The Identity Salon™

Identity Salon February 2025.v - Summary

In February 2025, *The Identity Salon*™ hosted our second virtual meeting, gathering identity and security professionals, researchers, and policymakers for a lively and insightful discussion on Artificial Intelligence (AI) and its impact on modern identity and access management (IAM). As with all salon events, the meeting was held under the Chatham House Rule to provide a platform for candid conversations about the challenges and opportunities shaping the field today. Here's a summary of the key takeaways.

Discussion

The discussion between Salon attendees focused on the threats and challenges of AI. We did note later that there are also opportunities and improvements possible with judicious use of AI; further discussion on that aspect of AI in Identity will need its own event.

The Expanding Role of AI in Identity and Security

The first virtual Identity Salon of 2025 focused on the expanding role of AI in identity and security, highlighting both opportunities and risks. As AI becomes increasingly embedded in digital systems, fraud and authentication challenges are evolving rapidly. People and businesses are now facing sophisticated AI-driven threats, including deepfakes, synthetic identities, and adversarial machine learning attacks. Fraud is no longer just a byproduct of AI advancement; it has become a service, where attackers develop and sell AI-powered tools to scale phishing, identity fraud, and other malicious activities.

Even discussing AI presents challenges, from agreeing on terminology to questioning underlying assumptions about AI-driven behaviors. A central debate emerging from the discussion was whether the default assumption should be that an entity is human unless proven otherwise, or that it is AI unless verified as human. This fundamental question will shape future identity verification models and personhood credentials. Participants also raised concerns about whether personhood credentials, as currently envisioned, risk a proliferation of enforced identity proofing, which would raise significant privacy concerns.

AI-Driven Fraud and Authentication Challenges

A real-world example of AI-driven fraud discussed was a 2024 incident in Hong Kong, where attackers used AI-generated deepfake video in a Zoom call to authorize $25 million in fraudulent wire transfers. Even though the employee followed all expected security protocols, the attack succeeded because AI allowed adversaries to mimic trusted individuals with near-perfect accuracy. This case underscored the growing need for stronger authentication mechanisms capable of distinguishing between human and AI-driven interactions.

One key question raised was whether businesses should prioritize verifying human identity or instead focus on detecting non-human entities. Should AI interactions be assumed unless verified otherwise? The industry has yet to reach consensus, and the implications for authentication are profound. Additionally, concerns were raised about OpenAI and WorldCoin’s push for personhood credentials, with some viewing this as an attempt to control a problem they have helped create.

Emerging AI Threats: Prompt Injection and Data Poisoning

Beyond fraud, the discussion explored prompt injection attacks and data poisoning, two emerging threats to AI models. While these techniques can manipulate AI outputs, their real-world feasibility remains debated. A major concern raised was the lack of precision in how these terms are used, leading to uncertainty about their practical impact on identity security. Participants clarified that while prompt injection does not alter the underlying learning model itself, it can still compromise the accuracy and reliability of AI-generated responses.

Data provenance emerged as another significant issue, particularly in verifying whether an interaction originates from a human, an AI, or a hybrid of both. This is an area of focus for groups like the Coalition for Content Provenance and Authenticity (C2PA) and the Content

Authenticity Initiative (CAI). Liability concerns are also a concern; what happens if an AI agent, acting on behalf of a human, makes a commitment that the individual later disputes? Should businesses be responsible for distinguishing AI from human users, and to what extent? How can individuals reliably identify an AI agent which is acting on behalf of an enterprise? Questions such as these remain unresolved but are becoming increasingly urgent as AI adoption accelerates.

The Challenge of Identity Delegation

The discussion also addressed the long-standing problem of identity delegation. Agentic AI exacerbates this issue, as these systems can operate independently, potentially undermining existing authentication frameworks such as passkeys. If AI begins managing authentication on behalf of users (something it's starting to do now), traditional security models will be disrupted.

Participants noted that delegation is already a challenge in business environments, particularly in call centers and marketplaces, where advisors act on behalf of customers. While some systems explicitly log "on-behalf-of" transactions, others grant blanket access, raising significant security concerns. Without clear policies and standards, businesses will find themselves navigating legal gray areas when interacting with AI-driven agents. Attendees agreed that the industry must proactively define liability frameworks and standards before regulations impose fragmented solutions. Unfortunately, at this stage, fragmented solutions seem more likely than unified industry responses.

AI-Driven Data Privacy and Leakage Risks

Another critical issue was AI-driven data privacy and leakage risks. As organizations increasingly use AI to extract personal information from structured and unstructured documents, concerns around data exposure continue to grow. The discussion highlighted a lack of industry consensus on what constitutes data leakage, with distinctions made between risks at the machine learning phase versus the deployment phase.

The EU Data Protection Board has issued guidance stating that models can be trained on personal data (PD) without that data being considered part of the model. However, this does not address whether the data can be reconstituted or leaked through indirect means. Some participants suggested that tokenizing PD before feeding it into AI models could mitigate leakage risks, though practical implementations remain uncertain; and it was noted that—at least under the GDPR—an item of PD which is rendered unidentifiable *per* *se* but which may still be made identifiable when combined with other data, is still PD. Since both the corpus of data an AI model uses and the way in which it might combine items of data are opaque, the risks of sharing even tokenized PD are not insubstantial. Questions about enforcement and practical implementation remain unresolved, particularly as businesses increasingly rely on AI for identity verification.

AI in Identity Systems and Regulatory Challenges

The use of AI in identity systems makes those systems increasingly vulnerable to data breaches and privacy violations. In the identity verification space, businesses are leveraging AI to extract information from unstructured documents (e.g., utility bills). Depending on jurisdiction and even the business's own terms of service, this practice may violate privacy laws.

Biometrics are another area of concern, as companies feed biometric data into AI systems without sufficient consideration of privacy implications. The push for efficiency and cost savings often means business leaders overlook the ethical and legal risks raised by technologists. Without stronger industry standards and regulatory engagement, businesses risk facing compliance challenges as policymakers begin crafting AI-specific regulations.

Next Steps for the Industry

Looking ahead, the industry must take decisive action in several areas:

- **Establishing Clear Liability Frameworks**: Businesses must understand their responsibilities when AI-driven interactions occur. At a minimum, organizations should update their terms of service to clarify liability when engaging with AI agents.
- **Developing Standards for AI in Identity**: The industry needs clearer guidelines for verifying AI-driven interactions and ensuring interoperability between authentication models. If left unchecked, AI's role in identity management could lead to fragmented and inconsistent solutions.
- **Addressing Identity Delegation at Scale**: Businesses must develop structured approaches to managing AI acting on behalf of users. Without addressing the delegation challenge, AI may become an unintended attack vector for fraud and identity theft.
- **Improving Data Provenance and Trust Indicators**: Verifying the authenticity of digital interactions must become a priority. Organizations need mechanisms to determine whether an entity is AI-driven or a verified human.

- **Exploring Tokenized Data Models**: Investigating AI models built on tokenized data may offer an approach to reducing privacy risks while maintaining AI's effectiveness in identity applications.
- **Staying Engaged with Policymakers**: To prevent a reactive and fragmented regulatory landscape, the industry must proactively shape AI-related identity policies. Engaging with regulators now can help establish practical frameworks that balance innovation with security.

Finally, a future discussion was proposed on whether IAM should build its own identity-focused LLM. Over the next twelve months, we hope to see tangible progress in these areas. The ability to define AI interaction transparency, delegation standards, and liability frameworks will determine whether identity practitioners can stay ahead of these rapidly evolving challenges.

Additional Reading

- "EDPB opinion on AI models: GDPR principles support responsible AI" — EU Data Protection Board (18 December 2025)
- "Authenticated Delegation and Authorized AI Agents" — Tobin South et al. (16 January 2025)
- "The Call Is About To Come From Inside The House" — Nishant Kaushik (3 February 2025)
- "AI Agent Security: A Framework for Accountability and Control" — Ryan Hurst (3 February 2025)

Side Note - How We Used AI to Help Generate This Report

In the spirit of transparency, we used AI to help structure and refine this report by feeding it the raw notes (unattributed, of course) from the meeting and the Zoom chat. While the insights and analysis come directly from you, AI-assisted tools were helpful for organizing key themes and summarizing the meeting notes. That said, any bad jokes, typos, extra semicolons, or formatting quirks are still entirely human-generated. We'll let AI take the credit for the structure… but we're keeping the humor for ourselves!