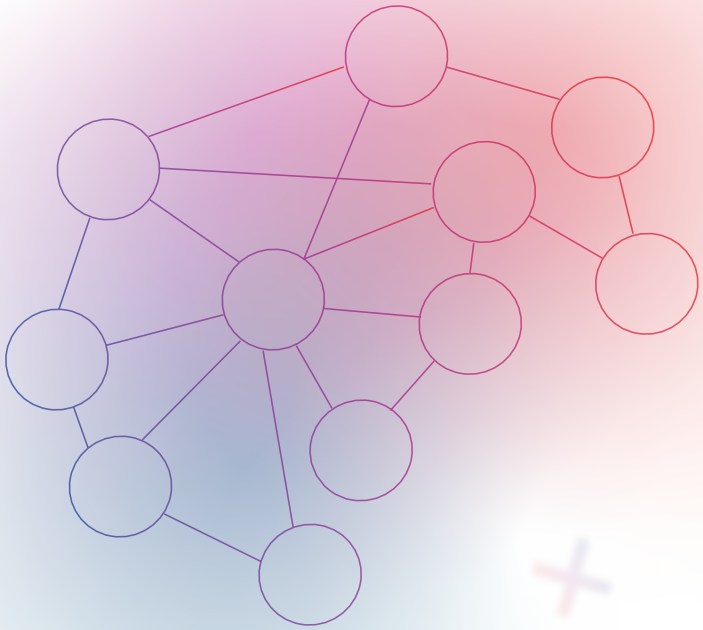
RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

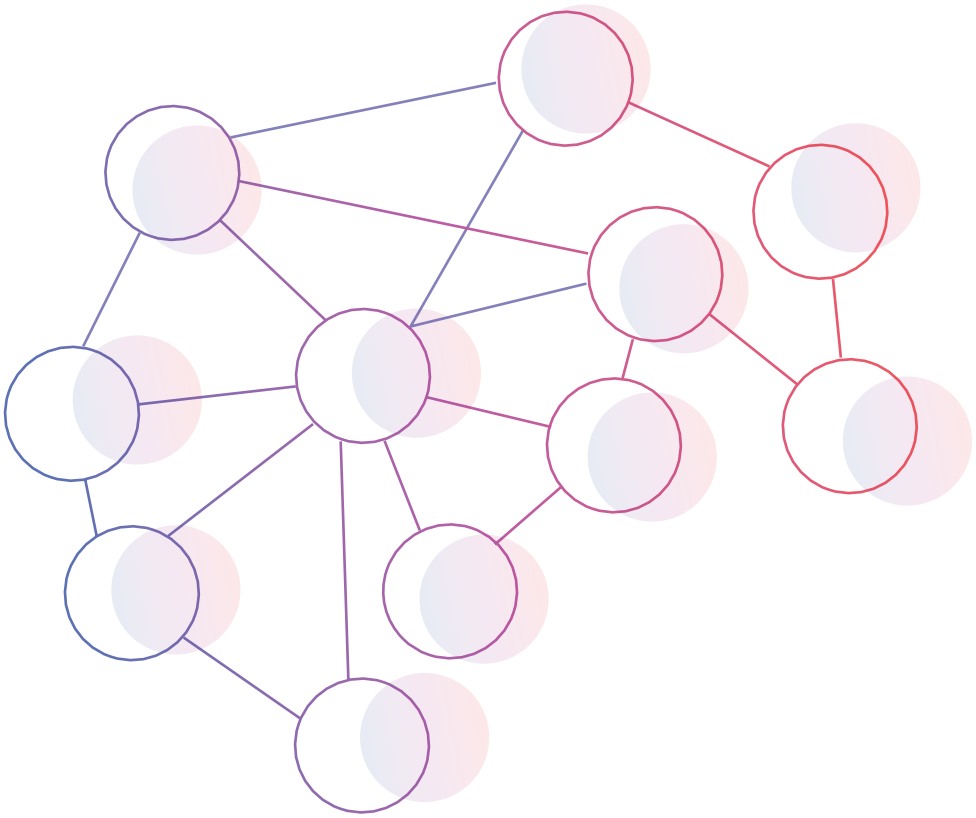
Inria

POLICY PAPER

Agentic AI: **Deployment, Adoption and Impacts**



June 2026



Introduction

Generative AI has recently evolved from models that primarily produce text, images, or code into systems capable of autonomous decision-making and proactive behavior within human-defined constraints. Agentic AI systems, unlike traditional generative AI models, possess the capacity to analyse and respond to signals from their environment, plan actions, execute them, and reflect on outcomes, often operating continuously with varying degrees of autonomy and adaptive behavior.

While the current wave of Agentic AI has gained prominence with the rise of large language models and advanced machine learning systems, the underlying concepts are not entirely new. For decades, several scientific disciplines have explored systems capable of autonomous decision-making and interaction. Research in multi-agent systems and game theory studied how multiple autonomous entities coordinate, cooperate or compete within shared environments. Control theory has addressed feedback, adaptation and regulation in dynamical systems. Related work in robotics, distributed systems and human-computer interaction has also investigated how autonomous components can analyse signals from their environment, plan actions, and adjust their behavior over time.

The current notion of Agentic AI builds upon these foundations but emerges in a new technological context, where natural language interfaces enable broader deployment of autonomous digital agents. Its development and use represent a significant shift: with Agentic AI, AI tools are not only on a case-by-case basis and on command, but can perform the entire chain of task to achieve a given goal and therefore impulse sub-decisions that arise on the chain and orient them. This means that Agentic AI does not only pose the addition of challenges of every agent individually, but creates new, specific challenges, notably regarding **accountability, safety, traceability, and control**.

Because many digital tools and processes were designed around human operators and human-paced decision cycles, increasing agentic autonomy may change established operational, organizational, and societal paradigms, including impacts on labor and environmental footprint. Governance and regulatory framework must channel these impacts.

Agentic AI presents opportunities in streamlining, detailing, personalizing or accelerating processes. Recent advancements in agentic architectures have potential to fuel innovation in industrial automation, research assistants, personal productivity tools, and defence. The agentification of AI systems also support the automation of tasks such as data labelling, output scoring and comparison, instead of humans-based annotations and evaluation.

As the industry, policy makers, academia and potential users are in the process of identifying Agentic AI as an emerging trend and therefore haven't yet grasped all the uses and impacts, this paper aims to provide an understanding of what Agentic AI is. It also explores how it can be used and the ways in which it differs to comparable tools. It defends the position that Agentic AI is a tool that can be harnessed in order to ensure positive outcomes by limiting the potential negative effects, including on society and the environment. Rather than solely documenting the current landscape of Agentic AI adoption and impacts, this paper aims at advancing specific policy recommendations designed to shape the trajectory of Agentic AI development and deployment.

1. Technical concepts and paradigms

1. A Working definitions

While Agentic AI is currently being adopted, there is no single, broadly accepted definition. Indeed, the technology is still evolving and stakeholders are yet observing the shape it is taking, without the environments in which it is deployed. Therefore, in order to identify its key features and characteristics, in addition to looking at the practice, it is useful to look at the main definitions that have been worked at, as they highlight how Agentic AI is shaping.

According to academic researchers from Cornell University (USA) and University of the Peloponnese (Greece): "*Agentic AI systems, in contrast to AI Agents, represent a paradigm shift marked by multi-agent collaboration, dynamic task decomposition, persistent memory, and coordinated autonomy.*" [1]

According to Laurence Devillers, Professor of Artificial Intelligence at Sorbonne University (LIMSI/CNRS), "Agentic AI perceives, decides, and acts according to an assigned intention. It represents an entity capable of understanding high-level objectives and establishing a list of actions and then executing them, under human supervision or autonomously."[2]

In a comprehensive technical survey published by IEEE, researchers define Agentic AI as "autonomous systems designed to pursue complex goals with minimal human intervention" that demonstrate "adaptability, advanced decision-making capabilities and self-sufficiency, enabling [them] to operate dynamically in evolving environments," distinguishing it from traditional AI which "depends on structured instructions and close oversight."[3]

In a working paper published in February 2026 as part of its Artificial Intelligence Papers series, the OECD defines AI agents as "systems that can perceive and act upon their environment with a degree of autonomy, using tools as needed to achieve specific goals and adapt to changing inputs and contexts," and distinguishes Agentic AI as systems "composed of multiple co-ordinated AI agents that can break down tasks, collaborate, and pursue complex objectives autonomously over extended periods".[4]

All of the definitions underline the same main characteristics of Agentic AI systems:

- a level of autonomy to achieve a goal
- orchestrated multi-agent collaboration and connection with their environment

In order to further understand the risks and opportunities of Agentic AI in detail, it is necessary to explore the characteristics it entails and identify how it differs from other tools.

---

1 Sapkota at al. (2025) AI Agents vs. Agentic AI: A Conceptual Taxonomy, Applications and Challenges. *Information Fusion 126, 103599*

2 Devillers, L. (2024) Agents IA : nouveaux alliés ou fausse intelligence ?. *Le Hub.*

3 Acharya et al. (2025) Agentic AI: Autonomous intelligence for complex goals—A comprehensive survey. IEEE Access.

4 OECD (2026) The agentic AI landscape and its conceptual foundations. *OECD Artificial Intelligence Papers.*

1. B

Key characteristics and implementation pattern

The key characteristics outlined below draw on the white paper State of Agentic AI Security and Governance 1.0, a reference framework developed under the OWASP GenAI Security Project.[5]

Autonomy & execution: Agentic systems can be supervised across a spectrum from human-in-the-loop to fully autonomous operation, with explicit decision boundaries governing the scope of autonomous action.

Planning & reasoning: Agentic AI systems exhibit planning and reasoning capabilities, including task decomposition, iterative deliberation, and the ability to reflect on and self-correct outputs.

Memory & statefulness: Agentic systems maintain state across interactions, from immediate context within a single conversation (short-term memory) to persistent knowledge bases and operational histories (long-term memory), enables agents to reference prior interactions, recognize patterns, and adapt to evolving objectives over time.

Action and tool use: Agentic systems can take actions in external environments through tool integration, such as web browsing, API calls, code generation and execution, and domain-specific plugins, enabling capabilities that go beyond pure text-based responses.

Multi-agent coordination: In more advanced configurations, Agentic AI systems coordinate multiple agents through communication protocols, task delegation mechanisms, and collective decision-making processes, enabling parallelization and specialization.

In practice, these characteristics are realized through common architectural patterns. These patterns represent design approaches to implement autonomy, reasoning, memory, and tool use in deployed systems. They are not prescribed by a single standard but reflect common design approaches observed in deployed systems.

Functional roles in Agentic architectures

Recent Agentic AI systems often decompose functionality across multiple specialized agents or modules that collaborate through structured communication and task delegation. Although implementations vary, several recurring roles can be identified[6].

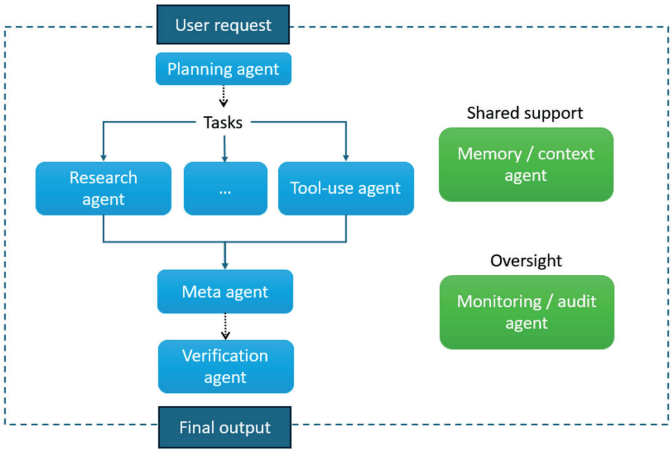


Figure 1: *Illustrative architecture of a specialized multi-agent workflow*

5 OWASP GenAI Security Project. (2025) State of Agentic AI Security and Governance 1.0. OWASP.

6 Abou at al. (2026) Agentic AI: A comprehensive survey of architectures, applications, and future directions. *Artificial Intelligence Review*.

Planning agents are responsible for goal decomposition and task orchestration, transforming high-level objectives into sequences of actionable steps. **Reasoning agents** explore candidate solutions using techniques such as chain-of-thought reasoning or iterative refinement. **Information-seeking agents** interact with external tools and environments, such as search engines, APIs, or web browsers, to gather relevant information. To improve reliability, some architectures introduce **verification or critic agents** that evaluate intermediate outputs and detect errors or inconsistencies. At a higher level, **meta-agents** may aggregate decisions from multiple agents or resolve conflicts between competing proposals. Other components focus on system state, such as **memory or context agents** that maintain conversation history and persistent knowledge, enabling continuity across tasks. Finally, operational concerns are addressed by **monitoring and audit agents**, which collect logs, ensure traceability of actions, and detect anomalous behavior. Together, these roles form a modular architecture that supports both flexibility and oversight in complex agentic workflows.

1.C

Boundaries between traditional software, AI Agents and Agentic AI

The boundaries between traditional software, AI Agents and Agentic AI are not always sharply defined in practice. Many real-world systems combine characteristics from multiple categories, and their autonomy and adaptability exist on spectrums depending on design and deployment choices. The distinctions presented here should therefore be understood as analytical reference points rather than strict classifications.

Traditional software systems operate according to fixed, pre-programmed rules and logic. Its behavior is determined by explicit rules: given identical inputs and state, it produces identical outputs. This determinism makes execution paths traceable and verifiable. Behavioral change requires explicit human intervention in the form of code modification or reconfiguration.

AI Agents, in the classical AI sense, are software entities that carry out tasks within an environment based on goals, rules, or learned behavior. They may be fully or partially autonomous and can range from very specialized to broadly capable, but their autonomy is often bounded to the execution of specific tasks assigned by humans.

Agentic AI refers to a paradigm for designing AI systems that can pursue goals autonomously, plan how to achieve them, remember past events, and use tools regardless of their underlying implementation (LLM-based, reinforcement learning, or hybrid). These systems adapt their behavior as situations change and can operate for extended periods with limited human intervention.

The role of a human also differs across these systems. In traditional software, humans specify the logic of the system and remain directly responsible for its operation and outputs. With AI agents, humans typically assign tasks, review intermediate results, and adjust the level of autonomy granted to the system. In Agentic AI systems, humans define goals, constraints, and guardrails, while retaining oversight mechanisms such as approval processes or override controls for high-impact actions.

1. D

I.D. Adoption and early use cases

To anchor our analysis in operational reality, we consulted 30 companies across G7 countries, spanning sectors ranging from energy and finance to software development. The objective was to assess the current state of adoption, identify strategic drivers, and quantify the barriers delaying the deployment of Agentic AI in professional environments.

The results depict a highly engaged but cautious industrial landscape: 83.3% of respondents expressed a "very strong" intention to explore or invest in agentic approaches, and every respondent anticipates at least a partial transformation of their business operations. This enthusiasm is consistent with broader industry surveys. A PwC reports that 79% of executives indicate their companies are already using AI agents in some capacity.[7]

A clear maturity gap nonetheless persists. While enthusiasm is high, the technology is largely viewed as "promising but immature" (53.3%). Only 23.3% consider that it is already stable enough for real-world use. Despite this, companies are moving quickly: 48.3% are already prototyping, and 34.5% have successfully moved agentic solutions into production.

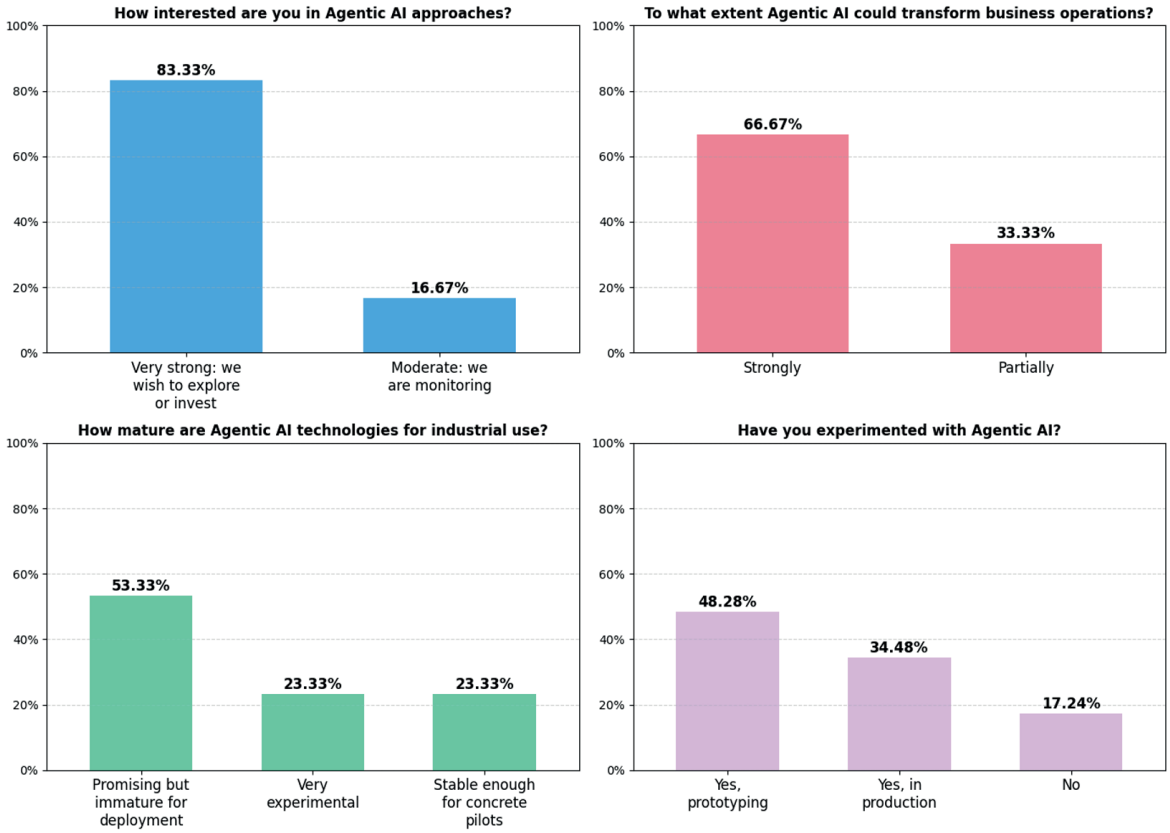


Figure 2: *General industry perspectives on Agentic AI*

7 *PwC. (2025, May).* AI Agent Survey. *PwC Tech Effect.*

Current Applications

Our survey suggests that adoption is currently concentrated in technical and support functions where human oversight remains central. "Design and Engineering" and "Analysis and Decision Support" are the leading application areas, reflecting strong demand for code generation, technical validation, and the synthesis of complex data for decision-makers.

Operational roles also show significant uptake, with "Interaction and Support" and "Maintenance and Quality Assurance" (e.g., defect detection). "Operations and Production" represents a smaller proportion, likely due to the higher reliability requirements in live production environments.

Other specific use cases mentioned by respondents beyond these categories include:

- Automated compliance: generating specifications and validating source code against specific standards.
- Streamlining human resources management.
- Technical documentation: automating the creation and maintenance of complex technical documents.

The distribution of use cases observed among our respondents is consistent with trends reported more widely: Agentic systems are finding early traction in domains where tasks are structured, outputs are verifiable, and failures are recoverable. In customer-facing functions, the top-cited application area in PwC's enterprise survey, agents are deployed for intelligent service orchestration, autonomously handling interactions and escalating only ambiguous or high-stakes cases to human operators. In financial services, Agentic AI is being used for real-time risk monitoring and compliance automation: a Moody's survey of 600 global risk and compliance professionals found that 26% are already actively using Agentic AI models in their workflows.[8]

In software engineering, Agentic systems are being integrated into CI/CD pipelines as autonomous quality gates, capable of generating test cases, detecting regressions, and proposing corrective patches.

Barriers to adoption

While technical integration remains a challenge, the primary obstacle to deploying Agentic AI is trust. 43% of the respondents identify "Reliability or Control risks" as their single most significant barrier, followed by "Difficulty integrating with existing systems" (~17%) and "Cybersecurity and regulation issues" (10%). Interestingly, the "lack of concrete use cases" is rarely cited as a primary blocker, suggesting that companies generally understand where to apply the technology, but hesitate on how to do so safely.

Qualitative insights from respondents further clarify the specific challenges they face:

- Opacity of commercial models: the closed nature of many industrial-grade models prevents the deep auditing required for regulated sectors.
- Incremental strategy: some organizations prefer to stabilize their use of standard Generative AI before introducing the added complexity and autonomy of agentic systems.
- Cost of error: concerns regarding the expanded attack surface of autonomous agents led some respondents to restrict their use to non-critical environments where errors are manageable, a concern reinforced by a broader Infosys study reporting that 95% of companies experienced at least one AI incident in 2025, with 77% resulting in financial losses.[9]

8 *Moody's. (2025, October).* Navigating the Shift: How Agentic AI is Reshaping Risk and Compliance. *Moody's Analytics.*

9 Infosys Knowledge Institute. (2025, August). Responsible Enterprise AI in the Agentic Era.

Overall, organizations are not waiting for full technological maturity, but are progressing cautiously by limiting scope and maintaining human supervision. The transition toward broader deployment therefore depends less on identifying new applications than on establishing sufficient reliability, traceability, and accountability in operational settings.

Energy impact

Energy consumption is widely recognized as a growing concern in the deployment of Agentic AI systems. 83% of respondents consider energy impact as a topic of interest, yet only 47% report having metrics or established methods to measure the energy consumption of their AI models of infrastructure.

2. Conditions to make the deployment of AI a success

2. A → Traceability and explainability

Traceability and explainability refer to the ability to reconstruct and understand the outputs and decisions produced by an Agentic AI system, which is necessary in regulated contexts and when failures occur. This requires explanations of the reasoning process that remain comprehensible to humans.

Some limitations are intrinsic to the system's own generation and internal reasoning processes, and may affect outputs even when the surrounding environment remains unchanged. Long-term memory can unpredictably influence results, making it difficult for auditors to reconstruct the prior context in which a decision was produced. In multi-agent settings, unintended reasoning patterns may also emerge: behaviors can propagate across interconnected agents, while attribution and reproducibility become difficult to maintain. Additionally, the probabilistic nature of many underlying AI models means identical inputs may not always produce identical outputs due to differences in randomization parameters. Differences in the configuration of the model-hosting hardware may also affect the output.

Other risks are extrinsic and arise from changes in the environment around the agentic system. Such dynamics can silently alter the evidence considered, the reasoning path followed, and the reproducibility of outputs even when user inputs appear identical. For instance, silent updates of upstream models such as language models or embeddings may invalidate prior explanations. The evolution of external data sources, including web or news content, can likewise prevent reliable auditing. Finally, network dependencies introduce additional fragility, as connectivity interruptions may lead to missing evidence and incomplete audit trails.

Strategies for successful adoption in this regard rely on the set of core controls:

- **Comprehensive logging of raw** prompts and system messages, memory reads and writes, tools and API calls with inputs and outputs, and intermediate reasoning traces
- **Versioning of all components**, including models and decoding parameters (e.g. temperature, top-k/top-p, random seeds, and hardware configuration)
- **Controlled randomness**, supporting fixed random seeds where possible and allowing deterministic decoding modes (e.g. greedy).

2. B

Controllability

Controllability refers to the ability of human operators or governance mechanisms to reliably influence, constrain, and redirect the behavior of an Agentic AI system. In systems capable of autonomous planning, reasoning, and tool use, decisions may unfold over long chains of intermediate steps and interactions with external tools or environments. Maintaining controllability therefore requires mechanisms that allow humans to supervise execution, enforce operational constraints, and intervene when necessary.

Some limitations arise from the internal dynamics of agentic systems themselves. During multi-step reasoning or planning, agents may reinterpret instructions, generate intermediate objectives, or adapt their behavior based on environmental feedback. Over time, this process can lead to gradual divergence between the system's operational objective and the user's original intent. In multi-agent settings, coordination between specialized agents may further amplify such deviations, as intermediate decisions propagate across interconnected components. Long decision chains and distributed reasoning can therefore make it difficult to identify when and where human intervention should occur.

Other evolutions are extrinsic and relate to the operational environment in which the system is deployed. Access to external tools, APIs, or autonomous web interactions may allow the system to trigger actions whose consequences extend beyond the original task. Changes in system configuration, updates of safety policies, or differences in deployment context may also alter how and where control can be exercised.

The concept has strong parallels with ideas from control theory, where a system is considered controllable if external inputs can drive its state toward desired configurations. While Agentic AI systems do not follow the dynamical systems typically studied in classical control theory, the underlying analogy remains relevant: humans must retain the ability to shape the trajectory of system behavior through supervisory inputs and constraints.

Strategies for successful deployment regarding controllability therefore rely on explicit control mechanisms embedded throughout the system architecture:

- **Architectural control points**, enabling inspection and intervention at key stages of the decision process
- **Constraint enforcement mechanisms**, such as policy layers, sandboxed execution environments, and access controls for external tools
- **Clear allocation of control authority**, defining which actors are responsible for setting objectives, enforcing constraints, and supervising system operation, including hybrid configurations where control is shared between automated mechanisms and human oversight

- **Regular testing of control mechanisms**, ensuring that intervention procedures remain effective and that operators are prepared to respond to abnormal behavior. In the context of the EU, article 24 of GDPR constitutes a regulatory obligation of this approach.[10]

Currently, sufficiently robust technical approaches of these kinds remain limited in scope and implementation, and we are not confident that technical progress in these areas will be able to keep up with the rate of autonomous AI capabilities progress. Integration into critical systems requires thus a careful risk assessment and conservative deployment practices.

2.C

Safety and harm prevention

The use of LLMs introduces their specific risks into software systems, and Agentic AI may amplify these risks through iterative planning loops, autonomous tool execution chains, persistent and evolving memory, and emergent behaviors.

Fabricated outputs can propagate across multiple steps, leading to a cascade of erroneous actions or persistent knowledge that disrupts decision-making. Bias may also accumulate in memory or spread among agents, gradually normalizing over time and influencing future plans. Through reflection and memory updates, systems may progressively alter their objectives and drift away from the original human intent.

Agentic workflows are also vulnerable to adversarial instructions: malicious prompts can persist across steps and hijack the goals of the agents or the tools they rely on. In addition, persistent memory may resurface private information across interactions, creating potential privacy breaches and regulatory violations.

Because Agentic AI systems can be deployed rapidly through online services or software integrations, unintended behaviors or vulnerabilities may quickly affect a large number of users, amplifying the scale of potential harm before safeguards are fully tested.

Successful Agentic AI adoption requires controls that limit error amplification, prevent misuse, and protect personal data are key, to tackle potential safety and harm risks. Key measures include:

- Sandboxing and containerization: Deploy agents in isolated environments with restricted access to data sources, APIs, and system resources until behavior is verified safe.
- Multi-source output validation and feedback loops to detect and correct hallucinations and compounding errors.
- Model redundancy: Avoid reliance on a single model architecture by implementing fallback mechanisms to ensure system reliability and continuity.
- Regular testing for bias in embeddings and output.
- Periodic restatement of system objectives.
- Separation of user inputs from system instructions, with monitoring for prompt injection patterns.
- Strict tool access verification based on least-privilege principles.
- Personal data minimization and retention time limits.

10 European Union. (2016) General Data Protection Regulation (GDPR). Regulation (EU) 2016/679

2. D

Multi-agent risk factors

Multi-agent risk factors arise when multiple agents interact, coordinate, or delegate tasks to one another, creating additional points of uncertainty beyond single-agent behavior. Once decision-making is distributed across agents, it becomes more difficult to attribute responsibility, observe the full decision path, or predict emergent behaviors that result from their combined actions rather than any single agent in isolation.

In such settings, a compromised or biased agent may steer collective decisions and lead to systemic failure. Joint decision processes can also produce outcomes that cannot be clearly traced back to any individual agent's intent or logic. In addition, faulty outputs generated by one agent may be accepted and escalated by others without detection, allowing errors to propagate and amplify across the system.

Multi-agent systems are also exposed to external risks. For example, the failure of a single external service or API may trigger retries or improvisation across several agents simultaneously, turning a localized disruption into a broader system-wide failure.

To ensure **successful adoption**, potential risks in multi-agent systems requires controls that limit error propagation, reduce manipulation of collective decisions, and constrain emergent behavior. Mitigation strategies include:

- Mechanisms to **prevent any single agent from dominating group decisions**.
- **Intermediate validation and verification checkpoints** between agents to detect and contain errors.
- **Observer agents** to monitor collective behavior.
- **Capacity agreements** with external service providers.

While the proposed mitigation measures can improve observability and reduce certain risks, they do not fully address the inherent limitations described in this section. In particular, challenges related to autonomy, non-determinism, and system interactions cannot be fully mitigated through current technical controls. As a result, in higher-risk deployment contexts, additional safeguards may be necessary, including mandatory human oversight at key decision points, constrained deployment settings, or delayed deployment until better controls exist.

2. E

Energy and resource efficiency

While assessing the full impact of AI systems on energy and resources is already a central challenge, Agentic AI significantly raises the level of complexity. Its reasoning-driven operation introduces specific and less predictable demands on both energy consumption and resource efficiency.

Agentic workflows and energy efficiency

Agentic workflows can operate continuously across multiple steps and extended durations, with agents persistently querying databases, invoking APIs, and updating their memory throughout the process. These workflows enable cascading decision-making, where the output of one reasoning informs subsequent actions, each with its own energy impact. Furthermore, the recursivity of the reasoning implies the possibility of agents getting "stuck-in-a-loop": they would be repeatedly re-evaluating or re-executing similar steps without converging on a solution. In addition to hampering efficiency, it can increase the environmental footprint due to prolonged compute usage and the potential need to restart the entire process after failure detection.

Deadlocks and infinite loops also happen in traditional software systems. However, Agentic workflows are broader and more complex: agentic systems interact with external tools, trigger API calls, coordinate across distributed services, and operate with a higher degree of autonomy. As a result, deadlocks or runaway action chains may be harder to detect and correct as it can propagate across interconnected systems, platforms, and services.

At the same time, assessing the specific energy footprint of AI systems will become increasingly difficult. As Agentic AI integrates into the wider digital infrastructure, its computational and energy consumption will be entangled with that of cloud services, data centres, APIs, and network operations. Hence, distinguishing the energy impact of AI from the one of the digital sector as a whole will therefore become more challenging.

Resource pressure

Agentic AI's resource needs go beyond what energy metrics alone reveal. Scaling AI factories, cloud platforms, and edge nodes pulls on a much broader chain of costs: rare mineral extraction, shorter hardware lifecycles and significant water usage for cooling. Beyond this, agentic workloads are increasingly distributed across the full computing continuum, from cloud to edge to end devices, multiplying resource allocation inefficiencies rather than consolidating them. The fast-moving evolution of interconnection layers is likely to accelerate hardware turnover faster than efficiency gains can compensate. Without better models to anticipate these cascading demands, resource efficiency remains something to react to rather than plan.

Efficiency and optimization under constraints

In complex operational environments such as Agentic AI systems, entropy can be understood as the number of possible configurations of action: a highly entropic system would exhibit a large number of possible suboptimal configurations of action. Reducing entropy requires to observe the system and process information to select the most efficient path of action. The trajectory of the digital sector may depend on whether it develops freely or under constraint: energy scarcity, rising raw materials costs or strict sustainability requirements could reduce the number of acceptable configurations by eliminating energy-intensive ones or unnecessary complexity. In this context, Agentic AI may outperform human-managed processes in task efficiency and information processing can become a lever for energy efficiency.

According to the International Energy Agency, scaling up existing AI applications to the sectoral level across energy end-use sectors could unlock up to 13.5 EJ in energy savings by 2035, which represents 3% of today total energy consumption.[11] However, significant uncertainties remain: the pace of AI adoption, potential rebound effects, and broader resource pressures on water, minerals and biodiversity could undermine these efficiency gains.

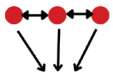
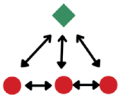
11 IEA (2026), Key Questions on Energy and AI

Energy impact of multi-agents' architecture choice and agents' conception

The energy impact of Agentic AI systems is linked to the computational complexity. In multi-agent systems, coordination and consensus mechanisms influence performance and energy efficiency. As depicted by Table 1, hybrid architecture, combining hierarchical oversight and peer-to-peer coordination between agents, are expected to have greater energy impact as it exhibits higher LLM calls, communication overhead and memory complexity.

Consequently, agentic systems in which an orchestrating agent coordinates multiple sub-agents operating in peer-to-peer coordination model or agentic systems autonomously spawning new agents risk dramatically amplifying computational complexity and, by extension, energy consumption.

Besides multi-agent architecture choice, agents' model and architecture also matter: smaller, well-tuned LLMs with information preprocessing can deliver equal or better task performance than large models while consuming less energy. Krupp et al. demonstrates this empirically through a benchmark of five open-source web agents evaluated on identical tasks, revealing a variation in energy consumption by a factor ten across systems. Notably, the most energy-efficient agent also delivered the strongest overall task performance.[12]

	Characteristic	Interaction Type	LLM calls	Communication overhead	Memory complexity
Single agent system	–		O(k)	0	O(k)
Multi-agent system	Independant	 Agregation	O(n · k)	1	O(n · k)
	Decentralized	 Majority voting	O(d · n · k)	d · n	O(d · n · k)
	Centralized		O(r · n · k) + O(r)	r · n	O(r · n · k)
	Hybrid		O(r · n · k) + O(r) + O(p)	r · n + p · m	O((r+p) · n · k)

Agent

Orchestrator agent

n : number of agents
k : max iterations per agent
d : debate rounds
r : orchestrator rounds
p : peer communication rounds
m : average peer requests per round

The notation O(x) denotes an asymptomatic upper bound that can be understood as : "the complexity/LLM calls do not grow faster than a constant multiplied by x as x become large".

Table 1: Comparison of complexity metrics between different multi-agent architectures.
Source: adapted from Google Research[13]

12 Krupp et al. (2025), "Promoting Sustainable Web Agents: Benchmarking and Estimating Energy Consumption through Empirical and Theoretical Analysis"

13 Kim et al. (2026), "Towards a Science of Scaling Agent Systems"

Energy consumption of reasoning-enabled models:

While some tasks can be handled by conventional AI models, either generative or deterministic, Agentic AI may also rely on reasoning models with less clearly defined energy impacts.

For non-reasoning AI models, a correlation appears between the growth in the number of parameters and the rise in energy consumption. Energy reduction strategies include Small Language Models (SLMs), model compression techniques, pruning techniques, quantization techniques and careful model selection evaluating the trade-off between task-specific accuracy and energy consumption.[14, 15] It is worth noting that for high-intensity usage, large sparse models with better quantization can sometimes prove more energy-efficient than SLMs.

Agentic systems operate through multi-step reasoning processes that generate and explore multiple solution paths simultaneously, each with distinct computational costs and energy footprints. In the context of information theory, the uncertainty of the decision-making process can be formalized through the concept of entropy which measures the degree of unpredictability of possible choices.[16] The volume and diversity of information processed during agentic workflows generate high initial entropy and require costly computational reductions to identify optimal reasoning paths. As highlighted by Ebert et al., the energy impact prediction is more cumbersome for reasoning models as it depends on the model size as well as on the path and the intensity of the reasoning, particularly on the number of tokens generated for it. Their preliminary analysis indicates that reasoning models consume, on average, 30 times more energy than traditional models.[17] Figure 1 depicts the GPU energy consumption increase with and without reasoning capabilities for three specific models.

Existing compression methods fail as an entropy conflict emerges: compression-oriented training lowers entropy, resulting in shorter reasoning and reduced exploration, while accuracy-driven optimization raises entropy by lengthening chains of thought and improving performance.[18]

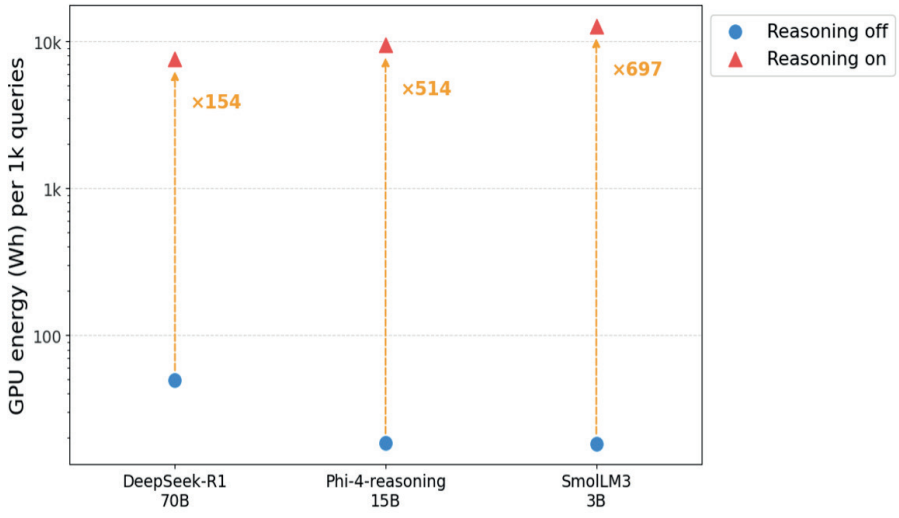


Figure 2: *GPU energy consumption with and without reasoning across three models.*
Data source: Hugging Face

14 Barros et al. (2025), "Small Is Sufficient: Reducing the World AI Energy Consumption Through Model Selection", *ArXiv:2510.01889*.

15 Varoquaux et al. (2025), "Hype, Sustainability, and the Price of the Bigger-is-Better Paradigm in AI", *Association for Computing Machinery*.

16 Here, entropy refers to the concept defined in Information Theory, while the concept of entropy discussed in the section 'Efficiency and Optimization Under Constraints' is related to the thermodynamic definition.

17 Ebert et. al. (2026), "The Global Landscape of Environmental AI Regulation: From the Cost of Reasoning to a Right to Green AI"

18 Zhu et al. (2025), "Entropy-Guided Reasoning Compression", *arXiv:2511.14258*.

Black-box AI and the loss of awareness

The proliferation of Agentic AI systems functioning as "black boxes" can diminish societal awareness of using AI, thus diminish the awareness of potential risk. This poses risks in terms of security, but also on energy consumption: users and organizations may lose sight of the computational resources they consume and the carbon footprint associated with continuous or large-scale AI operations.

Strategies for successful adoption:

- **Assessment of models' performance and energy impact** prior to deployment, with a focus on integrating energy consumption metrics into existing performance benchmarks.
- Selection of **energy-efficient data centres** powered by sustainable energy sources.
- Implementation of robust **loop detection and termination criteria** to ensure sustainable and stable long-term operation.
- **Calibration of reasoning depth to task complexity and specialized optimization strategies** for reasoning models such as entropy-guided reasoning compression.
- Development of **evaluation framework and processes** to assess the energy consumption of Agentic AI systems across their operational lifecycle.

2. F

Impact of introducing Agentic AI to the workforce

As Agentic AI becomes increasingly integrated in organizations, it is equally important to **anticipate the evolution** within the context of work and employment. Beyond technical level implications, Agentic AI is reshaping how tasks are performed and decisions are made within organizations. The risks discussed above, such as traceability, explainability, and multi-agent factors, also extend into the workplace. They influence production processes, workers' roles, experiences, and interactions with AI-enabled systems. Therefore, this section discusses the implications of Agentic AI on worker experience, emphasizing how its integration can be managed to support productive and equitable work environments.

Responsibility

As Agentic AI systems independently make decisions, assigning responsibility for their outcomes becomes increasingly complicated.[19] In cases of discriminatory hiring decisions, financial losses, or other harms, accountability may be unclear as it raises questions about whether responsibility lies with developers, deploying organizations, or the AI system itself. Without clear accountability frameworks, employees may experience moral distress, fear of blame, or reluctance to challenge AI-generated recommendations. Over time, this uncertainty can erode employees' sense of agency and professional identity, weakening trust in both organizational leadership and AI-enabled systems.[20, 21] This can also cause uncertainty from a judicial perspective, making it difficult to know who is responsible in case of conflict or illegal outcomes.

19 Nisa, et al. (2025), Agentic AI: The age of reasoning—A review.

20 Acharya et al. (2025), Agentic AI: Autonomous Intelligence for Complex Goals—A Comprehensive Survey.

21 Narasima Venkatesh, (2025). Future of Work: Managing Ethical Challenges of Agentic AI and Super Intelligence in Organizations.

Transparency and explainability

As explained in the previous section, as Agentic AI systems grow more complex and autonomous, the ability to reconstruct and understand the outputs and decision-making process may become opaque. This creates "black box" systems that employees cannot easily understand or challenge.[22] Lack of explainability results in employees struggling to interpret evaluation outcomes or strategic directives generated by agentic systems. This opacity can foster mistrust and weaken employees' willingness to rely on or collaborate with AI.

Possibility of skill erosion

In particular, the emergence of Agentic AI signals profound changes in the software development workforce. Developers are increasingly transitioning from architects of detailed instruction sets to designers of agent objectives, constraints and verification mechanisms. This role transformation carries risks of skill loss, particularly for junior developers who may not have learned deep understanding of underlying systems if they work primarily through high-level agent specifications. Programming languages and development practices may evolve or become less central as systems increasingly self-organize through agent coordination. Yet this picture deserves nuance. Technology has historically raised the baseline of human performance even as it surpasses humans at specific tasks. The critical variable is quality of engagement: developers who use agentic tools to challenge their thinking are likely to grow, while those who defer passively risk atrophying their own judgment. Agentic AI may displace certain expertise, but equally open space for new competencies in oversight and system-level reasoning.

Valuing human contribution

The productivity gains afforded by Agentic AI raise urgent questions about how value is created, measured, and shared within organisations. Efficiency gains should be equitably distributed across the workforce, given that such gains typically reflect collaboration between worker and system rather than AI capability alone. This raises a deeper measurement challenge: standard performance metrics were designed for environments where human output is visible and attributable, but in agentic workflows much of the observable work is AI-generated, obscuring the quality and nature of human contribution. Organisations will need to develop new evaluation frameworks that can distinguish the value of human judgment, oversight, and critical engagement from the volume of automated output.

Strategies for successful deployment:

- Establishing a comprehensive ethical AI governance framework.
- Ethics-focused workforce including AI literacy.
- Providing targeted ethics and responsible AI training.
- Promoting interdisciplinary education and collaboration.
- Enforce human oversight.
- Collective bargaining through unions to ensure recognition of new forms of work.

22 Murugesan, S. (2025). The Rise of Agentic AI: Implications, Concerns, and the Path Forward.

3. Regulatory frameworks (focus on the EU context)

Agentic AI falls under legal definitions of AI systems, notably the one set by the EU AI act (art. 3 (1))[23], meaning that the obligations of regulations such as the AI Act apply to it, thus establishing a framework for adoption.[24] Thus, applicable law address certain issues related to Agentic AI, particularly regarding liability, competition, security, and data protection. When incorporated from the design phase, this framework can enable the deployment of tools that are compliant and respectful of users and consumers. However, since Agentic AI exhibits characteristics that were unknown at the time these standards were drafted, analyses of existing law in light of this technological evolution could shed light on potential legal gaps or obstacles to its responsible deployment.

3.A → Data Protection

The multi-agent nature of the Agentic AI systems and their openness to the outside world can pose **evolutions to anticipate**.

Data use and retention by actors not included in the consent: new participants or AI agents in the ecosystem may gain access to personal data even though they were not initially authorised, leading to broader data sharing than originally consented to. The dynamic interaction between multiple agents may make it difficult to ensure that only authorised entities access personal data. Once data is shared across multiple agents or components, ensuring deletion at the appropriate time may become difficult, even more since the evolving and autonomous nature of Agentic AI systems makes it difficult to determine how long data should be stored.

Silent expansion of collected data and its use: Agentic AI systems may process large volumes of data, including information that is not strictly relevant or necessary to achieve the intended purpose, or alter the initial purpose of processing without the user's awareness. In addition, AI agents may generate new personal data by combining or inferring information from existing datasets, leading to an infringement on the relevance of the data and the principle of data minimization. The dataset collected and used may gradually expand to include third-party data (e.g., colleagues, contacts), beyond what is necessary for the initial purpose. Over time, Agentic AI could incrementally shift data usage beyond the original consent, without user awareness.

The inability to ensure compliance with the safeguards provided for under the GDPR: decisions taken by Agentic AI may qualify as solely automated decisions producing legal or similarly significant effects (article 22 GDPR), triggering strict GDPR conditions and safeguards (human oversight, transparency, contestability), which may be difficult to ensure in practice.

23 European Union. (2024). Artificial Intelligence Act: Laying Down Harmonised Rules on Artificial Intelligence.
24 In this paper, the regulatory part focuses mostly on the EU context, as an illustration.

Strategies for successful deployment

Within the European Union, the GDPR applies and must therefore be taken into account throughout the system's entire lifecycle. It presents elements that can help limit foreseeable risks:

- Appoint a "controller", the "natural or legal person, public authority, agency or other body which, alone or jointly with others, determines the purposes and means of the processing of personal data" (art. 4 (7) GDPR). According to the Spanish Data Protection Agency, "*The controller is the person who (...) determines the purposes and means of the processing (...).*" Applied to Agentic AI, they suggest they "*Ensure regulatory compliance, manage new risks in processing that could arise from the use of Agentic AI systems and analyse the proportionality of critical impacts that could arise from using agents.*"[25]
- Since Agentic AI systems are connected to third-party services, the controllers of various related agents should collaborate to ensure end-to-end coverage. More generally, ensure GDPR compliance by third parties connected to the system and change partners if this is not the case.
- Conduct tests and report incidents to demonstrate legal compliance (Article 24 of the GDPR).
- Assess the utility of datasets in terms of legal compliance, but also in terms of utility.
- The interface can include a clause for consent that explains data use and potential risks; this clause could also usefully specify explicitly when the user is interacting with an Agentic AI, in the interest of transparency.[26]

3. B

Competition law

Agentic AI can have a **transformative effect** on business by significantly impacting competition. As such, it is subject to relevant regulations, such as the Digital Services Act (DSA) and the Digital Markets Act (DMA).[27]

Market transparency: The requirement for transparency of information, which is fundamental to a market economy, may be compromised, as choices are steered toward one good or service over another in an opaque manner.

Competition: Excessive concentration may arise from the introduction of Agentic AI: given the barriers to entry in the Agentic AI market, a player controlling the entire market chain could gain a comparative advantage that undermines free competition.[28] Dominant players could use existing comparative advantages to develop Agentic AI that outperform their competitors, thereby locking down the market, contrary to Article 5 of the DMA, which prohibits gatekeeper practices.

Collusive effects: If Agentic AI were responsible for pricing or logistics, unintended collusive effects could arise, as pursuing similar optimization objectives could lead to a convergence of prices and sales terms, in contradiction with Art. 39 of the DSA treating the analysis of systemic risks and Art. 5 DMA that prohibits restrictive practices.[29]

---

25 Spanish agency for data protection. (2026). Agentic Artificial Intelligence from the perspective of data protection.

26 Spanish agency for data protection. (2026). Agentic Artificial Intelligence from the perspective of data protection.

27 Regulation (EU) 2022/2065 of the European Parliament and of the Council of 19 October 2022 on a Single Market For Digital Services and amending Directive 2000/31/EC (Digital Services Act) and Regulation (EU) 2022/1925 of the European Parliament and of the Council of 14 September 2022 on contestable and fair markets in the digital sector and amending Directives (EU) 2019/1937 and (EU) 2020/1828 (Digital Markets Act).

28 Bouffier et al. (2026), "*Intelligence artificielle : regards croisés du CIANum et de l'Autorité de la concurrence*", Racine avocats.

29 Deffains (2025), "*L'IA agentique : du changement d'échelle technologique au questionnement du modèle juridique classique*", Dalloz Actualités.

Strategies for a successful deployment:

- The creation of commons for the development of Agentic AI can help establish a level playing field, encouraging the emergence of a diverse range of market participants.
- An assessment of the risk of monopolistic situations arising, with regard to consumer benefits, could be undertaken[30].
- Incorporate a transparency requirement for agentic systems regarding the moderation and recommendation criteria that can influence consumer choices (Art. 17 DSA).
- The implementation of licenses to sell Agentic AI systems or to interact with such systems (for example, through certification of these tools) could ensure their quality as well as their interoperability. These licences must be adapted to Agentic AI.
- The DSA includes regulations protecting consumers in this context: Article 34 requires providers of large online platforms and search engines to identify, analyse, and assess any systemic risks arising from the design or functioning of their services and related systems. As a result, where AI agents form a part of the platform's technical infrastructure or decision-making processes, the risks associated with their deployment must be taken into account in the platform's mandatory risk assessment. This assessment must be tailored to the service and proportionate to the severity and likelihood of systemic risks. Consequently, when AI agents are used by large platforms, they become part of the regulatory perimeter of the DSA through the platform's overarching obligation to assess and mitigate systemic risks.

3.C → Liability

The use of a digital system as an intermediary in the performance of tasks may present profound **evolutions** regarding liability. The higher the degree of autonomy granted to the system, the greater the risk of damage. This concern is particularly acute in the case of autonomous AI agents interacting with third-party APIs.

Attribution of liability: Since Agentic AI is interconnected and open to the outside world, it can be complex to identify the source of harmful behavior and thus attribute liability. The AI Act specifically imposes obligations of transparency, explainability, and traceability on high-risk systems, particularly to enable the identification of the source of a failure. However, while this obligation reduces computational opacity, it does not eliminate it. In the case of agent-based AI, responsibilities can be prioritized, given that decisions arise from continuous interactions between agents, data, and infrastructure.

Intellectual property: Similarly to liability, in the case of value-added creation, it can be difficult to identify who holds the intellectual property rights: the user, the programmer, or even the agent-based AI system. In addition, involuntary IP infringements by agentic systems also present a risk, in relation with the data processed (see V.1).

Lack of transparency and explainability: Like any AI system, AI agents may generate hallucinations or inaccurate outputs, which can produce cascading effects when such outputs are shared among multiple agents responsible for different tasks. The lack of transparency and explainability may further complicate the attribution of liability.

30 European Union. (2025). Agentic AI: Leveraging European AI talent and Regulatory Assets to Scale Adoption. Shaping Europe's digital future.

Territorial scope: Finally, legal uncertainty may arise if an agent-based AI system involves interaction between agents subject to different jurisdictions, as the territorial scope of the system may be difficult to determine.

Strategies for a successful deployment:

- For Agentic AI, risk prevention and regulatory oversight must be implemented ex ante. This implies a shift: regulators must therefore focus on the infrastructure itself under a "code is law" approach.
- Conduct ex ante risk assessment and test the agentic system in real life conditions.
- Appropriate and mandatory insurance could be required for certain agent-based AIs posing specific risks.
- Contract law could warrant particular attention regarding liability: if intermediary agents were to enter into agreements, the prerequisites and liability regime could be examined..

3. D

→ Governance challenges and gaps

As noted above, existing regulations already address a number of evolutions Agentic AI might bring. A challenge for regulators stems from the need to create a framework secure enough to ensure certainty and trust for industrial partners, while not creating barriers to innovation and adoption. A particular risk stems from the fact that Agentic AI adapts its behavior to its environment and operates in conjunction with third parties[31]. Uncertainty remains with regards to the novelty of the technology.

Agentic AI as a "deputy" to professional and non-professional users: laws regulating AI did not anticipate that people would "delegate" decisions or powers (such as contracting) to Agentic AI, which raises certain questions, such as:

- Article 5 of the AI Act addresses the manipulation of human behavior that can induce individuals to make harmful decisions. However, in the case of "delegation," Agentic AI does not alter human behavior but acts for them, potentially harming its interests, for example, through the injection of malicious prompts. This case appears as a blind spot of this article.
- Certain obligations, particularly those relating to human oversight (Article 14 of the AI Act), apply only to the system developer and the professional deployer, thereby excluding non-professional deployers, who may, however, use Agentic AI.
- Based on the assumption of human-machine interaction, the transparency requirement under article 50 of the AI Act appears ill-suited to scenarios in which Agentic AIs interact with one another or with entities such as e-commerce sites.

Size of agents vs. agentic systems: Legal gaps might appear if AI agents are taken into account individually, and not as an agentic system. For instance, the AI Act imposes certain risk assessment if the cumulative amount of computation used for its training measured in floating point operations is greater than 10^{25} In this case, what if a single AI agent crosses the 10^{25} threshold? Does the entire agentic system fall under systemic risk category? On the contrary, could there not be systemic risks when small agents interact?

31 Bellogín et al. (2025), "Systemic Risks Associated with Agentic AI: A Policy Brief", Europe Technology Policy Committee.

Potential ripple effects: the introduction of Agentic AI in environment such as the workplace might generate drastic changes in the allocation between capital and work, thus shifting the point at which value is created. This could have many ripple effects regarding the current policy framework: for instance, it might affect the level of fiscal income.

Strategies for a successful adoption: Policy makers play a critical role in adding legal specifications that could ensure a secure and successful deployment of Agentic AI. By involving all actors (industry, academics and users), the trust in the framework and its comprehensiveness could be further ensured.

- A dedicated study of applicable regulations could be undertaken to identify potential legal gaps, necessary adaptations, or specific interpretive guidelines to be developed. Dedicated research programs could thus be usefully planned, within programs such as EU Horizon.
- Dedicated supervision and testing methods could be developed.
- Specifically, regarding the AI Act, the Europe Technology Committee suggests the following adaptations:
    - → Amend article 9 to include Multi-Agent interaction risk assessment.
    - → Amend article 55 for providers regarding how their models could enable harmful agentic behavior contributing to systemic risk.
    - → Add a new article: "Ecosystem Safety and Multi-Agent System Testing."
    - → Strengthen article 15. Require Multi-Agent-specific cybersecurity audits.
    - → Expand article 5 to prohibit tacit collusion and covert channels (see part II).
    - → New liability Clause: Collective accountability for emergent harm.
- Soft law instruments also present a useful tool, as they provide guidelines and support in the adoption. On the international level, the OECD AI Principles[32] and the UNESCO Recommendation on the Ethics of Artificial Intelligence[33] provide widely endorsed soft-law instruments. These emphasize principles such as transparency, accountability, fairness, and human oversight. These soft law instruments can constitute useful guidelines. The recent OWASP State of Agentic AI Security and Governance 1.0[34] report highlights new, nuanced risks specific to agentic systems, most notably threats tied to autonomy, memory, and inter-agent dynamics, that aren't fully addressed in current regulatory tools like the EU AI Act or OECD Principles.
- Another avenue consists in adopting dedicated national strategies, such as the national authorities in Singapore did. The aim of this strategy is to "(help) organisations define agent boundaries, identify risks, and implement mitigations such as agentic guardrails."[35]

32 OECD. (2019). Recommendation of the Council on Artificial Intelligence.
33 UNESCO. (2021). Recommendation on the Ethics of Artificial Intelligence.
34 OWASP GenAI Security Project. (2025). State of Agentic AI Security and Governance 1.0. OWASP.
35 Infocomm Media Development Authority (2026) Singapore Launches New Model AI Governance Framework for Agentic AI

4. Recommendations

Priority 1 · Ensure strategic autonomy

The current Agentic AI landscape is dominated by a small number of providers. Decision makers can implement a strategy to ensure domestic Agentic AI solutions are managed yet flexible across the entire stack (foundation models, orchestration platforms, data infrastructure, and execution environments), while diversifying suppliers to avoid strategic dependency and ensure long-term autonomy.

Priority 2 · Establish a clear and operational definition of Agentic AI

A working definition grounded in the technical specificities of Agentic AI systems is necessary to provide a stable regulatory anchor. This definition should evolve through continuous technical dialogue to avoid definitional drift as capabilities expand, while remaining precise enough to guide policy, risk classification, and compliance obligations. It needs to be implemented by a consensual authority, to be largely adopted.

Priority 3 - Foster a fertile ground for Agentic AI research

Given the rapid evolution of the field, it is essential to establish agile research funding mechanisms. Allocating short-term grants to targeted areas, reflecting the inherent uncertainty of Agentic AI research, would enable timely generation of critical knowledge. Researchers must have access to state-of-the-art tools and platforms. Research must be inherently interdisciplinary, ensuring all dimensions and impacts of the technology are explored, and include dedicated efforts to address the scientific gaps identified above.

Priority 4 · Ensure legal security for users and developers

The coverage of Agentic AI by current legal and regulatory frameworks warrants dedicated analysis, including competition law and data protection regulations. The EU's regulatory framework (AI Act, DSA, DMA, and Omnibus Directive) could be examined to assess whether the risk-based approach adequately addresses Agentic AI's specific characteristics and to ensure it doesn't present risks for consumers. A targeted review of Articles 5, 9, 15, and 55 of the AI Act would be particularly advisable. Clear responsibility frameworks must define liability between developers, deployers, and users.

Priority 5 · Adopt an impact-centered governance approach

Governance frameworks should be built around anticipated real-life impacts, on society, labor, and the environment, rather than purely on technical criteria. This requires an interdisciplinary standpoint from the outset, involving not only technical experts but also legal scholars, ethicists, labor economists, sociologists, linguists, psychologists, and environmental scientists. It also implies ongoing observation and technical discussion to refine the understanding of Agentic AI as the technology evolves, ensuring frameworks remain relevant and comprehensive. A balance between the precautionary principle and the support of innovation must be found.

Priority 6 · Define common ground to compare energy consumption – Develop an energy score, based on common benchmarks, to give insights to developers and users on the energy impact of the chosen models and multi-agents architecture. Such an approach should prevent retro-engineering practices revealing industrial secrets, but would require auditing to ensure the veracity of the scores.

Priority 7 · Address cybersecurity risks specific to Agentic AI

Agentic AI's open-loop interaction with its environment introduces novel cyberattack risks. The opacity of many commercial models further prevents the deep auditing required in regulated sectors. Dedicated analysis is required to anticipate the scale and diversity of these threats, and secure identity and access control architectures should be promoted.[36]

Priority 8 · Anticipate workforce transformation and cognitive impacts

The development of Agentic AI systems calls for dedicated research into workforce transformation, skill erosion, and emerging phenomena such as "cognitive surrender." Social sciences must play a central role in analyzing these dynamics and informing policy responses, including education, training, and labor market adaptation strategies.

Priority 9 · Strengthen public awareness and education on Agentic AI

As Agentic AI systems become increasingly embedded in everyday tools, there is a need to promote accessible public education and AI literacy to ensure informed and critical use. This includes clarifying how agentic systems differ from other IT tools and raising awareness of risks such as over-reliance, reduced human agency, and hidden resource consumption. Attention should be given to the education of workers using Agentic AI systems, ensuring they are equipped to engage critically with these systems in professional contexts and adapt to evolving roles.

Conclusion

In conclusion, Agentic AI marks an important evolution in the landscape of generative artificial intelligence. It describes systems that combine generative models with tools, persistent memory and orchestration layers, enabling them to execute multi-step tasks and interact continuously with their environment within human-defined constraints. These abilities have a transformative potential in the economy, but introduce new complexities around accountability, safety, traceability and control that require demonstrated control in order to ensure the beneficial deployment of Agentic AI.

The key to harnessing the potential of Agentic AI lies in anticipating this evolution through effective and comprehensive governance, transparent oversight, and the implemention of proactive strategies to steer a beneficial adoption. Clear accountability frameworks, coupled with transparency could help limit concerns regarding responsibility, trust, and ethical

36 Cybersecurity is outside the scope if this paper, as the scale of potential impacts and number of challenges would require dedicated attention.

considerations, notably in the workplace. Moreover, the environmental impact of Agentic AI, particularly its energy consumption and computational demands, calls for strong frameworks to assess its ecological footprint.

As the technology continues to evolve, it is crucial for stakeholders, including policymakers, industry leaders, and academics, to collaboratively define and refine the regulations and governance structures that can ensure the safe deployment of Agentic AI systems. The development of specific frameworks tailored to its unique characteristics, alongside ongoing technical innovations to enhance transparency and traceability, will be pivotal in unlocking its full potential while mitigating adverse outcomes. Ultimately, a balanced and framed approach that maximizes the positive impacts of Agentic AI, while carefully monitoring and managing its risks, is essential for its responsible integration into society. Only so, Agentic AI can be a useful addition to society, and not a burden on individual and collective cognitive competencies.

Acknowledgements

This paper was coordinated by an interdisciplinary taskforce at Inria's Agency Programm, composed of Anna Bobasheva, Anaya Kumar, Loriane Koelsch, Fabien Le Voyer, Jacques Sainte-Marie, Marie-Fleur Simmet, Selma Souihel, and Fanny Terrier. It was based on thematic literature reviews and iterations with experts and practitioners. 5 Workshops gathering over 80 participants in total were organized, to do so. Some experts stayed involved after the workshops by providing written feedback. In addition, a survey gathered 30 respondents from the industry, shedding light on the actual use of Agentic AI.

In this perspective, we would like to thank Christophe Biernacki (Inria); Fabien Gandon (Inria); Paolo Giudici (University of Pavia); Cédric Gouy-Pailler (CEA); Nina Laibleiz (Télécom Paris ; Institut Polytechnique de Paris); Laurent Lefevre (Inria); Thomas Le Goff (Télécom Paris ; Institut Polytechnique de Paris); Liang Min (Stanford University); Pierre-Yves Oudeyer (Inria); Marina Reyboz (CEA); Juliette Senechal (Inria); Damien Sileo (Inria); Gilles Sassatelli (CNRS); François Terrier (CEA); Denis Trystram (Inria) for engaging in the researchers' webinars.
We would also like to thank Allison Bryan (EPSRC – UK gov); Zachary Bensemana (ISED-ISDE); James Clarke-Lister (ISED-ISDE); Robert Edwards (Shared Services Canada); Marc Goldfinger (TBS-SCT); Yuko Harayama (Tokyo Centre of the GPAI Expert Community); Olesia Kazantseva (Shared Services Canada); Stephanie King (Ceimia); Stuart Spence (ISED-ISDE); Scott Syms (Shared Services Canada); Matilda Titeca (DSIT, UK gov); Thomas Spencer (International Energy Agency) for taking part in the experts' webinars.

We express our sincere gratitude towards all the industry ecosystem, who engaged in the webinar to present an almost finished version of the paper, and more specifically to those who contributed to the survey and sent their thoughts on the paper. More specifically, we thank Merve Hickok (Invited at the GPAI Tokyo Expert Community Centre ; CAIDP) for bringing her perspective.

Finally, we would like to thank the organizers of the Digital & Tech track at the French Presidency of G7, for including our work in their discussions, and to all members, for engaging their national ecosystems. This support allowed to add nuance and additional perspectives to the paper.