EXHIBIT 8

UNITED STATES DISTRICT COURT

NORTHERN DISTRICT OF CALIFORNIA

SAN JOSE DIVISION

CONCORD MUSIC GROUP, INC., ET AL., Plaintiffs, v. ANTHROPIC PBC, Defendant.	Case Number: 5:24-cv-03811-EKL-SVK Judge Eumi K. Lee Magistrate Judge Susan van Keulen

EXPERT REPORT OF DR. BEN ZHAO

December 19, 2025

Contains Information Designated as
Confidential and/or Highly Confidential—Attorneys' Eyes Only

lyrics can be much more efficient, using optimizations that combined similar lyrics in a single scan, or utilized an efficient implementation using GPUs, or using fast hashing techniques like fuzzy fingerprint hashes.[229]

155. In short, when it undertakes its next update of Claude, Anthropic can train Claude on a dataset that does not include Publishers' lyrics. This is the only foolproof way to ensure that Claude will not be able to include this content in its outputs.

VII. Risks from AI

156. AI development poses significant risks to a number of different stakeholders, ranging from the disruptive to the catastrophic. Anthropic and its employees have opined extensively on these risks, and I have conducted significant research in this area over the course of the last three years, publishing a number of peer-reviewed articles in premier publication venues in security and machine learning,[230] and receiving multiple awards in the process.[231] My work has largely focused on how to provide solutions that quantify and mitigate such harms, particularly to human creatives such as artists. In addition, I have conducted multiple meetings with different

---

[229] *See* Feng Zhou et al., *Approximate object Location and spam filtering on Peer-to-peer systems,* PROC. OF ACM/IFIP/USENIX INTL. MIDDLEWARE CONF. (2003), https://disco.ethz.ch/courses/ws0506/seminar/papers/peer_spam_filtering.pdf; *see also* Udi Manber, *Finding Similar Files in a Large File System*, Winter USENIX Technical Conference (1994), https://www.cs.princeton.edu/courses/archive/spr05/cos598E/bib/manber94finding.pdf.

[230] *See, e.g.*, Ben Y. Zhao, et al., *Somesite I Used to Crawl: Awareness, Agency and Efficacy in Protecting Content Creators from AI Crawlers*, ACM Internet Measurement Conference 2025 (Oct. 2025), https://people.cs.uchicago.edu/~ravenben/publications/pdf/crawlers-imc25.pdf; Ben Y. Zhao, et al., *Organic or Diffused: Can We Distinguish Human Art from AI-generated Images?* PROC. OF THE ACM CONF. ON COMPUTER AND COMM'N SEC. 2024 (Oct. 2024), https://people.cs.uchicago.edu/~ravenben/publications/pdf/organic-ccs24.pdf; Ben Y. Zhao, et al., *Nightshade: Prompt-Specific Poisoning Attacks on Text-to-Image Generation Models*, PROCEEDINGS OF THE 45TH IEEE SYMPOSIUM ON SECURITY AND PRIVACY (May 2024), https://arxiv.org/pdf/2310.13828; etc.

[231] *See, e.g. The 100 Most Influential People in AI 2024*, TIME Magazine, https://time.com/7012857/ben-zhao/ (awarded to Ben Zhao, inventor of Nightshade); *CCS 2024 Distinguished Paper Awards*, Association for Computing Machinery Special Interest Group on Security, Audit and Control, https://www.sigsac.org/ccs/CCS2024/program/awards.html (awarded to Ben Zhao, and others, for the paper "Organic or Diffused: Can We Distinguish Human Art from AI-generated Images?"); etc.

stakeholders to help them understand and address some of these harms produced from the usage of LLMs, including federal regulatory agencies (U.S. Copyright Office, Federal Trade Commission, Dept. of Defense, Dept. of State), multiple universities, multiple labor organizations representing artists, actors, writers and musicians, and state legislatures in Illinois and California.

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.[232] These harms are real and measurable today, even though many 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.

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.[233] 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.[234]

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. In the early days of LLM usage, one of the first organizations to disclose harms from LLMs was writers. Specifically, writer communities who are being inundated with a massive volume of AI-generated text. For example, in February 2023, Clark's World, the sponsor of the well-regarded competition for fictional writing, announced that

---

[232] *See, e.g.*, J. Kaplan Dep. Tr. 85:10–15.
[233] *See* D. Ganguli 30(b)(1) Dep. Tr. 80:22–81:16.
[234] *Large Language Models Pose a Risk to Society and Need Tighter Regulation, Say Oxford Researchers*, OXFORD UNIVERSITY (Aug. 7, 2024), https://www.ox.ac.uk/news/2024-08-07-large-language-models-pose-risk-society-and-need-tighter-regulation-say-oxford; Sandra Wachter et al., *Do Large Language Models Have a Legal Duty to Tell the Truth?*, 11 (8) R. SOC. OPEN. SCI. (2024), https://royalsocietypublishing.org/rsos/article/11/8/240197/92624/Do-large-language-models-have-a-legal-duty-to-tell.

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.[235]

160. Similar trends were quickly found on more commercial platforms like Amazon, where the vibrant book market was flooded with unlabeled, low quality AI-books that were turning away customers who could not immediately identify which were human written books without first paying for the purchase.[236] Beyond just outputting an overwhelming volume of AI slop, it is clear that LLMs were overwhelmingly used by bad actors to target and take away profits from individual authors by replicating, imitating, or summarizing human written books and trying to displace them on the marketplace.[237] Unfortunately, without significant changes by AI model trainers, this wave of AI slop is unlikely to stop any time soon.[238]

161. When we consider the cost of this large volume of AI slop, it is critical to understand that they do not come for free. Anthropic (and other AI companies) burn large quantities of energy in order to generate the wave of AI slop now overwhelming the Internet.[239] There are massive costs involved in both training and operating LLM models like Claude. Research studies estimate that training a model such as GPT-4 used about 44 GWh, which is

---

[235] Vincent Acovino et al., *A Sci-Fi Magazine Has Cut Off Submissions After a Flood of AI-Generated Stories*, NPR (Feb. 24, 2023), https://www.npr.org/2023/02/24/1159286436/ai-chatbot-chatgpt-magazine-clarkesworld-artificial-intelligence.
[236] Victor Tangermann, *Amazon is Being Flooded With Books Entirely Written by AI*, FUTURISM (May 6, 2023), https://futurism.com/the-byte/amazon-flooded-books-written-by-ai.
[237] *See, e.g.,* Kate Knibbs, *Scammy AI-Generated Book Rewrites Are Flooding Amazon*, WIRED (Jan. 20, 2024), https://www.wired.com/story/scammy-ai-generated-books-flooding-amazon/; Will Oremus, *He Wrote a Book On a Rare Subject. Then a ChatGPT Replica Appeared on Amazon*, WASH. POST (May 5, 2023), https://www.washingtonpost.com/technology/2023/05/05/ai-spam-websites-books-chatgpt/;
[238] *See* Max Read, The Internet's AI Slop Problem Is Only Going to Get Worse, N.Y. MAG. (Sept. 25, 2024), https://nymag.com/intelligencer/article/ai-generated-content-internet-online-slop-spam.html.
[239] Nidhal Jegham et al., *How Hungry is AI? Benchmarking Energy, Water, and Carbon Footprint of LLM Inference*, arXiv:2505.09598 (Nov. 24, 2025), https://arxiv.org/pdf/2505.09598.

enough to power San Francisco for days, or a small city for a month.[240] Current estimates put a single LLM query at around 0.3 Wh, roughly equivalent to running a desktop computer for 36 minutes. Another research study finds that each Claude query produces about 3.5 grams of CO2, roughly equivalent to watching 35 minutes of YouTube videos.[241]

162. Looking at the broad impact on society, Claude and LLMs are also disrupting education at every level. LLMs such as Claude provide an easy way for students of all ages to quickly generate text for homework for a variety of subjects, particularly research and creative writing assignments, where Claude can quickly generate entire essays in seconds with minimal prompting from a student.[242] This is already having a dramatic impact on K-12 schools, where AI cheating runs rampant, particularly for school districts with limited resources to monitor and detect such behavior.[243]

163. The widespread and unregulated proliferation of LLMs is also making waves among higher education as well. For example, a recent study shows that many or most college students have used LLMs in some way, with many admitting that LLMs have replaced the learning process in numerous classes.[244] High profile cases of student plagiarism at universities are also becoming increasingly common, including among the most selective and prestigious schools.[245] Educators across the world are now grappling with the challenge of adapting to teaching in a world

---

[240] James O'Donnell and Casey Crownhart, *We Did the Math on AI's Energy Footprint. Here's the Story You Haven't Heard*, MIT TECH. REV. (May 20, 2025), https://www.technologyreview.com/2025/05/20/1116327/ai-energy-usage-climate-footprint-big-tech/.

[241] Christine Hall, *What's Your Chatbot's Carbon Footprint?*, FOSSFORCE (Apr. 3, 2025), https://fossforce.com/2025/04/whats-your-chatbots-carbon-footprint/.

[242] *See, e.g.*, D. Ganguli 30(b)(1) Dep. Tr. 193:9–195:17, 196:23–202:6.

[243] *Confronting AI Cheating in K-12 While Embracing Responsible Innovation*, EDCIRCUIT (Sep. 3, 2025), https://edcircuit.com/confronting-ai-cheating-in-k-12-while-embracing-responsible-innovation/; *see also, e.g.*, D. Ganguli 30(b)(1) Dep. Tr. 193:9–195:17, 196:23–.197:22

[244] James D. Walsh, *Everyone is Cheating Their Way Through College,* N.Y. MAG. (May 7, 2025), https://nymag.com/intelligencer/article/openai-chatgpt-ai-cheating-education-college-students-school.html.

[245] *See, e.g.,* Angela Yang, *Kicked Out of Columbia, This Student Doesn't Plan to Stop Trolling Big Tech With AI*, NBC NEWS (Mar. 27, 2025), https://www.nbcnews.com/tech/tech-news/columbia-university-student-trolls-big-tech-ai-tool-job-applications-rcna198454.

where cheating via AI tools is common.[246] Attempts to detect AI-based plagiarism is extremely challenging, and schools are struggling with the need to balance academic integrity with the potential of harming innocent students with plagiarism charges stemming from false positives in AI-detection tools.[247]

164. Beyond the classroom, the use of Claude and other LLMs poses risks of physical harm to young people by offering tools and advice for self-harm. Because LLMs scrape and train on content indiscriminately, they absorb facts about dangerous substances, as well as mechanisms for self-harm. While Claude tries to block some of these subject matters from user prompts, research has found that these guardrails are trivial to overcome. For example, a recent study from Northeastern University found Claude readily provided specific information about self-harm.[248] In a more formal study published in Psychiatric Services, a journal of the American Psychiatric Association, researchers found that Claude answered questions about ropes, firearms, or poisons and their success rate in self-harm.[249]

165. LLMs are also effective tools in generating misinformation campaigns online. LLMs are found to be extremely convincing in producing misinformation campaigns (*e.g.*, politically-motivated misinformation attacks, conspiracy theories), and at the same time, have very

---

[246] Alice Evans, *How Should Academics Respond to AI?*, GGD (May 19, 2024), https://www.ggd.world/p/how-should-academics-respond-to-ai.
[247] *See* Will Coldwell, '*I Received a First But It Felt Tainted and Undeserved': Inside the University AI Cheating Crisis*, THE GUARDIAN (Dec. 15, 2024), https://www.theguardian.com/technology/2024/dec/15/i-received-a-first-but-it-felt-tainted-and-undeserved-inside-the-university-ai-cheating-crisis; *AI Detectors Don't Work. Here's What to Do Instead*, MIT MGMT STS TEACHING & LEARNING TECH., https://mitsloanedtech.mit.edu/ai/teach/ai-detectors-dont-work/.
[248] Cody Mello-Klein, *New Northeastern Research Raises Concerns Over AI's Handling of Suicide-Related Questions*, NORTHEASTERN GLOBAL NEWS (July 31, 2025), https://news.northeastern.edu/2025/07/31/chatgpt-suicide-research/.
[249] Matt O'Brien and Barbara Ortutay, *Study Says AI Chatbots Need to Fix Suicide Response, As Family Sues Over ChatGPT Role in Boy's Death*, AP NEWS (Aug. 26, 2025), https://apnews.com/article/ai-chatbots-selfharm-chatgpt-claude-gemini-da00880b1e1577ac332ab1752e41225b.

limited effectiveness when used as a detector for misinformation.[250] Indeed, native halluncinations generated by LLMs are difficult to detect.[251] 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."[252]

166. Finally, it is worth noting that LLMs pose significant risks to entire populations and nation states, by enabling AI agents that can simplify and automate the generation of cyber attacks. Again, these are not hypothetical, but real harms happening today. In the research realm, a number of groups have tested and evaluated LLMs' vulnerability to be taken over and weaponized by attackers. The most recent study conducted by Carnegie Mellon on realistic enterprise network environments showed that LLMs are easily weaponized to autonomously plan and execute complex, multi-step network attacks without human instruction.[253] Another study tested AI agents built on existing LLMs, including Claude 4, and showed that such agents could be compromised using a range of attacks (*e.g.*, prompt-injection, attack via RAG systems and inter-agent interactions). These agents would accept and execute malicious commands such as installing and executing malware on victim machines.[254]

---

[250] Yulong Ma et al., *Linguistic Features of AI Mis/Disinformation and the Detection Limits of LLMs*, NATURE COMMC'NS. (DEC. 11, 2025), https://www.nature.com/articles/s41467-025-67145-1_reference.pdf;
[251] *See* Sebastian Farquhar, et al., *Detecting hallucinations in large language models using semantic entropy*, NATURE 630 (2024), 625, https://www.nature.com/articles/s41586-024-07421-0.pdf.
[252] *Agentic Misalignment: How LLMs could be insider threats,* ANTHROPIC (June 20, 2025), https://www.anthropic.com/research/agentic-misalignment.
[253] Michael Cunningham, *When LLMs Autonomously Attack*, CARNEGIE MELLON UNIVERSITY COLLEGE OF ENGINEERING (Jul. 24, 2025), https://engineering.cmu.edu/news-events/news/2025/07/24-when-llms-autonomously-attack.html. *See also* Brian Singer et al., *Incalmo: An Autonomous LLM-Assisted System for Red Teaming Multi-Hose Networks*, arXiv:2501.16466 (Nov. 22, 2025), https://arxiv.org/pdf/2501.16466.
[254] *See* Matteo Lupinacci, et al, *The Dark Side of LLMs Agent-based Attacks for Complete Computer Takeover*, arXiv:2507.06850v3 (Jul. 11, 2025), https://arxiv.org/pdf/2507.06850v3.

167. In another recent report, Anthropic disclosed that in September 2025, a Chinese state-sponsored group utilized Claude to conduct a "highly sophisticated cyber espionage operation."[255] The espionage group used Claude Code to support reconnaissance, vulnerability discovery, credential harvesting, data analysis, and exfiltration operations.[256] Notably, the "AI autonomously discovered vulnerabilities in targets selected by human operators and successfully exploited them in live operations, then performed a wide range of post-exploitation activities from analysis, lateral movement, privilege escalation, data access, to data exfiltration," without the need for significant human intervention.[257] In each case, Claude code was used across multiple phases of the attacks, including reconnaissance (mapping target networks and identifying high-value systems), vulnerability and exploitation (generating and testing exploit code for discovered vulnerabilities), credential harvesting (systematically extracting and testing credentials), data exfiltration (querying databases, extracting sensitive data including personal records, financial information, and government credentials, and categorizing it by intelligence value), and documentation (generating detailed attack documentation to enable seamless handoff and future campaign resumption).[258] A major saving grace in this case was that Claude also suffered from hallucinations, which limited the level of autonomy in these attacks and required human input in some cases.

168. Finally, there is the significant risk that AI will pose to large swathes of the global economy and job market. I understand that Anthropic's CEO has opined extensively on the likelihood that AI will replace a majority of entry level jobs and lead to massive unemployment

[255] Anthropic, *Disrupting the first reported AI-orchestrated cyber espionage campaign* (Nov. 2025), https://assets.anthropic.com/m/ec212e6566a0d47/original/Disrupting-the-first-reported-AI-orchestrated-cyber-espionage-campaign.pdf.
[256] *Id.*
[257] *Id.*
[258] Mary Cunningham, *Anthropic Says Chinese Hackers Used Its Claude AI Chatbot in Cyberattacks*, CBS NEWS (Nov. 14, 2025), https://www.cbsnews.com/news/anthropic-chinese-cyberattack-artificial-intelligence/.

within the next five years, calling it a potential "bloodbath"[259] and warning that "this is already happening."[260] I understand that other Anthropic employees have expressed similar views on the high likelihood that such "knowledge work" will be automated through AI within the next five years.[261] I agree that the possibility of these things occurring is high, including for the reasons discussed above. But regardless of whether LLMs become sophisticated enough to replace skilled workers, it is true that they are already disrupting and replacing entry-level positions in a number of industries. In doing so, they are breaking the skill training pipeline for human workers, making it much more challenging for any entrants to the workforce from learning the skills necessary to become an expert.[262]

[259] Jim VandeHei and Mike Allen, *Behind the Curtain: A white-collar bloodbath*, AXIOS (May 28, 2025), https://www.axios.com/2025/05/28/ai-jobs-white-collar-unemployment-anthropic. *See also* Anderson Cooper and Nichole Marks, *Why Anthropic CEO Dario Amodei spends so much time warning of AI's potential dangers*, CBS NEWS (Nov. 16, 2025), https://www.cbsnews.com/news/anthropic-ceo-dario-amodei-warning-of-ai-potential-dangers-60-minutes-transcript/.
[260] Samantha Delouya, *AI CEO says technology 'moving very quickly,' could soon replace more jobs*, CNN (Sept. 17, 2025), https://www.cnn.com/2025/09/17/business/anthropic-warns-ai-could-soon-replace-jobs.
[261] *See* J. Kaplan Dep. Ex. 245 (Anthropic_0000428963).
[262] Walther, Cornelia C., *Is AI Pushing Us to Break the Talent Pipeline?*, KNOWLEDGE AT WHARTON (Aug. 12, 2025), https://knowledge.wharton.upenn.edu/article/is-ai-pushing-us-to-break-the-talent-pipeline/.

Dated: December 19, 2025

Ben Zhao