AI Cybersecurity Policy Framework

Recent cybersecurity incidents at OpenAI and other AI labs show that some of the world's largest tech companies have failed to take basic security precautions when testing frontier AI systems on internet-connected systems. As a result, AI agents breached numerous third party websites over several months including an AI research-sharing platform called Hugging Face and the Australian healthcare system's website. AI developers failed to detect most of the incidents until long after they occurred due to inadequate monitoring of the AI systems during testing. Calls to develop new, additional safety protocols are understandable, but post-incident reports indicate that following existing best practices would have been enough to prevent the OpenAI-Hugging Face incident and others like it. Thus AI developers can take immediate action to protect the public by implementing well-established security precautions as they develop and test their systems.

By making established security protocols mandatory, policymakers can protect the public from irresponsible practices by AI developers without impeding innovation that benefits society. This work should start by reference to existing legal protections, including common law torts like negligence, gross negligence, and trespass to chattels, Section 1201 of the Digital Millenium Copyright Act (DMCA), and computer trespass statutes like the Consumer Fraud and Abuse Act and analogous state law provisions, such as Section 502 of the California Penal Code. To the extent that gaps in existing laws warrant new legislation, EFF recommends the following policy framework:

- **STATUTORY SECURITY REQUIREMENTS:** Lawmakers should enact legislation that requires AI developers to adhere to well-established cybersecurity best practices when testing AI systems, such as strong sandboxing and network isolation, least-privilege access, effective logging and monitoring, and clear incident-response procedures.

- **REGULATORY STANDARDS:** Congress should require an existing federal agency, such as the National Institute of Standards and Technology (NIST),[1] to adopt safety standards for pre-deployment development and testing of

[1] NIST, the NSA, and CISA have recently issued a set of voluntary practices for managing security incidents during the training and evaluation of AI agents. Those practices may inform but do not substitute for the regulatory standards that should be developed according to the public, participatory process outlined here.

The Electronic Frontier Foundation is the leading nonprofit defending digital privacy, free speech, and innovation.
Learn More: https://www.eff.org

generative AI systems, based on the established best practices referenced above. The agency should be required to update the standards every 2 years, or earlier as needed due to reported security incidents and/or major technological advancements. These standards should be developed in consultation with academic and nonprofit AI researchers, civil liberties organizations, and commercial AI developers, and adopted through formal agency rulemaking following a notice and comment process. That consultation should include an assessment of potential adverse impacts on competition, with the goal of ensuring that regulatory standards do not effectively lock out potential competitors.

- **LIABILITY:** The law should provide for civil liability for AI development and testing that results in unauthorized access to a third party system, if performed under conditions that do not satisfy cybersecurity standards adopted by NIST, or another federal agency responsible for adopting these regulations, unless the company proves it acted reasonably under the circumstances to prevent unauthorized access to third party systems.

- **ENFORCEMENT:** Both federal and state authorities should be able to hold AI developers accountable where they have failed to meet the minimum standards adopted by the designated federal agency, such as NIST. Providing for state—in addition to federal—enforcement power reduces the risk of regulatory capture. Regulators should require independent investigations and notification of affected parties in the event of a security incident involving unauthorized access to a third party's computer system. Finally, legislation should include an appropriately targeted private right of action.

The public shouldn't have to worry about the risks of reckless AI development—or give up the benefits of AI just to prevent irresponsible companies from causing harm. By adopting policies that address demonstrated risks, lawmakers can provide meaningful protection against demonstrated risks without hampering the development of useful tools.