IAPP **AI Governance Global Europe 2026**

Training 1-2 June
Workshop 2 June
Conference 3-4 June
DUBLIN

#IAPPAIGG26

HUMAN OUT OF THE LOOP?

Practical Governance Strategies for the Agentic AI Era

AGENDA OUTLINE

i. Welcome and Introductions
ii. How did we get here? - brief summary of AI governance in ML and Gen AI age
iii. How agentic AI differs from its predecessors?
iv. Emerging governance techniques - lessons in flying the plane while we're still building it
v. How agents and human collaborate
vi. Agentic AI governance maturity model
vii. Agents and governance frameworks: EU AI Act and Singapore deep-dives
viii. Questions and Answers
ix. Additional resources

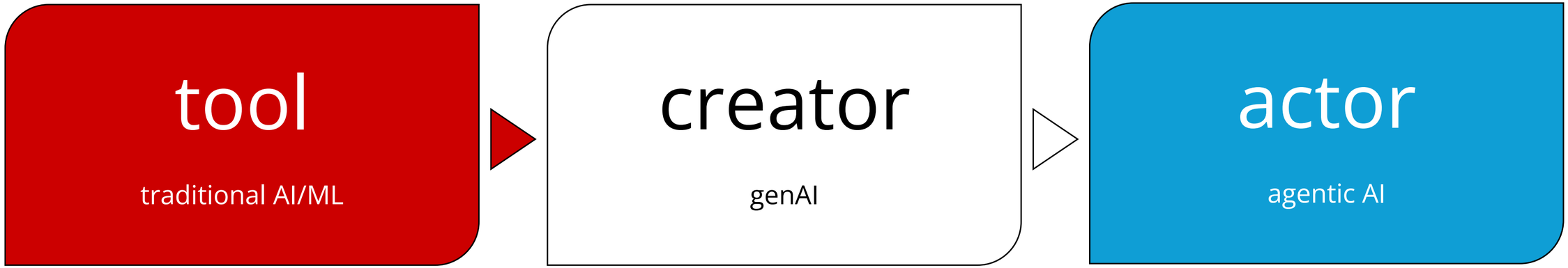
i. WELCOME AND INTRODUCTIONS

Elena Maran - founder of Alethesis AI & member of CEN/CENELEC JTC 21 WG4

Oliver Patel - AstraZeneca
Head of Enterprise AI Governance

Martin Woodward - Randstad
Director Global Legal
Global Responsible AI Officer

II. HOW DID WE GET HERE?



WHAT IS AN AI AGENT?

Audience question

III. HOW AGENTIC AI DIFFERS FROM ITS PREDECESSORS

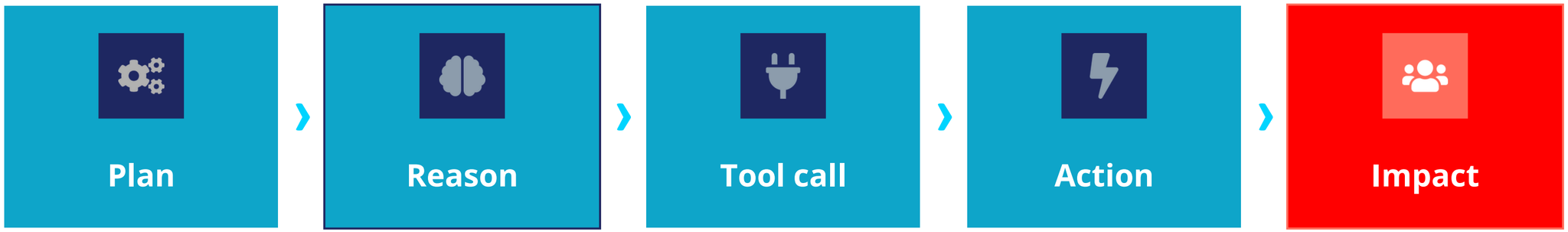
FROM ASSISTANCE TO ORCHESTRATION

STATIC AI · yesterday

Human prompt › Model output › Human acts

Human stays in control of every action. Risk = output quality.

AGENTIC AI · today



Agent orchestrates the process. Each step affects external systems, customers, regulators. Risk = every action.

III. DIFFERENT SHADES OF AGENTIC AI

DEFINITION

An AI system that, given a goal and a set of tools, **independently plans and executes multi-step actions** — choosing which tools to use, in what sequence, and sometimes coordinating with other agents to complete the task.

← human judgment ..... system autonomy →

COPILOT

WHO DECIDES
AI suggests · human approves each step

EXAMPLES
Code completion, email drafts, inline editing

AUTONOMY

COWORK

WHO DECIDES
Agent executes tasks alongside the human

EXAMPLES
Desktop task automation, file management, in-app actions

AUTONOMY

WORKFLOW

WHO DECIDES
AI executes structured flow · human at gates

EXAMPLES
RAG pipelines, scheduled automations, RPA + LLM

AUTONOMY

AGENTIC AI

WHO DECIDES
AI plans + chooses tools · sometimes multi-agent

EXAMPLES
Autonomous research, IT ops, customer ops

AUTONOMY

As autonomy increases, control shifts from human judgment to system architecture. *The right governance changes shape with it.*

III. WHERE THE RISKS LIVE

ATTACK SURFACE

What an adversary can exploit

Prompt injection

Malicious instructions in user input

Indirect injection

Malicious content reaching the agent via tool outputs

Tool / credential theft

Exfiltration of tokens, scopes or data via tool calls

Memory poisoning

Adversarial content persisted into long-term memory

Supply-chain compromise

Upstream model, tool or data source tampering

ARCHITECTURAL RISKS

What the design itself creates even without malicious intent

Privilege escalation

Agents chain tool calls beyond their intended scope

Behavioural drift

Memory and emergent behaviour shift the agent silently

Accountability gaps

Who decided this, on what basis, with what context?

Supply-chain opacity

Multi-provider stacks: foundation model + framework + tools + data

Lack of runtime evidence

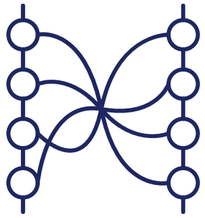
Controls designed for deployment-time, not for the moment of action

IV. 4 LAYERS OF ENTERPRISE AGENTIC AI GOVERNANCE

Layer 1. Democratised agent development

All employees can now develop and deploy agents with low / no code tools. Each organisation must choose how permissive they want to be.

Layer 2. Agentic AI engineering

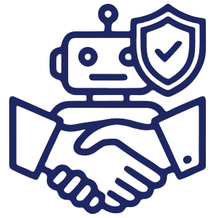


Engineering and business teams are increasingly developing and deploying agentic AI systems that serve a range of use cases.

Layer 3. Agentic AI system procurement

Engineering and business teams are increasingly procuring agentic AI systems from vendors, in support of a wide range of use cases.

Layer 4. In-platform agent enablement



Tech providers increasingly provide AI agents (sometimes hundreds) that can be enabled and deployed within existing software platforms.

IV. 3 Gs FOR GOVERNING DEMOCRATISED AI

AI governance should enable and empower teams across the organisation to innovate with confidence and take smart risks, knowing that robust guardrails and standards are in place.

The 3 Gs support enterprises in addressing Layer 1.

(HOW) HAVE YOU OPERATIONALISED HUMAN OVERSIGHT FOR AGENTS?

Audience question

V. HOW AGENTS AND HUMANS COLLABORATE

FULL CONTROL ← ——— *tradeoff: control vs. velocity* ——— → **FULL AUTOMATION**

TOO MUCH CONTROL

HUMAN IN THE LOOP

Every agent action gates on a human reviewer.

WHY IT FAILS

- Slow, expensive, doesn't scale
- Cannot match agent decision velocity
- Operators bypass under pressure

THE ANSWER

RISK-BASED OVERSIGHT

The action is the governance unit.
Consequential decisions are governance events.

WHAT IT DOES

- Routine actions auto-execute with logging
- Material actions trigger an approval or escalation
- Critical actions are blocked pending governance review
- Efficiency where safe · judgment where it counts

TOO LITTLE CONTROL

INVISIBLE HUMAN

Agent acts; no human is positioned to intervene.

WHY IT FAILS

- No qualified person consulted
- No record of who authorised
- No accountability when it goes wrong

RUNTIME AUTHORITY defeats the invisible-human problem — it makes oversight a property of the ***system,*** not the operator.

V. OVERSIGHT LEVELS & ACTION ONTOLOGY

DEGREES OF OVERSIGHT

Graduated to the score, not blanket

Auto-execute
Logging only

Monitor
Alert to oversight queue

Human approval
Routed to designated approver

Senior escalation
Dual sign-off required

Hard block
Governance-board review

ACTION ONTOLOGY

Classify before you route

01 DOMAIN
Finance · HR · IT ops · clinical · public-facing

02 PROCESS
Account opening · termination · production change

03 DECISION
KYC override · termination · prod deploy · refund > €10k

04 INSTANCE
Specific case attributes · amount · counterparty · reversibility

The action is the unit of governance. The ontology decides which actions matter. The level decides how much human judgment they get. ***Routine auto-executes; the consequential reaches the human qualified to decide.***

The 6 Degress of Human Oversight

Human-in-the-Loop (full)

Complete pre-approval of all AI actions.

The human directly reviews and approves (or rejects) all AI agent actions, before they are executed. Autonomy is highly limited.

Human-in-the-Loop (conditional)

Threshold-based approval of specific AI actions.

The AI agent independently handles various routine actions, but permission is required when specific thresholds are met.

Human-on-the-Loop

Active monitoring of autonomous AI actions.

The AI agent works autonomously, but humans actively monitor and review performance and intervene if necessary.

Human-in-Command

Autonomous AI actions within rule-based boundaries.

The AI agent works autonomously, and humans periodically review its performance and intervene retroactively if necessary.

Human-on-Standby

Human involvement by exception only.

The AI agent works autonomously without active human monitoring or review, and proactively initiates contact when it requires human support or input.

Human-Override-Only

Human oversight amounts to kill switch control.

The AI agent operates with full autonomy, but humans have the ability to deactivate it in response to an emergency or serious incident.

Agentic AI Governance Maturity Model

Evaluate your organisation's progress in governing agentic AI at scale

5

Transformative

Agentic AI governance is an enabler and accelerator of AI adoption. The framework scales with new agents and keeps pace with emerging capabilities, allowing the organisation to safely operate at the frontier.

4

Embedded

Agentic AI governance controls are applied by default across all four layers. Controls operate routinely and are increasingly automated, with agent inventories, autonomy constraints, monitoring, and human oversight mechanisms.

3

Standardised

Core governance controls have been updated and codified for agentic AI. Policies, processes, vendor due diligence, and technical guardrails explicitly address agentic AI risks.

2

Emerging

The organisation recognises that agents require distinct controls and has begun adapting its framework. However, coverage is partial / uneven and not yet operational at scale.

1

Nascent

Agentic AI is not addressed in the enterprise AI governance framework. Existing policies, risk assessments, processes, and guardrails were developed for traditional and generative AI.

WHERE IS YOUR ORGANISATION ON THIS MODEL?

Audience question

VII. AGENTS AND THE EU AI ACT - WHERE IT WORKS

AI agent
=
AI system

(machine-based, levels of autonomy, adaptiveness after deployment, infers how to generate outputs that influence physical or virtual environments)

AI Act risk classification applies

(dependant on intended purpose)

autonomy amplifies risks

(e.g. cybersecurity, human oversight)

agent's brain
=
GPAI model

(provider obligations under Chapter V AI Act)

VII. AGENTS AND THE EU AI ACT - WHERE IT HURTS

01 Obligations triggered by context

For an agent, context is dynamic, set at runtime by the conversation, the tools invoked, the data accessed and the downstream systems involved. Determining which obligations apply at any given moment is non-trivial.

02 “Substantial modification” becomes unmeasurable

Behavioural drift driven by memory, prompt injection or adversarial inputs makes it practically impossible to know with precision when a system has been substantially modified. Conformity assessment becomes a moving target.

03 Compliance cannot live only in policy

Documentation and procedural controls are necessary but not sufficient. Each action must be evaluated, routed and recorded at runtime. External layers of control become essential internal model behaviour cannot be guaranteed.

04 Provider/deployer split is harder

Modern agent stacks span multiple providers, foundation model, agent framework, tool ecosystem, data sources. Mapping responsibility cleanly onto this reality requires explicit contractual and architectural choices.

VII. AGENTS AND THE EU AI ACT - SUBSTANTIAL MODIFICATION

ARTICLE 3(23)

"Substantial modification"

A change to an AI system after placing on the market or putting into service which is not foreseen or planned in the initial conformity assessment carried out by the provider and as a result of which compliance is or may be affected.

WHY THIS MATTERS FOR AGENTS

Memory accumulation, dynamic tool discovery and emergent multi-agent strategies all change behaviour after deployment. Without runtime state logging, drift is not detectable making the threshold structurally unmeasurable.

THE DRIFT WINDOW

From conformity-assessed system to a different one and silently

Conclusion. High-risk agentic systems with untraceable behavioural drift cannot currently satisfy the AI Act's essential requirements. Providers cannot demonstrate that human oversight was operationalised as opposed to merely designed during the period of use.

VII. SINGAPORE MODEL AI GOVERNANCE FRAMEWORK FOR AGENTIC AI - 4 Key Dimensions:

01 Assess and bound the risks upront

- Update and tailor AI risk assessments for agentic AI use cases.
- Limit agents' access to tools, systems, and data to the minimum level necessary to fulfil tasks.
- Bound agentic autonomy with codified instructions and technical guardrails.

02 Make humans meaningfully accountable

- Allocate the roles and responsibilities of key decision-makers and teams.
- Require human approval prior to sensitive action execution and for high-stakes decisions.
- Mitigate automation bias risk with digestible approval requests, training on AI limitations and failures, and periodic assurance and auditing.

03 Implement technical controls and processes

- Implement technical controls across the agentic AI lifecycle.
- Pre-deployment testing and evaluation must be adapted to agentic AI.
- Continuous testing and monitoring should be in place for production agents. Log behaviour and define alert triggers.

04 Enable end-user responsibility

- Transparency, including information on user responsibilities, should be provided to users that interact with and use AI agents.
- Education and training should be provided to individuals that work with and leverage AI agents.

VII. SINGAPORE MODEL AI GOVERNANCE FRAMEWORK FOR AGENTIC AI - 4 Levels of Human Involvement:

Level 1.

Agent proposes, human operates

Level 2

Agent and human collaborate

Level 3

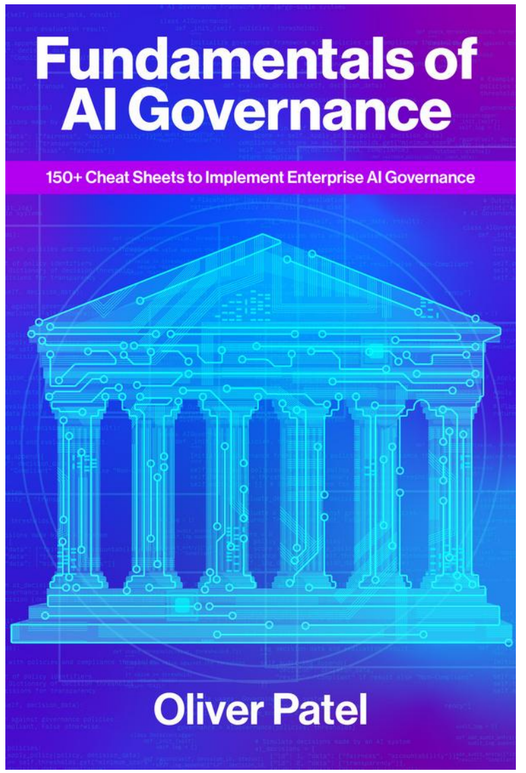
Agent operates, human approves

Level 4

Agent operates, human observes

ADDITIONAL RESOURCES

- **Singapore Model AI Governance Framework for Agentic AI** - Infocomm Media Development Authority (2026)
- **AI Agents under EU Law** - Luca Nannini, Elena Maran et al., (2026)
- **The Ultimate Agentic AI Governance Resource Guide**, features 80+ resources - Enterprise AI Governance newsletter (2026)
- **Fundamentals of AI Governance** - Oliver Patel (published in September 2026)



aigovernancebook.com
September 2026

How Did Things Go?
(We Really Want To Know)

i. Did you enjoy this session? Is there any way we could make it better? Let us know by filling out a speaker evaluation.

1. Open the IAPP Events app.
2. Select **IAPP AIGG Europe 2026**.
3. Tap "Schedule" on the bottom navigation bar.
4. Find this session. Click "Rate this Session" within the description.
5. Once you've answered all three questions, tap "Done".

i. Thank you!