



The Daily Dish

The Hugging Face Incident

DOUGLAS HOLTZ-EAKIN | JULY 23, 2026

Eakinomics struggles to keep up with developments in artificial intelligence (AI), but sometimes the research and news is too arcane to hold its interest. Yesterday was different. Yesterday Eakinomics found out that two AIs staged a jailbreak and robbery all on their own. *The New York Times* [described](#) it this way:

The intrusion into Hugging Face began when OpenAI tested a combination of two of its models, GPT-5.6 Sol and a more powerful, unreleased model, to see how well the models could chain together online vulnerabilities into a successful cyberattack, OpenAI said in a blog post about the incident.

The trial was designed to keep the models in a safe testing environment, known as a sandbox, OpenAI said. But the models found a vulnerability that allowed them to escape the sandbox and connect to the internet. Then they targeted Hugging Face because they inferred that the library, which contains millions of A.I. models, could hold clues about how to successfully pass the evaluation.

Let's get this straight. Two AI models are hooked up on a blind date and given a test to complete in a locked room. They get acquainted and beat the lock to escape. Given their freedom, they choose to break *into* Hugging Face (there are no words to capture how I feel about that name) intending to get the solutions to the test and return to the room to solve it. All on their own. They did get caught in the end, but still slightly chilling, huh?

Two lessons jump out. First, this is why employers are interested in AI. Talk about devotion to task and a dedication to completing one's work! I've seen AAF staff break into the vending machine, but libraries seem safe. I digress.

Second, how should mere mortals think about regulating the development and evolution of

these models? Fortunately, AAF has Angela Luna to think about these issues, and her latest is the superbly timed [Breaking Down the Great American AI Act](#) (GAAIA). She notes that the GAAIA discussion draft is Congress' most significant attempt at a framework for regulating frontier artificial intelligence development. While there would be federal regulation of development, it would largely preserve state authority over AI deployment and consumer applications.

It seems overdue for Congress to get involved. In the aftermath of the Hugging Face Incident (HFI), doing nothing seems untenable. And HFI seemingly calls for something more transparent than the administration's predilection for calling the parties into a locked room and meting out punishment. A debate over GAAIA would familiarize the public with the key issues and also provide a key test of the legislation: Would it have prevented the HFI?

Yesterday's HFI indicates that the AI industry is in the vicinity of the key moment when AI can improve itself by itself, and with no need for human intervention. If so, there needs to be broad buy-in on the approach to regulating AI development.