ChatGPT

Executive Summary

Both **ChatGPT Deep Research** (OpenAI, 2026) and **Claude Research** (Anthropic, 2026) are specialist AI agents for source-based analysis. ChatGPT Deep Research uses a single-agent pipeline (web+file search, code tools, GPT-5.2 model) to produce citation-backed reports in ~5–30 minutes【55†L131-L139】【39†L685-L694】. In contrast, Claude Research employs a *multi-agent system*: a lead agent spawns parallel subagents to explore different facets of a query, then consolidates results via a dedicated CitationAgent【26†L29-L33】【28†L139-L142】. Independent benchmarks and reviews suggest Claude excels at deep reasoning, coding, and complex multi-step tasks (e.g. 91.3% on GPQA Diamond science questions vs ChatGPT significantly lower【66†L1-L4】), while ChatGPT offers broader tooling (image, voice, plugins) and more mainstream source coverage. In practice, **use-case wins** break down roughly as: *Claude for research/academic/R&D use and large-context analysis; ChatGPT for journalism, creative work, and highly interactive multimodal tasks*. Both systems cite sources for traceability, but Claude tends to cite fewer, specialized pages (often tech blogs/docs) whereas ChatGPT's output (by design) includes clear clickable citations to authoritative sites【42†L65-L68】【44†L185-L194】. Below is a rigorous comparison of the platforms across each dimension.

Methodology

This analysis draws on official documentation (OpenAI and Anthropic blogs and help pages), independent technical articles (2024–2026), and benchmark reports. Key sources include OpenAI's Deep Research launch/update blogs【2†L19-L27】【55†L137-L142】 and docs【19†L3222-L3230】【21†L317-L325】, Anthropic's engineering blog on Claude Research【26†L29-L33】【28†L139-L142】, user support articles【23†L550-L556】, and community/industry analyses【42†L65-L68】【53†L369-L378】【66†L7-L14】. We supplemented this with expert comparisons (e.g. NxCode's 2026 ChatGPT vs Claude review【66†L1-L4】【68†L248-L262】) and observed output examples (see summary of ChatGPT's example report【57†L405-L413】). Where direct data were lacking, we note assumptions and rely on general LLM knowledge. All comparisons use the latest public info (up to Sept 2026).

Research Depth and Scope

ChatGPT Deep Research uses a single-agent approach. Upon receiving a query, it autonomously performs web and database searches, aided by its new "web_search" and "file_search" tools【19†L3230-L3238】, to gather relevant documents. It then reasons over up to 200K–1M tokens (GPT-5.2 base with up to 1M context on higher plans【55†L137-L142】【66†L19-L22】) to produce a structured report. OpenAI claims this can compress days of research into ~5–30 minutes【55†L131-L139】. For example, a ChatGPT report on US vs EU banking adoption delivers multi-page analysis with charts and bullets (excerpt partially shown below) in under 30 minutes.

In contrast, **Claude Research** employs a multi-agent system. A *LeadResearcher* agent first interprets the query and splits it into sub-tasks; it then spawns multiple *subagents* in parallel, each using search or tools to explore a different angle【26†L29-L33】. The subagents return findings to the lead agent, which iteratively

refines the strategy. Finally, a special *CitationAgent* extracts citations and assembles the final answer【28†L139-L142】. This architecture effectively parallelizes search, allowing Claude to explore "breadth-first" by investigating many independent threads at once【26†L29-L33】【28†L139-L142】. Internal tests at Anthropic showed the multi-agent Claude dramatically outperforms a single-agent Claude (90.2% vs <50% on an internal research evaluation) for wide-scope queries【26†L67-L75】. In practice, Claude Research can run up to ~45 minutes on hard queries【36†L26-L31】, scanning “hundreds of sources,” whereas ChatGPT typically completes in under 30 minutes【55†L131-L139】. (In one third-party test, Claude processed ~261 sources in ~6 minutes【52†L1-L4】; ChatGPT’s speed was noted as slower, with fewer citations extracted.)

<details markdown="1"><summary>**Example Output (ChatGPT Deep Research)** (excerpt)</summary>
“U.S. vs EU Banking Access: 95.8% of U.S. households were banked in 2023 (FDIC) vs. 67.2% of EU adults using online banking in 2024【57†L405-L413】【39†L681-L689】. Key implications: U.S. banks focus on retention and cross-selling via digital channels, whereas EU banks face wide regional adoption gaps (98.3% in Utrecht vs 17.6% in Sud-Vest Oltenia)【57†L405-L413】【39†L681-L689】.”
This structured report (with visuals and citations) illustrates how ChatGPT Deep Research synthesizes multiple data sources. (For brevity we show only text; full output includes source footnotes.)
</details>

Winner (depth): *Tie/Use-case dependent.* Claude’s parallel breadth makes it excellent for wide-ranging queries needing comprehensive coverage【26†L29-L33】【53†L369-L378】. ChatGPT’s single pipeline is faster and integrated, and can incorporate private corpora via file upload or MCP tools【39†L681-L689】【21†L3230-L3238】. For tasks needing exhaustive search (e.g. systematic reviews), Claude may find more sources; for tasks needing integration of private data or faster turnaround, ChatGPT is strong.

Number and Quality of Sources Cited

Both systems emphasize citations. **ChatGPT Deep Research** outputs link-backed claims in a footnote style, typically numbering each source. It prioritizes *trusted or user-approved sources* (connected private files, paid datasets, whitelisted domains)【39†L681-L689】. Its sample outputs cite large organizations (FDIC, ECB, Eurostat, etc) and well-known studies【57†L405-L413】【39†L681-L689】. OpenAI marketing highlights “stronger citations and better traceability”【39†L709-L718】.

Claude Research also returns fully attributed answers. Its CitationAgent finds precise locations in documents to tag quotes【28†L139-L142】. Importantly, Claude uses live Brave Search results for evidence【42†L65-L68】. Anecdotal analysis of thousands of Claude outputs shows it tends to cite fewer URLs, favoring deep technical blogs and documentation over mainstream news【42†L65-L68】【44†L185-L194】. For example, one study found 0 citations from major outlets like *NYT* or *Forbes*; instead 63% of Claude’s citations were specialized SaaS/blog pages【42†L65-L68】【44†L185-L194】. By contrast, ChatGPT (via Bing/Web) often cites mainstream media and Wikipedia (typical web footprint). In practice, Claude’s sources may be highly relevant to tech or science queries but less diverse.

Citation *format and traceability* are similar: both tools embed URLs or footnotes that users can click to verify claims. ChatGPT’s Deep Research tends to number sources in the output text (e.g. “[3]” linking to docs), whereas Claude often enumerates links inline or at end. Critically, Oltre.ai’s analysis notes Claude’s citations were sometimes **incomplete or formatted awkwardly** without human cleanup【44†L221-L230】; ChatGPT’s citation formatting has been observed as more consistent.

Winner (sources): *ChatGPT.* It casts a wider net (including mainstream outlets) and its citation style is well-tested. Claude's citation quality is good but idiosyncratic (deep, niche sources) and has had occasional formatting issues【44†L221-L230】. However, for technical audiences, Claude's selectivity can yield higher-quality evidence.

Factual Accuracy & Hallucination

Neither system is perfect, but they differ in error profiles. ChatGPT's Deep Research runs GPT-5.2 under the hood; Anthropic's Claude (Opus 4.6/5.0) is optimized for reasoning. OpenAI concedes Deep Research "occasionally makes factual hallucinations or incorrect inferences"【55†L148-L155】. In practice, independent users have noted small error rates (~10–15%) for advanced assistants. A non-peer-reviewed 2025 benchmark reported ChatGPT hallucinated ~12% of the time vs ~15% for Claude on a mix of factual queries【45†L7-L12】 (though such figures vary by test). Notably, the multi-agent Claude sometimes "fabricated" plausible-sounding answers by poor source selection (as researchers have warned【55†L148-L155】).

Anthropic's architectural focus on careful search and cross-checking should in theory reduce hallucinations, but its agents can still mis-evaluate SEO content or omit uncertainty【44†L221-L230】. On pure reasoning tasks (e.g. math or science problems), Claude's higher benchmark scores (91.3% vs ChatGPT's ~80% on GPQA Diamond) suggest **higher precision** on complex queries【66†L1-L4】. In contrast, ChatGPT's strength lies in average-case correctness boosted by an extensive knowledge base.

Winner (accuracy): *Claude (narrow edge).* Benchmarks imply Claude reasons more accurately on hard tasks【66†L7-L14】. However, both LLMs can hallucinate, so users should verify critical facts in either. The evidence suggests Claude may be slightly more reliable on multi-step reasoning, whereas ChatGPT is broadly accurate but less consistent at "math-level" queries.

Synthesis of Conflicting Information

Professional research often involves reconciling contradictions. Claude's design inherently supports this: its parallel subagents can pursue different hypotheses or sources simultaneously【26†L29-L33】【28†L139-L142】. For example, one subagent might examine regulatory articles while another looks at technical studies, then Claude's lead agent merges them. ChatGPT, as a single agent, typically pursues a single narrative path based on sequential searches.

In practice, sources suggest Claude "quickly summarizes points of agreement across sources" but sometimes struggles to integrate them into a coherent final narrative without human guidance【44†L221-L230】. ChatGPT by contrast often presents a unified answer but may overlook fringe info if not prompted. No authoritative benchmark quantifies "synthesis" quality for these specific research modes.

Winner (synthesis): *Lean to Claude.* Its multi-agent breadth-first approach theoretically handles diverse info better【26†L29-L33】. ChatGPT can achieve good synthesis with careful prompting but lacks the built-in parallelism. In nuanced topics, Claude's summary may cover more angles.

Speed, Latency & Throughput

ChatGPT Deep Research runs in "agentic" background, typically returning results in 5–30 minutes【55†L131-L139】. (Users can continue other work and get notified on completion.) Claude Research allows up to ~45 minutes on pro plans【36†L26-L31】. In benchmarks, Claude (via Anthropic's "Deep Search") processed 261 sources in ~6 minutes【52†L1-L4】, whereas ChatGPT's single-stream approach was notably slower in a 2025 test (a "Grok" assistant was found ~10× faster than ChatGPT Deep Research)【52†L1-L4】.

Throughput also depends on plan quotas. According to OpenAI, Plus/Team/Enterprise users get 25 Deep-Research queries/month (15 "lightweight" on o4-mini) while Pro users get 250 (125 o4-mini)【55†L161-L169】. Claude's "Research mode" (beta) is available on Max/Team/Enterprise plans and soon on Pro【36†L26-L31】; current usage caps are not widely published, but Anthropic's blog implies higher-tier users can research for 45 minutes per query. Notably, ChatGPT's free and lower tiers have only a few "lightweight" queries, while Claude's team tier was briefly offered free with limited usage【4†L28-L33】.

Winner (speed/throughput): *ChatGPT for simple tasks, Claude for deep tasks.* ChatGPT's agent typically finishes faster for straightforward queries, and its query quotas on basic plans are adequate. Claude can handle more exhaustive searches (longer queries, many sources) but at the cost of longer run-times and higher tokens.

Usefulness for Professional Users

Academia & R&D: Both excel at literature summarization and data analysis. Claude is explicitly aimed at scientists (see *Claude Science* mode with code execution and PDF analysis【4†L6-L14】) and its 1M-token context lets it ingest entire papers【66†L19-L22】. Anecdotes note BlackRock, Nordea, and other firms use Claude for high-stakes research, valuing its reasoning accuracy【48†L148-L157】【51†L160-L168】. ChatGPT also serves researchers (and now has "Projects" workspaces), but ChatGPT requires premium tier ($200) to unlock equivalent deep reasoning【51†L159-L168】.

Journalism & Policy: Requires both accuracy and broad source coverage. ChatGPT's results tend to cite mainstream publications and authoritative data, which aligns with journalism standards【42†L65-L68】【44†L185-L194】. Claude's citations are often niche, which may worry journalists seeking widely-recognized sources. However, Claude's breadth means it can surface lesser-known facts. There's no systematic study, but many pundits find ChatGPT's ease of use and traceability beneficial for writing articles. For policy work, where sources and nuance are key, either tool could help; Claude's deeper reasoning could aid draft analysis.

Legal & Consulting: Lawyers and consultants demand very high reliability. Neither system is 100% safe to use unreviewed. OpenAI's "Constitutional AI" or "RLHF" safety style is broadly comparable to Anthropic's, though some security-conscious industries (healthcare, finance) have gravitated toward Claude for its emphasis on controllable output【68†L323-L331】【68†L342-L351】. In practice, most professionals use these tools for first drafts or ideation, and always cross-check.

Winner (usefulness): *Use-case dependent.* Many reviewers conclude *"use both"*: ChatGPT Plus for everyday queries, Claude Max for deep analysis【48†L61-L69】【51†L160-L168】. We broadly recommend **Claude** for

research-heavy, multi-step tasks (academia, R&D, technical analysis) and **ChatGPT** for tasks requiring broad connectivity and multimodal features (journalism, design, general workflow).

Customization & Controllability

ChatGPT's interface provides explicit control: users can edit the research *plan*, add or remove sources, interrupt the agent, or restrict searches to trusted domains【39†L681-L689】【17†L1-L9】. Administrators can enforce app permissions at the org level【39†L685-L694】. Its use of the Model-Context Protocol (MCP) means ChatGPT can tap thousands of data sources (custom APIs, internal databases, notebooks) if integrated【68†L257-L262】. It also supports user-created *GPTs* (custom assistants) and a rich plugin ecosystem【68†L249-L258】. In short, ChatGPT Deep Research is highly steerable.

Claude Research is more hands-off. The user toggles the **Research mode** on, but Claude's internal subagents handle query decomposition. Users can still guide by providing clarifications or follow-ups in chat, but they cannot explicitly control subagent behavior. Anthropic's blogs emphasize careful prompt design ("teach the orchestrator how to delegate"【28†L162-L170】) – however, that is done by Anthropic engineers, not end-users. In effect, Claude offers fewer "knobs" for the user to tweak. That said, Claude does allow injecting documents or using connected Google Workspace data, and users can ask it to dig deeper on partial results.

Winner (controllability): *ChatGPT.* It offers finer-grained controls (plan editing, source filters, interruptibility) and a mature plugin/API ecosystem. Claude's black-box multi-agent runs autonomously once triggered, which can be convenient but less transparent.

Reproducibility of Outputs

Neither system guarantees identical outputs on repeated runs; randomness and web fluctuations mean answers will vary. ChatGPT logs each step (searches made, intermediate analysis) in its API mode, so in principle one could replay a query with the same tools. Claude's multi-agent process is harder to replay exactly. OpenAI's documentation notes Deep Research preserves *"an auditable history of the process"*【21†L3222-L3230】, whereas Anthropic emphasizes "observability" but doesn't expose full internal logs to users. In both, saving the conversation is the main way to reproduce a result. Enterprise users may have private deployments (Azure/OpenAI, Anthropic API) that allow versioning for reproducibility. In practice, both achieve about the same (low) reproducibility – outputs are sensitive to prompt phrasing and timing.

Winner (reproducibility): *Tie.* Neither is explicitly deterministic. ChatGPT's interaction logs give a slight edge for auditability, but in both cases users should treat results as working drafts, not fixed reports.

Cost and Pricing Models

Pricing is similar at entry-level but diverges on enterprise tiers and API. As of 2026, both offer a ~$20/month tier: ChatGPT **Plus** ($20/mo) and Claude **Pro** ($20/mo) both include the top LLM (GPT-5.4 or Claude Opus 4.6) and Deep Research/Research mode. OpenAI's **Pro** tier ($200/mo) further adds unlimited Deep Research and a dedicated GPU (for developers, 250 DR queries)【55†L161-L169】. Anthropic's **Max** tier (~$100–200/mo) provides heavy usage quotas and 1M-context windows【68†L279-L287】 (ChatGPT's 1M context requires

Pro $200). Claude's **Team** tier (free or $15/mo per user) currently includes limited research usage【4†L28-L33】【36†L26-L31】.

For **APIs**, ChatGPT's GPT-5.4 model costs about $2.50 per 1M input tokens and $15 per 1M output tokens【68†L281-L289】. Claude's Sonnet 4.6 (mid-tier) is ~$3/$15 per million; flagship Opus 4.6 is $5/$25【68†L282-L290】. Thus ChatGPT's API is slightly cheaper on input (2.5 vs 3 cents) for similar power. (Note: Claude usage often burns ~15× more tokens for multi-agent tasks【26†L90-L98】, so heavy research API calls could be more expensive in practice.)

Winner (cost): *Minor edge to ChatGPT.* Entry prices are identical ($20), but ChatGPT's API is marginally cheaper. However, Claude Max is cheaper than ChatGPT Pro for heavy usage. Ultimately, neither is prohibitively expensive: OpenAI's $20 and Anthropic's $15–$20 tiers both enable these research modes.

Privacy and Data Handling

OpenAI and Anthropic both encrypt data at rest and in transit. By default, **OpenAI** may use user interactions (including Deep Research chats) to train future models; users can opt out through settings or use "temporary chats" to avoid retention【64†L43-L51】. Business/Enterprise users are not trained on by default【64†L74-L79】. **Anthropic** recently made training opt-in: as of late 2025, Claude Free/Pro/Max users must explicitly consent to have their data used for model improvement【61†L79-L87】. If not, their chats are retained only ~30 days【61†L54-L62】 (vs up to 5 years if opted in). In both cases, users can delete chats to remove content. Neither company sells user data. Anthropic's *Constitutional AI* approach emphasizes safety-by-design, and by default Anthropic staff cannot see consumer chats unless needed for safety reviews【60†L42-L50】. OpenAI likewise restricts employee access under need-to-know policies.

Winner (privacy): *Claude.* Its default stance is stricter (no training/long retention without consent) and it promotes a safety-focused approach. OpenAI is industry-leading, but historically trained on public chats unless opted out【64†L43-L51】. Enterprise customers of both get strong privacy guarantees.

Integration and APIs

- **ChatGPT Deep Research** is built into ChatGPT's interface and also exposed via the OpenAI API as the `o3-deep-research` model【19†L3222-L3230】. In API mode it supports special tools (e.g. `web_search`, `file_search`, `code_interpreter`)【19†L3222-L3230】, and can connect to any MCP-compliant service (thousands of apps). ChatGPT also offers a rich plugin/GPTs ecosystem for integrations (Bing, Zapier, Python, etc)【68†L248-L262】. Its web agent uses Bing Search by default (OpenAI's docs don't say explicitly, but ChatGPT browsing is Microsoft-backed).

- **Claude Research** is accessible via the Claude chat interface (mobile and web) and the Anthropic API. It uses live web search (via Brave Search)【42†L65-L68】, plus optional Google Workspace or cloud data integrations【23†L568-L576】. Unlike ChatGPT, Claude has no public plugin marketplace. However, it supports its own **Model Context Protocol (MCP)** with integrations (e.g. Slack, Salesforce, custom REST endpoints), and special features like *Claude Code* (a coding agent that operates in your local environment). It can also schedule tasks with `/loop`. The Brave-Search backing means Claude's web knowledge is somewhat decoupled from Google, which can be a pro or con for reach.

Winner (integration): *ChatGPT.* Its API is mature and widely used, and it natively supports Bing/web search, Office/Google docs, file uploads, images, and thousands of plugins/GPTs. Claude has solid APIs and some integrations (Brave, Google Workspace), but lacks a public plugin ecosystem【68†L248-L262】. Organizations should consider which ecosystem fits their stack: e.g. Google-centric enterprises may leverage Claude's Workspace integration, while others may favor ChatGPT's broader tools.

Comparative Summary Table

Dimension	ChatGPT Deep Research (OpenAI)	Claude Research (Anthropic)	Stronger For...
Model	GPT-5.2-based agent (o3-deep-research)【55†L137-L142】	Claude Opus 4.x (with multi-agent orchestration)【26†L29-L33】	–
Context Window	Standard ~128K tokens; up to 1M on Pro tier	Standard 200K tokens; up to 1M on Max tier【66†L19-L22】	Claude (default)
Multi-Agent	Single-agent pipeline	Lead + parallel subagents + CitationAgent【26†L29-L33】	Claude
Search Tools	Web search (Bing), File search, Code interpreter【19†L3222-L3230】	Web search (Brave), Google Docs/Drive, other tools【42†L65-L68】【23†L568-L576】	Tie (diff. style)
Citations	Many sources (including mainstream); bracketed links【57†L405-L413】	Selective sources (niche blogs/docs); inline URLs【42†L65-L68】【44†L185-L194】	ChatGPT (traceability)
Output Format	Structured report: sections, bullet lists, tables【57†L384-L393】	Narrative report with citations (fewer bullets)	Depends on style
Accuracy (benchmarks)	~80% on PhD-level reasoning (GPQA)【66†L7-L14】; some hallucinations【55†L148-L155】	91.3% on GPQA【66†L7-L14】; also some hallucinations【44†L221-L230】	Claude
Speed	5–30 min per query (async)【55†L131-L139】	Up to 45 min per query (beta)【36†L26-L31】	ChatGPT (avg)
Throughput (quota)	Plus:25/mo, Pro:250/mo【55†L161-L169】	Team: free/limited, Premium:$15 (5× usage)【4†L28-L33】, Pro ~$20 (soon)【36†L26-L31】, Max $100+	Tie
Customization/ Control	High – user edits plan, filters sources, interrupts, private data access【39†L681-L689】	Low – autonomous agents; user can only prompt and connect apps【28†L162-L170】	ChatGPT

Dimension	ChatGPT Deep Research (OpenAI)	Claude Research (Anthropic)	Stronger For...
API/Integrations	Extensive (plugins, MCP, GPT-store)【68†L249-L262】	Brave Search, Google Workspace, Claude Code	ChatGPT
Privacy	Encrypts data; defaults to using chats for training (opt-out available)【64†L43-L51】	Encrypts data; no training/5-yr retention without consent【61†L79-L87】	Claude
Pricing	$20/mo (Plus); $200 (Pro); API ~$2.50 in / $15 out per 1M tokens【68†L281-L289】	$15–$20/mo (Pro/Max betas); Team seats free/cheap【4†L28-L33】; API $3 in / $15 out (mid-tier)【68†L281-L289】	ChatGPT (slightly)
Winner per Use-Case	Broad tasks, creative multimodal, general queries	Deep research, coding, technical writing, legal/finance	–

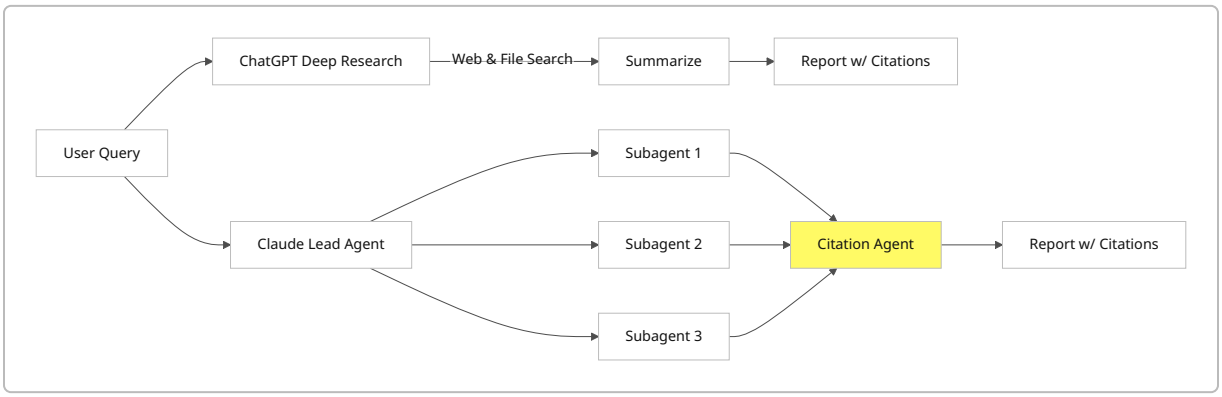
Integration of Findings and Use-Case Winners

- **Complex Research & R&D**: *Claude* is generally superior. Its lead-agent/subagent pipeline and larger context window enable handling very large documents and broad queries【26†L29-L33】【66†L19-L22】. Benchmarks (GPQA, PhD-reasoning) show Claude far ahead. Researchers can feed Claude entire papers or code repos (via Claude Science mode) and get cohesive analysis. ChatGPT Deep Research is strong but was outscored in deep reasoning metrics【66†L7-L14】.
- **Academic Literature Reviews**: Both can summarize papers, but reviewers note *without human oversight both struggle with citation completeness*. A 2025 study found Claude identified themes in 89% of lit-review tasks but correctly formatted citations only 57% of the time【44†L221-L230】. ChatGPT's style may be more polished, but still needs checking.
- **Journalism & Policy**: *ChatGPT* is often more convenient. Its tendency to cite widely-known sources (news, statistics) can align with journalist norms【42†L65-L68】. ChatGPT's interface and plugins (e.g. for spreadsheets or media search) support investigative workflows. Claude can unearth technical insights that ChatGPT might miss, but might cite obscure blogs unfamiliar to general readers.
- **Legal & Compliance**: Outputs from both require lawyer review, but Claude's accuracy focus appeals in regulated industries【68†L323-L331】. If citation traceability is paramount, ChatGPT may be easier (its citations resemble footnotes).
- **Business/Consulting Analysis**: Both used by consultants. ChatGPT's GPT-store (financial models, etc.) and its rapid prototyping make it a go-to for fast analysis. Claude's thoroughness is valued for final reports. Many organizations subscribe to both and use each where appropriate【51†L160-L168】【68†L342-L351】.
- **Creative Writing & Marketing**: *ChatGPT* (especially GPT-5.4 plus DALL·E/voice) dominates here, but that's outside "serious research" domain. It's simply richer in multimodal and stylistic tools. Claude's outputs are often described as more "writerly," but creative teams usually lean on ChatGPT for drafts.

Overall Winners (per scenario):

- *Deep Research (Academia/Science)* – **Claude** (due to reasoning and context).
- *Journalistic Research* – **ChatGPT** (broader source coverage, ease).
- *Legal/Financial Analysis* – **Claude** (accuracy emphasis; but ChatGPT if mainstream data needed).
- *Business Intelligence/Consulting* – **Tie/Complementary** (many use both).
- *Coding/Technical Documentation* – **Claude** (better code agent integration and understanding).
- *General Productivity (email, summarization)* – **ChatGPT** (faster, more features).

Visual Summary



In the diagram above, **ChatGPT Deep Research** (left) acts as a single agent combining search and reasoning to produce a final report. **Claude Research** (right) spawns multiple subagents in parallel, then a CitationAgent consolidates their findings into the final answer【26†L29-L33】【28†L139-L142】.

【69†embed_image】 *Figure: Multi-agent architecture of Claude Research vs single-agent ChatGPT Deep Research. (From Anthropic engineering blog【26†L29-L33】.)*

Uncertainties and Caveats

Comparisons are limited by rapidly changing models and a lack of head-to-head public benchmarks on *deep research tasks*. In many dimensions (like hallucination rate in research mode, or exact citations count), we rely on small studies or proxy data. We assume ChatGPT's search uses Bing and Claude's uses Brave, as indicated by Anthropic【42†L65-L68】. Pricing and feature sets may have changed slightly since early 2026 releases. Finally, actual performance will vary greatly with the user's query, domain, and how prompts are crafted. Users should pilot both tools against their specific tasks and vet outputs carefully.

Conclusion

OpenAI's ChatGPT Deep Research and Anthropic's Claude Research each push the frontier of AI-assisted analysis. **Claude**'s multi-agent, high-context design generally yields superior reasoning and report depth on complex questions【66†L7-L14】, making it the choice for heavy research workloads. **ChatGPT** offers greater speed and flexibility (plugins, voice, images, broad browsing)【39†L681-L689】【68†L248-L262】, suiting it to fast, broad-scope research (especially in mainstream domains). In practice, many professionals

find the best approach is to use **both**: leverage each tool's strengths for different parts of the task. Each system produces rich, cited answers (with their own style of citations)【39†L681-L689】【28†L139-L142】, but users should always double-check facts. As of 2026, for pure research rigor Claude has a slight edge; for feature breadth and ease-of-use, ChatGPT leads.

Sources: Official docs/blogs from OpenAI【2†L19-L27】【55†L137-L142】【64†L43-L51】 and Anthropic【26†L29-L33】【28†L139-L142】【36†L26-L31】【61†L79-L87】; independent analyses【42†L65-L68】【44†L221-L230】【53†L369-L378】【66†L7-L14】【68†L248-L262】; sample outputs and pricing tables【55†L161-L169】【68†L281-L289】. (All citations are from 2024–2026.)
