

Digital@UNGA 2026 Session Outcomes Report

Verifiable and Accountable AI: Trust Infrastructure for an Inclusive Agentic Era

26 September 2026

The session developed a practical trust framework for AI agents built around verifiable identity, bounded delegated authority, revocation, evidence-grade records and independent verification. This report records the discussion outcomes, the limits of confirmed quotation, participation data and publication information. No new formal partnerships or commitments were announced during the session.

Official title	Verifiable and Accountable AI: Trust Infrastructure for an Inclusive Agentic Era
Date and time	22 September 2026, 10:00–11:00 EDT / 16:00–17:00 CEST
Format	Online moderated panel discussion in English
Organizer	Management Company Elixir LLC, Ukraine
Moderator	Oleh Vasylenko, Director, Management Company Elixir LLC; Associate Member, ITU-T Study Group 17
Status	Independently organized Affiliate Session of the ITU–UNDP Digital@UNGA 2026 programme

Participation and geographic reach

Metric	Value	Note
Registration submissions	48	Includes repeat submissions
Unique registrants	43	Registration data
Countries in registration data	21	After normalization of country names
Participants at peak	12	Includes the moderator
Peak concurrent connections	13	Includes one automated notetaking assistant
Panelists	4	Luukas Ilves, Elena Sirbu, Gilles Fayad and Juliana Cafik

The registration data covered Bangladesh, Canada, China, Egypt, Estonia, France, India, Ireland, Japan, Mongolia, Nigeria, Peru, the Republic of Korea, Singapore, Switzerland, Tanzania, Thailand, Ukraine, the United Arab Emirates, the United Kingdom and the United States. Registrants represented government and regulators, the private sector, civil society, academia and research, youth, international organizations, the technical community, media, philanthropy and consulting. Registration figures do not imply that every registrant attended live.

Outcome summary

The session identified a practical trust framework for AI agents: verifiable identity, bounded delegated authority, revocation, evidence-grade records and independent verification. Panelists stressed that investigability must be designed before incidents occur by preserving provenance, context, configuration, permissions and external-system records. Cross-platform use requires portable trust anchors, open standards and interfaces so that smaller institutions and states can verify agents without dependence on closed ecosystems. Human accountability should be established through governance, monitoring, risk controls and escalation. No new formal partnerships or commitments were announced. The organizer is publishing the recording and report as open references for standards, testing and pilot discussions.

Key outcomes

- Trust in an agent cannot rest on a name or account alone. It requires a verifiable link between the agent, the person or organization authorizing it, the limits of that authority and the specific action.
- Delegated authority should be bounded by purpose, duration, transaction limits and the right to delegate further. Systems should support revocation and verify authority at the moment of action.
- Investigability must be designed before an incident. Reconstruction is possible only when systems preserve context, configurations, permissions, versions, timestamps, tool interactions and independent records from affected external systems.
- A log can serve as evidence, but its evidentiary weight depends on provenance, integrity, control of the system and independent corroboration. An agent's own report that it performed an action is not sufficient.
- Cross-platform verification requires portable trust anchors, self-contained signed artifacts, verifiable validity and revocation status, and open interfaces.
- Inclusion has an infrastructure dimension. If identity and verification are available only inside large closed platforms, smaller institutions and states become dependent on a provider for access to the agentic economy.
- Human accountability should be established through system governance, monitoring, risk controls and escalation, rather than nominal human approval of every individual decision.

Panelist contributions

Direct written quotations are separated from editorial summaries. Editorial summaries are based on the public recording and on text sent to the panelists for review on 22 September 2026.

Luukas Ilves

- Public institutions deploying their own agents have a heightened duty of care concerning data governance, the technical stack and security models. They must also prepare for external agents acting on behalf of citizens and businesses.
- Responsible interaction with an external agent requires a clear link to the person or organization it represents, together with defined limits on duration, transactions and delegated authority. Controls must account for incomplete knowledge of the external agent's model and security.
- The authority an institution needs to understand depends on the service. Read-only access creates different risks from a complex transactional service. The institution needs to know the applicable policy, permitted access rights and whether the policy was enforced at the relevant moment.
- Small institutions and states should not have to solve the full complexity of cross-system verification alone. Shared platform functions, national infrastructure and globally standardized protocols can absorb part of that burden.
- Meaningful human responsibility should be designed at the level of systems, governance, monitoring and escalation, rather than reduced to nominal human approval of each decision.

- Consumer organizations can identify systemic patterns of harm, but they need transparency and analyzable records. Governments can demonstrate stronger transparency and oversight practices that may later influence the private sector and regulation.

Over the next year, institutions should make one socially significant public service safe for external agents by authenticating, verifying, constraining and, when necessary, revoking delegated authority, while ensuring that actions can be reconstructed before the approach is scaled.

Organizer's editorial formulation based on the recorded closing remarks; not a verbatim quotation. No corrections were received by 25 September.

Elena Sirbu

- An investigation can reconstruct only what systems preserved. Relevant evidence may include user inputs, system instructions, context, agent outputs and actions, tool interactions, configurations, permissions, versions, timestamps and identifiers.
- An agent's claim that it sent an email, changed a database or performed another external action is not sufficient evidence. The agent trace should be compared with records from the email provider, database, API, tool provider or other affected system.
- There is no absolute boundary between a log and real evidence because a log may be evidence. Its evidentiary weight depends on provenance, integrity and corroboration: who created it, who controlled the system, whether it could be altered or destroyed, and whether independent corroborating records exist.
- Fraud and online-advertising cases require investigators to distinguish intentional misuse of an agent from misaligned agent behaviour. The evidence categories remain the same, but the incident may cross organizational boundaries and require external parties or law enforcement.

Organizations should institutionalize investigability by building the infrastructure for traceability and incident analysis in advance, ensuring that sufficient relevant records exist to reconstruct and assess AI incidents once they occur.

Direct written formulation sent by Elena Sirbu on 22 September 2026.

Gilles Fayad

- When one agent delegates to another across platforms, the chain should preserve the identity of the ultimately responsible person or organization, each agent's authority, the authenticity of the delegating agent and an auditable record of successive transfers.
- Current authorization mechanisms are designed mainly for a user granting access to one application, not for multi-step agent-to-agent delegation. Portable verifiable credentials or equivalent artifacts are needed to express the scope of authority and the right to delegate further.
- Portable agent identity is insufficient without bounded authority. Authorization should be checked at runtime and recursively across the full nested delegation chain.
- Delegation should be a structured, machine-readable and verifiable object recording who delegated, what was delegated, for how long, whether redelegation is allowed and how authority can be revoked. Revocation remains one of the hardest unresolved questions.
- Consumer trust requires mandatory auditability, a clear accountability rule linking autonomous action to a responsible person, and open interoperable identity infrastructure.
- Agent permissions, delegation history and revocation status should remain traceable across platforms so that a receiving system can identify and authorize the agent in context.

Institutions should begin building accountability into agent delegation chains now so that identity, authority, revocation and traceability move with agents across platforms before the scale of closed systems makes interoperable accountability difficult to add retrospectively.

Organizer's editorial formulation based on the recorded closing remarks; not a verbatim quotation. No corrections were received by 25 September.

Juliana Cafik

The following four passages are Juliana Cafik's written formulations of 24 September 2026 and appear in the main report in her own words. The two remaining points are identified as editorial summaries.

Current remote identity systems infer identity once, at onboarding, and lock it into a static credential or token that is treated as valid indefinitely. This construct fails against AI-driven attack because it collapses three separate facts whether an agent is authorized, whether it's the same agent, and whether a human is behind it — into one signal never designed to answer all three. Trust instead requires an issuer-proofed, verifiable credential carrying signed identity assurance claims, bound to a specific person through kinetic provenance a live, hardware-anchored proof of presence applied both at binding and refreshed at the moment of each subsequent action.

This binding is necessary as the trust anchor from which agentic identity must proceed, because every subsequent layer of an agentic system inherits whatever foundation this one provides. If that foundational binding is inferred rather than proven, every layer above it is procedurally sound while remaining substantively unproven.

Written formulation by Juliana Cafik, 24 September 2026.

Establishing that an interacting agent is the same agent, rather than merely using the same name or account, requires both technical-instance continuity and identity-assurance continuity: a cryptographic key bound to a specific runtime with proof of possession, together with identity-assurance continuity, confirmation that the identity assurance behind the agent remains current, closed specifically through kinetic provenance: a live, ephemeral, silicon-anchored proof that the same person originally bound to the credential remains the one controlling it now.

Written formulation by Juliana Cafik, 24 September 2026.

- Editorial summary by the organizer: The distinction between ordinary logging and trustworthy evidence is the distinction between assertion and attestation. Evidence becomes stronger when independent parties sign different parts of the action lifecycle, including runtime actions, human presence, credential issuance and the relying party's resulting action or refusal.

Cross-platform verification requires portable trust anchors, self-contained signed artifacts, transaction-specific context binding, publicly checkable validity and revocation status, and mechanisms capable of demonstrating human presence or authorization at the relevant moment, though the mechanism for demonstrating live human presence remains the least mature of the five, with no existing standard defining it yet.

Written formulation by Juliana Cafik, 24 September 2026.

If platforms control both the identity credential and the verification mechanisms for agentic identity, they hold the infrastructure that decides who gets recognized as a trustworthy actor, in an economy increasingly run by agents transacting with other agents. That is a form of de facto sovereignty over a basic condition of participation. Left unaddressed, this does not distribute evenly. Under-resourced states, institutions, and populations already furthest from formal identity systems become highly dependent on the continued goodwill of platforms. An organization is included when it can prove who it is and verify who it's dealing with without going through a large platform to do so. If the only way to participate is by becoming a platform's customer, licensee, or user, that is not inclusion - It is dependency.

Written formulation by Juliana Cafik, 24 September 2026.

- Editorial summary by the organizer: Trust solutions should be assessed through open standards, interoperability testing, proof-of-concept validation, relying-party adoption, regulatory alignment and, where appropriate, certification or trust-mark programmes.

Every agentic action needs undeniable proof, freely verifiable, by any trusted party anywhere, of whether a real person is bound to, and knowingly authorized a specific action.

Direct written formulation by Juliana Cafik, 24 September 2026.

Publication and contact

- The session recording and this report are published on the event page: <https://vasylenko.tel/events/digital-unga-2026>
- Written post-session additions received from panelists are incorporated and identified by source and date.
- For collaboration or follow-up, contact Oleh Vasylenko at oleg@vasylenko.tel.

Attribution method

- A direct quotation is used only when the panelist wrote or explicitly confirmed the text.
- An editorial summary conveys the substance of the public recording but is not presented as a verbatim quotation.
- The Luukas Ilves and Gilles Fayad closing formulations were prepared by the organizer from their recorded closing remarks. They are explicitly labelled and are not attributed as verbatim quotations.
- Elena Sirbu's 22 September written formulation is quoted as her own written statement. A response she did not provide to the later follow-up question is not reconstructed or attributed to her.
- Juliana Cafik's four written clarifications and closing sentence appear in the main text as her own wording. The two points she did not revise remain editorial summaries.
- Automated transcripts were used for navigation. No new verbatim quotation was added from them without checking the audio recording.

Sources

Public session recording <https://www.youtube.com/watch?v=WaUOURsyOUY>

Event page <https://vasylenko.tel/events/digital-unga-2026>

- Automated recording transcript and session text file, used to navigate and verify discussion themes.
- Organizer correspondence with the four panelists from 22 to 25 September 2026.
- Registration form and linked response sheet: 43 unique registrants and 21 countries.
- Eight Google Meet screenshots taken during the session.
- Digital@UNGA 2026 Affiliate Session criteria and Affiliate Session Outcomes Form.

Status disclaimer

This Affiliate Session was independently organized by Management Company Elixir LLC as part of the Digital@UNGA 2026 programme. Inclusion in the programme does not constitute endorsement of its content. The views summarized in this report belong to the organizer and participants and do not necessarily reflect the positions of ITU, UNDP or the United Nations.