Machine Learning Think Tank Warns of Serious Risks With Large Language Models

Researchers find 23 inherent security risks in black box LLM foundation models

BERRYVILLE, VIRGINIA, UNITED STATES, January 24, 2024 /EINPresswire.com/ -- The Berryville Institute of Machine Learning (BIML), a think tank dedicated to AI and machine learning security, today released new research that reveals serious security risks in the architecture of LLMs that put companies at risk and threaten election integrity and democracy.

Led by Dr. Gary McGraw, a 30-year security veteran who established the field of software security and is the author of twelve best-selling books on the subject, BIML has done a comprehensive deep dive into LLMs and discovered over 80 inherent risks, including issues not previously revealed publicly. The research concludes that there are serious consequences to relying on big tech's black box LLM foundation models, which are quickly being adopted as critical corporate business tools by technology executives who have no understanding of how the tech actually operates or what the consequences are to their company.

> We are big fans of AI and Machine Learning, but we are concerned about whether the current generation of LLMs are built with security in mind... The only answer is regulation."
>
> *Dr. Gary McGraw, co-founder of BIML*

To ensure the safe and secure use of the technology, the development of black box LLM foundational models should be regulated to require AI vendors to provide transparency into how they are constructed, the data they are constructed from, and how they work.

BIML's new research report, titled "An Architectural Risk Analysis of Large Language Models:

Applied Machine Learning Security," concludes that the black box LLM foundational models enable AI vendors to hide 23 critical risks that users can't control – and often don't even know about themselves. LLM users who adopt a black box foundation model are blindly trusting each LLM foundation model vendor to manage important risks for them without even knowing what those risks are. Most companies are forced to depend on these black box LLM foundation models because they lack the resources to build their own LLM's from scratch, primarily because vendors are siloing access to the ocean of data required for training LLMs—a kind of data feudalism.

"We are big fans of AI and Machine Learning," says Dr. Gary McGraw, co-founder of BIML, "but we are concerned about whether the current generation of LLMs are built with security in mind. In our view, the two biggest risks in the rampant spread of LLMs are recursive pollution (in which ML generated wrongness grows just like guitar feedback through an amp does) and data debt accumulated when enormous oceans of training data turn out to be full of poison, garbage, nonsense, and noise, much of which is difficult or impossible to scrub out. The only answer is regulation."

BIML is concerned that generative AI poses very real risks to election integrity in 2024, which will be a critical year for democracy all around the world. Generative models and LLMs are deep fake machines, and will be used to pollute the information atmosphere.

BIML believes that LLMs should be regulated, and that regulations should first target black box LLM foundation models, and only then target the downstream use of such models. Regulation can be used to open the black box. The new report lists specific targets for regulations that are technical and clear.

The "Architectural Risk Analysis of Large Language Models: Applied Machine Learning Security" is designed for use by security executives, machine learning engineers, developers and others who are creating applications and services that use LLM technologies. It also provides essential technical data for AI regulators. For more information about the report, visit https://berryvilleiml.com/results/BIML-LLM24.pdf.

About BIML

The Berryville Institute of Machine Learning (BIML) was created in 2019 to address security issues with ML and AI. The organization was founded by Dr. Gary McGraw, father of software security, best-selling author, world-renowned security expert and CTO of Cigital (acquired by Synopsys); Harold Figueroa, previous director of Machine Intelligence Research and Applications (MIRA) Lab at an intelligence community contractor; Richie Bonett, a Site Reliability Engineer at Verisign, and Katie McMahon, former President and COO at Native Voice and VP at Shazam who advises startups and holds patents in speech recognition, natural language understanding and augmented reality. BIML is headquartered in Berryville, Virginia. For more information, visit https://berryvilleiml.com/.

Contact:

press@berryvilleiml.com
Berryville Machine Learning Institute
+1 415-728-2821
email us here