SCHOLARS
STRATEGY NETWORK

The Critical Need for Transparency and Regulation amidst the Rise of Powerful Artificial Intelligence Models

Jim Samuel, Rutgers University-New Brunswick

As artificial intelligence (AI) technologies cross over a vital threshold of competitiveness with human intelligence, it is necessary to properly frame critical questions in the service of shaping policy and governance while sustaining human values and identity. Given AI's vast socioeconomic implications, government actors and technology creators must proactively address the unique and emerging ethical concerns that are inherent to AI's many uses.

Open Source versus Black Box AI Technologies

AI can be viewed as an adaptive "set of technologies that mimic the functions and expressions of human intelligence, specifically cognition, and logic." In the AI field, foundation models (FMs) are more or less what they sound like: large, complex models that have been trained on vast quantities of digital general information that may then be adapted for more specific uses. Two notable features of foundation models include a propensity to gain new and often unexpected capabilities as they increase in scale ("emergence"), and a growing predisposition to serve as a common "intelligence base" for differing specialized functions and AI applications ("homogenization"). Large language models (LLMs) that power applications like ChatGPT are foundation models with a focus on modeling human language, knowledge, and logic. Advanced AIs and foundation models have the potential to replace multiple task-specific or narrow AIs due to their scale and flexibility, which increases the risk of a few powerful persons or entities who control these advanced AIs gaining extraordinary socioeconomic power, creating conditions for mass exploitation and abuse.

ChatGPT and other large language model applications, that are growing in popularity, use a nontransparent "black box" approach; users of these technologies have little to no access to the inner workings or underlying AI models and can only observe outputs (such as an essay) that result from data inputs (such as a written or spoken prompt) to judge how these applications function. Such opaque foundation models have widespread future AI application potential, displaying homogenization where the base models can be adapted to serve a range of specialized purposes. These exponentials are the risks inherent in a system that allows for private control of advanced AIs. Open-source initiatives, on the other hand, prioritize transparency and public availability of the AI models, including code, process, relevant data, and documentation, so that users and society at large have an opportunity to understand how these technologies function.

For human society, the spirit of the open-source movement is one of the most valuable forces at play in the AI technologies arena. Research and development of AI ethics must emphasize the contributions of open data, open-source software, open knowledge, and responsible AI movements, and contrast these with the challenges presented by relatively opaque AI applications. Closely coupled with open-source, the open-data movement (which refers to "data that is made freely available for open consumption, at no direct cost to the public, which can be efficiently located, filtered, downloaded, processed, shared, and reused without any significant restrictions on associated derivatives, use, and reuse") can be a significant contributor to the development of responsible AI. Open-source initiatives distribute power and reduce the likelihood of centralized control and abuse by a few AI owners with concentrated power.

Questioning AI Ethics to Inform Regulatory Processes

Drawing from the open-source movement's prioritization of transparency, decentralization, fairness, liberty, and empowerment to the people, responses to the following five questions should be required of all companies creating opaque AI applications such as ChatGPT.

Is it fair to use an opaque black-box approach for AI technologies when the implications and impacts of complex AI technologies posit many significant risks? The consequences of AIs are great and must be associated with proportionately higher levels of accountability. The volatile impacts of AIs are expected to be exponentially greater over time than past technologies. Therefore, for human society, it would be relatively less risky in the long run for companies to embrace the open-source approach which has already demonstrated the critical value of transparency and open availability of source materials.

If building upon valuable, free, good-faith open-source research, is it then morally correct to build opaque black boxes for private profit? Significant open-source contributions, made in good faith, have laid the foundations for the present-day AIs. Many of the technological modules, such as transformers, used within applications like ChatGPT came from open-source research. Having reaped the benefits of open-source, privatization of critical AI models blocks societal innovation opportunities and even if presently legally permissible, should be considered ethically wrong.

Why should for-profit companies be allowed to deprive people of their right to know all specific details about what data an AI (that they are expected to use, compete against, and perhaps even be subject to in the future) has been developed with? Large language models are trained on vast quantities of data—in the interests of public benefit and transparency, all data "ingredients" used for training must be detailed in a testable manner. If the specific texts on which GPT3,4 /ChatGPT have been trained remain largely undisclosed, it would be difficult for the public or governments to audit for the fair use of data; to gauge if restricted, protected, private or confidential data have been used; and to see if a company has added synthetic data or performed other manipulations causing the AI to present biased views on critical topics. All training data must be declared in real-time.

Why should companies not be held responsible for transparency and be compelled to demonstrate the absence of deliberate bias mechanisms and output-manipulating systems within their applications? Combining complex, risk-inducing, hidden AI technologies with the opacity of data use gives "emergence" to the potential rise and spread of manipulative AIs. Companies must be required to show that their AIs provide a fair and unbiased representation of information, and they must be held responsible to prove in real-time that they are indeed reflecting "facts as they are." The absence of protections for the public will potentially facilitate mass manipulation of users—consider a future filled with powerful mind-bending manipulative AIs under the control of a few of the self-declared elites.

Why should for-profit AIs not be regulated by enhanced AI-appropriate laws and policies for consumer protection? ChatGPT Plus was released at $20 per month for privileged access in February of 2023. For a company that started as a "nonprofit to develop AI" with a self-declared purpose to "benefit humanity as a whole," this rush to monetization can appear opportunistic and shortsighted. When an AI is presented as a commercial product, the company profiting from this product should be held liable for full transparency and all the output the AI produces, and companies should not be allowed to actively or passively coerce users (again) into signing away all their fair rights. Governments must break the habit of acting only *after* powerful companies and wealthy investors have ensured positive returns on investment, and companies must be challenged to develop profit models which accommodate full transparency of methods and data.

Fear of abuse or global security concerns are feeble excuses for hiding general-purpose scientific discoveries, building opacity, and creating black boxes. Instead of waiting for harm to occur, governments must be proactive; for example, if large language model applications had been covered by proactive AI regulations before ChatGPT's launch, we would be able to use the technology more confidently and increase productivity with informed concerns on bias, limitations, and manipulation. Monetization and positive return on investment are important and it is necessary to have sustainable business models with some level of opacity of final production systems. However, given the nature and power of AI technologies, utmost care must be taken to provide open availability of AI foundation models and training data and to ensure transparency, fairness, accountability, application of open-source principles, and adoption of responsible AI practices. Supporting the integration of open data and AI, along with proactive policies built on the principles of the open-source movement, will serve as a sustainable value-creation strategy to ensure that the benefits of AI are truly disbursed equitably to all people. Using a framework derived from the open-source movement will ensure an optimal measure of public power over artificial intelligence, and lead to a much-needed improvement in accountability and responsible behavior by companies and governments who "own" these technologies.

This brief was produced under the *Acceptable AI* research initiative at Rutgers University's public informatics program.