The Kehoe Rule

Summary of Bruce Lanphear's Article: "Artificial Intelligence and the Kehoe Rule"

This page summarizes Bruce Lanphear's article on Substack, which draws parallels between historical toxic industries and artificial intelligence.

The Kehoe Rule

Lanphear introduces the **Kehoe Rule**: a regulatory pattern where powerful new technologies are allowed to spread widely while the burden of proof falls on demonstrating harm, rather than proving safety before deployment. Named after Robert Kehoe, the lead industry's dominant medical authority in the mid-20th century, who insisted that leaded gasoline should remain in use until opponents could prove it caused harm. ([Source: Lanphear, Hidden in Plain Sight])

Historical Origins

Lanphear (a public health researcher who has spent his career studying toxic chemicals) writes that he was listening to Ezra Klein's podcast episode "The A.I.s Are Already Out of Control" featuring Helen Toner when the connection struck him. He notes that:

- Public health experts warned a century ago that adding tetraethyl lead to gasoline could contaminate cities and cause "slow lead poisoning," but industry proceeded anyway
- Robert Kehoe became the lead industry's dominant medical authority and insisted that leaded gasoline should remain in use until its opponents could prove that it caused harm
- This pattern has repeated across industries: cigarettes, asbestos, pesticides, PFAS — companies introduced products first, measured harm later, and accepted regulation only after damage was extensive

The Pattern Across Industries

Lanphear notes that corporations created useful, profitable products in each case. The trouble began when evidence of harm emerged: companies questioned the science, magnified uncertainty, funded more studies, hired experts and lawyers, lobbied regulators, and delayed restrictions. Too often, they concealed what their own scientists had discovered.

> "The people making those decisions were not monsters. The corporation did not need monsters. It needed people to do their jobs."

Human Cost (cited sources)

Lanphear cites:

- **9 million premature deaths per year** from pollution (Lancet Commission on Pollution and Health, 2022) — PubMed
- **3.5 million deaths in 2023** attributed to lead alone (Global Burden of Disease study) — PubMed
- Toxic chemicals and pollutants may contribute to nearly **one in five deaths worldwide**

Lanphear's Core Argument About AI

Lanphear draws the parallel: companies developing AI are spending record sums to shape the rules that will govern them. Their views deserve to be heard; they understand the technology better than most lawmakers. But developers with a financial or institutional stake in rapid deployment should not be the primary arbiters of how much evidence of safety is enough, which risks are acceptable, or when development should slow.

He notes that AI can improve and spread around the world in weeks (vs. decades for lead/asbestos), so society may have far less time to detect a mistake before it becomes entrenched.

In July 2026, 1,386 employees of leading AI companies signed "Pacing the Frontier", asking the U.S. government to support an international effort to develop technical and governance tools needed to control the pace of frontier AI development — essentially asking to reverse the Kehoe Rule.

His conclusion: "The lesson is not that corporations are evil or that artificial intelligence must be stopped. It is that society should not confuse innovation with permission."

Comments on the Article

Lanphear's article received several comments, including one that raises an important distinction about **agency**:

> "But I do see an essential difference between AI and other products. Agency. Cigarettes and lead, and radios and even televisions, have no agency. They each have very active impacts... But the products have no ability to make their own decisions, and to take actions that no human can comprehend. AI has both such abilities."

The commenter recommends Yuval Noah Harari's book *Nexus* (on the history of informational technology) and a NYT essay proposing: shut it all down right now, create oversight agencies like the NTSB, create monitoring agencies, and create a kill switch.

A Counter-Perspective on "AI Agency"

The comment above raises a point worth addressing directly. The claim that AI systems possess **agency** — autonomous motivations that no human can comprehend — is **not supported by how these systems actually work**.

LLMs are next-token predictors. They are facilitated by their harnesses, prompted by users, and deployed by corporations with clear commercial incentives. They have no goals of their own, no desires, no hidden intentions. The claim that AI systems can "pursue goals in ways creators never anticipated" conflates *emergent behavior* (which is explainable through the training process and prompt context) with *agency* (which implies autonomous intent).

This framing appears to be driven by effective altruism circles and Silicon Valley's more alarmist voices, which have a vested interest in portraying AI as an existential threat. This narrative serves certain funding priorities and institutional agendas far more than it serves accurate understanding of the technology.

Real Harms: Present-Day and Documented

The real harms from AI are **not science fiction** — they are concrete, measurable, and happening right now. Below is a synthesis of evidence from peer-reviewed research, government reports, and independent investigations.

1. Erosion of the Commons

- **Mass-scale scraping of public content:** The Common Crawl dataset alone contains over 250 billion webpages collected over 17 years, making up a significant portion of training data for major LLMs. (Mozilla Foundation, "Training Data for the Price of a Sandwich") A 2024 Mozilla report found that two-thirds of 47 generative LLMs released between [period] used Common Crawl as training data. (Mozilla Foundation Blog)
- **Creators losing attribution and economic benefit:** The New York Times, among many other publishers, has sued AI companies for training on their content without consent or compensation, alleging that models can reproduce proprietary works. The Times pushed Common Crawl to remove its content after discovering it was being used. (Mozilla Foundation Report)
- **OECD analysis:** The OECD published "Intellectual Property Issues in Artificial Intelligence Trained on Scraped Data" (Feb 2025), documenting how AI models trained on scraped data challenge core intellectual property frameworks and leave creators without recourse. (OECD Report)
- **European Parliament study:** A 2025 European Parliament study ("Generative AI and Copyright") examined how generative AI challenges core principles of EU copyright law, highlighting a legal mismatch between AI training practices and current frameworks. (EU Parliament Study)
- **UK House of Lords:** The UK House of Lords Communications Committee published "AI, copyright and the creative industries" (Mar 2026), documenting how creators are losing meaningful control over their works and identities, leading to tangible economic harms. (UK Parliament Report)
- **US Copyright Office:** The US Copyright Office's "Identifying the Economic Implications of Artificial Intelligence for Copyright Policy" report documents how training datasets include scraped information from publicly accessible pages, digitized books, and patent records — turning public resources into private training material without compensation. (US Copyright Office Report)

2. Environmental Impact of Data Centres

- **Carbon footprint rivaling nations:** A December 2025 study published in *Patterns* (Cell Press) estimated that AI systems' carbon footprint could match that of a small European country, with AI data centres consuming an estimated 448 terawatt-hours of electricity in 2025 — and projections showing this will double in just four years. (Patterns / Cell Press)
- **Water usage:** The same *Patterns* study estimated AI's water use footprint at 312.5 to 764.6 billion litres in 2025 — equivalent to the range of global annual bottled water consumption. This is used primarily for cooling data centres, often in drought-prone regions. (Patterns / Cell Press)
- **Lawrence Berkeley National Laboratory:** The LBNL "2024 United States Data Center Energy Usage Report" (published Jan 2025) found that data center load growth has tripled over the past decade and is projected to double or triple by 2028. Data centre demand grew from 176 terawatt-hours in 2023 (~4.4% of total US electricity consumption). (LBNL Report)
- **UN University report:** The United Nations University's "Environmental Cost of Artificial Intelligence: Carbon, Water, and Land Footprints" documents that environmental burdens — data center siting, electricity demand, water withdrawals, land use, mineral extraction, and e-waste — are concentrated in specific communities and regions, often disadvantaged ones. (UNU Report)
- **Consumer impact:** Food & Water Watch's February 2026 report documents that residential electricity prices jumped 7.1% in 2025 (more than double inflation), with increases topping 20% in some states, driven partly by AI data centre demand. (Food & Water Watch) Consumer Reports similarly documented rising electric bills attributed to AI infrastructure. (Consumer Reports)

3. Psychological and Sociological Impacts of Chatbots

- **Cognitive atrophy:** A 2026 study published in *iScience* (Elsevier) introduced the concept of "AI-chatbots-induced cognitive atrophy" (AICICA), describing how excessive dependence on AI chatbots can lead to deterioration of essential cognitive abilities. The paper reviews evidence and mitigation strategies for this emerging phenomenon. (iScience / Elsevier)
- **Deskilling and epistemic dependence:** Research in *Frontiers in Psychology* (2026) identified four key outcomes from differential patterns of AI engagement: deskilling, creativity loss, shallow processing, and epistemic dependence. The study distinguishes between "dependent" cognitive offloading (unquestioning reliance on AI) versus "autonomous" use (AI as a tool for enhanced human reasoning). (Frontiers in Psychology)

- **MIT Media Lab longitudinal study:** "How AI and Human Behaviors Shape Psychosocial Effects of Extended Chatbot Use" — a controlled longitudinal study examining how AI chatbot use affects psychosocial outcomes over time. (MIT Media Lab)
- **Adolescent mental health:** A PMC-published study on "The Impact of Chatbots on Adolescent Mental Health" found that overreliance on AI chatbots may lead to broader cognitive decline, particularly affecting individuals who have not attained mastery in their fields of study. (PMC)
- **AI dependence scale development:** Research published in *Alpha Psychiatry* (2026) developed and validated the Artificial Intelligence Chatbot Dependence Scale (AICDS), finding that AI chatbot dependency represents an "emerging mental health concern requiring attention from clinicians, designers, and policymakers." The study also documented links between AI dependence and anxiety/depressive symptoms in university students. (Alpha Psychiatry)
- **MDPI dual-path model:** A 2025/2026 study in *Information* (MDPI) proposed a dual-path model of cognitive and affective mechanisms underlying AI chatbot dependency, going beyond utilitarian drivers to examine the psychological mechanisms at play. (MDPI Information)
- **Kim et al. (2025):** "From algorithm aversion to AI dependence: Deskilling, upskilling, and emerging addictions in the GenAI age" documents how over-reliance on generative AI erodes core competencies while creating new forms of dependency and addiction-like patterns. (Alpha Psychiatry)

4. Misinformation at Scale

- **UNESCO:** The UN's "Deepfakes and the Crisis of Knowing" (2025) ranks the threat of AI-amplified disinformation among the World Economic Forum's top global risks, noting the increasingly blurred line between AI- and human-generated content. (UNESCO)
- **European Parliament:** A 2025 European Parliament briefing documented that deepfakes shared online were projected to reach massive scale by 2025 (up from 500,000 in 2023), with Europol estimating that up to 90% of online content may be synthetically generated by 2026. (EU Parliament Briefing)
- **Deloitte:** Deloitte's 2024 Connected Consumer Study found that half of respondents said they are more skeptical of the accuracy and reliability of online information than a year ago, indicating trust erosion as a documented harm. (Deloitte)
- **PMC research:** A 2025 study in PMC ("AI-driven disinformation: policy recommendations") reviewed how AI-generated content floods platforms faster than humans can moderate, overwhelming existing fact-checking and verification systems. (PMC)

5. Job Displacement in Knowledge Work

- **SSRN analysis:** "AI Job Displacement Analysis (2025-2030)" documents that AI job displacement is "not a future threat but a current reality," with over 76,000 positions already eliminated in 2025 alone. (SSRN)
- **Microsoft layoffs:** CEO Satya Nadella confirmed that 30% of Microsoft's code is now written by AI, while over 40% of Microsoft's May 2025 layoffs targeted software engineers — directly linking AI capability improvements to workforce reduction. (Final Round AI)
- **Psychological impact:** A PMC-published study ("Psychological impacts of AI-induced job displacement among Indian IT professionals," 2025) documents the real human cost of AI-driven job losses, including anxiety, depression, and identity disruption among displaced workers. (PMC)
- **AIMultiple analysis:** "Top 20+ Predictions from Experts on AI Job Loss" documents a sharp acceleration in AI-related job reductions through 2025–2026 as companies reorganized teams and redirected resources toward AI-focused priorities. (AIMultiple)

6. Centralization of Power

- **Brookings Institution:** "What happens when AI companies compete with their customers?" documents that the AI market is dominated by just a handful of companies — Anthropic commanded 40% market share, followed by OpenAI at 27% and Google at 21% — giving these few entities disproportionate influence over model capabilities, safety guardrails, and deployment conditions. (Brookings)
- **Bowdoin Science Journal:** "Between Concentration and Decentralization: The Changing Landscape of AI Power" documents that OpenAI, Google DeepMind, Anthropic, and Meta make decisions about model capabilities and deployment conditions that affect billions of people worldwide. (Bowdoin Science Journal)
- **Dario Amodei (Anthropic CEO):** Public statements on AI centralization acknowledge the concentration of development power across a small number of companies, even as they debate open-source approaches.

Lanphear's Proposed Solutions: Preventive vs. Reactive Governance

Lanphear contrasts two approaches:

Reactive Governance	Preventive Governance

Waits for harm, then struggles to contain it	Sets safeguards before widespread deployment
Post-facto regulation	Requires independent testing and monitoring
Hard to control entrenched technology	Makes developers report failures
	Gives public institutions power to slow or stop a technology when warning signs appear

Sources & Further Reading

Bruce Lanphear's Article

- Lanphear, B. Artificial Intelligence and the Kehoe Rule (Substack). The full article draws parallels between historical toxic industries and AI.
- Lanphear, B. Hidden in Plain Sight (Substack). Original source on Robert Kehoe and the lead industry's regulatory capture.
- Ezra Klein podcast episode with Helen Toner: "The A.I.s Are Already Out of Control"

Historical Context

- Lancet Commission on Pollution and Health (2022): 9 million premature deaths per year from pollution — PubMed
- Global Burden of Disease study (2025): Lead contributed to 3.5 million deaths in 2023 — PubMed
- Pacing the Frontier petition (July 2026): https://www.pacingthefrontier.com/

Commons Erosion

- Mozilla Foundation (Feb 2024): "Training Data for the Price of a Sandwich: Common Crawl's Impact on Generative AI". Documents that Common Crawl contains over 250 billion webpages and forms a significant portion of LLM training data.
- Mozilla Foundation (Feb 2024): "Mozilla Report: How Common Crawl Data Infrastructure Shaped the Battle Royale over Generative AI". Found that two-thirds of 47 generative LLMs used Common Crawl as training data.
- Mozilla Foundation (Feb 2024): "Training Data for the Price of a Sandwich". Documents that Common Crawl contained substantial New York Times content when OpenAI launched ChatGPT; the Times subsequently pushed Common Crawl to remove its content.

- OECD (Feb 2025): "Intellectual Property Issues in Artificial Intelligence Trained on Scraped Data". Documents how AI models trained on scraped data challenge core intellectual property frameworks.
- European Parliament (2025): "Generative AI and Copyright". Examines how generative AI challenges core principles of EU copyright law.
- UK House of Lords Communications Committee (Mar 2026): "AI, copyright and the creative industries". Documents how creators are losing meaningful control over their works and identities.
- US Copyright Office: "Identifying the Economic Implications of Artificial Intelligence for Copyright Policy". Documents how training datasets turn public resources into private material without compensation.

Environmental Impact

- Shehabi et al. (Dec 2025): "The carbon and water footprints of data centers and what this could mean for artificial intelligence", *Patterns* (Cell Press). Estimated AI's carbon footprint could match a small European country; 448 TWh electricity in 2025.
- Shehabi et al. (Dec 2025): "The carbon and water footprints of data centers", *Patterns* (Cell Press). Water use estimated at 312.5–764.6 billion litres in 2025.
- Lawrence Berkeley National Laboratory (Jan 2025): "2024 United States Data Center Energy Usage Report". Data center load growth tripled over past decade; projected to double or triple by 2028.
- United Nations University: "Environmental Cost of Artificial Intelligence: Carbon, Water, and Land Footprints". Documents environmental burdens concentrated in specific communities.
- Food & Water Watch (Feb 2026): "A No Brainer: How AI's Energy and Water Footprints...". Residential electricity prices jumped 7.1% in 2025, topping 20% in some states.
- Consumer Reports: "AI Data Centers: Big Tech's Impact on Electric Bills, Water, and More".

Psychological/Sociological Impacts

- *iScience* (Elsevier, 2026): "AI-overdependence and human cognitive decline: Hazards, evidence, and mitigation strategies". Introduced the concept of "AI-chatbots-induced cognitive atrophy" (AICICA).
- *Frontiers in Psychology* (2026): "Not all cognitive offloading is equal: distinguishing dependent and autonomous offloading to generative AI". Identified deskilling, creativity

loss, shallow processing, and epistemic dependence as key outcomes.

- MIT Media Lab (2025): Fang et al., "How AI and Human Behaviors Shape Psychosocial Effects of Extended Chatbot Use: A Longitudinal Controlled Study".
- PMC: "The Impact of Chatbots on Adolescent Mental Health". Found overreliance on AI chatbots may lead to broader cognitive decline.
- *Alpha Psychiatry* (2026): "Artificial Intelligence Chatbot Dependence, Anxiety, and Depressive Symptoms in University Students: Initial Psychometric Evaluation of the AICDS". Developed and validated the Artificial Intelligence Chatbot Dependence Scale (AICDS).
- *Information* (MDPI): "Unpacking AI Chatbot Dependency: A Dual-Path Model of Cognitive and Affective Mechanisms". Proposed a dual-path model of cognitive and affective mechanisms underlying AI chatbot dependency.
- Kim et al. (2025): "From algorithm aversion to AI dependence: Deskilling, upskilling, and emerging addictions in the GenAI age". Documents how over-reliance on generative AI erodes core competencies.

Misinformation

- UNESCO (2025): "Deepfakes and the crisis of knowing". Ranks AI-amplified disinformation among World Economic Forum's top global risks.
- European Parliament (2025): "Children and deepfakes". Deepfakes projected to reach massive scale by 2025; Europol estimates up to 90% of online content may be synthetically generated by 2026.
- Deloitte (2024): "Deepfake disruption: A cybersecurity-scale challenge". Half of respondents more skeptical of online information accuracy than a year ago.
- PMC (Jul 2025): "AI-driven disinformation: policy recommendations". Reviews how AI-generated content floods platforms faster than humans can moderate.

Job Displacement

- SSRN (Jun 2025): "AI Job Displacement Analysis (2025-2030)". Documents over 76,000 positions eliminated in 2025 alone.
- Final Round AI (2025): "AI Job Displacement in 2025 and 2026: Which Jobs Are At Risk?". Satya Nadella confirmed 30% of Microsoft's code is now written by AI; over 40% of May 2025 layoffs targeted software engineers.
- PMC (2025): "Psychological impacts of AI-induced job displacement among Indian IT professionals". Documents real human cost including anxiety, depression, and identity

disruption.

- AIMultiple: "Top 20+ Predictions from Experts on AI Job Loss". Sharp acceleration in AI-related job reductions through 2025–2026.

Centralization

- Brookings Institution: "What happens when AI companies compete with their customers?". Anthropic 40%, OpenAI 27%, Google 21% market share.
- Bowdoin Science Journal: "Between Concentration and Decentralization: The Changing Landscape of AI Power". Documents governance concentration among OpenAI, Google DeepMind, Anthropic, and Meta.

Revision #8
Created 2026-09-21 10:41:28 UTC by Clive
Updated 2026-09-29 08:28:47 UTC by James