J. Random Hacker

Dr. Emma Carelton

PHIL 6900

9 December 2025

On Bullshit

Writeup

In his fine writing, Frankfurt mentions that bullshit is effectively unavoidable in modern society. Thanks to revelations in AI, machine learning, language synthesis, and model training, this problem is only getting worse. Frankfurt writes: "Bullshit is unavoidable whenever circumstances require someone to talk without knowing what he is talking about. Thus the production of bullshit is stimulated by whenever a person's obligations or opportunities to speak about some topic exceed his knowledge of the facts that are relevant to the topic" (Frankfurt 33). Of course, Frankfurt was originally referring to humans bullshitting, but, this is not the case anymore. In our modern age, we have AIs that are quite keen on making up all manner of information with no regard for its factual accuracy.

My question for discussion was "as we move into the AI age, the current large language models we have currently have a very high BS-propensity, and are very easy to gaslight into producing false information. What does this mean going forward? Why

do we blindly trust AIs to be accurate? Should the producers of AI products be held liable for their bullshit? Why or why not?" While this may be somewhat of a loaded question, I believe that it addresses many of the issues we are facing (or will face in a much stronger capacity) in our modern society.

As someone that has used (and developed) large-language models for a variety of purposes, it is amazing to me how unreliable these truly are. Current LLMs are designed to generate text "at all costs" -- this does not mean they are accurate by any means. Sure, they may certainly appear to be accurate, but, it would be unwise to assume they are experts by any stretch of imagination. The GPT4All project provides a "document consumption" mechanism that allows LLMs under its control to cite sources, but, it can still hallucinate the facts from the material given. This is because, as Frankfurt mentions, AIs are under no obligation whatsoever to actually produce factually accurate information. Furthermore, it is tremendously easy to trick or manipulate them into acting poorly or producing more bad information. As such, it is pivotal that future LLMs and model training techniques take into account information accuracy beyond what it currently is. Unfortunately, this means that a lot of human labor will be involved; we have already seen how OpenAI exploited their workers to train ChatGPT (while I will not cover the situation here, it is mired in its own ethical issues).

As for my own response to this, I think that going forward, people will have to understand that AIs are just one source in a collection of sources (as was bought up in

the class discussion). Regarding on why we trust AIs to be accurate, I believe it is more out of desperation and ignorance, rather than real trust. Very few people truly trust an AI to be accurate, and, this is where I think we should all stand. I do believe that companies that produce AI products should be held liable for any harmful or dangerously misleading information their LLMs produce -- while this may seem harsh, I think it may be necessary to impose such a rule. In the past, even as far back as the 1980s, companies that sold so-called "expert systems software" made it very clear and apparent that it was just "one gear in a mechanism and not a one-answer-solution to any problem it was designed to solve. If you were using expert systems software to design processors, you would have saved a lot of time but assumed some risk in doing so. In 1986, this was done -- the processor of the DEC VAX 9000 mainframe computer was the first large-scale complex device to be designed with the assistance of AI. Tremendous work was poured into tuning the expert system to be as accurate as possible in its design of CPU logic gates, and, each iteration resulted in more accurate design. I think this is the ideal use case for an AI -- acceleration. Granted, the DEC CPU-designing expert system had no natural language processing ability at all, but, this does not exclude it from being useful.

AI statement: I didn't use AI for this, and I haven't for any assignments in this class.

2025 Update

Despite my assertive position in the body of the above essay that AI was a useful tool in the toolchest, I now have come to the conclusion that AI is not useful for much

of anything -- furthermore, the usage of AI is unto itself, in many ways, problematic. For one, the current generations of generative AIs (text generators, image generators, video generators, and sound generators) were trained on entirely on stolen content. This is, of course, rather ethically questionable: I do not believe that there is any good-faith usage of AI based on these grounds. Since the creators of the content were not properly compensated (and we do live in a world where compensation is important) and the generative AIs can (and do quite well) accurately reproduce copyrighted materials without automatically compensating the original author. The current generation of AI company leaders scoff at the idea of not having to face any punishment for their blatant copyright infringement, but this will not last forever.

In conclusion, generative AI is an ethical nightmare and should not be considered for use.