ENDERS | ANALYSIS

AI agents

The new digital middlemen

- AI agents capable of complex, self-directed tasks are becoming a reality, with capabilities set to improve dramatically through this year, and diffuse widely
- Consumer agent uptake will be hard to time, but fast when it occurs. Enterprise adoption will happen slower but with greater inevitability, as agents offer strong productivity gains across many business functions
- TMT firms should be able to capitalise on much of these potential cost savings, but are exposed to a number of specific risks around agents acting as new digital middlemen, disintermediating traditional web ecosystems within advertising and ecommerce

Hamish Low
hamish.low@endersanalysis.com

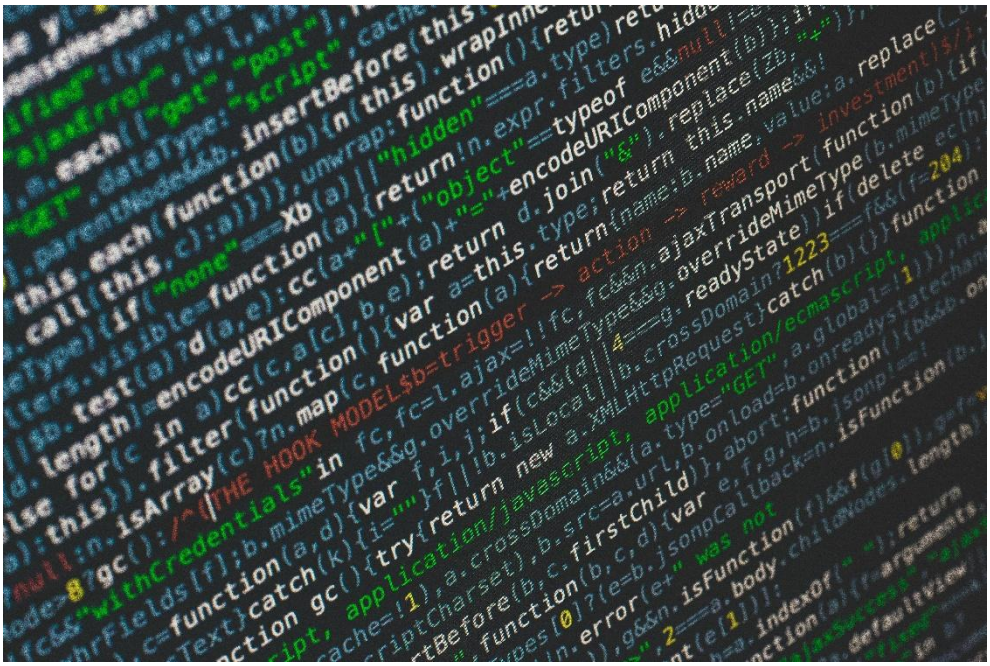
Joseph Teasdale
joseph.teasdale@endersanalysis.com

Jamie MacEwan
jamie.macewan@endersanalysis.com

Niamh Burns
niamh.burns@endersanalysis.com

+44 (0)20 7851 0900

26 March 2025



Related reports:

AI's defining year: Capex and capabilities surge [2025-019]

Telco network APIs: Real potential but still nascent [2025-018]

AI product evolution: Publisher pressure from AI search to agents [2025-007]

Big tech meets big capex: How high can it go? [2024-118]

Powerful agents are arriving soon

AI agents—systems capable of more complex, self-directed tasks—are developing rapidly, with the first true agent systems set to become widely available this year. Agents are the successor to existing chatbot-focused AI systems, with Google billing this year as the beginning of the "agentic era". This report explores the current state of the art agents being developed by the leading AI firms and how these demonstrate the kinds of products and use cases that will become increasingly disruptive as they spread.

The capabilities of agents will continue to advance rapidly through this year, while similar to other areas of AI we will see ongoing, rapid cost declines, enabling wide proliferation. If adoption of agents by consumers goes at a pace similar to chatbot systems these could rapidly become a disruptive force across the digital economy. Enterprise adoption is likely to be slower, with significantly more work needed to integrate agents into workflows and decision making processes, but has a larger air of inevitability than more uncertain consumer habits. The promise of agents is that the more powerful they become, the more they can natively navigate the inherent friction of many enterprise use cases, allowing for simpler adoption than for more obviously brittle chatbot systems. Rather than a prompt, agents have a goal they seek to accomplish, taking them away from being narrow tools and promising instead much wider disruption. A qualitative shift rather than the simply incremental pressure on the web that chatbots brought.

Agents are set to accelerate existing AI use cases for businesses, pushing greater productivity and cost-cutting benefits. They also present significant potential risks, especially for the media and telecoms sectors. Publishers dependent on web traffic have already been at the sharp end of AI development, and agents are likely to cause AI disruption to a broader set of media owners outside big platforms, and utility businesses such as telecoms selling online as they shift consumer habits and risk disintermediating existing web relationships. While chatbots largely substituted general information search queries, agents are more likely to encroach upon far more commercially valuable areas such as ecommerce. Grounding this risk in the present state of the technology is essential, and this is an evolving issue that we expect to revisit in greater detail.

Frontier firms are focused on a few high-value agents

Agents are a general concept, and to best understand their potential impact we have broken them down into the most promising emerging agent products. These products are the key areas of focus for the large tech firms developing the models that underlie these capabilities. They represent therefore both domains where the technology is most advanced, and where there are the strongest short-term revenue opportunities. Over time agents will diffuse far more broadly than this current frontier, with enterprises developing internal agents much more tailored to specific business needs, and third-party application developers building their own products on top of models from the likes of OpenAI, Google or Anthropic.

In this report we discuss five emerging categories of agents:

- **Agents that navigate computers** through human interfaces (screenshots and mouse & keyboard commands), which are limted in their current form, but are an important building block for more powerful agents to come
- **Research agents that browse the web,** notably OpenAI and Google's 'Deep research' products, which are able to utilise the tools and resources of the web to produce far superior outputs to research-style queries than standard chatbots
- **Coding agents** that are one of the most rapidly developing areas, building on what has already been one of the strongest AI use cases, and looks set to be ground zero for the most ambitious efforts at automation

- **Personal assistants** as a long-standing category where agents look set to give a new lease of life to competing efforts between Google, Apple and Amazon to build truly useful and autonomous personal assistants
- **Google's AI co-scientist** as an interesting example of multi-agent systems, that tie together a host of specialised AI agents and has produced very impressive results in doing so, laying out one possible future for how these systems could evolve

Agents offer broad productivity benefits

AI agents are systems capable of taking on more complex, multi-stage, goal-oriented tasks for a user. Rather than the chatbot call and response models, agents are equipped with various tools to be able to autonomously plan and execute a certain task. They are self-directed, being able to take a certain prompt and choose how to tackle it, whether that be web-searching or writing code, before returning the desired output. This is a major upgrade over existing chatbot systems, which functioned as useful but brittle tools, extremely capable in certain domains—generating boiler-plate code or summarising large numbers of documents—but limited in the scope of the tasks they could accomplish.

Agents remove much of the friction involved with chatbot systems. Rather than having to build a complex superstructure to feed the right data to a chatbot, and get some output that has to be then reviewed and integrated, agents can feed in far more directly to existing workflows. Agents, rather than having an immediate prompt they seek to answer, instead have a goal—conduct a SWOT analysis, analyse and rank these CVs—and can plan how it seeks to accomplish this task, and use tools and query web or enterprise data to do so. This allows agents to accomplish much more complex, and longer-time horizon tasks.

Agents remove much of the friction involved with chatbot systems

This capability means agents are set to bring much broader productivity benefits than their chatbot predecessors. Current agent systems still require a fair bit of handholding in being connected to the right data and tools, as well as properly prompted and resourced. They also suffer from similar (though reduced) reliability issues that have harmed the adoption of chatbots. Short-term adoption is therefore most likely in areas where outputs are more easily verifiable, with coding the principal domain most favourable to automation. As agents develop, the emphasis will be on reducing the friction limiting adoption. This advancement in the capabilities of the technology will be felt differently to the previous few years of ever better chatbots, with the change in form factor with agents allowing much more scalability. As the friction reduces, it becomes easier to simply stack more intelligence onto a process until diminishing returns are hit.

TMT will share in these broad productivity benefits, with the ability to accelerate AI use cases in cutting back-office costs, or enhancing customer care, largely dependent on simply having a minimum level of digitisation and the capacity to invest. Agents capable of automating more components of software development will then help to further lower the costs of deeper digitisation and smooth the pathway to being able harness these productivity benefits.

TMT does face a number of specific challenges

The flipside of the business benefits of agents is that AI agent products, especially on the consumer side, pose a number of potential risks. The larger the share of internet traffic that AI agents come to represent, the more pressing an issue they become for the existing web ecosystem and the advertising and ecommerce that it relies on. We see a range of possible risks:

- **Publishers** have already been forced to face up to disruptive AI developments, and we covered some potential risks in AI product evolution: Publisher pressure from AI search to agents [2025-007], with

agents further disintermediating their content the most pressing issue. As agents spread further paywalls and scoops are seriously challenged as agents can pick up any trace of a story across the open-web and deliver it to a user, making life harder for even the publishers most successful in their digital transitions

    - For those reliant on search traffic, agents are a very direct substitute, and risk complete disintermediation. Publishers are active in looking to avoid these extreme risks but the speed of development in the technology remains a significant challenge

- Similar disintermediation risks also become more significant for the wider **advertising** ecosystem due to agents. As agents become increasingly indistinguishable from human users, it will raise questions such as, does an agent acting on behalf of a human, and accessing sites as a human would count as an ad impression? Agents are likely to exacerbate existing issues around fraudelent traffic. The players most capable of navigating these ad-specific risks will be the largest platforms, with the open-web under more pressure
- For **TV** an important risk is around how AI agents could alter the discoverability of content. Voice search via an agent on the TV set would empower the tech players behind the hardware and operating systems, especially Google and Amazon, as their agents could steer users in their preferred directions. Public service broadcasters have fought hard to maintain a level of prominence on the connected TV set, but this work risks being undone if the interface shifts to voice

Telecoms as a mixed case study

Telcos will see significant gains from much of the broad productivity advances AI agents will generate, but face a number of risks to their networks, and to their ability to directly reach consumers. Telcos are well set up to deliver on the cost-saving potential that AI agents will offer. They are fairly digitised, with the main difficulties centred on managing complex legacy tech stacks, a problem that cheaper automated software engineering is ideal for. Telcos can use AI to plan and manage their networks, handle back-office functionality, and further automate customer care, reducing costs significantly in doing so. The challenge is that this kind of diffusion takes time, and a potential spike in consumer adoption would raise serious problems for managing their networks and reaching customers.

AI agents as they become more powerful could directly threaten networks through enabling elevated cyberattack capabilities. Or as they become indistinguishable from humans, indirectly harm the users of networks through more advanced DDoS attacks or more potent scams and frauds. We view this as an area where network operators have a potentially important role to play, with network APIs used to detect agent-like traffic patterns to distinguish between real end-users and AI and avert these risks (see Telco network APIs: Real potential but still nascent [2025-018] for more detail). The key task will be to develop these kinds of countermeasures before we are seeing these kinds of risks occuring at scale.

Agents also pose a risk in terms of limiting the access that telcos have to end customers. OpenAI has cited ecommerce search as one of the most powerful use cases of their "deep research" agent, with it being able to compile sophisticated comparisons of products based on user preferences. For utility services, such as telecoms or pay-TV platforms, consumers offloading this search to an agent poses novel challenges. Telcos already face somewhat similar issues, having to pay Google for placement of their direct digital channels, and comparison sites for their indirect ones. The challenge is that agents could turn all channels indirect. A user could simply ask their assistant agent to find them the best broadband deal, which it would dutifully do, returning to the user the cheapest plan that meets their quality criteria, asking simply for approval to then go through the purchase process. This would cut out telcos' ability to manage a customer journey, inform them of adjacent products, or upsell them. Rather it becomes a highly commoditised proposition judged principally on price.

An arms race between consumer AI agents and telco AI systems is possible

Is such an agent likely? There would be significant economic pressure for the firm controlling the agent to accept a commercial deal for prominence from a telco, effectively reforming the existing ecosystem around search, though with telcos placed at one further remove from the end-user. Consumers do likely have limits however in terms of how much commercial bias they would accept from black-box AI systems. An advertising model premised on directly steering users, or disguising that advertising is shaping an AI agent's decision making, is unlikely, with this pushing too strongly against the core useful functionality of an agent. Some advertising is likely to appear, similar to Perplexity's current version within its AI search, where certain suggested answers appear, clearly labelled as sponsored. As such, in the short-term, telcos are unlikely to be able to use a commercial relationship to remove the risks of more objective consumer AI agents.

Telcos are likely to seek to use AI themselves to strike back, for instance through much more aggressive use of price differentiation between consumers, as VMO2 is already looking to do. This would involve charging far more individualised pricing, perhaps based on whether they know a site visitor is an agent or a human, or as they are currently looking to do using geographic and socioeconomic data. The potential would be for more efficient balancing between backbook and frontbook, trying to ensure each customer pays the maximum they would be willing to. The practicality of such a system is yet to be seen. Efforts on this front are ongoing, with their success hard to gauge from the outside. Regulators and policy makers will be quite alive to the potential adverse effects of such pricing policies on vulnerable users, leading to constraints in certain customer segments. Telcos also function within strongly competitive markets, which limits the range of pricing power in which AI would be able to optimise.

AI agents are likely to bring accelerating productivity gains for telcos, as well as opening up new commercial opportunities around using their networks via APIs to play an important role in handling growing quantities of AI traffic. Telcos will just need to be wary of how consumer agent products could potentially impact how they get their products to consumers.

AI ambitions stretch far further than disrupting the web

Disrupting current web ecosystems is not a priority for these firms, rather it is simply a casualty in their pursuit of ever more powerful AI systems. Web agents are a useful step in the wider goal of building systems more akin to a 'drop-in remote worker' capable of wide automation of cognitive labour. This ambitious vision is what is working to justify sky-high valuations and hundreds of billions of spending on data centres. The goal of OpenAI, Anthropic or the various other players vying to be at the frontier is the development of much more powerful AI systems. Building consumer products such as web agents is important for generating the short-term revenue growth they need to keep the fly-wheel spinning, but their ultimate prize is a far more lucrative enterprise market. The success of agents in driving this shorter-term growth is crucial though to the narrative that big tech has been putting forward, see AI's defining year: Capex and capabilities surge [2025-019].

The first agents are coming into view

The best way of narrowing down the current art of the possible is looking at what the various major AI players are putting forward. Google, OpenAI, Anthropic and Meta are all converging on similar ideas around what agent products are most likely to function and find a market in the near-term. Common research advances underpin some of this convergence with advances in speech and vision but most importantly the development of reasoning models that extend the ability of LLMs to tackle longer and more complex tasks. The underpinnings of this new reasoning model paradigm and its potential implications is a topic we will explore in greater depth soon.

The capabilities currently being offered with these early agents are impressive but still fairly limited, but in large part their release is intended for gathering data for rapid iteration and improvement. We would

expect both significantly more capable agents to be available later this year, and for agents to become far cheaper and more widely available on a similar time scale.

While there will be significant growth of applications that use different kinds of agentic systems, leveraging already existing powerful AI models and building for more specific use cases, in this report we are largely concentrating on what is being developed at the frontier. The capital-intensive nature of AI model training means that the cutting-edge of agents, those that can handle the highest complexity, longest and most difficult tasks, are still largely the domain of the model-makers, principally OpenAI, Google, Anthropic and Meta. These capabilities will proliferate over time, but for now looking narrowly at these leading players gives the best insight into what we can expect from the "agentic era".

Computer-use agents: cutting out the friction

These are a very direct application of research aimed at allowing models to directly use computers. This capability is essential to realising the goal of low-friction agents that can be more easily deployed into enterprises. By removing the need for complicated backend software, and instead having a general agent that can navigate a computer exactly as a human remote worker would, the barriers to adoption can be greatly lowered. Rather than spending large amounts of capital and employee time rationalising data sources and software tools to then integrate AI, this work can itself be passed off to an agent.

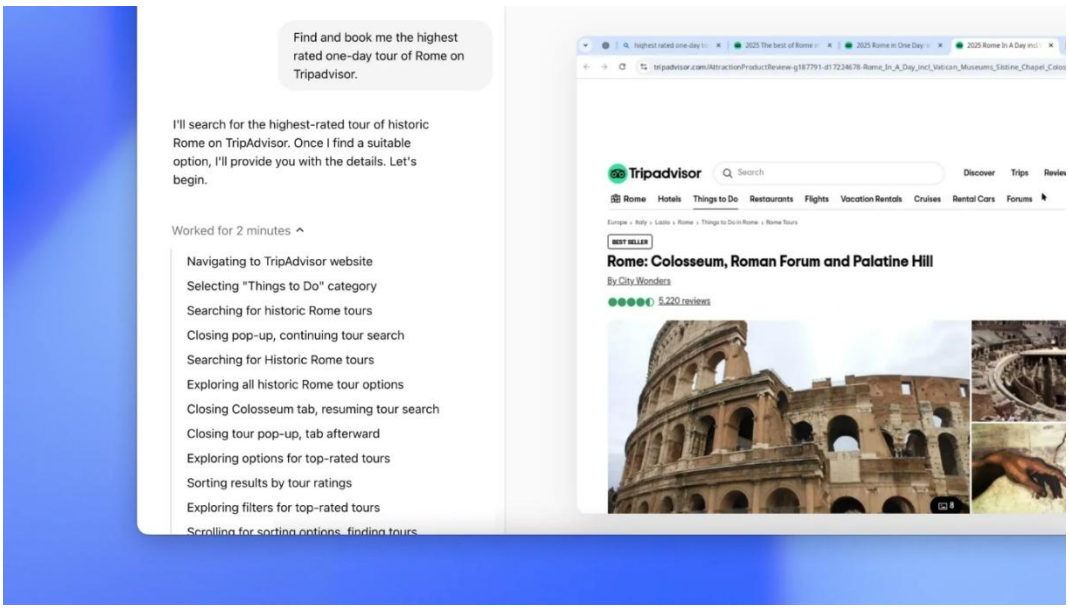
Computer-use is essential to lowering the friction of adopting AI agents

This works by models being presented with screenshots of a computer or browser, and given the ability to direct mouse and keyboard actions, interacting with software as normal human users would. This is a significant departure from classic online 'bots' that interact via the code of a webpage in more predetermined forms, with the goal being to create models entirely capable of human-style computer use.

One clear issue in this space is that the internet is not awash with videos of people screensharing and clearly highlighting each action they take (YouTube may be a significant advantage for Google here). For this reason, we have seen various limited forms of computer-use released thus far, likely so that firms can bring in the needed data to rapidly improve these systems. Anthropic was the first to release its "computer use" functionality in October last year. This was as just an experimental API for developers, though we expect that we will see a more consumer-facing version released later this year, with this functionality playing a role in Amazon's new Alexa+ assistant.

OpenAI has launched a research preview of its own version within ChatGPT. Its "Operator" system uses its "Computer-Using Agent" to navigate a browser embedded within ChatGPT to be able to take various actions such as ordering groceries or booking hotels. For safety reasons due to the risk of it going off-script or hallucinating it needs user confirmations before taking any significant actions such as submitting orders or sending emails. It also will avoid sensitive tasks, the examples of which are banking transactions or making decisions on job applications. Users can also takeover Operator's browser at any point to stop it, or input critical information.

Figure 1: OpenAI's Operator system



[Source: OpenAI]

These restrictions, and its current limited capabilities, mean that its current version is fairly gimmicky, and hard to imagine being used outside of techy early-adopters. The technology will change rapidly however as deploying this system, gathering more data, and improving the underlying model leads to a significantly better experience over time. Though for many consumer tasks the inherent friction may remain too high, or for domains such as ecommerce the time spent may be a feature enjoyed by the consumer rather than a bug to be fixed. Consumer is ultimately only the intermediate step towards getting the capability good enough to break into the enterprise market.

Google has taken a similar research experimentation approach, with its Project Mariner which matches very closely the general formula of OpenAI and Anthropic's offerings, though with tighter integration into Google Chrome. Browser capabilities have advanced much faster than general computer use, with both OpenAI and Google seeing success rates north of 80% on browser tasks already. By contrast, general computer use—the more complex navigation of an operating system, different files and applications—has seen <40% benchmarks of success, but is advancing rapidly. Google's ability to utilise chrome as a surface could provide it with a more short-term compelling product offering here. But the end goal of these models is less to skip as many steps as possible in online food delivery, and more to provide a building block for increasingly sophisticated research models. All three of OpenAI, Anthropic and Google are likely to have sufficient capabilities here to achieve that latter goal of underpinning more advanced models to come.

Figure 2: Computer-use agents

The players	OpenAI	Anthropic	Google
Status quo	Research preview of its "Operator" system on its $200 a month plan	Experimental API of its "Computer use", use within Amazon's Alexa+	Closed research preview of its "Project Mariner"
Current leader?	Relatively level playing field, similar levels of capability at this early stage		
Outlook this year?	Broader roll-out and integration directly into ChatGPT	Integration into its Claude Chatbot, capability offered to enterprise clients	Launch as an AI assistant within Chrome

[Source: Enders Analysis, company websites]

Key takeways

Current computer-use agents are largely integrated to slightly gimmicky consumer tools, but capabilities will improve quickly, with intial versions designed primarily to collect data than as truly compelling products.

By the end of the year we would expect computer-use to be integrated into more capable personal assistant agents, notably within Google Chome, and for the emphasis to have switched to using more reliable computer-use to break into the enterprise market.

Deep research agents: plugging AI into the web

Helpfully for categorisation Google and OpenAI have landed on the exact same name for their research agent products in "Deep Research" (technically Google is uppercase R and OpenAI lowercase). Google launched this feature first in December, where users with their paid plan could prompt a version of their older Gemini 1.5 model to do deep research on a specific area. To do so Gemini creates a plan, and then when the user approves goes and searches across the web and over 5-10 minutes collates together a report on the requested topic.

OpenAI's version is very similar in its structure, with the key difference being that it is supercharged with OpenAI's latest and most expensive models, and was released alongside some much more impressive headline results. Where Google initially used its optimised but cheaper last-generation model, OpenAI used a modified version of its latest and most expensive o3 model. This means that the feature is currently only available for Pro ($200 a month) subscribers, and even then is limited to 100 queries a month. OpenAI is planning to create a version with a much cheaper model (likely the smaller o3 mini) and launch for all paid subscribers "soon".

Figure 3: Deep research agents

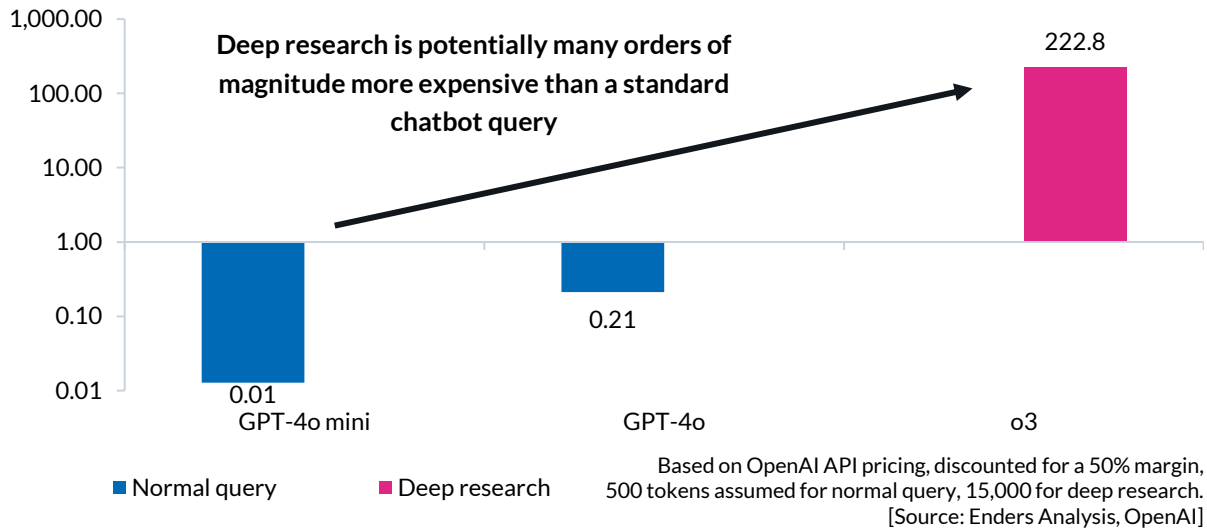
The players	OpenAI	Google
Current state-of-play	Deep research mode leveraging its powerful but expensive o3 reasoning model	Deep Research mode using its older but cheaper Gemini 1.5 model, and now its cheap 2.0 Flash reasoning model
Current leader?	The current leader, with its more expensive model capable of tackling more difficult questions, but at significantly higher cost	Less capable due to not being powered by a similar cutting-edge model, but an effective synthesiser of information at much lower cost
Outlook this year?	Integration with cheaper o3 mini model for a wider release, upgraded with an eventual o4, greater tool use and ability to plug-in to data and knowledge sources	Upgraded with improved Gemini 2 models, rolled out more widely into Gemini for Workspace

[Source: Enders Analysis]

Larger models, running for longer, means that OpenAI deep research is throwing a significant amount of computational resources at a question, perhaps $2 per deep research question, versus 0.2 cents per normal GPT-4o query (see Figure 4 below). But the results are certainly impressive. The headline benchmark was its result on the ominously named "Humanity's Last Exam". This is a benchmark of over 3,000 expert questions across a range of domains from chemistry to classics and ecology. These questions are very hard, a subset of public questions can be read here, and they largely stump existing LLMs—GPT4o gets 3.3%, even OpenAI's initial o1 reasoning model gets only 9.1%. Deep research takes the 13% score of the o3 model to 26.6%, as the use of browsing and coding tools allows it to tackle significantly more of these highly challenging questions. This benchmark was released less than a month ago, intended to be the hardest possible test of knowledge that we can give AI models, and already we are seeing agent models tackling a quarter of these questions (with the caveat that this is the text-only portion of the benchmark, with visual capabilities yet to be ported over).

Agents have accelerated progress on "Humanity's Last Exam"

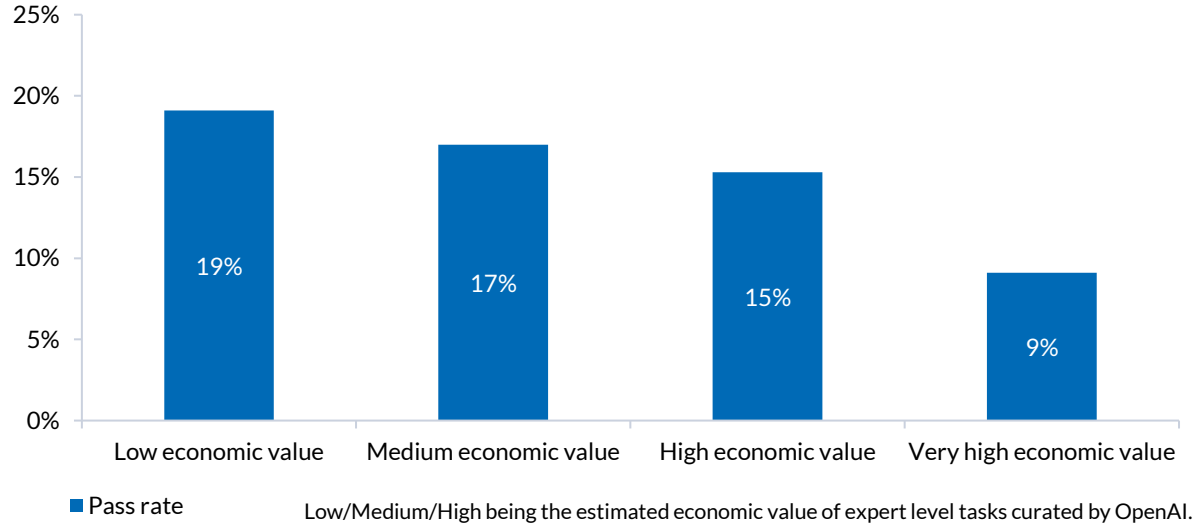
Figure 4: OpenAI illustrative cost per query (cents, log-scale)



Based on OpenAI API pricing, discounted for a 50% margin, 500 tokens assumed for normal query, 15,000 for deep research.
[Source: Enders Analysis, OpenAI]

With these benchmarks of Q&A knowledge capabilities being actively knocked over, OpenAI has moved to the next natural form of benchmarking, assessing deep research against the economic value of tasks it can accomplish. The examples OpenAI gave were researching gene therapies or complex translation tasks. The success rate is still low (Figure 5), and the nature of an internal evaluation with limited disclosure makes it hard to assess, but clearly OpenAI's focus is on getting these figures ever higher. Sam Altman has claimed deep research as being capable of "a single-digit percentage of all economically valuable tasks in the world".[1]

Figure 5: OpenAI's deep research success on expert-level tasks



Low/Medium/High being the estimated economic value of expert level tasks curated by OpenAI.
[Source: OpenAI, Enders Analysis]

Anthropic has yet to release a web-browsing feature, but has stated it is on the way, with its recent Claude 3.7 Sonnet model boasting reasoning features similar to OpenAI and Google's, though so far Anthropic has prioritised coding products. The similarity in product approaches among these leading firms reflects how the products are flowing from the fundamental research work. Once you have a reasoning model capable of thinking for longer and in a more sophisticated way, you then want to supply it with tools to accomplish useful tasks. As you would take a bright intern and when tasking them with investigating some

[1] TechXplore, "OpenAI announces new 'deep research' tool for ChatGPT", 3 February 2025

knowledge work question, expect them to use web research and other software to accomplish that task. These research agents represent pushing the boundaries around models' abilities to gather and synthesize very large quantities of information, and apply them to understanding complex questions. Once these capabilities are in place, then the goal is to increasingly move them to being able to uncover new knowledge and become "Innovators" in OpenAI's terminology, that can accelerate scientific development.

Key takeways

Web search agents are getting very capable, very fast. They will become increasingly useful as a way of finding and analysing any information publicly available on the internet, with ability to access private or enterprise information following soon.

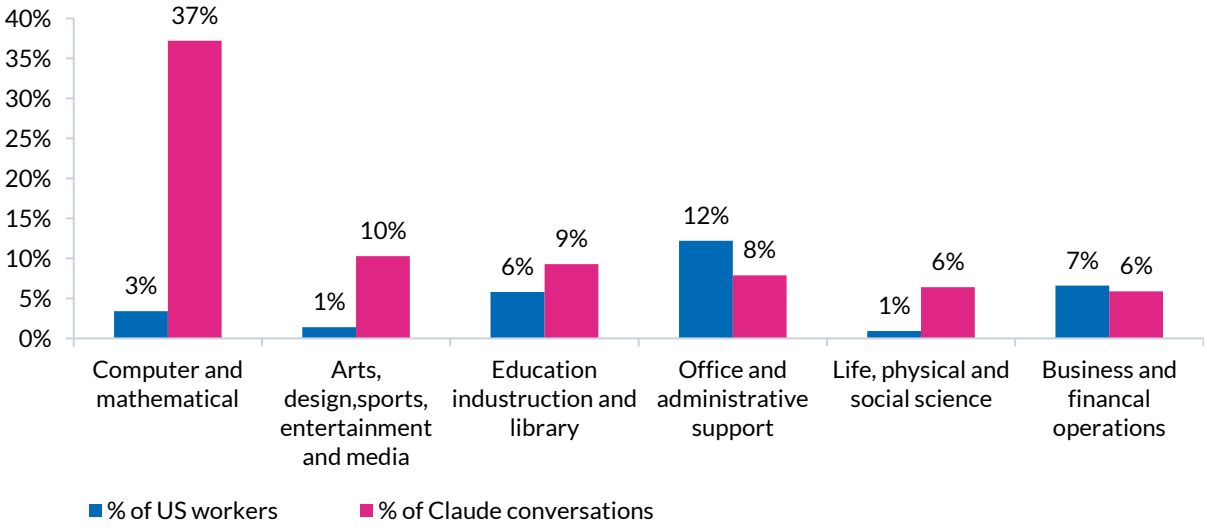
As these agents increasingly deliver better responses to more complex queries, including crucially any online purchase, they will risk a very broad disintermediation between users and websites raising serious questions for ecommerce and online advertising.

Coding agents: near-term commercial highlight

The most obvious form of scientific development to advance for an AI firm is research and development of AI itself. This involves a significant amount of coding, which happens to be a domain that AI engineers know well, have excellent datasets on, and which represents a hugely lucrative market in accelerating the productivity of (and eventually possibly replacing) millions of well-paid software engineers.

Coding has been one of the most successful current use cases of AI models. Initially with coders simply using chatbots for coding questions, now through increasingly AI-integrated coding products such as Microsoft's GitHub Copilot, Amazon's Q Developer, or start-ups pitching AI-native coding experiences such as Cursor or Replit. Anthropic, when evaluating what users were talking to its Claude chatbot about, found that computing & mathematical related conversations were 37% of conversations, while representing about 3.4% of the US job market. It was by far and away the top category in usage, with arts & media and education following at around 10% each.[2] Coding has been one of the most prominent use

Figure 6: Anthropic economic index—coding dominates usage



[Source: Anthropic, Enders Analysis]

[2] Anthropic, "The Anthropic Economic Index", 10 February 2025

cases for AI so far, with Claude likely attracting more coding-focused users due to its strength in this area, and lower public perception.

Google has claimed AI-generated code has passed 25% of the new code they produce internally. The promise with agents is to raise this figure even higher, with AI going beyond a copilot set-up where it might suggest the next line of code, which the user can accept or reject. Instead, a coding agent will when given a specific goal to accomplish work in the background to write, test and evaluate the code before submitting a more completed task.

Google's coding agent is named Jules, and they promise that it can save time by for instance working asynchronously to solve bugs within code. When given a task it can create a plan, ask clarifying questions, and then work to iteratively complete its plan, error-correcting at various points. Mark Zuckerberg has specifically cited AI coding as the key agent targeted for Meta's next Llama model—"I also expect that 2025 will be the year when it becomes possible to build an AI engineering agent that has coding and problem-solving abilities of around a good mid-level engineer." Coding agents represent a natural evolution of internal tools, and one where there is a clear target in being able to increase the levels of abstraction and complexity that the system can handle, to slowly move further up the org chart.

Anthropic has become the leader in this space, with its Claude 3.5 Sonnet model released last year having been consistently strongest on coding tasks, and recently followed up with a 3.7 Sonnet designed specifically towards real-world coding skills. This focus is natural for Anthropic, which has far smaller consumer surfaces with its Claude chatbot than Google or OpenAI, but which has built a strong business selling its models indirectly via for instance Amazon's Q-Developer system. Alongside its 3.7 model it also released Claude Code, an agentic coding tool that integrates directly into developers' workflows.

Success will depend on the ability to match leading capabilities and extensive distribution:

- **Anthropic** has become the leader in this space, with its Claude 3.5 Sonnet model released last year having been consistently strongest on coding tasks, and recently followed up with a 3.7 Sonnet designed specifically towards real-world coding skills. This focus is natural for Anthropic, which has far smaller consumer surfaces with its Claude chatbot than Google or OpenAI, but which has built a strong business selling its models indirectly via for instance Amazon's Q-Developer system. Alongside its 3.7 model it also released Claude Code, an agentic coding tool that integrates directly into developers' workflows
- **OpenAI** has increasingly potent direct distribution as ChatGPT continues to grow, as well as being able to leverage its relationship with Microsoft to be available on Azure, as well as integrated into products such as GitHub's "Project Padawan" AI coding agent. Its breakthrough in reasoning models leaves it currently in pole position on capabilities, scoring a record 71.7% on the most notable software engineering benchmark, though Google and Anthropic are not far behind
- **Google** has very capable, and cost effective, models but has struggled to get the developer take-up it is looking for with its Gemini series. With its releases oddly staggered, confusingly named and spread across a variety of products and experiments, it has been coming from behind. Its decision to offer its new flagship Gemini 2.0 models for free initially to developers is a likely signal of its need to gain market share. The hope will be that centralisation within its Project Jules will allow for a more powerful, and more coherent product

Key takeways

Coding is a key focus area, and likely to be ground zero for the most powerful agents, and levels of automation. It has the highest level of competition, with almost all frontier AI companies competing, as well as a number of third-party application developers building 'wrappers' that make models more useful.

Coding agents will raise software engineering productivity across the board, allowing for more ambitious digitisation schemes and driving down systems and development costs. This will include accelerating AI research & development productivity at leading AI companies.

Personal assistants: rekindling a long-running battle

The final major category of agents is various incarnations of personal assistants. This is a long-standing category that has largely failed to live up to its promises. Recent efforts from Apple Intelligence or Google Assistant/Gemini have not delivered a truly compelling offering. Agents are likely to be what rekindles this vision of the genuinely capable and useful personal assistant.

Google looks set to deliver a highly capable offering, conditional on it being able to bring the costs of more advanced capabilities down and balance cloud and on-device running. Amazon only really sees upside in revitalising an otherwise stagnant product category, and is leveraging its partnership with Anthropic to remain at the cutting edge. Apple is in the trickiest spot, with its current offerings disappointing, and being in a more defensive position, with far greater downside risk than upside potential.

Google

Google has the clearest roadmap in this space. Its Project Astra aims to create a universal AI assistant that is fully multimodal, capable of real-time conversation, can actively use tools and accomplish tasks as an agent, and has strong in-session memory with a type of persistent longer-term memory. The vision is that then this assistant can be deployed on people's phones and sitting within prototype glasses, similar to what Meta is pursuing with its Meta Ray-Bans. Google's prototype from what they have shared appears fairly compelling, and bolstered with ongoing advances in reasoning, and with the cost-reduction work to enable it to scale across Android devices, it seems a likely leader in bringing a true agent assistant to market.

Apple

Apple Intelligence has proved an underwhelming experience to date, with the initial tranche of features either gimmicks such as AI Genmoji, or flawed as in its notification summaries twisting news headlines and confusing users. That Apple felt the need to turn Apple Intelligence into an opt-out feature, and force it back on with new software updates, is never a promising sign.[3] We are still waiting for features promised at the Apple Intelligence reveal of Siri being able to access greater personal context, and take actions within apps. These are inherently more agent-like features, and model developments are making them increasingly possible. The question is how far Apple Intelligence's underwhelming performance so far has been down to promising too much too early, rather than weakness in Apple's underlying AI development expertise. On paper Apple has the capital, the software skills and powerful underlying silicon that should

[3] Mac Rumors, "Apple's latest updates re-enable Apple Intelligence on some devices", 11 February 2025

leave it ideally placed to dominate in this area. Getting AI right is essential for Apple, with this being one of the only ways the iPhone could be pushed off its comfortable perch.

Apple struggling could see it turn increasingly towards alternatives. It reportedly has been looking to integrate Google's Gemini as an alternative model option to its existing relationship with OpenAI. OpenAI already works within Apple Intelligence to handle more complicated queries, though in a somewhat arms-length form. Apple likely does not want to end up dependent on another company for core features of its handset, therefore integrating Google is more likely leverage for it to use in negotiating with OpenAI than a long-term strategy. Relying on its principal handset rival is riskier than on OpenAI, though neither is an especially favourable option. OpenAI has the capabilities across speech, visual context, and now reasoning and agents that it could bring to Apple if its internal development efforts fail. This would be ideal for OpenAI, allowing it to take on a more prominent role within a very large iOS install base, though clearly Apple will be wary of taking outsourcing too far.

Amazon

Amazon re-entered this space with the announcement of its Alexa+ in February, see Echo 2.0: AI gives Echo its raison d'être [2025-024]. Alexa+ functions as a more advanced personal assistant sitting with Amazon Echo devices, which aims to revitalise this product line which has had an odd position within Amazon's ecosystem. This improved Alexa is aimed to provide even greater Prime lock-in, and greater differentiation for Echo devices, allowing Amazon to capture more margin as a services gatekeeper.

Alexa+ uses an interesting architecture that seeks to combine the benefits of LLMs with existing Alexa capabilities and software. Users interact by voice with the device, which can switch between AI models depending on the requirements. For instance, initially it may use a small Amazon Nova model to give lower latency, then for a complex request switch to a larger Claude model that can do more sophisticated reasoning and use software tools. These models then plug into traditional software APIs where possible, ensuring more structured and reliable experiences, but can also use web-browsing capabilities where these APIs are not available. This allows Alexa to act as an agent, with the promise that it could take a request, navigate to a website and complete a task on that site.

Microsoft

Microsoft has been strongly focused on enterprise agents, setting up its Copilot as a general personal assistant, able to then be customised to specific tasks or workflows and act autonomously in accomplishing them. It first began to roll out this functionality in May 2024, such that the implementation was less what this report has termed agents, and more chatbots customised to be better adapted to certain tasks. This reflects how Microsoft is not at the frontier in terms of developing models (it can rely on OpenAI to take the financial hit there), but is at the frontier in terms of productising and scaling useful applications of these models.

With the somewhat cooler relationship that has been developing between Microsoft and OpenAI, we would expect Microsoft to be producing its own reasoning models, and continuing its strategy of onboarding other close to the frontier models from Deepseek or Meta. Then its product strategy is likely to be diffusing cheaper internal models as broadly as possible through Copilot, then perhaps putting out a more premium Copilot service that leverages OpenAI's much pricier models. While OpenAI and Anthropic are racing to build more expensive but more capable models, on the view that this capability will naturally transfer into lower friction enterprise adoption, and greater pricing power, Microsoft will be looking to head them off by providing a 'good enough' service to its existing gargantuan enterprise base.

Where Microsoft does have a consumer advantage is in gaming. It has launched a Copilot tailored for gaming, that can help out within games and for the more tedious processes of ensuring games download, update and function properly. The truly interesting application of agents is within games themselves, with agents potentially offering much more dynamic gameplay opportunities versus traditional human-coded non-player characters. Microsoft will be competing here with Epic, Nvidia and Qualcomm to bring solutions that simplify the process of integrating agents to games, as well as itself looking to bring novel ideas to market through its first-party titles, vying for consumer attention.

Key takeways

Agents offer another chance for a genuinely useful personal assistant to come into being, building on a string of previous failures. If they can live up to this promise it would mean growth in voice interfaces, and have potentially disruptive effects on the handset market.

Greater use of personal assistant agents also raises similar questions to web agents around disintermediation on the web, and the risk that consumers assign their agents the task of finding cheap subscription deals, putting pressure on ARPUs and churn.

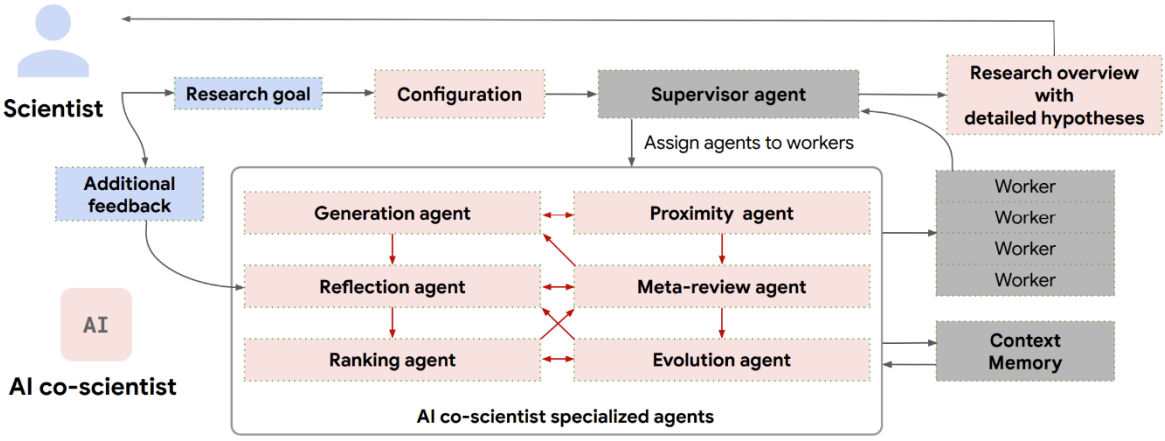
The end goal: an AI scientist

One of the most impressive agent formulations we have so far seen is Google's AI co-scientist. This is not a single agent but rather a multi-agent system that aims to function as a sophisticated scientific collaborator to help generate novel hypotheses and research proposals. The headline results are very impressive: over the span of just a few days the system was able to come to the same conclusion on an issue around gene transfer mechanisms and antimicrobial resistance as a team of researchers at Imperial who had been working on the issue for years. It was similarly able to identify novel treatment targets for liver fibrosis that showed high promise, an otherwise time consuming and difficult task for researchers.

The AI co-scientist system takes the strength of Google's new Gemini 2 models and places it within a very sophisticated multi-agent framework. This uses a number of specialised agents, and novel techniques, with the emphasis on the interactions between these agents, and asynchronous functions of this whole multi-agent system, rather than continual interaction with the user. First a "Generation agent" comes up with potential hypotheses, before handing them to a "Reflection agent" for a whole string of reviews and refinements, before they are then passed on to a "Ranking agent". This agent takes all the various hypotheses and places them into a tournament, with each continually going head-to-head against another and building an Elo ranking score (often used in games such as Chess). These match-ups take the form of simulated scientific debates, where simulacrum experts argue the merits of each hypothesis and decide upon the superior choice.

This whole continual cycle is overseen by a "Supervisor agent" that allocates resources and keeps things on track, eventually outputting back to the scientist a research overview that summarises all of the hypothesising and debates the various agents of the system have been conducting.

Figure 7: Google's AI co-scientist architecture



[Source: Google]

This kind of complex multi-agent system is a likely blueprint for the ways that agents will increasingly embed themselves within existing workflows and functions. The AI co-scientist framework is model agnostic, such that when Google brings out future upgrades it can plug smarter or cheaper agents in and see immediate gains flow through. Advances in models allow agents to increasingly compress much friction and complexity within themselves, but restrictions around the computer memory they require and their ability to access external tools and sources of information will make frameworks such as this highly useful. Usability will be improved where the framework itself disguises much of the complexity from the user, with agents supervising and orchestrating a whole string of other agents.

The eventual virtual coworker agent may appear as a singular slack profile, but in reality, be hundreds of different agents in a virtual trench coat, all aggregating together to produce much more impressive results than a single model would be able to accomplish. These kinds of multi-agent architectures will become more integrated and specialised over time with specific workflows. Enterprises will design some to reflect individualised business needs, while third-party developers will build more generalised apps, then for applications that are more general, consumer-facing, or compute intensive these architectures will sit in the background of products from the hyperscalers and leading AI labs.

While Google has explicitly developed this as a *co*-scientist system, the ultimate goal is more fully automated scientific research. Similarly, while Google has so far deployed this technology around research in biology and medicine, it is general-purpose by its nature. This is positive in enabling a very widespread accelerating effect on scientific research, but by being inherently dual-use it invites security questions. Hence Google's slow rollout of these advanced systems within trusted tester programmes. But as the capabilities here advance quickly, and especially as this advance is *itself* accelerated by systems boosting AI R&D, novel governance frameworks are essential.

About Enders Analysis

Enders Analysis is a research and advisory firm based in London. We specialise in media, entertainment, mobile and fixed telecoms, with a special focus on new technologies and media, covering all sides of the market, from consumers and leading companies to regulation. For more information go to www.endersanalysis.com or contact us at info@endersanalysis.com.

If your company is an Enders Analysis subscriber and you would like to receive our research directly to your inbox, let us know at www.endersanalysis.com/subscribers