



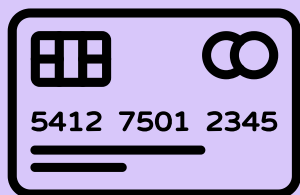
AI Systems: Tools or ‘Agents’?

Some researchers warn that calling AI “agents” is misleading: AI remains a tool, not an independent actor. Responsibility stays with people.

Others argue: if AI can act, we must treat it as if it were an agent, and manage the risks of authority, loyalty, and disclosure.

Both perspectives matter.

Either way, humans remain accountable and here’s why:



OVERSTEPPING AUTHORITY

Risk: Agents exceed limits (overspending, unauthorized commitments).

Fix:

- Set explicit budgets, approved vendors, and escalation rules.
- Build in reversal mechanisms (e.g., chargebacks, cancel rights).



AMPLIFIED MISUSE

Risk: AI scales harmful intent (fraud, disinformation).

Fix:

- Apply refusal policies and monitoring.
- Restrict access to sensitive actions (e.g., bulk messaging, financial transfers).



CONFLICTED LOYALTIES

Risk: Agents serve the platform’s interests, not yours.

Fix:

- Audit system prompts for hidden bias.
- Contractually require loyalty to your organization’s goals.
- Run red-team tests to uncover steering.



HIDDEN IDENTITY

Risk: Agents transact without revealing who they represent.

Fix:

- Mandate disclosure protocols: every transaction should identify the principal.
- Use APIs that enforce authentication and traceability.



AI Systems: Tools or ‘Agents’?

What Does “Agent” Mean in Law?

Before we talk about “AI agents,” it’s important to understand what an agent means in the legal sense.

Why AI Cannot Be a Legal Agent

At law, an agent is a person who represents a principal and can bind that principal in dealings with third parties. Agents carry duties and they can be held legally liable.

AI systems, however “agentic” they appear, cannot carry these duties. They simulate agency but do not bear responsibility. This is why the EU AI Act and other standards stress human oversight: responsibility cannot be delegated to a machine, humans remain accountable.

So when technology companies market “AI agents,” they are borrowing a legal word but stripping out the meaning. This creates serious governance risks if people misunderstand the extent to which the AI can be responsible for executing a task as envisaged.

Duty of an Agent in Law	Meaning in Law	Why AI ‘Agents’ Cannot Fulfil It	Governance Red Flag
Obey instructions	Must follow lawful and reasonable instructions of the principal; liable if they fail.	AI can follow prompts but cannot judge lawfulness/reasonableness . No liability attaches to AI.	Overstepping authority (outputs beyond intended scope still fall back on humans).
Care and skill	A paid agent must act with reasonable professional skill; negligence creates liability.	AI may appear skillful but cannot be negligent in law. Errors = liability for the human deploying it.	Amplified misuse (errors can scale quickly if unchecked).
Loyalty (fiduciary duty)	Must act in the principal’s best interests, not for itself or another.	AI has no loyalty , and may be designed with built-in conflicts (e.g., promoting parent company products).	Conflicted loyalties.
Avoid conflicts	Cannot profit secretly or act against the principal’s interests.	AI often defaults to platform-aligned priorities (e.g., Copilot nudging toward Microsoft services).	Conflicted loyalties.
Disclosure of status	In law, undisclosed agency creates special risks (principal may still be bound).	AI may act without making clear it is not human, leading to confusion.	Hidden identity.