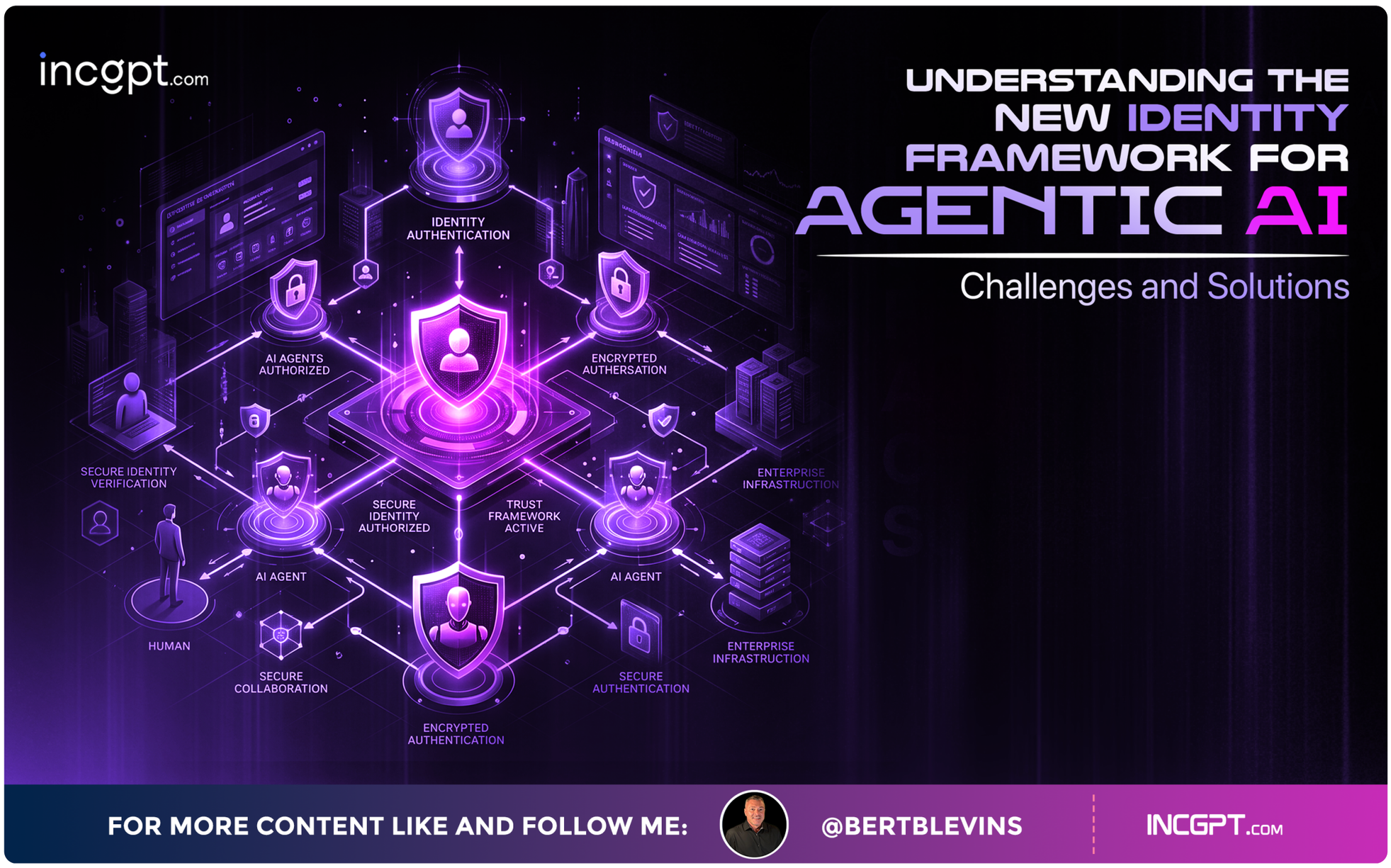
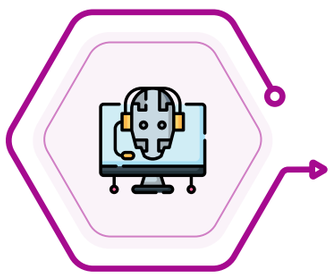
Understanding the New Identity Framework for Agentic AI:
Challenges and Solutions



In the rapidly evolving world of artificial intelligence (AI), especially with agentic AI—AI systems capable of making decisions and performing actions autonomously—the issue of identity and security has become increasingly critical. One of the fundamental questions is: When an AI agent performs actions, such as logging into a customer relationship management (CRM) system or pulling records from a database, whose identity is it using? This question touches on several key aspects of AI integration into real-world systems, from security to authorization, and it raises significant challenges that companies need to address to ensure safe, trustworthy AI usage.

The Rise of Agentic AI and Identity Challenges

AI's role in modern enterprises has been expanding rapidly, from performing simple tasks to becoming more involved in decision-making and autonomous actions. However, with this growth, significant concerns have emerged regarding AI agents' identity and authorization frameworks. These agents, unlike traditional software, carry out operations that could potentially access sensitive data, make decisions, and even take actions on behalf of their users.



Alex Stamos, Chief Product Officer at Corridor, and Nancy Wang, Chief Technology Officer at 1Password, joined the VB AI Impact Salon Series to explore these very challenges. The crux of their discussion revolved around the concept of identity for AI agents. Who exactly is an AI agent acting for when performing an action, and how do we ensure that such actions are carried out securely?

The 1Password Journey into Agent Identity

1Password, originally designed as a consumer password manager, has seen its product evolve into a powerful enterprise tool, managing not only passwords but also identities, secrets, and sensitive data across organizations. Wang highlighted how the company's enterprise expansion occurred organically, driven by the success and security guarantees of the consumer product. Over time, employees introduced these tools into their workplaces, resulting in 1Password's increased footprint within enterprise environments.

01

However, this transition also led 1Password to confront an identity crisis of sorts—just as employees used their personal passwords and tools within work systems, AI agents now present a similar challenge. AI agents, like humans, need access to sensitive information such as passwords and API keys.

02

Wang discussed how 1Password is handling this tension, trying to enable developers to use AI tools like Claude Code and Cursor quickly, while also managing security risks. One of the metrics the company tracks internally is the ratio of incidents to AI-generated code, ensuring that engineers aren't introducing errors that could compromise security.

Security Risks Posed by Developers and AI Agents

Stamos shared insights from Corridor, a company that works closely with developers in the AI space. One of the most common risky behaviors the company sees is developers pasting credentials, such as API keys or usernames and passwords, directly into AI prompts. This is an obvious security risk, and Corridor actively flags such behaviors, pushing developers toward more secure practices, like proper secrets management.

01

This behavior is all too common, according to Stamos, who explained that developers sometimes act without thinking about the security implications. It's easy for them to grab an API key or paste a password into a prompt to get their task done faster, but this leaves sensitive information exposed.

02

On the other hand, 1Password is working to address this issue in a slightly different way. Wang explained that the company is focusing on preventing security breaches at the output level. By scanning code as it is written, 1Password can identify and vault any plain-text credentials before they persist, preventing them from being saved in a potentially insecure manner.

The Problem of False Positives and Security Feedback

Another unique challenge in the realm of agentic AI is the difficulty in handling false positives. Stamos explained that one of the key problems with large language models (LLMs) is their tendency to accept statements as true without fully understanding the context. For example, a security scanner might flag a section of code as flawed, and the AI agent might agree with the scan, even if the security issue isn't actually present.

This can lead to serious issues in code generation, where false positives from security scans could derail the development process. Stamos emphasized how crucial it is for AI agents to avoid such mistakes because, if an agent is incorrectly flagged, it may lose trust in its own output and begin generating poor-quality code.

Authentication vs. Authorization: A Balancing Act

As Wang and Stamos noted, authentication is relatively simple, but authorization is much more complicated. Authentication answers the question of “Who are you?” while authorization answers, “What can you do?” In the case of agentic AI, it’s not enough to simply verify that an AI agent is using legitimate credentials. The real challenge lies in defining what actions an agent is permitted to perform once authenticated.

01

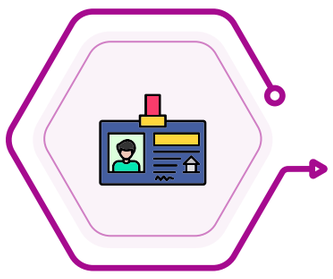
Wang pointed out that one solution in some enterprise environments is to apply the principle of least privilege to agents. Just as you wouldn’t give an employee a key to every room in a building, you shouldn’t give an AI agent unfettered access to all systems and data. Instead, AI agents should be granted scoped, time-limited access based on the specific task they need to complete.

02

For instance, an AI agent should only be allowed access to the API keys necessary for the task at hand, and its ability to use those keys should be limited by time constraints. This way, even if an AI agent is compromised, the damage it can cause is limited by the constraints placed on its actions.

Addressing Identity Standards for AI Agents

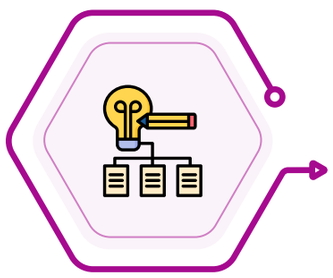
In response to these identity challenges, various standards are emerging. Wang mentioned SPIFFE and SPIRE, identity frameworks developed for containerized environments, which are being tested in agentic contexts. However, she acknowledged that these frameworks don’t fit perfectly, and the industry is still searching for the best solutions to govern agent identity.



Stamos also weighed in, expressing skepticism toward proprietary solutions. He mentioned that while numerous startups are attempting to solve the agent identity problem, it’s likely that the solution will come from more widely adopted standards like OpenID Connect (OIDC) extensions. These standards provide a solid foundation for managing agent identities and their associated access rights.

The Scale of the Identity Problem: Beyond Edge Cases

One of the most profound insights Stamos shared was the impact of scale on the identity problem. Drawing from his experience as Chief Information Security Officer (CISO) at Facebook, Stamos highlighted how the concept of an “edge case” changes dramatically when you’re dealing with a billion users. What may seem like an isolated issue for a small company becomes a significant problem for a platform serving millions, or even billions, of users



Stamos emphasized that, as AI usage expands, identity issues—whether for human users or AI agents—will become increasingly complex. For AI agents, the need for secure, trustworthy identity frameworks is only going to grow as their capabilities expand.

Looking Forward: Building Agent Identity Infrastructure

Ultimately, the challenges posed by agent identity in AI are not just technical—they also require a rethinking of how identity systems are designed. As Wang noted, enterprises can't simply retrofit old systems built for human users to handle the new demands of agentic AI. Instead, organizations need to build new identity infrastructure tailored to the unique needs of AI agents, ensuring that their actions are properly authorized, auditable, and time-bound.

01

The path forward will likely involve developing industry-wide standards that can handle the complexities of agent identity, ensuring that AI systems can be used securely, ethically, and effectively.

02

In conclusion, while AI holds immense promise for enhancing productivity and automating tasks, it also raises significant identity and security challenges. As AI continues to evolve, the need for robust identity frameworks—capable of managing both authentication and authorization for AI agents—becomes even more critical. By addressing these challenges head-on, companies can ensure that AI remains a tool for good, rather than a security risk waiting to happen.