MediaWell

News Item

External researcher access to closed foundation models: State of the field and options for improvement | Mozilla Foundation

By Esme Harrington & Mathias Vermeulen
October 22, 2024

> As foundation models become increasingly embedded in a wide array of downstream products and services, understanding their risks and vulnerabilities is more critical than ever to prevent negative impacts. External scrutiny can play a crucial role not only in forming a comprehensive understanding of these risks and vulnerabilities but also in ensuring that users, regulators, and the general public can trust that a foundation model has been rigorously tested.
>
> This raises questions concerning the minimum conditions for public scrutiny and public-interest research for those who choose to keep their model gated behind APIs or proprietary interfaces, including most dominant firms in the industry. Current policy initiatives in the EU, UK, and US have addressed this question only to a limited extent. The EU's AI Act introduces specific legal obligations for developers of foundation models, including red teaming and risk assessments, but falls short of spelling out minimum conditions. The UK and the US have included proposals around external researcher access into various non-binding policy frameworks, without mandating any form of external access.
>
> […]

Source: External Researcher Access to Closed Foundation Models