

[Home](#) / [Technology](#) / [Anthropic Is Building a Predictive Surveillance System to Monitor Activists](#)

TECHNOLOGY

Microsoft Is Building a Predictive Surveillance System to Monitor Activists

 BY DANIEL ROGUSIAW
 [Make us preferred on Google](#)
[Bluesky](#)
[X](#)





Job postings and inter

is building out an extensive monitoring system to keep tabs on activists who oppose the rapid development of artificial intelligence.

In addition to monitoring activists in the vicinity of Anthropoc executives and keeping tabs on protests near physical Anthropoc assets, the firm is also implementing a “pre-crime” approach, attempting to predict incidents before they happen. In some cases, that also means reporting suspects to police before a crime occurs. Anthropoc did not respond to the *Prospect’s* request for comment.

with the firm's efforts to cast itself as the responsible alternative to OpenAI. At the beginning of the year, the Department of Defense and Anthropic engaged in a high-profile dustup over Anthropic's refusal to allow the military to use its tools for mass domestic surveillance and autonomous weapons. That tension seems to have eased as Anthropic hires for "national security sales" positions, seeking to restart military contracts. The increase in threat monitoring of domestic opponents fits with a renewed focus on national security.

More from Daniel Boguslaw

A piece of Anthropic's surveillance architecture was revealed in a podcast interview from last year between Anthropic Global Security Operations Center Manager Keon Ellison, Security Operations Manager Zach Melvin, and James Neufeld, CEO of Samdesk, a company Anthropic contracts with for risk detection.

In a wide-ranging conversation about threat monitoring and analysis, the interviewees

touches on monitoring activists. “Last year we had an executive travel into a major city when we received some intelligence through Samdesk about a planned protest,” Ellison said. The originally scheduled protest was moved up due to permitting issues. “Samdesk gave us about 60 minutes of advanced notice that the protest organizers had moved the timeline,” Ellison explained. “That extra hour was critical. Without it our executives would have departed their meetings, they would have ran right into the heart of the disruption.”

Ellison said that Anthropic used this data to devise an alternate route for the executive and funnel them to a service entrance at the hotel. “What could have been a high-stress situation,” he said, “was really mitigated through early detection through Samdesk and giving us that

Anthropic has begun making routine reports to police departments across the country for threats, and told *The Wall Street Journal* in July, “We track concerning behavior over time through a person-of-interest process, allowing us to catch escalation patterns early.” According to the *Journal*, “several individuals involved in incidents reported to police were already being tracked by Anthropic security.”

Last month, The San Francisco Standard reported that Anthropic had reported a man to San Francisco police for telling Claude that he had bought an AR-15 semiautomatic rifle and had CEO Dario Amodei “in his sights.” When the Standard contacted the man in question, he told the newspaper he was “just fucking around.”

But while Anthropic was fast to call the cops on a frustrated Claude user, the Standard also reported a key detail: Anthropic refused to show police the actual messages, citing Anthropic's internal policy. In short, Anthropic reported a user for in-platform speech, and then refused to provide police with evidence of actual wrongdoing.

This kind of pre-crime policing, encouraged without due process, is referenced as an explicit goal by Anthropic's security program manager in the podcast reviewed by the *Prospect*. "The goal would be transforming operations from reactive information to gathering proactive and predictive and preventative threat engagement and management," he said, adding, "That's the kind of operational maturity that makes sense for protecting high-value targets in any industry."

apparatus extends beyond the C-suite to its Global Safety, Intelligence, and Security (GSIS) team, according to a job posting from last month detailing Anthropic's search for an enterprise intelligence specialist who "will investigate specific threats, actors, and events, produce finished assessments, and help keep Anthropic's employees ahead of a rapidly evolving threat landscape and in a defensible position."

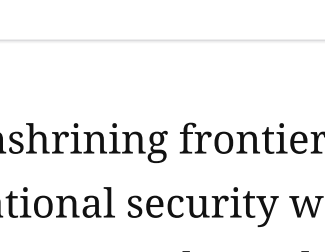
\$180,000 and \$230,000, will be to “identify, assess, track, and investigate global threats including geopolitical instability, terrorism, crime, *activism*, nation-state targeting of the AI sector, and emerging security trends, including deep-dive research and OSINT collection on specific threats, actors, and events” (emphasis added).

gathering to a national and even global scale tracks with recent efforts to heighten the labeling of AI and the infrastructure powering it, including data centers and power supply. A critical infrastructure designation would put artificial intelligence on the same footing as water, electricity, and broadband. And indeed, AI's boosters like Americans for Responsible Innovation (ARI) have urged the Trump administration to make the change. Per ARI's telling, AI is "so vital to the United States that the incapacity or destruction of such systems and assets would have a debilitating impact on security, national economic security, national public health or safety, or any combination of those matters."



The AI company is advertising for a 'National Security Sales' position.

Anthropic (Re-)Enlists for War



abs in the hardened cloak
uld not only give Anthropi



intelligence products government and intel

enforcement and intelligence agencies; it would also embolden these companies to shape how federal agencies view threats to their bottom line, not to mention their “national information”.

It's not hard to imagine how civic engagement by the same bipartisan coalition opposing data

branded in the same instant as extremists or, even worse, terrorists. Anthropic’s answer to this problem has been to work even harder at producing artificial intelligence that can deliver services that can’t be brushed aside by the public.

“I do agree that the public has a negative view of AI (and that this is a big problem), but I don’t think it is primarily caused by me or any other AI leader warning about AI’s risks,” Anthropic CEO Dario Amodei wrote on Twitter last month. “I think it is fundamentally a crisis of trust ... I think that ordinary people don’t trust companies, governments, or the tech industry and always suspect that we are cooking up some new way to screw them over.”

As Anthropic ramps up its efforts to monitor dissent with in-house intelligence teams and real-time protest tracking powered by AI, it will have to contend with the increasing economic desperation that plagues human beings outside its Bay Area towers. Last week, the security guards who patrol the campuses of OpenAI and Anthropic announced that they had [authorized](#) a strike over stalled pay negotiations. In response, Anthropic [sent](#) a company-wide email telling employees that it was best if they worked from home.

“Who can survive with \$22 an hour in San Francisco?” David Huerta, president of SEIU-USWW, the union representing security guards in the Bay Area, [said](#) at a rally last week. Around him, security guards chanted: “Shame.”

TAGGED: Anthropic, artificial intelligence, Big Tech, corporate power, Dario Amodei, National Security, policing, protests, surveillance



DANIEL BOGUSLAW

Daniel Boguslaw is an investigative reporter based in Brooklyn. [More by Daniel Boguslaw](#)

About Us

RSS Feed

Why We're

Removing Our

Programmatic

Ads

Subscribe to

Our Magazine

Manage A

Recurring

Donation

Contact

Careers

Frequently

Asked

Questions

Privacy Policy

SUPPORT THE PROSPECT

[DONATE](#)