



Bayerischer  
KI-Innovationsbeschleuniger

# Operationalising Trust: Unlocking AI AgentOps for Scalable Adoption and EU AI Act Compliance



Gefördert durch  
Bayerisches Staatsministerium  
für Digitales





# Content

**1.**

## **Introduction**

**2**

1.1 AI agent systems and regulatory oversight

**2**

1.2 Risks and the need for mitigation

**3**

1.3 Motivation and objectives

**4**

1.4 Scope and structure

**5**

**2.**

## **AI Agents: definition, types, and how they work**

**6**

2.1 Definition and capabilities

**6**

2.2 Design types of AI agents

**6**

2.3 Deep dive into each agent design types

**7**

2.4 Defining AI agents under the AI Act

**13**

**3.**

## **AgentOps and design principles for AI agents**

**15**

3.1 Understanding AgentOps

**15**

3.2 Embedding AgentOps in an EU AI Act Programme

**16**

3.3 Designing AI agent systems: balancing architecture and risk

**20**

3.4 Best practices for implementing AgentOps and compliance

**21**

**4.**

## **Discussion about Future Outlook and Impact**

**22**

4.1 Looking ahead (what we can expect) from AI Act and regulation perspective

**22**

4.2 What we can expect from european industry perspective

**23**

4.3 Conclusion

**26**

**Authors & contributors**

**27**

**About the appliedAI Institute for Europe**

**28**

**About the Bavarian AI Innovation Accelerator**

**29**

# 1. Introduction

## 1.1 AI agent systems and regulatory oversight

The field of artificial intelligence (AI) has, in recent years, experienced progress, primarily driven by Deep Learning, Generative AI and the emergence of powerful Foundation Models, including Large Language Models (LLMs) and Diffusion Models. These are pushing the boundaries of AI across various tasks and data modalities.

Within this evolving AI landscape, Agentic AI represents a significant paradigm shift in Generative AI (GenAI). Moving beyond the simple execution of instructions defined in single prompts, Agentic AI embodies systems capable of acting as autonomous, goal-oriented decision-makers. These systems are characterised by autonomous software agents capable of reasoning, planning, making independent decisions, and collaborating with other agents. Its transformative impact is increasingly evident across industries such as healthcare, finance, and manufacturing – enabling task automation, business decision support, and efficiency improvements. Furthermore, the development of low-code and no-code AI agent platforms enables business users to quickly customize and deploy these intelligent systems without programming knowledge, further accelerating their adoption.

However, the developers of AI agent systems need to navigate a regulatory landscape that is still grappling with how to respond to rapid advances in scientific methods and business solutions within Agentic AI. The European Union's AI Act is a key example of emerging regulation in this space. The legislation aims to establish a comprehensive framework to ensure the ethical development and deployment of AI technologies, including AI agent systems. It emphasises the importance of transparency, accountability, and fundamental rights protection in AI applications. A central challenge in the evolving conversation around AI governance is balancing the need to encourage innovation with the necessity of complying with new regulatory standards, while also maintaining public trust in these powerful technologies.

One of the key questions facing the drafters of the European Union's AI Act was how this legislation would hold up in a world where advancements in technology were likely to outpace regulatory initiatives. While the challenges of 'future-proofing' legislation are not new, it was a clear and pressing concern during the development of the AI Act.

At the start, lawmakers were confronted with the challenge of transforming high-level principles on trustworthy AI into concrete risk-based regulatory proposals. Unsurprisingly, this led to many controversies, including how best to formally define AI systems, how to apply product safety standards, which are traditionally designed to protect health and safety, to the protection of human rights, and how to integrate rules for AI into an already complex EU regulatory regime for products and digital services. As if the questions surrounding the AI Act weren't already complex, a major development further complicated matters: the release of ChatGPT in 2022. This event introduced the risk that the AI Act could become outdated even before it was formally enacted. Notably, the original 2021 Commission Proposal made no mention of generative AI. It was only after ChatGPT's debut that European legislative bodies began referencing general-purpose AI (GPAI) models in their discussions.

<sup>1</sup> <https://newsroom.accenture.com/news/2025/accenture-expands-ai-refinery-and-launches-new-industry-agent-solutions-to-accelerate-agentic-ai-adoption>

<sup>2</sup> <https://www.ibm.com/think/insights/ai-risk-management>

market after the Act entered into force. And more importantly for the purpose of this whitepaper, the notion of AI Agents - LLM based applications capable of reasoning, planning and autonomously executing tasks - will continue to stress the definitions, rationale, and purpose of the AI Act.

## 1.2 Risks and the need for mitigation

Although AI Agents hold substantial promise for streamlining processes and enabling autonomous actions, they inherently raise several concerns that demand attention. Challenges such as poor data quality, hallucinations, algorithmic bias, data security vulnerabilities, and the lack of explainability in decision-making processes pose significant barriers to the successful adoption and acceptance of AI technologies. While the deployment of AI agents across industries offers substantial efficiency gains, it simultaneously requires robust risk mitigation strategies. For example, financial institutions using AI agents for credit scoring must proactively address biases embedded in training data. A study by Stanford HAI revealed that predictive accuracy was 5-10% lower for low-income and minority borrowers due to 'thin' credit histories—highlighting the real-world impact of data-driven disparities. To combat this, organisations such as MIT and UNC are developing Less Discriminatory Algorithmic Models (LDAs) that integrate alternative data (e.g., rent payments) to improve fairness. Meanwhile, healthcare faces risks of diagnostic inaccuracies, exemplified by early iterations of IBM Watson Health's oncology tools, which struggled with limited training data diversity. Modern solutions now employ agent-based search and summarization systems, such as retrieval-augmented generation (RAG) or agentic search (i.e. deep-search functions in OpenAI, Perplexity), to dynamically pull and validate insights from updated medical databases, reducing reliance on static datasets.

Security remains critical when AI agent systems interact with external APIs or sensitive databases. For example, Mitre's AI-driven code management system processes legacy government software in a secure AWS Bedrock environment, ensuring compliance with data privacy laws e.g. GDPR in EU and HIPAA in USA, while mitigating data leakage risks. Similarly, autonomous physical AI agent systems, such as self-driving cars, face latency and safety challenges. Tesla's Autopilot employs hierarchical agent architectures where high-level planners delegate real-time obstacle avoidance to subordinate model-based reflex agents, balancing decision speed with safety.

Hallucination risks are acute in generative AI applications. Salesforce's Agentforce addresses this by integrating sentiment analysis and human-in-the-loop (HITL) validation during customer interactions, reducing errors in loan servicing and billing support. Meanwhile, Amazon Connect Contact Lens uses NLP to audit AI agent responses in contact centers, flagging inconsistencies for review. For transparency, the EU AI Act mandates explainability in high-risk systems such as credit assessments, pushing firms like JPMorgan to adopt ReAct frameworks that document agent decision trees and tool interactions. To address these challenges, technical approaches such as AgentOps emerge as a key framework. AgentOps can offer autonomous monitoring mechanisms that continuously track the performance, interactions and behaviour of AI agent systems, detecting emergent system behaviours and potential risks in real-time. Furthermore, it can enhance transparency by providing clear documentation and audit trails of agent actions and decisions. Compliance checks integrated into AgentOps frameworks ensure that systems adhere to relevant regulations and ethical standards throughout their lifecycle. By implementing these measures, AgentOps fosters increased trust in AI operations, enabling organisations to responsibly harness the power of agent-based AI while mitigating the associated risks.



## 1.3 Motivation and objectives

This whitepaper addresses the growing risks and challenges associated with the deployment of AI agent systems by focusing on a key question:

*Is AgentOps a viable solution to support compliance with the European Union's AI Act?*

This inquiry is particularly timely amid the rapid development and deployment of AI technologies. As AI becomes deeply integrated across industries, concerns about its potential risks are growing. With 70% of companies worldwide now adopting AI and U.S. private AI investment reaching \$109.1 billion in 2024, the need for scalable mitigation strategies – such as AgentOps – has become increasingly urgent<sup>1 2</sup>. Examples from leading companies, such as Booking.com's deployment of multi- agent orchestration for automated risk categorisation and dynamic data quality compliance<sup>3</sup>, highlight the potential of these approaches.

Additionally, AI leaders are investing twice as much in governance tools compared to their peers, prioritising techniques like retrieval-augmented generation (RAG) and continuous feedback loops to address a significant 62% of compliance challenges related to core business functions<sup>4</sup>. By operationalising these types of solutions, AgentOps may offer a pathway to reconcile the rapid growth of the AI market (projected 33.8% annual growth<sup>5</sup>) with evolving regulatory requirements, ensuring accountability without stifling innovation.

As AI agent systems become central to business operations, it is important to explore how companies should comply with emerging regulations. We chose to study how the adoption of AI agent systems affects compliance with the EU AI Act given that it is the world's first comprehensive AI legislation. A strong compliance framework is crucial for businesses to ensure that they can harness the full potential of AI while protecting themselves from potential challenges and disruptions.

The next critical step is the practical trial of these solutions. Real-world applications are essential, not only to draw more reliable conclusions but also to support continued development and iteration. It is important to explore various combinations of risk mitigation tools and identify best practices for their implementation. A comprehensive approach is necessary, as no single tool can adequately address the full spectrum of risks posed by generative AI. Instead, effective mitigation will require a blend of technical and socio-technical measures, tailored to each specific use case and shaped by an organisation's capabilities – both in terms of expertise and financial resources – as well as its product portfolio.

<sup>1</sup> <https://pureai.com/articles/2025/04/08/ai-index-2025-reveals-surge-in-adoption.aspx>

<sup>2</sup> <https://www.bcg.com/press/24october2024-ai-adoption-in-2024-74-of-companies-struggle-to-achieve-and-scale-value>

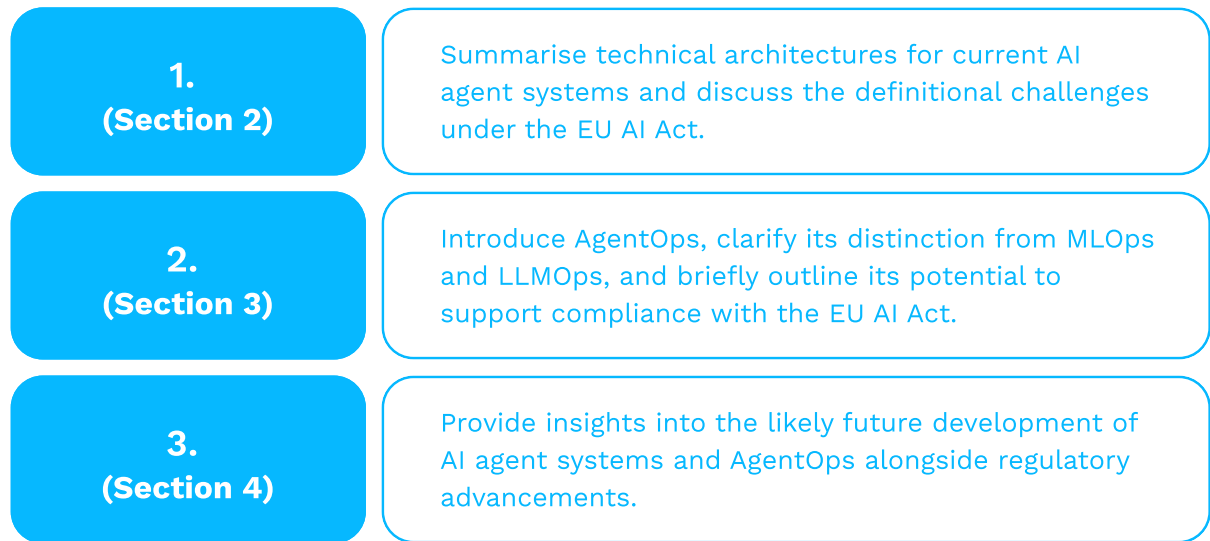
<sup>3</sup> <https://news.bloomberglaw.com/in-house-counsel/eus-ai-act-is-in-force-four-execs-share-their-best-practices>

<sup>4</sup> <https://ioni.ai/post/best-practices-in-building-compliance-ai-agents>

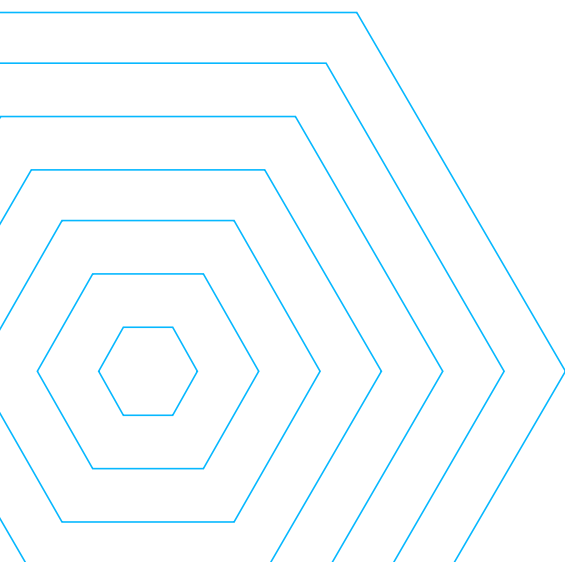
<sup>5</sup> <https://www.softwareimprovementgroup.com/eu-ai-act-summary/>

## 1.4 Scope and structure

This whitepaper aims to:



Our motivation is to establish a comprehensive understanding of the risk landscape for AI agent systems and evaluate the effectiveness of AgentOps as a governance strategy. This whitepaper will be particularly relevant for business leaders and industry practitioners looking to implement AI agent systems responsibly while maintaining competitive advantage and operational efficiency.



# AI Agents: definition, types and how they work

## 2.1 Definition and capabilities

An AI agent is a dynamic system that autonomously makes decisions and acts in real-world scenarios, unlike static workflows (predefined step-by-step sequences). For example, a workflow might automate invoice processing with fixed rules, while an agent adapts to unexpected inputs, like negotiating a contract clause. Modern LLM-based agents combine four core design patterns to enable this flexibility:



### Reasoning & Reflection Agent

**Thinking before acting:** Agents analyse their own logic (e.g., verifying if a medical diagnosis aligns with symptoms) to reduce errors.



### Planning Agent

**Breaking down complexity:** Agents decompose tasks (e.g., planning a trip by first booking flights, then hotels) using logic chains or tree-of-thought frameworks.



### Tool Use Agent

**Augmenting capabilities:** Agents interact with APIs/tools (e.g., pulling real-time stock data via Bloomberg Terminal integrations) to overcome LLM limitations.



### Memory Agent

**Learning from context:** Agents retain past interactions (e.g., tracking user preferences in a multi-day project) to manage long-term tasks.

Unlike rigid workflows, agents re-evaluate decisions based on new data (e.g., rerouting a delivery drone around sudden weather changes) — blending rules, learning, and real-time reasoning to handle uncertainty.

## 2.2 Design types of AI agents

Your design decision will directly affect the complexity and therefore the evaluation and regulation compliance difficulty. To be discussed in Section 3.

### Single Agent



- A single AI system **operates independently** to complete a **given task**.
- **Performs actions** and **reacts** to their results, including some error handling or unexpected results.
- Continues on its own until the **task either succeeds or fails**.

### Multi Agent



- Multiple agents, **each with their own specific subtask**, communicate and collaborate to solve the global task together.
- Endless **communication patterns**, depending on the global task.

## 2.3 Deep dive into each agent design types



### Reasoning & Reflection Agent

#### Single Agent

##### Goal

- Spend time thinking to improve accuracy.

##### Techniques

- Chain-of-thought prompting: e.g. 'think step by step'
- Reflection: feedback & refine iteratively
- Reasoning models: Reinforcement Learning on thought tokens

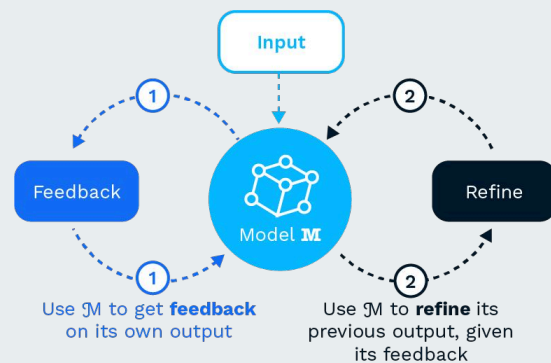


Figure 1. SELF-REFINE is instantiated with a language model and does not involve human assistance.

##### Pros

- Can potentially massively improve accuracy and quality (reduce hallucinations—see reflection part below).
- If thought is visible: slightly 'whiter box' interpretability.

##### Cons

- Higher latency and cost: need to think first and potentially for quite long.
- Less predictable: The more steps an agent reasons, the more unpredictable it becomes.  
→ This is a trade-off decision, related to reduce hallucination. please check our reflection part below.

##### Use

- When making difficult core decisions where accuracy and quality is more important than latency or cost, e.g. maths, coding, logic, and planning.

##### Don't Use

- When doing simple tasks, or where creativity (i.e. desirable hallucination) is wanted, or for real-time applications like voice.

##### Reflection

#### 1. The Challenge of Factual Inaccuracy in Generative AI

- The generation of factually incorrect information, or 'hallucinations', by AI models poses a significant operational challenge. Such outputs are generally undesirable, as they undermine AI system reliability and trustworthiness across applications.

#### 2. Enhancing Creative Problem-Solving through Reasoning and Reflection Agents

- In AI-driven problem-solving, creativity involves generating solutions that are both original and effective. Reasoning and reflection capabilities within AI agents are pivotal for this.
  - Exploration of Novel Paths: Reasoning enables AI agents to explore a broader spectrum of solution paths beyond conventional or probable trajectories, including unconventional reasoning lines. Reflection allows the agent to evaluate these paths and their performance.
  - Selection for Effectiveness: From explored paths, effective ones are selected. The interplay of expansive reasoning and critical reflection identifies responses that are original and practically effective, fulfilling creativity criteria.



### Goal

- Use reasoning to decompose a big task into smaller sub-tasks.
- Decide dynamically which steps to take based on context, instead of hard-coding.

### Techniques

- Decomposition-First: (1) Plan, (2) Execute step-by-step
- Interleaved: (1) Decide & execute step 1, (2) React, (3) Repeat or stop

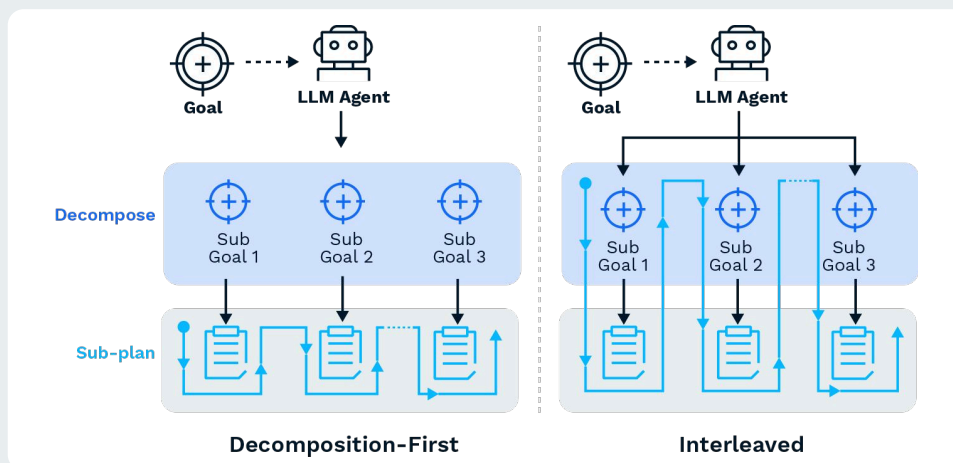


Figure 2. Example: Jina AI DeepSearch + more audience-relevant example

### Pros

- Can handle much more complex and longer tasks.
- Can enable some parallelisation (for independent steps) → faster.
- Each decomposed step can be delegated to appropriate 'experts' (e.g. tool, agent, different LLM or neural network, or even a human) → better performance, cost, speed, etc.

### Cons

- LLMs are not the best at planning yet, can introduce necessary complexity.
- If no backtracking is possible, might be stuck with a 'bad plan', i.e. deciding too early on how to solve the problem and realising too late (or never realising) that another step should have been taken, which leads to compounding errors.

### Use

- When you have a complex task that you cannot decompose in advance.

### Don't Use

- If you have a well-defined problem with known subtasks or steps, explicitly design that into a pipeline/ multi-agent workflow instead.



#### Goal

- Let the LLM interact with other computer systems as 'tools' as input, output, or more.

#### Common Types

- Retrieval (web search, DB/KB queries): fetch latest/detailed information
- Decision support (calculator, code interpreter): use more efficient/accurate tools
- Action (create alarm, generate images, run code): perform actions
- Communication (generate output to user, call other agents): send/receive information to/from humans or other agents

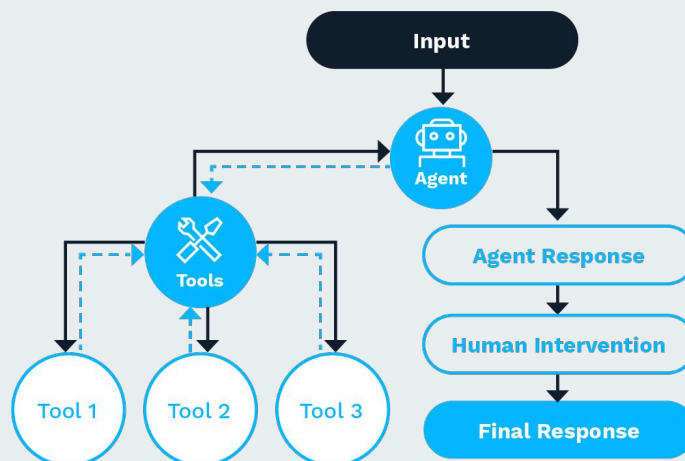


Figure 3. Overview diagram of standard tool use agent

#### Pros

- Potentially reduce hallucinations by providing facts, not forcing predictions (hallucinations are just a type of mispredictions of the LLM).
- Significantly broadens the action space (what they can do) from just text.
- Outsource tasks that LLMs are bad at, or are too inefficient for (e.g. arithmetic).
- Structured output that can interface with your system (now possible to guarantee correct/validate JSON).
- Each tool call is an opportunity for logging, verifying and steering the agent, including LLM-as-a-verifier, guardrails, human-in-the-loop.
- Most LLMs are now very good at tool use and quite reliable.
- Providing specific tools is often easier, more effective, and more predictable than just prompting.

#### Cons

- Agent hallucinations can now have real negative consequences.
- Higher cost, since each tool call leads to another LLM inference call.
- Raw text quality can sometimes be degraded.

#### Use

- For any interaction with external systems, **LLM interacts with other computer systems.**

#### Don't Use

- For simple text outputs without structured data requirements.



### Goal

- Self-improve in the next interaction by selectively saving relevant information. Unlike RAG, which retrieves external information, memory is usually self-managed, i.e. written and edited by the AI system.

### Examples

- ChatGPT with memory, character.ai, most customer support agents

### Techniques

- Long-term (external memory): Save information that could be useful for future tasks, either as general summarising information or specific episodic memory that is retrieved when relevant.
  - Example: Remembering user preferences or past mistakes.
- Implementation strategy: Use tools or multi-agent systems to enable memory functionality, either by integrating external memory tools (e.g., vector databases) or delegating memory tasks to specialised memory agents within a multi-agent framework.
  - Example: A tool-based agent queries a vector DB for past conversations, while a multi-agent setup includes a dedicated 'memory agent' that retrieves and updates relevant context.

### Pros

- Can perform better on the first try after having made a mistake.
- Can personalise the experience for the user.

### Cons

- If the memory contains hallucinations, it can compound the error.
- Grows over time, hard to tell which is relevant or not for the future.
- Increases cost to inject memory into the context
- System is no longer deterministic compare with traditional AI systems.
- Increased privacy concerns, since it'd be hard to remove one aspect of the memory.

### Use

- Only when necessary for the use case, due to potential instability in the long-term, or if a human (the user or from your business) can manually oversee and adjust the memory periodically.

### Don't Use

- Consistency and predictability are more important than personalisation, or privacy is very important.

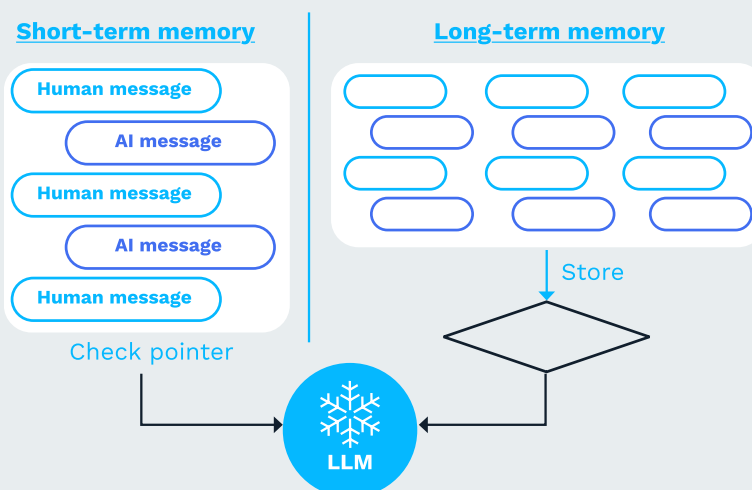


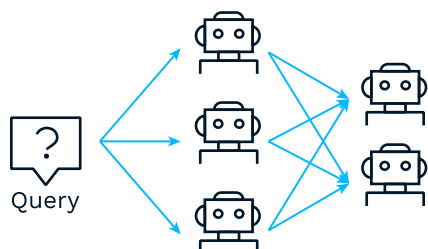
Figure 4. Difference of short-term agent and long-term memory agent



## Multi Agent

A multi-agent system refers to a collection of autonomous agents that collaborate or coordinate to achieve complex goals that are difficult for a single agent to handle alone. These agents can communicate, share memory, and divide responsibilities based on task specialisation or system design.

Common multi-agent architectures include the type below:

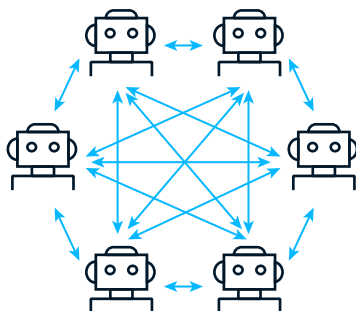


### Layered

Agents are organised hierarchically, with different layers handling planning, memory, and execution separately.

#### Examples

A top-level planner agent sets goals, a mid-level memory agent retrieves relevant context, and a low-level executor agent performs the actual task (e.g., calling APIs or generating responses).

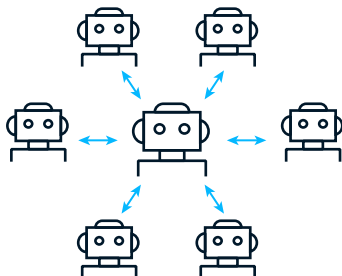


### Centralised

A central agent oversees and coordinates the activities of multiple sub-agents, managing memory and task distribution.

#### Examples

A central orchestrator delegates tasks to a search agent, summariser agent, and memory agent, then compiles the results into a unified response.

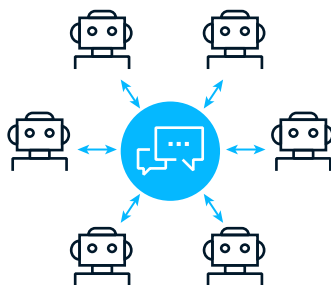


### Decentralised

Agents operate independently and communicate directly with each other, without a central controller.

#### Examples

A question-answering agent and a retrieval agent exchange messages to cooperatively respond to a query, each deciding its own actions.



### Grouped

Agents interact as participants in a shared conversation space, simulating human-like group discussions to collaboratively solve problems.

#### Examples

A 'researcher' agent, a 'critic' agent, and a 'writer' agent brainstorm together in a shared chat to generate and refine content.

## Key Trade-offs between single agent vs. multi agent design patterns

- **Cost vs. Resource Allocation:** Single-agent systems are more cost-effective but might be less efficient in distributed tasks, whereas multi-agents involve higher costs but better resource utilisation.
- **Specialisation vs. Generalisation:** Multi-agent systems can leverage specialised agents to optimize specific tasks, resulting in higher overall efficiency, whereas single agents must generalize their functionality.
- **Performance:** Multi-agent systems often perform better through parallel processing and task distribution, while single-agent systems can become performance bottlenecks.
- **Reliability:** Multi-agent systems offer greater reliability through redundancy and fault tolerance, single-agent systems can fail entirely if the sole agent fails.
- **Engineering Complexity:** Single-agent systems are simpler to design but may need extensive optimisation, while multi-agent systems require complex coordination but can be more robust and adaptive.
- **Hosting:** Single-agent systems are easier and cheaper to host locally, whereas multi-agent systems may benefit from distributed hosting solutions to manage resources more dynamically.

## Summary of advantages difference

### Single Agent



- A single AI system **operates independently** to complete a **given task**.
- **Performs actions** and **reacts** to their results, including some error handling or unexpected results.
- Continues on its own until the **task either succeeds or fails**.

### Multi Agent



- Multiple agents, **each with their own specific subtask**, communicate and collaborate to solve the global task together.
- Endless **communication patterns**, depending on the global task.

## 2.4 Defining AI agents under the AI Act

In the previous sections we explored the technical architectures for agentic systems and their business implications. Here we map those architectures to the terminology of the EU AI Act.

### Distinguishing between AI models and AI systems



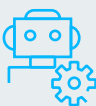
The EU AI Act does not explicitly mention AI agents – instead, it governs three categories: general purpose AI models (GPAI models), AI systems, and General-Purpose AI systems (GPAI systems).

Table 1 reproduces the Act’s formal definitions. In general terms, GPAI models are typically foundation models such as large language models, AI systems encompass machine-learning or logic-based applications, and GPAI systems are AI systems are built on top of GPAI models such that they can perform a variety of tasks.

It is important for enterprises to remember that a GPAI model on its own is not an AI system. It becomes part of a system only when additional engineering (APIs, UI, orchestration logic, data pipelines, etc.) is wrapped around it.

For instance, Open AI’s GPT-4o is a GPAI model. ChatGPT, which adds a conversational interface, user authentication, safety filters, etc. to GPT-4o, is an AI system. And, in this case, it is a general purpose AI system because it can be used for many purposes.

Table 1. Key definitions under the EU AI Act

	 GPAI Models	 AI Systems	 AI Systems
Definition	A GPAI model is defined as an AI model, including where such an AI model is trained with a large amount of data using self-supervision at scale, that displays significant generality and is capable of competently performing a wide range of distinct tasks regardless of the way the model is placed on the market and that can be integrated into a variety of downstream systems or applications, except AI models that are used for research, development or prototyping activities before they are placed on the market.	AI system means a machine-based system designed to operate with varying levels of autonomy, that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments.	A GPAI system means an AI system which is based on a general purpose AI model, that has the capability to serve a variety of purposes, both for direct use as well as for integration in other AI systems.
AI Act Reference	Article 3(63)	Article 3(1)	Article 3(66)
Example	GPT4o, Claude 4, Llama 4	Email spam classifier	ChatGPT, Perplexity

## AI agents as AI systems

AI agents — which typically integrate a general-purpose AI (GPAI) model with additional components such as tools, memory, or orchestration logic — will, at a minimum, be classified as AI systems under the EU AI Act. If these agents are capable of serving a wide range of functions, they may also fall under the category of general-purpose AI systems. However, current regulations offer no clear guidance or standardized criteria for determining whether a system qualifies as serving ‘a variety of purposes.’

It’s important to note that general-purpose AI systems are not exempt or distinct from broader regulatory obligations — they are simply a specific category within the broader AI system definition. Developers of such systems must still meet all applicable requirements set out in the AI Act.

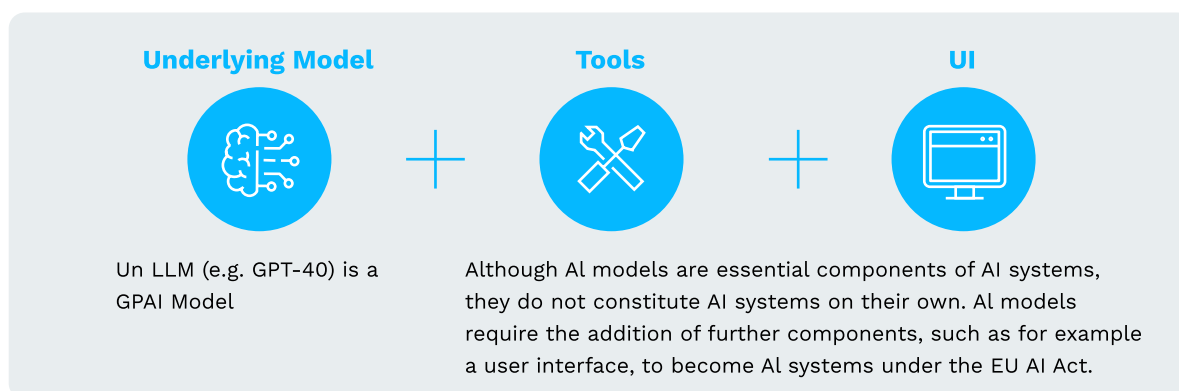


Figure 5. A simple AI Agent is an AI system under the AI Act

## Defining a boundary for AI agents

For regulatory purposes, it is important for enterprises to define the boundaries of their AI systems clearly. For more complex use cases, companies should distinguish between a single or multiple GPAI models operating as part of a single AI system and multiple, modular AI systems that may serve distinct purposes, but are capable of operating in concert when needed.

In practice, the EU AI Act does not offer guidance on how to set this boundary. It is up to each company to use their best judgement to classify what counts as ‘one’ AI system. We discuss the regulatory implications of GPAI models and AI systems in more detail in Section 3.

Table 2. Hypothetical examples of defining a system boundary

Boundary pattern	What it looks like in practice	Integrated LLM-only example
Single GPAI system that contains one or more GPAI models	All LLMs sit behind one shared governance stack.	A law-firm platform uses two specialised LLMs: one fine-tuned to summarise contracts and a second to classify risky clauses. Both models run inside the same micro-service, read from the same document repository, and surface results through one UI. The firm therefore registers and monitors them as one AI system under the AI Act.
Multiple, modular GPAI systems that can operate in concert	Each LLM-powered system has its own owner, compliance artefacts, and change-control process, but they exchange signals via secure APIs when a joint capability is required.	A SaaS vendor maintains three separate LLM systems: (1) a product-description generator, (2) a multilingual customer-support assistant, and (3) a regulatory-compliance reviewer that checks marketing copy for restricted claims. When the generator publishes new copy, it automatically pings the compliance reviewer for approval and then passes the approved text to the support assistant for translation. Despite this coordination, the firm documents each module an independent AI system for regulatory purposes.

# AgentOps and design principles for AI agents

## 3.1 Understanding AgentOps

**AgentOps** (Agent Operations) encompasses the specialized set of practices, tools, principles, and processes required for effectively building, deploying, managing, operating, and monitoring applications built using autonomous or semi-autonomous AI agent systems.

The objective of this paper is to provide a structured AgentOps framework concept, to ensure the robustness and reliability of both **single-agent and multi-agent systems** in development and operations, while enabling continuous improvement and mitigating risks.

AgentOps scope covers the **entire lifecycle of agent-based applications**, focusing on the control and monitoring of AI behaviour, interactions, and decision-making processes:

- Starting with the initial problem definition, includes the careful design of agent roles, capabilities, goals, and interaction protocols.
- Continuing through iterative development and thorough testing phases, and extending into deployment and operational monitoring.
- Combining design principles with operational oversight (based on automated, even AI-based, or humans) to mitigate data-related, user-driven, and technical risks.

## Comparison with other AI operations disciplines

AgentOps is distinct from, yet often complementary to, other operational disciplines within the Machine Learning and AI landscape:

### MLOps (Machine Learning Operations)

- MLOps is primarily concerned with streamlining and standardising the lifecycle of individual machine learning models, particularly within model-centric approaches.
- Its core concerns involve the efficient and reproducible building, training, validation, deployment, and monitoring of models, typically for predictive or analytical tasks.
- An example would be the pipelines and practices used to continuously train and deploy a video recommendation model for a platform like YouTube, emphasising automation, model versioning, and performance metric tracking (like accuracy or prediction speed).

### LLMOps (Large Language Model Operations)

- LLMOps addresses the specialised challenges associated with the development, training, fine-tuning, deployment, and serving of large AI foundation models, such as LLMs. It reflects the broader shift from model-centric to data-centric approaches.
- This field emphasises curation of massive datasets, distributed training infrastructure (e.g., large GPU clusters), sophisticated prompt engineering workflows, model versioning specific to LLMs, implementing content safety filters, and optimising the efficient serving and inference of these computationally intensive models.

### Understanding AgentOps at a Glance

- AgentOps builds upon concepts drawn from MLOps (for managing underlying models an agent might use) and LLMOps (for LLM-based agents), but its primary focus shifts significantly.
- Instead of focusing solely on individual models or base LLM infrastructure, AgentOps is uniquely concerned with both the development and **operational management of interactive, goal-driven agent systems**.
- It emphasises the **orchestration, monitoring, and safety of autonomous behaviours**, the reliability of coordinated task execution, the management of agent interactions, and ensuring safety and effectiveness of complex agent systems within its operational environment.

## 3.2 Embedding AgentOps in an EU AI Act programme

In this section, we describe how a good AgentOps framework can help enterprises build compliant AI Agents under the EU AI Act. We show how:

- The EU AI Act creates obligations for builders of AI systems.
- To abstract the requirements for high-risk AI systems.
- To embed regulatory requirements into an AgentOps workflow.

### How the EU AI Act classifies risk

When building and deploying AI Agents, enterprises must be familiar with the obligations that the EU AI Act imposes on providers of AI systems. Given that most companies are likely to integrate GPAI models into a system, the first step for companies is to classify the risk class of that GPAI system according to the EU AI Act's four-tiered classification system.

Table 3. AI system risk classes under the AI Act

Risk tier	Regulatory outcome	Typical conformity route
Unacceptable (Art. 5)	Use banned outright	n/a
High-risk (Art. 6)	Strict product-safety regime (risk management, data governance, logging, oversight, robustness, incident reporting)	Internal control or third-party assessment, plus harmonised standards when available*
Limited / transparency (Art. 50)	Disclosure, watermarking, user information duties	n/a
Minimal (Art. 95)	No specific legal duties	Voluntary best practice

## AgentOps for high-risk use cases

We focus on how AgentOps can support companies in meeting the requirements for high-risk AI systems for two key reasons. First, these requirements are the most stringent under the EU AI Act. A framework capable of enabling compliance at this level will inherently provide the flexibility needed to adapt to other, less demanding regulatory obligations – both within the EU and across other jurisdictions. Second, many enterprise-grade AI agents – such as those used in credit underwriting, HR management, medical decision support, or critical infrastructure – are likely to be classified as high-risk. As such, ensuring compliance in these contexts is both practically necessary and broadly applicable.

## Simplifying high-risk requirements

The most stringent requirements apply to providers of high-risk AI systems, who must follow the requirements for building AI systems (Chapter III, Section 2, Article 9-15 of the AI Act). To give companies a useful abstraction, we divide these articles into three ‘tiers.’

### Tier 1: Foundational Engineering Controls

The first is a ‘foundational’ set of engineering activities – these are common and scalable engineering practices that meet most of the requirements for high-risk systems.

<b>AI Act focus</b>	Articles 10, 12–15 (data governance, logging, transparency, human oversight, robustness & cyber-security)
<b>Core obligations</b>	<ul style="list-style-type: none"><li>• Establish and document data-quality rules, bias mitigation and provenance <b>(Art. 10)</b>.</li><li>• Capture relevant logs <b>(Art. 12)</b>.</li><li>• Supply user-facing instructions for safe integration and operation <b>(Art. 13)</b>.</li><li>• Ensure effective human-oversight and governance policies <b>(Art. 14)</b>.</li><li>• Validate accuracy, robustness and cybersecurity before release and continuously thereafter <b>(Art. 15)</b>.</li></ul>
<b>Why teams should care</b>	These duties will appear in almost every use case, regardless of risk, and form the baseline for any later conformity assessment.

### Tier 2: Risk-Driven Enhancements

The second are risk management ‘enhancements’ – these are additional engineering activities that will sit ‘on top of’ the previous tier that are unique to the use case’s risk profile.

<b>AI Act focus</b>	<b>Article 9</b> (risk management system)
<b>Core obligations</b>	Perform a continuous, documented risk assessment covering intended purpose, foreseeable misuse and post-market monitoring.
<b>Why teams should care</b>	High-risk systems must show that risks are identified and mitigated and monitored throughout the life cycle.

### Tier 3: Documentation & Evidence

The third is documentation – these tasks relate to capturing key information about the system and evidence of compliance with the first two tiers.

<b>AI Act focus</b>	<b>Article 11</b> (technical documentation)
<b>Core obligations</b>	Compile a technical file that proves conformity with Articles 9–15.
<b>Why teams should care</b>	Without a complete, auditable file, no presumption of conformity is possible.



## **A gentle caveat**

This ladder is a framework, not a guarantee. Final conformity depends on the system's purpose, the assessment pathway chosen, forthcoming harmonised standards and the judgment of EU market-surveillance authorities. AgentOps simply supplies the operational scaffolding that makes those formal steps faster, cheaper and more defensible while safeguarding innovation.

## **Additional considerations: choosing the right GPAI model**

As we discussed in Section 2.4, GPAI models and AI systems are distinct categories under the AI Act. However, if an enterprise is using a GPAI model (e.g. via an API or by downloading weights from a repository), they must consider two factors:

### **a. How the model provider complies with their obligations**

For the most part, the obligations for GPAI models must be fulfilled by the actor who trains and releases these models (eg., Meta, Anthropic, etc.).

At the time of writing, the AI Office has published the Code of Practice (CoP) for providers of GPAI models and is now waiting for EU Member States and the European Commission to assess its adequacy. Organisations who comply with the rules in this CoP will be presumed to have released compliant models.

When selecting a GPAI model for an AI agent application, companies should consider if the organisation providing the model has complied with these rules, as this might have a bearing on the compliance of the GPAI system.

### **b. The effect of fine-tuning or modifying a model**

In addition to the CoP, the EU Commission has also published guidelines on the scope of obligations for providers and modifiers of GPAI models. Under certain circumstances, actors who modify a model will be considered the providers of those models and will have additional obligations to fulfil.

## **Putting it together**

Starting with tier1 baseline controls, teams layer on tier2 risk analytics proportional to their application, then seal everything with tier3 evidence generation. By treating legal duties as incremental engineering checkpoints, AgentOps turns compliance from last-minute paperwork into a continuous, testable DevOps routine. Figure 6 demonstrates this through a visualisation of the AgentOps workflow.

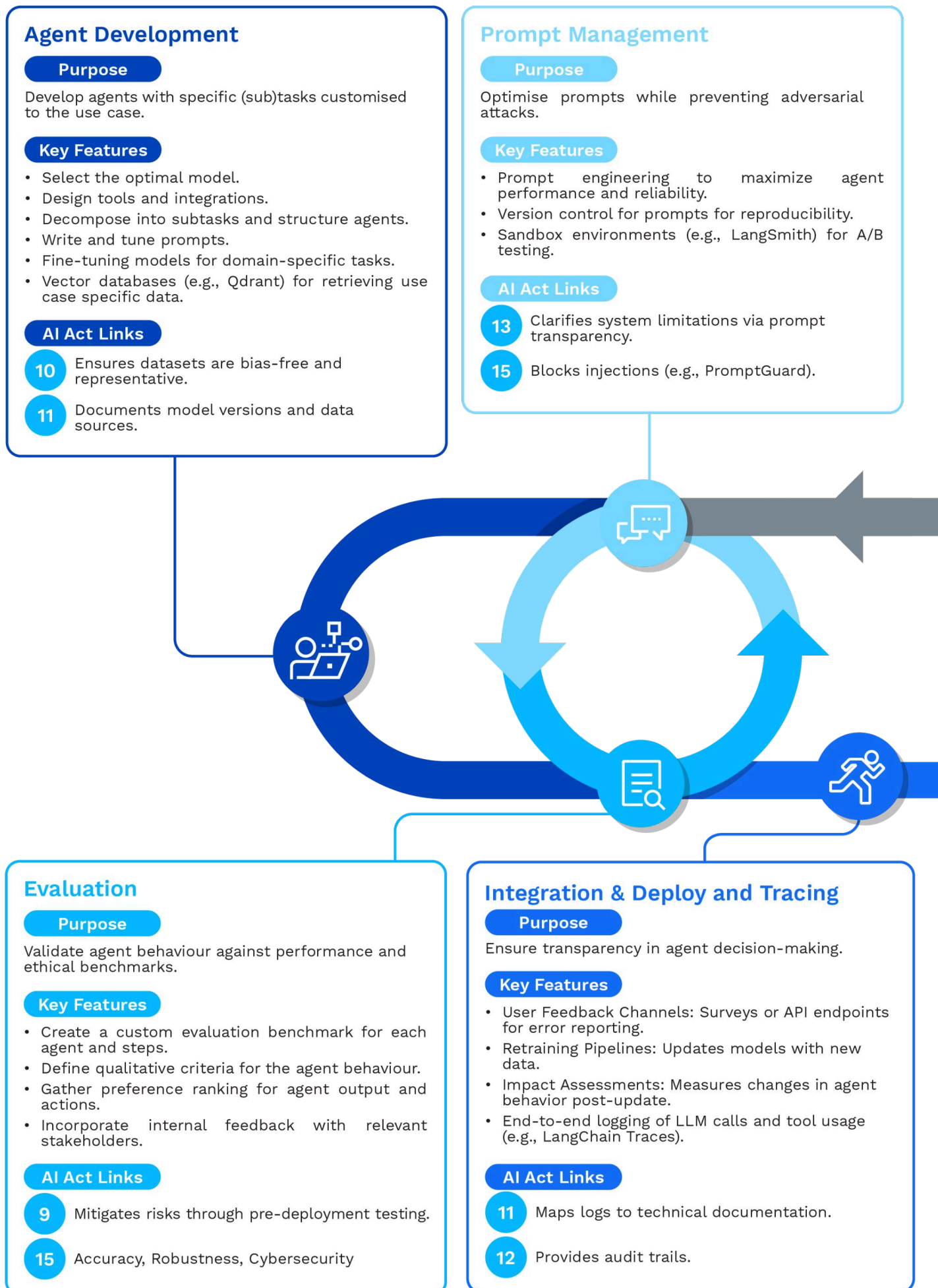


Figure 6. AgentOps concept overview

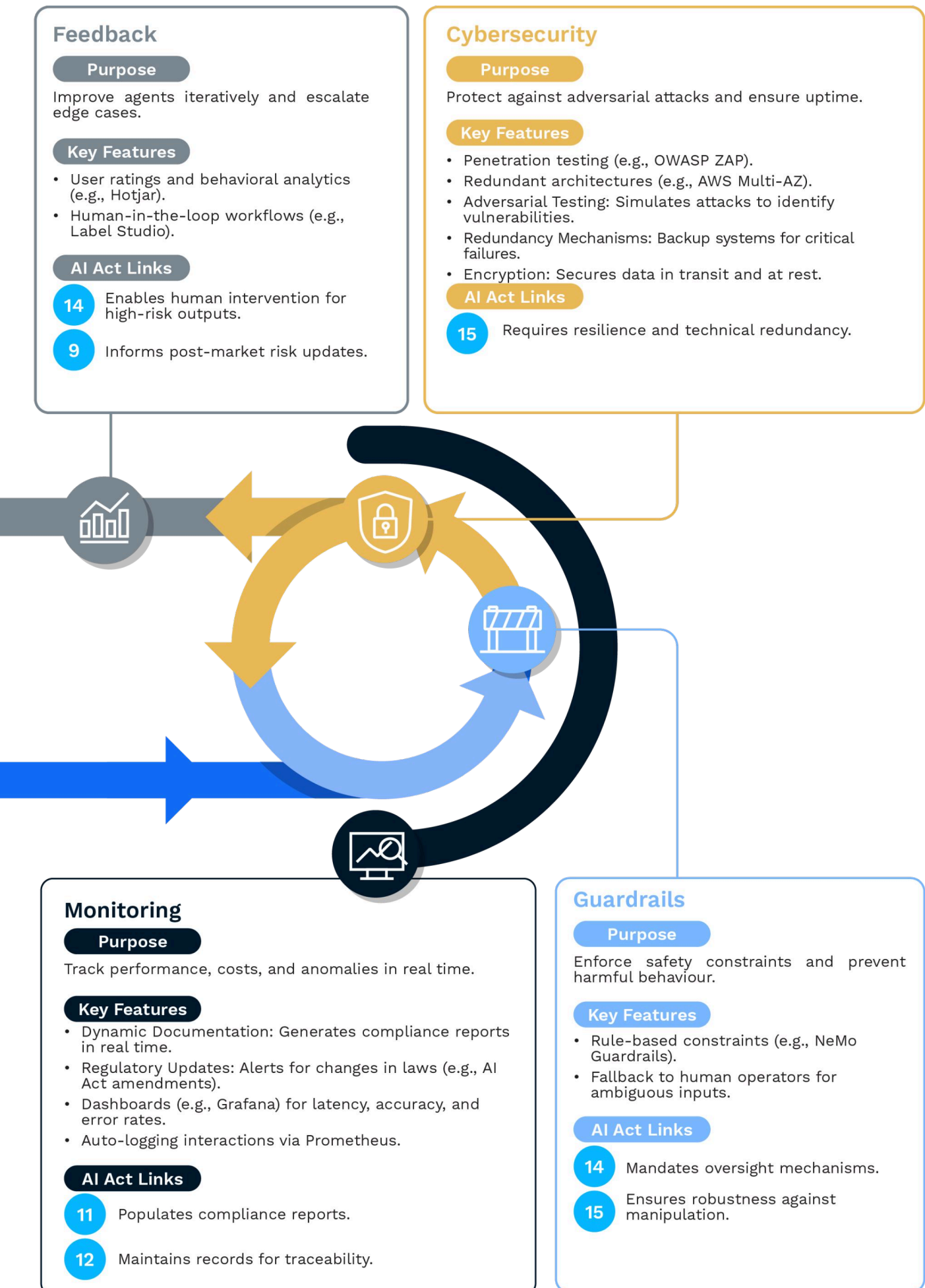


Figure 6. AgentOps concept overview

### 3.3 Designing AI agent systems: balancing architecture and risk

A fundamental architectural goal is to align the design of an AI agent system with AgentOps principles, ensuring that AI agent systems are inherently built to follow established operational best practices and incorporate necessary technical safeguards from the outset.

#### Key design considerations for AI agent systems:

- A fundamental architectural choice involves deciding between a **single-agent system**, where one AI manages multiple diverse tasks, or a multi-agent system, which utilises multiple specialised AIs, each potentially introducing unique interaction risks.
- Another critical design consideration is the **workflow structure**, determining whether processes will be **linear**, executing in a predefined step-by-step manner, or **iterative**, allowing for continuous feedback, adaptation, and potential refinement based on ongoing risk assessment.

#### Balancing system complexity and associated risk:

- Increasing an **AI system's complexity** may enhance its capabilities but also increase the risk of unintended behaviour and reduce traceability.
- **A holistic evaluation approach is vital** for understanding and managing risk in AI agent systems, particularly multi-agent systems, because their combined behaviours can create emergent properties and complex failures not seen in individual components.

#### Key architectural considerations:

- **Technical complexity assessment is key**, involving the evaluation of processing power, data throughput, and computational resources, which affects the AI system's feasibility, scalability, and maintainability.
- **Achieving resource efficiency is paramount**, meaning the AI agent architecture must optimise performance and responsiveness while minimising computational overhead, energy use, and operational costs.
- **The architecture must intrinsically support compliance and risk management** through its structure, enabling adherence to legal, ethical, and industry standards to mitigate harm and ensure accountability.

Ultimately, a well-defined design and architecture of an AI agent system is essential, as it proactively reduces security vulnerabilities, strengthens robustness against failures and security threats, and simplifies long-term management, maintenance, and system evolution – key factors in ensuring overall compliance and trustworthiness.

### 3.4 Best practices for implementing AgentOps and compliance

The following best practices provide a structured foundation for implementing AI agent systems based on AgentOps principles, ensuring operational robustness, compliance, and adaptability throughout the agent lifecycle.

**Foundation in design for operability and modularity:** From the outset, design AI agent systems (whether single or multi-agent) with operability, comprehensive observability (including logging, state tracking, and clear interfaces), and modularity in mind. Clearly define roles, responsibilities, capabilities, and communication protocols within modular designs to simplify management, debugging, fault isolation, and the assessment of interaction risks.

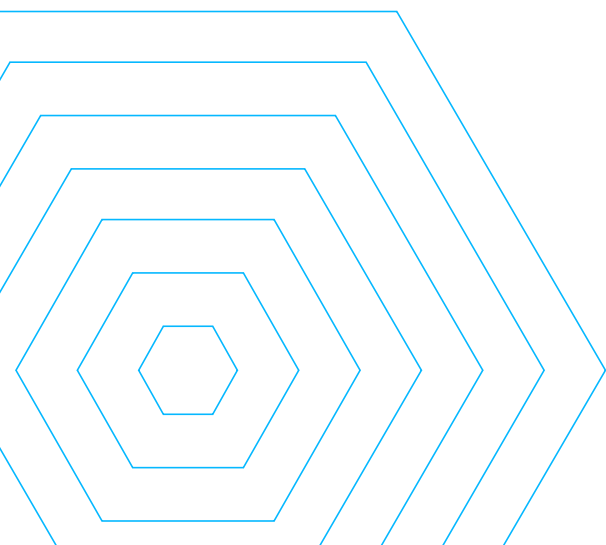
**Iterative development with continuous feedback:** Adopt an iterative development lifecycle for AI agent systems. This includes continuous deployment, multi-level monitoring of interactions and system health, and a consistent feedback loop for refinement and adaptation to real-world performance, evolving objectives, and identified risks.

**Comprehensive testing and validation:** Implement extensive testing regimes covering unit, integration, and particularly behavioural aspects. Test thoroughly against diverse, adversarial scenarios and for safety and ethical considerations, giving special attention to the emergent behaviours in complex multi-agent systems.

**Robust version control and traceability:** Maintain meticulous version control for all components of the agent system—including code, configurations, prompts, knowledge bases, models, tools, and communication protocols. This ensures reproducibility, full traceability, and the ability to perform safe rollbacks.

**Integrated governance:** Embed security safeguards (like access controls and secure data handling), ethical considerations, and compliance requirements directly into the AI agent and system architecture by design. Proactively and continuously identify, assess, and mitigate risks, evaluating the entire system holistically—especially emergent behaviours—rather than just its individual components.

**Resource efficiency and cost control by design:** Engineer AI agent systems and their components with a focus on resource efficiency. Implement and utilise mechanisms for diligent monitoring and active management of operational costs, ensuring system performance aligns with budgetary considerations.





# Discussion about future outlook and impact

## 4.1 Looking ahead (what we can expect) from AI Act and regulation perspective

While this paper has highlighted some of the challenges of interpreting the EU AI Act in light of developments in agentic AI, it is worth noting that there remain many moving parts to this law. At the time of writing, the EU Commission is still preparing the Code of Practice for general purpose AI models and guidance on what obligations actors who modify or fine-tune these models will face. Just as significantly, the harmonised standards that will prescribe compliance activities for providers of high-risk systems will only become available in late 2025 or early 2026. And finally, it remains to be seen how EU member states will implement their obligations under the Act in terms of market surveillance and oversight.

However, the EU is not the only jurisdiction where the governance of AI is a work in progress. AI law in other jurisdictions currently mimics the concerns that EU institutions have sought to address with respect to foundation models; namely issues around transparency, copyright, and safety and security. China, for example, recently published guidance for the watermarking of generative content<sup>1</sup>, which follows from its regulatory measures for GenAI services in July 2023<sup>2</sup>. Conversations are currently ongoing in China for a more comprehensive product safety based regulatory framework akin to the EU's AI Act.

AI regulation is in a state of flux in the US. At the federal level - among other things - the US National Science Foundation has issued a request for Information on the Development of an Artificial Intelligence (AI) Action Plan<sup>3</sup> that will substantially inform how foundation models are governed. In addition, the National Institute of Standards and Technology (NIST) has also issued several frameworks, guidelines, and profiles for GenAI and risk management. U.S. states are also drafting their own legislations, with some states like Colorado and Texas drawing inspiration from the EU AI Act.

Beyond existing regulation - the more agentic AI systems become, the more likely operators of such systems are to meet the limits of existing civil and criminal legislation. Regulators, civil society and industry will begin to debate contentious issues of how long-standing legal and regulatory principles should be amended to enable and constrain the development and deployment of agentic AI.

While there are an innumerable number of laws that will likely be affected in the future, some are predictably so. Consider, for example, the law of torts, which often depends heavily on standards of conduct. How should such standards be ascribed to agentic systems? Consider also, the law of contracts: should humans be legally bound to the actions of the AI agent systems? What about professional industry standards? How will agents and their operators be held to ethical standards when they provide advice that is typically reserved for licensed lawyers, financial advisers or doctors?

The governance of increasingly agentic AI remains a moving target across all major jurisdictions. Stakeholders should treat compliance as an iterative exercise, anticipating rapid updates to standards, statutes, and enforcement practices.

<sup>1</sup> [https://www.cac.gov.cn/2025-03/14/c\\_1743654684782215.htm](https://www.cac.gov.cn/2025-03/14/c_1743654684782215.htm)

<sup>2</sup> <https://www.chinalawtranslate.com/en/generative-ai-interim/>

<sup>3</sup> <https://www.federalregister.gov/documents/2025/02/06/2025-02305/request-for-information-on-the-development-of-an-artificial-intelligence-ai-action-plan>

## 4.2 What we can expect from european industry perspective

The widespread adoption of AI agent systems will reshape industries by introducing unprecedented automation, decision-making autonomy, and operational efficiency. However, this transformation demands robust frameworks like AgentOps to address technical complexities, regulatory compliance (notably the EU AI Act), and ethical imperatives. Below are some key domains poised for disruption, followed by a European industry outlook grounded in.

### 1. Real-world examples and challenges

**Search:** Jina AI predicts how the Search topic will be impacted and changed in the future because of Agent and other AI or new technologies. Surprisingly, many users still rely on outdated keyword-based systems. Why do these ‘old’ tools remain effective? Because humans compensate for their limitations. Users refine their queries, test synonyms, or chain searches together. This reveals a critical insight: the biggest gains in search won’t come from marginally improving retrieval models, but from employing search agents. Search engines have evolved dramatically over the years. We began with simple keyword-based systems, which matched queries to documents through literal terms. Then came semantic search, a leap forward that interpreted intent and context to deliver more relevant results. Yet, despite these advancements, we’ve hit a ceiling: even the most sophisticated search engines today struggle to answer complex questions in a single query.

To break through this barrier, we need agentic search — a paradigm where the system doesn’t just retrieve answers but actively strategises. For example, decomposing a query into sub-questions, iterating through hypotheses, or synthesising insights across multiple sources. This approach doesn’t just tweak accuracy by a few percentage points; it redefines search.

Use cases for agentic search span the full spectrum of a company’s data landscape — from deeply private internal records (user profiles, transaction logs, incident reports) all the way out to the broad expanse of public information (websites, news feeds, forums). They cover every function where information lives and decisions are made: product planning, customer support, competitive intelligence, regulatory compliance, and R&D. Here are two concrete examples that illustrate how an agentic system can shoulder the heavy lifting—and where human judgment still comes into play.

#### Example 1: Who is our main competitor?

A classic strategic question, answered today by tedious manual synthesis. With agentic search, some but not all of the tasks can be taken over by the agentic system.



#### Agent

- Find previous analysis.
- What products do we offer?
- Where are our markets.
- Finding other companies that offer the same products in the same markets.
- Check if they address the same user segment.
- Find predictive data that indicate the direction competitors are moving to.
- Write an email to someone requesting information.



#### Human

- Send the email.
- Talk to colleagues.
- Look into physical documents.



### Example 2: How did we fix our slow SQL queries last year?

A troubleshooting rewind that today means digging through scattered archives and hallway conversations. An agent can take over a lot of work for that task. However, certain tasks need to be done by humans.



#### Agent

- Find previous incident analysis.
- Find internal documentation.
- Find logs, error reports, and system alerts.
- Search relevant archived internal communications (emails, meeting minutes, forums).



#### Human

- Talk to colleagues who were directly involved in diagnosing or resolving the issue.

**Knowledge management:** AI agent systems will revolutionise how organisations process, retrieve, and synthesise information. For instance, agent-based search and summarisation systems can dynamically pull data from fragmented sources (e.g., internal databases, research papers) to generate actionable insights. In healthcare, agents could cross-reference patient histories with global medical research in real time, enabling faster diagnoses<sup>1</sup>. However, challenges like hallucination risks and data accuracy require rigorous validation layers to ensure reliability<sup>2</sup>.

**Software development:** Autonomous coding agents (e.g., GitHub Copilot's successors) will streamline workflows by generating code, debugging, and integrating APIs. For example, multi-agent systems could collaboratively design software architectures, with specialised agents handling frontend, backend, and security testing<sup>3</sup>. Yet, security vulnerabilities in AI-generated code—such as inadvertent use of unlicensed libraries—demand strict governance via tools like static analysis and human-in-the-loop reviews<sup>4,5</sup>.

**Customer service:** Fully autonomous agents will resolve 80% of routine inquiries by 2029, per Gartner predictions. Companies like Air Canada already deploy AI agent systems for real-time support, though missteps (e.g., incorrect bereavement policy advice) highlight the need for fail-safes and audit trails<sup>6</sup>.

**Supply chain optimisation:** AI agent systems will dynamically reconfigure logistics in response to disruptions (e.g., geopolitical events, natural disasters). For example, Symbolic's warehouse robots use agentic AI to optimise inventory placement, reducing latency by 30%<sup>6</sup>. However, reliance on real-time data APIs introduces cybersecurity risks that require zero-trust architectures<sup>7</sup>.

**Regulatory compliance:** Agents will automate GDPR and EU AI Act compliance tasks, such as data lineage tracking and bias audits. Salesforce's Agentforce already reduces hallucination risks in compliance workflows by 40% through human validation loops<sup>8</sup>.

<sup>1</sup> <https://elunion.com/2025/03/10/turning-ai-into-decision-making-agents-opportunities-challenges-and-whats-next/>

<sup>2</sup> <https://www.holistica.ai/blog/llm-agents-use-cases-risks>

<sup>3</sup> <https://www.galileo.ai/blog/introduction-to-agent-development-challenges-and-innovations>

<sup>4</sup> <https://builtin.com/artificial-intelligence/hidden-risks-ai-agent-adoption>

<sup>5</sup> <https://www.responsible.ai/from-genai-to-ai-agents-preparing-for-the-next-evolution-in-artificial-intelligence/>

<sup>6</sup> <https://tepp.perspectives.cmu.edu/all-articles/the-ethical-challenges-of-ai-agents/>

<sup>7</sup> <https://fpf.org/blog/minding-mindful-machines-ai-agents-and-data-protection-considerations/>

<sup>8</sup> <https://www.responsible.ai/from-genai-to-ai-agents-preparing-for-the-next-evolution-in-artificial-intelligence/>

## 2. Adoption trends and challenges

<b>High-adoption sectors</b>	<b>Healthcare<sup>1</sup></b> <p>Europe's stringent regulations (e.g., EU AI Act's 'high-risk' classification for medical AI) will drive adoption of compliant agents. For example, Babylon Health is piloting diagnostic agents that align with Article 10's data quality requirements, using retrieval-augmented generation (RAG) to validate outputs against peer-reviewed studies.</p>
	<b>Manufacturing<sup>2</sup></b> <p>German automakers are integrating physical AI agents into assembly lines. These agents monitor equipment health, predict failures, and autonomously order replacements — reducing downtime by 25%. The EU's focus on sustainability will further incentivise energy-efficient agent designs, such as adaptive activation to minimise computational waste.</p>
	<b>Financial services<sup>3</sup></b> <p>AI agents will dominate fraud detection and risk analysis. Dutch fintech Adyen uses agentic systems to flag suspicious transactions in real time, leveraging the EU's PSD2 open banking framework. However, compliance with the AI Act's transparency mandates (e.g., documenting decision trees) remains a hurdle.</p>
<b>Moderate/low-adoption sectors</b>	<b>Public sector<sup>4</sup></b> <p>While AI agents could streamline bureaucratic processes (e.g., visa processing), ethical concerns about bias in decision-making slow adoption. France's recent pause on AI-driven welfare eligibility systems reflects this caution.</p>
	<b>Education<sup>5</sup></b> <p>Despite potential for personalised learning agents, GDPR constraints on child data processing limit scalability. Finland's pilot with AI tutors in schools requires particular manual approval workflows for data access.</p>
	<b>Creative industries<sup>6</sup></b> <p>EU copyright laws complicate AI-generated content ownership. For instance, Italy's ban on AI-authored journalism underscores the tension between innovation and intellectual property rights.</p>

<sup>1</sup> <https://smythos.com/ai-agents/ai-agent-development/ai-agent-ethics/>

<sup>2</sup> <https://www.holisticai.com/blog/llm-agents-use-cases-risks>

<sup>3</sup> <https://smythos.com/ai-agents/ai-agent-development/ai-agent-ethics/>

<sup>4</sup> <https://tepperspectives.cmu.edu/all-articles/the-ethical-challenges-of-ai-agents/>

<sup>5</sup> <https://fpf.org/blog/minding-mindful-machines-ai-agents-and-data-protection-considerations/>

<sup>6</sup> <https://www.responsible.ai/from-genai-to-ai-agents-preparing-for-the-next-evolution-in-artificial-intelligence/>

## 4.3 Conclusion

The rapid evolution of artificial intelligence has ushered in the era of AI agent systems, bringing with it transformative opportunities – and significant regulatory challenges – for industries across Europe. These advanced autonomous systems promise substantial business value, enabling enhanced decision-making, personalised user experiences, and the creation of intelligent products and services. However, their accelerated development raises a critical concern: can regulatory frameworks, particularly the EU AI Act, keep pace with the complexity and unique capabilities of these systems?

Navigating the intricacies of the EU AI Act in the context of AI agent systems could appear daunting, however, achieving robust compliance is not an insurmountable obstacle. As this paper argues, the adoption of specialized frameworks is essential. Concepts such as agent design patterns – which embed safety, transparency, and control directly into the architecture of AI agents – and comprehensive AI AgentOps practices offer a viable path forward. AgentOps supports compliance by enabling tailored governance, ongoing monitoring, full lifecycle traceability, and mechanisms for accountability. Together, these frameworks empower organizations to meet legal requirements while fostering trust and responsible deployment of AI agents.

Looking ahead, the regulatory environment for AI agent systems will introduce new uncertainties that demand ongoing vigilance. Companies will need to ‘keep their eye on the ball,’ staying ahead of regulatory developments and interpretations as they emerge. However, from a technology and operational standpoint, a proactive approach is the optimal strategy. The early and thoughtful implementation of frameworks like AI AgentOps, complemented by sound agent design patterns, will be important. Such proactive measures will not only facilitate compliance with current and future regulations but also foster a culture of responsible AI development and deployment.

While the essential role of AgentOps frameworks and agent design patterns in achieving AI compliance and deploying reliable, trustworthy AI agent systems is evident, it's equally important to recognize that standardised practices in this space are still in their infancy. At present, companies face a significant challenge: the lack of widely established codes of conduct or universally accepted best practices tailored specifically to AgentOps. This absence of clear guidance can hinder effective implementation and lead to inconsistencies across the industry.

To address this, there is a growing consensus within the evolving AI ecosystem on the need for formalised principles and structured blueprints. Developing foundational guidelines for the design, development, deployment, and ongoing operation of AI agent systems would provide organisations with the clarity they need to move forward confidently.

Collaborative efforts that promote responsible AI agent design, robust operational governance, comprehensive risk management, and transparent accountability mechanisms will be crucial. These frameworks not only support smoother adoption and compliance with the EU AI Act but also pave the way for a future in which innovation and ethics are fully aligned.

By embedding these principles into their operational DNA, organisations can harness the power of AI agent systems responsibly, strengthen their competitive position, and contribute meaningfully to a trusted, AI-first future.

## Authors



**Mingyang Ma** 

Head of Agentic AI Solutions Development  
appliedAI Initiative GmbH  
[m.ma@appliedai.de](mailto:m.ma@appliedai.de)

Mingyang Ma is a leading expert in applying Large Language Models (LLMs), AI agents, and multimodal human-AI alignment. With over eight years of experience in natural language processing and AI product development, she brings deep expertise in conversational AI, particularly in building scalable platforms and DevOps infrastructures for complex, industry-grade applications.



**Akhil Deo** 

Senior AI Regulatory Expert  
appliedAI Institute gGmbH  
[a.deo@appliedai-institute.de](mailto:a.deo@appliedai-institute.de)

Akhil Deo is a Senior AI Regulatory Expert at the appliedAI Institute, specialising in the intersection of emerging technologies and public policy. With six years of experience navigating complex policy landscapes, he brings a strong background in AI governance and strategic communications. Prior to joining appliedAI, Akhil served as a Communications Specialist at the Future of Life Institute and as a Junior Research Fellow at the Observer Research Foundation, where he focused on technology policy and global AI governance.



**Bernhard Pflugfelder** 

AI Strategy and Coordination Lead  
Rohde & Schwarz  
[bernhard.pflugfelder@rohde-schwarz.com](mailto:bernhard.pflugfelder@rohde-schwarz.com)

Bernhard Pflugfelder works as AI strategy and Coordination Lead at Rohde & Schwarz. Bernhard has 15 years of experience in the fields of Data Science, Natural Language Processing (NLP), as well as data and AI across different companies such as BMW Group or Volkswagen Group. He is renowned for his expertise especially in the field of AI in general, NLP and generative AI in particular.



**Joong-Won Seo** 

Junior GenAI Engineer  
appliedAI Initiative GmbH  
[j.seo@appliedai.de](mailto:j.seo@appliedai.de)

Joong-Won is a master student at Technical University Munich (TUM) and also as a working student, Joong-Won has already dedicated two years to the role of Junior GenAI Engineer. He is a creative and self-driven individual with significant hands-on experience in cutting-edge Generative AI techniques and the development of AI agents.

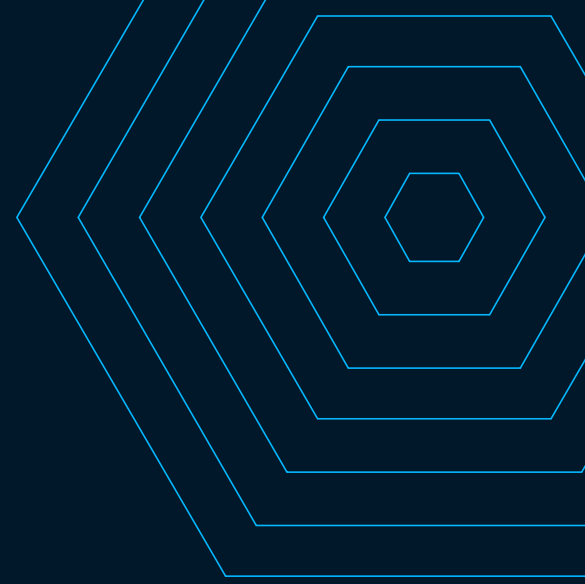
## Contributors



**Florian Hoenicke** 

Principal Engineer  
Jina AI  
[florian.hoenicke@jina.ai](mailto:florian.hoenicke@jina.ai)

Florian has more than 10 years of experience in the Field of AI, working at Axel-Springer, Deloitte, and SoundCloud. He works as a Principal Engineer at Jina AI, rapidly prototyping AI solutions. His expert domain is Agentic Search and synthetic data generation. Florian is serving as an AI policy advisor, providing explanations and insights to members of the European Parliament to enhance their understanding of artificial intelligence.



## About the appliedAI Institute for Europe

The appliedAI Institute for Europe is dedicated to strengthening the European AI ecosystem by advancing knowledge, promoting trustworthy AI tools, and offering high-quality educational and interactive formats.

Established in Munich in 2022 as a non-profit subsidiary of the appliedAI Initiative — a joint venture between UnternehmerTUM and IPAI — the Institute is led by Dr. Andreas Liebl and Dr. Frauke Goll.

At its core, the Institute is focused on people. Its mission is to build a shared European AI community and make AI knowledge accessible to all. By promoting the responsible and trustworthy use of AI, the Institute helps accelerate adoption and foster public confidence in AI technologies.

Through its emphasis on knowledge development, trusted tools, and engaging educational experiences, the appliedAI Institute for Europe serves as a valuable hub for companies, organisations, and individuals seeking to deepen their understanding of AI. It enables cross-sector collaboration and encourages the exchange of expertise.

The Institute welcomes companies, startups, institutions, and AI enthusiasts to explore and benefit from its wide range of programs, tools, and community initiatives.



Gefördert durch  
Bayerisches Staatsministerium  
für Digitales



## About the Bavarian AI Innovation Accelerator

The Bavarian AI Innovation Accelerator is a pioneering initiative designed to drive AI innovation and entrepreneurship across Bavaria. By supporting early-stage startups and fostering cutting-edge research, the Accelerator plays a key role in transforming Bavaria into a leading hub for artificial intelligence in Europe.

Spearheaded by the Bavarian Ministry of Economic Affairs, Regional Development and Energy, and powered by leading partners in academia, industry, and the startup ecosystem, the Accelerator provides targeted resources, mentorship, and funding opportunities. It offers a structured program to help AI-driven startups scale faster—from proof of concept to market-ready solutions.

The Accelerator emphasises applied innovation, focusing on real-world AI applications that solve pressing challenges across sectors such as manufacturing, healthcare, mobility, and sustainability. Through access to technical expertise, business coaching, and collaboration networks, participants are equipped to build impactful, trustworthy, and competitive AI solutions.

With a strong commitment to responsible AI development, the Bavarian AI Innovation Accelerator also supports alignment with regulatory frameworks such as the EU AI Act. By nurturing a vibrant AI ecosystem rooted in ethics, excellence, and entrepreneurship, the Accelerator empowers the next generation of innovators to shape Europe's AI future.

For more information, please visit [www.appliedai-institute.de](http://www.appliedai-institute.de).



**Bayerischer  
KI-Innovationsbeschleuniger**

**appliedAI Institute for Europe**

Freddie-Mercury-Straße 5

80797 München

Germany

[www.appliedai-institute.de](http://www.appliedai-institute.de)



Gefördert durch  
**Bayerisches Staatsministerium  
für Digitales**

