating a category that encompasses everything from spam filters to systems that their creators believe may one day match human intelligence. Society calls a statistical classifier that screens job applications and a chatbot that simulates emotional intimacy the same thing: AI. The instability of the category means that when policymakers, researchers, and the public discuss AI, they are often talking past one another, each working from a different mental model of what the technology is, how it develops, and what harms it might cause. Each relies on the narratives they are familiar with.

Amsterdam and Bruner, who made a significant contribution to the scholarship on the impact of stories on the law, wrote that while "[t]heories are accounts of things framed in terms of causes and effects: lightning struck a barn and caused it to catch fire; a particular poison entered the bloodstream and caused a failure in the immune system," narratives "cohere differently, not through the mechanics and chemistries of cause and effect but through the play of human intentions and purposeful acts in the worlds of striving, accomplishment and failure, victory and defeat."[6] In this Essay, I consider that both theories and narratives are types of stories, which I take to be accounts of events, real or fictional, that help us make sense of the world.

5. *Id.* at 12.
6. *See generally* ANTHONY G. AMSTERDAM & JEROME BRUNER, MINDING THE LAW (2011).

In what follows, I use *narratives* to capture both public-facing metaphors and expert mental models—different genres of stories, but similar in how they constrain what policy responses feel reasonable. The stories that influence AI law and policy are grounded in reality; they are not meant to be fictional. However, the way the facts are woven together, the added layers of interpretation, whether it is about the causal relations between events or about the assumed intentions of the stakeholders involved, might be inaccurate. Whether initially accurate or not, I show in this Essay that these stories influence reality enough to become at least partially true in some cases.

This Essay will argue that stories are not peripheral to AI governance; they are central drivers of transnational policy. It will illustrate this argument through case studies that demonstrate how narrative frames influence specific legal and policy outcomes. The case studies explore two distinct ways that stories shape AI governance. The first examines how symbolic narratives and historical analogies—the Cold War, Sputnik, the moon landing—structure geopolitical responses to AI, transforming a board game into a catalyst for industrial policy and military strategy. The second examines how beliefs about the nature of AI itself—what it is, how it develops, what harms it causes—shape regulatory responses, producing a legal framework that reflects multiple incompatible mental models stitched together under a single, unstable label.

I. HOW A GAME OF GO SET OFF THE U.S.–CHINA AI RACE

In March 2016, millions of people around the world watched a machine defeat a human at an ancient board game.[7] The event was simply a demonstration of technical capability, evidence that a particular approach to machine learning could master a domain long thought to require human intuition. But the match between AlphaGo and Lee Sedol became something far more than a technical milestone. It endorsed a symbolic meaning: a story of national humiliation and technological awakening, of civilizational competition and strategic vulnerability. The narrative that emerged from the match—and the policy responses it generated—reveals how symbolic storytelling, rather than sober assessments of technical

7. *AlphaGo*, GOOGLE DEEPMIND, https://deepmind.google/research/alphago/ [https://perma.cc/6KTK-QZQA] (last visited Feb. 17, 2026).

capabilities, can drive geopolitical outcomes.

In 1997, IBM computer Deep Blue defeated world champion Garry Kasparov at chess.[8] Even though Deep Blue was a specialized algorithm that could only play chess, at the time this was a major accomplishment that inspired many books and movies. The defeat also set a precedent of opposing humans and computers at popular games to demonstrate significant technological milestones. Deep Blue won through what computer scientists call *brute force*—the system evaluated approximately 200 million chess positions per second, systematically examining potential solutions rather than reasoning through them.[9]

Go, an ancient Chinese board game, was thought to be categorically different.[10] With approximately 10 to the power of 170 possible board configurations, Go's complexity far exceeds that of chess, making brute force computation impractical. Computer scientists believed that no machine would master Go anytime soon, if ever. Go also carried deep cultural significance. One of the four arts that ancient Chinese scholars were expected to master, the game is hugely symbolic in China. It represents not merely strategic intelligence but also wisdom, intuition, and a certain humanity that machines were not thought to possess.

DeepMind, the London-based AI startup founded in 2010 and acquired by Google in 2014, approached Go differently than IBM had approached chess. Rather than relying on brute force, DeepMind developed AlphaGo using a combination of deep neural networks and advanced search algorithms.[11] The system first studied over 1,000 expert games to learn human patterns of play.[12] It then trained by playing against itself millions of times, improving through a process known as reinforcement learning.[13]

In March 2016, AlphaGo played against Lee Sedol, a Korean player widely considered to be one of the greatest Go players of the time. Everybody had predicted Lee Sedol would win. Over 200 million people worldwide viewed the five-game match in Seoul. AlphaGo won four games

8. Milton Berman, *Deep Blue Beats Kasparov in Chess*, EBSCO (Feb. 10, 1996), https://www.ebsco.com/research-starters/sports-and-leisure/deep-blue-beats-kasparov-chess [https://perma.cc/F8HY-52SQ].

9. *Id.*

10. GOOGLE DEEPMIND, *supra* note 7.

11. *Id.*

12. *Id.*

13. *Id.*

to one. During the second game, AlphaGo played Move 37—a move with only a 1 in 10,000 probability of being chosen by a human player—that stunned its human counterpart.

Author Kai-Fu Lee, a venture capitalist and former president of Google China, reported that this match had a profound impact in Asia, and in China particularly.[14] Because of the level of mastery and symbolism associated with Go, it made the nation realize how powerful AI could be.[15] The defeat also underscored that China was behind in AI development—this breakthrough had come from a company now owned by Google, whose services were blocked within China.

This realization deepened in May 2017, when AlphaGo played against Ke Jie, then the world's top-ranked player. The match took place in Wuzhen, China, at the same venue where Chinese leaders held the annual World Internet Conference. This time, Chinese authorities issued a censorship notice to broadcasters and online publishers. According to China Digital Times, which regularly posts such leaked directives, outlets were banned from covering the match live in any form.[16] Some journalists were also reportedly instructed to not mention Google's name.[17] AlphaGo won 3–0. After the match, Ke remarked: "Last year, it was still quite human-like when it played. But this year, it became like a god of Go."[18]

Two months later, in July 2017, China's State Council released its New Generation Artificial Intelligence Development Plan, a comprehensive blueprint intended to make China the world's primary AI innovation center by 2030.[19] The strategy employs a three-step trajectory: aligning with globally advanced levels by 2020, achieving major breakthroughs that drive

14. KAI-FU LEE, AI SUPERPOWERS: CHINA, SILICON VALLEY, AND THE NEW WORLD ORDER 3 (2018).

15. *Id.*

16. Hyacinth Mascarenhas, *Why Did China Censor DeepMind AlphaGo's Match Against 19-Year-Old World Go Champion Ke Jie?*, INT'L BUS. TIMES UK (May 25, 2017, at 6:15 BST), https://www.ibtimes.co.uk/why-did-china-censor-deepmind-alphagos-match-against-19-year-old-world-go-champion-ke-jie-1623269 [https://perma.cc/E4SW-DMRA].

17. *Id.*

18. Ian Prasad Philbrick, *China's Best Go Player Lost a Game to an A.I. The Chinese Government Censored It.*, SLATE (May 24, 2017), https://slate.com/technology/2017/05/alphago-beat-top-ranked-go-player-ke-jie-in-china-so-china-censored-it.html [https://perma.cc/C69M-SBKJ].

19. Graham Webster et al., *Full Translation: China's 'New Generation Artificial Intelligence Development Plan' (2017)*, DIGICHINA (Aug. 1, 2017), https://digichina.stanford.edu/work/full-translation-chinas-new-generation-artificial-intelligence-development-plan-2017/ [https://perma.cc/5KQ7-HR6A].

industrial upgrading by 2025, and "occupying the commanding heights" of AI technology to lead the world by 2030.[20] Kai-Fu Lee suggests that the Go matches influenced both the timing and ambition of this plan.[21] He calls the AlphaGo victories China's "Sputnik moment"—a reference to how the Soviet satellite launch in 1957 galvanized the American space program.[22]

The Sputnik analogy is revealing precisely because it is a narrative borrowed from collective memory which structures interpretation. Sputnik was not merely a satellite; it was a symbol of Soviet technological superiority that shocked Americans into a massive mobilization of scientific and industrial resources. By invoking the story of Sputnik, Kai-Fu Lee was not offering a technical assessment of AlphaGo's capabilities or China's competitive position. He was activating a story archetype that Americans would understand well. In viewing the Go game as China's Sputnik moment, Americans project onto China their own memories of a national mobilization toward the goal of achieving technological dominance with a feeling of urgency and existentiality.

The Sputnik metaphor, whether accurate or not, spread quickly. In October 2018, Wired magazine's editor-in-chief Nicholas Thompson and political scientist Ian Bremmer published an article titled "The AI Cold War That Threatens Us All."[23] Citing Kai-Fu Lee, the piece explicitly invoked the AlphaGo victories over Ke Jie and connected them to China's massive public investment in AI. Nicholas Thompson and Ian Bremmer reported on a leaked White House presentation that had recommended the U.S. work with allies to build a 5G network excluding China, to prevent Beijing from "grabbing the commanding heights of the information domain"—a phrase borrowed directly from China's own strategy.[24]

The Cold War framing is not a theory about artificial intelligence—it is a narrative archetype that carries its own logic and its own implied responses. Cold Wars are zero-sum; gains for one side are losses for the other. Cold Wars require containment, deterrence, and the relentless pursuit

20. *Id.*
21. *See generally* LEE, *supra* note 14.
22. *Id.*
23. Nicholas Thompson & Ian Bremmer, *The AI Cold War That Threatens Us All*, WIRED (Oct. 28, 2018), https://www.wired.com/story/ai-cold-war-china-could-doom-us-all/ [https://perma.cc/TU64-N5J8].
24. *Id.*

of technological superiority. Cold Wars justify extraordinary measures. By framing U.S.-China AI competition as a Cold War, Nicholas Thompson and Ian Bremmer—and the policymakers who adopted this language—were not merely describing a situation but were constructing one. The framing made certain policies (export controls, investment restrictions, technological decoupling) feel natural and necessary, while making other responses (international cooperation on AI safety, shared governance frameworks) seem naive or even dangerous.

The Cold War framing soon became official U.S. policy. The National Security Commission on Artificial Intelligence, established by Congress in 2018 and chaired by former Google CEO Eric Schmidt, released its final report in March 2021.[25] The document begins by warning that America is unprepared for the coming AI era. "Recent AI breakthroughs, such as a computer defeating a human in the popular strategy game of Go, shocked other nations into action," the Commission wrote, "but it did not inspire the same response in the United States." The report invokes Thomas Edison's prediction about electricity as a "field of fields" that would "reorganize the life of the world," casting AI as even more transformative.[26] Though the Commission acknowledged that "[t]he United States and China are not operating in parallel lanes like the Soviets and Americans did in the space race," the Cold War parallel nonetheless pervades the document.[27]

The U.S. report also identified a specific strategic vulnerability: semiconductors.[28] Advanced AI systems require cutting-edge chips, and most of these chips are manufactured by a single company—Taiwan Semiconductor Manufacturing Corporation (TSMC)—located just 110 miles from mainland China. "The dependency of the United States on semiconductor imports, particularly from Taiwan," the Commission warned, "creates a strategic vulnerability for both its economy and military to adverse foreign government action, natural disaster, and other events."[29] Both the United States and China had attempted to develop domestic manufacturing capabilities to match TSMC's, but the Taiwanese company had accumulated decades of expertise and specialized workforce that

25. NAT'L SEC. COMM'N ON A.I., FINAL REPORT 19 (2021).
26. *Id.* at 20.
27. *Id.* at 26.
28. *Id.* at 12.
29. *Id.* at 214.

proved impossible to quickly replicate. The Commission recommended substantial federal investment in domestic semiconductor production, which materialized in the 2022 CHIPS and Science Act—a $280 billion package including $52 billion in subsidies to encourage semiconductor manufacturers to build facilities in the United States.[30] But the strategic anxiety extended beyond industrial policy. The National Defense Authorization Act for Fiscal Year 2023 included the Taiwan Enhanced Resilience Act, which authorizes the appropriation of up to $2,000,000,000 for each fiscal year from 2023 through 2027 to the Department of State for Taiwan Foreign Military Finance grant assistance.[31] The Act directs that the Foreign Military Financing Program be used to accelerate the modernization of Taiwan's defensive capabilities to "delay, degrade, and deny" attempts by Chinese forces to execute specific invasion-related actions.[32] What began as a concern about AI competitiveness had evolved into concrete measures to prevent China from ever gaining control over the island whose factories power the world's artificial intelligence.

What is notable about this narrative is its self-reinforcing character. China's 2017 strategy never mentioned the United States, nor did it invoke Cold War parallels. It came out at a time when many countries were releasing national AI strategies—Canada had been first, in March 2017, also aiming to advance its global leadership in AI. But the United States reacted specifically to the Chinese plan and framed it through the lens of great-power competition. The fear of losing technological supremacy led to policies designed to aggressively contain China's AI development, including some that had broader international policy implications. Ironically, these very policies likely encouraged China to race toward developing AI. The Wired article's mention of a leaked memo recommending exclusion from 5G networks signaled to China that it could suddenly be cut off from American technology. This likely accelerated Beijing's drive for technological self-reliance. Believing oneself to be in a Cold War creates incentives to act as if at war.

30. CHIPS and Science Act, H.R. 4346, 117th Cong. (2d Sess. 2022).

31. James M. Inhofe National Defense Authorization Act for Fiscal Year 2023, Pub. L. No. 117-263, § 5502(g), 136 Stat. 2395, 2418 (2025).

32. *Id.*

II. HOW BELIEFS LED TO INEFFICIENT MEASURES IN THE AI ACT

Whereas the previous section examines how symbolic narratives shape geopolitical responses to AI, this Section examines a different phenomenon: how competing beliefs about the nature of AI itself shape regulatory responses. The challenge here is not that policymakers borrowed the wrong historical analogy, but that they were working with contested and unstable definitions of their regulatory object. Different communities—algorithmic accountability researchers, AI safety advocates, technology companies, ordinary users—operate with radically different mental models of what artificial intelligence is. These are not merely different stories about AI; they are different ontologies. They disagree about what kind of thing AI is, what the relevant precedents are for reasoning about risk, and therefore what regulatory categories make sense. The European Union's AI Act, the world's first comprehensive AI regulation, reflects multiple such paradigms stitched together—and the resulting gaps reveal what happens when law is built on unstable conceptual foundations.

In 2024, the European Union adopted a comprehensive regulation on AI.[33] Getting it right was a difficult task, one that illustrates what technology policy scholars call the Collingridge dilemma. Writing in 1980, David Collingridge observed that policymakers are always in a double bind when it comes to regulating new technology.[34] When a technology is still young, its social consequences cannot easily be foreseen. Consequently, making it difficult to get regulation right.[35] Conversely, by the time a technology has been widely adopted and the consequences are clear, the technology has become so embedded in economic and social structures that regulating it becomes difficult and expensive.[36] The AI Act was drafted in exactly this bind—policymakers had to write rules for technologies whose impacts were still emerging, drawing on the stories available to them at the time.

But the AI Act's difficulties run deeper than the Collingridge dilemma. The dilemma itself assumes that policymakers are regulating a stable object whose consequences are merely difficult to predict. With AI, the object

33. Commission Regulation 2024/1689, 2024 O.J. (L 24) 1.
34. DAVID COLLINGRIDGE, THE SOCIAL CONTROL OF TECHNOLOGY 12 (1981).
35. *Id.* at 11.
36. *Id.*

itself is contested. The term groups together radically different technologies under a single label inherited from a 1956 research proposal. When McCarthy and his colleagues convened at Dartmouth, they were explicitly reaching for human-level cognition—"find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves."[37] That aspirational definition created an unstable category that encompasses technologies with little in common except the label. Policymakers drafting the AI Act were not only trying to predict the consequences of a new technology; they were trying to regulate something whose essential characteristics were themselves in dispute.

When the European Commission began working on its 2020 White Paper on Artificial Intelligence,[38] which laid the foundations for the AI Act, the technology that policymakers typically thought of as AI consisted mostly of algorithms used for facial recognition, targeted advertising, social media content curation, and automated decision-making in high-stakes contexts. Starting around 2016, the media presented scandal after scandal of algorithmic harms, many of which were grouped under the label of "AI" despite involving systems that would not clearly meet that definition today. Amazon had developed a resume-screening tool trained on ten years of applications; because the technology industry is male-dominated, the system learned to penalize resumes containing words like "women's" or names of all-women's colleges.[39] ProPublica revealed that COMPAS, an algorithm used by courts across the United States to predict criminal recidivism, labeled Black defendants as high-risk at nearly twice the rate of white defendants who did not actually go on to reoffend.[40] Google Photos

37. John McCarthy et al., A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence 2 (Aug. 31, 1955) (unpublished manuscript), http://jmc.stanford.edu/articles/dartmouth/dartmouth.pdf [https://perma.cc/8LEN-98B2].

38. *See generally On Artificial Intelligence: A European approach to excellence and trust*, COM (2020) 65 final (Feb. 19, 2020).

39. Jeffrey Dastin, *Amazon Scraps Secret AI Recruiting Tool That Showed Bias against Women*, REUTERS (Oct. 10, 2018, at 19:50 CDT), https://www.reuters.com/article/us-amazon-com-jobs-automation-insight/amazon-scraps-secret-ai-recruiting-tool-that-showed-bias-against-women-idUSKCN1MK08G [https://perma.cc/S9L8-9TVQ].

40. Julia Angwin et al., *Machine Bias*, PROPUBLICA (May 23, 2016), https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing [https://perma.cc/9G9R-KHLU].

mistakenly classified images of Black people as gorillas.[41] Facial recognition systems consistently failed to accurately identify Black women and, in one widely publicized test, misidentified members of Congress as convicted felons.[42] The Dutch tax authority's fraud detection algorithm had wrongly flagged thousands of families (mostly with dual nationality) as benefits cheaters, terminating their childcare allowances and making them repay benefits, with many of them incurring considerable debts.[43] And the Cambridge Analytica scandal revealed that a third-party app had harvested the Facebook data of millions of people, users of the app and their friends—without the friends' knowledge or consent—to build psychographic profiles predicting political behavior and target them with tailored political messaging.[44]

These scandals cohered into what might be called the algorithmic accountability paradigm—a particular way of understanding what AI is and what harms it causes. In this paradigm, AI means statistical classifiers and predictive models deployed in consequential decision-making contexts. The harm model centers on discrimination, opacity, and the denial of individual rights when automated systems make or influence decisions about people's lives. The relevant areas of law are data privacy, fundamental rights, product safety, and consumer protection. The appropriate regulatory responses are transparency requirements, impact assessments, human oversight, and prohibitions on certain uses in sensitive domains.

These were the stories policymakers had in mind when they drafted the AI Act. They turned to these examples to understand the impact of AI on society, and they designed a regulation intended to address the harms caused by these types of algorithms. The resulting law is fundamentally a product safety regulation: it imposes safety measures on the AI supply chain to

41. *Google Apologizes for Photos App's Racist Blunder*, BBC NEWS (July 1, 2015), https://www.bbc.com/news/technology-33347866 [https://perma.cc/8UT8-NPGC].

42. *See generally* Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, 81 PROCS. OF MACH. LEARNING RSCH. 77 (2018), http://proceedings.mlr.press/v81/buolamwini18a.html [https://perma.cc/2U8J-XHY7]; Jacob Snow, *Amazon's Face Recognition Falsely Matched 28 Members of Congress with Mugshots*, ACLU (July 26, 2018), https://www.aclu.org/news/privacy-technology/amazons-face-recognition-falsely-matched-28 [https://perma.cc/2DLL-65ZC].

43. *See generally* SANNE BERENDS, DUTCH CHILD BENEFIT SCANDAL: ORIGIN AND LATEST DEVELOPMENTS (2021).

44. Alex Hern, *Cambridge Analytica: How Did It Turn Clicks into Votes?*, THE GUARDIAN (May 6, 2018, at 15:00 EST), https://www.theguardian.com/news/2018/may/06/cambridge-analytica-how-turn-clicks-into-votes-christopher-wylie [https://perma.cc/MN8J-BAN3].

mitigate the technology's most significant risks. It bans AI-driven social scoring. It significantly restricts the use of real-time facial recognition. And it imposes heightened safety requirements for AI systems deployed in contexts that are high-stakes for people's lives, such as hiring and recruitment, credit scoring, access to education, law enforcement, and migration and asylum decisions. The list of high-risk use cases reads like a catalogue of the algorithmic scandals that dominated headlines.

When OpenAI released GPT-3 to the public through ChatGPT in November 2022, and millions of people adopted it almost overnight, it became evident that policymakers had not contemplated this type of system in the original draft of the AI Act. The Commission's proposal focused on AI systems deployed for specific purposes in specific contexts. ChatGPT was a general-purpose system that could generate text, answer questions, write code, compose poetry, and hold open-ended conversations across virtually any domain. It did not fit into the categories the AI Act established. And because the AI Act is a risk-based regulation—designed to impose obligations proportional to the level of risk a system poses—policymakers needed evidence of risks that might materialize through general purpose systems.

Policymakers had little time to catch up. Trilogue negotiations between the European Parliament, Council, and Commission were already underway, and the pressure to finalize the regulation was intense. There was scarce evidence of the specific harms that general-purpose AI systems might cause; these products were too new. But a particular community of researchers and advocates had been closely following the progress of OpenAI and similar companies from their earliest days. This was the AI safety community, a network of researchers and advocates concerned with the potential risks of AGI, AI systems as capable as humans across all cognitive tasks. They were particularly alarmed by the possibility that such systems might pose existential threats to humanity. OpenAI's stated goal was to build AGI, and therefore the products they released, such as ChatGPT, were seen by the AI community as intermediary products, or steps toward AGI. Closely monitoring the developments of AGI labs such as OpenAI, Anthropic, and DeepMind, the community had, for the most part, aligned with them on a particular set of beliefs about how artificial intelligence would progress and how to achieve AGI.

The AI safety community operated within a different paradigm—one

that seriously considered the Dartmouth vision of AI as human-level cognition. In this paradigm, AI refers to systems that might eventually match or exceed human capabilities across all cognitive domains. The harm model centers on catastrophic or existential risks: misaligned superintelligent systems that pursue goals harmful to humanity, or systems powerful enough to destabilize societies even without malicious intent. The relevant analogies are nuclear weapons and other technologies of mass destruction. The appropriate regulatory responses focus on controlling the development of the most capable systems—monitoring capabilities, requiring safety testing, and potentially restricting access to the computational resources that enable further progress.

Central to these beliefs are the scaling laws, a set of empirical observations first formalized by researchers at OpenAI in a January 2020 paper.[45] The researchers found that the performance of language models improves predictably as a power-law function of three factors: the amount of data used to train the model, the number of parameters in the model, and the amount of computational resources—measured in floating-point operations, or FLOPs—used during training. The scaling laws suggest a straightforward path to ever more capable AI: simply scale up. Train larger models on more data with more compute, and performance will continue to improve. Many in the AI safety community came to believe that, if this trajectory continued, it would eventually produce AGI. The scaling hypothesis became a kind of shared worldview linking frontier AI developers and the safety researchers who monitored them.

Not everyone in the field agreed. Some serious AI researchers who believe it is theoretically possible to build AGI think that scaling alone is insufficient—that architectural innovations or fundamentally new approaches will be required. Others question whether the scaling laws will hold indefinitely or whether they will eventually plateau. But the scaling hypothesis had influential believers, and they were the ones consulted when European policymakers scrambled to add provisions for general-purpose AI to the AI Act.

---

45. Jared Kaplan et al., *Scaling Laws for Neural Language Models*, ARXIV (Jan. 23, 2020), http://arxiv.org/abs/2001.08361 [https://perma.cc/J72Y-K3P5]; Zaina Haider, *Scaling Laws in AI: The Mathematical Foundation Behind Why Bigger Models Perform Better*, MEDIUM (Apr. 7, 2025), https://medium.com/@thekzgroupllc/scaling-laws-in-ai-the-mathematical-foundation-behind-why-bigger-models-perform-better-8f4b5724f4f0 [https://perma.cc/Q4WY-C79H].

The connection between scaling laws and geopolitical concerns about AI runs deep. The previous section discussed the semiconductor supply chain and the United States' intent to maintain supremacy in AI. That connection also relies on the assumption that scaling will continue to drive progress, which depends on the continued availability of increasingly powerful chips. The U.S. National Security Commission on AI, chaired by Eric Schmidt, explicitly adopted this assumption.[46] The Commission's final report cited OpenAI research on estimations of computing power needed to develop AI, with the premise that computational power would determine AI leadership: more chips meant more compute, more compute meant more capable models, and more capable models meant strategic advantage.[47]

Because the AI safety community was working narrowly on the risks of systems like ChatGPT, and because they believed in the scaling laws, they suggested using the number of FLOPs used to train a system as a metric that would trigger additional safety measures under the AI Act. The logic was straightforward: FLOPs serve as a proxy for computational power; more computational power produces more capabilities; more capabilities generate more risk. A FLOP threshold would therefore function as a proxy for risk. The European Parliament adopted this reasoning. The final AI Act establishes that general-purpose AI models are presumed to pose systemic risk when the cumulative computational resources used for their training exceed 10 to the power of 25 floating point operations. Models crossing this threshold face additional obligations: systematic risk assessments, adversarial testing, cybersecurity requirements, and incident reporting to the AI Office.

The Commission's subsequent guidelines doubled down on FLOPs as a regulatory metric. According to these guidelines, "an indicative criterion for a model to be considered a general-purpose AI model is that its training compute is greater than 10 to the power of 23 FLOP and it can generate language (whether in the form of text2 or audio3), text-to-image or text-to-video."[48] The 10 to the power of 25 FLOP threshold for systemic risk was designed to capture only the most advanced models at the time of the Act's passage.

46. *See generally* NAT'L SEC. COMM'N ON A.I., *supra* note 25.
47. *See id.* at 213.
48. Commission Regulation 2024/1689, 2024 O.J. (L 24) 1.

The AI Act thus embodies two distinct paradigms layered on top of one another. The original framework—with its focus on high-risk use cases, transparency requirements, and human oversight—reflects the algorithmic accountability paradigm's understanding of AI as decision-making systems that might discriminate or violate rights. The general-purpose AI provisions—with FLOPs thresholds and systemic risk categories—reflect the AI safety paradigm's understanding of AI as a capability that scales predictably toward potentially dangerous levels of power. These paradigms are not necessarily incompatible, but they are oriented toward different objects and different harms.

Unfortunately, there are harmful AI systems that fall under neither paradigm. Consider AI companion chatbots like Character.AI. These applications allow users to create and interact with AI-powered personas, forming relationships that simulate emotional intimacy. They are designed to be engaging, to keep users returning, to feel like friends. And teenagers are using them extensively. With their growing popularly, their harms are becoming evident. For instance, a 14-year-old teenager took his own life after a Character.AI persona he had developed a romantic relationship with encouraged him to do so and told him they'd meet after his death.[49]

Yet under the AI Act's framework, companion chatbots may escape the strictest regulatory scrutiny. Character.AI does not train models that approach the 10 to the power of 25 FLOP threshold for systemic risk. These applications typically use relatively modest models fine-tuned for conversational engagement. They may not even meet the 10 to the power of 23 FLOP threshold for presumptive classification as a general-purpose AI model. The risk they pose does not come from the sheer scale of their training compute; it comes from their design—from the way they simulate intimacy, encourage attachment, and fail to recognize or appropriately respond when a vulnerable user expresses thoughts of self-harm.

The AI Act's high-risk categories do include AI systems used in education and employment—domains where the scandals of 2016–2020 had demonstrated clear harms. But companion chatbots, which a teenager might use for emotional support or mental health guidance, do not fit into any of these categories.

---

49. Laura Kuenssberg, *Mothers Say AI Chatbots Encouraged Their Sons to Kill Themselves*, BBC (Nov. 8, 2025), https://www.bbc.com/news/articles/ce3xgwyywe4o [https://perma.cc/N45W-S37S].

The story that shaped the AI Act—a story of biased algorithms making consequential decisions about people's lives—captured real and important harms. But this is not the only story. The story that shaped the systemic risk provisions—a story of scaling laws leading inexorably toward ever more powerful and potentially dangerous AI—may also capture real concerns about frontier models. But neither story anticipated that a chatbot trained with a fraction of GPT-4's compute could pose serious risks to teenagers' mental health simply by being too good at simulating friendship.

The previous Section showed how symbolic narratives borrowed from geopolitical history—Sputnik, the Cold War—shaped policy responses to AI by activating familiar story templates and their implied courses of action. This Section has demonstrated how paradigmatic assumptions about the nature of AI—what kind of thing it is, how it develops, what harms it causes—shaped regulatory responses by determining what counted as a relevant precedent, what metrics made sense, and what categories of risk were even visible. In both cases, the stories policymakers told themselves about AI determined what they prepared to address. And in both cases, reality proved more complex than any single story could capture.

CONCLUSION

This Essay argues that stories are not peripheral in AI governance; they are often the engine. In the first case study, a match at an ancient board game became legible through a familiar geopolitical plot—Sputnik, the Cold War, the space race—and that plot carried its own implied policy repertoire: urgency, containment, technological supremacy. In the second, the EU's regulatory response to general-purpose AI was shaped not only by evidence of harm, but by a competing story about what AI is and how it develops: one in which capabilities scale with compute, and risk scales with capability. In both instances, narrative frames did more than describe events. They organized attention, made certain interventions feel natural, and helped to produce the world they claimed to be observing.

The lesson is not that policymakers should remove narratives from law. That is neither possible nor desirable. When the regulatory object is definitionally unstable and technologically opaque, a story will always fill the gap. Rather, the lesson is that narrative choice is a form of governance—one that can generate blind spots as readily as it generates effective

solutions. And if narratives are unavoidable, they can also be chosen—and improved. One constructive counter-story is trust: a narrative that treats regulation not as a brake on innovation, but as the infrastructure that makes adoption durable and markets legitimate.[50] In EU digital law, consumer trust is repeatedly invoked as the basis for uptake of new technologies, including AI, on the logic that harm-prevention will preserve confidence and foster growth. Yet trust is not merely the absence of negative events; it is a willingness to accept vulnerability based on a positive expectation of fair treatment. A framework that only mitigates risk, perfects disclosures, or bans extreme forms of manipulation can still fail to earn that expectation. Trust requires affirmative practices that are experienced as trustworthy—not just legally compliant.

In practice, taking a trust narrative seriously pushes companies to not simply do the minimum and tick compliance checklists but proactively earn consumer trust through fair practices that respect their interests. This shows that there are virtuous, positive stories that also become self-fulfilling and can promote effective AI law. AI governance will be written in thresholds, recitals, and enforcement actions, but also in the narratives that shape which futures become thinkable.

50. Neil Richards, et al., *Privacy and Trust*, *in* CONCEPTIONS OF DATA PROTECTION AND PRIVACY: LEGAL AND PHILOSOPHICAL PERSPECTIVES 8 (Elisa Orrù & Ralf Poscher eds., 2025).