OAISIS

Third Opinion Launches To Help AI Professionals Evaluate Concerning Developments

Frontier AI professionals can receive anonymous, secure opinions from independent experts to help them judge if a development they're seeing is a cause for concern

Berlin – OAISIS, an independent non-profit organization, today announced the launch of Third Opinion – a secure, confidential online platform that empowers AI professionals to get expert guidance on a wide range of technical, ethical, and organizational concerns.

AI experts like Yoshua Bengio, Stuart Russell, and multiple (former) employees of frontier AI labs have recently highlighted[1] the importance of ensuring that individuals working at the frontier of artificial intelligence have the freedom and support to raise concerns about AI risks.

However, many of the leading innovators in artificial intelligence face a common challenge - they witness activities or behaviors that raise serious questions, but lack a trusted, confidential way to validate their professional instincts. Constrained by nondisclosure agreements, a culture of secrecy, and uncertain protections for reporting concerns, these AI practitioners often feel isolated and unsure how to fulfill their duties responsibly. Reaching out to friends, approaching external experts directly, or speaking publicly to clarify a concern can be unsafe and not yield the desired outcome of helping them evaluate their concerns.

"AI researchers and engineers are on the frontlines, seeing issues that could pose substantive risks. But without a safe path to get an outside expert opinion, they're left wondering if their concerns are warranted," said Karl Koch, co-founder of OAISIS. "Third Opinion gives them that confidential path to clarify the facts and find the right course of action - all while preserving their identity."

Through Third Opinion, AI professionals can submit questions relating to a wide range of potential issues in the development or deployment of AI models. Questions can range from technical safety in development to misuse of models, from concerning organizational practices to potential cybersecurity vulnerabilities, from political bias to lobbying and more - all without sharing confidential information. Third Opinion then carefully vets the inquiries and orchestrates anonymous information exchanges with trusted, independent experts to provide qualified, unbiased assessments of their questions.

A year of research by OAISIS found that the lack of a trusted, neutral resource is a major obstacle preventing important AI issues from being addressed effectively. "Each individual

---

[1] https://righttowarn.ai/

retains all the power of how to proceed, if at all, with the opinions they have received" said Max Nebl, OAISIS co-founder. "We are focused solely on helping AI professionals understand their situation and make smarter risk management decisions."

William Saunders, a former OpenAI research engineer who spoke up on the industry's non-disparagement clauses and recently testified in the US Senate, added, *"AI labs are headed into turbulent times. I think OAISIS is the kind of resource that can help lab employees navigate the tough ethical challenges associated with developing world-changing technology, by providing a way to safely and responsibly obtain independent guidance to determine whether their organizations are acting in the public interest as they claim."*

About OAISIS

OAISIS is an independent nonprofit dedicated to supporting AI professionals who seek to address critical concerns in the public interest. The organization's mission is to enable responsible innovation by equipping concerned individuals with the resources required to fulfill their duties. A network of senior advisors and collaborators in AI, the law, and journalism support OAISIS' work.

For more information, please visit Third-Opinion.org [Embargoed/WIP until 12th of December 2024].