EXHIBIT B

IN THE UNITED STATES DISTRICT COURT
FOR THE NORTHERN DISTRICT OF CALIFORNIA

CONCORD MUSIC GROUP, INC., et al.,

Plaintiffs,

Case No. 24-cv-03811-EKL-SVK

v.

ANTHROPIC PBC,

Defendant.

EXPERT REBUTTAL REPORT OF RUSS ALTMAN

JANUARY 23, 2026

Highly Confidential – Attorneys' Eyes Only

TABLE OF CONTENTS

I. Introduction 1

A. Assignment 1
B. Summary of opinions 2

II. Professor Zhao mischaracterizes technical, market, and governance challenges as intrinsic "risks" of AI 4

A. Professor Zhao's claims of educational harm ignore institutional adaptations to LLMs 16
B. Professor Zhao misrepresents Anthropic's mission and safety-first approach to LLM development 25

III. Professor Zhao ignores the significant benefits and transformative nature of LLMs 29

A. Professor Zhao ignores the ways in which Claude is benefiting education 30
B. Claude is transforming science and medicine 30
C. Claude has made a positive impact in creative sectors and public civil engagement 33

IV. Conclusion 34

Appendix A: Materials Considered 36

I. INTRODUCTION

A. Assignment

1. I have been retained by Wilmer Cutler Pickering Hale & Dorr LLP ("Counsel") to provide expert testimony in connection with the matter *Concord Music Group, Inc. et al., v. Anthropic PBC,* No. 5:24-cv-3811-EKL (N.D. Cal.).

2. On December 19, 2025, I submitted an expert report in this matter (the "Altman Opening Report" or my "Opening Report"). My qualifications are provided in my Opening Report and my curriculum vitae was included in Appendix A of that report.[1] Capitalized terms not defined herein have the same meanings given in the Altman Opening Report.

3. For this report, Counsel has asked me to evaluate and respond to the opinions and conclusions on the "risks of AI" provided by Plaintiffs' expert, Professor Ben Zhao, in his expert report dated December 19, 2025 (the "Zhao Opening Report").[2] My lack of a response to any particular opinion in the Zhao Opening Report should not be interpreted as agreement with that opinion.

4. The scope of my opinions, research, and report is limited to the issues described in the paragraphs above. I have not been asked to opine on issues beyond the scope of my assignment.

5. As noted in my Opening Report, AI, in particular Large Language Models ("LLMs"), is a rapidly evolving field. My opinions expressed in this report reflect the state of knowledge and evidence that was available to me at the time of writing this report. My work in this matter is ongoing. I reserve the right to supplement or revise my opinions as any new information becomes available.

6. In addition to the data and documents cited in my Opening Report, additional data and materials considered are listed in Appendix A of this report.

[1] Expert Report of Russ Altman, Concord Music Group, et al., v. Anthropic PBC, Case No. 24-cv-03811-EKL-SVK (Dec. 19, 2025) ("Altman Opening Report"), Section I.B ("Qualifications") and Appendix A: CV.

[2] Expert Report of Ben Zhao, Concord Music Group, et al., v. Anthropic PBC, Case No. 24-cv-03811-EKL-SVK (Dec. 19, 2025) ("Zhao Opening Report").

7. I am being compensated for my work in preparing this report at a rate of $2,000 per hour. I had assistance in this matter from staff at Keystone Strategy, LLC, who worked under my direction. All the opinions and conclusions stated in this report are my own. My compensation does not depend upon the outcome of this matter or the substance of any opinion I reach.

B. Summary of opinions

8. My opinions to date are contained in the body of this report and of my Opening Report. I present a summary of my opinions in this report herein.

9. While Professor Zhao and I agree on certain elements of the "risks" of AI (e.g., the challenges faced by educators), we diverge on how those challenges should be interpreted, evaluated, and addressed, specifically given the significant benefits of this technology, as well as their relevance to the transformative nature of AI's applications.[3]

10. Professor Zhao's analysis fails to distinguish between the challenges of deployment, governance, and institutional adaptation commonly observed in the early phases following the introduction of a new technology, and intrinsic technological risk. The majority of the "risks" identified by Professor Zhao, such as misuse by bad actors, content proliferation, and educational disruption, are not evidence of inherent flaws of LLMs. Rather, they are predictable and historically common challenges that arise during the early adoption of transformative technologies.

11. Many of the specific issues Professor Zhao highlights are already being actively addressed. For example, educational institutions have been, and continue to, adapt assessment methods, classroom policies, and academic integrity standards to account for LLM tools. These responses undermine Professor Zhao's claim that LLMs are being deployed in an uncontrolled manner.

12. Professor Zhao fails to recognize even one of the innumerous benefits of LLMs or the ways in which LLMs are already being used to drive improvements in productivity and well-being across society. My Opening Report enumerated a variety of early and concrete

[3] The economic literature describes transformative technologies as ones that reshape broad swaths of the economy through "pervasiveness, inherent potential for technical improvements, and 'innovational complementarities'" that amplify downstream productivity and new applications. *See* Timothy F. Bresnahan and M. Trajtenberg, General purpose technologies 'Engines of growth'? (Jan. 1995), https://doi.org/10.1016/0304-4076(94)01598-T.

examples, which evidence the transformative benefits of AI across many disciplines and industries, and are benefits which will continue to proliferate in the future.[4] Although, the long-run risks of AI remain largely theoretical, speculative, and highly susceptible to evolving mitigation strategies, LLMs are currently providing demonstrable and concrete benefits in the real world. For many reasons, including the ones I discuss in my Opening Report and in the sections below, it stands to reason that AI will continue to provide new benefits in the future as the technology improves.

13. In my opinion, the appropriate focus is not whether LLMs present challenges, because—like any new technology—they do, but rather whether those challenges are manageable through technical innovation, market adaptation, and reasonable governance, while also allowing for the significant value that these LLMs introduce to continue growing.[5] Based on my experience and the historical record of other notable transformational technologies, I conclude that the challenges identified by Professor Zhao do not demonstrate that LLMs pose unmitigable risks to society. Instead, as with any new technology, these are indicative examples of the growing pains that can and should be addressed through prudent governance. By bringing into focus oversight efforts by a range of stakeholders in the use of LLMs, I maintain that there is a path to fostering substantial economic value and visible improvements to people's everyday lives through the use of Generative AI ("GenAI") technologies, like Claude.

14. Accordingly, I find no basis to conclude that the early-stage effects cited by Professor Zhao bear any impact on the substantial long-term potential benefits of LLMs or justify characterizing these technologies as inherently harmful. Almost every new technology has potential for misuse, in addition to its beneficial uses, and it is up to key stakeholders, from users and developers to policymakers and ethicists, to: (i) foster discussion and policies about how to encourage beneficial uses; and (ii) realistically recognize and mitigate the harmful uses. As with prior transformative technologies, *see* Section II, the long-term

[4] AI and specifically LLMs are a new technology and the benefits we are seeing from early applications are still an early version of the technology's full potential. These kinds of applications usually improve with use and evolution in response to feedback, new ideas, and new capabilities.

[5] It is worth noting that although the AI risks Dr. Zhao mentions are likely manageable through solutions developed by risk-focused AI companies (like Anthropic), markets, technologies, institutions, and governance bodies (including policymakers and regulators), he does not suggest, nor have I seen evidence that these risks are susceptible to mitigation or resolution by copyright law specifically.

impact of LLMs will depend on how institutions, markets, and policies evolve alongside them.

15. Based on the historical trajectory of transformative technologies, as well as the trajectory of LLMs in general, it is my opinion that the transformative applications for and benefits of this technology will continue to expand, particularly in science, medicine, and other public-serving domains.

II. PROFESSOR ZHAO MISCHARACTERIZES TECHNICAL, MARKET, AND GOVERNANCE CHALLENGES AS INTRINSIC "RISKS" OF AI

16. All new technologies present challenges that need to be considered and managed. Where Professor Zhao and I differ is not on whether AI presents challenges, but rather on how those challenges should be interpreted, evaluated, and addressed. As the Neubauer Professor of Computer Science, Professor Zhao's analysis reflects a perspective informed primarily by his computer science background, and as a researcher in educational settings.[6] My work, by contrast, concentrates on the applications of AI and LLMs across a wide range of real-world domains, including medicine, biomedical research, public health, policy, and other high-stakes societal contexts in which AI and LLM systems are already being deployed at scale.

17. Due to our differing vantage points, Professor Zhao and I diverge sharply in our assessment of whether the challenges presented by AI demonstrate intrinsic technological risk or instead reflect issues of deployment, policy design, and institutional adaptation, all of which are commonly observed in the early periods of new technological innovations. In my view, Professor Zhao also largely ignores the substantial and growing benefits of LLMs, which in my opinion, must be considered alongside any discussion of the potential harms presented by the technology.

18. First, Professor Zhao does not distinguish between the different types of risks associated with complex technologies. In a technical sense, risk is typically characterized as a function of the likelihood and severity of harm in a given context.[7] Intrinsic risk refers to harms that arise from the inherent properties of a technology and that persist even when

[6] Zhao Opening Report, ¶ 1.

[7] National Institute of Standards and Technology (NIST), Glossary: Risk, https://csrc.nist.gov/glossary/term/risk.

reasonable safeguards, governance mechanisms, and best practices are in place.[8] By contrast, deployment risk arises from how, where, and by whom a technology is used, while governance and institutional risks reflect gaps in oversight, incentives, or regulatory frameworks that shape real-world outcomes. These distinctions are important because these categories differ substantially in their predictability, ability to be mitigated, and implications for long-run social impact.

19. Against this backdrop, the issues Professor Zhao repeatedly labels as "risks of AI" should overwhelmingly be characterized as risks of deployment, governance, and institutional adaptation rather than intrinsic technological risks. Nearly every example of harm raised by Professor Zhao is contingent on specific patterns of misuse and has an associated set of mitigations that will develop over time. For example, concerns of students cheating with LLMs reflect vulnerabilities in certain assessment formats, not an inherent inability to evaluate learning. In response, teachers continue to develop best practices for student evaluation that assess complex reasoning and detailed thinking and are not easily bypassed by students.[9] The "risks" that Professor Zhao identifies indicate that companies, markets, technologies, and in some cases, policy frameworks must evolve alongside LLMs. Treating these challenges as evidence of intrinsic social harm conflates fundamentally different categories of risk and represents a central analytical flaw in Professor Zhao's report.

20. Second, throughout Section VII of the Zhao Opening Report, Professor Zhao advances the view that LLMs and their continued development pose significant risks to a varied group of stakeholders, with potential harms ranging from disruptive to catastrophic.[10] He emphasizes that these harms are already observable, measurable, and ongoing, and he asserts that LLMs, *if* left "unregulated and uncontrolled," can inflict widespread societal damage,[11] ignoring that companies like Anthropic are already taking steps to mitigate

[8] In the AI field, recent guidance emphasizes that trying to completely eliminate risk is unrealistic. The NIST AI Risk Management Framework (2023) explicitly warns that "attempting to eliminate negative risk entirely can be counterproductive in practice because not all incidents and failures can be eliminated." See NIST AI Resource Center, NIST AI Risk Management Framework (Jan. 26, 2023), https://airc.nist.gov/airmf-resources/airmf/1-sec-risk/.

[9] *See* analysis below, ¶¶ 42-44.

[10] Zhao Opening Report, ¶ 156. ("AI development poses significant risks to a number of different stakeholders, ranging from the disruptive to the catastrophic.").

[11] Zhao Opening Report, ¶ 157. ("Since their arrival, LLMs have demonstrated through real events that unregulated and uncontrolled AI can inflict a variety of harms on broad subsets of the population. These harms are real and measurable today…").

known and even theoretical harms of AI.[12] Professor Zhao further highlights extreme scenarios, including the potential misuse of LLM-enabled knowledge in areas such as bioweapons development, national security, and the erosion of shared social truth in democratic societies.[13] Taken together, his framing presents LLM-related risks as both severe and inevitable, while expressing skepticism that the promised societal benefits of systems such as Claude have yet been realized.[14]

21. I do not disagree with Professor Zhao that LLMs introduce meaningful challenges and their risks warrant serious attention. In fact, addressing these challenges is central to Anthropic's founding purpose, as demonstrated by its explicit acknowledgement of such risks,[15] and its extensive, proactive efforts to anticipate and prevent them.[16] I have written

[12] *See, e.g.,* Anthropic_0000570411 (discussing Anthropic's Responsible Scaling Policy); Anthropic_0000570419 (describing the creation of a Long-Term Benefit Trust with authority over board composition); Anthropic_0000570027 (outlining Anthropic's framework for building safe and trustworthy AI agents); Anthropic_0000570384 (highlighting research on interpreting the internal reasoning of large language models); Anthropic_0000569592 (describing Anthropic's layered, end-to-end safeguards across Claude's lifecycle); Anthropic_0000570351 (discussing Anthropic's research on AI risks and opportunities and its effort to promote a "race to the top" on safety); Anthropic_0000570340 (discussing Anthropic's research on methods that can defend AI models against universal jailbreaks); and Anthropic_0000570435 (describing Anthropic's key processes, programs, and practices for responsible AI development).

[13] Zhao Opening Report, ¶ 158. ("AI models are capable of having more net knowledge than any human, and that knowledge can be exploited by humans to create extremely dangerous things, such as bioweapons or nuclear bombs. In light of all these present-day harms, many are openly questioning the value of LLMs and their societal harms, particularly their cumulative, long-term risks to science, education and shared social truth in democratic societies.").

[14] Zhao Opening Report, ¶ 157. ("[M]any of Anthropic's justifications for its products and potential benefits for humanity remain unrealized years after its first release of Claude. All this stands in stark contrast to Anthropic's stated mission to help mankind through AI, and its goal of differentiating itself from its competitors by more ethical design and practice.").

[15] The CEO of Anthropic, Dario Amodei, has consistently maintained that safety is one of the company's top priorities. *See* Deposition of Dario Amodei (Jan. 8, 2026) at 146:10-147:22 ("Anthropic is not developing AI models that are just like its competitors. Anthropic measures its models for misalignment risk. And as of the most recent models, Anthropic models perform substantially better than other models on misalignment risk. Anthropic is substantially more transparent, […] being honest about the risks of these models in disclosing the risks of these models. Anthropic has also advocated for regulation, for legislation, forcing companies to disclose the risk of these models and take safety measures against the risk of these models. Now, that's actually a very costly action. Talking about the risk of AI models is an incredibly costly action […] But […] I am still willing to talk about these risks, because it's so important to society to do it."). Anthropic was founded on the premise that it could make LLM technology safer than its competitors. Anthropic_0000593638, Anderson Cooper and Nichole Marks, 60 Minutes, CBS News, Why Anthropic CEO Dario Amodei spends so much time warning of AI's potential dangers (Nov. 16, 2025), https://www.cbsnews.com/news/anthropic-ceo-dario-amodei-warning-of-ai-potential-dangers-60-minutes-transcript/. ("Dario Amodei is 42 and previously oversaw research at what is now a competitor, OpenAI, working under its CEO Sam Altman. He left along with six other employees, including his sister, Daniela, to start Anthropic in 2021. They say they wanted to take a different approach to developing safer artificial intelligence.").

[16] Anthropic builds layered, end-to-end safeguards into Claude's entire lifecycle. This ranges from policy design and training to testing, real-time enforcement, and ongoing monitoring. For example, Anthropic has operationalized its Responsible Scaling Policy to manage catastrophic risks from increasingly powerful AI models. See, e.g. Anthropic_0000569592 and Anthropic_0000570411.

extensively about the need for regulation of AI in medicine.[17] In my research with colleagues, we recognize that scientific advances have long presented dual-use risks—that is, tools developed for beneficial purposes can be repurposed for harm.[18] These risks are not evidence of intrinsic technological defects, nor are they simple deployment failures. Instead, they are possible consequences of powerful technologies, whose impacts can extend beyond their intended use. As such, there exist capability-specific mitigation strategies that can be implemented by individual research teams, without hindering the significant value of this technology.[19] Best practice mitigation strategies developed by researchers, AI specialists, and policymakers are already starting to form actionable guidance. One example of such guidelines can be found in the U.S. Food and Drug Administration's ("FDA") recent January 2026 publication, "Guiding Principles of Good AI Practice in Drug Development."[20]

22. While I agree with Professor Zhao that challenges with emerging technologies like LLMs do exist and need to be managed, I disagree with the implicit undertone of his analysis, which suggests that these challenges are largely not addressable and that early-stage risks will continue unabated over time. In my opinion, Professor Zhao does not adequately account for, and in several instances overlooks, the substantial ongoing efforts, from stakeholders including model developers like Anthropic to policy makers and regulators, aimed at technical mitigation, institutional learning, and market and regulatory adaptation.

---

[17] *See, e.g.,* Artem A. Trotsyuk, et al., Toward a framework for risk mitigation of potential misuse of artificial intelligence in biomedical research (Nov. 26, 2024), https://doi.org/10.1038/s42256-024-00926-3.

[18] *See* Artem A. Trotsyuk, et al., Toward a framework for risk mitigation of potential misuse of artificial intelligence in biomedical research (Nov. 26, 2024), https://doi.org/10.1038/s42256-024-00926-3. ("The notion that scientific advances can be misused for harm is not a new one. From the Dr. Frankenstein's monster of literature to the real-world atomic bomb and lab-created pathogens, the scientific community has long been intertwined with risk of misuse. As AI proliferates through academia and industry, scientific research can be further misused for harm, and researchers have an obligation to prevent it.").

[19] Specifically, we develop a multi-pronged mitigation framework that integrates: (1) existing ethical and regulatory guidance from AI, biosecurity, and dual-use research; (2) established off-the-shelf AI safety practices such as adversarial testing, red-teaming, and transparency measures; and (3) design-specific safeguards including differential privacy, controlled access, watermarking, and federated learning. We encourage researchers to weigh risks against benefits at various stages of research and design. To make things concrete, we apply this framework across three domains in biomedicine where AI uniquely shapes misuse risk: ambient intelligence in healthcare, generative models for synthetic data, and AI-enabled chemical and drug discovery. See further in Artem A. Trotsyuk, et al., Toward a framework for risk mitigation of potential misuse of artificial intelligence in biomedical research (Nov. 26, 2024), https://doi.org/10.1038/s42256-024-00926-3.

[20] In it, the FDA explicitly recognizes that there are "complex and dynamic processes involved in developing, deploying, using, and maintaining AI technologies." It calls for a human centric design, clarity in context of use, and utilizing risk-based performance assessment measures among other things to "inform, enhance, and promote the use of AI for generating evidence across all phases of the drug product life cycle." *See* U.S. Food and Drug Administration, Guiding Principles of Good AI Practice in Drug Development (Jan. 2026), https://www.fda.gov/media/189581/download.

By evaluating present-day potential harms largely in isolation, his analysis does not establish long-run social harm.

23. As a result, Professor Zhao's analysis implicitly assumes: (i) no meaningful improvement in safeguards over time; (ii) no learning or behavioral adaptation by users, institutions, or platforms; and (iii) no effective response by AI companies, markets, technologies, or governance bodies, including policymakers and regulators. These assumptions are inconsistent with both historical experience and the observed evolution of complex transformative technologies, particularly dual-use technologies, whose deployment and oversight mature through a combination of technological innovation, market adaptation, institutional learning and, in some cases, policy response.

24. Based on my experience studying and advising on complex technological systems, these assumptions are inconsistent with how transformative technologies evolve in practice. Historically, early-stage challenges prompt iterative improvements in technical design, governance structures, and institutional norms, rather than persisting unchanged. Numerous examples come to mind that illustrate how initial disruptions associated with new technologies are addressed over time with technical innovation, market adaptation, good governance, and in some cases, regulatory intervention.

25. The widespread rollout of electrical power in the late 19th and early 20th centuries introduced significant new risks, including electrocution and electrical fires. Inconsistent standards and limited understanding of electricity led to frequent accidents, property damage, and public fear, particularly as electrical systems were rapidly installed without uniform regulation.[21] Early critics made claims that high-voltage alternating-current electricity was inherently dangerous and unsuitable for dense urban environments.[22] New

[21] *See* Richard Hirsh, Virginia Tech, Electrification in History (2024), https://hirsh.history.vt.edu/history-of-resuscitation-techniques/. ("It turns out that the emerging electric utility industry in the 1880s encountered a serious practical and public-relations problem. In particular, too many of its workers and customers died from electrical shocks. Linemen and power plant operators often inadvertently contacted high-current cables, while customers sometimes died after touching inadequately insulated wires or poorly grounded appliances.")

[22] *See, e.g.,* David L. Morton, Jr., Reviewing the history of electric power and electrification (June 1, 2002), https://doi.org/10.1016/S0160-9327(02)01422-9. ("At the moment of its invention, there were competing ideas about the sale and use of electricity that had to be worked out."). *See also* Thomas P. Hughes, Harold P. Brown and the Executioner's Current: an Incident in the AC-DC Controversy (1958), https://doi.org/10.2307/3111700. ("During the closing years of the decade witnessing the introduction of the competing systems, the issue of alternating versus direct current dominated the arena where the campaigns for the electric lighting market were fought. Not only were the engineers, scientists, and businessmen involved, but lawmakers and the public were also to play a part. Contemporaries called this the "battle of the currents." […] Efforts would also have been made to outlaw by legislation the advantageous features of the alternating-current system, and to have firmly fixed in the public mind through capital punishment by electrocution the association of alternating current with death").

products were developed to reduce these risks such as fuses, and later circuit breakers. Furthermore, the development of electrical safety codes (such as those established by the International Electrotechnical Commission ("IEC")), standardized product settings (such as Underwriters Laboratories ("UL") standards), and professional licensing requirements drastically reduced many of the initial risks.[23] These advancements enabled electricity to become a foundational technology for the 20th century, leading to advancements in industry, healthcare, communications, and daily life.

26. Following the discovery of x-rays and radioactive materials in the 1890s, radiation was quickly adopted across medicine, industry, and consumer products, often without a full understanding of its biological risks.[24] Early uses included unshielded medical imaging, consumer devices, and even novelty "glow-in-the-dark" products that exposed individuals to excessive radiation.[25] As adverse health effects became apparent, scientific and medical institutions developed rigorous safety standards, exposure limits, professional training requirements, and oversight mechanisms.[26] These adaptations did not eliminate

[23] *See* IEEE Standards Association, The National Electric Code and the National Electronic Safety Code – Partners for a Safer Tomorrow (Sep. 2016), https://standards.ieee.org/wp-content/uploads/import/documents/other/nec_nesc_case_study.pdf. ("The fact is—more times than not—we have been successful in … meeting our common commitment to the highest standards of electrical safety for our people and the general public."). *See also* International Electrotechnical Commission, How & why the IEC was started, https://www.iec.ch/history/how-why-iec-was-started. ("[O]ver time a much wider interest had developed in creating a coherent system of units for electricity. For this reason two "commissions" of the IEC were created: Electric units and standards; Nomenclature and characteristics of electrical machines and apparatus.."). See further UL Standards & Engagement, What is a UL standard?, https://ulse.org/standards/. ("A UL standard is a published set of best practices for testing and evaluating the safety, security, and sustainability of a product or system. Our standards are developed through an accredited process involving experts across industries and interests."). *See also* LegalClarity, Do Electricians Need a License? Requirements and Penalties (July 3, 2025), https://legalclarity.org/do-electricians-need-a-license-requirements-and-penalties/. ("Electrical work is a regulated profession due to safety risks. Performing this work requires specialized knowledge to prevent hazards like fires and electrocution. For this reason, governments establish strict licensing rules to ensure that individuals performing electrical services are qualified...In the United States, there is no national electrician license; regulation occurs at the state or even local level.").

[24] *See* Environmental Protection Agency, Significant Discoveries and the History of Radiation Protection (2018), https://www.epa.gov/sites/default/files/2018-12/documents/significant_discoveries_history_radiation_protection-worksheet_rp_1.pdf. ("Scientists learned that radiation was not only a source of energy and medicine; it could also be a potential threat to human health if not handled properly. In fact, early pioneers in radiation research died from radiation- induced illnesses from too much exposure.").

[25] *See* Environmental Protection Agency, Significant Discoveries and the History of Radiation Protection (2018), https://www.epa.gov/sites/default/files/2018-12/documents/significant_discoveries_history_radiation_protection-worksheet_rp_1.pdf. ("After the war was over, the industry that supported this technology changed its focus to paint glow-in-the- dark clocks and watch faces. The young women who painted these items would form a fine point on their paint brushes by pulling the freshly-dipped brushes between their lips before applying the paint onto the watch faces. Unknowingly, they were swallowing small amounts of radium and damaging their bodies. Several of the women died of unexplained anemia and disease complications with their mouth, teeth and jaw.").

[26] *See* Environmental Protection Agency, Significant Discoveries and the History of Radiation Protection (2018), https://www.epa.gov/sites/default/files/2018-12/documents/significant_discoveries_history_radiation_protection-worksheet_rp_1.pdf. ("By 1915, the British Roentgen Society had adopted a resolution to protect people from overexposure to x-rays. This was one of the first organized efforts in radiation protection. American organizations

radiation-based technologies. Instead, they enabled their safe and targeted use in medicine, including diagnostic imaging (e.g., x-rays and CT scans) and cancer treatment, resulting in substantial public health benefits.[27]

27. The Global Positioning System ("GPS") was initially developed as a military technology.[28] Early concerns suggested that widespread civilian access could jeopardize national security, enable misuse, or undermine military advantages. Consequently, GPS signals were intentionally degraded for civilian users for many years through a policy known as "Selective Availability."[29] Over time, however, the U.S. government determined that the societal benefits of broader access outweighed these risks. Once Selective Availability was discontinued, GPS evolved into a ubiquitous civilian technology.[30] Today, robust governance and international coordination manage security risks while allowing GPS to revolutionize navigation, emergency response, transportation, agriculture, and scientific research.[31]

28. The introduction of the automobile led to an increase in pedestrian and driver fatalities and injuries, as well as higher incidents of property damage. Instead of banning automobiles, regulatory and safety measures were implemented, including the introduction of speed limits (in 1901),[32] roadway design changes to improve pedestrian safety, and the adoption

---

had adopted the British protection rules by 1922. Awareness and education continued to grow. Throughout the 1920s and 1930s more guidelines were developed, scientists were studying the effects of radiation on living organisms, and various organizations were formed to address radiation protection in the United States and overseas.").

[27] *See* U.S. Centers for Disease Control and Prevention, Facts About Imaging Procedures (Feb. 15, 2024), https://www.cdc.gov/radiation-health/data-research/facts-stats/imaging-procedures.html. ("While additional exposure to radiation can be harmful to human health in certain doses, the benefits of diagnosing or treating health problems with ionizing radiation usually outweigh the risks involved.").

[28] Sameer Jumar and Kevin B. Moore, The Evolution of Global Positioning System (GPS) Technology (Mar. 2002), https://doi.org/10.1023/A:1013999415003. ("The GPS is a dual-use (military–civil) satellite based radio navigation system initially developed and currently maintained and operated by the U.S. Department of Defense").

[29] GPS.gov, Selective Availability, https://www.gps.gov/selective-availability. ("Selective Availability (SA) was an intentional degradation of public GPS signals implemented for national security reasons.").

[30] GPS.gov, Selective Availability, https://www.gps.gov/selective-availability. ("In May 2000, at the direction of President Bill Clinton, the U.S government discontinued its use of Selective Availability in order to make GPS more responsive to civil and commercial users worldwide.").

[31] GPS.gov, What can GPS Do, https://www.gps.gov/what-can-gps-do. ("The free, open, and dependable nature of GPS has led to the development of hundreds of applications affecting every aspect of modern life."). *See also* LandAirSea, Missing Persons Statistics: How GPS Revolutionizes Search and Rescue, https://landairsea.com/blogs/consumers/missing-persons-statistics-how-gps-revolutionizes-search-and-rescue. ("GPS revolutionizes search and rescue by allowing authorities to home in on a person's exact position or the location of a vehicle or personal belongings. With the right trackers in place, GPS technology can play a vital role in reuniting missing people with their friends and families.").

[32] *See* History.com, This Day in History, Connecticut Enacts First Speed-Limit Law (Nov. 13, 2009), https://www.history.com/this-day-in-history/may-21/connecticut-enacts-first-speed-limit-law. ("On May 21, 1901,

of vehicle safety advancements, such as seat belts (in 1968)[33] and airbags (in 1998),[34] to reduce fatalities and injuries. Safety improvements were also driven by responsible market actors. For example, in 1959, Volvo pioneered the modern three-point seat belt and made the design freely available, accelerating industry-wide adoption.[35] This combination of company-led innovation, regulatory standards, and market adoption improved safety outcomes over time, and demonstrates how the risks associated with transformative technologies are often mitigated through institutional and market responses.

29. For high-impact technologies, the identification, documentation, and mitigation of potential risks are widely recognized as best practices. These practices are routinely encouraged by stakeholders such as LLM developers, market participants, researchers, and policymakers.[36] The presence of safety guardrails in LLMs, including Anthropic's Claude models, and the fact that they are continuously refined and improved, undercuts Professor Zhao's characterization of LLM systems as "uncontrolled."[37] Iterative improvement is a defining feature of mature technological governance, particularly for transformative technologies whose applications evolve over time. Anthropic, along with other leading LLM developers, has implemented multiple evolving measures to identify, mitigate, and manage potential harms of general-purpose AI-technology. Additionally, Anthropic's engineers and support teams constantly monitor, test, and update the processes involved

Connecticut becomes the first state to pass a law regulating motor vehicles, limiting their speed to 12 mph in cities and 15 mph on country roads.").

[33] *See* Liam Cope, Engineer Fix, When Did Seatbelts Become Mandatory in Cars? (Jan. 9, 2026), https://engineerfix.com/when-did-seatbelts-become-mandatory-in-cars/. ("The 1968 federal law ensured that a seatbelt was available in every new car").

[34] *See* History.com, This Day in History, Federal legislation makes airbags mandatory (May 27, 2025) https://www.history.com/this-day-in-history/September-1/federal-legislation-makes-airbags-mandatory. ("On September 1, 1998, the Intermodal Surface Transportation Efficiency Act of 1991 finally goes into effect. The law required that all cars and light trucks sold in the United States have airbags on both sides of the front seat.").

[35] *See* Abed Serhan, Engineerine, The Innovation That Saved Millions: The Story of the Three-Point Seatbelt, (Jan. 6, 2025), https://engineerine.com/three-point-seatbelt-by-nils-bohlin-volvo/. ("Despite holding the patent, Volvo chose not to monetize the invention. Recognizing its potential to save lives, the company opened the patent to all car manufacturers. This unprecedented move ensured that the three-point seatbelt became a universal standard, benefiting millions worldwide. This decision reflects Volvo's commitment to safety as a core value.").

[36] *See, e.g.,* discussion of Anthropic's Responsible Scaling Policy in Anthropic_0000570411. *See also* Anthropic_0000570419, which discusses Anthropic's creation of a Long-Term Benefit Trust (LTBT), an independent governance body with authority over board composition.

[37] Zhao Opening Report, ¶ 157. ("Since their arrival, LLMs have demonstrated through real events that unregulated and uncontrolled AI can inflict a variety of harms on broad subsets of the population.").

in building Anthropic's AI product offerings.[38] It is not accurate to describe these processes and their resulting products as "uncontrolled".

30. More broadly, many of the concerns Professor Zhao raises reflect governance questions rather than intrinsic technological defects. Challenges related to education, platform misuse, and content moderation are appropriately addressed through technical mitigation and oversight, institutional design, and established governance best practices. Treating these challenges as inherent flaws of LLMs, and Claude in particular, misallocates responsibility away from the mechanisms that historically shape how dual-use technologies are deployed and evolve. Professor Zhao repeatedly cites early misuse, including AI-generated content, academic cheating, and misinformation as evidence of enduring harm.[39] Contrary to Professor Zhao's assertion, temporary overload or early misuse does not imply permanent degradation. In fact, periodic waves of low-quality or surplus content are a recurring outcome of sharp reductions in production cost, and not evidence of systemic market failure.[40] Historically, companies and markets have developed mitigation mechanisms such as curation, labeling, and quality differentiation.[41]

---

[38] *See, e.g.,* Anthropic, Building effective agents (Dec. 19, 2024), https://www.anthropic.com/engineering/building-effective-agents. ("Over the past year, we've worked with dozens of teams building large language model (LLM) agents across industries…In this post, we share what we've learned from working with our customers and building agents ourselves, and give practical advice for developers on building effective agents."). *See also* Claude Docs, What's new in Claude 4.5, https://platform.claude.com/docs/en/about-claude/models/whats-new-claude-4-5.

[39] Zhao Opening Report, ¶ 159. ("One of the major concrete harms of LLMs is the massive proliferation of what is now commonly called "AI slop," i.e., automated content that has little meaning other than being low qualify mimics of human created content.[…] For example, in February 2023, Clark's World, the sponsor of the well-regarded competition for fictional writing, announced that they would discontinue their competition because they received many times greater submissions than they could possibly evaluate, and the overwhelming majority of them were quickly generated AI-text samples.").

[40] *See* Deni Ellis Bechard, The Slop Cycle—How Every Media Revolution Breeds Rubbish and Art (Nov. 9, 2025), https://www.scientificamerican.com/article/ai-slop-how-every-media-revolution-breeds-rubbish-and-art/. ("Mass-produced culture has a long, messy history. The biblical book of Ecclesiastes—thought to have been written between 300 and 200 B.C.E.—laments, "Of making many books there is no end." This was in response to the flood of philosophical writings in the ancient Near East and Hellenistic world. Since then, whenever new tools to crank out communications have become available, somebody has flooded the zone with the fastest, most imitative material that could garner attention.").

[41] For purposes of this discussion: "curation" refers to the act of selecting and organizing content for presentation or distribution, rather than presenting all available content indiscriminately. *See* Merriam-Webster, Curation, https://www.merriam-webster.com/dictionary/curation. ("[T]he act or process of selecting and organizing (something, such as articles or images) for distribution or publication") "Labeling" refers to the provision of written or visual information accompanying content to furnish identifying or descriptive information about that content. See Merriam-Webster, Label, https://www.merriam-webster.com/dictionary/labeling. ("[W]ritten or printed matter accompanying an article to furnish identification or other information.") To better understand how mitigation mechanisms like labeling are currently being used on content platforms like TikTok, *see*, *e.g.*, TikTok, About AI-generated content, https://support.tiktok.com/en/using-tiktok/creating-videos/ai-generated-content. ("TikTok may automatically apply the "AI-generated" label to content we identify as completely generated or significantly edited with AI. This may happen when a creator uses TikTok AI effects or uploads AI-generated content that has Content Credentials attached, a technology from the Coalition for Content Provenance and Authenticity (C2PA). Content Credentials attach metadata to content that we can use to recognize and label AIGC instantly.") "Quality

Similar phenomena long predate the emergence of LLMs. For example, unsolicited physical mail, email spam, and low-quality online content emerged with earlier technologies and reflected misuse of broadly beneficial systems by bad actors rather than inherent defects in the underlying technologies themselves.[42] The printing press enabled a flood of pamphlets, cheap photography produced amateur snapshots, cable TV expanded filler programming, and email resulted in spam.[43] Over time, technical, institutional, and market-based responses substantially reduced these harms without abandoning the core technologies.

31. It is also important to acknowledge that the initial introduction of LLMs has been accompanied by a period of rapid and enthusiastic adoption, during which some potential uses were not fully anticipated or evaluated – as is the case with most if not all new technologies. Over the past few years, however, a significant portion of industry, academic, and policy-focused work has been directed at identifying the shortcomings of early deployments, and articulating principles and safeguards for responsible, safe, and effective use.[44] My own research, and specifically my involvement on task forces and committees,[45] reflects this broader effort to move from early experimentation toward more

differentiation" refers to the practice of marking or signaling distinctions among content based on perceived attributes such as source, reliability, or production method, consistent with Merriam-Webster's definition of "differentiate". *See* Merriam-Webster, Differentiate, https://www.merriam-webster.com/dictionary/differentiate. ("[T]o mark or show a difference in").

[42] *See, e.g.,* Aaron Gordon, How we ended up with all this junk mail (Oct. 15, 2020), https://www.vice.com/en/article/how-we-ended-up-with-all-this-junk-mail/. ("The biggest increase in junk mail occurred in the 1980s for two reasons […] The second change, which direct mail industry executives like to talk about a lot more than the subsidy they were handed by the USPS, was the technological breakthrough in the computer industry that allowed mailers to do much more sophisticated analyses about who to send what mail to based on ZIP codes, census tracts, income levels, recent purchases, and so on […] In 1972, junk mail was roughly 25 percent of all mail pieces delivered. In 1982, that jumped to 32 percent, according to Leonard. The number of catalogs sent to US addresses ballooned from 5.8 million in 1980 to 13.4 billion by 1989, a 131 percent increase, a study on the US direct mail industry found.").

[43] *See* Carlo Kopp, Going deep: Understanding the mechanics of deception and disinformation (Feb. 24, 2025), https://lens.monash.edu/the-science-of-disinformation-and-deception/. ("History shows that newer and faster media for the distribution of information will be exploited for purposes both good and bad."). *See also* Deni Ellis Bechard, The Slop Cycle—How Every Media Revolution Breeds Rubbish and Art (Nov. 9, 2025), https://www.scientificamerican.com/article/ai-slop-how-every-media-revolution-breeds-rubbish-and-art/. ("The 20th-century cinema boom followed a similar pattern. By 1908 roughly 8,000 nickelodeons—five-cent storefront theaters—ran nonstop shows. The production demands produced a lot of rubbish…").

[44] Anthropic was founded to reduce the risks of LLMs. *See* Deposition of Dario Amodei (Jan. 8, 2026) at 145:14-17 and 146:1-9. ("[Question] So you say in this Anthropic memo that you believe there's a 20 to 30 percent chance that AGI will lead to the destruction or permanent stunting of humanity, right? [Answer]: Anthropic exists to try to change that trajectory, to try and change the development of AI in -- in a direction that reduces or eliminates the risks and increases the benefit. The whole purpose of Anthropic's mission is to identify the risks and act in such a way that reduces those risks, to develop the technology in a different and alternative way").

[45] As I describe in my Opening Report, I am personally involved with the Stanford Institute for Human-Centered AI (HAI) and Responsible AI for Safe and Equitable Health (RAISE Health). These organizations work together to guide the responsible use of AI across biomedical research, education, and patient care. *See* Altman Opening

mature governance frameworks that preserve the transformative value of LLMs while mitigating risk.

32. Professor Zhao further presents a portrayal of potential harms that places substantial emphasis on extreme hypothetical scenarios, such as the use of LLMs in the creation of bioweapons or nuclear warfare. While it is appropriate to acknowledge that, like other transformative dual-use technologies, LLMs may be tested by nefarious actors for such purposes, and that even low-probability scenarios warrant serious consideration, it is not the case that there are no mechanisms for addressing these risks. I have personally served on task forces focused on these issues and on identifying mitigation strategies, specifically in my role as co-organizer of the AI & Biotechnology theme of the 2025 "Spirit of Asilomar" summit, where these risks are already under active discussion.[46] Anthropic has itself published information about how it is addressing these risks.[47] While these scenarios are theoretically conceivable and are worthy of consideration, presenting them as representative of the typical or expected impacts of LLM deployment mischaracterizes how these risks are assessed and managed in practice. Realization of these extreme scenarios depends on numerous external factors such as access to specialized materials and equipment, and domain specific knowledge, that extends far beyond the output provided by an LLM. Moreover, these risks are already the subject of active oversight, monitoring, and governance.[48] In the case of bioweapons, there are international treaties,

Report, ¶ 51 and Anthropic_0000595304, Stanford Responsible AI for Sage and Equitable (RAISE) Health, https://med.stanford.edu/raisehealth.html.

[46] I served as a co-organizer for The Spirit of Asilomar and the Future of Biotechnology Summit which took place in February 2025. A critical focus of this summit was to discuss potential risks and mitigation strategies related to biosecurity and bioweapons use in AI technology, including LLMs. As part of this summit, I co-authored an entreaty written on "proposing and assessing risks, risk management strategies, governance frameworks, and challenges in regulating AI tools." *See* Yana Bromberg, Russ Altman, Michael Imperiale et al., Spirit of Asilomar Entreaty 2025.3.1 Artificial Intelligence & the Future of Biotechnology, Spirit of Asilomar (Mar. 1, 2025), https://repository.rice.edu/server/api/core/bitstreams/652ee6ad-e5df-44cc-a00f-7bd2b2c1b8e5/content.

[47] Anthropic, Detecting and Countering Misuse of AI: August 2025 (Aug. 27, 2025) https://www.anthropic.com/news/detecting-countering-misuse-aug-2025. ("Our Threat Intelligence report discusses several recent examples of Claude being misused, including a large-scale extortion operation using Claude Code, a fraudulent employment scheme from North Korea, and the sale of AI-generated ransomware by a cybercriminal with only basic coding skills. We also cover the steps we've taken to detect and counter these abuses.").

[48] The U.S. Nuclear Regulatory Commission ("NRC"), created by Congress with the specific purpose of regulating and ensuring the safe use of nuclear power plants and materials, is one such independent agency which is actively monitoring threats to national nuclear security and maintaining safeguards. On September 29, 2025, the NRC published its FY26 Artificial Intelligence Strategic Plan. This policy specifically identifies enhanced AI governance and AI-specific threat modeling and penetration testing as key cybersecurity action items for the NRC, as part of its broader governance and risk management objectives. *See* U.S. NRC, FY26 Artificial Intelligence Strategic Plan (Sep. 29, 2025), https://ww2.nrc.gov/docs/ML2526/ML25269A196.pdf.

regulatory controls, and built-in surveillance systems for detecting dangerous molecules at companies that synthesize these chemicals.[49]

33. Historical experience demonstrates that effective oversight of dual-use technologies is possible. For example, the 1975 Asilomar conference established voluntary safety guidelines for recombinant DNA research at a formative stage in the technology's development.[50] Similarly, the Chemical Weapons Conference ("CWC"), established a governance model that eliminated an entire class of weapons while preserving the peaceful use of chemistry in science and industry.[51] Comparable oversight mechanisms have been developed and continue to evolve with other dual-use technologies such as genetic engineering, chemical design and synthesis, genome sequencing, and many other technologies I work with regularly. These examples illustrate that the presence of extreme-use risks does not hold back responsible development or substantial societal benefits.

34. Focusing disproportionately on extreme scenarios without an equal consideration for existing mitigation strategies and governance mechanisms diverts attention from the more relevant questions of how risks are managed in practice and how they compare to the substantial public benefits introduced by these technologies.[52] Over time, as with prior transformative technologies, continued technical advancement, broader institutional adaptation, and ongoing governance efforts are likely to expand the beneficial applications of LLMs. With this in mind, the appropriate focus is how LLMs' evolving capabilities can be responsibly guided to maximize public benefit while managing and minimizing foreseeable risks. As I discuss in greater detail in Section II.B, I understand these views to be consistent with Anthropic's overall mission and organizational philosophy. Anthropic acknowledges the risks associated with advanced AI systems and focuses on managing

[49] *See* Department of Homeland Security, CBD Focus Areas – Biosurveillance (Apr. 10, 2025), https://www.dhs.gov/science-and-technology/biosurveillance. ("Biosurveillance primarily focuses on developing effective surveillance, prevention and operational capabilities for detecting and countering biological threats").

[50] J. Benjamin Hurlbut, Taking responsibility: Asilomar and its legacy, Science (Jan. 30, 2025), https://www.science.org/doi/10.1126/science.adv3132. ("In February 1975, leading molecular biologists gathered at the Asilomar conference center on the California coast to evaluate risks of the emerging technology of recombinant DNA and to establish guidelines to govern research.").

[51] *See* Organization for the Prohibition of Chemical Weapons ("OPCW"), OPCW confirms: All declared chemical weapons stockpiles verified as irreversibly destroyed (July 7, 2023), https://www.opcw.org/media-centre/news/2023/07/opcw-confirms-all-declared-chemical-weapons-stockpiles-verified.

[52] Furthermore, many of these extreme tail risks are the first things to be considered and managed against. For example, the meeting on biosecurity, bioweapons, and AI in biotech for which I served as co-organizer looked at these risks as part of our agenda. *See* The Spirit of Asilomar, Why Now? Why Asilomar? https://www.spiritofasilomar.org.

those risks through a governance-oriented approach that emphasizes deliberate system design, oversight, and continuous evaluation to safely develop and deploy LLMs.

A. Professor Zhao's claims of educational harm ignore institutional adaptations to LLMs

35. Professor Zhao characterizes the use of LLMs in education as inherently destructive, asserting that Claude and other LLMs are "disrupting education at every level."[53] As an educator, I have firsthand experience with some of the issues Professor Zhao describes. LLMs are prompting changes across education contexts. Many teachers are being asked to reconsider curriculum design, assessment practices, and evaluation methods without having been trained or adequately supported to do so. In this sense, the current state of affairs reflects not a loss of learning capacity, but a temporary state of flux where emerging technologies and existing instructional and assessment frameworks are adapting to one another. More precisely, the "disruption" that Professor Zhao identifies reflects the persistence of outdated assessment methods rather than a loss of learning capacity. Historically, education systems have responded to comparable shifts introduced by new technologies by gradually adapting their instructional and evaluation frameworks, shifting assessment toward formats that better measure comprehension, reasoning, and applied knowledge, rather than by attributing the disruption to inherent harm caused by the technology.[54] Thus, the early challenges that AI is currently creating in the classroom may ultimately spur better pedagogy and learning in the long run.

36. As part of this adaptation process, I led the AI at Stanford Advisory Committee at Stanford,[55] which in March of 2024 was tasked to "assess the role of AI at Stanford in administration, education, and research to identify potential policy gaps and other needs to advance the responsible use of AI at Stanford."[56] The culmination of our committee's work was a set of guiding principles to govern AI use, reflecting the recognition that it is

[53] Zhao Opening Report, ¶ 162.

[54] One example of this is the evolution of calculus that came from the introduction of handheld calculators. I discuss this example in more detail below, ¶ 40.

[55] The Provost asked me to lead a university faculty and staff committee to evaluate the areas in which Stanford policies may be lagging behind or not appropriate in the AI age. We looked at education, administration and research. Our job was not to set policy but to recommend where policies would be useful. The Provost then asked various units across the university to consider our report and formulate policy. *See* Russ Altman et al., Report of the AI at Stanford Advisory Committee (Jan. 9, 2025), https://provost.stanford.edu/sites/g/files/sbiybj29391/files/media/file/ai-at-stanford-report.pdf.

[56] Russ Altman et al., Report of the AI at Stanford Advisory Committee (Jan. 9, 2025), https://provost.stanford.edu/sites/g/files/sbiybj29391/files/media/file/ai-at-stanford-report.pdf.

impossible to anticipate every potential application and that a principled framework is necessary to support reasonable adoption, not unlike Anthropic's Constitutional AI.[57]

37. More broadly, Professor Zhao's framing of the integration of LLMs into educational settings as inherently disruptive reflects a familiar pattern associated with prior technological shifts in education. Similar concerns were raised with the introduction of calculators, online search engines, and collaborative digital tools, each of which initially prompted fears about erosion of learning and academic integrity.[58]

38. The introduction of handheld calculators during the 1970s, and their subsequent rise in popularity immediately sparked debate about their use in educational settings. Critics speculated that calculators may be used as a "crutch" and that down the line "core skills" would need to be re-taught.[59] Institutional responses varied, for example, The College Board approved handheld calculators on the Advanced Placement ("AP") Calculus Exam in 1983, only to walk back the decision two years later.[60] A decade later, in 1994, the test was redesigned and calculators were reintroduced with clear limitations, including restricted calculator sections, standardized calculator capabilities, and continued requirements for students to show analytical work and reasoning.[61] Since then, calculators

---

[57] *See* Anthropic_0000594126, Constitutional AI, What is Constitutional AI? (accessed Dec. 16, 2025), https://constitutional.ai/.

[58] *See* Harsh Singh, Forbes, The Evolution of Education: From Calculators To Generative AI (Nov. 4, 2024), https://www.forbes.com/councils/forbesbusinesscouncil/2024/11/04/the-evolution-of-education-from-calculators-to-generative-ai/. ("From calculators to the internet, innovations tend to spark debate on whether these tools will hinder or help the learning process. Today, generative AI is at the center of this conversation—much like the calculator was in the 1970s and search engines in the early 2000s."; "When electronic calculators were first introduced into classrooms, educators and parents worried they would cripple students' fundamental arithmetic skills. Critics argued that students would no longer need to learn basic math operations, fearing a loss of the ability to think critically about numbers."); ("Similarly, when Google became widely accessible … [many] feared that students would stop thinking for themselves, becoming passive recipients of information rather than active learners.").

[59] *See* Charlotte Libov, New York Times, Calculators in Exam Debated (Dec. 8, 1985), https://www.nytimes.com/1985/12/08/nyregion/calculators-in-exam-debated.html. ("I think we are going to find down the line if we go to calculators not only problem-solving skills will be done on them but, 5 to 10 years from now, we are going to have to start all over and reteach the core skills").

[60] *See* Scott Bennett, Governors State University, The Evolution of the AP Calculus AB Test: 1955-2018 (2019), https://opus.govst.edu/cgi/viewcontent.cgi?article=1000&context=theses_math. ("Test takers could use calculators for the first time in 1983 […] The use of calculators was abandoned two years later. An issue here could have been that students were not allowed to use a calculator in their day to day calculus course. Then as students could use them on the test, it would have caused confusion and students also felt that they had to use them and that led to more time being used on questions.").

[61] *See* College Board, AP Calculus 2007-2008 Professional Development Workshop Materials, https://apcentral.collegeboard.org/media/pdf/ap-sf-calculus-approximation.pdf. ("Approximation techniques involving derivatives, integrals, and Taylor polynomials have been tested on the AP Calculus Exams from the very beginning. With the transition to the use of graphing calculators and the changes to the AP Calculus Course

have been permitted on the AP Calculus exam and the curriculum has evolved such that the use of calculators is an "integral" aspect of the AP Calculus curriculum.[62] Calculators have become a commonplace tool that supports conceptual understanding while preserving rigorous assessment of core skills.

39. Search engines and collaborative digital tools followed a similar trajectory. I recall, when internet search engines became widely available in the late 1990s and early 2000s, educators raised concerns that students would rely on easy access to information rather than develop independent research, critical thinking, and writing skills. Early responses included outright bans or strict limitations on internet use for assignments.[63] Over time, however, educational institutions adapted by redesigning curriculum and assessments to emphasize source evaluation, synthesis, and original analysis rather than simple information retrieval.[64] Collaborative digital tools, such as shared documents and online discussion platforms, were likewise initially viewed as facilitating academic dishonesty but are now widely integrated into coursework to support group collaboration, peer review, and real-world skill development. Today, both search engines and collaborative technologies are commonplace educational tools that, when appropriately governed, enhance learning outcomes while maintaining academic standards.

40. When calculators, search engines, and digital collaboration tools were introduced, they exposed the limitations of prevailing assessment models rather than flaws in the tools

---

Description in the mid-1990s, however, the emphasis on approximations became a more fundamental component of the course.").

[62] College Board, AP Calculus: Use of Graphing Calculators, https://apcentral.collegeboard.org/courses/resources/ap-calculus-use-of-graphing-calculators. ("The use of a graphing calculator in AP Calculus is an integral part of the course. Teachers should use this technology on a regular basis with students so that students become adept at using their graphing calculators.").

[63] *See* Rebecca M. Howard and Laura J. Davies, Plagiarism in the Internet Age (Mar. 2009), https://pdo.ascd.org/LMSCourses/PD15OC002M/media/GuidingStudentResearchProjectsGrades6-12_M3Reading1.pdf. ("In an age when students gravitate to online sources for research--and when tremendous amounts of both reputable and questionable information are available online--many have come to regard the Internet itself as a culprit in students' plagiarism. Some teachers go so far as to forbid students from researching online, in the mistaken assumption that if students are working from hard-copy sources only, the problem will disappear.").

[64] A useful historical analogue comes from the academic community's early engagement with Wikipedia. When Wikipedia first emerged, educators observed that students frequently cited it directly, raising concerns about reliability and scholarly rigor given its existence as an open-source encyclopedia. Over time, institutions and professors like me, have taught students to use Wikipedia as a starting point, a tool for orientation and discovery, while emphasizing verification through primary and peer-reviewed sources. This approach eventually became the norm and reflects how educational standards evolve to incorporate new knowledge technologies rather than rejecting them outright. See additional details at Neil L. Waters, Why You Can't Cite Wikipedia in My Class Communications of the ACM (Sep. 1, 2007), https://cacm.acm.org/opinion/why-you-cant-cite-wikipedia-in-my-class/.

themselves. Over time, educational institutions adapted by revising instructional design and evaluation methods, not by rejecting the technologies. LLMs present a more rapid and complex challenge, but they likewise call for thoughtful adaptation, institutional innovation, and shared responsibility.

41. Professor Zhao notes that tools such as Claude enable students to rapidly generate written assignments, and claims that this leads to widespread cheating that is difficult to control, particularly in under-resourced schools.[65] But framing LLMs as uniquely enabling uncontrollable cheating overstates the problem and understates the agency of educational institutions and stakeholders as well as the ways in which LLMs can help overcome pre-existing challenges. Well before the introduction of LLMs, students were able to access homework and exam solutions, including explanations and direct answers, through services such as Chegg and similar platforms.[66] These tools amplified existing weaknesses in certain forms of take-home or un-proctored assessments but did not result in the collapse of educational evaluation. Instead, they are prompting positive changes in assessments design and instructional practices through innovations like open book tests, which promoted higher order thinking and problem-solving skills rather than rote memorization. LLMs may amplify existing weaknesses in assessment formats that rely heavily on unsupervised written output, but they do not eliminate the possibility of meaningful evaluation. Rather, they underscore the need for deliberate, supported approaches in how learning is assessed. Importantly, LLMs can also support this evolution by enabling new forms of instruction, feedback, and assessment that strengthens learning outcomes rather than undermine them.[67]

---

[65] Zhao Opening Report, ¶¶ 162-163.

[66] *See* Edmund Pickering and Clancy Schuller, Chegg's Growth, Response Rate, and Prevalence as a Cheating Tool: Insights From an Audit within an Australian Engineering School (July 31, 2024), https://link.springer.com/article/10.1007/s10805-024-09551-6. ("Alarmingly, we found over half of audit units had cheating content on Chegg is broadly used to cheat and 50% of questions asked on Chegg are answered within 1.5 h. This makes Chegg an appealing tool for academic misconduct in both assignment tasks and online exam.").

[67] I have found that student engagement with these issues improves when discussions are framed in terms of long-term skill development and professional competitiveness rather than rule enforcement alone. For example, at Stanford, my students respond strongly to the argument that higher education represents a substantial investment, and that exclusive reliance on widely available tools such as LLMs does not differentiate them in the labor market. Framing LLMs as complements to, rather than substitutes for, foundational skills encourages students to reflect on how they can use these tools to augment their own expertise and competitiveness as well as making themselves more valuable to future employers. Where maladaptive uses persist, my experience is that they tend to arise from identifiable and addressable pressures rather than indifference to learning. Students may perceive certain assignments as low-value, feel acute time pressure and use LLMs as a shortcut without meaningful engagement, or believe that peers are gaining an unfair advantage by using LLMs despite formal prohibitions. These dynamics are not static; they are actively being studied and addressed by educators as this remains a top priority for educators like me.

42. Much like the risks of other technologies that students have used throughout history, the risks of GenAI that Professor Zhao identifies are largely addressable and schools and instructors have already begun implementing measures to counter the ways students can use LLMs for cheating purposes. Educational institutions have already adapted their testing and academic integrity practices to account for the availability of AI tools and to leverage these tools to advance the quality of and access to education.[68] For example, in my class, *Representations and Algorithms for Computational Molecular Biology*,[69] the final exam is taken on pen and paper, and requires not just regurgitation of facts, but synthesis of concepts from lectures into coherent written answers. This medium has been effective, and I plan to increase the percentage of the grade represented by the final exam, as it provides a secure and reliable way to measure student understanding. As the head of the Stanford AI Task Force, I support Stanford in their governance efforts both in preventing knowledge depletion and opening channels for expanded/improved learning. This learning is important not just for students, but also among faculty, researchers, and administrators. In recent years, the Stanford AI Task Force has provided key recommendations about the use of AI in education, including instructor justification and disclosure of AI use in student assessments, revising honor codes to provide finer distinctions in the ways students and faculty may use AI in their work, and providing university-sanctioned AI services to Stanford workers.[70]

43. I understand that Anthropic provides a version of Claude which is specially designed for use by students and educators at institutions of higher education, which targets some of

---

[68] There are several examples regarding how schools from New York, Washington, California, and many other states, are embedding AI in their curriculum and learning methodologies. *See* Bree Dusseault and Justin Lee, CRPE, How Districts Are Responding to AI—and What It Means for the '23-'24 School Year (Sep. 2023), https://crpe.org/study-how-districts-are-responding-to-ai-and-what-it-means-for-the-new-school-year/. For other examples, see Columbia University, Office of the Provost, Generative AI Policy, https://provost.columbia.edu/content/office-senior-vice-provost/ai-policy; and Johns Hopkins, New Guidelines for Responsible Use of Generative AI (GenAI) (May 2025), https://it.johnshopkins.edu/featured-articles/new-guidelines-for-responsible-use-of-generative-ai/.

[69] *See* Stanford University Bulletin, BMDS214 Representations and Algorithms for Computational Molecular Biology, Course Description, https://bulletin.stanford.edu/courses/1410301.

[70] Dan Schwartz, Dean of the Graduate School of Education ("GSE"), noted that "Students have been exploring it much more than the faculty, and that's why it's important to find ways to educate the faculty, which I think is also true on the research side". Schwartz also said "The question is slowly changing from what are our students going to cheat with, to a more sophisticated question of what counts as cheating, how they can use AI in their work, and how to make it part of the assignment." *See* Stanford University, Report Outlines Stanford Principles for Use of AI (Jan. 9, 2025), https://news.stanford.edu/stories/2025/01/report-outlines-stanford-principles-for-use-of-ai. To support this more nuanced approach, the GSE has already created the AI Tinkery, part of the Stanford Accelerator for Learning, to provide a collaborative space for educators to explore the possibilities of AI. More details can be found at Stanford University, Stanford AI Tinkery, https://ai-tinkery.stanford.edu; and Stanford University, Stanford Accelerator for Learning, https://acceleratelearning.stanford.edu.

the early challenges that GenAI has presented in education.[71] Using *Learning mode*, for example, Anthropic aims to guide students in their thinking, rather than providing answers directly.[72] Anthropic has also partnered with leaders at Northeastern University "to build best practices for AI integration in higher education", and to work on ways to address concerns about the use of AI in the classroom.[73] One purpose of Anthropic's partnerships with universities is "to transform teaching [and] research", if not also to "shape how AI can positively transform education and society."[74]

44. As a part of Northeastern University's partnership with Anthropic, the university community received access to Claude.[75] In spite of the purported harms, universities understand the tremendous value that LLMs bring to higher education and are working to mitigate the risks of LLMs while giving their community members (students and teachers) the benefits provided by this cutting-edge technology.[76] For example, some Northeastern classes have begun to incorporate LLM's into their curriculum. In the Journalism School, an AI and Media Industries class "get[s] students working with AI to understand the uses — and limits — of these tools on journalism, social media and public relations."[77] While the professor of the class acknowledges the shortcomings of AI, specifically the tendency to

[71] *See* Anthropic, Introducing Claude for Education (Apr. 2, 2025), https://www.anthropic.com/news/introducing-claude-for-education. ("Today we're launching Claude for Education, a specialized version of Claude tailored for higher education institutions.").

[72] *See* Anthropic, Introducing Claude for Education (Apr. 2, 2025), https://www.anthropic.com/news/introducing-claude-for-education. ("In Learning mode, Claude helps students develop their independent thinking by: Guiding rather than answering: Asking "How would you approach this problem?" instead of providing immediate solutions.").

[73] Anthropic, Introducing Claude for Education (Apr. 2, 2025), https://www.anthropic.com/news/introducing-claude-for-education. ("We're working with Northeastern University to build best practices for AI integration in higher education, new AI-powered education tools, and frameworks for responsible AI adoption in educational settings.").

[74] Anthropic, Introducing Claude for Education (Apr. 2, 2025), https://www.anthropic.com/news/introducing-claude-for-education.

[75] *See* Anthropic_0000569650, Northeastern University, Claude.ai at Northeastern University, https://claude.northeastern.edu/. ("All active Northeastern students, faculty, and staff can access Claude, which includes a full range of models, advanced features, large context windows, and high usage limits.").

[76] *See, e.g.,* David Nordman, Northeastern Global News, Northeastern and Anthropic to lead in responsible AI innovation in higher education (Apr. 2, 2025), https://news.northeastern.edu/2025/04/02/anthropic-ai-partnership/. ("AI is a tool that will empower our students, faculty and staff, and together with Anthropic we will develop and provide the tools, resources and training that they need to effectively integrate AI into their work. We are also committed to the responsible use of AI and proud to work alongside a partner that shares this fundamental commitment.").

[77] Cody Mello-Klien, Northeastern Global News, How a journalism professor got his class to put AI to the test and think like an AI model in the process (June 23, 2025), https://news.northeastern.edu/2025/06/23/faculty-staff/journalism-class-ai-media-industry/.

hallucinate, he believes "the models are helping [them] in important ways, and [he] would like to think [they] learned more than [they] could've without them".[78]

45. Many instructors promote new and improved methods for learning and knowledge consumption using LLMs, while adjusting assessment formats to reduce or eliminate opportunities for misuse. These approaches include in-class, pen-and-paper examinations, device-restricted and proctored assessments, blue book exams, lockdown browsers, and internet or phone bans.[79] These approaches do not rely on detecting LLM outputs after the fact but instead prevent its use during assessments altogether, while acknowledging that LLMs can help students better engage with material during the learning process.[80] Universities are also seeing success by partnering directly with students to address the challenges posed by LLMs on academic integrity and engaging them in the development of student guidelines for their use.[81] Studies have shown that

[78] Cody Mello-Klien, Northeastern Global News, How a journalism professor got his class to put AI to the test and think like an AI model in the process (June 23, 2025), https://news.northeastern.edu/2025/06/23/faculty-staff/journalism-class-ai-media-industry/.

[79] *See, e.g.,* Jocelyn Gecker, The rise of AI tools forces schools to reconsider what counts as cheating (Sep. 12, 2025), https://apnews.com/article/ai-cheating-school-chatgpt-4f89a552e9093ce2180471b4d4736675. ("Cuny's students at Valencia High School in southern California now do most writing in class. He monitors student laptop screens from his desktop, using software that lets him 'lock down' their screens or block access to certain sites."). *See also* Perry Chiaramonte, Is It Real or Is It AI? Professors Turn to Ancient Practice of Oral Exams To Find Out (Dec. 12, 2025), https://www.nysun.com/article/is-it-real-or-is-it-ai-professors-turn-to-ancient-practice-of-oral-exams-to-find-out. ("Some professors have fought back with detection software, though the available tools struggle to produce reliable results. Others have gone decidedly analog, reverting to in-class handwritten exams, spurring a resurgence in the use of 'blue books' — those paper booklets that dominated college testing in the late 20th century.").

[80] For example, Dr. Fischer explains that GenAI can function as a "study buddy," and in many respects may be more effective than a student peer. *See* Toby S. Fischer, 5 Engaging AI Classroom Activities to Try With Your Students (June 3, 2025), https://www.edutopia.org/article/5-engaging-ai-classroom-activities-try-your-students/. ("Students aren't always the best study partners for each other. Some overestimate what they know, others would rather chat about nonacademic topics, and few are trained in using evidence-based study strategies like retrieval or spaced practice. An AI study buddy, though, never gets tired, distracted, or off-topic. Whether students need help quizzing themselves, breaking down complex topics, or identifying where their understanding falls short, AI can provide real-time support. Ideally, for this task, they're using a teacher-created chatbot that you specifically instruct to only provide hints without directly giving the answer ... Alternatively, students can ask the AI to simplify and explain tricky concepts using terms they'll understand. For example: "I'm a fifth grader who loves baseball. Can you explain photosynthesis to me in terms I'll understand?").

[81] *See, e.g.,* Carolyn Stott, To Gen AI or not to Gen AI: Collaborating with student partners on the two-lane assessment approach (Dec. 18, 2025), https://educational-innovation.sydney.edu.au/teaching@sydney/interactive-oral-assessment-in-practice/. ("Working in discipline teams, [faculty and students at the University of Sydney School of Languages and Cultures (SLC)] redesigned learning and teaching activities across the four language competencies (listening, speaking, reading and writing) that could be enhanced by the incorporation of gen AI […] Assessment tasks were also redesigned to be either open or secure. The teams collaborated to produce a set of SLC student guidelines for gen AI use, as well as some SLC staff guidelines for gen AI use.").

students' use of GenAI as a study tool has often changed student outcomes for the better.[82]

46. In parallel, instructors are expanding the use of oral and real-time assessments. For example, the University of Sydney has described its success adopting interactive oral assessments, which consist of short, structured conversations between a student and an instructor, conducted live or via synchronous video. During these sessions, students are required to discuss and clarify their submitted work, allowing instructors to evaluate whether students can explain their reasoning and demonstrate genuine understanding of the material.[83]

47. Many institutions have revised their honor codes and academic integrity rules to address concerns around the improper use of AI and LLMs. These rules often address plagiarism concerns and include a requirement of formally disclosing any form of AI assistance. For example, Syracuse University, who partnered with Anthropic and distributes Claude enterprise licenses to its students,[84] maintains rules for AI-use specific to each course.[85] Professors can include language which either entirely bans the use of LLMs, restricts the use of LLMs to certain assignments, or allows the use of LLMs as long as the student discloses its use.[86]

48. State education departments, accrediting bodies, and international organizations have also issued guidance on how AI may be integrated in education, which signals a broader

---

[82] *See, e.g.,* Shuzchen Chen and Alan Cheung, Effect of generative artificial intelligence on university students learning outcomes: A systematic review and meta-analysis (version of record Nov. 3, 2025), https://doi.org/10.1016/j.edurev.2025.100737. ("This meta-analysis comprehensively incorporates 57 studies and 97 estimations and confirms that Gen-AI has a positive and large effect on university students' learning outcomes, particularly large effect sizes in language skills, academic achievement, affective-motivational status, and a medium effect size in higher-order thinking.").

[83] Leanne Stevenson, Benjamin Miller and Clara Sitbon, Interactive Oral Assessment in practice (June 3, 2025), https://educational-innovation.sydney.edu.au/teaching@sydney/interactive-oral-assessment-in-practice/. ("Interactive Oral Assessments (IOAs) are defined as secure assessments in the new Sydney assessment framework. With the rise of generative AI, concerns about academic integrity have grown. One way to address this is by using more in-person assessments, like exams, oral exams, and IOAs, which help confirm that students are doing their own work").

[84] *See* Syracuse University, Getting Started with Claude Enterprise, https://its.syr.edu/getclaude/.

[85] *See* Syracuse University, Syllabus Requirements, https://teachingexcellence.syr.edu/academic-integrity/syllabus-requirements/.

[86] *See* Syracuse University, Syllabus Requirements, https://teachingexcellence.syr.edu/academic-integrity/syllabus-requirements/. ("Option 1: Zero tolerance for artificial intelligence use… Option 2: Some Artificial intelligence use… Option 3: Open Artificial Intelligence use").

institutional shift toward governance,[87] adaptation, and setting of standards, rather than passive inaction. Having consulted with the Stanford Center for Teaching and Learning during the development of their AI Teaching Strategies,[88] I can speak to the guidelines that have been given to Stanford instructors on the different ways to use AI in teaching. These guidelines lay out pedagogical strategies that apply across all teaching approaches to AI, with template rules for instructors to use as a starting point for fine-tuning their class rules based on their judgement. The Stanford Center for Teaching and Learning further formulated three educational approaches to the use of GenAI in teaching and learning, labelled as "AI-assigned", "AI-limited" and "AI-prohibited" approaches.

49. In addition to these specific strategies, my colleagues at Stanford have also developed and begun implementing an Artificial Intelligence Teaching Guide, with modules for instructors that focus on exploring the instructional uses of AI, analyzing the implications of AI in learning environments and integrating AI into assignments.[89] The Stanford faculty senate and Board on Conduct Affairs are considering the ways in which the Stanford Honor Code intersects with AI. As part of these efforts, the Stanford Office of Community Standards released a "Generative AI Policy Guidance" in 2023.[90] This guidance will evolve over the next few years to encompass new GenAI tools and ensure these tools are used to meet key learning objectives.

50. Finally, while I agree with Professor Zhao's claim that AI-enabled cheating is a cause for concern, I also believe that the potential for AI to democratize access to knowledge in communities facing barriers to high-quality educational resources can outweigh these risks, when AI tools are supported by good governance. As discussed above, instructors in under-resourced schools can already rely on strategies such as adjusting testing

[87] Many states are implementing state-specific playbooks outlining how schools can adopt AI safely and effectively. For example, the North Carolina Department of Public Instruction has issued step-by-step guidance on how schools can integrate AI appropriately, including using AI as a "co-pilot" for certain assessments to enhance creativity. *See* Alyson Klein, EducationWeek, State Outlines Guidance for Different Levels of AI Use in Classrooms (Jan. 29, 2024), https://www.edweek.org/technology/state-outlines-guidance-for-different-levels-of-ai-use-in-classrooms/2024/01. Similarly, California's guidelines emphasize using AI to enhance peer collaboration and complement educators, while Washington's guidance promotes a "Human-AI-Human" (H AI H) model that prioritizes human inquiry and empowerment while incorporating AI assistance. *See* AI for Education, State AI Guidance for K–12 Schools (updated Oct 28, 2025), https://www.aiforeducation.io/ai-resources/state-ai-guidance.

[88] *See* Stanford University, Center for Teaching and Learning, AI Teaching Strategies (Sep. 2, 2025), https://ctl.stanford.edu/aimes/ai-teaching-strategies.

[89] Stanford University, Teaching Commons, Artificial Intelligence Teaching Guide (2023), https://teachingcommons.stanford.edu/teaching-guides/artificial-intelligence-teaching-guide.

[90] Stanford University, Office of Community Standards, Generative AI Policy Guidance (Feb. 16, 2023), https://communitystandards.stanford.edu/generative-ai-policy-guidance.

frameworks to directly evaluate students' understanding of subject matter without discouraging the responsible use of AI by students who might not otherwise have access to high-quality learning tools. In my Opening Report and in Section III.A below, I highlight several ways in which LLMs like Claude are unlocking value in the education space, and how, once the oversight and governance frameworks improve, those benefits will continue to materialize.

51. Overall, none of the challenges raised by Professor Zhao make me any less enthusiastic about LLMs or the transformational value they bring to education. As such, I conclude that, as with any other transformational technology, risks do exist, but the risks mentioned by Professor Zhao are neither novel nor unmanageable. These risks can be mitigated through well-established institutional responses, many of which are already being implemented. Educational institutions are actively adapting assessment design, academic integrity frameworks, and governance structures to account for the availability of AI tools, demonstrating that LLMs can be integrated into education in a manner that preserves academic rigor while expanding access to high-quality learning opportunities.

B. Professor Zhao misrepresents Anthropic's mission and safety-first approach to LLM development

52. Professor Zhao's characterization of Claude overlooks the stated mission of Anthropic as well as the substantial controls, safeguards, and governance mechanisms of the company.[91] Anthropic develops and deploys LLMs with an explicit focus on managing the risks associated with advanced AI systems.[92] Rather than delaying or halting development due to these risks, Anthropic takes a governance-oriented approach that emphasizes deliberate system design, oversight, and continuous evaluation. A range of controls and safeguards are deployed to identify and mitigate potential harms, including model evaluation, red-teaming, usage restrictions, and ongoing safety research. These controls, safeguards and governance mechanisms are not static, but rather, are ever-evolving in response to new information about their strengths and weaknesses.[93] This approach

[91] Zhao Opening Report, ¶ 157. ("[M]any of Anthropic's justifications for its products and potential benefits for humanity remain unrealized years after its first release of Claude. All this stands in stark contrast to Anthropic's stated mission to help mankind through AI, and its goal of differentiating itself from its competitors by more ethical design and practice.").

[92] Deposition of Dario Amodei (Jan. 8, 2026) at 146:6-9. ("The whole purpose of Anthropic's mission is to identify the risks and act in such a way that reduces those risks, to develop the technology in a different and alternative way.").

[93] Deposition of Dario Amodei (Jan. 8, 2026) at 64:12-13 ("We are always upgrading and improving our safeguards.") *See also* Anthropic_0000570027, which notes Anthropic's framework for building safe and trustworthy AI agents;

reflects my view that the challenges associated with this technology can be addressed through responsible development practices and iterative governance.

53. Consistent with this safety-first mission, Anthropic has made a deliberate choice to conduct and publish safety research, including internal evaluations and stress-testing of its models.[94] Professor Zhao cites these disclosures as further evidence that LLMs pose significant dangers.[95] This transparency allows risks to be scrutinized and mitigated rather than obscured before they materialize in real-world deployment. The intended purpose of publishing its safety research, Anthropic has explained, is to improve its own systems and to enable other researchers, developers, and institutions to build upon and strengthen existing safety techniques.[96] It encourages collective learning and to accelerate the development of more effective safeguards. In my opinion, this kind of transparency has the potential to enhance and accelerate the pace of safety developments in GenAI.

54. Anthropic's practices around transparency, red-teaming, and safety disclosures exemplify responsible governance and demonstrate the company's commitment to safety through its deliberate integration of risk mitigation directly into model development and deployment.[97] I was particularly impressed with Anthropic's commitment to AI Safety specifically through their goal to "ignite a race to the top on safety," in being a safety-first company which tries to "constantly set the industry bar for AI safety and security and drive

---

and Anthropic_0000570384, which highlights Anthropic's work on interpreting the internal reasoning of large language models.

[94] *See* Anthropic, Agentic Misalignment: How LLMs could be insider threats (June 20, 2025), https://www.anthropic.com/research/agentic-misalignment. ("[O]ur red-teaming efforts are targeted at finding evidence of risks now to provide early warning of the potential for future harm and help us develop mitigations in advance of risks appearing in real-world situations. To do this, we conduct experiments in artificial scenarios designed to stress-test AI boundaries and understand how models might behave when given more autonomy. While it seems unlikely that any of the exact scenarios we study would occur in the real world, we think they are all within the realm of possibility…").

[95] Zhao Opening Report, ¶ 165, ("In a recent paper, Anthropic disclosed that during a stress-test of a number of AI models, including its own Claude, the models resorted to "malicious insider behaviors when that was the only way to avoid replacement or achieve their goals— including blackmailing officials and leaking sensitive information to competitors."").

[96] *See* Anthropic, Agentic Misalignment: How LLMs Could be Insider Threats (June 20, 2025) https://www.anthropic.com/research/agentic-misalignment. (The purpose as Anthropic describes it, is to "identify potentially risky agentic behaviors before they cause real harm …[and to help others]…identify ways to improve current safety techniques"). *See also* Anthropic, Core Views on AI Safety: When, Why, What, and How (Mar. 8, 2023), https://www.anthropic.com/news/core-views-on-ai-safety.

[97] *See* Deposition of Dario Amodei (Jan. 8, 2026) at 146:22-147:5. ("Anthropic is substantially more transparent, as in the Anderson Cooper interview that you described, in being honest about these -- these -- being honest about the risks of these models in disclosing the risks of these models. Anthropic has also advocated for regulation, for legislation, forcing companies to disclose the risk of these models and take safety measures against the risk of these models."). *See also* Altman Opening Report, ¶ 20.

others to do the same."[98] A central element of this approach is Anthropic's Constitutional AI framework, in which models are trained and guided according to explicit principles that prioritize safety, helpfulness, and respect for human values and ethics.[99] To emphasize its ongoing commitment to AI safety, Anthropic has evolved its model training principles to ensure that Claude's new constitution plays a more central role in its training.[100] Anthropic employs both internal and external red-teaming to proactively identify vulnerabilities and potential misuse scenarios.[101] The results of this red-teaming are then provided to the company's model fine-tuning and safeguards teams to assess how the findings should be integrated into model training, model development, and the deployment of safety and enforcement strategies.[102] Anthropic has further demonstrated this priority by publicly committing to de-deploy or restrict its own models if they are found to pose significant risks, and by collaborating with government bodies such as the U.S. Center for

---

[98] As part of their goal to make AI systems that users can rely on, Anthropic has published research on both the opportunities and risks of AI as well as the ways in which Anthropic is working to establish industry precedent for managing those risks. One of the key principles that stood out to me was the "Ignite a race to the top on safety" whereby Anthropic outlined how building reliable, trustworthy, and secure systems was seen not just as a goal but a responsibility. *See* Anthropic_0000570351.

[99] *See* Anthropic_0000594126, Constitutional AI, What is Constitutional AI? (accessed Dec. 16, 2025), https://constitutional.ai/. ("The key ideas behind Constitutional AI are aligning an AI's behavior with a 'constitution' defined by human principles, using techniques like self-supervision and adversarial training, developing constrained optimization techniques, and designing training data and model architecture to encode beneficial behaviors."). *See also* Anthropic, Claude's Constitution, Our vision for Claude's character, https://www.anthropic.com/constitution. ("Anthropic wants Claude to be genuinely helpful to the people it works with or on behalf of, as well as to society, while avoiding actions that are unsafe, unethical, or deceptive.").

[100] Claude's new constitution is built around the principles of safety, ethicality, compliance, and helpfulness. *See* Anthropic, Claude's new constitution (Jan. 21, 2026), https://www.anthropic.com/news/claude-new-constitution ("Our previous Constitution was composed of a list of standalone principles. We've come to believe that a different approach is necessary. We think that in order to be good actors in the world, AI models like Claude need to understand *why* we want them to behave in certain ways, and we need to explain this to them rather than merely specify *what* we want them to do. … We use the constitution at various stages of the training process. This has grown out of training techniques we've been using since 2023, when we first began training Claude models using Constitutional AI. Our approach has evolved significantly since then, and the new constitution plays an even more central role in training.").

[101] *See* Altman Opening Report, ¶ 20. Professor Zhao is also aware and even acknowledges that Anthropic employs red-teaming to prevent unwanted output. See Zhao Opening Report, ¶ 108.

[102] *See* Anthropic, Anthropic's Transparency Hub: Voluntary Commitments (Aug. 28, 2025), https://www.anthropic.com/transparency/voluntary-commitments. ("Results from red teaming are provided to our model finetuning and safeguards teams to assess for integration back into model training, model development, and deployment of safety and enforcement strategies. For example, we used feedback from child safety experts at Thorn around signals often seen in child grooming to update our classifiers, enhance our Usage Policy, fine-tune our models, and incorporate these signals into testing of future models."). *See also* Anthropic_0000569592. ("This is where our Safeguards team comes in: we identify potential misuse, respond to threats, and build defenses that help keep Claude both helpful and safe.") and ("We partner with external domain experts to identify areas of concern, and then stress-test these concerns against our policies by assessing the output of our models under challenging prompts. Our partners include experts in terrorism, radicalization, child safety, and mental health. The findings from these stress tests directly shape our policies, training, and detection systems. For example, during the 2024 U.S. election, we partnered with the Institute for Strategic Dialogue to understand when Claude might provide outdated information. We then added a banner pointing Claude.ai users seeking election information to authoritative sources like TurboVote.").

AI Standards and Innovation ("CAISI") and the UK AI Security Institute ("AISI").[103] Through these collaborations, Anthropic leverages expertise in areas including cybersecurity, intelligence analysis, and threat modeling to strengthen security measures and improve model performance. Anthropic's efforts reinforce the company's belief and commitment to the fact that the risks associated with advanced LLM systems can and should be addressed through governance, research, and responsible market adaptations.

55. In my opinion, any assessment of LLM-related risks must be evaluated in the context of their surrounding safety and governance practices. LLMs should not be judged monolithically; rather, the nature and magnitude of potential risks and benefits vary substantially depending on how, where, and by whom they are deployed. In my own experience, the safety considerations for LLMs that directly impact patient care are materially different, and more rigorous, than those for LLMs used to support scientific biomedical discovery, where risks are not zero but are generally lower and more readily managed through existing research protocols and guidelines.

56. This context-sensitive approach allows for a more precise evaluation, recognizing that in some applications risks may be limited and manageable, while in others they may be more serious and require stricter safeguards, governance, or oversight. Importantly, the presence of risks, while it does warrant governance, does not diminish the technology's value. Rather, it underscores the importance of aligning deployment practices with appropriate controls. Based on my review, Anthropic's approach to safety research, transparency, and governance reflects an understanding that managing risks and enabling benefits are not opposing objectives, but interdependent ones. Anthropic's emphasis on safeguards, principled model design, red-teaming, and engagement with external experts and institutions is consistent with how responsible companies have historically approached other transformative technologies and aligns squarely with the historical trajectory of technologies that ultimately deliver significant net social benefits, as I discuss in greater detail below.

[103] *See* Anthropic, Strengthening our safeguards through collaboration with US CAISI and UK AISI (Sep. 12, 2025), https://www.anthropic.com/news/strengthening-our-safeguards-through-collaboration-with-us-caisi-and-uk-aisi. ("Over the past year, we've collaborated with the US Center for AI Standards and Innovation (CAISI) and UK AI Security Institute (AISI), government bodies established to measure and improve the security of AI systems.") and ("[Goverments'] feedback helps us improve our security measures so our systems can withstand some of the most sophisticated attempts at misuse.").

III. PROFESSOR ZHAO IGNORES THE SIGNIFICANT BENEFITS AND TRANSFORMATIVE NATURE OF LLMS

57. Critically, Professor Zhao's analysis gives little weight to the significant transformative benefits associated with LLMs.[104]

58. Professor Zhao's analysis of LLM-related risks does not engage with the central question raised by transformative technologies of this nature: whether the challenges identified meaningfully constrain or negate the technology's capacity to generate broad transformative societal value. Rather than assessing how the risks he describes interact with, limit, or impede the transformative potential of LLMs such as Claude, Professor Zhao treats the existence of early-stage harms and deployment challenges as if they were inherent and impossible to manage. This framing is analytically incomplete. The relevant inquiry is not whether LLMs present challenges, which they do, but instead whether those challenges bear weight on the technology's ability to deliver durable, large-scale benefits across domains.[105]

59. As I explain below, the risks identified by Professor Zhao do not diminish, and in many cases are orthogonal to the substantial and growing evidence, which I highlight in my Opening Report, that LLMs are already transforming education, science, medicine, creative work, and public-interest applications.

---

[104] For example, recent work provides evidence that AI can be understood as a "General Purpose Technology" and as a method of invention, with the potential to raise overall productivity levels in the economy. *See* Martin Neil Baily et al., Generative AI at the Crossroads: Light Bulb, Dynamo, or Microscope? (June 27, 2025), https://www.federalreserve.gov/econres/feds/files/2025053pap.pdf. ("We conclude there is substantial evidence that genAI is both a GPT [general-purpose technology] and an IMI [inventions of methods of invention], an encouraging sign for higher productivity in the future."). Similarly, there is evidence that Generative AI has already improved productivity of workers. For example, see Alexander Bick et al., Federal Reserve Bank of St. Louis, On the Economy Blog, The State of Generative AI Adoption in 2025 (Nov. 13, 2025), https://www.stlouisfed.org/on-the-economy/2025/nov/state-generative-ai-adoption-2025. ("Our survey asks generative AI users how many additional hours they would need to work to get the same amount done if they had not had access to generative AI. We pooled the data from our February, May and August 2025 surveys, and then calculated the aggregate time savings. Factoring in all workers, including nonusers, workers report generative AI time savings equivalent to 1.6% of all work hours. When we feed these estimates into a standard aggregate production model, this suggests that generative AI may have increased labor productivity by up to 1.3% since the introduction of ChatGPT.").

[105] Melvin Kranzberg, a renowned historian of technology, formulated Kranzbeg's First Law in his 1986 paper, Technology and History: "Kranzberg's Laws", wherein he underscores the need for societal oversight given the dual-uses of technology. *See* Melvin Kranzberg, Technology and History: Kranzberg's Laws (July 1986), https://www.jstor.org/stable/3105385. ("Kranzberg's First Law reads as follows: Technology is neither good nor bad; nor is it neutral. By that I mean that technology's interaction with the social ecology is such that technical developments frequently have environmental, social, and human consequences that go far beyond the immediate purposes of the technical devices and practices themselves, and the same technology can have quite different results when introduced into different contexts or under different circumstances.").

A. Professor Zhao ignores the ways in which Claude is benefiting education

60. Although Zhao raises concerns about the risks of LLMs in the education space, he makes no mention of the transformational benefits introduced by the tools. In my Opening Report, I discuss how LLMs such as Claude are already delivering measurable educational benefits in several respects, including improved student support and learning outcomes, with Claude serving as a tutoring aid by generating study guides, explaining complex concepts, and assisting students with problem-solving in subjects such as mathematics.[106,107] I also describe how LLMs contribute to the democratization of access to knowledge by expanding access to educational support and expertise in underserved communities, including through social enterprises such as Rising Academics, which uses Claude to power a WhatsApp-based mathematics tutor serving more than 150,000 students across multiple countries.[108]

61. In addition, I discuss how LLMs can increase efficiency in teaching and assessment by automating repetitive tasks such as grading and lesson planning and by enabling real-time feedback through educational technology applications, allowing educators to devote more time to direct instruction and student engagement rather than administrative work.[109] One example, MagicSchool,[110] an AI copilot for educators and students, uses Claude to transform education by streamlining lesson planning, providing personalized student feedback, and creating an abundance of new and engaging content for students. The platform's ability to save teachers time has improved student outcomes.[111]

B. Claude is transforming science and medicine

62. Based on my experience working at the intersection of medicine and AI, I have experienced firsthand the valuable changes that Claude, alongside other LLM tools, has already brought to scientific fields, including its potential to accelerate scientific discovery,

---

[106] Altman Opening Report, ¶¶ 87-89.

[107] As I discuss in my Opening Report, lower development costs enabled by LLMs have accelerated the deployment of educational tools that provide individualized, real-time learning support. *See* Altman Opening Report ¶ 120.

[108] Altman Opening Report, ¶ 90.

[109] Altman Opening Report, ¶ 93.

[110] *See* Anthropic_0000569970.

[111] MagicSchool has had tremendous reach, already improving learning accuracy, education alignment, as well as personalizing learning for students and reducing teacher burnout. *See* Anthropic_0000569970. ("With Claude, MagicSchool supports 3 million educators actively using the platform, over 100 million AI-powered engagements, 5,500+ school & district partnerships, and 1 million students though teacher-guided AI interactions.").

improve medical research and clinical care, enhance productivity, and expand access to information and expertise. In my Opening Report, I described seven specific use cases of Claude in this domain, and I expect the returns from Claude to continue to accelerate as models further evolve.[112]

63. In October 2025, Anthropic introduced Claude for Life Sciences as a targeted offering to help researchers use LLM technology for the advancement of scientific discovery.[113] Researchers can leverage Claude through the end-to-end research and discovery process – from carrying out literature reviews to developing hypotheses, analyzing data, and drafting regulatory submissions – improving not only efficiency but pushing the frontier of complex drug discovery.[114]

64. Another notable example is Biomni, a Stanford designed, Claude-powered general-purpose biomedical agent that is able to access hundreds of databases, software packages and protocols which were otherwise fragmented and difficult to analyze.[115] Biomni integrates complex LLM reasoning to enable biomedical workflows, as well as to dynamically compose and execute critical research tasks. Early trials have already shown that Biomni helps scientists dramatically enhance research output and generate testable hypotheses.[116] Also at Stanford University, some of my recent work has introduced how machine learning models, such as Claude, can analyze protein structures to predict how specific regions react to biological function, helping scientists to more quickly understand how proteins behave and interact. More specifically, this research has shown that AI systems can systematically identify important functional regions within proteins,

[112] In my Opening Report, I describe how Claude has delivered measurable benefits when deployed as a scientific assistant in public and civic domains. Drawing on multiple case studies, I show that Claude accelerates research and analysis and provides technical capabilities that enable real-time, evidence-based decision-making, thereby fundamentally transforming public and civic decision-making processes. *See* Altman Opening Report, §§ III.A and III.C.

[113] *See* CNBC, Anthropic launches Claude Life Sciences to give researchers an AI efficiency boost (Oct. 20, 2025), https://www.cnbc.com/2025/10/20/anthropic-claude-life-sciences-research-ai.html.

[114] *See* Aragon Research, Anthropic Enters Life Sciences with Claude for Research Acceleration (Oct 20, 2025), https://aragonresearch.com/anthropic-enters-life-sciences/. ("By embedding its technology into the daily workflows of researchers, the company aims to bring measurable efficiency gains to one of the world's most data-intensive industries—while setting the stage for AI's deeper role in the future of scientific innovation.").

[115] *See* Anthropic, How scientists are using Claude to accelerate research and discovery (Jan. 15, 2026), https://www.anthropic.com/news/accelerating-scientific-research.

[116] *See* Anthropic_0000593996, Kexin Huang et al., Biomni: A General-Purpose Biomedical AI Agent (June 2, 2025), https://doi.org/10.1101/2025.05.30.656746. ("Biomni envisions a future where virtual AI biologists operate alongside and augment human scientists to dramatically enhance research productivity, clinical insight, and healthcare. Biomni is ready to use at https://biomni.stanford.edu, and we invite scientists to explore its capabilities, stress test its limits, and co-create the next era of biomedical discoveries.").

accelerating the interpretation of complex biological data and supporting downstream applications such as drug discovery, filling a computational gap that existed with previous computational methods.[117]

65. Other notable research and policy organizations like the Broad Institute,[118] a leading genomics research center affiliated with MIT and Harvard, have already drawn on the benefits of LLMs to expand their science research and analysis capabilities. Today, The Broad Institute uses AI agents (built on Claude) to sift through genomic datasets, run analyses, and draft reports, acting as a research assistant. According to Broad Institute's Head of Data Science, Claude unlocks "entirely new scale" to how scientists explore questions, effectively turbocharging research productivity.[119] Other companies have also used Claude in their services to allow for non-linear problem solving, improve research workflows and accelerate scientific discovery.[120]

66. In the biomedical research setting, including oncology and rare disease research, Claude and other LLMs have been used to support medical development and scientific research. For example, Claude introduces capabilities in genomic and pharmaceutical data analysis, discovery, and literature review to improve accessibility in scientific data interpretation and enable faster delivery of life-saving medicine and therapies.[121] These capabilities depend

---

[117] *See* Alexander Derry et al., Collapse: A representation learning framework for identification and characterization of protein structural sites (Dec. 15, 2022), https://doi.org/10.1002/pro.4541; and Alexander Derry et al., Protein functional site annotation using local structural embeddings (June 25, 2025), https://doi.org/10.1101/2023.10.13.562298.

[118] *See* Broad Institute, This is Broad, https://www.broadinstitute.org/about-us.

[119] *See* Anthropic, Claude for Life Sciences (Oct. 20, 2025), https://www.anthropic.com/news/claude-for-life-sciences. ("Broad Institute scientists purse the most ambitious questions in biology and medicine, creating tools to empower scientists everywhere. We're working with Manifold on Terra Powered by Manifold. AI agents build on Clade enable scientists to work at entirely new scale and efficiency, exploring scientific domains in previously impossible ways.").

[120] *See, e.g.,* Claude, Novo Nordisk accelerates clinical documentation and drug development with Claude, https://claude.com/customers/novo-nordisk. ("Our work with Anthropic and Claude has set a new standard—we're not just automating tasks, we're transforming how medicines get from discovery to the patients who need them"). *See also* Claude, Accelerate science, from discovery through translation, https://claude.com/solutions/life-sciences. ("Claude's commitment to mechanistic interpretability provides a safe, secure, and steerable model to empower our science-first approach.").

[121] According to Sanofi's Chief Digital Officer, Claude (integrated with the company's knowledge bases) is now used daily by most of Sanofi's 100,000+ employees via an internal Concierge app. *See* Anthropic, Claude for Life Sciences (Oct. 20, 2025), https://www.anthropic.com/news/claude-for-life-sciences. ("We're seeing efficiency gains across the value-chain, while our enterprise deployment has enhanced how teams work. This collaboration with Anthropic augments human expertise to deliver life-changing medicines faster to patients worldwide."); Genmab's Global Head of Data, Digital and AI stated that Claude will help bring life-saving therapies to market, faster. *See* Anthropic, Claude for Life Sciences (Oct. 20, 2025), https://www.anthropic.com/news/claude-for-life-sciences. ("We see tremendous potential in Claude streamlining how we bring drugs to market. The ability to pull from clinical data

on training on scientific and medical sources, enabling Claude to recognize patterns, terminology, and contextual relationships. When used by trained professionals, Claude can connect insights across various data sources, resulting in novel hypotheses thereby allowing medical professionals to focus their time on experimental design and patient care. In this capacity, Claude has already delivered tangible public benefits through its role in accelerating diagnostics, medical research and therapy development.

C. Claude has made a positive impact in creative sectors and public civil engagement

67. In my Opening Report, I also discuss how Claude has made a positive impact on creative sectors in multiple ways including: (i) streamlining repetitive tasks and freeing up time on actual creative tasks; (ii) creating new content; and (iii) expanding content creators' reach and impact.[122]

68. Notably, artists and creatives have found success using Claude to manage the "less creative" aspects of running an art business, allowing them to devote more time to creative work.[123] Many independent artist users of Claude, from music producers to amateur songwriters such as myself, also engage the general assistance of LLMs for brainstorming and transforming their day-to-day approach to content creation.[124] When it comes to bridging the gap across different stages of the creative process in the film, television and music industries, Claude has also crucially helped creators amplify the visibility of their content by increasing their exposure through search output and recommendations.[125]

69. As I discuss in my Opening Report, Claude played a critical role in civil efforts such as the Tahoe Lead Removal Project by identifying relevant information from decades old technical and legal documentation that would have been a barrier to informed action. Claude's ability to synthesize this information aided organizers in mobilizing community support ultimately leading to the removal of six miles of toxic lead cable from the bottom

sources and create GxP-compliant outputs will help us bring life-changing cancer therapies to patients faster while maintaining the highest quality standards.").

[122] Altman Opening Report, § III.D.

[123] Altman Opening Report, ¶¶ 108-110.

[124] Altman Opening Report, ¶¶ 111-112.

[125] Altman Opening Report, ¶¶ 114-116.

of Lake Tahoe.[126] Claude also contributed to the preservation of the Rwandan gorilla population as part of the Tehanu Project, by enabling a scalable and data driven approach to wildlife preservation. Claude helped researchers better understand the species' behavior and needs, as well as quantify the economic value of this vulnerable population to increase conservation awareness.[127] These are just two of many examples that demonstrate the ways in which the transformative benefits of Claude extend into public domains such as conservation, climate research, and policy development.

IV. CONCLUSION

70. After reviewing the Zhao Opening Report, my opinions expressed in my Opening Report remain unchanged. While there are some points of agreement between Professor Zhao and me, these shared views only underscore the inadequacy of his assessment with respect to both the risks of LLMs and of the transformative benefits of LLMs.

71. Professor Zhao repeatedly characterizes a range of downstream social and institutional challenges as intrinsic "risks" of AI. This framing conflates the properties of the technology itself with the policy, regulatory, and governance mechanisms that shape its real-world deployment. Historically, such challenges are neither unusual nor indicative of net social harm for transformative technologies but rather signal the need for thoughtful adaptation, institutional innovation, and shared responsibility. For the reasons stated throughout my Opening Report and this rebuttal report, it is still my opinion that Claude has already demonstrated significant public benefits across several domains, and that Claude is a transformative, foundational technology.

72. I provide the opinions herein based on the information currently available. I may supplement or revise my opinions if new information arises, or in response to other expert opinions. I may also use demonstrative or visual aids to explain my views.

[126] Altman Opening Report, § III.C ¶100. ("The Tahoe Lead Removal Project is another example of how Claude accelerated the research process to help a group of volunteers achieve the removal of six miles of decommissioned lead-clad telecommunication cables from Lake Tahoe.").

[127] Altman Opening Report, § III.C ¶ 103. ("Another example comes from the Tehanu Project, which in 2024 set out to advance conservation science for Rwanda's mountain gorillas.").



Dr. Russ Altman

APPENDIX A: MATERIALS CONSIDERED

To the extent not already included below, the following list incorporates by reference as though set forth fully herein the identification information for all sources cited or considered in the Rebuttal Report of Russ Altman, dated January 23, 2026.

Table 1: Materials Considered

#	Internal Documents
1	2025-11-07 Rao, Vinay Deposition Transcript 30(b)(1)
2	2025-11-07 Rao, Vinay Deposition Transcript 30(b)(6)
3	2025-11-12 Amodei, Daniela Deposition Transcript
4	2026-01-08 Amodei, Dario Deposition Transcript
5	Anthropic_0000568974
6	Anthropic_0000569549
7	Anthropic_0000569592
8	Anthropic_0000569650
9	Anthropic_0000569688
10	Anthropic_0000569947
11	Anthropic_0000569970
12	Anthropic_0000570027
13	Anthropic_0000570123
14	Anthropic_0000570154
15	Anthropic_0000570159
16	Anthropic_0000570317
17	Anthropic_0000570340
18	Anthropic_0000570351
19	Anthropic_0000570384
20	Anthropic_0000570396
21	Anthropic_0000570411
22	Anthropic_0000570419
23	Anthropic_0000570435
24	Anthropic_0000593638
25	Anthropic_0000593996
26	Anthropic_0000594126
27	Anthropic_0000595304
	Expert Sources
28	2025.12.19 - Expert Report of Dr. Ben Zhao - HC- AEO
29	2025.12.19 Expert Report of Russ B. Altman - HC-AEO

Academic Sources

30 Alexander Derry et al., Collapse: A representation learning framework for identification and characterization of protein structural sites (Dec. 15, 2022)
31 Alexander Derry et al., Protein functional site annotation using local structural embeddings (June 25, 2025)
32 Artem A. Trotsyuk, et al., No Time to Lose: Stronger Artificial Intelligence Regulation Is Needed in the U.S., Especially for Biomedicine (Sep. 13, 2023)
33 Artem A. Trotsyuk, et al., Toward a framework for risk mitigation of potential misuse of artificial intelligence in biomedical research (Nov. 26, 2024)
34 David L. Morton, Jr, Reviewing the history of electric power and electrification (June 1, 2002)
35 Edmund Pickering and Clancy Schuller, Chegg's Growth, Response Rate, and Prevalence as a Cheating Tool: Insights From an Audit within an Australian Engineering School (July 31, 2024)
36 J. Benjamin Hurlbut, Taking responsibility: Asilomar and its legacy, Science (Jan. 30, 2025)
37 Kexin Huang et al., Biomni: A General-Purpose Biomedical AI Agent (June 2, 2025)
38 Martin Neil Baily et al., Generative AI at the Crossroads: Light Bulb, Dynamo, or Microscope? (June 27, 2025)
39 Melvin Kranzberg, Technology and History: Kranzberg's Laws (July 1986)
40 Rebecca M. Howard and Laura J. Davies, Plagiarism in the Internet Age (Mar. 2009)
41 Sameer Jumar and Kevin B. Moore, The Evolution of Global Positioning System (GPS) Technology (Mar. 2002)
42 Shuzchen Chen and Alan Cheung, Effect of generative artificial intelligence on university students learning outcomes: A systematic review and meta-analysis (version of record Nov. 3, 2025)
43 Thomas P. Hughes, Harold P. Brown and the Executioner's Current: an Incident in the AC-DC Controversy (1958)
44 Timothy F. Bresnahan and M. Trajtenberg, General purpose technologies 'Engines of growth'? (Jan. 1995)

Public Sources

45 Aaron Gordon, How we ended up with all this junk mail (Oct. 15, 2020), https://www.vice.com/en/article/how-we-ended-up-with-all-this-junk-mail/
46 Abed Serhan, Engineerine, The Innovation That Saved Millions: The Story of the Three-Point Seatbelt, (Jan. 6, 2025), https://engineerine.com/three-point-seatbelt-by-nils-bohlin-volvo/
47 AI for Education, State AI Guidance for K–12 Schools (updated Oct 28, 2025), https://www.aiforeducation.io/ai-resources/state-ai-guidance
48 Alexander Bick et al., Federal Reserve Bank of St. Louis, On the Economy Blog, The State of Generative AI Adoption in 2025 (Nov. 13, 2025), https://www.stlouisfed.org/on-the-economy/2025/nov/state-generative-ai-adoption-2025

49 Alyson Klein, EducationWeek, State Outlines Guidance for Different Levels of AI Use in Classrooms (Jan. 29, 2024), https://www.edweek.org/technology/state-outlines-guidance-for-different-levels-of-ai-use-in-classrooms/2024/01

50 Anderson Cooper and Nichole Marks, 60 Minutes, CBS News, Why Anthropic CEO Dario Amodei spends so much time warning of AI's potential dangers (Nov. 16, 2025) https://www.cbsnews.com/news/anthropic-ceo-dario-amodei-warning-of-ai-potential-dangers-60-minutes-transcript/

51 Anthropic, Agentic Misalignment: How LLMs Could be Insider Threats (June 20, 2025) https://www.anthropic.com/research/agentic-misalignment

52 Anthropic, Anthropic's Transparency Hub: Voluntary Commitments (Aug. 28, 2025), https://www.anthropic.com/transparency/voluntary-commitments

53 Anthropic, Building effective agents (Dec. 19, 2024), https://www.anthropic.com/engineering/building-effective-agents

54 Anthropic, Claude for Life Sciences (Oct. 20, 2025), https://www.anthropic.com/news/claude-for-life-sciences

55 Anthropic, Claude's Constitution, Our vision for Claude's character (Jan. 21, 2026), https://www.anthropic.com/constitution

56 Anthropic, Claude's new constitution (Jan. 21, 2026), https://www.anthropic.com/news/claude-new-constitution

57 Anthropic, Core Views on AI Safety: When, Why, What, and How (Mar. 8, 2023), https://www.anthropic.com/news/core-views-on-ai-safety

58 Anthropic, Detecting and Countering Misuse of AI: August 2025 (Aug. 27, 2025) https://www.anthropic.com/news/detecting-countering-misuse-aug-2025

59 Anthropic, How scientists are using Claude to accelerate research and discovery (Jan. 15, 2026), https://www.anthropic.com/news/accelerating-scientific-research

60 Anthropic, Introducing Claude for Education (Apr. 2, 2025), https://www.anthropic.com/news/introducing-claude-for-education

61 Anthropic, Strengthening our safeguards through collaboration with US CAISI and UK AISI (Sep. 12, 2025), https://www.anthropic.com/news/strengthening-our-safeguards-through-collaboration-with-us-caisi-and-uk-aisi

62 Aragon Research, Anthropic Enters Life Sciences with Claude for Research Acceleration (Oct 20, 2025), https://aragonresearch.com/anthropic-enters-life-sciences/

63 Bree Dusseault and Justin Lee, CRPE, How Districts Are Responding to AI—and What It Means for the '23-'24 School Year (Sep. 2023), https://crpe.org/study-how-districts-are-responding-to-ai-and-what-it-means-for-the-new-school-year/

64 Broad Institute, This is Broad, https://www.broadinstitute.org/about-us

65 Carlo Kopp, Going deep: Understanding the mechanics of deception and disinformation (Feb. 24, 2025), https://lens.monash.edu/the-science-of-disinformation-and-deception/

66 Carolyn Stott, To Gen AI or not to Gen AI: Collaborating with student partners on the two-lane assessment approach (Dec. 18, 2025), https://educational-innovation.sydney.edu.au/teaching@sydney/interactive-oral-assessment-in-practice/

67 Charlotte Libov, New York Times, Calculators in Exam Debated (Dec. 8, 1985), https://www.nytimes.com/1985/12/08/nyregion/calculators-in-exam-debated.html

68 Claude Docs, What's new in Claude 4.5, https://platform.claude.com/docs/en/about-claude/models/whats-new-claude-4-5
69 Claude, Accelerate science, from discovery through translation, https://claude.com/solutions/life-sciences
70 Claude, MagicSchool transforms K-12 education for 3 million educators and their students with Claude, Anthropic_0000569970
71 Claude, Novo Nordisk accelerates clinical documentation and drug development with Claude, https://claude.com/customers/novo-nordisk
72 CNBC, Anthropic launches Claude Life Sciences to give researchers an AI efficiency boost (Oct. 20, 2025), https://www.cnbc.com/2025/10/20/anthropic-claude-life-sciences-research-ai.html
73 Cody Mello-Klien, Northeastern Global News, How a journalism professor got his class to put AI to the test and think like an AI model in the process (June 23, 2025), https://news.northeastern.edu/2025/06/23/faculty-staff/journalism-class-ai-media-industry/
74 College Board, AP Calculus 2007-2008 Professional Development Workshop Materials, https://apcentral.collegeboard.org/media/pdf/ap-sf-calculus-approximation.pdf
75 College Board, AP Calculus: Use of Graphing Calculators, https://apcentral.collegeboard.org/courses/resources/ap-calculus-use-of-graphing-calculators
76 Columbia University, Office of the Provost, Generative AI Policy, https://provost.columbia.edu/content/office-senior-vice-provost/ai-policy
77 Constitutional AI, What is Constitutional AI? (accessed Dec. 16, 2025), https://constitutional.ai/
78 David Nordman, Northeastern Global News, Northeastern and Anthropic to lead in responsible AI innovation in higher education (Apr. 2, 2025), https://news.northeastern.edu/2025/04/02/anthropic-ai-partnership/
79 Debra A. Valentine, Privacy on the Internet: The Evolving Legal Landscape, Federal Trade Commission (Feb. 11, 2000) https://www.ftc.gov/news-events/news/speeches/privacy-internet-evolving-legal-landscape
80 Deni Ellis Bechard, The Slop Cycle—How Every Media Revolution Breeds Rubbish and Art (Nov. 9, 2025), https://www.scientificamerican.com/article/ai-slop-how-every-media-revolution-breeds-rubbish-and-art/
81 Department of Homeland Security, CBD Focus Areas – Biosurveillance (Apr. 10, 2025), https://www.dhs.gov/science-and-technology/biosurveillance
82 Environmental Protection Agency, Significant Discoveries and the History of Radiation Protection (2018), https://www.epa.gov/sites/default/files/2018-12/documents/significant_discoveries_history_radiation_protection-worksheet_rp_1.pdf
83 GPS.gov, Selective Availability, https://www.gps.gov/selective-availability
84 GPS.gov, What can GPS Do, https://www.gps.gov/what-can-gps-do
85 Harsh Singh, Forbes, The Evolution of Education: From Calculators To Generative AI (Nov. 4, 2024), https://www.forbes.com/councils/forbesbusinesscouncil/2024/11/04/the-evolution-of-education-from-calculators-to-generative-ai/

86 History.com, This Day in History, Connecticut Enacts First Speed-Limit Law (Nov. 13, 2009), https://www.history.com/this-day-in-history/may-21/connecticut-enacts-first-speed-limit-law
87 History.com, This Day in History, Federal legislation makes airbags mandatory (May 27, 2025) https://www.history.com/this-day-in-history/September-1/federal-legislation-makes-airbags-mandatory
88 IEEE Standards Association, The National Electric Code and the National Electronic Safety Code – Partners for a Safer Tomorrow (Sep. 2016), https://standards.ieee.org/wp-content/uploads/import/documents/other/nec_nesc_case_study.pdf
89 International Electrotechnical Commission, How & why the IEC was started, https://www.iec.ch/history/how-why-iec-was-started
90 ISO, About ISO, https://www.iso.org/about
91 Jocelyn Gecker, The rise of AI tools forces schools to reconsider what counts as cheating (Sep. 12, 2025), https://apnews.com/article/ai-cheating-school-chatgpt-4f89a552e9093ce2180471b4d4736675
92 Johns Hopkins, New Guidelines for Responsible Use of Generative AI (GenAI) (May 2025), https://it.johnshopkins.edu/featured-articles/new-guidelines-for-responsible-use-of-generative-ai/
93 Kristine De Abreu, The Perils of Early Aviation, ExplorersWeb (Oct. 28, 2022), https://explorersweb.com/the-perils-of-early-aviation/
94 LandAirSea, Missing Persons Statistics: How GPS Revolutionizes Search and Rescue, https://landairsea.com/blogs/consumers/missing-persons-statistics-how-gps-revolutionizes-search-and-rescue
95 Laura Laster, Early Accidents Rates and Dispatchers (2024), https://www.theaviationvault.com/early-accident-rates
96 Leanne Stevenson, Benjamin Miller and Clara Sitbon, Interactive Oral Assessment in practice (June 3, 2025), https://educational-innovation.sydney.edu.au/teaching@sydney/interactive-oral-assessment-in-practice/
97 LegalClarity, Do Electricians Need a License? Requirements and Penalties (July 3, 2025), https://legalclarity.org/do-electricians-need-a-license-requirements-and-penalties/
98 Liam Cope, Engineer Fix, When Did Seatbelts Become Mandatory in Cars? (Jan. 9, 2026), https://engineerfix.com/when-did-seatbelts-become-mandatory-in-cars/
99 Merriam-Webster, Curation, https://www.merriam-webster.com/dictionary/curation
100 Merriam-Webster, Differentiate, https://www.merriam-webster.com/dictionary/differentiate
101 Merriam-Webster, Label, https://www.merriam-webster.com/dictionary/labeling
102 National Institute of Standards and Technology (NIST), Glossary: Risk, https://csrc.nist.gov/glossary/term/risk
103 National Institute of Standards and Technology, Framework for Improving Critical Infrastructure Cybersecurity (Feb. 12, 2014), https://nvlpubs.nist.gov/nistpubs/CSWP/NIST.CSWP.02122014.pdf

104 Neil L. Waters, Why You Can't Cite Wikipedia in My Class Communications of the ACM (Sep. 1, 2007), https://cacm.acm.org/opinion/why-you-cant-cite-wikipedia-in-my-class/

105 NIST AI Resource Center, NIST AI Risk Management Framework (Jan. 26, 2023), https://airc.nist.gov/airmf-resources/airmf/1-sec-risk/

106 Northeastern University, Claude.ai at Northeastern University, https://claude.northeastern.edu/

107 Organization for the Prohibition of Chemical Weapons ("OPCW"), OPCW confirms: All declared chemical weapons stockpiles verified as irreversibly destroyed (July 7, 2023), https://www.opcw.org/media-centre/news/2023/07/opcw-confirms-all-declared-chemical-weapons-stockpiles-verified

108 Perry Chiaramonte, Is It Real or Is It AI? Professors Turn to Ancient Practice of Oral Exams To Find Out (Dec. 12, 2025), https://www.nysun.com/article/is-it-real-or-is-it-ai-professors-turn-to-ancient-practice-of-oral-exams-to-find-out

109 Peter Dizikes, Study: Flying Keeps Getting Safer, MIT News (Aug. 7, 2024), https://news.mit.edu/2024/study-flying-keeps-getting-safer-0807

110 Pilots Collective, The Evolution of Aviation Safety: Learning from the Past to Protect the Future (Oct. 8, 2025), https://www.thepilotscollective.org/post/the-evolution-of-aviation-safety-learning-from-the-past-to-protect-the-future

111 Richard Hirsh, Virginia Tech, Electrification in History (2024), https://hirsh.history.vt.edu/history-of-resuscitation-techniques/

112 Russ Altman et. al, Report of the AI at Stanford Advisory Committee (Jan. 9, 2025), https://provost.stanford.edu/sites/g/files/sbiybj29391/files/media/file/ai-at-stanford-report.pdf

113 Scott Bennett, Governors State University, The Evolution of the AP Calculus AB Test: 1955-2018 (2019), https://opus.govst.edu/cgi/viewcontent.cgi?article=1000&context=theses_math

114 Stanford AI Tinkery, https://ai-tinkery.stanford.edu

115 Stanford Responsible AI for Sage and Equitable (RAISE) Health, https://med.stanford.edu/raisehealth.html

116 Stanford University Bulletin, BMDS214 Representations and Algorithms for Computational Molecular Biology, Course Description, https://bulletin.stanford.edu/courses/1410301

117 Stanford University, Center for Teaching and Learning, AI Teaching Strategies (Sep. 2, 2025), https://ctl.stanford.edu/aimes/ai-teaching-strategies

118 Stanford University, Office of Community Standards, Generative AI Policy Guidance (Feb. 16, 2023), https://communitystandards.stanford.edu/generative-ai-policy-guidance

119 Stanford University, Report Outlines Stanford Principles for Use of AI (Jan. 9, 2025), https://news.stanford.edu/stories/2025/01/report-outlines-stanford-principles-for-use-of-ai

120 Stanford University, Stanford Accelerator for Learning, https://acceleratelearning.stanford.edu

121 Stanford University, Stanford AI Tinkery, https://ai-tinkery.stanford.edu

122 Stanford University, Teaching Commons, Artificial Intelligence Teaching Guide (2023), https://teachingcommons.stanford.edu/teaching-guides/artificial-intelligence-teaching-guide
123 Syracuse University, Getting Started with Claude Enterprise, https://its.syr.edu/getclaude/
124 Syracuse University, Syllabus Requirements, https://teachingexcellence.syr.edu/academic-integrity/syllabus-requirements/
125 The Spirit of Asilomar, Why Now? Why Asilomar? https://www.spiritofasilomar.org
126 TikTok, About AI-generated content, https://support.tiktok.com/en/using-tiktok/creating-videos/ai-generated-content
127 Toby S. Fischer, 5 Engaging AI Classroom Activities to Try With Your Students (June 3, 2025), https://www.edutopia.org/article/5-engaging-ai-classroom-activities-try-your-students/
128 U.S. Centers for Disease Control and Prevention, Facts About Imaging Procedures (Feb. 15, 2024), https://www.cdc.gov/radiation-health/data-research/facts-stats/imaging-procedures.html
129 U.S. Food and Drug Administration, Guiding Principles of Good AI Practice in Drug Development (Jan. 2026), https://www.fda.gov/media/189581/download
130 U.S. NRC, FY26 Artificial Intelligence Strategic Plan (Sep. 29, 2025), https://ww2.nrc.gov/docs/ML2526/ML25269A196.pdf
131 UL Standards & Engagement, What is a UL standard?, https://ulse.org/standards/
132 Yana Bromberg, Russ Altman, Michael Imperiale et al., Spirit of Asilomar Entreaty 2025.3.1 Artificial Intelligence & the Future of Biotechnology, Spirit of Asilomar (Mar. 1, 2025), https://repository.rice.edu/server/api/core/bitstreams/652ee6ad-e5df-44cc-a00f-7bd2b2c1b8e5/content