

AI Security Warning from Signal app's Meredith Whittaker – The Hidden Dangers of Agentic AI

2 MIN • YOUTUBE • [HTTPS://WWW.YOUTUBE.COM/WATCH?V=JE_CNEzJV70&T=2S](https://www.youtube.com/watch?v=jE_CNEzJV70&t=2s)

https://www.youtube.com/watch?v=jE_CNEzJV70&t=2s

SUMMARY

The discussion highlights the emerging concept of agentic AI, which offers advanced capabilities to automate tasks by accessing personal data across various platforms. While this technology promises convenience, it raises significant concerns regarding privacy and security, as it requires extensive permissions that could compromise sensitive information.

- *Agentic AI is a new trend that automates tasks by integrating with personal data systems.*
- *It can perform complex actions like booking tickets and messaging friends without user intervention.*
- *Such systems require extensive access to personal data, including calendars, credit card information, and messaging apps.*
- *The processing of data is likely to occur in the cloud, raising concerns about data security and privacy.*
- *There is a risk of undermining privacy protections as these systems may blur the lines between different applications and data sources.*
- *The convenience offered by agentic AI could lead users to relinquish control over their personal information.*
- *The hype surrounding these technologies may overshadow the potential dangers associated with surveillance and data misuse.*

constructed. We can get into that as well. But um an agent agentic AI this the thing that is the new buzzword that everything that used to be an assistant now is um I think there's a real danger that we're facing and signal is clocking this pretty closely of the introduction of this sort of notion of agentic AI into our devices and and lives in part because what we're doing is giving so much control to these systems that are going going to need access to data. So if you you read the kind of you you read the marketing on this, the the value ad is something like can look up a concert, book a ticket, schedule it in your calendar and message all your friends that it's booked, right? And so we can just put our brain in a jar because the thing is doing that and we don't have to touch it, right? And so what would it need to do that? Well, it would need access to our browser, an ability to drive that. It would need our credit card information to pay for the tickets.

It would need access to our calendar, everything we're doing, everyone we're meeting. It would need access to signal to open and send that message to our friends. And it would need to be able to drive that across our entire system with something that looks like root permission, accessing every single one of those databases, probably in the clear because there's no model to do that encrypted. And if we're talking about a sufficiently powerful model AI model that's powering that, there's no way that's happening on device. Even though ondevice isn't a prof prophylactic isn't going to really solve a lot of those issues, that's almost certainly being sent to a cloud server where it's being processed and sent back. So there's a profound issue with security and privacy that is

haunting this sort of hype around agents and that is ultimately threatening to break the blood brain barrier between the application layer and the OS layer by conjoining all of these separate services, muddying their data and doing things like undermining the privacy of your signal messages because hey, the agents got to get in, the agents got to text your friends, the agents got to pull the data out of your text and got to summarize that so that again your brain can sit in a jar and you're not doing any of that yourself.

You're doing something else. So I think we need to be really careful, you know, when I think about the immediate concerns, not simply the history of AI and the fact that it's, you know, predicated on this larger surveillance model. There's a real issue right now of the undermining that AI systems are poised to do in these privacy and security guarantees in the name of this sort of, you know, magic genie bot that's going to take care of the exigencies of life.