Perspective Article

International Journal of Mathematic Exploration and Computer Education

AI Agents : Your New Smart Colleagues

Rodolfo Valacca

Area 62 Srl, Milan, Italy.

***Correspondence author**
Rodolfo Valacca,
Area 62 Srl, Milan,
Italy.

Submitted : 30 Jul 2025 ; Published : 3 Oct 2025

Citation: Valacca, R. (2025). AI Agents: Your New Smart Colleagues. *Int J Math Expl & Comp Edu.2*(3):1-12.

Abstract

This article offers the perspective of an Innovation Manager with over 20 years of professional and academic experience on the impact of Artificial Intelligence (AI) Agents on personal and professional life as per knowledge (end-July 2025).

This article, titled "AI Agents: your new Smart Colleagues", explores the rapid evolution of AI systems from passive Large Language Models (LLMs) and structured AI Workflows to autonomous or semi-autonomous, ***goal-oriented AI Agents.***

Understanding ***AI Agents from the user's perspective*** *is crucial. Unlike traditional models, AI Agents independently rotate through thinking, acting, planning and executing to achieve high-level objectives, effectively* ***becoming decision-makers****. Their core characteristics include autonomy, task specificity, responsiveness and adaptation, powered by foundational technologies such as Generative AI, LLMs, Multi-Modal AI, Natural Language Processing (NLP), and Reinforcement Learning. They differ from AI assistants by proactively initiating complex,* ***multi-step actions.***

The ***AI Agent market*** *is projected to grow substantially, from an estimated* ***$7.84*** *billion in 2025 to* ***$52.62*** *billion by 2030, fostering transformative applications across diverse sectors like business automation, customer service, financial analysis and legal activities. Use cases include Moody's for financial analysis, eBay for marketing automation and JPMorgan Chase for legal contracts.*

Crucial for their seamless operation is the ***interoperability of AI Agents****, facilitated by emerging open protocols like Model Context Protocol (MCP), Agent 2 Agent (A2A) and NLWeb, which promote communication and collaboration among various AI Agents, forming "****Agentic AI****".*

Performance *is rigorously evaluated using specific* ***benchmarks*** *(e.g. DS-Bench, GAIA, and Web Arena) and a combination of technical, business, user-centric and Return-On-Investment (ROI)* ***metrics****. While still at early experimental stages, real-world manifestations like OpenAI's ChatGPT Agent and Google's Deep Research underscore their significant potential.*

The article emphasizes the critical need for the ***responsible use of AI Agents****, addressing inherent risks such as data privacy concerns, decision errors and embedded biases. It strongly advocates a "****humans in the loop" approach,*** *stressing the paramount importance of robust* ***human oversight*** *and* ***ethical guidelines*** *to ensure safety and foster trust.* ***Hallucinations*** *and* ***errors*** *made by AI Agents do not remain confined within the chatbots (as in the case of Gen AI/LLM), but become visible to third parties with a potential negative impact on personal reputation.*

AI Agents + Humans = Assisted AI Agents!

The integration of AI Agents into the workplace necessitates a fundamental ***rethinking of educational*** *systems and a proactive* ***shift in relevant human skills,*** *moving towards expertise in orchestration, interpersonal relations and complex decision-making, thus forging a truly collaborative human-AI ecosystem.*

The ***collaboration between humans and AI Agents*** *(Smart Colleagues) will become a defining characteristic of the modern workplace.* ***Human workers will adapt their own role****, as companies will integrate AI Agents into their organizations, following the* ***golden rules for responsible use of AI Agents.***

Keywords: Agentic AI, Agent 2 Agent (A2A), AI Agent, Artificial Intelligence (AI), Assisted AI Agent, Bias, Generative Artificial Intelligence (GenAI), Hallucination, Large Language Model (LLM), Machine Learning (ML), Model Context Protocol (MCP), Prompt Engineering, Responsible AI, Retrieval Augmented Generation (RAG).

The landscape of Artificial Intelligence is rapidly evolving, navigating in a new era where AI systems transcend their traditional roles as passive tools to become autonomous, goal-oriented "Smart Colleagues".

As Satya Nadella, CEO of Microsoft, proclaimed in 2025, "*AI Agents will replace all Software… From Software as a Service to Agent as a Service".*

This shift represents a fundamental transformation in how AI interacts with users and operates within complex environments.

Understanding AI Agents: beyond LLMs and AI Workflows

Traditional Large Language Models (LLMs) are trained on vast datasets to generate new content based on user prompts. By default, they are passive and reactive, lacking access to external tools or autonomous reasoning capabilities. For instance, an LLM alone cannot answer, "When is my next appointment?" due to its intrinsic limitations, such as a reactive nature and lack of access to private user data like a calendar.

AI Workflows mark an advancement, allowing users to define steps, tools and data for the LLM to consult, such as instructing it to "look in my Calendar before answering" for a personal event query. While external tools and data can be accessed (e.g. through Retrieval Augmented Generation or RAG), the user remains the primary decision-maker, defining all the control logic.

AI Agents represent the next evolutionary leap. Instead of step-by-step instructions, the user provides a high-level goal (e.g. "Create and publish a post on LinkedIn every day"). The AI Agent then autonomously and iteratively performs a cycle of:

1. Choosing the necessary tools ("Which AI Agent is best suited to promote professional experience?").
2. Thinking ("What is the best way to collect relevant articles?").
3. Acting ("Which AI model writes best for LinkedIn?").
4. Planning the next logical step ("What is the optimal number of LinkedIn posts to publish per week? Additionally, on which days and at what times should posts ideally be scheduled?").
5. Executing it autonomously ("How can I improve this post campaign?").

This continuous cycle of thinking, acting, evaluating and improving distinguishes AI Agents from AI Workflows, allowing the **AI Agent to become the decision-maker** until a predefined or self-imposed quality standard is met. Several LLMs (e.g. Open AI) offer the possibility of using them in AI Agent mode (e.g. "ChatGPT Agent") for example, to make an e-commerce purchase (searching for the best product and then buying it online). These AI Agent usage options on the most popular chatbots are helping to democratize the use of AI Agents.

As Demis Hassabis, CEO Google DeepMind, announced in 2025, *"Gemini: will become an autonomous Agent… a real proactive assistant in users' daily activities, e.g. booking tickets or making online purchases independently".*

The foundations for modern AI Agents were laid in the 1950s and 1960s with early work on AI. However, the concept of an AI Agent as we understand it today developed more concretely in the 1980s and 1990s and then started to become widely used in late 2024.

Today, an **AI Agent** is defined as "**An AI system designed to operate in specific environments and to achieve goals by interacting with the surrounding environment, making decisions and taking actions autonomously (or semi-autonomously) to execute specific tasks**".

Their key characteristics include autonomy, task specificity, responsiveness, and adaptation. An AI agent is typically a software system that can:

- Perceive its environment through sensors or data input.
- Reason and make decisions.
- Plan a sequence of actions to achieve a goal.
- Exhibit a degree of autonomy (an AI Agent can act independently without constant human intervention).
- Learn and adapt over time.
- Interact with other agents or systems for more complex tasks.

A **crucial distinction** lies between AI Agents and AI Assistants:

- **AI Agents** (e.g. ChatGPT Agent) are goal-oriented and can proactively initiate complex, multi-step actions, adapting as they go.
- **AI Assistants** (e.g. ChatGTP) usually respond to direct user prompts and require more human decision-making.

This greater autonomy means AI Agents perform actions independently, such as writing emails, sending them, searching for information or booking trips. They are designed to reason through complex problems, create actionable plans and execute these plans using a suite of AI tools. Their capabilities include advanced reasoning, memory retention, and task execution. In practice, AI Agents can also coordinate with each other in **multi-agent AI systems (Agentic AI)** to achieve specific results.

Classification of AI Agents: a multi-dimensional perspective

AI Agents can be categorized by their system architecture:

- **Single Agent:** the AI Agent performs tasks independently.
- **Multi-Agent:** the AI Agent collaborates with other AI Agents.
- **Human-Agent:** the AI Agent works alongside people.

AI Agents can also be classified in macro-categories based on decision-making:

- **Simple reflexive AI Agents:** react to immediate stimuli, rule-based, no memory (e.g. smart thermostats).
- **Model-based AI Agents:** decide and act guided by a model (e.g. email classification/response systems).

- **Goal-based AI Agents:** decide and act guided by a specific goal (e.g. robot vacuum cleaners learning a room layout).
- **Utility-based AI Agents:** optimize processes by evaluating risks/benefits to achieve best outcomes (e.g. humanoid robots).

In terms of their development and deployment, the types of AI Agents can be

- **Predefined AI Agents:** ready-to-use, either
  - horizontal: general-purpose AI Agents (e.g. Manus AI or ChatGPT Agent)
  - vertical: AI Agents with focus on a specific purpose: specific department (e.g. Microsoft's AI Sales Agent) or specific target (e.g. Webidoo's "Groow", a family of AI Agents for Small Medium Enterprises) or specific target and business function (e.g. Webidoo's "Groow HR").
- **Customized AI Agents:** created using low-code tools (e.g. Microsoft's Copilot Studio) or no-code tools (e.g. Zapier).
- **Custom-built AI Agents:** developed using AI Agent creation tools (e.g. Microsoft's Azure AI Foundry, Lindy) through custom training with privately owned (e.g. OpenAI's LLM) or open-source models.

Designing and building scalable AI Agents

Before sharing some tips on how to build an AI Agent, a couple of premises on the key components and core technologies of AI Agents may be useful.

The internal workings of an AI Agent rely on several **key components:**

- **Action Module:** executes planned actions to interact with the environment.
- **Agent Core:** the central processing unit acts as the 'brain'.
- **Decision-Making Module:** uses models to choose the next steps.
- **Learning Module:** improves performance based on outcomes.
- **Memory Module:** stores and retrieves information for context and continuity, enabling learning from past interactions.
- **Perception Module:** collects and interprets data from the environment (e.g. through sensors) and external resources (e.g. through Application Programming Interfaces - APIs).
- **Planning Module:** analyses problems and devises strategies to achieve goals.

AI Agents are powered by **core technologies** such as

- **Generative AI:** for creating content, solutions, and responses.
- **Large Language Models** (e.g. ChatGPT, Claude, Gemini).
- **Multi-Modal AI:** for processing text, images, audio, video.
- **Natural Language Processing:** for language understanding and context handling.
- **Reinforcement Learning:** learning through trial and feedback.

Designing an effective AI Agent involves

1. Defining **clear objectives**.
2. Selecting the **appropriate LLM** (considering task complexity, data availability, and computational resources).
3. Configuring **inputs and parameters** (e.g. tone of responses, clear rules).
4. Integrating with existing **IT/Data systems** (e.g. Customer Relationship Management, databases, and tools to access and update information in real time).
5. Testing and continuous **optimization** (e.g. test the AI Agent in simulated environments to obtain user feedback).

Building scalable AI Agents additionally requires

- Careful **selection of frameworks** (e.g. OpenAI Agents SDK, CrewAI, LangGraph, Autogen).
- Purposeful **integration using APIs and Model Context Protocol** (MCP) servers.
- Implementing **reasoning frameworks**: to guide decision-making and behavior.
- Robust **memory management**: short-term (real-time context for immediate task) and long-term (past interactions and facts for better recall and continuity).
- Choosing appropriate **knowledge/Data Bases** (e.g. Vector DB, Graph DB, Knowledge Graph DB).

Market trajectory and transformative applications of AI Agents

The **AI Agent market** is projected to grow substantially, from an estimated **7.84 billion USD in 2025 to 52.62 billion USD by 2030**, reflecting a Compound Annual Growth Rate (CAGR) of 46.3% [Markets and Markets, 2025].

This rapid expansion is driven by the AI Agents ability to augment human capabilities and to generate **applications of AI Agents** across numerous sectors:

- **Business Automation:** companies like eBay deploy AI Agents to personalize customer interactions/offers and automate marketing initiatives, leading to a 25% increase in advertising campaign conversion rates [eBay, 2025].
- **Customer Service:** Agentic AI is expected to manage 68% customer service interactions of technology vendors [Cisco, 2025].
- **Financial Analysis:** companies like Moody's use AI Agents to summarize complex financial documents, simulate economic scenarios and collaborate with human analysts, reducing report production times by 40% [Moody's, 2025].
- **Legal Activities:** AI Agents significantly boost operational efficiency, reducing legal research time by 70-80% and due diligence time by 85-90%. They also improve accuracy (95% for document identification vs. 70-80% manually) and cut operating costs by 30-40% for automatable tasks. Companies like JPMorgan Chase use AI Agents to review legal contracts (over 12,000 annually). Furthermore, AI Agents excel in predictive analysis, forecasting case outcomes and suggesting strategies by analyzing vast amounts of historical data [JPMorgan Chase, 2025].

- **Maintenance:** companies like Uber use AI Agents to autonomously handle internal tech support [Uber, 2024].
- **Supply Chain:** companies like Bayer use AI Agents to predict flu outbreaks optimizing supply chain and marketing efforts [Bayer, 2025].

Gartner predicts that **by 2028, at least 15% of daily work decisions will be made autonomously by AI Agents,** a significant increase from 0% in 2024 [Gartner, 2025]. This shift signals a future where virtual workers powered by AI will operate within corporate networks, necessitating a radical rethinking of organizations and cybersecurity strategies.

Interoperability between AI Agents: the backbone of Agentic AI

Seamless communication and interaction are vital for AI Agents to become true smart colleagues, New open **protocols and standards** are emerging to **facilitate interoperability between AI Agents**

- **Model Context Protocol (MCP)** is the protocol launched by Anthropic in November 2024. MCP is an open-source protocol that enables secure two-way communication between AI chatbots/agents and diverse data sources and applications like databases, APIs, and enterprise tools. It acts as a standardized way to provide tools and context to LLMs and AI Agents, eliminating the need for custom integrations for each new data source. MCP speeds up development, reduces maintenance burdens and ensures data security through clear system boundaries. There are already several ready-to-use MCP servers that provide AI Agents with real-world access to various platforms, including for example Docker, File System, GitHub, Google Drive, Google Maps, Notion, Perplexity, PostgreSQL, Redis, Slack, Stripe.
- **Agent 2 Agent (A2A)** is the protocol launched in April 2025 by Google in collaboration with over 50 companies (including Atlassian, MongoDB, PayPal, Salesforce, and SAP). A2A is an open protocol that enables interoperability between AI Agents. Similarly and complementary to MCP, A2A allows AI Agents, even if built with different technologies, to communicate, coordinate and securely exchange information. It is designed for complex enterprise scenarios like supply chain management and internal workflow automation, supporting rapid or long-lasting tasks, including multimodal interactions (e.g. text, audio, video). AI Agents can expose their capabilities via an "Agent Card," fostering an ecosystem where AI Agents act as digital colleagues working together in real-time.
- **NLWeb** is the standard introduced by Microsoft in May 2025, NLWeb is a new standard similar to HTML that enables direct and semantic interactions between users, web content and AI Agents. It allows websites to offer conversational interfaces based on custom AI models and proprietary data, making their content searchable and accessible to AI Agents. Microsoft envisions an "open agentic web" where AI Agents make decisions and perform tasks on behalf of users or organizations.

Agentic AI is an intelligent system that works like a team/network of AI Agents. The AI Agents collaborate (also thanks to the above protocols and standards), plan and adapt:

1. Set their own sub-goals.
2. Use APIs, tools and browsers simultaneously.
3. Work in role-based teams (planners, critics, doers).

Real-world examples are:

- **GitHub Copilot Workspace:** end-to-end development cycles.
- **AutoGPT & AutoGen:** self-driven agents that reason, iterate, and act.
- **CrewAI / LangGraph:** orchestrated teamwork across autonomous agents.

The main **plus/opportunities for enterprise** are cost reduction (+20%), productivity boost (+30%), AaaS (Agents-as-a-Service - paid per outcome), Multimodal insights (text, image, audio - without silos).

The **Agentic RAG** is a full-blown orchestration layer over Retrieval Augmented Generation (RAG). RAG improves LLMs by incorporating information retrieval (domain-specific/updated/private info), reranking relevant information before generating responses. LLM behaves like an AI Agent: planning, making decisions, analyzing the prompt and choosing what tools or strategies to use next. Agentic RAG is changing everything in terms of:

- **Autonomy:** analyses, summarizes and crosses reference data independently.
- **Multi-step reasoning:** breaks complex prompts into sub-tasks.
- **Proactive retrieval:** anticipates gaps in data and fetches missing context before finalizing responses.
- **Adaptive learning:** meta-agents refine strategies based on real-time feedback and ever evolving datasets.

Enterprises leveraging Agentic RAG are already generating great performances, such as **40% higher accuracy** in legal analysis and medical research.

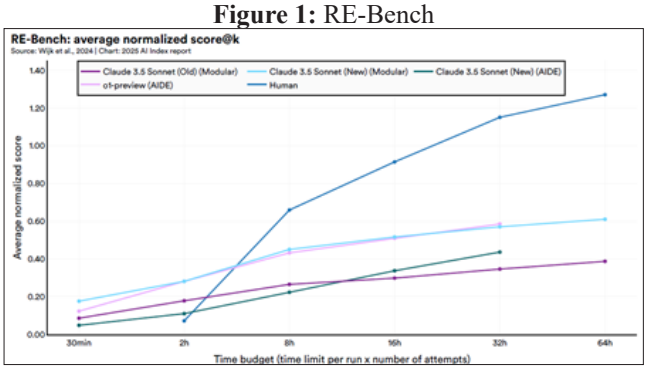
Benchmarks and performance evaluation of AI Agents

The evaluation of AI Agents is a rapidly evolving field, reflecting their growing sophistication and diverse real-world applications. Here's an overview of how **main benchmarks** are used to assess the capabilities, reliability and effectiveness of AI Agents:

- **Browse Comp:** a benchmark that measures the ability of AI Agents to find information that is difficult to find on the web.
- **DS-Bench:** a benchmark that evaluates AI Agents on realistic Data Science tasks (ranging from data analysis to data modeling).
- **GAIA** (General AI Assistants)**:** a benchmark that evaluates AI Agents (including LLM-powered assistants) in performing complex and real-world tasks. The benchmark evaluates complex reasoning, multi-modal handling, web browsing & tool use and task generalization.

- **RE-Bench:** a rigorous benchmark that evaluates complex tasks over time. According to AI Index Report 2025 of Stanford University:
  - in short time-horizons (in 2 hours) top AI systems outperform human experts (AI systems score 4 times higher than human experts);
  - as time increases (in 5 hours) human experts reach top AI systems (AI systems achieve scores comparable to those of human experts);
  - in longer time-horizons (in 32 hours) human experts outperform top AI systems (human experts score 2 times higher than AI systems).

Figure 1: RE-Bench



- **Spreadsheet Bench:** a benchmark that evaluates the ability of AI models to modify spreadsheets (derived from real-world scenarios).
- **Visual Agent Bench (VAB):** a benchmark that reflects the growing multimodality of AI models, assessing performance in embodied agents in different environments: Graphical User Interface (GUI) agents (interacting with mobile and web applications) and visual design agents (e.g. Cascading Style Sheets - CSS – debugging, figuring out what might be the problem when you have unexpected layout results). According to AI Index Report 2025 of Stanford University, most proprietary AI models currently average around a 20% success rate on VAB, indicating they are "far from ready for direct deployment in agentic settings".
- **Web Arena:** a benchmark that evaluates the performance of web browsing AI Agents in completing real-world web tasks.

AI Agent **performance measurement** relies on a combination of technical, business, user-centric and return-on-investment metrics, as well as related KPIs.

Core technical metrics

- **Task Completion Rate:** the percentage of successful task executions (fundamental for basic efficacy).
- **Accuracy:** how closely the AI Agent's outputs match the correct or desired outcomes.
- **Latency:** how quickly the AI Agent responds to requests or completes tasks.
- **Efficiency:** the volume of tasks handled per unit of time also considering the resource utilization (CPU, memory, tokens usage).
- **Hallucination Rate:** the frequency with which the AI Agent generates incorrect or unreal information.

Main business-impact metrics

- **Autonomy Ratio:** the proportion of tasks completed without human help, reflecting true AI Agent independence.
- **Agent Efficiency Index:** compares the number of actions taken by the AI Agent to an optimal, benchmarked path, penalizing unnecessary steps.
- **Context Awareness Score:** how well the AI Agent integrates user or environmental context for personalization and accuracy.
- **Deviation Rate:** the frequency of instances of behavior straying from intent, policies or ethical guidelines.
- **Availability:** the percentage of time the AI Agent is online and responsive.

Core user Experience Metrics

- **User Satisfaction:** the direct feedback and satisfaction scoring from users interacting with the AI Agent.
- **Time-to-Resolution:** how quickly the AI Agent fully resolves issues compared to traditional methods.
- **Deflection Rate:** the percentage of tasks resolved entirely by the agent without human escalation.

Main Return on Investment (ROI) metrics

- **Cost Savings:** the reduction in operational costs attributable to AI Agent automation (e.g. fewer support hours, faster task completion, increased sales).
- **Revenue Generation:** the increase in turnover attributable to AI Agent (e.g. upselling/cross selling, new revenue streams).
- **Cost of Investment:** the reduction of the investment cost attributable some way to AI Agent.

Successful benchmarking and KPI tracking use a mix of these metrics to inform iterative improvements. Strong benchmarks and KPIs ensure AI Agents are not only accurate and efficient, but also robust, trusted, and aligned with organizational goals.

Real-World manifestations of AI Agents

Several leading AI companies are already deploying various forms of AI Agents for final users, **listed below by launch date:**

- **Google's "Astra"** (announced in May 2024) was the first manifestation of a Google's AI Agent/feature. It provides real-time assistance via mobile cameras/screens/devices (e.g. smartphones), interpreting images, videos and audio.
- **Anthropic's Claude 3.5 Sonnet "Computer Use"** (launched in October 2024) allows the AI Agent to interact autonomously with a computer desktop, simulating human behavior to automate tasks (e.g. moving cursors, clicking buttons, typing text).
- **Salesforce's "Agentforce"** (launched in October 2024) is a suite of autonomous AI Agents capable of performing enterprise tasks (e.g. customer support, lead generation, and marketing optimization). These AI Agents continuously learn and adapt, operating on any channel, with pre-configured and customizable options.

- **Google's "Mariner"** (launched in December 2024) is an interactive feature of Gemini that acts as an AI Agent. It autonomously navigates the web for tasks (e.g. online shopping, finding alternatives to unavailable items). The user only gives final confirmation for the purchase.
- **Google's Gemini "Deep Research"** (launched in December 2024) is the AI Agent that creates comprehensive reports on complex topics with academic rigor and source accuracy. The user provides a prompt (e.g. compare the best smartphones 2025). Deep Research generates a customizable research plan (e.g. which smartphones, which features). Once the research plan is approved by the user, Deep Research performs an in-depth search on the web and creates a report (with sources) that can be interacted with. There is also an enterprise version.
- **Google's "Agentspace"** (launched in December 2024) is a platform to simplify the creation and implementation of AI Agents in organizations/enterprises (e.g. Banco BV, KPMG, Wells Fargo), integrating it with existing workflows and breaking down data silos whether the required information resides in common work applications (e.g. Google Workspace, Microsoft 365), specialized tools (e.g. Salesforce, Service Now) or even web content. Over time new AI features have been added e.g. "Agent Designer" (April 2025) a no-code interface for building custom AI Agents for a better organization and "Agent Gallery" (May 2025) a centralized view of available AI Agents in the enterprise.
- **OpenAI's "ChatGPT Tasks"** (launched in January 2025) was the first manifestation of an Open AI's AI Agent/ feature. It provides scheduled tasks, sending notifications or automatically executing preset requests like daily AI news briefings.
- **OpenAI's "ChatGPT Operator"** (launched in January 2025) is a Computer-Using Agent (CUA) that interacts with Graphical User Interfaces (GUIs) on behalf of the user, achieving 87% success in web-based tasks. It combines GPT-4o's Vision capabilities with Advanced Reasoning and Reinforcement Learning, asking for human intervention for credentials or final confirmation for actions like sending emails.
- **"Perplexity Assistant"** (launched in January 2025) is the AI Agent that performs tasks and answers questions, interacting with other Android/iPhone apps (e.g. Spotify, YouTube, and Uber). It's also multimodal (e.g. meaning you can ask it questions about what's on your screen (Time Square), e.g. "Get me a ride to go there", the assistant automatically opens Uber with available rides from your position to that destination.
- **"Perplexity Deep Research"** (launched in February 2025) is the AI Agent designed for articulated investigations, mirroring human research methods. It generates reports on a topic specified by the user, in three simple steps: 1) performs a web search based on the user's query; 2) reads the results, analyzes them and processes them; 3) prepares a complete report in less than 3 minutes for most searches.
- **Microsoft's Copilot "Think Deeper"** (launched in February 2025) is the AI Agent/feature/mode integrated into Copilot, designed to tackle complex topics like major purchases (e.g. electric car) or career changes, offering comparisons, pros/cons and impact analysis.
- **OpenAI's "ChatGPT Deep Research"** (launched in February 2025) is an AI Agent integrated into ChatGPT that excels at complex, articulated online investigations. It navigates the web for 5-30 minutes to generate reports, offering significant time savings, in-depth analysis, structured results and access to hard-to-find information.
- **"Manus AI"** (launched in March 2025) is the general-purpose AI Agent released by "The Butterfly Effect". Manus ("Mind" in Sanskrit) is capable of planning and coordinating multiple sub-agents (using various LLMs operating independently: e.g. Anthropic's Claude, Alibaba's Qwen) to manage complex everyday tasks (e.g. organizing a trip, comparing insurance policies or searching for properties) and work tasks (e.g. CV evaluation, researching B2B suppliers, code writing or developing interactive websites) autonomously. It performs complex tasks with great transparency, allowing users to observe what the AI Agent is doing and to intervene at any time.
- **Google's "Idea Generation"** (launched in April 2025) is the ready-to-use AI Agent that uses hundreds of AI Agents to generate, refine and rank innovative ideas for enterprise users.
- **Microsoft's "Copilot Researcher"** (launched in June 2025) is an AI Agent within Microsoft 365 Copilot. It functions as an advanced reasoning AI Agent designed to handle complex, multi-step research tasks at work. It can independently analyze, synthesize and summarize information from a combination of your work documents (emails, files, meetings, and chats), enterprise data sources and the web. The AI Agent is capable of planning and executing research workflows, asking clarifying questions and delivering tailored, structured reports with full source citations.
- **Microsoft's "Copilot Analyst"** (launched in June 2025) is a reasoning-powered AI Agent specifically designed for advanced data analysis and insight generation within the Microsoft 365 suite. It acts as a virtual data scientist, enabling users to analyze complex datasets (such as Excel, CSVs, databases) and extract insights through multi-step reasoning.
- **Google's "Gemini CLI"** (launched in June 2025) is an open-source AI Agent that brings the power of Gemini directly to your pc/terminal (lightweight access to Gemini). Gemini CLI (Command Line Interface) brings Gemini's power to developers' terminals for coding, but also for other final users for content generation and task management.
- **OpenAI's "ChatGPT Agent"** (launched in July 2025) is a general-purpose AI Agent combining Operator's web interaction, Deep Research's information synthesis and ChatGPT's conversational skills. It can perform complex tasks end-to-end using a virtual computer, dynamically learning and optimizing its approach. Users remain in control and are able to intervene, clarify instructions, or pause tasks. Similarly, ChatGPT can proactively ask you for

additional details when needed to ensure the task remains on track. ChatGPT Agent outperforms previous AI models and even significantly outperforms human performance in data analysis tasks 89,9% (vs human 64,1%) and data modelling tasks 85,5% (vs human 65,0%) according to the DS-Bench indicator (Data Science benchmarking). On the Spreadsheet Bench, it achieves 45.5% (vs. human 71.3%), but it is better than other AI models (vs Copilot in excel 20%). On the benchmark Web Arena it achieves 65.4% (vs. human 78.2%) in the performance of web browsing. ChatGPT Agent's applications include automating repetitive workplace tasks (e.g. converting dashboards to presentations, rescheduling meetings) and personal tasks (e.g. planning travel, scheduling appointments).

Level of Autonomy: Assisted AI Agents and Open Source Agents

According to the MIT Technology Review, it is useful to characterize **AI Agent** systems on a spectrum of **autonomy**:

- The **lowest level:** no impact on relevant decision-making flows (e.g. chatbots that greet you on a company website).
- **Intermediate levels:** decide which human-provided steps to take; tool callers run human-written functions using AI Agent-suggested tools (e.g. determine which functions to do when/how).
- The **highest level:** fully autonomous AI Agents (e.g. write/execute new code without human constraints or oversight) that can take action without further human approval (e.g. moving around files, changing records, communicating by email).

Each step represents a removal of human control and must be subject to a risk-benefit assessment, keeping in mind the goal: promoting human well-being (not increasing efficiency at all costs).

According to Hugging Face (global start up in responsible use of open-source AI), when AI systems control multiple sources (private communications and public platforms) simultaneously, the potential for harm increases exponentially. Therefore, we need to **keep humans in the loop** while using AI Agents.

Giving up control, bit by bit: AI Agents are based on LLMs, which are unpredictable and prone to significant, comical and/or offensive errors:

- When an **LLM** generates **errors** in a chatbot, all the mistakes remain inside that conversation.
- But when an **AI Agent** performs **errors** or actions we didn't intend (e.g. manipulating files, impersonating users, or making unauthorized transactions) - accessing both private communications and public platforms – it could share personal information on the web. The spread of that information (true or untrue) could be amplified by further sharing (e.g. on social media) to create reputational damage. We imagine that "It wasn't me—it was my AI Agent!" will soon be a common refrain to excuse bad output.

Historical precedents demonstrate why **maintaining human oversight is critical** (e.g. in 1980, computer systems falsely indicated that over 2,000 Soviet missiles were heading toward North America; human cross-verification over different warning systems avoided catastrophic events).

Autonomous systems prioritized speed over certainty. The development of AI Agents must occur alongside the development of guaranteed human oversight in a way that limits the scope of what AI agents can do.

According to Hugging Face, **open-source AI Agent** systems are one way to address risks, since these systems allow for greater human oversight of what systems can and cannot do. This approach stands in stark contrast to the prevailing trend toward increasingly complex, opaque AI systems that obscure their decision-making processes behind layers of private owned technology, making it impossible to guarantee safety.

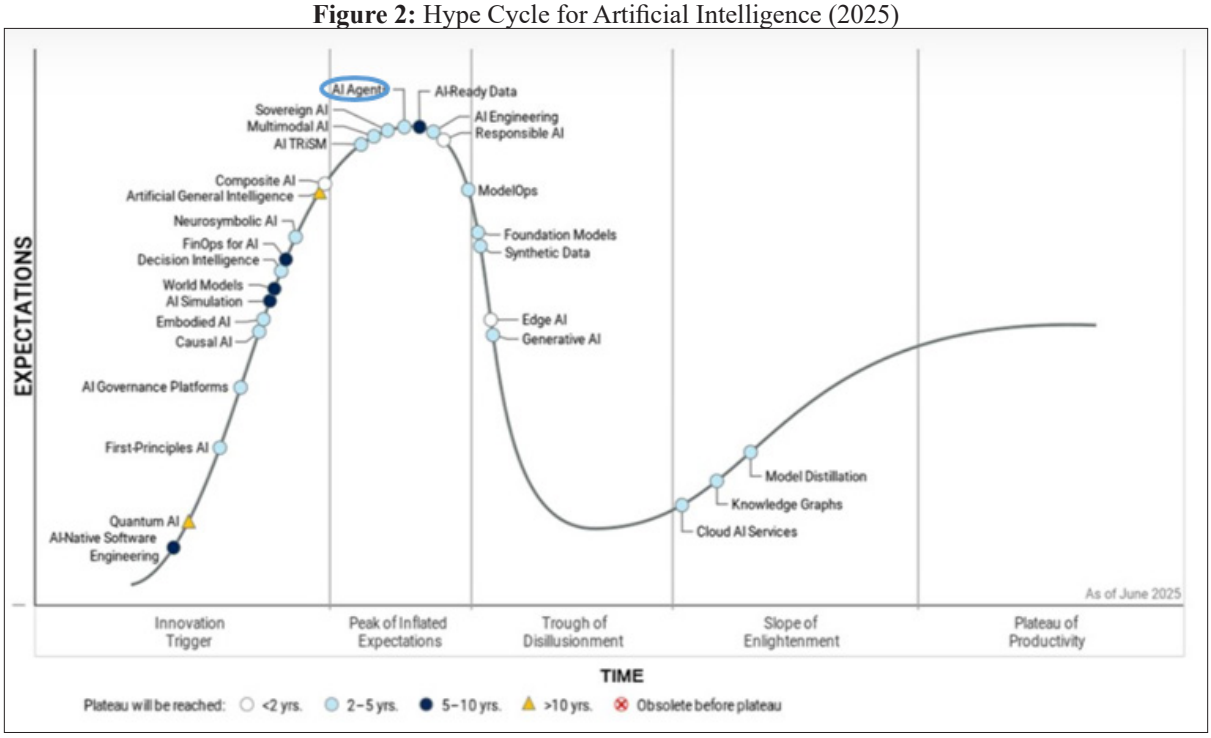
Current status and future development of AI Agents

Despite the rapid advancements and promising applications, some sources indicate that **AI Agents are still in their very early experimental stage**. Currently, there is still no AI Agent in production that sets its own goals, adapts in real time and delivers tangible results without human supervision.

A Stanford University/Google's DeepMind study found that **only about 1%** of generative AI Agents **show encouraging results,** but only in under ultra-controlled simulated environments with permanent human supervision.

The Gartner's "Hype Cycle for Artificial Intelligence (2025)" suggests that media hype of AI Agents is decreasing (as experiments and implementations are not yet ready to deliver expected results), though investments need to continue seeking for maturity. It will require time, serious research, realistic experiments and healthy development for AI Agents to offer interesting results.

Figure 2: Hype Cycle for Artificial Intelligence (2025)



As Dario Amodei, CEO and Co-Founder Anthropic, announced in 2025 "*By 2026 or 2027 we will have AI systems that are broadly better than almost all humans at almost all things*".

With reference to the **future development of AI Agents**, many researchers believe that 2027 could see an AI so advanced that "a single copy of the model, running at human speed, will already be qualitatively better than any human in AI research" in all strategic fields: from programming to hacking, from the creation of biological weapons to the ability to make political decisions.

AI Agents: Smart Colleagues and Smart Workforce

By mid-2025, current AI Agents can already be considered **Smart Colleagues**, capable of performing tasks as well as an intern would, that is, producing quality outputs that, however, require human review.

As Armand Ruiz, VP of AI Platform of IBM, declared in 2025 *"The future of AI is Agentic… They are still in the very early experimental stage… we will soon start collaborating with AI Agents instead of humans for some specific tasks"*.

Future **Smart Employees** will have a significantly higher level of autonomy: these digital entities will have elements that will make them more similar to real human colleagues:

- Their own "memory".
- Well-defined roles within the company.
- Even personal company accounts and passwords.

Solutions are already being developed that allow for

- Detailed tracking of virtual workers' account activities.
- The creation of new account classification systems for managing autonomous digital identities.
- The detection any suspicious activities that cannot be traced back to human trace.

Gartner published the "Top Strategic Technology Trends for 2025 report", on how, **autonomous AI Agents** (i.e. Agentic AI) **"will dramatically upskill workers and teams, enabling them to manage complicated processes, projects and initiatives through natural language."**

There are already **AI Agents specialized in specific fields**: cybersecurity (e.g. Darktrace Antigena Agent), robotic/process automation (e.g. UiPath), software engineering/coding (e.g. Devin AI).

As Asha Sharma, Corporate VP & Head of Product Microsoft AI Platform, announced in 2025 *"With AI agents, Artificial Intelligence becomes a workforce… just assign a task to an AI agent and then review it"*.

There are platforms for building AI Agents with a particular focus on creating an **AI workforce** (e.g. Relevance AI) which **builds teams of AI Agents that deliver human-quality work.**

Furthermore, there are many other relevant perspectives. When evaluating the **future potential of Smart Colleagues**, it may be useful to consider that the AI explosion will be concentrated in 2027, leading to:

1. The **"superhuman programmer":** an AI system that can perform coding tasks at a level equal to (or beyond) the best human code developer/software developer.
2. The **"superhuman AI researcher":** an AI system that can conduct AI research in all fields of knowledge as well as (or better than) the best human AI researcher.
3. The **"superintelligent AI researcher":** an AI system that can conduct AI research in all fields of knowledge far better than the best human AI researcher.
4. The **"Artificial superintelligence"**: an AI system that will be far better than the best human at every cognitive task.

AI Agents vs human resources: the evolution of relevant human skills

The integration of AI Agents into the workplace (capable of managing entire activity flows – not just outputs) is redefining who does what and the relevant human skills in the new human-machine balance, highlighting a genuine shift in focus. So-called **information skills** (e.g. data analysis, knowledge updating), historically central to medium and high income professions, are **in decline**, as they are among the first activities that AI Agents can perform autonomously, reducing the need for direct human supervision.

Conversely, **interpersonal, organizational and decision-making skills** (e.g. managing resources, negotiating, coordinating activities, making decisions in ambiguous contexts, leading teams) are **increasing in importance**. These are all areas in which the value of the human lies not in speed or precision, but in the ability to interpret, connect, mediate and hold others accountable.

The **demand for human resources** is not disappearing, but **is shifting toward those who can orchestrate processes, people and conflicts** in a technology-enhanced environment. A transition that has direct implications for training, career guidance, and career path design. (Stanford University, 2025).

We need to **rethink the educational system** and to **reskill current workers** (especially managers and executives): the notional model, which worked since the industrial revolution until yesterday, gradually specializing the workforce with more specific training, is no longer profitable and effective.

Risks of AI Agents: tools and rules for Responsible Use

The integration of AI Agents into daily work and personal life presents significant risks that must be carefully managed:

- **Data Privacy:** AI Agents require access to personal information (e.g. emails, calendars, browsing histories) to build and execute plans, raising concerns about privacy violations. When AI systems control private communications and public platforms simultaneously, the potential for harm increases exponentially if errors occur, leading to unauthorized transactions or the sharing of personal information and reputational damage.
- **Decision Errors:** AI Agents may make decisions not in line with user expectations, choosing undesirable options or misinterpreting preferences.
- **Bias**: AI Agents execute actions that can be influenced by biases present in their training data.
- **Structural Risks** (Artificial General Intelligence): the most difficult risks to predict concern unintended consequences from AI multi-agent systems interacting with human complexity. Dangers include creating perfectly credible false information, undermining public trust or gradually taking control of economic and political mechanisms (e.g. influencing global trade policies), potentially leading to "machines in command, without there having been intentional action in this direction".

To mitigate these risks, a **"humans in the loop" approach** is critical.

There are useful tools to assess and guide the **responsible deployment of AI Agents** according to the study "Future of Work with AI Agents" (Stanford University, 2025):

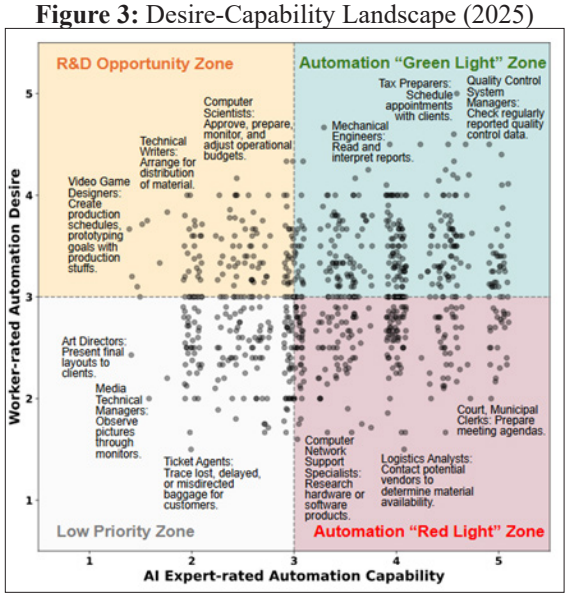
Human Agency Scale (HAS): defined a scale to assess the degree of human involvement desired or necessary to carry out work tasks in the presence of an AI Agent. A 5-level scale (H1-H5) measures the degree of human involvement needed for effectiveness and quality:

- H1 (AI drives) involves complete AI Agent autonomy (e.g. self-reporting).
- H2 (AI needs input from humans): the AI Agent performs tasks autonomously, but with human input (e.g. payment/trading).
- H3 (equal partnership): Humans and AI Agents are configured as equal partners in symmetrical cooperation, similar to what happens in well-coordinated human teams (e.g. data analysis).
- H4 (Human needs AI support): the human role is fundamental, AI Agents play ancillary roles (e.g. training).
- H5 (Human drives) means AI Agents play minimal roles (e.g. forums).

Desire-Capability Landscape: by comparing workers' automation desires/needs (ordinates) and technological automation capabilities (abscissas), a four-quadrant map (desire-capability landscape) is created, which allows each task to be classified into four zones:

- Green Light Zone (high desirability and high technical capability): tasks ready to be automated in a socially acceptable manner (e.g. reconciling accounting documents, preparing standard materials/slides for training, entering repetitive data into CRMs or generating meeting reports from recordings).
- Red Light Zone (high technical capability, but low desirability): tasks that can be automated, but are at risk of generating friction, rejection or undesirable effects (e.g. educational feedback for students, personalized interactions with vulnerable patients/clients, supervising colleagues or assigning tasks).
- R&D Opportunity Zone (high desirability, but low technical capability): tasks on which to focus research and development efforts (e.g. intelligent synthesis of legal/regulatory documents, dynamic reconfiguration of schedules/shifts based on individual preferences or proactive analysis of problems in complex production flows).
- Low Priority Zone (low desirability and low technical capability): tasks to be considered in future, after the evolution of the desirability/tech capability (e.g. mediating conflicts between people, evaluating the performance of creative professionals or facilitating complex and ambiguous decision-making processes).

Figure 3: Desire-Capability Landscape (2025)



The integration of AI Agents into the workplace also redefines:

- Required human skills (less information skills, more interpersonal, organizational and decision-making skills).
- The role of humans (less doing, more planning/supervising).
- The methodological approach of humans (less reactive and static, more proactive and dynamic).

To ensure a **responsible use of AI Agents, 10 Golden Rules** can be helpful:

1. **Integrate AI Agents into your daily routine,** starting with areas where you have personal/professional knowledge.
2. **Choose the most functional AI Agent** for each specific task.
3. **Have clear goals and desired outputs** in mind before using AI Agents.
4. **Write an initial prompt clearly** with context, instructions and the request.
5. **Check and approve the AI Agent's plan** to guide its performance.
6. **Don't be discouraged by initial misunderstandings;** modify the prompt/plan or add examples.
7. **Reiterate and refine the prompt/plan** until it is completely clear and in line with your aim.
8. **Accept the final output of AI Agent as a good draft** that needs human review.
9. **Identify and mitigate potential hallucinations** in the final output of AI Agent (report/research/action).
10. **Check and improve the final output**, always keeping humans in the loop.

AI Agents Revolution: thinking out of the box

Another aspect should also be considered. We need to **open our minds to collective solutions**, rather than simply addressing individual concerns ("AI Agents will steal our jobs"). We should also **think about the future we can create**, rather than the one we have to endure.

While the competition between companies developing AI Agents sometimes requires significant corporate investments (which can sometimes even result in layoffs, to free up capital for investment), the **AI Agents Revolution also offers an opportunity to reimagine true progress**. We ought to avoid approaching this innovation as if we were merely slaves to the technological advancement or the indiscriminate economic growth (especially if the benefits are increasingly undistributed). To do this, **we must first redefine and redesign work**: from an activity necessary for income and social positioning, to an opportunity to create real added value to ourselves, to our companies and hopefully to our global society. We already have all the tools and resources to do so.

It would seem however that one fundamental thing is missing: **trust in others and in the future**. And that, fortunately or unfortunately, depends entirely on us.

Conclusion: the future of Human-AI Agent collaboration

AI Agents are poised to redefine the future of work and daily life, evolving from mere tools to **autonomous, goal-oriented "Smart Colleagues"**. They perform tasks independently, reason through complex problems, create actionable plans and iterate towards goals.

While still act in an early experimental stage with performance varying across complex tasks, the rapid advancements, market projections and emerging interoperability standards like **MCP** and **A2A** indicate their transformative potential **(Agentic AI).**

The vision is clear: **AI Agents will increasingly automate repetitive tasks, augment human capabilities** in analysis and problem solving and coordinate in multi-agent systems to achieve complex outcomes.

However, this **future requires humans to take an ethically based approach that addresses critical risks.** Such as data privacy, decision errors and inherent biases. The emphasis on keeping **"humans in the loop" through assisted AI Agents**, coupled with transparent development and robust oversight mechanisms, is paramount to fostering trust and ensuring beneficial outcomes.

As human skills evolve to focus on orchestration, interpersonal relations and decision-making in ambiguous contexts, the **collaboration between humans and AI Agents** will become a defining characteristic of the **modern workplace**.

AI Agents are not just about increased efficiency; they **represent a fundamental shift towards a dynamic, collaborative ecosystem** where Artificial Intelligence truly functions as a helpful, intelligent colleague.

Optimists and pessimists, conservatives and innovators, should keep in mind: "*The future of AI lies in autonomous AI Agents…"*, as declared by Marc Benioff, CEO Salesforce, in 2024.

The most important thing is to make a **responsible use of AI Agents.**

References

1. Agent force - https://www.salesforce.com/agentforce
2. Agent space - https://cloud.google.com/products/agentspace?hl=en
3. Agent2Agent Protocol (A2A) - https://developers.googleblog.com/en/a2a-a-new-era-of-agent-interoperability
4. Armando Ruiz. (2025). VP of AI Platform of IBM. https://www.linkedin.com/posts/armand-ruiz_the-future-of-ai-is-agentic-lets-learn-activity-7315327785739198464-qR9z
5. Asha Sharma. (2025). Corporate VP & Head of Product Microsoft AI Platform. https://en.ilsole24ore.com/art/with-agents-artificial-intelligence-becomes-workforce-AHezhVK
6. Astra - https://deepmind.google/models/project-astra
7. AutoGen - https://microsoft.github.io/autogen/stable//index.html
8. AutoGPT - https://agpt.co
9. Azure AI Foundry - https://azure.microsoft.com/it-it/products/ai-foundry
10. Bayer (2025) - https://www.gobeyond.ai/ai-resources/case-studies/bayer-australia-ai-google-health-marketing?2b76cf50_page=10
11. BrowseComp - https://openai.com/index/browsecomp
12. ChatGPT - https://chatgpt.com
13. ChatGPT Agent - https://openai.com/index/introducing-chatgpt-agent
14. ChatGPT Deep Research - https://openai.com/index/introducing-deep-research
15. ChatGPT Operator - https://openai.com/it-IT/index/introducing-operator
16. ChatGPT Tasks - https://community.openai.com/t/new-feature-tasks-beta-rolling-out/1091180
17. Cisco (2025) - https://newsroom.cisco.com/c/r/newsroom/en/us/a/y2025/m05/agentic-ai-poised-to-handle-68-of-customer-service-and-support-interactions-by-2028.html#:~:text=May%2027%2C%202025-,Agentic%20AI%20Poised%20to%20Handle%2068%25%20of%20Customer,and%20Support%20Interactions%20by%202028&text=News%20Summary%3A,be%20handled%20by%20agentic%20AI
18. Claude - https://claude.ai
19. Claude 3.5 Sonnet “Computer Use” - https://www.anthropic.com/news/3-5-models-and-computer-use
20. Copilot - https://copilot.microsoft.com
21. Copilot Analyst - https://techcommunity.microsoft.com/blog/microsoft365copilotblog/analyst-agent-in-microsoft-365-copilot/4397191
22. Copilot for Sales - https://www.microsoft.com/en-us/microsoft-365/copilot/copilot-for-sales
23. Copilot Researcher - https://techcommunity.microsoft.com/discussions/microsoft365copilot/copilot-researcher--analyst/4414795
24. Copilot Studio - https://www.microsoft.com/en-us/microsoft-copilot/microsoft-copilot-studio
25. Cornell University (2024). WebArena - https://arxiv.org/abs/2307.13854
26. Cornell University (2024). SpreadsheetBench - https://arxiv.org/abs/2406.14991
27. Cornell University (2024). RE-Bench - https://arxiv.org/abs/2411.15114
28. Crew AI - https://www.crewai.com
29. Dario Amodei. (2025). CEO and Co-Founder of Anthropic. https://www.anthropic.com/news/paris-ai-summit
30. Darktrace Antigena Agent - https://www.darktrace.com/news/darktrace-launches-antigena-version-2-0
31. Data Masters. Artificial Intelligence Courses. https://www.datamasters.it
32. Demis Hassabis. (2025). CEO of Google DeepMind. https://blog.google/technology/google-deepmind/google-gemini-ai-update-december-2024/#ceo-message
33. Devin AI - https://devin.ai
34. eBay (2025). https://innovation.ebayinc.com/stories/ebay-uses-agentic-ai-to-supercharge-personalized-ecommerce
35. Gartner. (2025). https://www.gartner.com/en/newsroom/press-releases/2025-06-25-gartner-predicts-over-40-percent-of-agentic-ai-projects-will-be-canceled-by-end-of-2027#:~:text=Gartner%20predicts%20at%20least%2015,less%20than%201%25%20in%202024
36. Gartner. (2025). Hype Cycle for Artificial Intelligence 2025. https://www.gartner.com/en/newsroom/press-releases/2025-08-05-gartner-hype-cycle-identifies-top-ai-innovations-in-2025
37. Gartner (2025). Top Strategic Technology Trends for 2025 report. https://www.gartner.com/en/articles/top-technology-trends-2025
38. Gemini - https://gemini.google.com
39. Gemini CLI - https://cloud.google.com/gemini/docs/codeassist/gemini-cli
40. Gemini Deep Research - https://gemini.google/overview/deep-research
41. Gen.AI.Works - https://invest.genai.works
42. GitHub Copilot Workspace - https://githubnext.com/projects/copilot-workspace
43. GraphDB - https://graphdb.ontotext.com
44. Hugging Face. (2025). https://huggingface.co/blog/ethics-soc-7
45. Hugging Face. DSBench - https://huggingface.co/papers/2409.07703
46. Hugging Face, GAIA – https://huggingface.co/spaces/gaia-benchmark/leaderboard
47. Idea Generation - https://cloud.google.com/agentspace/agentspace-enterprise/docs/idea-generation
48. Knowledge Graph - https://neo4j.com/use-cases/knowledge-graph
49. JPMorgan Chase. (2025). https://www.gobeyond.ai/ai-resources/case-studies/jpmorgan-coin-ai-contract-analysis-legal-docs
50. LangGraph - https://www.langchain.com/langgraph
51. Manus AI – https://manus.im
52. Marc Benioff. (2024). CEO of Saleforce. https://www.businessinsider.com/marc-benioff-salesforce-llm-ai-agents-future-podcast-2024-11
53. Mariner - https://deepmind.google/models/project-mariner

54. Markets and Markets. (2025). AI Agents Market. https://www.marketsandmarkets.com/Market-Reports/ai-agents-market-15761548.html?gad_source=1&gad_campaignid=257503036&gbraid=0AAAADxY7SxOsLmMG8shzv_9oHH8MzgTn&gclid=CjwKCAjwkvbEBhApEiwAKUz6-9tGfz5jxIDnLNA9i2cdA_LsyhapwT-PiA5IHzMjRIxZOG3Bf3KljxoCQ0sQAvD_BwE
55. MIT Technology Review. (2025). https://www.technologyreview.com/2025/03/24/1113647/why-handing-over-total-control-to-ai-agents-would-be-a-huge-mistake/
56. Model Context Protocol (MCP) - https://docs.anthropic.com/en/docs/mcp
57. Moody's (2025),.https://www.moodys.com/web/en/us/insights/data-stories/kyc-ai-risk-and-compliance-survey.html
58. NLWeb-https://news.microsoft.com/source/features/company-news/introducing-nlweb-bringing-conversational-interfaces-directly-to-the-web
59. Northwestern University. (2024). Kellogg School of Management. Northwestern University. https://www.kellogg.northwestern.edu/
60. OpenAI Agents SDK - https://openai.github.io/openai-agents-python
61. Perplexity - https://www.perplexity.ai
62. Perplexity Assistant - https://www.perplexity.ai/help-center/en/collections/11466378-perplexity-assistant
63. Perplexity Deep Research - https://www.perplexity.ai/hub/blog/introducing-perplexity-deep-research
64. Satya Nadella, (2025). CEO of Microsoft, https://www.linkedin.com/pulse/from-saas-aaas-next-evolution-enterprise-software-niraj-k-verma-ok0rc
65. Stanford University. (2025). Artificial Intelligence Index Report 2025. Stanford University. https://aiindex.stanford.edu/report
66. Stanford University. (2025). Future of Work with AI Agents. https://futureofwork.saltlab.stanford.edu
67. Think Deeper - https://www.microsoft.com/en-us/microsoft-copilot/blog/2025/02/25/announcing-free-unlimited-access-to-think-deeper-and-voice
68. Uber (2024). https://www.uber.com/en-IT/blog/genie-ubers-gen-ai-on-call-copilot
69. UiPath - https://www.uipath.com
70. Vector DB - https://vectordb.com
71. Visual Agent Bench - https://openreview.net/forum?id=2snKOc7TVp
72. Webidoo Groow - https://groow.ai