Early access: Evaluating GPT-5.5 for agentic legal work

Written by Angel Faus



We have conducted extensive early access evaluations of GPT-5.5 within Vincent.

GPT-5.5 is OpenAI's latest frontier model, with improvements in reasoning, context utilization, and output consistency across complex tasks. We have been testing it within Vincent as part of our ongoing early access evaluations of models for legal workflows.

Evaluation methodology

Our early access evaluations show GPT-5.5 delivering measurable gains across both research and document-intensive tasks. These improvements are most visible in workflows that require sustained reasoning over multiple sources, including authority-grounded legal research, contract analysis, and multi-document review.

Clio's evaluation suite, authored by legal experts, covers the legal work our customers actually do every day: legal research, document analysis, drafting, discovery, and scenario-based advisory. On that proprietary metric, Vincent with GPT-5.5 achieved an overall score

of 87.2% when run inside Vincent's full system, the highest score we have recorded to date across all models evaluated under the same conditions.

When powered by GPT-5.5, Vincent delivered the top overall benchmark score we recorded at 87.2%, higher than any other frontier model we tested.

Performance on complex legal tasks

Two categories of work push frontier models hardest in our evaluation, and they are where GPT-5.5 moves furthest from the prior generation.

The first is legal research that requires citing the controlling authority, including the specific case, the exact statutory section, and the leading commentary, rather than merely describing the rule. On these tasks, GPT-5.5 delivers a roughly 20% relative improvement over the prior generation, closing gaps that earlier systems consistently left open.

The second is difficult, open-ended document work, which includes contract analysis, deal-point extraction, multi-document review, and discovery across large file sets. Earlier models would reliably surface the right answer but could miss the qualifying language, scope clauses, and secondary requirements that might alter its legal meaning. GPT-5.5 reads further into the document and captures key information: the survival periods, the fraud carve-outs, the jurisdictional conditions, the conditions of exercise. Across our document-analysis scenarios, this translates to a ~7% relative improvement over the prior generation, and the difference is even larger at higher reasoning effort. The result is an answer that is not merely directionally correct but more legally complete.

Context utilization

GPT-5.5 is markedly more efficient in how it uses tokens during reasoning. It spends fewer tokens deliberating internally for the same quality of answer than other frontier models we tested. In one comparison, it used ten times fewer reasoning tokens per tool call. In practice, this means two concrete things for our customers: faster responses, and more headroom in the context window for Vincent to retain context in long, multi-turn sessions and autonomous agent work.

System-level performance

These results reflect system-level performance. All evaluations are run within Vincent, where models operate with orchestration, retrieval, and access to structured legal data, including Clio Library. Outputs are generated against retrieved context, with citations tied to sources accessed during execution and document analysis performed on full inputs. Performance at this level depends on how effectively the model retrieves, integrates, and acts on available context.

Ongoing evaluation

As with prior models, some behaviors remain under active evaluation. We continue to observe cases where responses are more detailed than required for the task, particularly in simpler queries. This is an area we are monitoring as part of ongoing testing.

GPT-5.5 is currently in early access within our evaluation pipeline. We will share more details as it becomes available within Vincent.

About Clio

Clio is the global leader in legal AI, powering legal work across the profession, from law firms of every size to corporate legal teams, government agencies, and courts. Its Intelligent Legal Work Platform unites the business and practice of law, connecting the people, information, and processes behind legal work. Clio's AI brings the context together with legal sources drawn from Clio Library, a growing collection of more than one billion legal documents. Over 400,000 legal professionals in more than 130 countries rely on Clio to run their legal world. Learn more at www.clio.com.