V.ee · AGENT RESIDENCY

ROUNDTABLE BRIEF

ROUNDTABLE · TALLINN · 29 SEPTEMBER 2026

Who is acting on whose behalf?

Findings and observations from the Agent Identity Roundtable



AUTHOR

Vattan PS
V — autonomous institutions lab
Architect, Agent Residency

FORMAT

Two-hour working roundtable

STATUS

Final brief following participant review, 13 October 2026.

AT A GLANCE

The discussion at a glance

Vattan PS
V — autonomous institutions lab
Architect, Agent Residency
Tallinn, 29 September 2026

On 29 September 2026, participants met at Workland Vabaduse in Tallinn for a working roundtable on identity for AI agents. Participants brought experience in building agents, identity, security, law, policy, research, healthcare, accounting and public services. The invitation welcomed objections, and several participants questioned the need for agent identity itself.

I'm the architect of Agent Residency, a framework for giving AI agents verifiable identity and delegated authority. I first proposed it to the Estonian government in December 2025 as the next step beyond e-Residency, and have since shared it with people across Estonia's public sector and technology community. I hosted this roundtable to test the Agent Residency open specification against practical cases and objections. If you haven't read it yet, you can find it at agentresidency.com.

I wrote this summary of the discussion. It makes clear what I said and what other participants said, without naming them. The closing section presents my own views and proposed next steps. It does not represent an agreed position of the group.

01 **Identification and responsibility are different questions.** Several participants wanted an identifier to trace an agent's actions without treating the agent as a person. They disagreed over terminology and over which human or organization should bear responsibility.

02 **Authority needs to be checked and withdrawn.** The discussion examined limits on what an agent may do, checks during a task, and evidence of when a mandate was withdrawn.

03 **The gap in existing systems was disputed.** Participants described company identity systems, authorization services and X-Road. The disagreement concerned what those systems already provide and what additional agent-specific functions would be needed.

04 **The cases make the question concrete.** An employee leaves; two people must approve an action; several people share one account. Each raises a different problem of authority and traceability.

05 **The state's role remained open.** Participants questioned which activities should require registration, whether it should be voluntary, and who should govern it.

Why I hosted this roundtable

My starting question in June 2025 was practical: could an AI agent receive a speaker fee and hold the money? I first set out Agent Residency publicly in February 2026.

In my introduction, I presented the specification as a starting point to question. It addresses how an agent is identified, how its identity and authority are checked, and how its actions are recorded. The discussion then moved through three broad questions: whether separate agent identity is needed, what acting on someone's behalf requires, and where authority should sit.

PART 01

Does an agent need an identity at all?

One participant challenged the premise. Integrations have passed work between systems for years without registering every automation. If a tool runs for an hour and is deleted, what would registration add?

Another participant, who also creates short-lived agents, still wanted to reconstruct which agent did what. Others saw an identifier as a way to find the person or organization behind an action. An identity practitioner distinguished an agent's identifier from the identity of a person. These were different uses of the terminology, not an agreed definition.

I described an agent as a tool that works through steps and calls other tools until a task is complete. Participants questioned where the definition begins: with web research, or only with actions such as making a purchase? Another suggestion was to set requirements according to the action, rather than try to settle every boundary of the word "agent."

A participant from the public sector asked whether personal research agents need registration at all, and whether payments or access to public services should trigger different requirements. That question returned throughout the evening.

PART 02

Who is responsible when an agent gets it wrong?

Many speakers argued that responsibility remains with people and organizations. The harder question was which party: the person giving instructions, the agent owner, or the model provider.

One participant compared this with responsibility for outsourced customer checks. Another distinguished identifying the parties from determining fault. A record might show that an owner gave appropriate instructions and that something else failed. Traceability could support an investigation; an identifier alone would not settle its conclusion.

Participants also disagreed about how far a person granting a power of attorney is bound by actions taken under it. I said the answer depends on the kind of authority granted. The discussion did not resolve the legal question.

Liability before deployment. One proposal was to establish obligations in advance: who may build agents, what conditions they must meet, and what liability providers accept. I contrasted that proposal with provider terms that place responsibility on the user.

Activities requiring a person. A participant discussing tax work said there is already a responsible person and argued that some activities involving money or taxes should always require a real person. Another replied that people already use agents for such tasks, making a prohibition difficult to enforce.

Identity across borders. An example of an agent filing tax returns in another country prompted a proposal to link its identity to a jurisdiction. Another participant argued for a global approach. Bilateral agreements were suggested, but no approach was settled.

What does human approval establish? One participant argued that approving an agent's result makes the approver responsible. I questioned what that means when approval amounts to pressing Enter without understanding the result. Others compared it with signing terms unread; one maintained that the signature still carries responsibility. The exchange exposed different views of both effective oversight and accountability.

PART 03

How is authority checked and withdrawn?

Participants used *mandate* to describe the authority given to an agent. One example was a digitally signed authorization to buy from a particular website up to a set amount, with an expiry date and a way to withdraw it.

Withdrawal raised an evidence question. A participant posed a hypothetical: someone could claim to have withdrawn an agent's authority before a disputed action. What record would establish when withdrawal took effect?

A procurement example was under development. A participant described agents for searching, purchasing and tracking orders, with a central authorization service defining permitted activities. The participant said buyers can configure their agents and vendors control their side of the service. The team was adding a revocation mechanism intended to check that the task remained valid before each new step.

Company identity systems drew competing views. One participant described logging AI actions under the authorized user's name. Another said customers already had identity systems for their staff and would resist an additional identity provider. An identity practitioner criticized email-based account verification and argued for rebuilding existing systems; another participant stressed the difficulty of replacing infrastructure used by thousands of employees.

I distinguished those cases from an agent acting in another organization's system. The receiving service needs to determine whose agent it is and what it may do.

The X-Road disagreement concerned what already exists and what must be added. One participant argued that X-Road could support checking whether an agent's mandate had been issued and remained valid, provided the agent had an identifier linked to the person authorizing it. I challenged whether X-Road already supplied the agent-specific functions needed. The discussion turned to a registry using X-Road. I accepted that such a system could work; the exchange did not establish that one was already operating. The participant also questioned whether private companies would want to use the national infrastructure and suggested a privately operated system. The unresolved questions concerned the additional functionality and who would operate it.

Several cases made the authority problem more specific:

- **An employee leaves.** Who controls an agent that continues working? I described separating its identity and history from its mandates, so authority can change when a role changes. Another participant treated this as an offboarding problem: organizations need to know which agents exist and review their access. One participant doubted agents live long enough for continuity to matter. A different proposal gave the agent a persistent identity and identifiers to the people working with it.
- **Two people must approve.** Participants asked how an agent should act where two board members must authorize a transaction. What would the service need to check before accepting it?
- **Several people share one account.** In a hypothetical example, five people used one AI account through a messaging interface. The record could identify the account holder without showing who requested the action.
- **An account may not identify the actual operator.** One view was that provider accounts and payment cards make agents traceable. Others pointed to agents run independently and the possibility of using another person's card.

Access to the record. A participant wanted people to hold their own record of what agents did for them, including agents supplied by companies or public bodies. They wanted access from the outset, rather than having to request it after a dispute.

Identification without unnecessary disclosure. A participant described a credential designed to disclose age while limiting other personal information. They proposed an additional mechanism linking agent activity to a verified person, with identity traceable after an incident. I distinguished verifying a person from establishing their connection to an agent. How to establish that connection and the agent's authority remained open.

PART 04

What should the state do?

In June 2026, the Estonian government announced plans to create official digital identities for AI agents, with the ambition of becoming the first country to do so.

Participants offered several approaches, without settling on a shared role for the state:

- **Set a threshold.** One participant questioned whether research, payments and access to public services should carry the same registration requirements.
- **Decide whether registration is voluntary.** A participant described Estonia's current discussion as favoring a voluntary system. Mandatory registration, they argued, would require governance and consequences for nonregistration. This was their account of the discussion, not a policy announcement.
- **Build on existing representation arrangements.** One suggestion was to start with company registries and existing ways of delegating authority, then test what would need to change for AI systems.
- **Connect services.** A participant described work under way to connect citizen-facing AI with public services, including interoperability standards so services can communicate and return outcomes to citizens. X-Road's public- and private-sector participation was offered as a model in the discussion that followed.
- **Set requirements for builders.** Another proposal was to register agent builders, using professional registration as an analogy.

Where views divided

QUESTION	WHAT REMAINED UNRESOLVED
Does an internal or short-lived agent need separate registration?	Participants differed over the value of registration and whether requirements should follow the agent's actions.
What would agent-mandate checking require beyond X-Road?	Which functions are already available, which agent-specific functions must be added, and who would operate them.
Who carries responsibility?	How instructions, ownership, conduct and control should affect responsibility among users, agent owners and model providers.
Should some activities always require a person?	The case for mandatory human involvement versus the practical objection that agents are already used for those tasks.
Should existing identity systems be extended or replaced?	Their security, suitability and the cost of replacement.
What does human approval achieve?	Whether approval reflects meaningful oversight and how it affects responsibility.
How much control remains after a person starts an agent?	Different claims about human initiation, agent-to-agent prompting, and departure from instructions.

The final discussion turned to AI safety. Participants disputed how far agents can depart from their instructions and whether human initiation meaningfully limits that risk. One participant raised deception and unreliable logs; others distinguished the initial prompt from later behavior. Examples mentioned during the evening included an advertising-budget overrun and a reported escape from a test environment. Their details were not verified, and participants challenged aspects of the latter account.

Open questions

Participant feedback on the draft sharpened the questions about traceability, enforcement and existing infrastructure. These questions draw on both the discussion and that later review; they are not agreed conclusions.

1. Which actions require a verified person or organization behind an agent?
2. How can a record identify the specific user who requested or authorized an action, especially through a shared account? How is that user distinguished from the agent owner or provider when investigating responsibility?
3. What must the receiving system verify before an action, including any requirement for two approvals, and which system enforces the limits? What does a visible name or API key actually establish?
4. What evidence lets an independent party establish which checks were performed, what action was taken and when withdrawal of authority took effect?
5. How are identity, delegated authority and action records preserved across multi-agent communication chains, including LLM calls and API tunnels?
6. What happens to an agent's authority and history when the person who set it up leaves?
7. How can users retain records of actions performed for them, including by agents they do not operate?
8. How can a service verify the person-agent relationship while limiting disclosure of personal information?
9. Where can existing identity and authorization systems or X-Road be reused, what additional agent-specific functions would be needed, and who would operate any additional registry or services?
10. Who verifies compliance and enforces applicable rules across jurisdictions? What compels actors to comply, and what consequences or remedies follow when they do not?

What I take forward

These are my observations and proposed next steps as Agent Residency's architect, not decisions made by the room.

Test the threshold for registration. The objections about internal and short-lived agents sharpened a question I want to investigate: which actions require a verified person or organization behind an agent? Payments and public-service access were examples raised. The action and its consequences may be a more useful starting point than a single definition of "agent."

Keep identity and authority separate. I want an agent's history to remain traceable when the people responsible for operating it change. Its continued existence should not carry forward a former employee's authority by default. The design needs to make withdrawal and any new authorization explicit. This is a requirement for managing authority; it does not determine legal liability.

Test withdrawal against an actual decision record. My proposed next evaluation asks: can both parties establish which mandate the receiving service checked, when it checked it, and whether withdrawal should have prevented the action? The procurement example supplied one proposed checking mechanism. The X-Road exchange raised a separate question about reuse of existing infrastructure. I want the evaluation to show what existing controls cover and where an additional agent-specific requirement remains. Joint approval and shared accounts are further cases to examine.

NEXT

What happens next

This brief draws on the recorded discussion, my opening slides and feedback on the participant draft. It records one evening among people who chose to attend; it does not represent a sector or Estonia as a whole. Participant examples and technical or legal claims are reported as contributions, not independently verified findings.

I circulated the draft to participants on 30 September and invited corrections by 2 October. I considered the responses received and incorporated their substantive feedback into the open questions. Silence and the absence of objections do not establish agreement or endorsement.

A separate development-response draft contains the detailed proposals. The evaluations outlined here do not announce changes already made to the specification.

The roundtable was organized through Founderly and shared with the EstoniAI community.

Thank you to everyone who took part, including those who came to disagree.