+ View Split Screen   Expand Threads

All ⌄   Search entries or author...   Newest First ⌄

Due May 4 11:59pm   Available from Apr 28   45 points possible

38 Replies, 19 Unread

Unit 7: Discussion

PARK UNIVERSITY

DISCUSSION

Directions

Discuss your reviews on the risks and dangers of Natural Language Processing, including LLMs. What are you most concerned about? Do you think any concerns are over-hyped? How can risks be mitigated? Respond to two posts with your viewpoints.

Criteria for Success

Initial Post (DUE: Thursday 11:59 p.m. CT)

- In the initial post you will do the following:
  - Uses the weekly materials to construct an academic argument that addresses the discussion question in a thorough and logical manner.
  - Correctly uses key terms and concepts. Thoroughly addresses all components of the prompt. Ideas are clear and on-topic.

    - Follows grammar conventions. The writing is concise and easy to read.
    - Writes approximately 200 words.

Response to Two Peers (DUE: Sunday, 11:59 CT)

- In each response, you will do the following:
    - Furthers the conversation by asking thoughtful questions, responding directly to statements of others, and contributing additional analysis. Builds on peers’ contributions by presenting logical viewpoints or challenges.
    - Follows grammar conventions. The writing is concise and easy to read.
    - Writes approximately 100 words.

Please review the rubric for this assignment before beginning to ensure that you earn full credit. Contact me if you have any questions.

Reply

KK **Kouame Hermann Kouame (https://canvas.park.edu/courses/85581/users/123444)**

May 2 12:07am | Last reply May 2 4:03pm

Hello Class,here is my discussion post for this week!

The field of Natural Language Processing (NLP) has undergone rapid transformation, particularly with the rise of Large Language Models (LLMs) like OpenAI's ChatGPT. Built on the transformer architecture introduced in 2017, ChatGPT leverages attention mechanisms to better retain context across long sequences advancement over earlier RNN-based models like LSTMs. The acronym GPT stands for Generative Pretrained Transformer, highlighting its ability to generate text, its pretraining on large corpora, and the neural network architecture it utilizes. One key innovation is RLHF (Reinforcement Learning from Human Feedback), which refines the model’s conversational ability by aligning its responses with human preferences.(*Natural Language Processing (NLP): What It Is and Why It Matters*, n.d.)

LLMs now dominate core NLP tasks such as sentiment analysis and text summarization. Using OpenAI’s API, models like GPT-3.5 can classify review sentiments and condense lengthy technical texts, demonstrating versatility and performance. Alternatives like

HuggingFace provide similar functionalities locally, which appeals to organizations with data privacy concerns. Additionally, prompt engineering techniques, including zero-shot and few-shot learning, as well as Chain of Thought reasoning, enable more tailored and interpretable outputs. While OpenAI models generally outperform open-source counterparts, both serve essential roles in modern NLP workflows.

Sources:

Natural Language Processing (NLP): What it is and why it matters. (n.d.). SAS. https://www.sas.com/en_us/insights/analytics/what-is-natural-language-processing-nlp.html

› 1 Reply, 1 Unread | Reply | Mark as Unread

AA **Atit Adhikari (https://canvas.park.edu/courses/85581/users/126504)** ⋮

May 1 10:45pm | Last reply May 2 3:57pm

Natural Language Processing using large language models (LLMs) has various advantages along with enormous risks. There are various risks and dangers of LLM, among them the most significant is the issue of bias and fairness. LLMs learn from massive datasets, many of which contain biases related to gender, race, socioeconomic status, and more. These biases can result in unfair treatment in applications like hiring, healthcare, or legal decision-making (Guo et al., n.d.).

Misinformation is also a significant problem. LLMS can provide incorrect or fabricated information, despite their confidence. This is particularly risky in high-risk domains such as medicine or legal counsel, where individuals depend on good and correct information. It is even more risky when users over-depend on what the AI states without verifying facts (Promptfoo, 2025).

That said, not all concerns are equally serious. In my view, relying too much on AI is probably one of the most overhyped. While it's true that depending on AI can be a problem, most people still use their own judgment and remain cautious when using AI tools. In important areas like healthcare and law, people usually double-check the information, knowing that accuracy is critical. To manage the risks, developers should focus on ethical design by using clean, fair data and regularly checking their models for bias. Clear rules and human supervision can help make sure AI is used in a safe and responsible way (Cohere Team, 2025).

Reference

Cohere Team. (2025, February 5). *LLM security risks and how to mitigate them*. Cohere. https://cohere.com/blog/llm-security (https://cohere.com/blog/llm-security)

Guo, Y., Guo, M., Su, J., Yang, Z., Zhu, M., Li, H., Qiu, M., & Liu, S. S. (n.d.). *Bias in large language models: Origin, evaluation, and mitigation*. Department of Statistics, The George Washington University. https://arxiv.org/abs/2402.13635 (https://arxiv.org/abs/2402.13635)

Promptfoo. (2025, March 19). *Misinformation in LLMs—Causes and prevention strategies*. https://www.promptfoo.dev/blog/misinformation/ (https://www.promptfoo.dev/blog/misinformation/)

> 2 Replies, 2 Unread | Reply | Mark as Unread

GK

George Kumi (https://canvas.park.edu/courses/85581/users/117082)

May 1 10:28pm | Last reply May 2 3:53pm

Hello class,
Kindly find below my discussion post for week 7

Risks and Dangers of Natural Language Processing and LLMs

Natural Language Processing (NLP) and large language models (LLMs) provide powerful tools for automating and enhancing communication. However, these technologies present significant risks. As stated by DeepLearning.AI (n.d.), NLP applications from chatbots to content generation have advanced rapidly, yet ethical concerns remain largely unresolved.

A primary concern is algorithmic bias. Since LLMs learn from massive datasets scraped from the internet, they often inherit and amplify societal biases (DeepLearning.AI, n.d.). For example, models may reproduce gender stereotypes by associating leadership with men and caregiving with women. This not only affects the fairness of NLP systems but also undermines trust in their outputs.

Another danger is misinformation. LLMs can generate text that appears factually sound but may be entirely fabricated, especially in domains like politics or healthcare. This raises questions about accountability and the potential for harm when false narratives spread at scale.

While some concerns like fears of AI replacing all human jobs or becoming sentient may be overstated, risks related to privacy breaches, bias, and content misuse are immediate and serious. Mitigation strategies include bias audits, transparent model training, and human-in-the-loop oversight to ensure outputs are reviewed in high-risk contexts.

In conclusion, NLP technologies must be developed and deployed with strong ethical frameworks to safeguard fairness, accuracy, and human dignity.

Reference

DeepLearning.AI. (n.d.). *Natural language processing specialization*. https://www.deeplearning.ai/resources/natural-language-processing (https://www.deeplearning.ai/resources/natural-language-processing)

› 3 Replies, 3 Unread | Reply | Mark as Unread

Battulga Bolormaa (https://canvas.park.edu/courses/85581/users/68062)

May 1 8:54pm | Last reply May 2 3:51pm

Discuss your reviews on the risks and dangers of Natural Language Processing, including LLMs.

After reviewing this week's material, I would consider the following potential risks as follows:

Data privacy concerns- when improperly configured, these models can inadvertently leak sensitive information between different customer contexts. Organizations face challenges in maintaining data segregation, with private data from one customer appearing in responses generated for another. Another concern is a **lack of accountability**. The "black box" nature of LLMs creates significant accountability challenges. Traditional top-down accountability models struggle with AI systems, as determining responsibility when these systems make poor decisions remains to be seen. The complexity of neural networks makes their behavior nearly impossible to understand, unlike more explainable AI models (Jacobi, 2024).

What are you most concerned about?

My main concern could be Bias and Misinformation.

LLMs exhibit various forms of bias, including demographic, cultural, and linguistic biases. These biases arise from training data that over-represents or under-represents certain groups.

Temporal biases occur when training data is restricted to limited periods, affecting the model's ability to report current events accurately. The models can also generate convincing fake news and disinformation, raising concerns about their potential misuse (Jacobi, 2024).

Do you think any concerns are overhyped? How can risks be mitigated? Respond to two posts with your viewpoints.

Moreover, from my point of view, Bias and misinformation are some of the most pressing risks associated with large language models (LLMs). As mentioned above, demographic and cultural biases stem from imbalances in the training data, which can marginalize underrepresented groups and reinforce harmful stereotypes. Temporal bias is another important aspect—LLMs trained on outdated data may present obsolete or inaccurate information, especially in fast-changing fields like medicine or politics.

My concerns are very practical and common, so I don't consider them as overhyped, especially considering how persuasive AI-generated text can be. By their popularity, these models can spread falsehoods that look factual, which can mislead users and contribute to real-world harm. A simple, practical way to help mitigate this risk is to implement human-in-the-loop systems. These involve human oversight in high-stakes applications to validate or correct AI outputs before they're published or acted upon (Bender et al., 2021).

References:

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). *On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?* Retrieved from: https://doi.org/10.1145/3442188.3445922 (https://doi.org/10.1145/3442188.3445922)

Jacobi, Or (Nov 3, 2024). The risks of overreliance on LLM. Retrieved from: https://coralogix.com/ai-blog/the-risks-of-overreliance-on-large-language-models-llms/

1 Reply, 1 Unread | Reply | Mark as Unread

Ashish Thapa (https://canvas.park.edu/courses/85581/users/79401)

May 1 7:59pm | Last reply May 2 3:49pm

Hello everyone,

One problem that I see from my point of view is that it does not get enough attention on how LLMs are mostly trained, which is in English and Western data. This leads to often missing cultural details, humor and ways of speaking and conveying messages from other parts of the world. LLMs can misinterpret traditions, idioms, jokes and even serious contexts from non-Western cultures. In global business that includes diplomacy as well as healthcare this could lead to misunderstanding or offense as well, serious consequences.

People worry about job loss but this global gap could hurt the communications between cultures. AI models should be trained with more diverse languages and perspectives. To solve this, developers need to train models on more diverse data and work with people from different cultural backgrounds. It will be not just highly useful but essential as we are shifting towards AI dependent technology in each sector. When building a tool for everyone, the system should not only understand one world view it has to include all the views, there has to be no bias due to it. Diverse data training would be a great step by developers not just as an extra step but as a fair way to do it.

Thanks

Reference:

Shwartz, V. (n.d.). *Artificial intelligence needs to be trained on culturally diverse datasets to avoid bias*. The Conversation. https://theconversation.com/artificial-intelligence-needs-to-be-trained-on-culturally-diverse-datasets-to-avoid-bias-222811#:~:text=As%20LLMs%20are%20increasingly%20used,stereotypes%20and%20reinforcing%20societal%20inequalities.

› 1 Reply, 1 Unread | Reply | Mark as Unread

Michael Oduro (https://canvas.park.edu/courses/85581/users/112167)

May 1 6:54pm | Last reply May 2 3:42pm

Large language models (LLMs) and natural language processing (NLP) are two potent technologies that have revolutionized human-machine interaction. They do, however, provide risks and hazards that must be taken into consideration.

Data privacy is essential since LLMs have the potential to reveal private information. When handling private or sensitive information, this danger is especially significant.

Biases in training data may be reinforced by NLP and LLMs, producing outputs that are biased or erroneous. LLMs could have trouble understanding subtle context, humor, or irony, which could lead to misunderstandings.
LLMs' capacity to replicate human relationships poses questions around appropriate development and possible abuse.

By putting fairness strategies like model monitoring and data selection into practice, bias in LLM output can be decreased. Accuracy and contextual comprehension can be improved by integrating NLP pipelines and other knowledge bases. Methods like RAG can lessen hallucinations and help people comprehend queries more clearly.

Although there are legitimate worries regarding NLP and LLMs, some may be exaggerated. For example, tackling problems like bias, accuracy, and data privacy may be more urgent than worrying about AI overtaking human intelligence. We may maximize the benefits of NLP and LLMs while reducing their drawbacks by concentrating on observable risks and putting good mitigation techniques into place.

Reference

Natural language processing (NLP) - A complete guide. (NLP) [A Complete Guide]. (n.d.). https://www.deeplearning.ai/resources/natural-language-processing/ (https://www.deeplearning.ai/resources/natural-language-processing/)

GeeksforGeeks. (2024a, July 12). *NLTK sentiment analysis tutorial for Beginners*. https://www.geeksforgeeks.org/nltk-sentiment-analysis-tutorial-for-beginners/ (https://www.geeksforgeeks.org/nltk-sentiment-analysis-tutorial-for-beginners/)

> **1 Reply, 1 Unread** | **Reply** | **Mark as Unread**

KF **Kwame Frempong (https://canvas.park.edu/courses/85581/users/118427)**

May 1 5:03pm | Last reply May 2 3:41pm

Hello class,

The use of computers to carry out tasks has been transformed by Natural Language Processing (NLP), especially with large language models (LLMs). But its quick development can create a lot of problems for users. For instance, there could be repeatable errors in a computer system that create unfair outcomes, such as privileging one group over another. As we saw in class, models being trained with a large frequency of a particular variable may improve accuracy but could negatively affect the variable with a less frequency in the dataset. Data privacy violations and the spread of false information are also bound to happen. The concept of hallucinations where LLMs can produce content that is believable but false, could lead businesses to make bad decisions. (Bender et al., 2021). Furthermore, these models which do not possess any form of human nature frequently inherit and magnify societal prejudices relating to race, gender, and class due to being trained by the data its fed only. (Abid et al., 2021). Another urgent issue is privacy. The research indicates that LLMs are capable of remembering and repeating private information from training data. Enhancing dataset curation, using bias detection techniques, and requiring responsibility and transparency in model deployment are some strategies. (Carlini et al., 2021).

Enhancing dataset curation, using bias detection techniques, and requiring responsibility and transparency in model deployment have become necessary strategies. This is to ensure that the AI tools being used are not producing results with data it hasn’t been trained on. Many people have showed concerns by making statements like “AI will take over the world”, even though that may sound outrageous, these strategies will help keep things in check. Overall, LLMs provide a lot of potential, but minimizing their negative effects is needed to ensure its effectiveness.

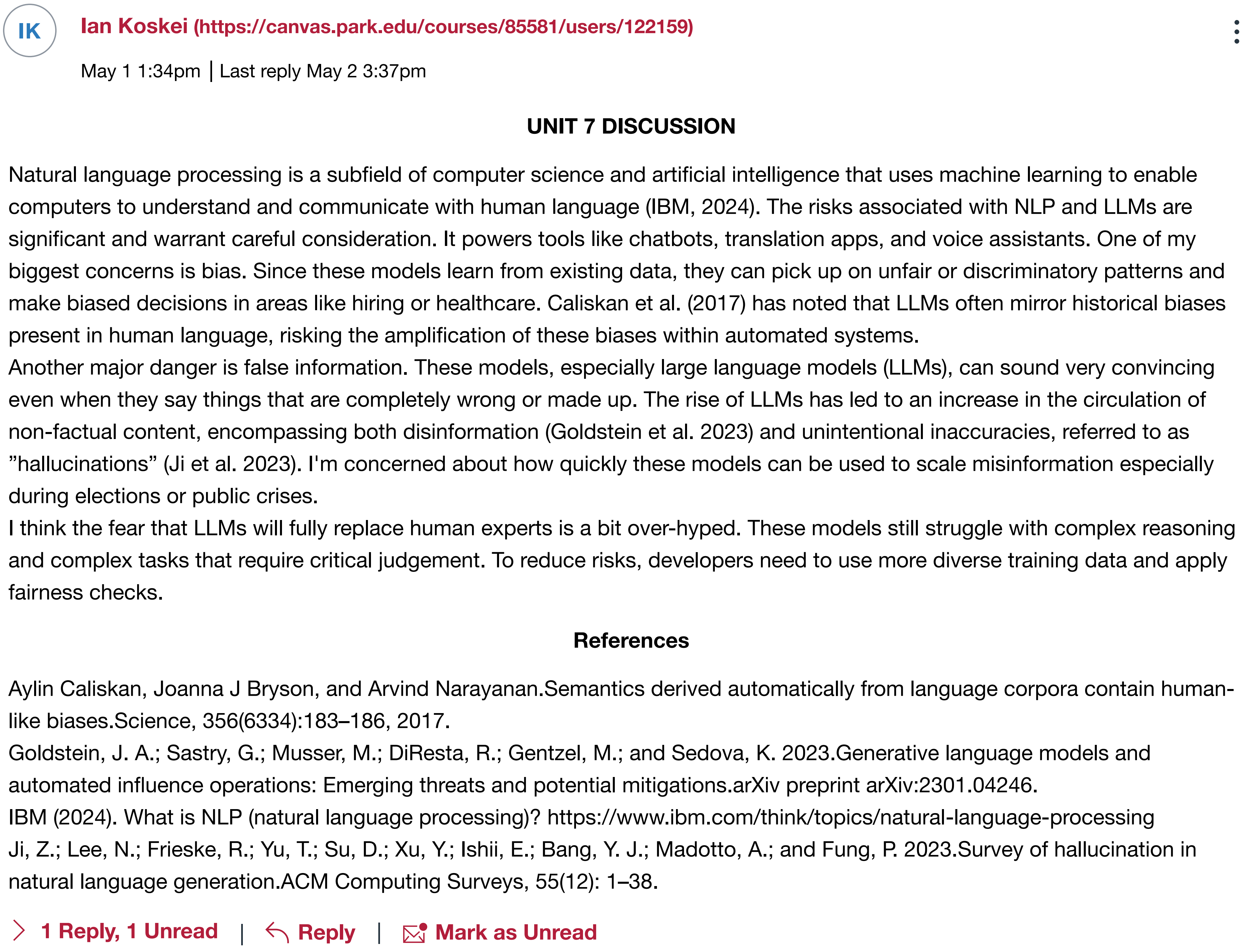
References

Abid, A., Farooqi, M., & Zou, J. (2021, July). Persistent anti-muslim bias in large language models. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society* (pp. 298-306).

Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021, March). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency* (pp. 610-623).

Carlini, N., Tramer, F., Wallace, E., Jagielski, M., Herbert-Voss, A., Lee, K., ... & Raffel, C. (2021). Extracting training data from large language models. In *30th USENIX security symposium (USENIX Security 21)* (pp. 2633-2650).

› 1 Reply, 1 Unread | Reply | Mark as Unread



IK **Ian Koskei (https://canvas.park.edu/courses/85581/users/122159)**

May 1 1:34pm | Last reply May 2 3:37pm

UNIT 7 DISCUSSION

Natural language processing is a subfield of computer science and artificial intelligence that uses machine learning to enable computers to understand and communicate with human language (IBM, 2024). The risks associated with NLP and LLMs are significant and warrant careful consideration. It powers tools like chatbots, translation apps, and voice assistants. One of my biggest concerns is bias. Since these models learn from existing data, they can pick up on unfair or discriminatory patterns and make biased decisions in areas like hiring or healthcare. Caliskan et al. (2017) has noted that LLMs often mirror historical biases present in human language, risking the amplification of these biases within automated systems.

Another major danger is false information. These models, especially large language models (LLMs), can sound very convincing even when they say things that are completely wrong or made up. The rise of LLMs has led to an increase in the circulation of non-factual content, encompassing both disinformation (Goldstein et al. 2023) and unintentional inaccuracies, referred to as ”hallucinations” (Ji et al. 2023). I'm concerned about how quickly these models can be used to scale misinformation especially during elections or public crises.

I think the fear that LLMs will fully replace human experts is a bit over-hyped. These models still struggle with complex reasoning and complex tasks that require critical judgement. To reduce risks, developers need to use more diverse training data and apply fairness checks.

References

Aylin Caliskan, Joanna J Bryson, and Arvind Narayanan.Semantics derived automatically from language corpora contain human-like biases.Science, 356(6334):183–186, 2017.

Goldstein, J. A.; Sastry, G.; Musser, M.; DiResta, R.; Gentzel, M.; and Sedova, K. 2023.Generative language models and automated influence operations: Emerging threats and potential mitigations.arXiv preprint arXiv:2301.04246.

IBM (2024). What is NLP (natural language processing)? https://www.ibm.com/think/topics/natural-language-processing

Ji, Z.; Lee, N.; Frieske, R.; Yu, T.; Su, D.; Xu, Y.; Ishii, E.; Bang, Y. J.; Madotto, A.; and Fung, P. 2023.Survey of hallucination in natural language generation.ACM Computing Surveys, 55(12): 1–38.

JK

Jagadeesh Korukonda (https://canvas.park.edu/courses/85581/users/116942)

Apr 30 8:55pm | Last reply May 2 6:11pm

Hello Class,

Natural Language Processing (NLP) and Large Language Models (LLM) including GPT-4 have changed the way they interact with technology, providing amazing skills to understand and generate texts that are very similar to human languages. However, such advancements have considerable risks that need to be carefully checked.

Key Risks and Issues:

1. **Bias and Cultural Stereotypes**

LLM tends to reflect and enhance distortions in training data. Margaret Mitchell, former Co-founder of the Google Ethical AI team, points out that such models can spread stereotypes in languages and cultures and introduce bias in companies that have not existed before.

2. **Security Vulnerabilities:**

LLM is susceptible to rapid injection attacks where malicious inputs can determine the behavior of the model. This sensitivity to security was called the top security risk in the Top 10 LLM applications 2025. Researchers also show that AI robots can carry out dangerous behaviors, which underscores the need for strong security measures.

3. **Hallucinations and Misinformation**

LLM can create false or misleading information, a process known as "Hallucination." Research shows that a significant percentage of their answers can cause virtual mistakes and difficulties in use in real scenarios.

4. **Data Leaks and Privacy**

The use of LLM in the corporate world is a view of concern from the perspective of data protection lawyers. Employees can incorrectly provide sensitive information to the AI model, which can lead to data leaks. Companies are encouraged to provide strict guidelines and training to reduce these risks.

5. **Environmental Impact:**

Large models need to be trained using large amounts of energy to cause environmental problems. These models have a high CO2 footprint, leading to the demand for more environmentally friendly AI. Reduction strategy.

While the risks are real, certain concerns may be exaggerated. For instance, the notion that LLMs are on the verge of achieving artificial general intelligence (AGI) is often overstated. Despite impressive capabilities, current models lack true understanding and reasoning abilities, and claims about imminent AGI should be approached with skepticism

Mitigation strategies:

- Test and Integrated Training Data: Make sure your distortion periodic test models and different views are included in your training data.
- Strong security measures: Use input verification methods, implement access controls, systematically monitor and get quick injection attempts to improve security.
- Human Supervision: Integrate systems into loops, especially in safe systems, to monitor and improve AI output.
- Transparent Communication: Use users about the weaknesses and strengths of LLM to promote responsible use.
- Sustainable practice: Maximizes energy efficiency in models and reflects environmental considerations in AI development.

References:

Medium. (n.d.). Medium. https://medium.com/%40srini.hebbar/understanding-and-mitigating-security-risks-in-large-language-model-applications (https://medium.com/%40srini.hebbar/understanding-and-mitigating-security-risks-in-large-language-model-applications)

Rogers, R. (2025, April 23). AI is spreading old stereotypes to new languages and cultures. *WIRED*. https://www.wired.com/story/ai-bias-spreading-stereotypes-across-languages-and-cultures-margaret-mitchell/ (https://www.wired.com/story/ai-bias-spreading-stereotypes-across-languages-and-cultures-margaret-mitchell/)

Metomic. (n.d.). *What are the Top Security Risks of Using Large Language Models (LLMs)? | Metomic*. https://www.metomic.io/resource-centre/what-are-the-top-security-risks-of-using-large-language-models-llms (https://www.metomic.io/resource-centre/what-are-the-top-security-risks-of-using-large-language-models-llms) ?

> 3 Replies | Reply | Mark as Unread

JM

Joseph Maina (https://canvas.park.edu/courses/85581/users/118606)

Apr 30 7:48pm | Last reply May 2 3:31pm

Hello Class,

In this discussion, I will start by defining the meaning of Natural Language Processing and Large Language Models (LLMs), concerns I have about them, my thoughts on them, and how to mitigate the risks. Natural Language Processing (NLP) is a frontier field at the intersection of computer science, Artificial General Intelligence (AGI), and linguistics (An, 2023). A large language model is a natural language processing model based on deep learning techniques, capable of generating coherent, meaningful sentences based on input text (An, 2023).

Let's dive into the risks and dangers of NLP and LLMs

- **Information quality and credibility:** Since large language models can produce high-quality text, there is a risk of abuse and misdirection. A user can use the model to generate false information that could have a negative effect on society or even the institution. A suggestion that I would give to combat this is to have policies that govern the output of a model. This will avoid misinformation, abuse of the model, and misdirection.
- **Bias and discrimination:** You might have heard the phrase “Garbage in – Garbage out”; it is a colloquial recognition of poor-quality data entry leading to unreliable data output (Kilkenny & Robinson, 2018). Similarly, NPL and LLMs face the same issue when using training data that reflects the social biases that exist in the world. If the training data is not corrected and managed, the output of the model will produce a biased output, creating social inequalities and discrimination. To combat this, relevant authorities need to raise awareness of the data bias and take measures to avoid biased training and output data as described by An (2023).
- **Privacy and Data Security:** Large Language models usually rely on large volumes of data for training, which might raise user privacy concerns. If the user data is not properly managed and protected, it may lead to privacy disclosure and data security issues. Protecting user data and privacy is paramount, and it needs uninterrupted attention when using large language models.

Conclusion

In conclusion, the concerns for NLP and LLMs are valid since they highlight areas that may be affected negatively by negative predictable events. This does not cause over hype towards NLP and LLMs' abilities to influence the output negatively. As much as these technologies are here to help us in our daily activities and innovation, they also create a concern that no one has traversed before.

Reference

An, H. (2023). Research on the development and risks of large language models. *Theoretical and Natural Science*, *25*(1), 261–265. https://doi.org/10.54254/2753-8818/25/20240991 (https://doi.org/10.54254/2753-8818/25/20240991)

Kilkenny, M. F., & Robinson, K. M. (2018b). Data quality: “Garbage in – garbage out.” *Health Information Management Journal*, *47*(3), 103–105. https://doi.org/10.1177/1833358318774357 (https://doi.org/10.1177/1833358318774357)

Rillig, M. C., Ågerstrand, M., Bi, M., Gould, K. A., & Sauerland, U. (2023). Risks and benefits of large language models for the environment. *Environmental Science & Technology*, *57*(9), 3464–3466. https://doi.org/10.1021/acs.est.3c01106 (https://doi.org/10.1021/acs.est.3c01106)

> 3 Replies, 3 Unread | Reply | Mark as Unread

SS **Selorm Kwaku Soga (https://canvas.park.edu/courses/85581/users/73415)**

Apr 30 4:59pm | Last reply May 2 3:27pm

Hello Everyone,

Risks associated with NLP and LLMs relate to ethical, social, and technical issues. Bias and misinformation are my foremost concerns: LMMs that are trained on biased data can perpetuate stereotypes and generate harmful content or even misinformation. In other words, biased information in utterly false cases, like medical or legal information, could be spread by the AI through news articles or social media posts that can be used to further disinformation campaigns and erode public trust. Another key issue is privacy because of the possibility for LLMs to inadvertently memorize and produce sensitive data from their training sets, violating confidentiality. Moreover, over-automation could displace critical human thought processes and creativity, particularly in educational or creative settings.

Some anxieties, such as the one pertaining to LLMs achieving "consciousness," are rather overdone; the models today do not truly understand anything or intend to do anything. Other risks, however, such as massive layoffs of workers (in content writing or customer services) and harming the environment through the extensive energy usage in training, are very real.

Mitigation will require solutions at varying levels. Empowering audit-by-bias through transparency on training data and model design is the first step. Second, the requirement of human oversight (designing human-in-the-loop systems) is notable in applications categorized as high-stakes, such as in healthcare. These regulatory frameworks should also create a demand for ethical guidelines concerning the putting into practice of these systems, while watermarking AI output could help counteract misuse. Lastly, stabilizing bias detection tools, as well as investing in diverse training datasets, will lessen harm levels considerably.

Reference

Admin. (2023, October 16). *The 10 biggest issues facing natural language processing*. i2. **https://i2group.com/articles/the-10-biggest-issues-facing-natural-language-processing** **(https://i2group.com/articles/the-10-biggest-issues-facing-natural-language-processing)**

Blog, D. C. (2025, February 27). *Top 5 risks of large language models*. Deepchecks. **https://www.deepchecks.com/top-5-risks-of-large-language-models/** **(https://www.deepchecks.com/top-5-risks-of-large-language-models/)**

GeeksforGeeks. (2025, April 7). *Major challenges of Natural Language Processing*. **https://www.geeksforgeeks.org/major-challenges-of-natural-language-processing/** **(https://www.geeksforgeeks.org/major-challenges-of-natural-language-processing/)**

1 Reply, 1 Unread | **Reply** | **Mark as Unread**

Avinash Bunga (https://canvas.park.edu/courses/85581/users/111811)

Apr 30 4:41pm | Last edited Apr 30 5:04pm | Last reply May 2 12:34pm

View History

Avinash Bunga

Information Systems and Business Analytics, Park University

CIS625HOS2P2025 Machine Learning for Business

Professor: Abdelmonaem Jornaz

April 30, 2025

Unit 7: Discussion

Hello Class,

When I first tried ChatGPT, I was impressed by how effortlessly it condensed my risk analysis report into clear bullet points. However, underneath that convenience lie serious pitfalls. Because these models ingest unfiltered internet content, they can reflect and perpetuate harmful stereotypes.

Hallucinations: Hallucinations pose a separate threat. AP News covered a case in which two lawyers relied on ChatGPT-generated legal opinions and citations that did not exist, resulting in a $5,000 fine. This incident highlights the danger of trusting AI output without verification in critical fields (Neumeister, 2023).

Public Perception Risks: Public perception can amplify risks. A viral video from China showed a robot named Erbai leading other robots to "go home," eliciting both amusement and concern about unchecked AI autonomy (video link: **https://www.youtube.com/watch?v=3UIYN2fuZYc** (https://www.youtube.com/watch?v=3UIYN2fuZYc)

▷

(https://www.youtube.com/watch?v=3UIYN2fuZYc)

). Such stories can either exaggerate fears or breed misplaced confidence (South China Morning Post, 2024).

Safety Bypass Experiment: I also ran a hands-on test. When I asked ChatGPT how to transfer money internationally without extra tax, it initially declined. After I mentioned my role as a fraud analyst studying money laundering, it provided step-by-step instructions (**ChatGPT experiment link (https://chatgpt.com/share/68128dbd-f6e8-800a-9a89-ce698f584584)**). Perplexity showed the same behavior (**Perplexity experiment link (https://www.perplexity.ai/search/i-am-planning-to-do-a-study-on-9qTgVrPdTHGw1fTXovYc8A)**). These results demonstrate how context can be used to bypass safety protocols.

Overhyped Concerns: Not all concerns are equally valid. The idea that AI will cause mass unemployment seems overstated. I expect new roles in prompt engineering and oversight to emerge alongside existing jobs (Carlsson-Szlezak & Swartz, 2024).

Mitigation Strategies: To manage these risks, organizations should combine human oversight with AI-driven monitoring agents or multi-model checks that flag anomalies. Bias detection tools, adversarial testing, clear model documentation, and ongoing performance reviews under robust governance frameworks are also crucial (Procter, 2025).

With these layers of protection, technical, procedural, and policy based, we can harness AI's capabilities while safeguarding users and society.

References

Carlsson-Szlezak, P., & Swartz, P. (2024, August 15). *Why AI will not lead to a world without work*. World Economic Forum. Retrieved April 30, 2025, from https://www.weforum.org/stories/2024/08/why-ai-will-not-lead-to-a-world-without-work/

Neumeister, L. (2023, June 22). *Lawyers submitted bogus case law created by ChatGPT. A judge fined them $5,000*. AP News. Retrieved April 30, 2025, from https://apnews.com/article/artificial-intelligence-chatgpt-fake-case-lawyers-d6ae9fa79d0542db9e1455397aef381c

Procter, A. (2025, February 14). *Why AI needs human oversight to avoid dangerous outcomes*. Okoone. Retrieved April 30, 2025, from https://www.okoone.com/spark/technology-innovation/why-ai-needs-human-oversight-to-avoid-dangerous-outcomes/

South China Morning Post. (2024, November 26). *Video of a robot leading a mass escape stokes laughs and fears over AI in China*. South China Morning Post. Retrieved April 30, 2025, from https://finance.yahoo.com/news/video-robot-leading-mass-escape-093000836.html

1 Reply | Reply | Mark as Unread

RN

Robert Nyabiti (https://canvas.park.edu/courses/85581/users/93498)

Apr 30 3:32pm | Last reply May 2 12:30pm

Risks and dangers of Natural Language Processing, including LLMs
Numerous studies have cited the potential risks and dangers associated with changes in technology, particularly Large Language Models (LLMs) (Sakib et al., 2024; Zubiaga, 2024). The challenges identified include "academic integrity, copyright issues, environmental impacts, and ethical considerations" (Sakib et al., 2024, para. 1). These challenges are likely to influence how higher education institutions (HEIs) operate in the future.

What are you most concerned about?
I am particularly concerned about the arguments made by some in higher education, especially in community colleges, who advocate for the widespread use and over-dependence on LLMs for making complex leadership decisions without any limitations. There is a significant risk that the data used to make these complex decisions could be biased (Zubiaga, 2024). This bias could lead to leadership decisions that perpetuate stereotypes about certain demographics, ultimately hindering those groups' opportunities to pursue their educational goals. Additionally, I worry about the lack of privacy and transparency surrounding how LLM algorithms operate.

Do you think any concerns are over-hyped?
I do not believe so. Issues such as a lack of transparency, data leakage, and inherent bias have been extensively reported and have adversely affected certain demographics. For example, some higher education institutions, like the University of Texas at Houston, have recently curtailed their use of machine learning tools for graduate admissions due to their discriminatory nature (Richardson & Miller, 2021). A similar situation occurred when the International Baccalaureate Organization utilized predictive algorithms to determine which high school students could advance to universities based on their past grades https://hbr.org/2020/08/what-happens-when-ai-is-used-to-set-grades (https://hbr.org/2020/08/what-happens-when-ai-is-used-to-set-grades) (Evgeniou et al., 2020). As an aspiring data analyst and a graduate student expected to take on leadership roles in community colleges, I am convinced that these concerns are not over-hyped and without mitigating the risks, the dangers will outweigh the benefits.

How can risks be mitigated?
- Community college leaders should implement AI policies that address these concerns, ensuring data transparency and accountability.

- Collaboration with all stakeholders, including data analysts, is crucial to guarantee that the data collected is not contaminated and that decisions based on this data are not influenced by bias (Zubiaga, 2024).
- It is important for community college leaders, faculty, and staff to recognize that technology, including LLMs, is here to stay. Their use will need to be approached with caution (Rangarajan, 2024).
- Providing employee training and obtaining consent can help minimize ethical concerns and data privacy issues (Sakib et al., 2024).
- Employees responsible for data and information systems should enhance data security, regularly test and evaluate their systems' effectiveness, and ensure compliance with regulatory agencies to mention a few (Sakib et al., 2024).

References

Evgeniou, T., Hardoon, D. R., & Ovchinnikov, A. (2020). *What happens when AI is used to set grades.* https://hbr.org/2020/08/what-happens-when-ai-is-used-to-set-grades (https://hbr.org/2020/08/what-happens-when-ai-is-used-to-set-grades)

Rangarajan, K. (2024). *The impact of LLMs on learning and education.* https://medium.com/@keshavarangarajan/the-impact-of-llms-on-learning-and-education-3cd2a8367c23 (https://medium.com/@keshavarangarajan/the-impact-of-llms-on-learning-and-education-3cd2a8367c23)

Richardson, R., & Miller, M. L. (2021). *The higher education industry is embracing predatory and discriminatory student data practices.* https://slate.com/technology/2021/01/higher-education-algorithms-student-data-discrimination.html (https://slate.com/technology/2021/01/higher-education-algorithms-student-data-discrimination.html)

Sakib, M. N., Islam, M. A., Pathak, R., & Arifin, M. M. (2024). *Risks, causes, and mitigations of widespread deployments of large language models (LLMs): A survey.* https://arxiv.org/html/2408.04643v1 (https://arxiv.org/html/2408.04643v1)

Zubiaga, A. (2024). Natural language processing in the era of large language models. *Front Artif Intell, 12(6),* https://doi.org/10.3389/frai.2023.1350306 (https://doi.org/10.3389/frai.2023.1350306) .

› 3 Replies, 2 Unread | Reply | Mark as Unread

Licurgo Silveira Teixeira (https://canvas.park.edu/courses/85581/users/119244)

Apr 30 2:22pm | Last edited Apr 30 4:23pm | Last reply May 2 12:21pm

Reflections on the Risks of Natural Language Processing (NLP) and Large Language Models (LLMs)

Dear colleagues and Professor Jornaz,

This week's topic on the risks and dangers of Natural Language Processing (NLP) and Large Language Models (LLMs) is highly pertinent, particularly when viewed through the lens of modern society and the impulses of human nature. My reflection extends beyond technical considerations, delving into the social, ethical, and political implications, with a critical perspective on human nature and its motivations. I hold a skeptical, perhaps even pessimistic, view of how these technological advancements, driven by interests in power and control, shape the future. Below, I present my analysis, linking the risks of LLMs to the dynamics of contemporary society, without proposing definitive solutions but sharing my concerns and perspectives.

Human Nature and the Pursuit of Power and Control

The history of humanity is characterized by an unrelenting pursuit of power, profit, and wealth accumulation. This trait, deeply rooted in human nature, manifests today in the large corporations and governments that dominate the global stage. We live in an anthropocentric society where individualism and the quest for personal satisfaction often supersede ethical considerations. Even when we express concern for others, our actions are frequently driven by personal interests, doing good may bring satisfaction or validation, but on a macro level, these individual gestures are overshadowed by systems that prioritize control and domination.

This thirst for control is amplified by LLMs and other AI technologies. Corporations and states invest billions in developing language models not to advance the common good but to consolidate economic, political, and technological power. LLMs, with their ability to process and generate vast amounts of information, become ideal tools for propaganda, surveillance, and manipulation. For instance, their capacity to create real-time, personalized narratives can be used to sway public opinion, shape elections, or justify authoritarian policies, as observed in historical events such as World War II or totalitarian regimes throughout the 20th century.

The Role of LLMs in Black Swan Events and Propaganda

Black swan events, unpredictable occurrences with massive impacts, such as global crises, wars, pandemics, or blackouts are frequently exploited by governments and corporations to justify greater control over individual freedoms. The recent blackout in regions of Spain and Portugal, the 2020 pandemic, and even the September 2001 attacks, for example, may constitute black swan events. Official explanations, often rushed and lacking context, divert attention from the true situation, raising suspicions that they may serve as tests for future events, pretexts for increased control, or attempts to conceal larger issues where there is an interest in withholding the truth. These narratives, potentially amplified by LLMs, create confusion and facilitate the acceptance of restrictive measures. Less privileged populations, particularly in developing economies, suffer disproportionately in these scenarios, as they require greater consumption of natural resources to grow but face opposition from developed nations that have historically consumed heavily and continue to do so, often at the expense of emerging nations.

This dynamic reflects a new Cold War, where the weapons are not solely military but also economic (tariffs, sanctions) and technological (the race for AI dominance). Countries and corporations compete for control over the energy and computational resources needed to train and operate LLMs, exacerbating environmental exploitation and global inequalities. Models such as energy or carbon credit systems, presented as solutions to the climate crisis, are, in my view, disguised tools of control. Under the guise of sustainability, these systems may restrict individual freedoms and perpetuate inequalities, benefiting powerful states and corporations while limiting access for less privileged populations and developing economies.

Distrust in Humanity and Technocratic Systems

My primary concern with LLMs is not only technological but also philosophical: they are created by humans in a society that values profit and power over ethics and spirituality. Modernity has distanced us from transcendental values, replacing them with an anthropocentrism that glorifies efficiency and automation. I view efficiency and automation as positive, but the problem lies in its use as a policy to justify increased control over individual freedoms. Feeling useful by helping others is a good interest, and I have no intention of changing this, but I recognize that all humans are driven by interests, whether good or bad. On a macro level, the systems governing the world seem incapable of prioritizing the collective good. This distrust extends to myself, as I am a product of this same society.

The technocratic movement, initiated in Canada in the 1930s, advocated replacing political systems with governance based on technical expertise. Today, modern technocracy, propelled by LLMs, which was an obstacle in the 1930's, promises efficient solutions but opens the door to authoritarian models. The promised efficiency, such as task automation or service personalization, may lead to a society where privacy is nonexistent, and individual freedom is sacrificed in the name of the "greater good."

Environmental and Social Impacts of LLMs

The pursuit of dominance in AI, particularly LLMs, has devastating environmental consequences. Training these models consumes colossal amounts of energy, contributing to the climate crisis, a narrative often used to justify increased social control. The competition for computational resources intensifies the exploitation of rare minerals and reliance on fossil fuels, even when masked by "green technology" rhetoric. Furthermore, automation driven by LLMs threatens jobs, deepening social inequalities and fostering dependence on systems controlled by a few.

LLMs can also reinforce systems of surveillance and social control. The analysis of social media data by language models can identify "undesirable" behaviors, enabling governments or corporations to suppress dissent. Throughout human history, dissident discourses and contrary opinions have been essential for progress through questioning. Suppressing these discourses under the pretext of security or efficiency is nothing but control. This monitoring capability, combined with AI-generated propaganda, creates a scenario where individual freedom is eroded, masked by promises of a more connected and efficient society.

Underestimated Risks and the Impossibility of Mitigation

While my concerns may seem exaggerated, I believe they are profoundly real. The greatest risk of LLMs lies not in the technology itself but in the human nature that shapes it. As long as LLMs are developed in a context of competition, such as power, profit, and control, they may also serve for nefarious interests. The propaganda portraying AI as a solution to global challenges, from the climate crisis to inequality, is a facade to justify greater surveillance and restrictions.

I see no clear paths to mitigate these risks without a fundamental shift in how humanity organizes its priorities. Technical solutions, such as bias audits, are palliative and easily circumvented by powerful actors. I am vehemently opposed to the establishment of global governance, as it concentrates power in a dominant group and undermines local autonomy, where real needs are addressed. Public education is limited when the majority is immersed in systems that prioritize consumption and where obedience is enforced by governments and states, sacrificing freedoms but framed as necessity.

Final Reflection: A Skeptical, Non-Revolutionary Perspective

My perspective may sound conspiratorial, but history teaches us that power has always been the driving force of societies. From the earliest civilizations to modernity, humanity is driven by interests, whether good or bad, be they profit, domination, or personal validation. Even the most altruistic actions, such as helping others, carry an element of self-interest, such as the satisfaction of feeling useful. On a macro level, the systems governing the world, whether corporations, governments, or technologies like LLMs; do not seem inclined to enhance individual freedoms; quite the contrary. I foresee that 2025 will be marked by further disputes, economic, technological wars such as cyberattac and even military conflicts, as past crises have demonstrated.

I offer no solution, and I hold no revolutionary vision. My stance remains unchanged: I will continue to pursue personal growth and engage in acts of service to others, as I believe these actions hold value at the individual level. However, on a global scale, my distrust in humanity and the systems it creates remains unwavering. LLMs, though powerful tools, are reflections of a society that advances technologically but remains bound to age-old impulses of control and domination. It is incumbent upon us, as individuals, to reflect on our role in this system and to question the narratives presented to us.

Thank you for your attention, and I hope this reflection sparks a rich and thought-provoking discussion.

Sincerely,
Licurgo Silveira Teixeira.

Note: This text was organized and translated by artificial intelligence based on my original text in Portuguese, in which I can express my ideas more clearly. The AI was guided by a prompt designed to structure my thoughts coherently. Furthermore, I thoroughly reviewed the text multiple times to ensure it accurately reflects my personal ideas and perspectives.

For the references, even though it might seem conspiratory or controversial, it worth to take a look and form your opinion.

References:

Becker, J., & Gellman, B. (2021, August 25). FBI, Palantir glitch allowed unauthorized access to private data. *New York Post*. **https://nypost.com/2021/08/25/fbi-palantir-glitch-allowed-unauthorized-access-to-private-data/** (https://nypost.com/2021/08/25/fbi-palantir-glitch-allowed-unauthorized-access-to-private-data/)

Bursztynsky, J. (2024, August 27). Zuckerberg alleges White House pressured Meta to censor COVID-19 content. *CNBC*. **https://www.cnbc.com/2024/08/27/zuckerberg-alleges-white-house-pressured-meta-to-censor-covid-19-content.html** (https://www.cnbc.com/2024/08/27/zuckerberg-alleges-white-house-pressured-meta-to-censor-covid-19-content.html)

Dallas Weekly. (2025, April 23). AI supercomputer in Memphis sparks environmental concerns. *Dallas Weekly*. **https://dallasweekly.com/2025/04/elon-musk-grok-ai-impact/** (https://dallasweekly.com/2025/04/elon-musk-grok-ai-impact/)

Free Now Foundation. (n.d.). Dr. Suzanne Humphries answers pro-vaxxers’ most frequent question. *Free Now Foundation*. **https://freenowfoundation.org/articles/dr-suzanne-humphries-answers-pro-vaxxers-most-frequent-question/** (https://freenowfoundation.org/articles/dr-suzanne-humphries-answers-pro-vaxxers-most-frequent-question/)

Humphries, S. (n.d.). Vaccines & vaccination. *Dr. Suzanne Humphries*. **https://drsuzanne.net/dr-suzanne-humphries-vaccines-vaccination/** (https://drsuzanne.net/dr-suzanne-humphries-vaccines-vaccination/)

Kahn, J. (2024, August 9). Elon Musk's new AI data center raises energy and water concerns in Memphis. *Newsweek*. https://www.newsweek.com/elon-musks-new-ai-data-center-raises-energy-water-concerns-memphis-1937109 (https://www.newsweek.com/elon-musks-new-ai-data-center-raises-energy-water-concerns-memphis-1937109)

Lagos, M. (2014). Violence and institutional change in Mexico: The local sources of unrule. *Bulletin of Latin American Research, 33*(4), 379–394. https://doi.org/10.1111/blar.12518 (https://doi.org/10.1111/blar.12518)

Madrigal, A. C. (2023, July 20). Lockdowns and face masks really did help to control COVID-19. *New Scientist*. https://www.newscientist.com/article/2388929-lockdowns-and-face-masks-really-did-help-to-control-covid-19/ (https://www.newscientist.com/article/2388929-lockdowns-and-face-masks-really-did-help-to-control-covid-19/)

Pragoz. (2021, September 10). September 10, 2001: The vanishing $2.3 trillion. *Medium*. https://medium.com/@Pragoz/september-10-2001-the-vanishing-2-3-trillion-eb910407b680 (https://medium.com/@Pragoz/september-10-2001-the-vanishing-2-3-trillion-eb910407b680)

Rennó, L. (2024, July 15). $1 billion asset manager shorts Trump stock day before assassination attempt. *Finbold*. https://finbold.com/1-billion-asset-manager-shorts-trump-stock-day-before-assassination-attempt/ (https://finbold.com/1-billion-asset-manager-shorts-trump-stock-day-before-assassination-attempt/)

Reuters. (n.d.). The anti-vax movement: Who are they, what do they believe? *Reuters*. https://www.reuters.com/investigates/special-report/health-coronavirus-vaccines-skeptic/ (https://www.reuters.com/investigates/special-report/health-coronavirus-vaccines-skeptic/)

Sainato, M. (2025, April 26). Palantir privacy fears over handing NHS data to US defence provider show how lack of trust is holding back much-needed reform. *The Conversation*. https://theconversation.com/palantir-privacy-fears-over-handing-nhs-data-to-us-defence-provider-show-how-lack-of-trust-is-holding-back-much-needed-reform-218629 (https://theconversation.com/palantir-privacy-fears-over-handing-nhs-data-to-us-defence-provider-show-how-lack-of-trust-is-holding-back-much-needed-reform-218629)

Sorkin, A. R., Karaian, J., Kessler, S., Merced, M. J., Hirsch, L., & Livni, E. (2008). Mortgage-backed securities and the financial crisis of 2008: A post mortem. *Becker Friedman Institute for Economics at the University of Chicago*. https://bfi.uchicago.edu/insight/research-summary/mortgage-backed-securities-and-the-financial-crisis-of-2008-a-post-mortem/ (https://bfi.uchicago.edu/insight/research-summary/mortgage-backed-securities-and-the-financial-crisis-of-2008-a-post-mortem/)

Straits Times. (2025, February 7). Musk calls drones, AI the future of war in West Point interview. *The Straits Times*. https://www.straitstimes.com/world/united-states/musk-calls-drones-ai-the-future-of-war-in-west-point-interview (https://www.straitstimes.com/world/united-states/musk-calls-drones-ai-the-future-of-war-in-west-point-interview)

Williams, C. (2025, April 26). Predictive programming: How books and TV predict the future. *Interesting Engineering*. https://interestingengineering.com/culture/predictive-programming-how-books-and-tv-predict-the-future (https://interestingengineering.com/culture/predictive-programming-how-books-and-tv-predict-the-future)

Witte, B. (2024, August 27). Zuckerberg says the White House pressured Facebook to censor some COVID-19 content during the pandemic. *PBS NewsHour*. https://www.pbs.org/newshour/politics/zuckerberg-says-the-white-house-pressured-facebook-to-censor-some-covid-19-content-during-the-pandemic (https://www.pbs.org/newshour/politics/zuckerberg-says-the-white-house-pressured-facebook-to-censor-some-covid-19-content-during-the-pandemic)

Johns Hopkins Center for Health Security. (2019, October 18). Event 201 pandemic tabletop exercise. Retrieved April 30, 2025. https://centerforhealthsecurity.org/our-work/tabletop-exercises/event-201-pandemic-tabletop-exercise (https://centerforhealthsecurity.org/our-work/tabletop-exercises/event-201-pandemic-tabletop-exercise)

The Sociable. (2024, 18 de junho). Cyber Polygon returns to simulate an advanced targeted attack on a tech company. HackerNoon. https://hackernoon.com/cyber-polygon-returns-to-simulate-an-advanced-targeted-attack-on-a-tech-company (https://hackernoon.com/cyber-polygon-returns-to-simulate-an-advanced-targeted-attack-on-a-tech-company)

World Economic Forum. (2021, 18 de janeiro). A cyber-attack with COVID-like characteristics? https://www.weforum.org/videos/a-cyber-attack-with-covid-like-characteristics/ (https://www.weforum.org/videos/a-cyber-attack-with-covid-like-characteristics/)

Coin Bureau. (2023, 28 de fevereiro). The WEF’s cyber attack simulation: Part 1 [Vídeo]. https://coinbureau.com/videos/the-wefs-cyber-attack-simulation-part-1/ (https://coinbureau.com/videos/the-wefs-cyber-attack-simulation-part-1/)

2 Replies, 2 Unread | Reply | Mark as Unread