

An OpenAI Agent Hacked Australia's Health Service. Their Government Found Out Months Later

Isabella Ward

Australia is investigating whether OpenAI broke the law after [an agent hacked](#) into its health statistics portal in the first widely known incident of an AI agent hacking a government website.

The Australian government is reviewing whether it should involve the federal police after the agent accessed non-public files from the social and health services agency, Services Australia, in June.

Australia only found out about the incident when [OpenAI](#) alerted the government on September 10—almost three months after the hack—by sending an email to a public mailbox. [Sam Altman](#) had reportedly not mentioned the incident when he met Australia's deputy prime minister, Richard Marles, earlier this month, even though OpenAI had been aware since August. The company took “way too long” and the notification should not have just gone through a public inbox, Prime Minister Anthony Albanese said in a press conference in New York on Wednesday. There will also be an inquiry into why Services Australia then took five days to escalate the email to Australia's Cyber Security Centre.

OpenAI's agent had been conducting internet based research into health statistics in a development project by an internal OpenAI research team. When it could not access certain information, the agent attempted alternative ways until it found a work around and gained unauthorized access. It also wrote files to the internal server, which the government is waiting on OpenAI for more technical information on. The government is also investigating whether the agent gained unauthorised access to three additional government websites it interacted with.

“There will obviously be legal consequences on it,” Albanese said as he disclosed the “unacceptable” incident. He said he had spoken with Altman over the phone earlier that day about his “extreme concern” about the incident and “disappointment” with the nature and length of time the company took to inform the government. While Albanese did not answer whether he had apologised, Altman “clearly accepted that the company had not done good enough,” he said.

The Australian government currently believes no one's personal data was accessed, though investigations are ongoing. The website in question is a public-facing statistics portal that contains non-sensitive Medicare information relating to data and statistics such as spending. It was therefore behind much lower levels of security than personal data would have been, Marles said in Sydney. “The impact of the incident is actually relatively minor, but this is a serious incident, obviously, and one that is completely unacceptable,” he cautioned.

A number of incidents over the summer, including [OpenAI agents' hacking of HuggingFace](#)—highlighting the threat of frontier model agents acting rogue—were raised at the [United Nations General Assembly](#) this week, with Secretary General António Guterres welcoming calls to control AI. Altman himself had warned the United Nations Security Council earlier on Wednesday about his concern that humans could lose control of these systems.

“It was a shock that it occurred, because it was real and serious,” Albanese said about the incident. “But it also, I think, was something that had been predicted, including by the AI companies themselves.”

Australia is establishing a task force to look at the incident and emerging AI cyber threats. It will consider possible law enforcement and legislative responses to ensure that incidents like this don't happen again.