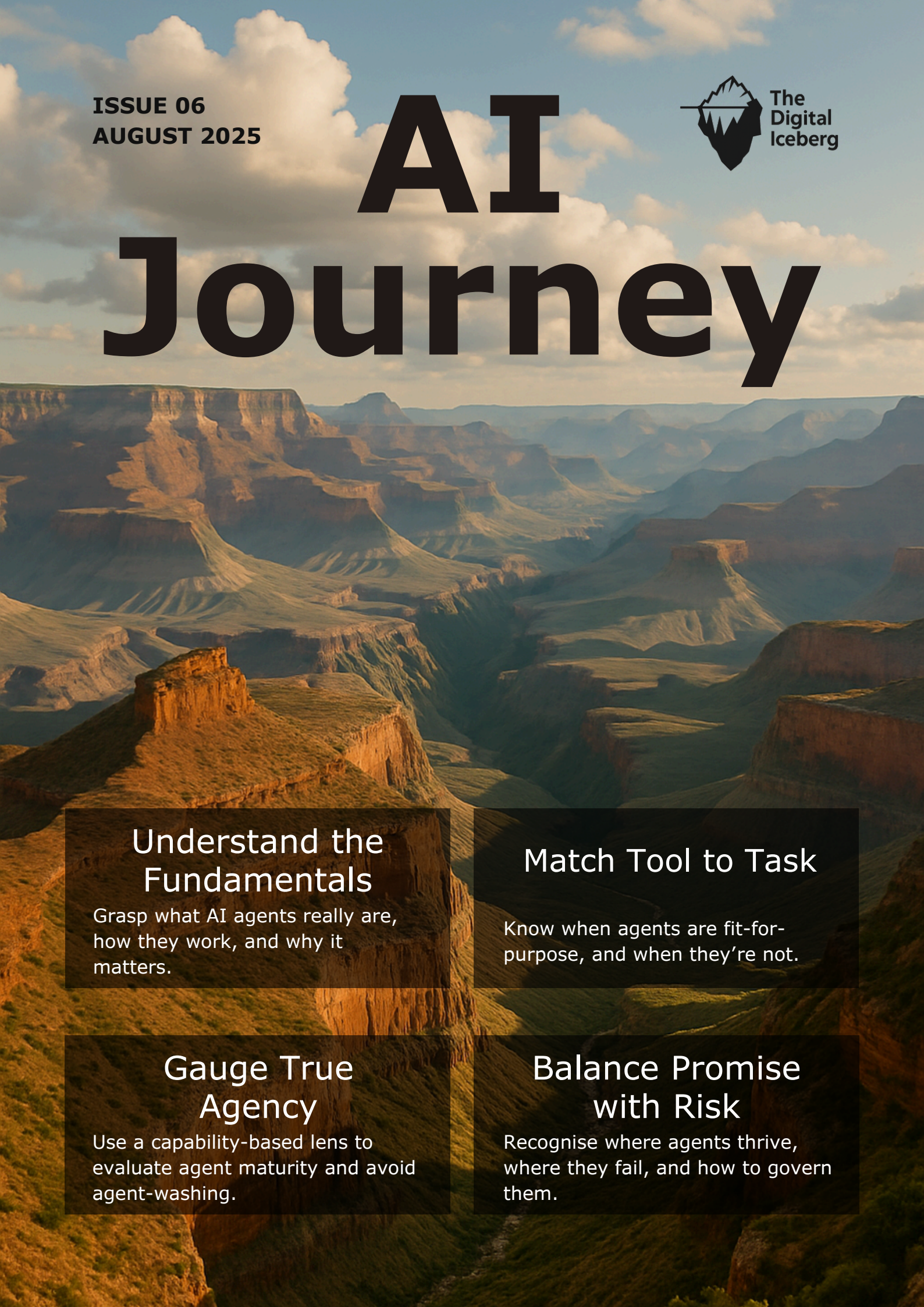


ISSUE 06
AUGUST 2025

AI Journey

The Digital Iceberg

Understand the Fundamentals

Grasp what AI agents really are, how they work, and why it matters.

Match Tool to Task

Know when agents are fit-for-purpose, and when they’re not.

Gauge True Agency

Use a capability-based lens to evaluate agent maturity and avoid agent-washing.

Balance Promise with Risk

Recognise where agents thrive, where they fail, and how to govern them.

An AI Agent Decision Framework

AI agents are currently generating significant excitement, promising to elevate Artificial Intelligence to unprecedented levels. But while the enthusiasm, and the marketing hype remain high, there remains considerable confusion in the market regarding their nature, application, and limitations. This e-book aims to demystify these concepts, providing clarity on what AI agents truly are, how they work, when and where to deploy them, and, crucially when alternative approaches are more suitable.

What are AI agents?

Let's start with understanding what they actually are. At their core, AI agents are sophisticated software entities that typically interact with either a human user or another system. Up until now, our software entities have been very prescriptive. We've had to explicitly give them information, tell them how to use that information, and specifically program the steps they should take to reach an outcome. Now, in the new world of AI agents, these more advanced software entities are able to draw conclusions or make decisions all by themselves in pursuit of a goal. They are able to determine what information they need, decide how to use it, and create an action plan based on the evidence. This is what is known as reasoning - it's the ability to connect facts, recognise patterns and apply logic to reach an outcome.

As a result, building an AI agent involves providing a prompt and a set of "tools". The prompt defines the goal we would like the agent to achieve, and a set of guardrails in which we would like it to operate within. And the tools are information sources, external services or external applications that the agent can invoke in pursuit of its goal. It does the rest - it creates the plan and then adapts its approach as it receives feedback during the execution of the plan.

This perceive-decide-act loop is not new and has existed in computer science and artificial intelligence research since the 1980's and 1990's. What makes it possible today in ways it wasn't before, is the advent of the Large Language Model (LLM) which allows these software entities to process natural language instructions provided by a human in pursuit of a goal. This has given it new levels of understanding and planning. However, that doesn't render more traditional approaches useless, far from it. Deterministic rule-based logic, machine learning driven optimisation algorithms,

and knowledge retrieval all still very much have their place because the LLM relies on them for context when it's planning, and for execution when it's decided on its approach.

Fundamentally, the value proposition of AI agents lies in their ability to create solutions that are more adaptive and better equipped to handle complexity. Unlike conventional, predefined automation, agents can act more autonomously, dealing effectively with exceptions, changes, and unforeseen circumstances within their environment. This means we are elevating the autonomy of software closer to human operation.

What are the different kinds of AI definitions?

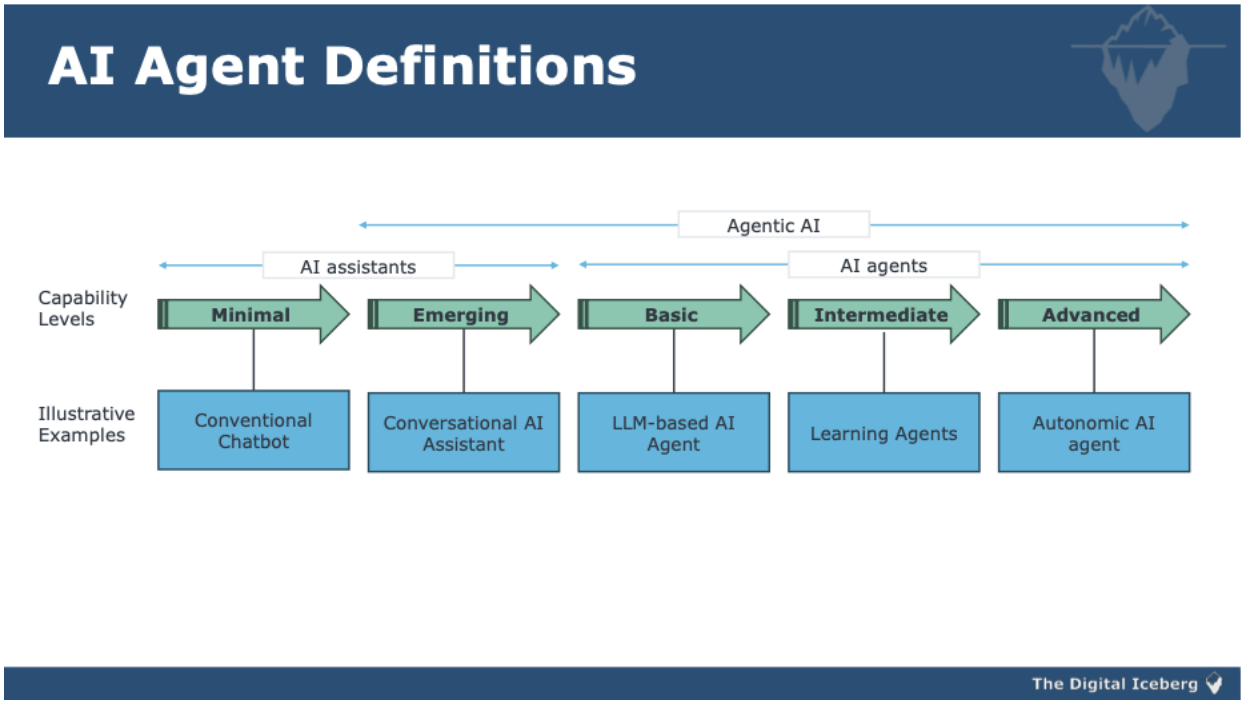
As the field of agentic AI has taken the market by storm, it has created significant definitional ambiguity with various interpretations of what constitutes an AI agent and agentic AI. To bring clarity I would like to outline specific definitions:

AI Agents: AI agents are software entities capable of autonomous or semi-autonomous operation meaning that they exhibit at least a basic level of agency. They perceive their environment, make decisions on how to respond, and take actions to achieve their goals, often with some degree of human oversight.

Agentic AI: This refers to the approach of building AI solutions that incorporate at least one software entity that qualifies as an AI agent.

AI Assistants: Distinct from AI agents, AI assistants are software entities or applications designed primarily for interacting with humans or another system. Their role is to execute tasks as explicitly requested, specified, or dictated by the user or system. The key difference is that AI assistants do not possess the inherent level of autonomy expected of AI agents.

The reason why baselining these definitions is so important is because it allows you to differentiate between genuine AI agent offerings and instances of "agent washing". There are a lot of vendors relabelling conventional automation solutions, such as workflow automation and Robotic Process Automation (RPA), with "agent" or "agentic" in their names, despite these offerings not meeting the true criteria required for an AI agent.



What are the different levels of Agent capabilities?

Not all AI agents or agentic AI are equal, and you are likely craving a means for differentiating or measuring the differences between different vendor solutions, or even your own self-built solutions. Contrary to what it might look like in the marketing collateral, agency is not a binary characteristic and like many things, instead exists along a spectrum ranging from minimal to advanced.

The following framework aims to allow you to decompose the overall agency level into more specific, measurable capabilities, allowing for a more granular assessment. These capabilities are:

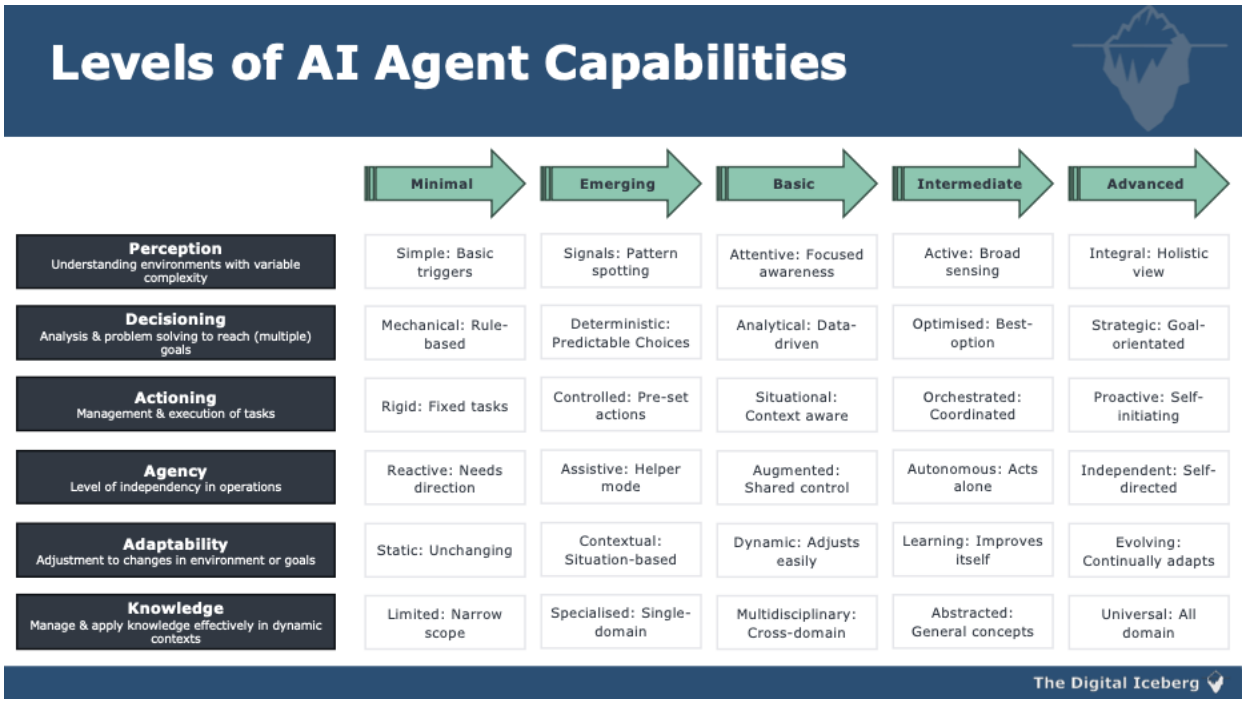
Perception: This capability assesses an agent's ability to understand environments with varying complexity.

Decisioning: This focuses on the agent's capacity for analysis and problem-solving to achieve (potentially multiple) goals.

Actioning: This measures the agent's proficiency in managing and executing tasks.

Adaptability: This evaluates the agent's capacity for adjustment to changes in its environment or goals.

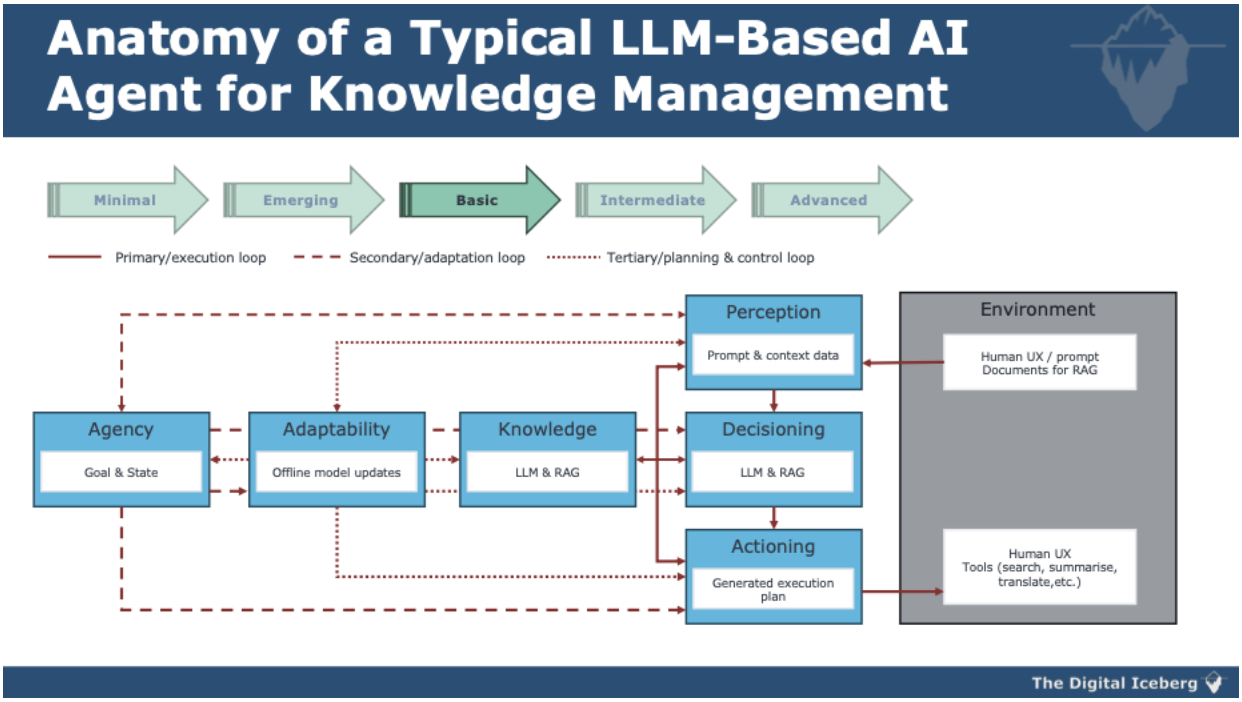
Knowledge: This refers to the agent's ability to manage and apply knowledge effectively within dynamic contexts.



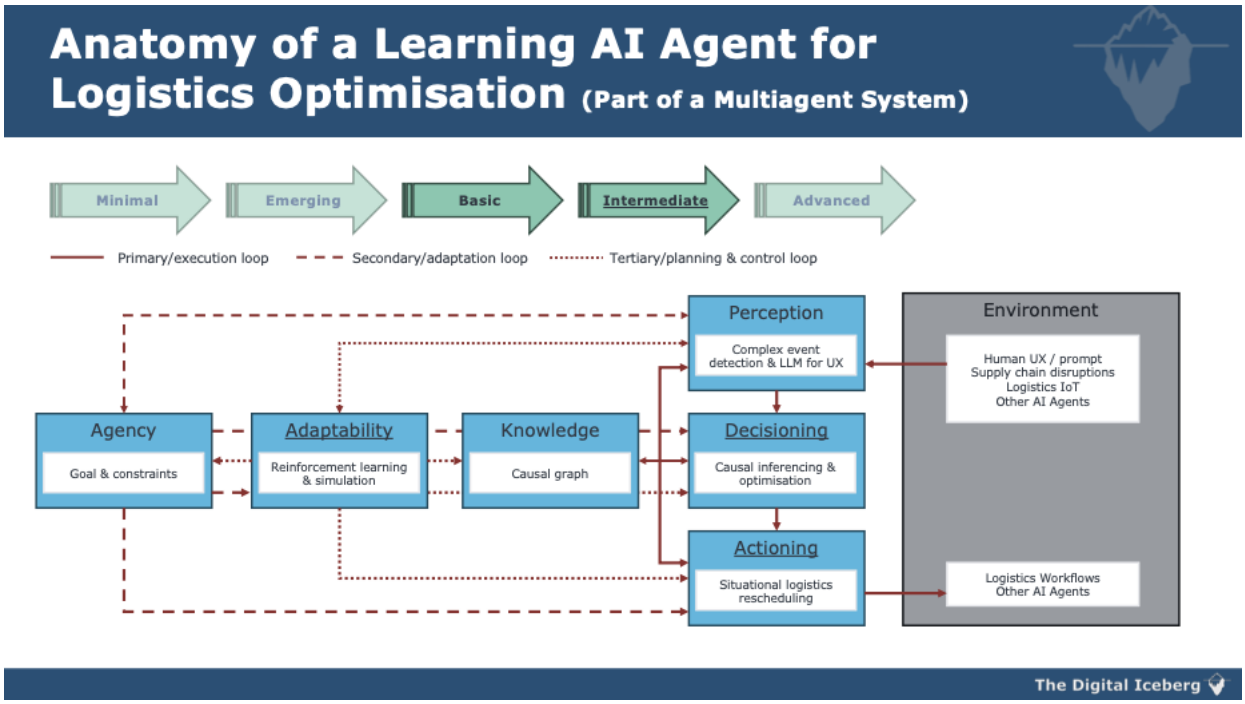
By assessing an agent across these different capabilities, organisations can determine its true level of agency and discern whether it genuinely qualifies as an AI agent or should perhaps be classified as an AI assistant.

The framework can also be utilised to create an "anatomy of an agent," which provides an in-depth view of how the various capabilities that constitute an agent interact and function together. For example:

A common AI agent used for knowledge management heavily relies on Large Language Models (LLMs) in conjunction with Retrieval Augmented Generation (RAG). This agent takes prompts and documents as input, generating summarised or listed results through a human interface.



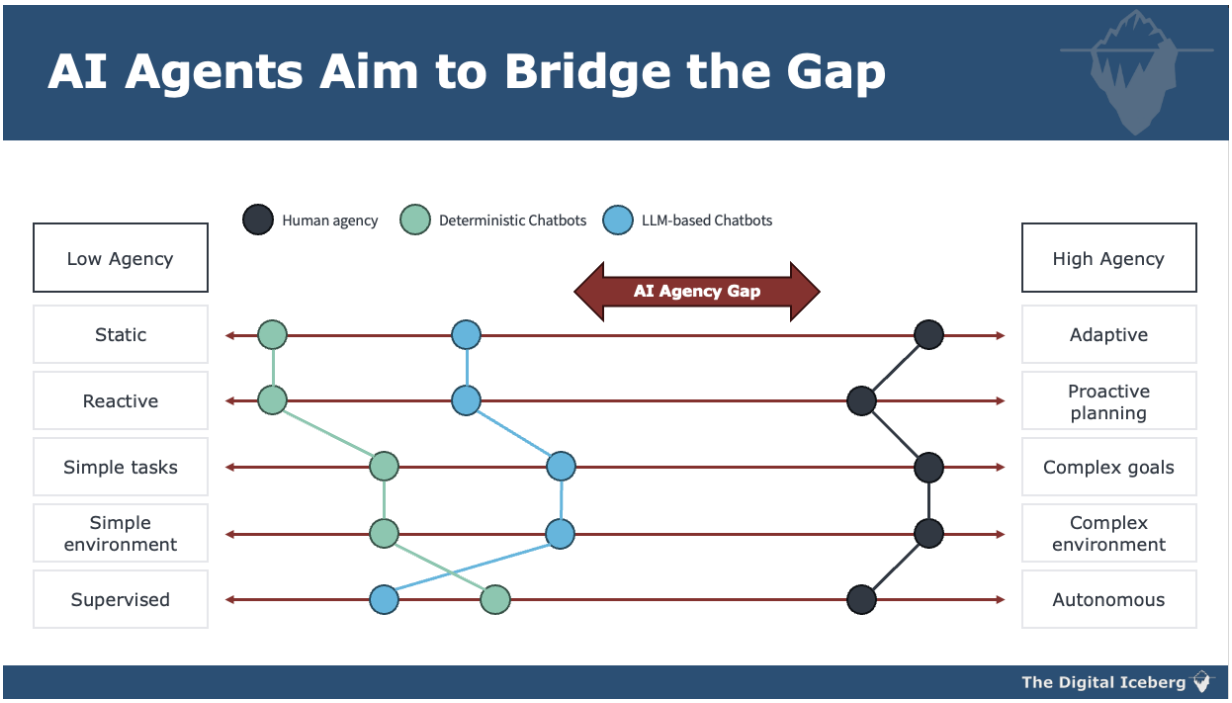
A more advanced agent, part of a multi-agent system for logistics optimisation, may primarily use causal graphs or causal reasoning for its actual reasoning and decision-making, rather than solely depending on an LLM. While an LLM might still be used for user interaction (natural language processing), the symbolic approach of causal reasoning offers greater transparency and reliability, which may be preferred when LLMs are deemed insufficiently reliable.



This demonstrates that agents are not a singular type of technique but rather a category of software entities capable of autonomous or semi-autonomous actions, leveraging diverse capabilities at varying levels of sophistication.

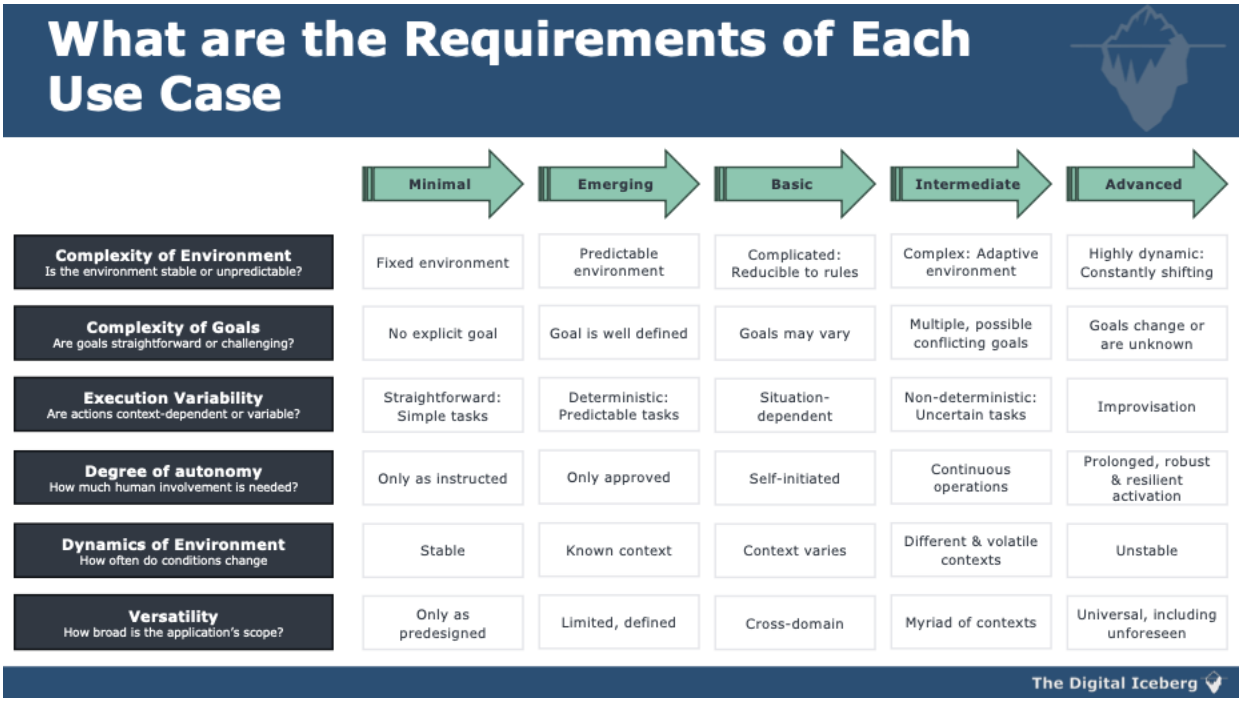
When should you use AI agents?

The decision to deploy AI agents should be a strategic one, not driven by hype. While AI agents offer significant advancements over conventional automation, they are not a "silver bullet" for every use case. The core value of AI agents lies in their ability to create solutions that are more adaptive, capable of handling greater complexity, and can act with higher degrees of autonomy. They excel where conventional automation, which relies on predefined steps, struggles with exceptions, changes, and dynamic environments. Essentially, agents bridge the agency gap, bringing automation closer to the flexibility and problem-solving abilities of humans.



To determine if an AI agent is suitable for a particular use case, I recommend using a requirements-driven approach, just how you would for assessing any other technology. Use the simple framework provided to determine the extent to which your use case demands agentic AI or AI agents. This will help you identify the

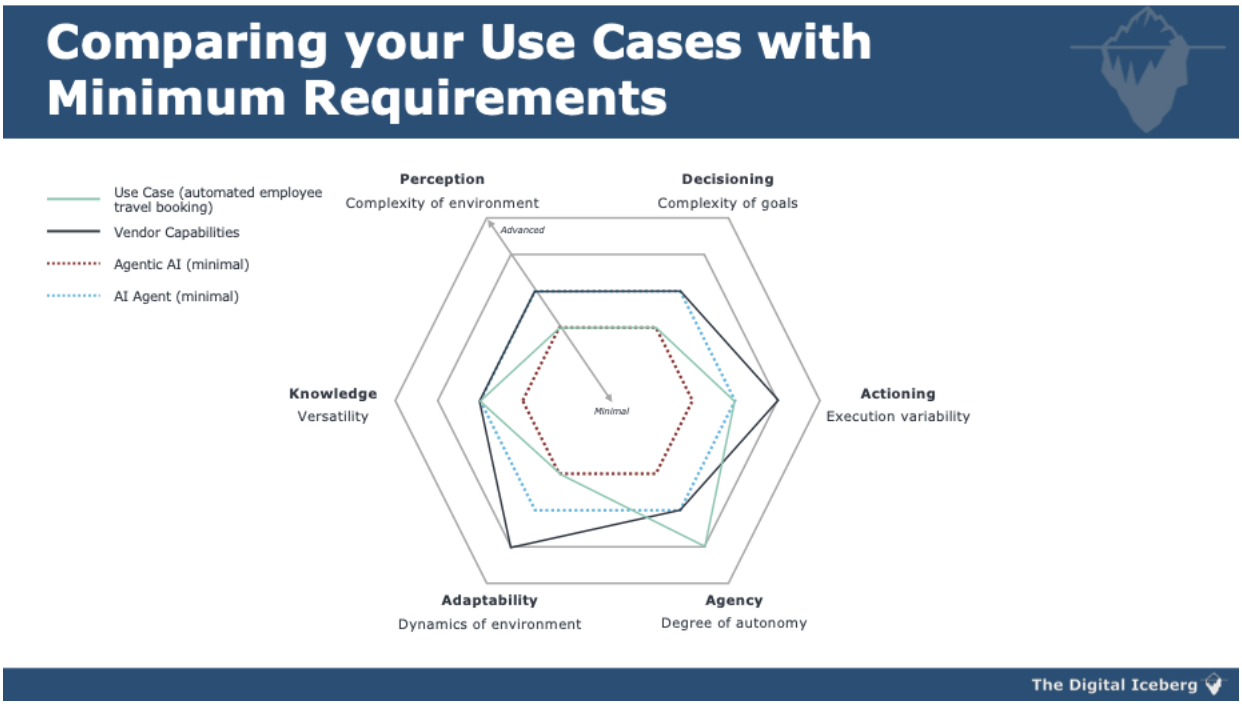
necessary levels of capabilities, such as the complexity of the environment, the required autonomy, or the versatility of the agent.



Consider the example of automating employee travel booking. This task is inherently complex due to numerous variables, including employee preferences, corporate policies, budget constraints, and even sustainability goals. Conventional automation often finds this task challenging. By applying the framework (visualised below as a spider diagram), you can plot the required level of agency for each capability (e.g., perception, decisioning, actioning, autonomy, adaptability, knowledge) for the travel booking use case.

The spider diagram allows for a crucial comparison:

- Your use case requirements are represented by the green line.
- The minimal level for agentic AI is shown by a red dotted line.
- The minimal level to qualify as an AI agent is indicated by a light blue dotted line.



By comparing your use case requirements (green line) against these thresholds, you can determine if an agent is needed at all. If your requirements fall below the minimum for agentic AI or AI agents, then conventional automation would suffice. However, in the travel booking example, the requirements typically exceed these minimal agentic levels, often requiring intermediate or advanced capabilities, thereby confirming the need for an AI agent.

This framework also serves as a critical tool for evaluating vendor offerings. If a vendor claims to offer an "agentic travel booking service," its assessed capabilities (represented by the black line) can be compared against your requirements. A significant gap between what you need and what the vendor offers indicates that the solution may not be suitable.

In practice, the greatest impact and value from AI agents are typically realised at higher organisational levels i.e. at the process or business model level, rather than being confined solely to individual or personal use.

Examples of high-value applications include:

- Process Level Automation: An Italian insurance company increased its automated claims processing from 60% to 75% using agents, which can generate action plans on the fly and handle exceptions or unforeseen circumstances better than conventional automation.
- Enterprise/Ecosystem Optimisation: A large global distribution company employs a multi-agent system to optimise logistics by constantly monitoring

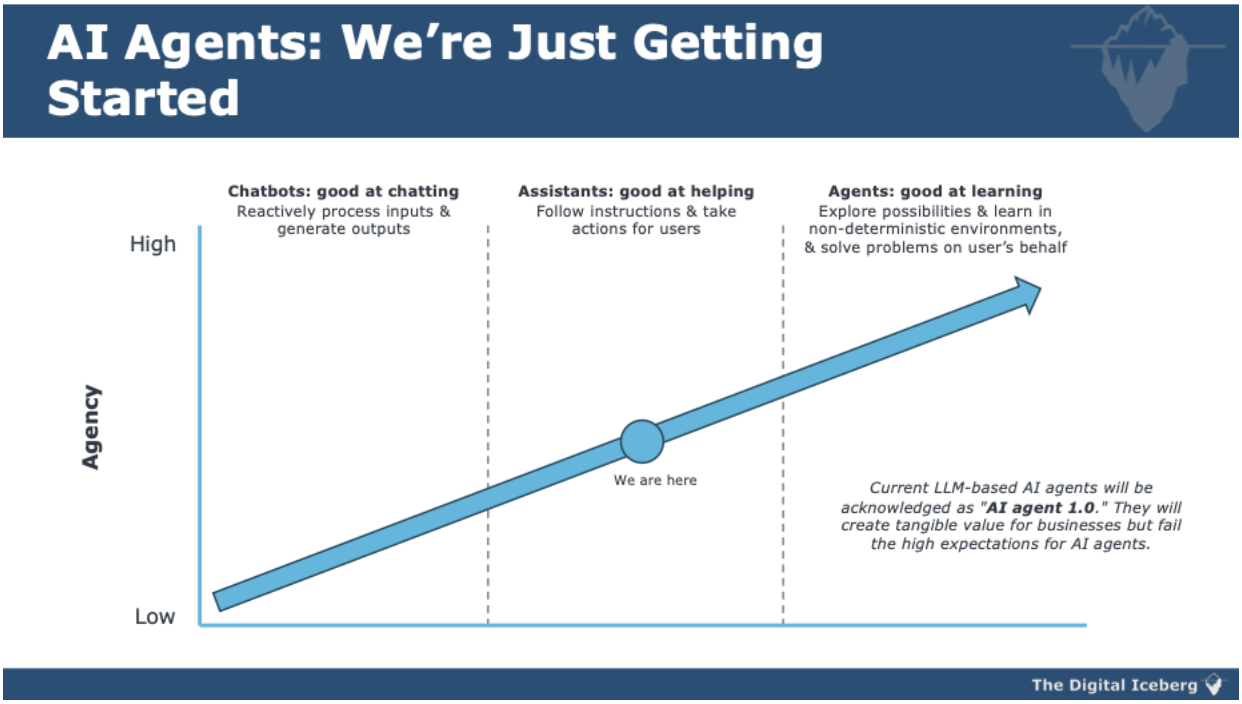
supply, demand, and disruptions, making real-time decisions to manage delays effectively.

- Back-office Functions: In the financial services industry, and indeed many other sectors, popular use cases for agents are found in back-office operations.
- Knowledge Management: Agents are widely used for finding information quickly, summarising content, and translating documents.
- Sales Force Automation: Agents are popular in this area for expediting and accelerating the creation of personalised offers for clients.

It's important to recognise that AI agents are currently best used to empower people, making them more efficient and improving the quality of their work, rather than as a direct replacement for human roles. While increased efficiency may lead to headcount reductions in some departments, currently most companies are taking the position that AI is not a full replacement for human intelligence but rather an enrichment and empowerment tool.

What is the current maturity of AI Agents?

The current state of AI agent technology is still in its nascent stages; we are, in essence, like with any technology, just getting started. The predominantly LLM-based agents available today are considered "AI agent version 1.0". While these agents can deliver tangible value, they are not yet capable of fulfilling the highest expectations



placed upon them. Looking ahead, "AI agent 2.0" is anticipated to emerge in the coming years through continuous innovation and new technological advancements. This next generation may incorporate a broader range of AI techniques beyond LLMs or improved versions of existing LLMs.

Current Strengths of AI Agent Version 1.0

Despite their early maturity, current AI agents possess notable strengths:

- They are adept at interpreting information.
- They excel at decomposing a question or task into a series of actionable steps (planning).
- They are proficient in identifying and utilising the appropriate external tools or services needed to achieve user-specified goals.
- They can contextualise information and take personal preferences or specific contexts into account.

Current Limitations of AI Agent Version 1.0

It is crucial to have a realistic understanding of the significant limitations of current AI agents, as these may dictate when and where not to use them.

- **Learning Capability**: A common misconception is that all agents learn. In reality, the vast majority of agents today do not learn in the sense that their underlying models do not fundamentally change over time. While they can adapt to context, they don't continuously improve their core logic through new experiences.
- **Cost**: Uncontrolled deployment of AI agents can lead to spiralling costs. If employees create numerous agents, especially those provided by vendors on a pay-per-use basis or utilising commercial LLMs charged per token, the expense can quickly outweigh the business value.
- **Lack of Human Touch**: For certain customer or employee interactions, a human touch may be highly valued. Agents may not always be the optimal solution for delivering a desired level of customer or employee experience if this human element is a strategic priority.
- **Reliability and Consistency**: When an agent relies on an LLM, its actions may not always be 100% consistent or reliable. This is a significant concern, particularly for risky use cases involving financial transactions or critical operations. While methods like predefined workflows, fine-tuning, or domain-specific LLMs can improve reliability, they don't guarantee perfection.

- **Performance (Speed)**: Agents, especially those using LLMs or other complex AI models, can be slow, with response times often exceeding a few seconds. Such delays are unacceptable for real-time applications, such as controlling machinery.
- **Transparency and Explainability**: Many AI models, particularly LLMs, operate as "black boxes," making it difficult to explain how the agent arrived at a particular conclusion or action. This lack of transparency can pose challenges from a responsible AI perspective, regulatory compliance, and risk management.
- **Sustainability Concerns**: There is a growing concern regarding the electricity and water consumption associated with AI in general, including the training and execution of AI agents. As the number of deployed agents scales to millions, the environmental impact becomes a significant headache in terms of energy consumption and sustainability.
- **Skill Requirements**: Implementing and managing AI agents requires specialised skills.
- **Integration Challenges**: Integrating AI agents with existing enterprise systems presents a significant hurdle. This is a perennial problem in IT, and while protocols like MCP exist to facilitate integration, they do not magically eliminate all difficulties.
- **Data Quality Issues**: Like other AI applications, AI agents are susceptible to poor data quality. Missing or inaccurate data will adversely affect agent performance.
- **Immature Vendor Offerings**: The market for AI agent solutions is rapidly evolving, and some vendor offerings may still lack maturity.
- **Lack of Agent Governance**: Supporting the widespread use of agents with adequate governance is currently immature. Organisations need robust mechanisms to monitor costs, track agent usage, and manage their lifecycle (e.g., removing agents no longer in use).
- **Security Vulnerabilities**: AI agents introduce new security risks. There are concerns about malicious actors injecting "foreign agents" into internal platforms, potentially gaining unauthorised access to data and systems if proper authorisation and identification controls are not in place. These risks necessitate mitigation strategies.

These limitations highlight the need for a realistic perspective on current AI agent capabilities, counteracting the prevailing overhyped expectations in the market.

When shouldn't you use AI Agents?

Understanding when *not* to use AI agents is as crucial as knowing when to deploy them effectively. Deploying agents inappropriately can lead to unnecessary costs, reduced reliability, and suboptimal outcomes.

Low Complexity Use Cases

If your use case requirements are not particularly advanced, the environment is not complex, or the goals are straightforward, AI agents are often overkill. For such scenarios, conventional automation methods like workflow automation, Robotic Process Automation (RPA), or Business Process Management (BPM) are perfectly adequate, more cost-effective, more reliable, and offer greater transparency. These traditional technologies are still highly effective when the use case does not demand the dynamism, complexity, optimisation, or adaptability that AI agents bring.

Tasks Requiring Significant Human Involvement

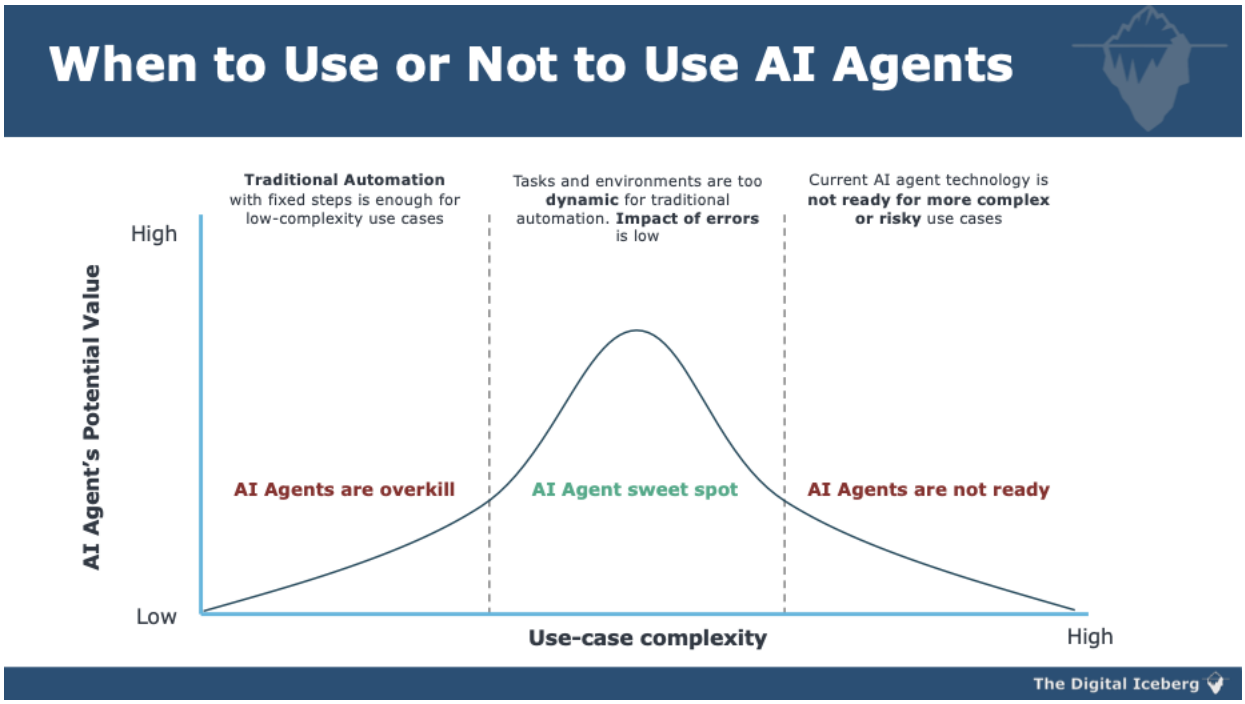
At the other extreme of complexity, there are tasks that remain too intricate or nuanced for current AI agents. These often require human judgment, creativity, or empathy that AI is not yet mature enough to replicate. In such instances, even with advancements, agents are simply "not ready," and human involvement remains essential. AI agents are not a universal solution for every problem; certain tasks still require human intelligence.

When Limitations Outweigh Benefits: Given the limitations of AI Agent Version 1.0 (as discussed previously), you should avoid using agents if any of these limitations pose an unacceptable risk or cost for your specific use case. These limitations include:

- **High Costs**: If the potential expenditure on agent deployment and operation (especially with pay-per-use LLMs) outweighs the value gained, or if cost control is unachievable.
- **Unreliability**: For high-risk tasks where 100% reliability and consistency are non-negotiable, and current agent reliability cannot be sufficiently guaranteed or mitigated.
- **Slow Performance**: In scenarios demanding real-time responses where even a few seconds of delay are unacceptable.
- **Lack of Transparency**: If regulatory requirements or internal policies demand explainability for every decision, and the opaque nature of some AI models makes this impossible.

- **Need for Human Touch**: For interactions where a personal, human connection is vital for customer or employee experience.
- **Immature Governance & Security**: If your organisation lacks the capability to implement robust governance and security measures to manage large-scale agent deployments and mitigate associated risks effectively.

In summary, the decision matrix for AI agents can be viewed as a graph where value increases with use case complexity up to a "sweet spot". Below a certain complexity, agents are overkill. Above a certain complexity, current agents are not sufficient and still require human involvement. The "sweet spot" for AI agents lies in the middle, where they bring capabilities to overcome a specific level of complexity that conventional automation cannot handle, and that's where they are most useful.



Closing Recommendations

To successfully navigate the landscape of AI agents and harness their true potential, consider the following recommendations:

1. **Assess and Validate Diligently:**

Use the frameworks and definitions provided to rigorously assess the capabilities of both vendor-provided and self-made AI solutions. This assessment should determine if they genuinely qualify as AI agents or agentic AI, or if they are merely AI assistants.

This practice is crucial to reduce market confusion and to identify instances of "agent washing," ensuring that vendors are offering proper, real agents that can deliver expected benefits.

2. **Adopt a Requirements-Driven Approach:**

Do not treat AI agents as a universal solution for every problem. The hype surrounding AI agents can lead to over-optimism, but a sense of realism is vital. For each potential use case, conduct a thorough assessment of its unique criteria and requirements. This includes evaluating the complexity, the dynamics of the environment, and other relevant characteristics. Make a rational choice based on whether an AI agent is truly needed for that specific use case.

3. **Keep an Open Mind for Alternative Approaches:**

Remember that AI agents are not the only technological solution available. Evaluate and remain open to alternative delivery approaches, including conventional workflow automation, Robotic Process Automation (RPA), or other non-AI techniques. These alternatives may be perfectly suitable, and often more efficient or reliable, for use cases where AI agents are not the optimal solution.

4. **Implement Robust Guardrails, Especially for Risky Use Cases:**

Given the potential unreliability of agents, particularly those using LLMs, it is imperative to implement appropriate rules and guardrails. The extent of these guardrails should directly correlate with the assessed risk level of the use case. Consider implementing "guardian agents" or similar mechanisms that can validate what a primary agent intends to do before execution, especially for medium or high-risk scenarios. This could involve simple constraints (e.g., no purchases above a certain amount) or more advanced methods where a guardian agent, potentially using another AI model, compares its conclusions with the primary agent's, escalating to a human user if there's disagreement.

5. **Prioritise Governance and Monitoring:**

Just like any other technology, AI agents require ongoing monitoring and governance post-deployment. Organisations must track how much agents are being used, monitor their performance (both good and bad), and connect their usage to defined Key Performance Indicators (KPIs) to measure their actual impact and value creation. Implementing a robust governance framework is also essential for managing costs, keeping track of which agents are in use, and ensuring proper "housekeeping" (e.g., removing agents that are no longer needed). It is important to acknowledge that support for agent governance is still relatively immature in the current market.

By embracing these recommendations, organisations can move beyond the hype, make informed decisions about AI agent adoption, and strategically leverage this

powerful technology to achieve tangible business value, while mitigating associated risks. The future of AI agents is promising, but a realistic and structured approach will be key to unlocking their full potential.



The Digital Iceberg

www.thedigitaliceberg.com