ITU | AI for Good | MeetLoyd
The platform for the agentic era

ITU-T SG17 WORKSHOP · AI FOR GOOD · GENEVA · 9 JULY 2026

Workshop Designing the Trust Management for Agentic AI

Session 1 Challenges for Trust Management in Agentic AI

# Introducing the Agentic AI Control Plane

**Trust is a property of the control plane - not the model.**


SESSION 1 · MODERATOR & SPEAKER

**Laurent BEAUMOIS**

Founder, MeetLoyd **laurent@meetloyd.com**

# Trust is not a property of the model - it's a property of the control plane.

*"…an entire trust framework, including an **agentic AI trust control plane** to keep humans in the loop." - Arnaud Taddei, Chair, ITU-T SG17 (Jan 2026)*

**MANAGEMENT PLANE · AUTHORSHIP** Humans author intent & stay accountable — **HUMAN**

**CONTROL PLANE · ENFORCEMENT** — **DETERMINISTIC - TRUST LIVES HERE**

Binds every action to the declared intent - the standardizable surface

1

### Identity & delegation

Name the actor - and the chain it acts on behalf of.

2

### Authority

Bound what it may do - least-privilege, dynamic, not a static allowlist.

3

### Attribution & audit

Across a multi-agent workflow: who did what, with what data, why.

4

### Autonomy & containment

How far it may act without a human - earned and revoked, always inside the envelope.

**DATA PLANE · ACTION** Agent reasoning · tools · memory - can't be trusted directly — **NON-DETERMINISTIC**

Every challenge this panel names is a gap in the control plane. **Trust lives in the deterministic layer.**

# Autonomy is a dial, not a switch - and it must be earned.

## THE AUTONOMY DIAL · the human's role - hands come off the wheel, one step at a time

| LEVEL 1 | LEVEL 2 | LEVEL 3 | LEVEL 4 |
|---|---|---|---|
| **Human leads** | **Human approves** | **Human oversees** | **Human governs** |
| *human works, AI assists* | *AI proposes, human signs off* | *AI acts, human can step in* | *AI runs, human sets the rules* |

*#1 - Even at full autonomy the human governs the envelope (identity · authority · audit · verification) and **stays accountable**.*
*#2 - 'Ungoverned autonomy' is the failure mode the control plane exists to prevent.*
*#3 - **Governance at agentic speed can't be human**, the human sets intent, PVP verifies every actions at agentic speed.*

## PROGRESSIVE AUTONOMY · the closed loop that moves the dial - earned and revoked continuously

PATENT-PENDING

### Coherence

Is it acting within its declared role, tools and intent? Drift is flagged.

### Loyd Solver

Is each plan policy-legal before it runs? Checked pre-execution, with a cryptographic certificate.

### PVP verification

Sample actions like radar on a highway - a mathematically guaranteed catch-rate, proving at agentic speed.

### → The dial moves

Verdicts grow or shrink the agent's autonomy - auto-approved actions and spend limits. Revoked on drift; kill-switch always.

- Running today under this loop: our self-healing Support and Agentic SOC detect and remediate autonomously - inside the verified envelope.

# Panel Questions

1. *What is genuinely new about **trust in multi-agent systems** (agent-to-agent delegation, cross-org trust, multi-objective and emergent behaviour) that single-model assurance doesn't address?*

2. *How should we define **trust vs. trustworthiness** for agentic AI - and is trust a binary you grant once, or a graduated, measured, revocable property an agent earns?*

3. *How should a **threat taxonomy** be structured to cover both known and emergent, agent-specific threats end-to-end - including realistic attacker capabilities like indirect prompt injection, memory poisoning, and tool misuse?*

4. *As agents move **from human-approved to autonomous**, how do we keep humans accountable - and what oversight and audit trail does that demand?*