EXHIBIT C

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 REPLY REPORT OF RUSS ALTMAN

FEBRUARY 13, 2026

Highly Confidential – Attorneys' Eyes Only

TABLE OF CONTENTS

I. Introduction 1

A. Assignment 1
B. Summary of opinions 2

II. My expertise is the product of decades of work with computational tools, including AI and LLMs 4

A. My expertise is fundamentally rooted in computer science training and informs the classes I teach on AI systems 5
B. My opinions are informed by my experience with real-world applications of LLMs in medicine and scientific research 9
C. An interdisciplinary perspective is appropriate for evaluating the benefits and risks of LLMs 11

III. LLMs offer valuable applications in science and medicine 14

IV. Professor Zhao misrepresents the ways in which LLMs are impacting education 18

V. Professor Zhao disregards the connection between diverse training data and LLM performance 24

VI. Conclusion 25

Appendix A: Materials Considered 27

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]

3. Following my Opening Report, I submitted a rebuttal report on January 23, 2026 ("Altman Rebuttal Report" or my "Rebuttal Report").[2] The Altman Rebuttal Report addressed the opinions and conclusions in the expert report submitted by Plaintiffs' expert Professor Zhao, dated December 19, 2025 (the "Zhao Opening Report"). Capitalized terms not defined herein have the same meanings given in my Opening Report and Rebuttal Report.

4. For this report, Counsel has asked me to evaluate and respond to the opinions and conclusions provided by Plaintiffs' expert Professor Zhao in his expert report dated January 23, 2026 (the "Zhao Rebuttal Report").[3] I address only the portions of the Zhao Rebuttal Report that respond to my opinions and conclusions. My lack of a response to any opinion in the Zhao Rebuttal Report should not be interpreted as agreement with that opinion.

5. 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.

6. As noted in my Opening Report, the field of Artificial Intelligence ("AI"), and Large Language Models ("LLMs") in particular, is rapidly evolving. My opinions expressed in this

[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 Rebuttal Report of Russ Altman, Concord Music Group, et al., v. Anthropic PBC, Case No. 24-cv-03811-EKL-SVK (Jan. 23, 2026) ("Altman Rebuttal Report").

[3] Expert Rebuttal Report of Ben Zhao, Concord Music Group, et al., v. Anthropic PBC, Case No. 24-cv-03811-EKL-SVK (Jan. 23, 2026) ("Zhao Rebuttal Report").

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.

7. In addition to the data and documents cited in my Opening and Rebuttal Reports, additional data and documents considered are listed in Appendix A of this report.

8. 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

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

10. After reviewing the Zhao Rebuttal Report, my opinions expressed in my Opening and Rebuttal Reports remain unchanged.

11. My credentials include decades of experience evaluating computational tools in real-world scientific, medical, policy, and regulatory contexts. My ability to opine on AI, and LLMs in particular, is a function of the system-design and model-architecture expertise I have developed over more than 40 years of interdisciplinary research and applied scientific practice. My decades of studying how computational tools, now recently including LLMs, are used in a myriad of ways across industries have uniquely positioned me to understand, explain and evaluate the real-world impacts and benefits of LLMs. In fact, the U.S. Senate Committee on Health, Education, Labor, and Pensions invited me to testify on the ways in which AI can improve healthcare due to my expertise in this area.[4]

[4] As part of my Senate hearing, I presented on the ways in which AI can improve healthcare through improved patient care and accelerated drug discovery. *See* Senate Testimony "AI's Potential to Deliver Better Patient Care and Accelerate Drug Discovery" presented to U.S. Senate Committee on Health, Education, Labor, and Pensions, (Oct. 9, 2025), https://www.help.senate.gov/imo/media/doc/5176fe8a-b2bd-9e83-d9e6-ce010854bd6a/Altman%20Testimony.pdf. ("Today I want to share insights on two critical areas: the state of AI applications for delivering better patient care and how AI is accelerating the drug discovery process. I will highlight opportunities for congressional action to make these efforts stronger, safer, and human-centered. I should note that I am here in my personal capacity, and the views I share reflect my own professional experience.").

12. As I explained in my Opening and Rebuttal Reports, and reaffirm here, many of the concerns that Professor Zhao raises across education, industry, and scientific practice are characteristic of the early adoption phases of new and novel technologies. The challenges Professor Zhao highlights are not evidence that LLMs are incompatible with the high-stakes domains in which they are being deployed, nor do they negate the substantial and growing body of evidence that points to the contrary. Rather, LLMs provide beneficial and transformative applications across a wide variety of sectors. Both the Zhao Opening and Zhao Rebuttal Reports showcase the ways in which he misunderstands or minimizes the applications of LLMs, like Claude, which are already delivering important, and often lifesaving, value across science and medicine.

13. The Zhao Rebuttal Report places a disproportionate emphasis on issues that are largely manageable or transitory, such as hallucinations and experimental failures, while overlooking how LLMs are used in practice—namely, as tools that augment human expertise, accelerate research, support hypothesis generation, assist with literature synthesis, and improve end-to-end efficiency in regulated domains. When appropriately designed, validated, and governed, LLMs complement—not replace—established scientific and clinical methods. Based on my review, Anthropic has adopted a safety-oriented approach to LLMs and has advocated for regulations that would improve AI transparency. Nothing in either the Zhao Opening or Zhao Rebuttal Reports alters my assessment that systems such as Claude, when responsibly managed, have the potential to deliver significant and wide-ranging benefits.

14. Professor Zhao similarly adopts an overly pessimistic framing of the impact of LLMs on education. While legitimate challenges exist, including misuse and academic integrity concerns, Professor Zhao's analysis fails to account for the active and ongoing institutional adaptations that are already underway to address these concerns. He similarly fails to acknowledge how LLMs have the potential to greatly improve issues of accessibility and quality. As with prior technological transitions in education, these challenges are being addressed through evolving assessment practices, governance frameworks, and pedagogical approaches. Treating current frictions as evidence of permanent educational harm mischaracterizes both historical experience and the capacity of present-day responses.

15. Taken together, the assertions in the Zhao Rebuttal Report do not change the core conclusions of my prior reports. In my opinion, the evidence continues to overwhelmingly show that LLMs, such as Claude, are already generating substantial public benefits across education, science, medicine, creative, and public interest applications. Whilst the need for thoughtful deployment, governance, and institutional adaptation remains, Professor Zhao is wrong to reject the transformative potential of LLMs.

II. MY EXPERTISE IS THE PRODUCT OF DECADES OF WORK WITH COMPUTATIONAL TOOLS, INCLUDING AI AND LLMS

16. Professor Zhao adopts an unduly narrow view of the expertise relevant to the opinions I offer.[5] The opinions I offer in this report are informed by three complementary forms of expertise: formal training and a background in computer science; direct, hands-on experience with GenAI and LLMs; and an interdisciplinary perspective grounded in applying AI systems in high-stakes, real-world domains. Together, this combination of expertise provides an appropriate and reliable foundation for evaluating the benefits, risks, and impacts of LLMs.

17. I am trained in computer science and have decades of hands-on experience applying computational and data-driven methods across a wide range of domains, including medicine, genomics, drug discovery, and FDA regulatory processes.[6] As described in my Opening and Rebuttal Reports, I regularly publish on AI-related topics, including on the use of generative models in scientific and medical contexts and the challenges they present,[7] and I use LLMs in my day-to-day work for tasks such as scientific literature synthesis, regulatory drafting, data interpretation, and education.[8] As an active scientist, engineer, and practitioner, keeping up to date with advances in LLMs is a routine and necessary part of my work. I have kept apprised of relevant technologies in my field since receiving my PhD and continue to do so to this day.

---

[5] Professor Zhao challenges both my expertise and qualifications to offer opinions concerning LLMs and GenAI, suggesting that my background does not include sufficient training in computer science or AI, that my appointments in computer science are only by courtesy, and that my recent research has focused primarily on deep neural networks rather than generative models such as LLMs. He further assumes that my work with neural network architectures diminishes the relevance of my experience with generative models. Professor Zhao also questions the value of my discussion of real-world applications of LLMs outside the immediate context of this litigation. *See* Zhao Rebuttal Report, ¶¶ 126-128.

[6] I have taken formal computer science coursework beginning with my undergraduate studies up until the completion of my PhD. See discussion below in §§ II.A. and II.B.

[7] *See* Altman Opening Report, ¶¶ 28-33; and Altman Rebuttal Report, ¶ 64.

[8] Please refer to Section II below, which further outlines my relevant qualifications.

18. In addition, I hold multiple institutional leadership roles focused on AI research, governance, and responsible deployment, such as my appointment as Senior Fellow at the Stanford Institute for Human-Centered AI ("Stanford HAI"),[9] and my involvement with other interdisciplinary initiatives that bring together stakeholders across academia, industry, government, and civil society.[10] In these roles, I work with other thought leaders to address questions of AI development and challenges, as well as to advance AI research, education, policy, responsible deployment and practice. This cross-functional perspective directly informs the opinions I offer, which are grounded in my practical experience with the deployment, evaluation, and oversight of LLMs in regulated and high-stake domains.

19. In this context, the opinions I offer address how LLMs are used, the benefits they provide, the challenges they present, and how those challenges can be managed in practice. A proper understanding of these topics requires a perspective that can only be achieved from taking a broader interdisciplinary approach.[11] This is the perspective that informs my opinions throughout my reports. To the extent Professor Zhao suggests a narrowly defined computer science background is the only appropriate perspective to evaluate the use cases of LLMs, that view does not reflect the nature of the questions presented or how expertise is assessed for other complex technologies.[12]

A. My expertise is fundamentally rooted in computer science training and informs the classes I teach on AI systems

20. I hold a PhD in Medical Information Sciences, a field centered on the application of computer science, statistics, and AI to biological and medical problems. My formal training includes substantial coursework and research in theoretical and applied computer science, applied mathematics, technical statistics, and AI.[13] My doctoral training was

[9] *See* Stanford HAI, About - Who We Are, https://hai.stanford.edu/about.

[10] I provide additional details of my experiences in my Opening Report §§ II.A and II.B; my Rebuttal Report § II.A; as well as Section II within this report.

[11] By relying on sources from multiple disciplines in his reports, Professor Zhao implicitly acknowledges the relevance of views from educators, commentators, and jurists outside the field of computer science when assessing LLMs and GenAI models.

[12] *See* further discussion below in Section II.C.

[13] I earned a PhD in Medical Information Sciences, a field centered on the application of computer science, statistics, and artificial intelligence to biological and medical problems. *See* Russ B. Altman, Exclusion Methods for Determination of Protein Structure from Experimental Data (1990). My dissertation involved the development of novel computational algorithms to infer three-dimensional molecular structure from sparse and noisy constraints.

explicitly interdisciplinary but fundamentally computational. During my doctoral studies, I completed extensive formal coursework in core computer science subjects, including numerical analysis, data structures and algorithms, programming languages, file and database systems, artificial intelligence, decision analysis, probabilistic modeling, and statistical data analysis. My dissertation advisor was Professor Bruce G. Buchanan, a renowned professor of Computer Science and a foundational figure in AI, particularly in knowledge-based systems and biomedical AI.[14]

21. At Stanford, I am a chaired professor of Bioengineering, which is a technical field and not a medical specialty. While I am a board-certified physician, I am primarily a researcher in biomedical computer science. My teaching and research focus on modern AI systems such as attention mechanisms, transformer architectures, and foundation models, particularly as they apply to protein language models.[15] These models are architecturally analogous to LLMs, meaning they share the same core computational principles, but differ primarily in how input sequences are tokenized (amino acids vs. words or sub-words), in the data they are trained on (biological sequence databases vs. natural language text),[16] and in certain aspects of their objectives (LLMs are typically auto-regressive while many protein models adopt a masked, bidirectional training objective).[17] Following my

This work combined algorithm design, numerical optimization, probabilistic reasoning, and large-scale system integration, core competencies that remain directly relevant to modern AI systems, including large language models.

[14] *See* Bruce Buchanan, Google Scholar Profile, https://scholar.google.com/citations?user=EFEC_toAAAAJ&hl=en.

[15] Recently, I have co-authored a paper in which we use a novel approach based on state-of-the-art protein language models to identify high attention sites within protein primary sequences. *See, e.g.*, Gowri Nayar, Alp Tartici and Russ B Altman, Paying attention to attention: High attention sites as indicators of protein family and function in language models (Sep. 12, 2025), https://doi.org/10.1371/journal.pcbi.1013424. ("Protein Language Models (PLMs) use transformer architectures […] [A]t the core of PLMs is the attention mechanism […] we provide insight into how PLMs process proteins and offer a way to improve function prediction").

[16] *See, e.g.,* PipeBio, Protein language models: promises, pitfalls and applications (June 18, 2024), https://pipebio.com/blog/protein-language-models. ("[G]ood performance of these LLMs was mainly attributed to the unique transformer-based deep neural network structures and the attention feature [...] scientists are applying similar approaches to develop protein language models (PLMs). In PLMs – Analogously to LLMs – amino acid sequences are treated as a stretch of text to train machine learning models.").

[17] *See* Charlotte Hughes, ESM-2 vs ProtBERT: A Comprehensive Performance Comparison for Protein Function Prediction and Biomedical Applications, Protein Engineering Insights (Nov. 26, 2025) https://www.proteineng.com/posts/esm2-vs-protbert-a-comprehensive-performance-comparison-for-protein-function-prediction-and-biomedical-applications. ("The application of the Transformer architecture, originally developed for natural language processing (NLP), to protein sequences represents a paradigm shift in computational biology. Protein language models (pLMs) like ESM-2 and ProtBERT treat amino acid sequences as textual sentences, enabling deep learning models to capture complex evolutionary, structural, and functional patterns from massive protein sequence databases. These models utilize self-supervised pre-training objectives, particularly masked language modeling (MLM), where the model learns to predict randomly masked amino acids within sequences, thereby internalizing fundamental principles of protein biochemistry without explicit supervision.").

appointment to the faculty at Stanford, I was granted a "courtesy appointment"[18] in the Computer Science Department. Contrary to what Professor Zhao suggests, a courtesy appointment at Stanford does not signify a lack of training or expertise in computer science. In fact, the opposite is true. This appointment recognized the extensive research, expertise, and contributions I made, and continue to make, to the Computer Science department, including by teaching computational and algorithmic courses with substantial enrollment from computer science students for decades.[19] Moreover, my appointment as a Senior Fellow at the Hoover Institution[20] recognizes my research at the forefront of artificial intelligence, machine learning, and biomedical data science, where I have focused on applying advanced computing and informatics to deepen understanding of drug action, genetics, AI-driven insights in medicine, etc.—work that bridges core AI/ML research with transformative real-world impact.

22. In the three-plus decades since I joined Stanford's faculty, I have published hundreds of peer-reviewed papers introducing and evaluating new computational methods for

[18] Courtesy faculty appointments follow a formal process, rather than departments granting the title arbitrarily. *See* Stanford University, Faculty Handbook, Chapter 2: Appointments, Reappointments and Promotions in the Professoriate (last updated Jan. 8, 2026), https://facultyhandbook.stanford.edu/index/chapter-2-appointments-reappointments-and-promotions-professoriate. ("2.6.5 Courtesy Appointments […] Faculty members often make substantial contributions to departments other than their own, but in ways less formal than would justify a joint appointment. These contributions are sometimes recognized by means of courtesy appointments […] A recommendation to appoint an existing member of Stanford's professoriate to a courtesy appointment is initiated by the department or school wishing to offer it.").

[19] For approximately thirty years, I have taught Computer Science 274, Representations and Algorithms for Computational Molecular Biology. *See* Stanford, Bulletin ExploreCourses, CS 274: Representations and Algorithms for Computational Molecular Biology, https://explorecourses.stanford.edu/search?view=catalog&filter-coursestatus-Active=on&page=0&catalog=&q=CS+274%3A+Representations+and+Algorithms+for+Computational+Molecular+Biology&collapse=. ("This is a graduate level introduction to bioinformatics and computational biology, algorithms for alignment of biological sequences and structures, computing with strings, phylogenetic tree construction, hidden Markov models, basic structural computations on proteins, protein structure prediction, molecular dynamics and energy minimization, statistical analysis of 3D biological data, integration of data sources, knowledge representation and controlled terminologies for molecular biology, microarray analysis, chemoinformatics, pharmacogenetics, network biology.") This course also covers AlphaFold2, a deep-learning AI system that predicts a protein's three-dimensional structure from its amino acid sequence with near-experimental accuracy. *See also* Google DeepMind, AlphaFold, https://deepmind.google/science/alphafold/; and AlphaFold, Protein Structure Database, https://alphafold.com/. I further teach Protein Language Models, which are architecturally analogous to LLMs, differing primarily in tokenization and training data rather than in fundamental computational principles. These models employ the same transformer-based architectures and generative training paradigms as LLMs, and their study informs an understanding of GenAI more broadly. *See also* Abel Chandra et. al., Transformer-based deep learning for predicting protein properties in the life sciences (Jan. 18, 2023), https://doi.org/10.7554/eLife.82819. ("There is hope that deep learning can close the gap between the number of sequenced proteins and proteins with known properties based on lab experiments. Language models from the field of natural language processing have gained popularity for protein property predictions and have led to a new computational revolution in biology, where old prediction results are being improved regularly. Such models can learn useful multipurpose representations of proteins from large open repositories of protein sequences and can be used, for instance, to predict protein properties.").

[20] *See* The Hoover Institution, Profiles, Russ B. Altman, https://www.hoover.org/profiles/russ-b-altman.

biological and clinical data analysis, as documented in my curriculum vitae.[21] In terms of contributions to general computer science literature, I have also published technical analysis of LLMs as a co-author of an early influential paper that discussed the trends in and impacts of foundational models.[22] Professor Zhao asserts that my publications in this space are of little relevance because they were published in collaboration with others and concerned deep neural networks ("DNNs"), not LLMs.[23] First, collaborative scholarship strengthens, rather than weakens, the relevance and credibility of such research.[24] Second, Professor Zhao does not appear to acknowledge that LLMs are indeed *a version of* DNNs that followed from technologies such as Recurrent Neural Networks ("RNNs") and Long Short-Term Memory networks ("LSTMs"), and their increasing use of attention mechanisms, as part of the lineage of research in this area that attempted to improve performance on sequential data.[25] Ultimately, LLMs primarily differ from other DNNs in scale, training objectives, and output modality rather than in underlying computational principles.[26] Treating LLMs as unrelated to, and not part of the lineage of, prior deep-learning systems therefore misses the opportunity to benefit from the gained experience

[21] *See* Altman Opening Report, Appendix A: CV. In addition to my own research and teaching, I have served on the PhD dissertation defense committees of approximately 20 CS doctoral students. In my lab, I have supervised another 40 PhD dissertations focused on the development and evaluation of novel computational algorithms applied to biological and medical data. I currently mentor a Stanford Computer Science PhD student working on causal reasoning methods for bias detection and synthetic data generation. Additionally, recent peer-reviewed work from my group of students includes research on conditional generative models for synthetic tabular data, causal inference methods for detecting clinician bias, and multiple publications directly addressing the capabilities and applications of LLMs. These include studies using LLMs for pharmacovigilance, mechanistic analysis of attention in biological language models, agentic LLM systems for automating gene-editing experiments (CRISPR-GPT), and large-scale analyses of social media language for public health surveillance. Collectively, this work reflects sustained engagement with generative modeling, causal reasoning, and LLM-based systems at both methodological and applied levels.

[22] *See* Rishi Bommasani et al., On the Opportunities and Risks of Foundation Models (July 12, 2022), https://doi.org/10.48550/arXiv.2108.07258.

[23] *See* Zhao Rebuttal Report, ¶¶ 126 - 127.

[24] In particular for biology, which is my field of expertise, the main focal point of research is dominated by sequential data and DNA sequencing in the last few years is one of the few technologies that has exceeded Moore's Law in the creation of data. Therefore, it is absolutely imperative that I follow the development of these technologies.

[25] *See* Ashish Vaswani et. al., Attention Is All You Need (Aug. 2, 2023), https://doi.org/10.48550/arXiv.1706.03762. ("Recurrent neural networks, long short-term memory […] have been firmly established as state of the art approaches in sequence modeling […] such as language modeling […] the Transformer is the first transduction model relying entirely on self-attention to compute representations of its input and output without using sequence-aligned RNNs").

[26] *See, e.g.,* Renjie Li at. al, What is next for LLMs? Pushing the boundaries of next-gen AI computing hardware with photonic chips (Oct. 6, 2025), https://doi.org/10.1515/nanoph-2025-0217. ("Existing LLMs are all based on a DNN […] relying on an attention mechanism […] widely known as the Transformer architecture"). *See also* Cole Stryker, IBM, What are large language models (LLMs)?, https://www.ibm.com/think/topics/large-language-models. ("Large language models (LLMs) are a category of deep learning models trained on immense amounts of data, making them capable of understanding and generating natural language and other types of content to perform a wide range of tasks. LLMs are built on a type of neural network architecture called a transformer which excels at handling sequences of words and capturing patterns in text.").

of how to assess the value of these systems. The distinction Professor Zhao draws between so-called "discriminative" and "generative" models may be relevant for certain technical discussions but is less relevant to the general task of evaluating the performance characteristics of a computational system. In evaluating the suitability of LLMs for use in different settings, it is critical to understand how users incorporate them in practice, how challenges with LLMs are (or can be) managed, and whether LLMs can provide meaningful benefits in real-world settings.

23. More importantly, evaluating the impact of LLMs does not rely solely on their technical underpinnings, but also on the observed application, performance, and governance of the technology. Professor Zhao's framing also distracts from the practical questions of how these tools are used by scientists, physicians and educators. In any event, my opinions are substantially informed by my direct experience studying (and teaching) both how LLMs work (including their transformer architecture) as well as testing and using LLMs in my research, which I discuss further below.

B. My opinions are informed by my experience with real-world applications of LLMs in medicine and scientific research

24. My research has directly engaged with LLMs in applied, high-stakes scientific and regulatory contexts, including projects such as CRISPR-GPT, FDA bioequivalence evaluation, and PathwayQA. Across these projects, my research has explored how LLMs can advance biomedical discovery and clinical science more broadly.

25. One example is CRISPR-GPT, an LLM agent my colleagues and I developed to automate and enhance the design process of CRISPR-based gene-editing experiments. Gene-editing technologies such as CRISPR hold transformative potential for treating a wide range of diseases, but designing effective experiments is labor-intensive and time-consuming, requiring deep biological and computational expertise. CRISPR-GPT is augmented with domain-specific knowledge and external tools to assist with tasks such as selecting CRISPR systems, designing guide RNAs, cellular delivery methods, drafting protocols, and designing validation experiments to confirm editing outcomes-all critical steps in the gene-editing process. By supporting these components of experimental design, the system accelerates research workflows and reduces complexity while leaving empirical validation and scientific judgment firmly in human hands.

26. In ongoing FDA bioequivalence-related research, my collaborators and I have developed rigorous benchmarks to evaluate LLM performance on the review of complex regulatory documents.[27] This work required not only application-level familiarity with LLMs, but also deep engagement with LLM fine-tuning procedures, dataset construction, evaluation methodology, and error analysis. In this context, my laboratory applies LLMs to bioequivalence evaluation, which assesses whether generic drugs perform equivalently to their brand-name counterparts.[28] This research aims to build LLMs to review and summarize extensive pharmacokinetic data, study designs, and statistical analyses for completeness and consistency, while maintaining rigorous human oversight as part of a process that is critical for FDA approval. In another study, my co-author and I developed and validated a methodology that leveraged LLMs to construct an expanded lexicon of colloquial and informal terms associated with drugs of abuse for use in social-media-based pharmacovigilance.[29] The study involved using LLMs to extract, structure, and evaluate complex information from unstructured textual data, with rigorous human oversight, in support of advancing public health research. The study required applied expertise in LLMs, including controlled output generation, systematic refinement of outputs, and assessment of model performance in a domain-specific scientific context.[30]

27. My team is also developing a formal benchmark for evaluating LLM performance and is in the process of fine-tuning an open-weight LLM to operate against that benchmark, reflecting continued technical engagement with model development and evaluation in a regulated domains.[31] In this line of work, my laboratory systematically evaluated the

[27] *See* Altman Opening Report, ¶ 33. ("Our goal is to use LLMs to read and summarize these reports and to check for completeness and consistency. Automating these steps with LLMs could help the FDA improve the efficiency and consistency of bioequivalence evaluations while supporting its mission to ensure that safe and effective drugs are available to the American public, typically at a lower cost than their brand-name counterparts.").

[28] *See* Altman Opening Report ¶ 33.

[29] Kristy A. Carpenter and Russ B. Altman, Using GPT-3 to Build A Lexicon of Drugs of Abuse Synonyms for Social Media Pharmacovigilance (Feb. 17, 2023), https://doi.org/10.3390/biom13020387.

[30] In fact, building further on this work, my collaborators and I are using LLMs to disambiguate words that have multiple meanings, as part of our research in monitoring drug epidemics. We have found that LLMs are effective at looking at the context of word usage and identifying which meaning is being used. For example, we created prompts and analytical methods to recognize whether the word "fenty" is being used in the drug context, as a street name for fentanyl, or to reference a fashion or beauty product, with respect to singer Rihanna's cosmetics brand. This is valuable in social media monitoring of opioid usage, and we have outlined our general approach of leveraging LLMs to refine word searches in a recently published article. *See* Delaney A. Smith et al., Monitoring the opioid epidemic via social media discussions (May 15, 2025), https://doi.org/10.1038/s41746-025-01642-x at 6. ("An advantage of the high volume of drug discussions on Reddit combined with the availability of LLMs to streamline Natural Language Processing tasks means that systems can be built relatively cost-effectively and quickly to spot emerging drug trends that are not yet monitored by the CDC or NFLIS…").

[31] This research is available as a preprint. *See* Anthropic_0000594245.

performance of multiple LLMs on the accuracy of their responses on biochemical reactions, including cellular reaction pathways that are essential to understanding biological systems.[32]

28. These efforts demonstrate not only practical use of LLMs but also hands-on involvement in their development, evaluation, benchmarking, and error testing of LLMs in contexts where reliability, safety, and accountability are essential. This work goes beyond exploratory use and basic familiarity. My experience reflects the type of applied technical knowledge necessary to assess how LLMs perform in practice, what their limitations are, and how those limitations can be managed through design and governance.

29. Beyond my academic work, I have direct, hands-on experience with LLMs both as a user and an evaluator in real-world settings.[33] Prior to the release of ChatGPT, I was invited by researchers at Microsoft to preview and evaluate their LLM and its implications for medicine and biology. In addition, I have long-standing engagement with the AI research community and ongoing hands-on experience applying LLMs in regulated, high-stakes domains, as described above. Accordingly, my experience with LLMs reflects sustained technical engagement with their design, limitations, and deployment in regulated, high-stakes domains.

C. An interdisciplinary perspective is appropriate for evaluating the benefits and risks of LLMs

30. Fluency in computer science alone is not sufficient to provide a complete analysis of the benefits, risks and societal impacts of LLMs. Computer science is an important lens but, on its own, provides an incomplete toolkit to assess the many ways in which AI has the potential to change how humans perform tasks. Professor Zhao appears to erroneously believe that the only people qualified to study and critique a technology are those who design it.[34] That framing is inconsistent with how expertise is assessed for other complex

[32] *See* Anthropic_0000594245. ("We systematically curate Reactome into a question-answer dataset, and evaluate current LLMs for their ability to generate data related to protein pathways…").

[33] I also have experience with LLM development, including fine-tuning open-source foundation models. In one case, I adapted the Qwen model by modifying its weights through supervised training on thousands of FDA-focused question–answer pairs to achieve domain-specialized performance. Although this effort did not constitute de novo model creation or model evaluation, it involved substantive development work to tailor an LLM for specialized regulatory use cases.

[34] *See* Zhao Rebuttal Report, ¶ 126.

technologies. For example, physicians routinely evaluate the safety and effectiveness of medical devices without having designed the hardware, and regulators assess the impact of pharmaceuticals without having synthesized the molecules themselves. In those contexts, informed, hands-on experience with the applications, limitations, and governance in real-world settings shapes expertise. Consistent with my Opening and Rebuttal Reports, I explore the beneficial deployment of LLMs in different high-stake domains, across the fields of education, science, the arts, and public civil engagement.[35] Whether it concerns the way in which LLMs are improving the delivery of care,[36] expediting new drug discovery,[37] or helping drive conservation efforts in Lake Tahoe,[38] my opinions have always focused on the applications of LLMs, rather than limited technical questions of the model architecture. In order to conduct this thorough evaluation of the benefits, challenges, and governance strategies associated with the use of LLMs, I draw on a combination of both my technical and practical expertise. This approach is reflected in the broader industry and academic community, who consistently look to a range of experts for their commentary and analysis of LLMs.[39]

31. My applied perspective is reinforced by my institutional leadership roles focused on AI research, governance, and responsible deployment. As described in my Opening Report, these include serving as Associate Director and Senior Fellow at the Stanford Institute for Human-Centered AI ("HAI"), Co-Lead of RAISE-Health, Faculty Director of the 100-Year Study of Artificial Intelligence ("AI100"), Chair of the AI at Stanford Advisory Committee to the University Provost, and Co-Chair of the UCB–Stanford Digital Health Research

---

[35] *See* Altman Opening Report §§ II.A and III; and Altman Rebuttal Report § III.

[36] *See* my discussion of how LLMs improve delivery of critical care services in Altman Opening Report, ¶¶ 43-45.

[37] *See* examples of how Claude facilitates early-stage research and the development of new drugs in Altman Opening Report, ¶¶ 71-76.

[38] *See* my discussion of the LLM-facilitated research process used throughout the Tahoe Lead Removal Project in Altman Opening Report, ¶ 100; and Altman Rebuttal Report, ¶ 69.

[39] One example of collaboration across industry, academia and policy experts was the establishment of the Artificial Intelligence Safety and Security Board ("AISSB") by the U.S. Department of Homeland Security in April 2024, pursuant to an Executive Order by President Biden. *See* Executive Office of the President, Executive Order 14110, Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence (Oct. 30, 2023), https://www.govinfo.gov/content/pkg/FR-2023-11-01/pdf/2023-24283.pdf; and U.S. Department of Homeland Security, Notice, Establishment of the Artificial Intelligence Safety and Security Board (Apr. 29, 2024), https://www.govinfo.gov/content/pkg/FR-2024-04-29/pdf/2024-09132.pdf. The 24 members of the AISSB included AI developers, such as Anthropic's CEO and founder, Dario Amodei, as well as prominent academics, business leaders, civil activists and policymakers. AISSB set recommendations for the safe and secure development and deployment of AI in critical infrastructure. *See* U.S. Department of Homeland Security, Roles and Responsibilities Framework for Artificial Intelligence in Critical Infrastructure (Nov. 14, 2024), https://www.dhs.gov/sites/default/files/2024-11/24_1114_dhs_ai-roles-and-responsibilities-framework-508.pdf.

Collaborative.[40] I am a long-standing member of the Association for the Advancement of Artificial Intelligence ("AAAI"), where I first presented peer-reviewed research in 1987 and have remained an active participant since joining. As part of an interdisciplinary group of leading academics at Stanford HAI, I have consistently contributed perspectives on developments in GenAI, including on the impact of introduction of ChatGPT into the mainstream, and commentary on the significance of new models.[41] I have also presented testimony in front of the U.S. Senate Committee on Health, Education, Labor, and Pensions in which I discussed AI's potential to deliver better patient care and accelerate drug discovery, as well as congressional action required to realize AI's full benefits.[42] As part of my work, I present at and contribute to regular boot camps for congressional staff, who are involved in drafting government policies, to equip them with knowledge on the latest developments, benefits, and risks of AI.[43] These roles require continuous engagement with the evolving capabilities, limitations, and risks of AI systems across multiple domains as well as direct involvement in shaping policies and practices for their responsible deployment.

32. LLMs are now used by individuals, institutions, and organizations with diverse goals and constraints. As a result, evaluating their benefits, challenges, and impacts requires an interdisciplinary perspective that extends beyond the technical design considerations. Taken together, my training, research, institutional leadership, and experience developing, using, and evaluating LLMs in regulated, high-stakes settings provide the basis for the

---

[40] *See* Altman Opening Report, ¶ 7.

[41] *See* Anthropic_0000594350. ("We believe that interdisciplinary collaboration is essential in ensuring these technologies benefit us all. The following are perspectives from Stanford leaders in medicine, science, engineering, humanities, and the social sciences on how generative AI might affect their fields and our world."). My contribution focused on the potential for GenAI to improve the efficiency of trials, drug development and diagnostic testing. *See also* Stanford HAI, How Disruptive Is DeepSeek? Stanford HAI Faculty Discuss China's New Model (Feb. 13, 2025), https://hai.stanford.edu/news/how-disruptive-deepseek-stanford-hai-faculty-discuss-chinas-new-model?mc_cid=5056be677e.

[42] See Russ Altman, Senate Testimony "AI's Potential to Deliver Better Patient Care and Accelerate Drug Discovery" presented to U.S. Senate Committee on Health, Education, Labor, and Pensions, (Oct. 9, 2025). https://www.help.senate.gov/imo/media/doc/5176fe8a-b2bd-9e83-d9e6-ce010854bd6a/Altman%20Testimony.pdf. In this paper, I discuss how AI can "improve patient care", "the relationship between patients and their physicians", and "the patients' understanding and control of their care." Doing so requires setting up interdisciplinary teams and congressional action to "improve privacy preserving data collection", "invest in public compute infrastructure and research", "strengthen the public sector's capacity to evaluate AI-accelerated drug discoveries", and "develop a workforce equipped for safe and robust AI-driven drug discovery".

[43] *See* Stanford HAI, Congressional Boot Camp on AI, https://hai.stanford.edu/policy/policymaker-education/congressional-boot-camp. ("The Stanford Institute for Human-Centered AI (HAI) specifically designed the Congressional Boot Camp on Artificial Intelligence to explore the latest in AI developments, equipping participants with the comprehensive knowledge needed to think critically about regulating and governing this emerging technology.")

opinions I offer in this report, as well as my Opening and Rebuttal Reports. Assessing the benefits, challenges, and impacts of LLMs is fundamentally an applied and interdisciplinary inquiry that requires understanding how these tools are used, governed, and integrated into critical and societally impactful domains such as medicine, education, science, and public institutions. My opinions are grounded in a combination of technical understanding and extensive experience, which have enabled me to assess the real-world applications, governance, and impacts of LLMs. To the extent Professor Zhao suggests otherwise, his characterization reflects a deep misunderstanding of both the nature of the issues presented and the expertise required to evaluate them.

III. LLMS OFFER VALUABLE APPLICATIONS IN SCIENCE AND MEDICINE

33. As described in my Opening and Rebuttal Reports, LLMs offer valuable benefits as applied in a variety of fields.[44] I have witnessed the early benefits of these valuable applications firsthand in the fields of science and medicine, notwithstanding the potential, albeit manageable, risks of today's LLMs. Professor Zhao correctly asserts that hallucination errors are one such risk but incorrectly concludes that hallucinations by LLMs render their use incompatible with scientific discovery.[45] While hallucinations are a known limitation of current LLMs, the models themselves are non-linear functional approximations trained primarily on data. I disagree with the conclusion that this limitation makes LLMs unfit for scientific use. Although hallucinations are likely to persist at the model level, I believe that they can be minimized and managed through safeguards and other mechanisms when deployed. Claude and other LLMs have already brought significant valuable advances to scientific fields, particularly in accelerating scientific discovery, improving medical and clinical care, enhancing productivity, and expanding access to information and expertise. Beyond that, recent medical and scientific research has made significant progress in addressing and understanding what drives hallucinations, particularly in my fields of expertise and application areas.[46]

---

[44] *See* Altman Opening Report, §§ III.A, III.B., III.C. and III.D.

[45] Zhao Rebuttal Report, ¶ 130. ("Predictive models that make mathematical errors and hallucinate facts are not conducive to factual scientific discovery.") *See also* Zhao Rebuttal Report ¶ 131. ("While LLMs analyze large volumes of text faster than humans can, they are also known to make biased mistakes, as well as hallucinate non-existent statements that do not follow the text input.").

[46] Risk mitigation techniques such as retrieval-augmented generation ("RAG"), domain-specific models, source verification, and continuous evaluation are employed to preserve scientific integrity. *See, e.g.,* Amazon Web Services, What is RAG?, https://aws.amazon.com/what-is/retrieval-augmented-generation/. ("Retrieval-Augmented Generation (RAG) is the process of optimizing the output of a large language model, so it references an authoritative knowledge base outside of its training data sources before generating a response. Large Language

34. Medicine and healthcare have always been receptive to the integration of new technologies, despite any initial imprecision or limitations. The most pertinent example has been the integration of Electronic Health Records ("EHRs") or Electronic Medical Records ("EMRs") which, when first introduced, saw healthcare professionals face issues navigating the systems and ensuring adequate data security.[47] Now, these systems are standard practice and have continued to improve healthcare.[48] History has consistently demonstrated that there is the ability to improve the integration of these tools through proper clinician training, effective risk management, and above all, active research partnerships "between those who study the diffusion of information technology and those who study the human dimensions of medical care."[49]

35. Established diagnostic tools routinely exhibit false positive and false negative rates.[50] As medical professionals and researchers, we are trained to understand the limitations of diagnostic tools and methods, assess associated risks, and manage them through established protocols. There are also existing institutional safeguards present in the medical and scientific fields which aim to ensure standards and benchmarks for diagnostic tests as experimental methods develop.[51] It is a common goal of medical researchers to evaluate the performance of diagnostic tests, especially those methods which are in their

---

Models (LLMs) are trained on vast volumes of data and use billions of parameters to generate original output for tasks like answering questions, translating languages, and completing sentences. RAG extends the already powerful capabilities of LLMs to specific domains or an organization's internal knowledge base, all without the need to retrain the model. It is a cost-effective approach to improving LLM output so it remains relevant, accurate, and useful in various contexts."). *See also* Subhankar Maity and Manob Jyoti Saikia, Large Language Models in Healthcare and Medical Applications: A Review (June 9, 2025), https://doi.org/10.3390/bioengineering12060631. ("Key methodological aspects—such as domain-specific data acquisition, large-scale pre-training, supervised fine-tuning, prompt engineering, and in-context learning—are explored in the context of healthcare use cases.").

[47] *See* Michele L. Kreidler, The Development of Healthcare and Information Technology (2023), https://www.ebsco.com/research-starters/information-technology/development-healthcare-and-information-technology#full-article. ("Doctors were challenged in two ways when implementing EHRs; they had to learn how to navigate the EHR system and also ways of integrating the use of the EHR into the doctor-patient relationship…").

[48] *See* Ahti Valtteri, Why Imperfect Tools are Often Better Than None: A Business Perspective (Aug. 15, 2024), https://opteamyzer.com/knowledge-center/insights-articles/why-imperfect-tools-are-often-better-none-business-perspective. ("Nevertheless, the transition to EMRs significantly improved the quality of medical care by providing faster access to patient information and reducing the number of errors in treatment.").

[49] Richard Frankel et al., Effects of exam-room computing on clinician-patient communication (Aug. 2005), https://doi.org/10.1111/j.1525-1497.2005.0163.x.

[50] This is a fundamental concept that underlies scientific practice. *See* University of Hawai'i, College of Education, Practices of Science: False Positives and False Negatives, https://manoa.hawaii.edu/exploringourfluidearth/chemical/matter/properties-matter/practices-science-false-positives-and-false-negatives. ("All tests have a chance of resulting in false positive and false negative errors. They are an unavoidable problem in scientific testing. This creates problems in data analysis in many scientific fields").

[51] *See, e.g.,* U.S. FDA, Statistical Guidance on Reporting Results Evaluating Diagnostic Tests – Guidance for Industry and FDA Staff (Mar. 13, 2007), https://www.fda.gov/regulatory-information/search-fda-guidance-documents/statistical-guidance-reporting-results-studies-evaluating-diagnostic-tests-guidance-industry-and-fda.

experimental or developmental stages.[52] Part of engaging in the diagnostic process is also acknowledging that there are limitations to medical data.[53]

36. Likewise, in the context of scientific research, LLMs function as imperfect but useful tools that must be applied with appropriate awareness of their limitations. The research community is widely aware that hallucinations can occur, just like it is aware of other sources of potential error such as human mistakes in coding or flaws in experimental design. Therefore, as a user of these models and tools, it remains my responsibility to understand that LLM-generated output requires routine mitigation through human oversight, verification, and auditing. Professor Zhao should know that these safeguards have always been, and remain, essential components of the scientific process.

37. Professor Zhao's characterization of "LLM hallucinations in medicine as inevitable"[54] ignores the trends of improvement on the current state of LLM technologies. Medical research has already seen significant progress through the successful application of LLMs across a variety of fields, many of which I have previously highlighted.[55] A recent study, which analyzed the integration and performance of LLMs in complex medical scenarios, found that Claude's ability to provide concise answers dominated "in relevance and completeness, suggesting its superior ability to generate contextually pertinent and thorough responses," when compared to three other leading AI models.[56] In addition to evaluating the performance of these LLMs in various healthcare contexts, the authors' findings reinforce the need for medical professionals and researchers to continue applying nuanced and context-based knowledge to their decision-making.[57] A recent study of user

[52] A recent illustrative example would be studies which looked at the efficacy of Covid-19 antigen testing. *See, e.g.,* Filip Obradovic, Measuring Diagnostic Test Performance Using Imperfect Reference Tests: A Partial Identification Approach (Aug. 2024), https://doi.org/10.1016/j.jeconom.2024.105842. This study found that a widely used COVID-19 antigen test tended "to produce significantly more false negative results than what the currently cited figures suggest…".

[53] *See* Venktesh Ramnath, The Conversation, Medicine is an imperfect science – but you can still trust its process (Sept. 7, 2021), https://theconversation.com/medicine-is-an-imperfect-science-but-you-can-still-trust-its-process-166811. ("Doctors must then engage in thoughtful analysis of the strengths and weaknesses of scientific methods and data to optimize and tailor our recommendations for individual patients – without the luxury of perfect or even complete datasets to rely on.").

[54] Zhao Rebuttal Report, ¶ 132.

[55] *See* Altman Rebuttal Report § III.B.

[56] Vamsi Krishna Uppalapati and Deb Sanjay Nag, A Comparative Analysis of AI Models in Complex Medical Decision-Making Scenarios: Evaluating ChatGPT, Claude AI, Bard, and Perplexity (Jan. 18, 2024), https://doi.org/10.7759/cureus.52485.

[57] *See* Vamsi Krishna Uppalapati and Deb Sanjay Nag, A Comparative Analysis of AI Models in Complex Medical Decision-Making Scenarios: Evaluating ChatGPT, Claude AI, Bard, and Perplexity (Jan. 18, 2024),

interactions found that Claude is actively used for a measurable subset of healthcare-related tasks, often in patient-facing roles.[58] Claude is also being integrated with healthcare data systems to provide HIPAA-ready products to healthcare providers, payers, and health tech companies.[59] Industry reviews emphasize Claude's capacity to handle complex medical information responsibly and reliably.[60] Studies like these should make it clear that scientists and companies are working hard to understand the limitations of LLMs as well as improve their capabilities for logical inference, mathematical reasoning, and first principles thinking. Further illustrative examples of how researchers across medical and scientific fields are focusing on how to properly harness the power of these technologies can be seen in the work by scientists building hybrid models of physics-based reasoning with LLMs,[61] as well as AI work in causal modeling, to which I have contributed.[62] Contrary to Professor Zhao's assertions, the presence of hallucinations does not render LLMs unusable or incompatible with high-stakes domains. His analysis conflates model mechanics with scientific practice. In practice, LLMs are not substitutes for experimentation, deterministic computation, or empirical validation. Rather, they support core aspects of scientific workflows, including literature synthesis, hypothesis generation, experimental planning, data interpretation, and regulatory documentation.[63] Examples of this approach, as previously discussed in my Opening and Rebuttal Reports, include the development of CRISPR-GPT,[64] the use of Biomni by biomedical

https://doi.org/10.7759/cureus.52485. ("Although AI yields helpful information, it should reinforce rather than supersede human judgment, especially in sophisticated health situations.").

58 Gabriel Alain et al., Evaluating User Interactions and Adoption Patterns of Generative AI in Health Care Occupations Using Claude: Cross-Sectional Study (May 30, 2025), https://www.jmir.org/2025/1/e73918.

59 Anthropic, Advancing Claude in healthcare and the life sciences (Jan. 11, 2026), https://www.anthropic.com/news/healthcare-life-sciences.

60 *See* Alexis Balvair, HealthPoint, Claude AI in Healthcare – Review (Jan. 15, 2026), https://healthpoint.com/technology/claude-ai-in-healthcare-review.

61 *See, e.g.,* Yinggan Xu et al., Advancing AI-Scientist Understanding: Multi-Agent LLMs with Interpretable Physics Reasoning (Aug. 18, 2025), https://doi.org/10.48550/arXiv.2504.01911. ("Large Language Models (LLMs) have become increasingly popular for tackling complex physics problems […] we develop a novel multi-agent LLM physicists framework that enhances the interpretability, transparency, and verifiability of LLM outputs in physics problem-solving").

62 *See, e.g.,* Kara Liu, Russ Altman and Vasilis Syrgkanis, Detecting clinician implicit biases in diagnoses using proximal causal inference (2025), https://doi.org/10.1142/9789819807024_0024. ("[M]easuring implicit bias is challenging […] In this work, we propose a computational tool to detect clinician implicit bias in observational datasets by measuring the causal effect of patient attributes, like race, SES, and other social determinants of health (SDoH), on medical diagnoses.").

63 *See* Altman Opening Report §§ III.A. and III.C.

64 *See* Altman Opening Report, ¶¶ 31-32.

researchers,[65] genomic analysis using Claude-built AI agents by the Broad Institute,[66] and the targeted introduction of Claude for Life Sciences to be used broadly for the advancement of scientific discovery.[67] In this way, LLMs accelerate the human components of scientific work while leaving validation and decision-making grounded in established scientific methods.

38. It is also important to recognize the role of coordination at the industry and government levels in addressing these challenges. For example, Anthropic has engaged in efforts to understand the nature of these risks and develop systematic frameworks to address public safety and accountability concerns.[68] Their leadership has been active in encouraging industry-wide safety efforts. Most recently, Dario Amodei has publicly supported coordination between AI developers and policymakers, including calls for transparency legislation, which would require AI companies to disclose development and safety information about their models.[69] In my opinion, such approaches reflect a broader recognition that well-designed regulatory oversight will play an important role in the continued development and use of LLMs, particularly in high-stakes domains such as science and medicine.

IV. PROFESSOR ZHAO MISREPRESENTS THE WAYS IN WHICH LLMS ARE IMPACTING EDUCATION

39. I have taught classes at the collegiate and postgraduate levels for more than three decades including several courses in the Computer Science, Bioengineering, Genetics, and Biomedical Data Science departments as well as the Ethics & Society program at

[65] *See* Altman Rebuttal Report, ¶ 64.

[66] *See* Altman Rebuttal Report, ¶ 65.

[67] *See* Altman Rebuttal Report, ¶ 63.

[68] *See* Anthropic, The need for transparency in Frontier AI (July 7, 2025), https://www.anthropic.com/news/the-need-for-transparency-in-frontier-ai. ("AI is advancing rapidly. While industry, governments, academia, and others work to develop agreed-upon safety standards and comprehensive evaluation methods—a process that could take months to years—we need interim steps to ensure that very powerful AI is developed securely, responsibly, and transparently. We are therefore proposing a targeted transparency framework, one that could be applied at the federal, state, or international level, and which applies only to the largest AI systems and developers while establishing clear disclosure requirements for safety practices.").

[69] *See* Dario Amodei, The Adolescence of Technology - Confronting and Overcoming the Risks of Powerful AI (Jan. 2026), https://www.darioamodei.com/essay/the-adolescence-of-technology. ("Anthropic's view has been that the right place to start is with transparency legislation […] As more specific and actionable evidence of risks emerges (if it does), future legislation over the coming years can be surgically focused on the precise and well-substantiated direction of risks, minimizing collateral damage. To be clear, if truly strong evidence of risks emerges, then rules should be proportionately strong."); and ("California's SB 53 and New York's RAISE Act are examples of this kind of legislation, which Anthropic supported and which have successfully passed.").

Stanford.[70] In addition, as I mention in both my Opening and Rebuttal Reports, I have served and continue to serve on many academic committees and task forces, whose existence relates to overseeing policy related to the role of AI in education. These experiences make me acutely aware of the ways in which LLMs have impacted education.

40. Professor Zhao adopts a pessimistic framing of the impact of LLMs on education, emphasizing potential harms such as diminished reasoning and increased cheating while dismissing or minimizing evidence of existing and prospective educational benefits.[71] He argues that the adoption of LLMs has led to widespread AI-enabled cheating across educational institutions and characterizes this phenomenon as pervasive and difficult to control.[72] By positing that LLMs are "simply killing education",[73] he catastrophizes a set of real but manageable challenges being faced by both teachers and students.[74]

41. Cheating during periods of technological transition long predates the introduction of LLMs and has historically reflected transitional challenges in assessment design rather than incompatibility between new technologies and education.[75] As I discuss in my Rebuttal Report, similar concerns about cheating and academic integrity arose with earlier

[70] As noted in my Opening Report, I co-teach a Bioengineering Ethics course that enrolled approximately 240 students last spring. The course focuses on ethical reasoning applied to novel biomedical technologies, drawing on major ethical frameworks including consequentialism, Kantian deontology, common morality principles (*e.g.*, the Belmont Report), and casuistry, among others. The course is designated as a "writing in the major" requirement and includes instruction on academic writing, including the appropriate and responsible use of LLMs in written work. *See* Altman Opening Report, ¶ 41. In addition, I currently teach three other courses which are offered across several departments at Stanford. These are "Representations and Algorithms for Computational Molecular Biology" (see https://explorecourses.stanford.edu/search?view=catalog&filter-coursestatus-Active=on&page=0&catalog=&academicYear=&q=CS+274+Altman&collapse=); "Introduction to Biomedical Informatics Research Methodology" (see https://explorecourses.stanford.edu/search?view=catalog&filter-coursestatus-Active=on&page=0&catalog=&academicYear=&q=CS+272+Altman&collapse=); and "Principles of Pharmacogenomics" (see https://explorecourses.stanford.edu/search?view=catalog&filter-coursestatus-Active=on&page=0&catalog=&academicYear=&q=Genetics+224+Altman&collapse=). The first two courses in particular involve programming assignments in which, for some assignments, students can and are encouraged to use LLMs for coding.

[71] *See* Zhao Rebuttal Report, ¶¶ 135-138.

[72] Zhao Rebuttal Report, ¶ 135. ("Perhaps even more impactful than shallow learning, the adoption of LLMs is creating a boom in AI-based plagiarism across all levels of education.").

[73] Zhao Rebuttal Report, ¶ 137.

[74] From my recent experience teaching over 400 students across 4 classes, I have observed that students are acutely aware of the need to develop competitive skill sets for the job market and also understand that reliance on LLMs will not lead them to excel in the classroom nor beyond.

[75] One example Professor Zhao cites, at Zhao Rebuttal Report, ¶ 137, as evidence of widespread AI cheating actually involved no AI at all, as the students in that case cheated without using AI. *See* Neil Vigdor and Hannah Ziegler, Their Professors Caught Them Cheating. They Used A.I. to Apologize, The New York Times (Oct. 29, 2025), https://www.nytimes.com/2025/10/29/us/university-illinois-students-cheating-ai.html. ("It appeared that the students had been tipped off about the questions and when they had to respond, according to the professors, who sent emails to more than 100 students telling them that the ruse was up.").

technologies, including calculators, online search engines, homework-assistance platforms, and free online encyclopedias.[76] In each case, institutions initially took time to respond, but eventually developed assessment practices and governance frameworks that preserved academic rigor while accommodating new tools.

42. In my opinion, cheating itself is in no way exclusively tied to LLMs. As long as there have been schools, students have found creative ways to circumvent academic expectations, from handwritten cheat sheets to online encyclopedias (*e.g.*, Wikipedia) to outsourcing work to third parties (*e.g.*, Chegg and similar services).The proliferation of GenAI tools like LLMs has changed the ways in which some students attempt to cheat, but this pattern mirrors prior technological transitions. While LLMs may expand the set of available methods available for cheating, they primarily affect students already inclined to disregard integrity policies rather than creating that inclination.

43. More importantly, as I describe in both my Opening and Rebuttal Reports, LLMs are general-purpose technologies with many legitimate educational uses. Whether a particular use constitutes "cheating" depends primarily on user choice as well as institutional rules and expectations, rather than on the mere existence of the tool itself. Research on academic integrity has consistently shown that contextual factors, especially social norms, perceived severity of penalties, and the presence of governance structures such as academic honor codes, play a central role in shaping student behavior.[77] In this sense, the tendency to cheat is driven more by student incentives, peer norms, and institutional policy design, rather than by the availability of any particular technology. A study by my colleagues at Stanford demonstrated that the proliferation of GenAI did not cause major increases in cheating across American high schools.[78] This aligns with my own

[76] *See* Altman Rebuttal Report, ¶¶ 38, 39 and 40.

[77] *See, e.g.,* Donald L. McCabe et al., Honor Codes and Other Contextual Influences on Academic Integrity: A Replication and Extension to Modified Honor Code Settings (June 2002), https://link.springer.com/article/10.1023/A:1014893102151 at 360. ("Thus, peer perceptions concerning academic dishonesty can be an important influence as students make individual decisions about cheating in the classroom."); and at 368: ("[T]he level of academic dishonesty is highest at colleges that do not have honor codes…").

[78] *See* Victor R. Lee et al., Cheating in the age of generative AI: A high school survey study of cheating behaviors before and after the release of ChatGPT (June 12, 2024), https://doi.org/10.1016/j.caeai.2024.100253 at 9. ("Although AI chatbot technology has been a cause for alarm, especially with respect to students potentially using it for cheating and to produce writing to pass as their own, we do not see indications in this study that it has dramatically changed the prevalence of cheating – including specifically in behaviors that are ones that are thought to be most likely to be encouraged by AI chatbots."). *See also* Carrie Spector, What do AI chatbots really mean for students and cheating? (Oct. 31, 2023), https://ed.stanford.edu/news/what-do-ai-chatbots-really-mean-students-and-cheating. ("For years, long before ChatGPT hit the scene, some 60 to 70 percent of students have reported engaging in at least one "cheating" behavior during the previous month. That percentage has stayed about the

observations at the college level. While LLMs like Claude offer many valuable educational benefits for the average student, they likely play less of a role in increasing overall cheating rates for the students who are already inclined to misuse such tools.

44. Moreover, Professor Zhao's analysis gives insufficient weight to the ways educational institutions have begun to respond and adapt, and it treats a current state as evidence of a permanent and unmanageable condition. As I also describe in my Rebuttal Report, institutions are already experimenting with and implementing changes to assessment design and academic integrity frameworks in response to LLMs.[79] These include revised assessment formats, updated honor codes, and disclosure requirements governing the appropriate use of AI tools. While these adaptations are still evolving and not yet uniformly adopted, their emergence demonstrates that the educational response to AI-enabled cheating is in an early phase of adjustment and many of the current concerns with LLMs in education are unlikely to persist long term.

45. Professor Zhao characterizes LLM-based cheating as a "plague upon education," citing an opinion piece with a title containing the aforementioned phrase.[80] For the reasons I give here and in my other reports, I disagree with this characterization of LLMs as a plague upon education. Moreover, upon reviewing the article, written by a professor at Nova Southeastern University, I find that many of the experiences and opinions described in the article are in fact consistent with my own opinions on how LLMs can impact education. Importantly, the article was written only a few months after the initial release of ChatGPT in late 2022, at a time when educators had not yet had the opportunity to adapt assignments or assessment practices to a post-LLM environment.[81] In that context, the author warns fellow professors about the potential for LLMs to undermine assignments that were not developed with these tools in mind. His conclusion is that the arrival of LLMs marks a "pivotal moment in the history of education" and creates an opportunity for meaningful change in education.[82]

---

same or even decreased slightly in our 2023 surveys, when we added questions specific to new AI technologies, like ChatGPT, and how students are using it for school assignments.")

[79] *See* Altman Rebuttal Report § II.A.

[80] *See* Zhao Rebuttal Report, ¶ 135.

[81] Jeremy Weissman, ChatGPT Is a Plague Upon Education, Inside Higher Ed (Feb. 8, 2023), https://www.insidehighered.com/opinion/views/2023/02/08/chatgpt-plague-upon-education-opinion.

[82] Jeremy Weissman, ChatGPT Is a Plague Upon Education, Inside Higher Ed (Feb. 8, 2023), https://www.insidehighered.com/opinion/views/2023/02/08/chatgpt-plague-upon-education-opinion.

46. Professor Zhao also relies heavily on journalistic accounts describing AI-driven cheating as "rampant."[83] Academic misconduct is shaped by a range of contextual factors, including institutional policies, assessment formats, student incentives, and enforcement practices, rather than by LLMs alone. As I discussed in my Rebuttal Report, these challenges similarly arose with the introduction of previous technologies like calculators and the internet but were subsequently addressed.[84] Framing AI-enabled cheating as solely a technological problem risks obscuring these broader determinants and overstates the extent to which current challenges can be attributed to LLMs themselves.

47. Even in cases where Professor Zhao references educators who argue that AI may blunt students' ability to think, those same sources acknowledge that educators can create "tech-free spaces" and incentivize students to spend time in them.[85] This perspective aligns with my view that educators should carefully consider the appropriate conditions for the use of LLMs across educational stages, from early learners through college students and beyond. Part of this process involves educators taking deliberate steps to redesign assessments to mitigate perceived risks, including AI-enabled cheating.

48. In my opinion, the present moment should be understood as an early stage in an ongoing process of institutional adaptation. This does not imply that all risks have been resolved or that further work is unnecessary. Rather, it reflects the well-established pattern by which educational systems respond to new technologies through iterative changes in pedagogy, assessment, and governance. Whilst there is nothing speculative about the numerous examples that I have drawn from initial real-world uses of LLMs in the classroom and beyond, whether LLMs ultimately undermine or support educational objectives will depend on how these adaptations mature over time, not solely on conditions observed at an early point in their adoption.

---

[83] Zhao Rebuttal Report, at footnote 117, citing N.Y. MAG., James D. Walsh, Everyone Is Cheating Their Way Through College (May 7, 2025), https://nymag.com/intelligencer/article/openai-chatgpt-ai-cheating-education-college-students-school.html.

[84] *See* Altman Rebuttal Report, ¶¶ 37-40.

[85] Anastasia Berg, Why Even Basic A.I. Use Is So Bad for Students, New York Times (Oct. 29, 2025), https://www.nytimes.com/2025/10/29/opinion/ai-students-thinking-school-reading.html. ("Higher education aims to create cognitively mature adults, which in turn requires us to ensure students learn to read, think and write all on their own. It is easier than we think: Creating tech-free spaces and incentivizing students to spend time in them requires no new resources. All it takes is will. Many of our students still have it. Do their teachers?").

49. Professor Zhao similarly claims LLMs are reducing critical thinking and learning for both students and knowledge workers. He cites a survey of knowledge workers' use of GenAI to argue that increased reliance on these tools correlates with diminished critical thinking.[86] However, the study's own findings are far more nuanced. Rather than concluding that GenAI eliminates critical thinking, the authors observe a redistribution of cognitive effort toward tasks such as information verification, response integration, and task stewardship. They do not characterize this shift as an unmitigated risk but instead frame it as a design challenge and an opportunity to develop GenAI tools for knowledge work with more effective guardrails.[87] Other studies have shown that by integrating the use of LLMs with traditional teaching methods, students demonstrated "significant learning gains."[88]

50. There is no doubt that unrestricted use of LLMs by students can disrupt education. Educators and schools address these issues in a variety of ways. As discussed in my Rebuttal Report, many of the methods and approaches used by Stanford and other schools create an environment for taking appropriate advantage of LLMs, while also ensuring student learning outcomes.[89] The resources that we have developed aim to help educators adapt to new technologies, and as I have consistently maintained, this is a precedented outcome throughout history when it comes to new transformational technologies.

51. The same pattern of institutional adaptation that appears in educational settings is also evident in industry adoption. As with the educational usage of LLMs, initial challenges in industry reflect an adjustment period to a new tool rather than a definitive signal about its long-term effects. Companies who engage with LLM technologies are, and should be, treated as test cases for how to validate and deploy these applications. Given that broad

---

[86] *See* Zhao Rebuttal Report, ¶ 134. ("Controlled user studies involving LLMs show that students using LLMs experience lower cognitive load, and students using LLMs showed weaker reasoning and shallower understanding. Similarly, other controlled studies show that the more workers trust generative AI, the less there is critical thinking involved by the workers.").

[87] Hao-Ping (Hank) Lee, The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers (Apr. 25, 2025), https://doi.org/10.1145/3706598.3713778. ("Qualitatively, GenAI shifts the nature of critical thinking toward information verification, response integration, and task stewardship. Our insights reveal new design challenges and opportunities for developing GenAI tools for knowledge work.").

[88] Tarik El Fathi et. al., Integrating generative AI into STEM education: enhancing conceptual understanding, addressing misconceptions, and assessing student acceptance, Disciplinary and Interdisciplinary Science Education Research (Mar. 31, 2025), https://link.springer.com/article/10.1186/s43031-025-00125-z.

[89] *See* Altman Rebuttal Report, ¶ 42.

access to LLMs is less than four years old, these early proofs of concept indicate that continued refinement is likely to yield more reliable and broadly useful applications across many domains. This is also consistent with historical experience and observed patterns in other transformative technologies.

52. Professor Zhao places disproportionate weight on early failures and risks associated with LLM adoption while minimizing evidence of emerging successes.[90] Consistent with historical experience, a combination of technological innovation, market adaptation, institutional learning and, in appropriate cases, policy response can guide the deployment and oversight of these LLM tools and technologies as they mature..

V. PROFESSOR ZHAO DISREGARDS THE CONNECTION BETWEEN DIVERSE TRAINING DATA AND LLM PERFORMANCE

53. Finally, Professor Zhao implies that I do not adequately connect any of the benefits of LLMs to the inclusion of song lyrics in LLM training data.[91] This framing suggests a misunderstanding of how LLMs are developed and perform in practice. LLMs derive their capabilities from training on and exposure to vast amounts of heterogeneous text pulled from diverse sources reflecting the full range of human language use, including creative, metaphorical, and non-standard expressions.[92] Training on diverse linguistic sources enables LLMs to develop more general representations of language and the world, [93] which in turn supports robustness, flexibility, and safe deployment across a wide range of applications.[94] Diversity in training data also plays an important role in improving model

[90] For example, Professor Zhao cites an MIT study reporting that a high percentage of generative AI pilots fail but omits the study's conclusion that these failures are primarily driven by organizational learning gaps, poor integration with existing systems, and misaligned expectations, rather than by inherent limitations of the technology itself. See Zhao Rebuttal Report, ¶ 141.

[91] *See* Zhao Rebuttal Report, ¶ 142. ("Dr. Altman does not connect any of the purported benefits of LLMs he describes to the inclusion of song lyrics (either Publishers' lyrics specifically or song lyrics more generally) in LLM training data or their output by Claude. Nor would there be any reason to expect that including lyrics would significantly affect the use cases.").

[92] *See* Altman Opening Report, ¶ 26. ("As models have drawn from larger training sets, with larger context windows, more parameters, and more advanced alignment methods, they have developed sophisticated language understanding and reasoning that has allowed the models' predictive capabilities to become increasingly specific and nuanced.").

[93] *See* Deposition of Jared Kaplan (Nov. 21, 2025), at 155:1-4. ("[THE DEPONENT:] I agree that it's valuable to have a diversity of different forms of linguistic expression so that Claude understands the world as generally as possible.").

[94] *See* Deposition of Ben Mann (Nov. 20, 2025), at 115:23-116:18 ("Q: If Anthropic doesn't consider it important to include song lyrics in Claude's training data, why does Anthropic include song lyrics in Claude's Training Data? A: So, in general, as I mentioned, our goal is to train helpful, harmless, and honest models, and push the frontier of safety research so that humanity's transition from society today to one in which there's super-intelligent AI is a safe

performance and reliability, including by reducing the likelihood that an LLM over-relies on or memorizes a single category of content. [95] My team and I used a similar approach in fine-tuning our FDA bioequivalence model. We used LLMs to generate diverse paraphrased versions of the same underlying content to improve generalization and minimize the risk that the model would memorize or over-rely on any single category of training material.

54. From this perspective, the value of any individual class of training material, including song lyrics, cannot be evaluated in isolation, but must be viewed as part of a broader ecosystem of language data that collectively supports generalization and responsible use. From this perspective, the value of any individual class of training material, including song lyrics, cannot be evaluated in isolation, but must be viewed as part of a broader ecosystem of language data that collectively supports generalization and responsible use. The same holistic understanding of training applies equally to assessments of benefits and potential harms, which emerge from model behavior in context rather than from any single category of training input.

VI. CONCLUSION

55. After reviewing the Zhao Rebuttal Report, the opinions I expressed in my Opening Report and Rebuttal Report remain unchanged.

56. As discussed throughout this report, my opinions are grounded in decades of interdisciplinary experience evaluating computational tools in real-world scientific, medical, regulatory, and policy contexts, as well as direct engagement with LLMs in high-stakes domains. While LLMs present legitimate challenges that warrant careful responsible deployment and ongoing refinement, the issues Professor Zhao emphasizes are characteristic of early technological adoption and do not negate the substantial and growing evidence of their beneficial applications. When appropriately designed, validated,

---

and beneficial one for humanity. In order to do that, we need to train frontier models with the best capabilities from a knowledge/work standpoint, from a hard sciences standpoint. Many of our customers work on hard sciences, and we believe that if things go well, then this technology could transform people's lives for the better. Those are the kinds of things that we really care about, and we've talked about that in length at public.").

[95] *See* Deposition of Jared Kaplan (Nov. 21, 2025), at 152:6-22. ("Well, I wrote in the declaration that 'this is yet another reason to have a diverse and voluminous corpus - having a high volume of diverse training data decreases the likelihood the model will inadvertently memorize certain content, and the more distinct works are included in the training set, the less insight each individual work provides as a proportion of the entire corpus.' So I think what I said there is that each individual work is really of vanishing importance, diversity is important, and song lyrics are simply an example that may be useful and that one can discuss, among millions of other possibilities.").

and governed with human oversight, LLMs such as Claude augment expertise, support scientific discovery, improve efficiency in regulated domains, and expand educational and public-serving capabilities. In my opinion, the overall body of evidence continues to demonstrate that, if responsibly managed, LLMs have significant potential to deliver meaningful benefits.

57. 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.



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 Reply Report of Russ Altman, dated February 13, 2026.

Table 1: Materials Considered

Internal documents	
1	Deposition of Ben Mann, Transcript (Nov. 20, 2025)
2	Deposition of Jared Kaplan, Transcript (Nov. 21, 2025)
Academic Sources	
3	Abel Chandra et. al., Transformer-based deep learning for predicting protein properties in the life sciences (Jan. 18, 2023)
4	Anthropic_0000594245
5	Ashish Vaswani et. al., Attention Is All You Need (Aug. 2, 2023)
6	Charlotte Hughes, ESM-2 vs ProtBERT: A Comprehensive Performance Comparison for Protein Function Prediction and Biomedical Applications, Protein Engineering Insights (Nov. 26, 2025)
7	Delaney A. Smith et al., Monitoring the opioid epidemic via social media discussions (May 15, 2025)
8	Donald L. McCabe et al., Honor Codes and Other Contextual Influences on Academic Integrity: A Replication and Extension to Modified Honor Code Settings (June 2002)
9	Filip Obradovic, Measuring Diagnostic Test Performance Using Imperfect Reference Tests: A Partial Identification Approach (Aug. 2024)
10	Gabriel Alain et al., Evaluating User Interactions and Adoption Patterns of Generative AI in Health Care Occupations Using Claude: Cross-Sectional Study (May 30, 2025)
11	Gowri Nayar, Alp Tartici and Russ B Altman, Paying attention to attention: High attention sites as indicators of protein family and function in language models (Sep. 12, 2025)
12	Hao-Ping (Hank) Lee, The Impact of Generative AI on Critical Thinking: Self-Reported Reductions in Cognitive Effort and Confidence Effects From a Survey of Knowledge Workers (Apr. 25, 2025)
13	Kara Liu, Russ Altman and Vasilis Syrgkanis, Detecting clinician implicit biases in diagnoses using proximal causal inference (2025)
14	Kristy A. Carpenter and Russ B. Altman, Using GPT-3 to Build A Lexicon of Drugs of Abuse Synonyms for Social Media Pharmacovigilance (Feb. 17, 2023)
15	Renjie Li at. al, What is next for LLMs? Pushing the boundaries of next-gen AI computing hardware with photonic chips (Oct. 6, 2025)
16	Richard Frankel et al., Effects of exam-room computing on clinician-patient communication (Aug. 2005)
17	Rishi Bommasani et al., On the Opportunities and Risks of Foundation Models (July 12, 2022)
18	Russ B. Altman, Exclusion Methods for Determination of Protein Structure from Experimental Data (1990)
19	Subhankar Maity and Manob Jyoti Saikia, Large Language Models in Healthcare and Medical Applications: A Review (June 9, 2025)

20	Tarik El Fathi et. al., Integrating generative AI into STEM education: enhancing conceptual understanding, addressing misconceptions, and assessing student acceptance, Disciplinary and Interdisciplinary Science Education Research (Mar. 31, 2025)
21	Vamsi Krishna Uppalapati and Deb Sanjay Nag, A Comparative Analysis of AI Models in Complex Medical Decision-Making Scenarios: Evaluating ChatGPT, Claude AI, Bard, and Perplexity (Jan. 18, 2024)
22	Victor R. Lee et al., Cheating in the age of generative AI: A high school survey study of cheating behaviors before and after the release of ChatGPT (June 12, 2024)
23	Yinggan Xu et al., Advancing AI-Scientist Understanding: Multi-Agent LLMs with Interpretable Physics Reasoning (Aug. 18, 2025)
Public Sources	
24	Ahti Valtteri, Why Imperfect Tools are Often Better Than None: A Business Perspective (Aug. 15, 2024)
25	AlphaFold, Protein Structure Database
26	Altman Testimony, Senate Testimony "AI's Potential to Deliver Better Patient Care and Accelerate Drug Discovery" presented to U.S. Senate Committee on Health, Education, Labor, and Pensions, (Oct. 9, 2025)
27	Amazon Web Services, What is RAG?
28	Anastasia Berg, Why Even Basic A.I. Use Is So Bad for Students, New York Times (Oct. 29, 2025)
29	Anthropic, A framework for AI development transparency (July 7 2025)
30	Anthropic, Advancing Claude in healthcare and the life sciences (Jan 2026)
31	Anthropic_0000594350
32	Bruce Buchanan, Google Scholar Profile
33	Carrie Spector, What do AI chatbots really mean for students and cheating? (Oct. 31, 2023)
34	Cole Stryker, IBM, What are large language models (LLMs)?
35	Dario Amodei, The Adolescence of Technology - Confronting and Overcoming the Risks of Powerful AI (Jan. 2026)
36	Executive Office of the President, EO 14110, Safe, Secure, and Trustworthy Development and Use of AI (Oct. 30, 2023)
37	Google DeepMind, AlphaFold
38	Alexis Balvair, HealthPoint, Claude AI in Healthcare – Review (Jan. 15, 2026)
39	James D. Walsh, Everyone Is Cheating Their Way Through College (May 7, 2025)
40	Jeremy Weissman, ChatGPT Is a Plague Upon Education, Inside Higher Ed (Feb. 8, 2023)
41	Michele L. Kreidler, The Development of Healthcare and Information Technology (2023)
42	Neil Vigdor and Hannah Ziegler, Their Professors Caught Them Cheating. They Used A.I. to Apologize, The New York Times (Oct. 29, 2025)
43	PipeBio, Protein language models: promises, pitfalls and applications (June 18, 2024)
44	Stanford, Bulletin ExploreCourses, BMDS 224: Principles of Pharmacogenomics
45	Stanford, Bulletin ExploreCourses, CS 272: Introduction to Biomedical Informatics Research Methodology
46	Stanford, Bulletin ExploreCourses, CS 274: Representations and Algorithms for Computational Molecular Biology
47	Stanford HAI, About - Who We Are
48	Stanford HAI, Congressional Boot Camp on AI

49	Stanford HAI, How Disruptive Is DeepSeek? Stanford HAI Faculty Discuss China's New Model (Feb. 13, 2025)
50	Stanford University, Faculty Handbook, Chapter 2: Appointments, Reappointments and Promotions in the Professoriate (last updated Jan. 8, 2026)
51	The Hoover Institution, Profiles, Russ B. Altman
52	U.S. Department of Homeland Security, Notice, Establishment of the Artificial Intelligence Safety and Security Board (Apr. 29, 2024)
53	U.S. Department of Homeland Security, Roles and Responsibilities Framework for Artificial Intelligence in Critical Infrastructure (Nov. 14, 2024)
54	U.S. FDA, Statistical Guidance on Reporting Results Evaluating Diagnostic Tests – Guidance for Industry and FDA Staff (Mar. 13, 2007)
55	University of Hawai'i, Practices of Science False Positives and False Negatives
56	Venktesh Ramnath, The Conversation, Medicine is an imperfect science – but you can still trust its process (Sept. 7, 2021)
Expert Sources	
57	Expert Report of Ben Zhao, Concord Music Group, et al., v. Anthropic PBC, Case No. 24-cv-03811-EKL-SVK (Dec. 19, 2025)
58	Expert Report of Russ Altman, Concord Music Group, et al., v. Anthropic PBC, Case No. 24-cv-03811-EKL-SVK (Dec. 19, 2025)
59	Expert Rebuttal Report of Ben Zhao, Concord Music Group, et al., v. Anthropic PBC, Case No. 24-cv-03811-EKL-SVK (Jan. 23, 2026)
60	Expert Rebuttal Report of Russ Altman, Concord Music Group, et al., v. Anthropic PBC, Case No. 24-cv-03811-EKL-SVK (Jan. 23, 2026)