919

Adaptive Trust Infrastructures for AI: From an AI ClinicalTrials.gov framework to a Human-AI Workforce

Samuel Hill

Deutsches Forschungszentrum für Künstliche Intelligenz, Germany

Description

The former U.S. Surgeon General warned about social media's mental health risks, citing links to anxiety and depression. Similarly, AI is reshaping health, culture, and society, yet current governance frameworks remain fragmented, reactive, and largely focused on technical compliance rather than psychological, ethical, and societal impacts. This gap leaves high-risk AI systems under-evaluated before deployment, amplifying potential harms and eroding public trust.

We propose AI ClinicalTrials.gov - a novel, centralized system inspired by the NIH's ClinicalTrials.gov - to institutionalize responsible AI development and deployment. AI ClinicalTrials.gov would require organizations introducing new AI products or major updates to register and assess psychological, ethical, and societal impacts prior to release. This framework shifts the prevailing "move fast and break things" ethos toward a "test, validate, and deploy responsibly" paradigm. Drawing on principles of medical trial transparency, the system would enable trust-risk simulations, ethical impact statements, and conditional release protocols. Evaluations need not mirror the cost or duration of medical clinical trials; they could involve small-scale pilots, simulations, and longitudinal studies. For high-risk AI systems, deployment would be contingent on meeting predefined safety and trust thresholds alongside regulatory compliance.

Beyond risk mitigation, AI ClinicalTrials.gov would foster interdisciplinary collaboration, enhance public trust, and create a shared foundation for responsible innovation. Crucially, it would embed feedback loops across social sciences, health, law, policy and governance to ensure these fields evolve and adapt with the rapid pace of AI development.