Trusted Identities for AI Agents: An Opportunity for Europe

AI agents are reshaping how people and organizations interact. They already assist with bookings, payments, and contracts. Soon, they will execute transactions, control industrial systems, and engage in agreements at scale. By 2030, billions of AI agents will operate across borders. The upside is significant, but does not come without risk. The rapid scaling of the use of AI agents requires a corresponding evolution in digital infrastructure. To ensure long-term safety and accountability, integrated trust frameworks are needed to complement existing governance. Without further trusted identities and strong guardrails it becomes increasingly complex to hinder unauthorized deals and fake agents as well as risks to production lines, grids, or vehicles.

Shaping the emerging agentic landscape

Europe is well positioned to shape this emerging agentic landscape. With the European Digital Identity Framework (EUDIF), the European Digital Identity Wallet (EUDIW), and the development of the European Business Wallet (EUBW), the EU has created a unique legal and technical foundation. States can reliably establish the identity of humans, businesses, and public bodies, while preserving privacy through mechanisms like selective disclosure. These identities can be carried in a technical concept called verifiable credentials, enabling cryptographically secure communications and signatures. This allows humans to act safely and ensure accountability on behalf of humans, businesses, and public bodies.

Extending this trust framework to AI agents allows nonhumans to act safely and accountably on behalf of humans, businesses, and public bodies. For example, it is important to define the roles that an AI agent can and should play in a digital communication flow of providers, users and relying parties.

Securing the payment chain

One of the most frequent and fundamental digital transactions where AI agents will have a profound effect is in payments. The industry must reduce fraud without sacrificing convenience. Fraud spans true fraud, when the payer did not authorize the purchase, and friendly fraud, when the purchase is made but later disputed.

Agentic commerce adds complexity:

- It introduces a new actor in the payment chain, the AI agent.
- It requires cryptographically provable intent and mandates.

Example: an AI agent is instructed to buy a pair of shoes of a specific brand and model, size 37, up to EUR 100. The agent must only transact with genuine merchants, proven by a valid merchant AI agent or equivalent evidence.

This adds a new chain of trust, AI agent to merchant AI agent, on top of existing payment roles. At scale, failures can multiply into astronomical fraud. Securing the payment chain is essential. Cryptography and verifiable credentials are core to trustworthy transactions and a stable foundation for autonomous financial activity.

The EU offers concrete capabilities to address these challenges:

- Validate the identity of the payment credential holder to confirm the intended account owner.
- Confirm the relationship between a human and their AI agent.
- Confirm genuine merchants and merchant AI agents, solving today's issue where merchants block AI agents because they resemble bots.
- Enable mutual authentication between AI agents, so payer and payee agents, and the humans or merchants behind them, are cryptographically verifiable. This creates a trusted agent-to-agent interaction model.

One compromised digital identity can disrupt the payment ecosystem. Now extend this to the European energy grid, where AI agents manage physical systems or the movement to connect data spaces, Internet of things and AI agents. A single compromised agent could trigger large-scale physical harm, alongside financial loss and data breaches.

A trusted internet

The same reasoning applies to today's Internet and what we can trust when we are online. We already see AI-generated images and videos, fake passports, receipts, and other valuable documents being tampered with. With a EU Digital Identity Wallet, content can be digitally signed to prove there is a human behind it, even if AI helped create it. This is a step to help provide a trust layer to the Internet, an "Authenticated Internet" that protects free speech with accountability, while allowing the use of pseudonyms. It could facilitate a safer place for children, too.

Work is currently underway to establish rights issuance frameworks for consumer-mandated AI agents, alongside securing chains of proof to ensure clear accountability in the event of system errors or failures. With the EU's Digital Identity Wallet and Business Wallet frameworks, Europe is building a trust layer that improves efficiency of the EU, while also extending naturally to AI agents to form first-class elements of AI governance.

Europe now has a clear opportunity to lead in a trusted Internet where AI agents provide value, and control remains with humans.

What we suggest

We suggest that:

- The European Union convenes stakeholders to form a strategy for safe AI agents based on the EDIF and the Business Wallet framework.
- Standards bodies create working groups on interoperability between EU Digital Identity Wallets and AI agents.
- The European Union prioritises testing and pilots, and regulates only where strictly necessary.

David Magård, Co-coordinator WE BUILD

Peter Busch, Director Technology Mgmt, Robert Bosch GmbH

Dennis Hannemann, Director Regulatory Strategy & Digital Identity, Bundesanzeiger Verlag

Willem Scalongne, Program manager AI, Netherlands Chamber of Commerce

Bo Fjelkner, Director Technology Regulation, Ericsson

Stavan Parikh, VP/GM Payments, Google

Carsten Stöcker, CEO & Founder, Spherity GmbH

Laurent Bailly, Digital Identity Business Development Director, Visa Europe

Contact information:

Joost Fleuren
Coordinator WE BUILD
joost.fleuren@kvk.nl

David Magård
Co-coordinator WE BUILD
david@siros.org

WE BUILD Website
https://www.webuildconsortium.eu/