

ITU-T SG17 Workshop | Chongqing, 7 September 2026

# Confidential Computing for Agentic AI

What we can hand over — and what attestation cannot prove

---

**Shiqi Li**

Lead of Security Standardization, Alibaba Group | Editor, ITU-T X.LLMCC

# Four layers, one agent

Where my layer sits relative to the other three panellists.

## Model & planning

Goal integrity, long-horizon tasks

## Identity

What agent ID protocols settle — and do not

## Infrastructure & execution

Verifiable runtime, data protected in use

## Device

On-device trust control plane

**The other three layers all assume the runtime enforcing them is intact.**

Attestation is what makes that assumption checkable.

# Autonomy moves the secrets into the running process

An agent that answers leaks content. An agent that acts holds authority.

01

## Live credentials

Delegated identity and keys,  
whole session

02

## Persistent memory

Context outlives the request;  
poisoning sticks

03

## Real actions

Code, cloud operations,  
payments

04

## Wider supply chain

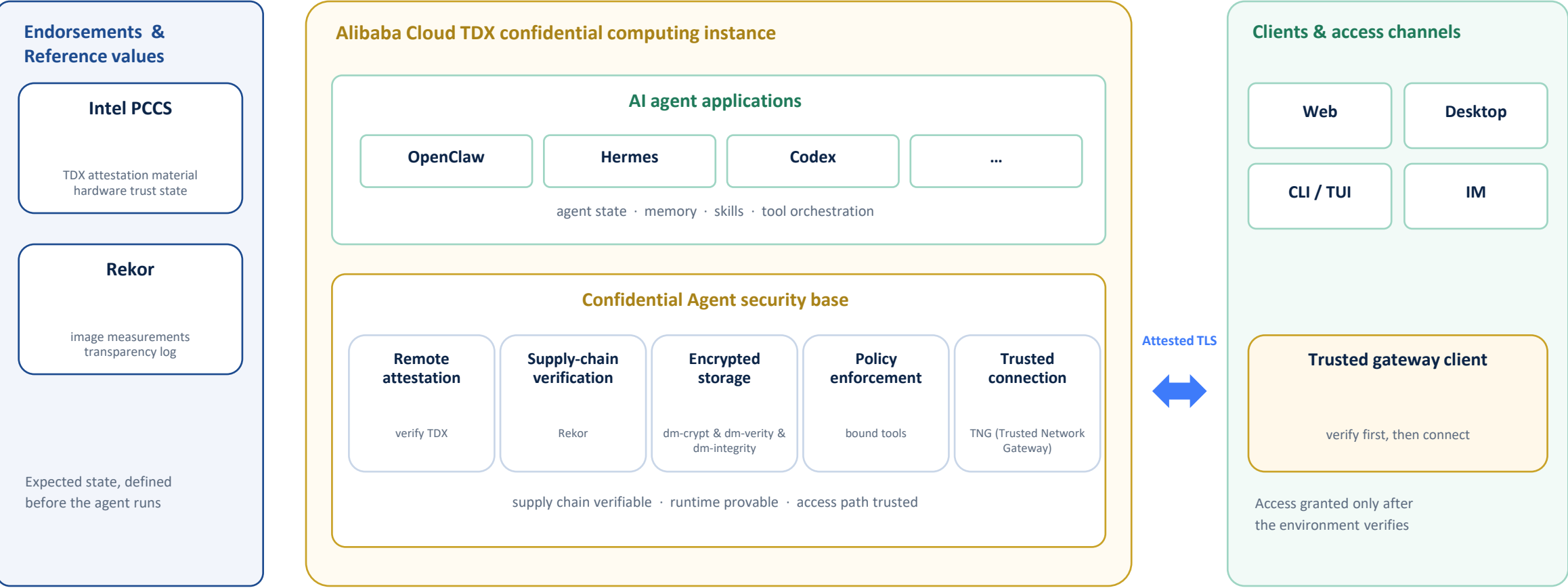
Plug-ins, MCP, peer agents

**At rest: encrypted. In transit: TLS. In use: exposed.**

Credentials, memory and tool context sit decrypted in memory — visible to whoever controls the host, hypervisor or management plane.

# Confidential Agent — trusted runtime architecture

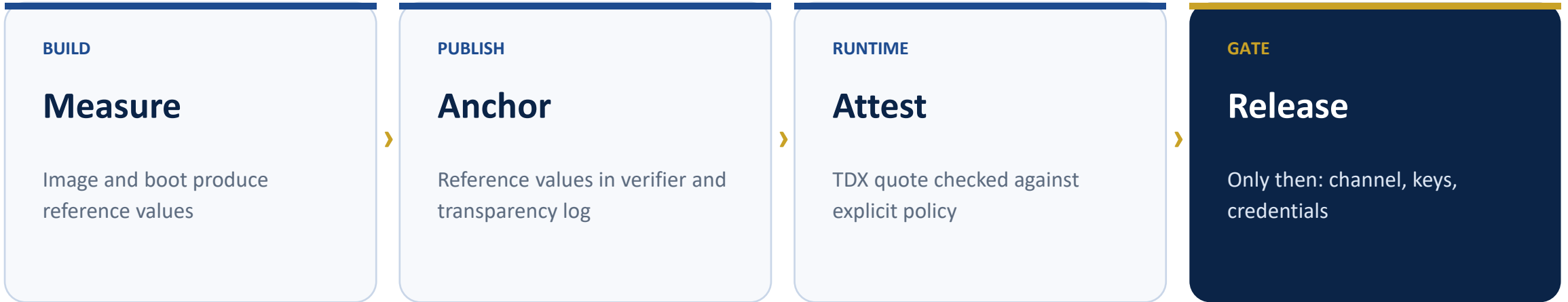
Real agents (OpenClaw, Hermes, ...) on Alibaba Cloud TDX instances; trust chain built on remote attestation and attested TLS.



Adapted from the internal Confidential Agent architecture (Alibaba Cloud × Intel).

# Verify first, then connect and release

The gate is what makes the architecture more than isolation.



**No verified environment, no key.**

Keys, models and credentials are released only on a passing attestation.

**Fail-closed.**

A failed or stale verification ends in no session, not a weaker one.

# Ready to reuse: draft ITU-T X.LLMCC

LLM data security based on confidential computing — Q16/17

## Core content

- Reference architecture: stakeholders, core components
- Security threats — in training and inference
- Controls from policy to attested execution to release
- Appendix: risk → control → component mapping

## What you can take

- The mapping table works as an integrator checklist
- Attestation policy gives procurement a testable question
- Built on RATS and TEEP — composes with identity work
- Vendor-neutral: written against TEE properties

# What we ask of the ecosystem

Three problems no single vendor can close.

**01**

## **A minimum attestation evidence profile for agent runtimes**

What an integrator is entitled to demand from an agent, tool or model supplier.

**02**

## **Attestation policy that is portable across TEE vendors**

Without it, cross-organization agent collaboration falls back to trust by contract.

**03**

## **Bind attestation to agent identity and authorization**

Environment evidence and delegated-authority evidence are standardized in separate rooms.