

# Generative AI in Tax Practice: Trusted Colleague or Compliance Risk?

By Marcus McCarty

Heart of America Tax Institute 2025

## Introduction

In June 2023, a federal judge in Manhattan sanctioned two attorneys for filing a brief that cited half a dozen judicial decisions which simply did not exist.<sup>1</sup> This wasn't a case of a fraudulent or intentional misrepresentation, instead it was an example of misplaced reliance on a large language model (LLM) – and early version of ChatGPT, to generate legal citations of relevant legal authority. That case, *Mata v. Avianca, Inc.*, has become a cautionary tale, signaling to the professionals that generative AI tools—while powerful—carry unique risks unlike those associated with traditional research platforms. Since then, courts across the country,<sup>2</sup> including the U.S. Tax Court,<sup>3</sup> have confronted similar problems in pleading, striking filings that relied on AI-generated hallucinations and in many cases disciplining the attorneys responsible. One source listed over 300 cases (as of October 2025) in Courts throughout the United States, including several from Kansas and Missouri that involve AI-generated hallucinations.<sup>4</sup>

The rise of generative artificial intelligence – and more specifically large language models or “LLMs,” present extraordinary opportunities and serious challenges for legal and tax practitioners. Exploiting the opportunities they create, while at the same time successfully navigating the challenges, requires a basic understanding of how these tools work and their potential risks. Yet given the continued example of hallucinations in court filings (and doubtless many other examples in client advice and communications) many practitioners still struggle to appreciate the risks and limitations. That may be due at least in part on the difficulty in finding reliable “nontechnical” relevant information written for lawyers and tax professionals explaining why this is a problem.

This article is intended to help. It is written for legal and tax professionals, by a lawyer who began practicing before email, cell phones, or desktop computers existed in most firms. The article's thesis is relatively straightforward:

---

1 *Mata v. Avianca, Inc.*, No. 22-cv-01461 (S.D.N.Y. 2023), <https://law.justia.com/cases/federal/district-courts/new-york/nysdce/1:2022cv01461/575368/54/>.

2 Damien Charlotin, AI Hallucination Cases Database, <https://www.damiencharlotin.com/hallucinations/> (last visited Oct. 28, 2025).

3 *Thomas v. Commissioner*, T.C. Dkt. No. 10795-22 (Order Oct. 23, 2024); *Westlake Housing, L.P. v. Commissioner*, T.C. Dkt. No. 478-24L (Order Jan. 13, 2025) See also <https://www.jwells.tax/p/thursday-thoughts-how-not-to-use>

4 Damien, AI Hallucination Cases Database, *Supra*.

By understanding how these new tools work, their strengths and their weaknesses, lawyers and tax professionals can better identify and address the ethical constraints that must govern their use in their practice, and at the same time identify ways these tools can make them a better and more efficient.

The article starts by laying the groundwork needed to understand how this new technology works and how it continues to be augmented and improved. From this base, the practical and ethical risks associated with the use of this technology will be identified along with suggestions professionals can use to mitigate these risks.

Of course, some argue that all this is unnecessary, that the best course of action, at least for now is to avoid using generative AI and LLMs in a tax practice entirely. At the risk of dating myself, I remember hearing that same advice being given with respect to the use of e-mail, cell phones, computer-based research, texts, and desk top computers. My strong opinion is that professionals that shun generative AI technologies will only steepen the learning curve they eventually will encounter to become proficient in their use. In fact, I believe failing adopt the technology may mark the beginning of the end of their professional career in some instances. That is a strong statement, but it is backed by over 40 years of antidotal real-world experience.

In addition, it is questionable whether avoiding AI is even a realistic option. In just the past two years, generative AI has become imbedded in almost every internet-based tool used, and as the next section illustrates, that trend likely will accelerate in the coming months and years.

Regardless, my hope is that “technology skeptics” will maintain an open and curious mind, if for no other reason than to appreciate the work that a few visionaries have already brought us! While I share many of the concerns regarding the misuse of generative AI such as LLMs, I’m optimistic that on balance the advantages of the technology will far outweigh their potential harm.

## Part I

### A. What Are These Tools?

To understand artificial intelligence or any new technology, the first step is to define it. One way to do this is to start with broadest terms and proceed to narrow the focus, highlighting terms that are most relevant to lawyers and tax professionals, and bypassing others that while relevant in other industries and professions, are less applicable to a lawyer or tax professional.

The term artificial intelligence has been defined: “as the tangible real-world capability of non-human machines or artificial entities to perform, task solve, communicate, interact,

and act logically as it occurs with biological humans.”<sup>5</sup> A more understandable definition might be “...technology that enables computers and machines to simulate human learning, comprehension, problem solving, decision making, creativity and autonomy.”<sup>6</sup>

Either definition covers a huge waterfront of very familiar computer products some of which have been around for many decades.<sup>7</sup> Generative Artificial Intelligence (generative AI) is a special subcategory of artificial intelligence: one that not only can simulate human capabilities (such as learning how to play chess and improving over time), but also can generate new content that simulates human creativity. These programs are much more recent, and in addition to written or spoken text can produce music, art, photography and video.

The one category of generative AI technology most relevant here is the large language model or (LLM). Like other forms of generative AI, LLMs also can create new content, but instead of music, art or photography, LLMs usually are used in tax practices to produce written content. Because that written content is newly created, these programs interact seamlessly with humans and can mimic human conversation or text communications.

Most recently LLMs (and other forms of generative AI) have been modified and adapted to better serve the needs of consumers. This is often referred to as “retrieval augmented generation” (RAG) and it often is implemented using tools such “application programming interfaces” (API), and artificial intelligence “agents” (AI Agents). API and AI agents allow an LLM to interact with other software and tools.

## B. How Do LLMs Work?

Like artificial intelligence, the path to the development LLMs, can be traced several decades. One early example is a program called Eliza, developed in 1966 in an MIT-based by Artificial Intelligence Laboratory, by Joseph Weizenbaum.<sup>8</sup> The idea behind the Eliza program was to prompt a computer to recognize certain key preloaded words and phrases as they were typed by a human user, and then to respond with one of several preloaded open-ended responses.

---

5 Gil de Zúñiga, H., Goyanes, M., & Durotoye, T. (2023). A Scholarly Definition of Artificial Intelligence (AI): Advancing AI as a Conceptual Framework in Communication Research. *Political Communication*, 41(2), 317–334. <https://doi.org/10.1080/10584609.2023.2290497>

6 Stryker, C and Kavlakoglu, E “What is AI” [https://www.ibm.com/think/topics/artificial-intelligence?utm\\_source=chatgpt.com](https://www.ibm.com/think/topics/artificial-intelligence?utm_source=chatgpt.com).

7 The History of AI: A Timeline of Artificial Intelligence, Coursera (Oct. 14, 2025), <https://www.coursera.org/articles/history-of-ai> (expert systems, rule-based reasoning, natural language processing, machine learning and neural networks have resulted in chat-boxes, applications for medical diagnosis, speech recognition etc.).

8 For an entertaining video showing Eliza at work see YouTube, <https://www.youtube.com/watch?v=8jGpkdPO-1> (last visited October 28, 2025).

For example, if the human user typed “I don’t get along well with my mother” the computer might recognize the word “mother” in its preloaded language database and respond with a programmed open-ended question in response such as “Tell me more about your family.” Even though processing power and machine memory was quite limited, Eliza could appear to carry on a conversation, at least for short time.

However, the development of LLMs capable of achieving true human like conversational responses required several other developments in both hardware and software that have occurred within the past 5-10 years. One important advance was the development of a program architecture capable of developing a probability-based relationship model between how words or appear in given context.

For example, the phrase “I went to the bank and ...” might be finish with the words “went fishing” or “deposited my paycheck.” The one that is most correct depends on the context (the sentences and paragraphs that came before and after the phrase “I went to the bank and”). Of course, my example offered only two choices and there are likely an almost infinite number of alternatives that might be more correct depending on the context -- everything from: “contemplated the meaning of life;” to “had a conversation with Polly.”

How would one determine the most likely correct response, the next word or phrase? One way would be to examine a large quantity of textual material that included the phrase “I went to the bank and” and then assign probabilities of the next words – for example a fishing trip or visit to financial institution, based on what was in that data (the context in which the phrase appeared). While the mathematics needed to handle all of these relative probability relationships was developed early in the 20<sup>th</sup> century the process for applying that math to develop an LLM is often credited to a paper first published in 2017.<sup>9</sup> The idea behind this programming is to allow a single program (a “model”) to develop a set of probability relationships between words used in different contexts based on its exposure to a vast amount of written material.

While this explains the ideas behind the construction of an LLM, it doesn’t explain the contribution that the development of more powerful hardware played. Logically, the more data (books, articles, web pages, etc.) a model can use to establish and refine probable relationships between words or phrases, the more likely it will produce a response that corresponds to what was written in text having a similar context selected from the words and phrases in the documents on which it was trained; text that presumably is logical and reflects the underlying intelligence of the individuals that authored it.

How much data is needed? Well, LLMs today have likely been exposed and trained on a sizable percentage of all the data currently available on the internet, and that of course is

---

9 Vaswani et al. “Attention Is All You Need” [https://arxiv.org/abs/1706.03762?utm\\_source=chatgpt.com](https://arxiv.org/abs/1706.03762?utm_source=chatgpt.com) (2017).

a large percentage of all human knowledge. This process of exposing a model to data so that it can establish these probability relationships is called “training” and it requires enormous amounts of data and computer processing power. As luck would have it for company named Nvidia, this training works best on computer processors originally developed for computer video games. They were first adapted for use in training LLMs around 2017.<sup>10</sup>

All this led eventually led to OpenAI’s public release of its LLM ChatGPT 3.5 in late 2022.<sup>11</sup> I decided to “interview” for a blog I wrote in early 2023.<sup>12</sup> I remember being impressed by the model’s ability to carry on a conversation, while at the same time troubled by several shortcomings.

Most important among these, I thought, was the model’s inability to tell me where and how it got the information it provided and the admission that its “training” had ended several months prior and therefore did not reflect the most current information. What I didn’t appreciate, and missed in that blog, was the fact that LLMs at their core, are just super-sophisticated probability machines. They digest whatever words and phrases the user types or uploads and then runs it through a probability “model” that has been built based on billions of words of previously published text. It assigns probability relationships based on this training and uses it to choose the most likely next series of words and phrases it should produce in a response. The response sounds like a coherent “human” response, because the program is basing its probability calculations on the works of hundreds of millions of human authors.

What I also didn’t fully appreciate at the time was that this characteristic, one that is at the heart of how all LLMs work, is the primary reason LLMs can make truly bonehead mistakes or just “make stuff up.”

### C. How Do LLMs Differ from “Traditional” Computer Assisted Research Tools?

It’s worth considering how LLMs differ from traditional computer assisted research tools that have been available for half a century. I remember my excitement as a law student when I learned that I might get access to a brand new computerized legal research tool called “Lexis.”<sup>13</sup> I was surprised how quickly it could pull up hundreds of cases by simply typing in a key word or phrase. Lexis was followed a few years later by computer research software for Westlaw and many other legal texts and loose leaf services, and these

---

10Tweaktown, NVIDIA Tesla V100 Tested: Near Unbelievable GPU Power, <https://www.tweaktown.com/news/59155/nvidia-tesla-v100-tested-near-unbelievable-gpu-power/index.html> (last visited Oct. 28, 2025).

11 Wikipedia, ChatGPT, <https://en.wikipedia.org/wiki/ChatGPT> (as of Oct. 28, 2025, 17:01 GMT)

12 McCarty, M “My Interview with ChatGPT-4, MoBroadband.org, <https://mobroadband.org/my-interview-with-chatgpt-4/> (last visited Oct. 28, 2025).

13 See History of Legal Tech: LexisNexis Spotlight, LinkSquares Blog, <https://blog.linksquares.com/history-of-legal-tech-lexisnexis> (last visited Oct. 28, 2025)

supplemented traditional word searches with topic and key word searches.<sup>14</sup> Of course, these tools have vastly improved over the years, and for the most part have replaced traditional books and looseleaf services for tax and legal research.

However, even though we access LLMs and these services the same way, it's important to emphasize that they operate in fundamentally different ways. Search engines do not create new content, and they don't "make stuff up." Instead, they search an extensive preset database for words and/or terms that best match the user's input and accurately report that content. While it's certainly possible that a user's search terms might not catch every relevant statute, case, or administrative ruling, those that are retrieved do exist.

Lexis, Westlaw, and Bloomberg Law, etc. are also comparable to "traditional" internet search engines like Google or Bing.<sup>15</sup> In both instances they retrieve existing, authoritative content, and do not generate new material. However, Google and Bing search the open webpages from the internet and rank results based on popularity or relevance,<sup>16</sup> while legal research systems operate within curated databases of statutes, cases, regulations, and legal commentary. In this respect their search tools emphasize accuracy, structure, and citation linkage, enabling users to locate the precise legal text or interpretation they need. Though each system uses its own proprietary indexing and query logic, all three share the same fundamental purpose: to deliver verifiable legal sources, not to synthesize or create new content.

#### D. RAGs, APIs and Agents the Future of Generative AI

Even though its barely three years from the public introduction of ChatGPT, the line between LLMs and traditional computer assisted research tools already has blurred substantially. Large language models have advanced rapidly since their initial release, not only in scale, but in how they access and integrate information. The earliest systems—Chat GPT-2, ChatGPT3 and Claude were closed models: they could generate fluent text but relied entirely on their internal training data. This meant their knowledge was static (frozen at the date of training) and somewhat more prone to "hallucinations."

---

14 Hellyer, Assessing the Influence of Computer-Assisted Legal Research, William & Mary Law School Scholarship Repository, <https://scholarship.law.wm.edu/cgi/viewcontent.cgi?article=1008&context=libpubs> (last visited Oct. 28, 2025)

15 Most recently, Bing and Google have begun using specialize LLMs (Copilot and Gemini) to provide "snippet" summaries preceding the webpages retrieved, so in that sense they are no longer "pure" search engines.

16 Google's PageRank system, for example, measures authority by counting links from other pages. Similarly, LexisNexis and Westlaw rank results by editorial tagging and citation frequency. However, these programs are inherently referential—they direct users to preexisting sources. They may omit relevant material, but they do not fabricate the content produced in a search. For more details on how these search, See Google PageRank Algorithm System: How Google Ranks Pages, GTECH, <https://www.gtechme.com/insights/google-pagerank-algorithm-system-uncovered-the-fundamentals-of-how-google-ranks-pages/> (last visited Oct. 28, 2025)

Beginning around 2023, developers began connecting generative AI models with “traditional” computer search engines, using application programming interfaces (APIs). A similar technique was deployed to supplement the LLMs original training data with proprietary data such as legal case databases or a law firm’s set of model documents. These techniques are examples of “retrieval-augmented generation” (RAG), so named because the LLM retrieves additional data from an external source and uses it to “augment and generate” its final response.

This means that before many “modern” LLMs generate a response, the model first looks up more current or more relevant information from some other location (the internet, a special database, or a law firm’s internal memos or documents). The model uses that fresh information to provide, hopefully, a more accurate and relevant response. An API is like a door or bridge that lets LLM interact with another application (e.g., a database or a web search engine).<sup>17</sup>

For example, if you ask an LLM: “What’s the marginal rates for federal income taxation for couples in 2026?” A RAG enabled LLM might use an internet search-engine accessed through an API to fetch the most recent published tax rates and use that data to write its answer. Supplementing traditional large language models with RAG does improve the usefulness and accuracy of model. The responses will almost certainly be based on the most up to date information online. However, that doesn’t change the fundamental fact that in crafting the response, the LLM is “guessing” what the most likely next word should be, based on the prompts provided and its underlying training.

I came to understand this better shortly after ChatGPT began offering internet search RAG to augment its responses in 2024. I was working on an income tax planning idea, and I needed to know the marginal federal income tax rate for a couple with taxable income of \$370,000. Since I had Chat open on my computer, I decided to ask it, rather than consult Google or Edge to find the appropriate IRS webpage for the information. Instantly, ChatGPT responded that the rate was 32% and (as I had requested for all its responses) it produced multiple web-based sources (URLs) consulted to derive its answer.

But when I examined the tables on the web pages, I saw the correct marginal rate was 24%, not 32%! I flagged the discrepancy, and ChatGPT “apologized,” explaining that it “went too fast” and should have double-checked before responding. Of course, that explanation is absurd, but the experience underscored for me a critical distinction between LLMs and traditional search engines, no matter how accurate the underlying

---

<sup>17</sup> A RAG-enabled LLM might query the Westlaw or Lexis API to retrieve relevant cases, statutes, or law review articles. Those documents are then used as context so the LLMs answer can cite real authorities. But in some setups, instead of using a public Lexis API, a firm may host its own local database of those cases, and the RAG system fetches from there.

data used to create an LLM is, fundamentally the accuracy of the response depends on whether the model “guesses” correctly, and that in turns is based on the relative probability of the next word built into the model initially. That confident, friendly LLM assistant can and will hallucinate legal or numeric results — in contrast to traditional research tools that never invent a law or tax rates.<sup>18</sup>

One likely new frontier in the development of generative AI tools, will be the increased availability of AI agents. AI agents are computer programs designed to allow an LLM to carry out tasks on their own once they understand a user’s goal. For example, an AI agent connected to an LLM might take a completed memorandum it generated, format and save it as a Word document. Another AI agent might load an LLM’s text to an Outlook email, ready for review. These kinds of integrations make working with an LLM faster and more seamless, since the user no longer must switch between applications or repeat steps manually. However, they also raise new concerns: an AI agent could, for instance, not only could just draft that email, it also could send it to the client, all without human review.

#### E. LLMs, APIs, and AI Agents Data Storage and Training

Lawyers and tax professionals may think that anything they upload to consumer or free generative AI tools is private and not reasonably accessible by others. But that assumption can be dangerously wrong. Some LMM providers treat user input—messages, file uploads, client prompts—as additional data for training their LLMs.<sup>19</sup> For example, if you uploaded an confidential client memorandum discussing a particular legal or tax strategy to a public LLM, there is a possibility that the strategy and the client’s use of it might show up in a later response to another party. The similar risk exists for businesses and professional practices that opt to use RAG to incorporate firm-created data (memos, contracts, other documents) as part of a private LLM. Since the LLM will “train” on these documents, failing to scrub them of client confidential information runs the risk that this information will show up later in a response to another unrelated request. Another risk when using LLMs, APIs and AI agents, is that one or all of them may lack the most robust level of security in the transmission of data and its storage.

---

<sup>18</sup> Compounding the hallucination issue, within the past 18 months, major search engines such as Google and Bing began integrating LLMs (or generative AI layers) into their search infrastructure. Rather than only listing links, they now often provide a short, synthesized answer (sometimes labeled as “AI Overview” or “AI Mode”) drawn from indexed web content or other data sources. In effect, the search engine is acting as a hybrid—combining traditional retrieval with generative LLM summarizations.

<sup>19</sup> For example, ChatGPT’s free version uses data imputed by users for training by default, although users can elect to override that setting. See OpenAI, <https://openai.com/consumer-privacy/> (last visited Oct. 28, 2025).

## Part II

### A. The Ethical Framework

The American Bar Association published Formal Opinion 512 (2024) (Model Opinion 512) last July in the wake of several highly published examples of LLM “fabrications” of legal authorities in court filings.<sup>20</sup> Not surprisingly, the Opinion confirmed that the existing ABA Model Rules of Professional Conduct (the Model Rules)<sup>21</sup> apply squarely to the use of all generative AI, including LLMs. The Opinion provided useful touchstones highlighting the challenges and risks associated with the use of the technology, but it wisely steered clear of setting many bright line rules, given the speed at which these tools and their uses are evolving.

Since state bar ethical rules largely adopt the Model Rules, it is worthwhile to review to core areas of emphasis in Model Opinion 512, recognizing of course that the State bars<sup>22</sup> and courts<sup>23</sup> have been active in the area as well. The American Institute of Certified Public Accountants (AICPA) also has focused on issues similar in many respects to those flagged by Model Opinion 512,<sup>24</sup> as have regulatory bodies in other countries.<sup>25</sup>

---

20ABA Standing Committee on Ethics and Professional Responsibility, Formal Opinion 512, Generative Artificial Intelligence Tools (July 29, 2024),

[https://www.americanbar.org/content/dam/aba/administrative/professional\\_responsibility/ethics-opinions/aba-formal-opinion-512.pdf](https://www.americanbar.org/content/dam/aba/administrative/professional_responsibility/ethics-opinions/aba-formal-opinion-512.pdf).

21ABA Comm. on Ethics & Prof'l Responsibility, ABA Model Rules of Professional Conduct, [https://www.americanbar.org/groups/professional\\_responsibility/publications/model\\_rules\\_of\\_professional\\_conduct/model\\_rules\\_of\\_professional\\_conduct\\_table\\_of\\_contents/](https://www.americanbar.org/groups/professional_responsibility/publications/model_rules_of_professional_conduct/model_rules_of_professional_conduct_table_of_contents/) last accessed October 28, 2025 (here after “Model Rules”)

22 See e.g., California: State Bar of Cal., Standing Comm. on Prof'l Responsibility & Conduct, Practical Guidance for the Use of Generative Artificial Intelligence in the Practice of Law (July 2024), <https://www.calbar.ca.gov/Portals/0/documents/ethics/Generative-AI-Practical-Guidance.pdf>; Florida: Fla. Bar Prof'l Ethics Comm., Fla. Ethics Op. 24-1 (2024) <https://www.lawnext.com/wp-content/uploads/2024/01/FL-Bar-Ethics-Op-24-1.pdf>.

23 Standing Order, Use of Generative AI (N.D Ohio July 2024)

<https://www.ohnd.uscourts.gov/sites/ohnd/files/Boyko.StandingOrder.GenerativeAI.pdf> (Prohibits use in pleadings); See also, Jessica Chasmar, Federal Judges Acknowledge Court Ruling Errors Tied to Staffers' AI Use After Grassley Inquiry, FOX News (Oct. 24, 2025), <https://www.foxnews.com/politics/federal-judges-acknowledge-court-ruling-errors-tied-staffers-ai-use-after-grassley-inquiry> (Law clerks prohibited from using generative AI for research or pleadings after hallucinated content in Court orders).

24 See e.g. “AICPA addresses the ethics of AI in audits”

[https://www.americanbar.org/groups/professional\\_responsibility/publications/model\\_rules\\_of\\_professional\\_conduct/model\\_rules\\_of\\_professional\\_conduct\\_table\\_of\\_contents/](https://www.americanbar.org/groups/professional_responsibility/publications/model_rules_of_professional_conduct/model_rules_of_professional_conduct_table_of_contents/) July 26, 2024.

25 For example, in 2024, the European Union adopted the Artificial Intelligence Act, Regulation (EU) 2024/1689, Artificial Intelligence Act, 2024 O.J. (L 265) 1. The Act classifies AI systems used in legal advice and representation as 'high-risk' applications, subject to strict requirements including risk assessment, human oversight, and documentation obligations (Id. Annex 3).

Finally, representation and practice before the IRS and the United States Tax Court imposes similar restrictions as well.<sup>26</sup> In particular, Internal Revenue Code Section 6713 imposes a strict civil penalty for disclosure of taxpayer information in a representation, and Section 7216, adds criminal penalties for reckless disclosure. The exact scope of these provisions as they relate to the use of generative AI in a tax practice is untested, but because Section 6713 does not seem to excuse even inadvertent disclosures, tax practitioners need to be particularly mindful of any use of AI that transmits or stores unencrypted data.<sup>27</sup>

Yet three primary ethical concerns are embodied to some degree in all these pronouncements: “Competence and Candor;” “Confidentiality and Client Consultation;” and “Supervision.” The balance of this article focuses on these, first focusing on the peculiar compliance risks posed by the use LLMs and other generative AI tools in a professional practice and then offering some practical ideas for meeting these challenges.

## B. Competence and Candor

These two concerns are embodied in several of the Model Rules. Model Rule 1.1 and Comment 8, require knowledge, skill and thoroughness and preparation reasonably necessary for each representation undertaken, and Comment 8 makes clear that this includes “keeping abreast of relevant technology.” The requirement for “candor,” is embodied in Model Rules 3.1, 3.3 (before administrative and judicial bodies and 8.4 (assertion of frivolous claims and positions). Together these prohibit attorneys from asserting knowingly make an erroneous statement of law or fact, taking frivolous positions, or engaging in conduct involving fraud or misrepresentation or prejudicial to the administration of justice.

Judicial and administrative bodies supplement these rules by requiring attorneys to certify that they have taken steps to assure to the accuracy of pleadings, as well as positions taken on tax returns. In addition, some courts have issued special rules governing the use of generative AI in pleadings or prohibited its use all together.<sup>28</sup>

---

26 31 C.F.R. §§ 10.22 (Due diligence), 10.33 (Best practices – client notice), 10.35 (Practitioner competence), 10.36 (Written Policies for use of LLMs and AI), 10.37 (Use of flawed or unsupported LLM material), 10.51 (Issuing a false or reckless opinion created by an LLM) (2024). Treasury Department Circular No. 230, Regulations Governing Practice Before the Internal Revenue Service, 31 C.F.R. pt. 10 (2024), <https://www.irs.gov/pub/irs-pdf/pcir230.pdf>.

27 See generally, Edward R. Jenkins & James Sansone, Tax Ethics and Use of Generative AI Systems, *The Tax Adviser* (Feb. 2024), <https://www.thetaxadviser.com/issues/2024/feb/tax-ethics-and-use-of-generative-ai-systems/>; Barclay R. Taylor, When Is Consent Enough? Tax Return Information in the Uncharted AI Landscape, 181 *Tax Notes Fed.* 1580 (Nov. 27, 2023), [https://www.mmmlaw.com/resources/When%20Is%20Consent%20Enough%20-%20Tax%20Return%20Information%20in%20the%20Uncharted%20AI%20Landscape\\_QzLqZ8E.pdf](https://www.mmmlaw.com/resources/When%20Is%20Consent%20Enough%20-%20Tax%20Return%20Information%20in%20the%20Uncharted%20AI%20Landscape_QzLqZ8E.pdf) .

28 See Footnote 23, *infra*.

*Mitigation Strategy - Always Keep in Mind How LLMs Work and Check Their Output!*

The numerous examples of AI generated hallucinated cases and legal citations in court filings,<sup>29</sup> including pleading before the Tax Court,<sup>30</sup> and even some Judicial Opinions<sup>31</sup> demonstrate that using an LLM and the related tools described in Part 1, without understanding their limitations, will almost certainly cause attorneys and tax professionals to fall short of their obligation to accurately state the facts and relevant law in pleadings. Yet more of us now use generative AI and LLMs in other areas of our lives each day. Within that last few months many now regularly rely on the LLM-created summary answers (snippets) that are imbedded in Microsoft Bing and Google<sup>32</sup> to answer factual questions, without bothering to consulting the list of underlying websites sources provided. While there are some accuracy and hallucination problems with these tools, particularly when they are asked to weigh in on controversial questions, antidotal experience suggests they can do well in answering some factual questions.<sup>33</sup>

As more of us use these tools largely without incident in everyday settings, it invariably becomes easier to trust them more. However, there is evidence that LLMs – particularly “general purpose” LLMs that have not been exposed to specific law or tax data struggle with legal analysis. An article published in the Journal of Legal Analysis in early 2024 found that early general purpose LLMs could provide erroneous answers to open ended legal questions up to 88% of the time.<sup>34</sup>

Admittedly, the LLMs tested in this article was several generations behind those in use today, and they lacked the ability to access the internet or other RAG enhancements such as dedicated legal databases to improve accuracy. However a similar study of legal based databases in published in mid 2024 showed hallucination rates even for LLMs built on

---

29 See AI Hallucination Cases Database, <https://www.damiencharlotin.com/hallucinations/> Footnote 2, *infra*.

30 Footnote 3, *infra*.

31 *Id.*; See Footnote 24, *infra* and Bloomberg Law, Judges Called Out for Nonfactual Rulings Admit to Use of AI, <https://news.bloomberglaw.com/business-and-practice/judges-called-out-for-nonfactual-rulings-admit-to-use-of-ai>

32 See text accompanying Footnote 18, *infra*.

33 See, e.g., Alice Li & Luanne Sinnamon, Generative AI Search Engines as Arbiters of Public Knowledge: An Audit of Bias and Authority, Proc. Ass’n for Info. Sci. & Tech. (2024), <https://arxiv.org/abs/2405.14034>; Yifan Hu et al., Evaluating Robustness of Generative Search Engine on Adversarial Factual Questions (2024) (preprint), <https://arxiv.org/abs/2403.12077>; Jason Wei et al., Long-Form Factuality in Large Language Models (2024) (preprint), <https://arxiv.org/abs/2403.18802>; Ryan Heinz et al., Search Engines in an AI Era: The False Promise of Factual Answer Engines (Oct. 2024) (preprint), <https://arxiv.org/html/2410.22349v1>; Daniel Fernández-Pichel et al., Evaluating Search Engines and Large Language Models for Answering Health Questions, NPJ Digital Medicine (2025), <https://www.nature.com/articles/s41746-025-01546-w>

34 Matthew Dahl, Varun Magesh, Mirac Suzgun & Daniel E. Ho, Profiling Legal Hallucinations in Large Language Models, arXiv preprint arXiv:2401.01301v2 (June 21, 2024), <https://arxiv.org/pdf/2401.01301v2>

proprietary legal databases ranged from 17% to 33%!<sup>35</sup> While results hopefully would be better today – and these models seem to continue to improve with the advent of specialized tools designed to spot mis-citations,<sup>36</sup> the admonition to NEVER file any pleading before an administrative or judicial tribunal, or send any client a substantive legal memorandum or email that was produced in whole or in a part by an LLM unless it has been thoroughly reviewed by a “human” for accuracy, continues to remain true today.

#### *Mitigation Strategy – Use Good “Prompts” to Improve the Odds of an Accurate Response.*

The adage, “garbage in garbage out” applies when working with LLMs. LLMs are more prone to generating inaccurate, irrelevant, or hallucinatory responses when given short open-ended questions than when presented with precise and well-constructed instructions. “Prompting” is the term used to describe the instructions provided to an LLM when to initiate an interaction. Understanding the anatomy of a good prompt can appreciably increase your chance of getting accurate and relevant information from the LLM.<sup>37</sup> There are many articles on how to create a good prompt<sup>38</sup> but at times the technical jargon used to describe various prompting techniques, can unnecessarily obscure some simple principles. Good prompting does not require training in a special computer code, and all good prompts seem to address three elements “Roles,” “Goals,” and “Limitations.”

#### *Writing an Effect Prompt -- Roles: Defining Who’s - Who in the Conversation*

Assigning clear roles in your prompt guides the LLM to adopt the right tone, expertise, and style when responding, and it can also help the model search out the best training data that it will use when formulating its response. Defining roles involve two perspectives:

- The LLM’s Role: This is the persona or expertise you want the LLM to embody. For example, “Act as a financial advisor,” or “You are a law school professor.” This helps the model tailor its language, depth, and focus, producing responses that are contextually appropriate and that are aligned with your expectations.
- The User’s Role: This is the “role” of the person who will use the information the LLM will produce, most often you or the client. For example, if you are drafting a memorandum to a client explaining “wash sale rules,” is that client an experienced investor or a novice?

---

35 Varun Magesh, Faiz Surani, Matthew Dahl, Mirac Suzgun, Christopher D. Manning & Daniel E. Ho, Hallucination-Free? Assessing the Reliability of Leading AI Legal Research Tools, arXiv preprint arXiv:2405.20362v1 (May 30, 2024), <https://arxiv.org/abs/2405.20362> (last visited Nov. 1, 2025)

36 See e.g., K. Rattray, “6 Best Legal Proofreading Software Options for 2025: Top Picks for Law Firms” (Oct. 2025) <http://www.clio.com/blog/best-legal-proofreading-software>

37 See e.g., Prompt Engineering in Higher Education: A Systematic Review to Help Practices Move from Hype to Understanding, Int’l J. Educ. Tech. Higher Educ. (2025), <https://educationaltechnologyjournal.springeropen.com/articles/10.1186/s41239-025-00503-7>.

38 See List of Prompting Articles (Appendix B).

### Writing an Effective Prompt -- Goals: Clarifying the Desired Outcome

Every effective prompt should be anchored in a clearly articulated set of objectives or goals. For example, before you enter your prompt in an LLM consider:

- Are you seeking a summary, a detailed explanation, a memorandum, a list of pros and cons, or something else?
- Should the response include specific hyperlinks to supporting material available on websites?
- Should the website consult specific memoranda, statutes or authority you have uploaded to generate a response?
- Do you want a Word formatted document?

Stating objectives explicitly in the prompt helps the underlying LLMs underlying algorithm generate a response that is more closely aligned to your needs.

### Writing an Effective Prompt -- Limitations: Setting the Rules and Boundaries

Finally, effective prompts need to detail acceptable limits you want to impose on the LLM as it does its work. This can help control the quality and reliability of the output. Here, think in terms of how you would advise a human assistant who was doing the work for you.

Examples might include –“seek out the most current information available on the web” or “only consult peer reviewed articles from recognized journals.” As you focus on this element, it is best not to assume the LLM will act rationally. For more involved projects, it also can save time and effort if you instruct the LLM to break the overall task up into various steps, and proceed further, only after receiving your approval.

It also may be helpful to ask the LLM not to hallucinate or speculate in providing a response or to stress the importance of verifying the accuracy of its proposed response before finalizing the response. Positive requests such as “provide sources for any conclusion that does not involve common knowledge” or “if you have inadequate data state a conclusion, say so -- do not speculate” may be useful as well, particularly if you are using a general purpose LLM.

### Writing an Effective Prompt -- A Practical Example:

To illustrate these elements, the following is an example of a prompt that applies these three elements of a prompt to generate a draft tax memorandum for a tax professional.

*Roles:*

*You are acting as a legal research assistant trained in U.S. federal income tax law. Your task is to analyze a potential deduction issue under I.R.C. §162(a) and provide a concise, professionally written memorandum for an attorney or CPA review.*

*Goals:*

*You have access to online legal sources and should include hyperlinks to authoritative materials such as:*

- *Cornell Law School's Legal Information Institute ([law.cornell.edu](http://law.cornell.edu))*
- *IRS.gov (revenue rulings, regulations, and guidance)*
- *Casetext.com, Justia.com, or other official case databases*

*Your tone should be professional, objective, and consistent with the writing style of a tax memorandum.*

*Your objective is to determine whether a settlement payment to a former employee for wrongful termination is deductible as an ordinary and necessary business expense under §162(a).*

*Specifically, you must:*

1. *Identify the governing legal standard under §162(a) and summarize the key requirements.*
2. *Cite and link to authoritative sources—cases, Treasury Regulations, and IRS rulings—relevant to deductibility of settlement payments and the “ordinary and necessary” test.*
3. *Explain how courts and the IRS distinguish between business vs. personal and ordinary vs. capital expenses.*
4. *Apply those authorities to the following facts:*
  - *The taxpayer operates a small real estate brokerage.*
  - *The taxpayer paid a one-time settlement to a former employee to resolve a wrongful termination claim.*
  - *The payment was compensatory, not punitive, and was made to avoid litigation.*
5. *Present your response using this structure:*
  - *Issue*
  - *Rule / Authority*
  - *Application*
  - *Conclusion*
6. *Include embedded hyperlinks for each authority used.*
7. *Prepare the result as a professional memorandum in Microsoft Word (.docx) format with standard 1-inch margins, 12-point font, and preserved hyperlinks.*

*Limitations:*

- *Do not fabricate authorities, citations, or links.*
- *Use only verifiable, publicly accessible sources.*
- *Do not provide tax advice or make client-specific recommendations—focus on legal research and analysis.*
- *If the law is unclear or conflicting, explain the uncertainty rather than guessing.*
- *Avoid overly technical explanations of AI or computational methods; maintain focus on legal reasoning.*
- *Keep the tone neutral and factual; avoid advocacy.*

The purpose of including this prompt, is not to suggest it as the “perfect model” for this situation. Instead, it is intended to illustrate that prompting can be accomplished without resorting to any specific technical jargon. The unedited text of the memo produced by Microsoft Copilot using this prompt is attached as Appendix B.

It is easy to forget that we are at the very early stages of learning how best to interact with LLMs. There are many experts out there with ideas about how to work with them. At the same time programmers are constantly working to improve LMMs, to make them more intuitive, responsive and less prone to hallucination. In fact, for unusual projects, it is worth considering providing the model a draft of your prompt and asking if additional information or instructions would improve it.

While good prompting can make LLM responses “better” they will not eliminate the risk of hallucinations, and you cannot even assume that the LLM will follow the instructions in the prompt. Most recently, I requested that the LLM provide case law interpreting the “true object test” -- a doctrine used throughout the United States to determine whether a sale is subject to sales tax (a transfer of tangible personal property) or a nontaxable transfer of a service or personal property. My prompt specifically requested that only Missouri cases be cited in the response, and I asked for URLs to any cases cited.

Yet the LLM cited a California case (that did involve the true object test) in its response! Compounding the problem, it used the correct case name but added a hallucinated Southwest 3d citation to a Missouri Supreme Court decision on an unrelated topic! While all this is frustrating, I was quickly able to quickly spot the error because I had requested a website link to all cases.

For now, those of us that use LLMs for legal research and drafting must accept this shortcoming. When accuracy is important – and it always is in a professional practice – you simply must check (and in some cases double check) an LLMs response to make sure it isn’t “making stuff up.”

*Mitigation Strategy -- Go Slowly When Exploring Practical Uses and Strategies for LLMs in Your Practice*

Within just a couple of years, large-language models (LLMs) have moved from pilot projects to daily utilities across many legal and tax practices. Surveys from Thomson Reuters reflect this trend, with adoption and concrete use-cases rising year over year.<sup>39</sup> Uses for LLMs include document and zoom meeting summaries; first drafts of letters, emails, and memos; analysis of cases and statutes; editing memorandum, correspondence and pleadings; document generation and even role playing for discovery or brainstorming ideas.

The level of experience, typical tasks performed in individual practices, and the professionals’ willingness to explore new technology will vary, so suggesting specific

---

39 Thomson Reuters Future of Professionals Report 2025, at 5 (2025), available at <https://www.thomsonreuters.com/content/dam/ewp-m/documents/thomsonreuters/en/pdf/reports/2025-generative-ai-in-professional-services-report-tr5433489-rgb.pdf> .

strategies a lawyer or tax professional should use to identify optimal use cases for LLMs in their practice is difficult. However, some general guidance will apply in most situations.

First and foremost, it is important to select the best LLM tool for the task. Here, confidentiality can be a significant concern. The risk that an LLM or a third-party API or AI Agent might disclose client information is a real, and strategies to protect against the disclosure are discussed in the next section, but most start with analyzing if the LLM has robust cybersecurity in place and does not use LLM prompts for training purposes. Many professionals find it also is best to begin working with LLMs that reside as part of tools you already trust. For example, enterprise licenses Microsoft 365 can employ Copilot in Outlook and Word for internal email drafting, proofreading, and formatting. For legal analysis, one can try LLM features inside Westlaw, LexisNexis, or Bloomberg Law (rather than a general-purpose chatbot). Alternatively, tax professionals might consider professional or enterprise versions of ChatGPT or another general purpose LLM and structure prompts to require the model to include embed links to legal authority relied on in its response.

### C. Confidentiality and Client Consultation

#### *Confidentiality*

Few professional obligations are more deeply rooted than a lawyer's duty to preserve client confidences, and that obligation is shared by other tax professionals, particularly when taxpayer specific information is used to prepare a return.<sup>40</sup> Web-based or cloud-hosted generative AI tools capable of drafting, summarizing, or analyzing legal text within seconds are increasing exponentially, and each implicate Model Rule 1.6's command that a lawyer "shall not reveal information relating to the representation of a client unless the client gives informed consent"<sup>41</sup> as well as even stricter statutory prohibitions that specifically relate tax professionals.<sup>42</sup> For lawyers, it is worth remembering that public filings do not lose their confidential character simply because they can be found online.<sup>43</sup>

Rule 1.6(c) does not make lawyers guarantors of secrecy. Instead, it requires "reasonable efforts to prevent the inadvertent or unauthorized disclosure of, or unauthorized access to," client information.<sup>44</sup> What is "reasonable" depends on the sensitivity of the data, the likelihood of disclosure, and the cost and feasibility of safeguards.<sup>45</sup> Formal Opinion 477R urges lawyers to conduct risk analyses before transmitting information electronically, weighing the sensitivity of the data, the likelihood of disclosure if additional safeguards

---

40 See text accompanying footnotes 26-27.

41 Model Rule 1.6, *supra*.

42 IRC Sections 6713 and 7216 (note that Section 6713 is an absolute prohibition, with no accommodation even if reasonable steps are taken to safeguard the information or has obtained client consent).

43 Model Rule 1.6, *supra*, cmt. [3].

44 But see, IRC Section 6713 (Penalty apparently can be imposed for certain disclosures of taxpayer information without regard to steps taken to secure it).

45 Model Rule 1.6(c) cmt. [18].

are not employed, the cost of employing additional safeguards, and the difficulty of implementing them.<sup>46</sup>

Logically, this means that the same cyber security standards used protect to email communications, the storage of data in the “cloud,” or the use of third-party software applications, also applies to generative AI systems (LLMs, RAGs, APIs and AI agents). It follows then, that at a minimum, before using any of these tools to transmit client information, lawyers must understand enough about those systems to evaluate the disclosure risk they pose.<sup>47</sup> Applied to generative AI and related tools, that means knowing where the data goes, who controls it, and whether it is stored or used for model training.

Unless a client has consented, the requirement for confidentiality covers information relating to the representation, whatever its source, even a client’s identity publicly filed documents can remain confidential. For example, unless the client has consented, uploading a draft pleading to a free online LLM for rewriting could still be a Rule 1.6(a) disclosure if the model that uses the uploaded data for training or if it lacks robust cyber security and retains the text of that uploaded information.

#### *Communication and Client Consent Model Rule 1.4*

Rule 1.4 requires a lawyer to “keep the client reasonably informed about the status of the matter” and to “explain a matter to the extent reasonably necessary to permit the client to make informed decisions regarding the representation.”<sup>48</sup> When a lawyer’s use of generative AI could materially affect the handling of client data or the nature of services provided, that duty may require disclosure and, in some instances, informed consent.

For routine, low-risk uses of artificial intelligence within secure environments, client consent may be implied. Lawyers have long relied on grammar checkers, document-management systems, and cloud-based email servers without specific client approval. Similarly, using a private RAG enabled LLM that never transmits data outside the firm’s network is likely within the implied authorization necessary to carry out the representation --- assuming the documents used to create and update that tool are carefully purged of any client identifying information. The question in either case, is whether the client would reasonably expect the lawyer to use this technology in delivering competent representation, and of course that will vary with the nature of the technology and the client’s desires.

However, even when disclosure is not strictly required, lawyers (and for that matter any tax professional) should consider informing clients in its engagement materials that the firm may use AI to improve the quality of services provided or to more efficiently provide

---

46 ABA Formal Op. 477R (2017), Securing Communication of Protected Client Information available at <https://docs.tbpr.org/pub/aba%20formal%20opinion%20477.authcheckdam.pdf>

47 See also, American Bar Association, Formal Opinion 498, Virtual Practice (Mar. 10, 2021), available at <https://inns.innsofcourt.org/media/196625/aba-formal-opinion-498-1.pdf>.

48 Model Rule 1.4(a)–(b).

those services, subject to reasonable safeguards. This transparency satisfies Rule 1.4 and mitigates potential misunderstandings if a client later learns AI tools were used in the representation without their knowledge.

For example, Model Rule 1.5 prohibits unreasonable fees and expenses. In an hourly billing arrangement, if an LLM allows a lawyer or other professional to complete a task more quickly, it would certainly be unreasonable to bill the client as if the work had taken longer, and of course firms cannot double-charge clients for both attorney time and the cost of the AI tool unless disclosed and agreed to in advance. Formal Opinion 512 emphasized that if LLMs are used, fees charged to client must reflect actual time spent and value delivered, not the mere use of technology. Disclosing that generative AI and other tools may be used to deliver client services more effectively, can create a more transparent working relationship by assuring that the client understands that while AI may be used in the engagement, it will not be used to overcharge them for the work performed.

#### *Mitigation Strategy – Use Enterprise Level LLMs and Related Applications*

It is easy to fill several pages of text with specific technical privacy security protocols lawyers and other tax professional should use to protect the privacy of client information. However, a more useful approach is to compare the privacy risks associated with using generative AI tools to what most all law firms traditionally follow to protect email communications and cloud-based storage of client information. Most every firm transmits confidential client information each day through the internet and store it offsite, on servers provided by third party providers. Over the years lawyers have become comfortable that provided certain protocols are followed, this does not create any substantial disclosure risk of disclosure, since the data is encrypted in transit (from the professional's computer to the remote party's cloud storage location) and it remains encrypted until it is retrieved for use. In this respect, so long as an LLM, including any RAG enabled application or AI Agent meet those same standards, the risk of disclosure are similar.

However, any generative AI (including an LLM) does differ from traditional applications that transmit and store data off-site. The primary difference is that LLMs and similar applications process and generate data, in addition to transmitting and storing it. In other words, they take the information (e.g., text or files uploaded in prompts) and manipulate it to create new output (answer the question, draft the document, or summarize files you uploaded, etc.). This is an additional privacy concern that attorneys and tax professionals need to analyze before using any generative AI tool or application.

Of course, as mentioned several times already, one important step, is to only use LLMs that do not use the data you upload or input to the LLM to “train” the AI model. Why? Because once that data is used in training, it is possible that the LLM will use it to generate

a response to a similar line of questioning it does exist, at least in the context of AI's used to produce computer software.<sup>49</sup>

This means training attorneys and tax professionals, staff, and IT Personnel on the need to focus on what happens to the data (the prompts and interactions created) by an LLM or other generative AI product. The core questions include whether the created data is encrypted, who holds it, how long it is retained, and how is it used. In most cases, there are no hard and fast “rules” here, but most paid business enterprise grade paid subscriptions maintain similar levels of encryption and security and typically do not maintain user data or output for longer than 30 days, unless specifically requested by the user. These platforms offer enhanced security, data isolation, and contractual safeguards that consumer or free-tier LLMs and their tools may lack.<sup>50</sup>

### *Mitigation Strategy – Anonymize Interactions with LLMs Whenever Practical.*

The unfortunate truth is that no matter what level of security used by an LLM provider, no system is 100% foolproof. For this reason, a second important mitigation strategy is simply to be judicious when deciding what information will become part of an LLM prompt. For many tasks there is little need to provide client names, taxpayer identification numbers or other highly sensitive information.<sup>51</sup> Instead make it a practice to anonymize documents or information contained in prompts.<sup>52</sup>

---

49 Judge Throws Out Majority of Claims in GitHub Copilot Lawsuit, Legal.io, <https://www.legal.io/articles/5516216/Judge-Throws-Out-Majority-of-Claims-in-GitHub-Copilot-Lawsuit> (2024). (Lawsuit against Microsoft claiming that parts of copyrighted software programming was reproduced by Copilot).

50 For example, Microsoft 365 Copilot provides enterprise-grade data protection, including tenant isolation, encryption at rest and in transit, and strict privacy controls. User prompts, responses, and organizational data are not used to train foundation models. Copilot supports GDPR, ISO/IEC 27018, and other compliance standards. See <https://learn.microsoft.com/en-us/copilot/microsoft-365/microsoft-365-copilot-privacy>. (Last accessed on October 30, 2025)

51 This also includes proprietary documents that a firm decides to train a RAG enabled LLM. To avoid cross-exposing client information from a prior engagement in that private LLM, the existing documents used to build the RAG model must be stripped of client identifying information.

52 Practical redaction guidance is available from sources such as the ABA Law Technology Today blog and the International Association of Privacy Professionals, which publish anonymization best practices. See ABA Law Technology Today, Best Practices for Confidentiality in the Age of AI-Powered Legal Tools, <https://www.jdsupra.com/legalnews/best-practices-for-confidentiality-in-1379976/> (last visited Oct. 30, 2025); International Association of Privacy Professionals, De-Identification of Personal Information, <https://iapp.org/resources/article/de-identification-of-personal-information/> (last visited Oct. 30, 2025).

## *Mitigation Strategy – Inform Clients in Engagement Letters of Your Policies Surrounding the Use of LLMs and Related Applications*

While we eventually may get to the point where standards for the use of LLMs and other generative AI tools are commonplace and universally accepted, at least at this point for attorneys and tax practitioners should routinely disclose their policy of using generative AI in representations. Even if that communication is not required in a particular case, proactive communication builds trust and can preempt misunderstandings about accuracy, billing, and data protection.

A short clause might read:

Our firm employs secure artificial-intelligence tools to assist in document drafting, research, and analysis. These systems are used only under conditions that protect client confidentiality, and no client information is shared with vendors for model training.”

Language such as this if coupled with clear internal controls, demonstrates diligence under Rules 1.4 and 1.6. Several examples of other versions are included at Appendix C.

### D. Supervision

#### *Supervising Attorneys and Staff –Model Rules 5.1 and 5.3*

Model Rule 5.1 addresses the responsibilities of partners, managers, and supervisory lawyers to ensure that other lawyers in the firm comply with the Rules of Professional Conduct.<sup>53</sup> Lawyers with managerial authority also must make reasonable efforts to establish internal policies and procedures designed to provide reasonable assurance of compliance.<sup>54</sup>

The rapid adoption of large language models (LLMs) in nearly every aspect of our “online” lives introduce a new layer of complexity to these duties since these tools can produce convincing but inaccurate or misleading outputs. When attorneys or staff use LLMs without adequate training or a clear understanding of the risks—such as hallucinated case law, breaches of confidentiality, or unauthorized practice of law—they risk inadvertently violate ethical rules or compromise client interests, and almost certainly they risk damaging the reputation of their firm. Yet because an LLMs output looks correct, it can be extremely difficult for a supervisor to recognize.

---

53 ABA Model Rule 5.1, Responsibilities of a Partner or Supervisory Lawyer, [https://www.americanbar.org/groups/professional\\_responsibility/publications/model\\_rules\\_of\\_professional\\_conduct/rule\\_5\\_1\\_responsibilities\\_of\\_a\\_partner\\_or\\_supervisory\\_lawyer/](https://www.americanbar.org/groups/professional_responsibility/publications/model_rules_of_professional_conduct/rule_5_1_responsibilities_of_a_partner_or_supervisory_lawyer/)

54 ABA Model Rule 5.3, Responsibilities Regarding Nonlawyer Assistance, [https://www.americanbar.org/groups/professional\\_responsibility/publications/model\\_rules\\_of\\_professional\\_conduct/rule\\_5\\_3\\_responsibilities\\_regarding\\_nonlawyer\\_assistant/](https://www.americanbar.org/groups/professional_responsibility/publications/model_rules_of_professional_conduct/rule_5_3_responsibilities_regarding_nonlawyer_assistant/)

### *Mitigation Strategy – Proactive Education and Written Policies*

The firm's use of generative AI should be reduced to a written policy covering the use of these tools in the firm's practice. Of course, it isn't feasible to create a one size fits all policy for every law firm or tax practice. However, any written policy statement should communicate expectations, and it must be understandable. Most should consider including a list of authorized LLMs and related RAGs, APIs and AI agents that meet standards for client security; forbid input of client-identifiable information into any unapproved or public LLM; set out the minimum requirements for identifying new tools (for example reviewing the vendor's terms of service, privacy policy, and security certifications (e.g., SOC 2 Type II, ISO 27001), and mandate human oversight and review any work product created for clients with generative AI for hallucinations and error, with particular attention on written material filed with a court or administrative tribunal. Examples of effective supervision policies are provided in the Appendix D.

Just as important is the need to educate all employees and professionals how LLMs work, the natural tendency to trust their output, and the mitigation strategies the firm uses to guard against inadvertent disclosure or LLM hallucination. Beyond emphasizing caution and compliance, these training sessions can emphasize good habits such as effective prompting, selection of the best AI tool, and anonymizing inquiries whenever possible as well. This training provides the "why" behind the firm's written procedures and expectations, and it can help develop a culture where everyone works to use generative AI tools responsibly. After all, the only thing worse than not having a written policy for the use of generative AI and its related tools, is to have one that isn't understood or followed.

### **Conclusion**

Generative AI tools are transforming legal and tax practices. Their ability to generate text, summarize complex materials, and accelerate drafting presents genuine opportunities for efficiency. But they also create risks unlike those posed by traditional research and practice tools. The lesson for professionals is straightforward: the technology is new, but the ethical framework is not. Competence requires understanding both the benefits and the limitations of AI. Candor prohibits submitting AI hallucinations to a tribunal. Confidentiality requires careful attention to vendor policies and client consent. Communication builds trust and foster effective delivery of the services client's demand. Effective supervision ensure that all members of the firm use generative AI wisely.

Circular 230 adds a further layer for tax professionals demanding diligence as to accuracy and privacy and reinforce the duty to independently verify AI-generated work. The U.S. Tax Court's recent orders striking LLM hallucinated filings serve as cautionary tales that reliance on this technology is no substitute for professional judgment.

Ultimately, LLMs are best understood not as a replacement for human expertise, but as a new category of assistant—powerful, fallible, and in need of supervision. Lawyers and tax professionals who understand this can integrate this technology into their practices responsibly and capture its benefits while avoiding ethical pitfalls. Those who fail to adapt risk sanctions, reputational harm, and, most importantly, diminished trust from clients and courts.

Generative AI can become a “trusted partner” for a tax professional, but it also can easily become a “compliance risk.” The outcome does not lie in the technology; rather it depends on the care taken by the human professionals that use it.

## Appendix A: Articles on Legal Prompting

1. The Ultimate Guide to Legal Prompt Engineering, Brightflag (Apr. 10, 2025), <https://brightflag.com/resources/legal-prompt-engineering/> (comprehensive best practices and frameworks for legal prompt engineering).
2. Introduction to Legal Prompt Engineering in Generative AI, Deloitte Legal Briefs (Jan. 28, 2024), <https://legalbriefs.deloitte.com/post/102iwka/introduction-to-legal-prompt-engineering-in-generative-ai> (overview of prompt engineering methods and impacts for legal professionals).
3. Prompt Engineering Is The New Drafting, Artificial Lawyer (Apr. 29, 2025), <https://www.artificiallawyer.com/2025/04/30/prompt-engineering-is-the-new-drafting/> (analytical essay on prompt engineering as a new legal drafting skill).
4. ADVANCED PROMPT ENGINEERING: A Guide for Lawyers, With Examples, AI for Lawyers Substack (Oct. 17, 2024), <https://aiforlawyers.substack.com/p/advanced-prompt-engineering-a-guide> (practical guidance and real-world examples for law).
5. Unleashing the potential of prompt engineering for large language models, 2025 PMC 12191768, <https://pmc.ncbi.nlm.nih.gov/articles/PMC12191768/> (scientific review of prompt engineering methods for LLMs, including legal contexts).
6. Legal Prompt Engineering Guide, Centre for Legal Innovation (May 4, 2025), <https://www.cli.collaw.com/resource-hub/2025/05/05/legal-prompt-engineering-guide> (detailed guide for law students and practitioners on prompt design and quality control).
7. Exploring the Effectiveness of Prompt Engineering for Legal Reasoning Tasks, in Proc. ACL Findings (July 2023), <https://aclanthology.org/2023.findings-acl.858.pdf> (peer-reviewed research evaluating prompt engineering for legal tasks).
8. Artificial Intelligence Legal Holds: Preserving Prompts & Outputs, Nat'l L. Rev. (Sept. 4, 2025), <https://natlawreview.com/article/when-chatgpt-meets-legal-hold-survival-guide-house-counsel-who-didnt-sign> (legal requirements for preserving AI prompts and outputs as ESI).
9. A Guide to Legal Prompt Engineering, Juro (Oct. 16, 2023), <https://juro.com/learn/legal-prompt-engineering> (introduction and foundational guide to crafting prompts for legal AI tools).
10. The Legal Professional's Guide to Prompt Engineering, Relativity (2025), <https://resources.relativity.com/rs/447-YBT->

[28/images/The%20Legal%20Professional%E2%80%99s%20Guide%20to%20Prompt%20Engineering.pdf](#) (industry guide summarizing legal prompt workflows and survey results).

11. The Art of AI Prompting in Law and Dispute Resolution Practice, SSRN (Sept. 28, 2025), [https://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=5544018](https://papers.ssrn.com/sol3/papers.cfm?abstract_id=5544018) (practical and theoretical discussion for dispute resolution and practice).

12. Introduction to Legal Prompt Engineering (LPE), Legal Prompt Guide (Oct. 14, 2025), <https://www.legalpromptguide.com/1.-introduction-to-legal-prompt-engineering-lpe> (step-by-step guide and frameworks for effective legal prompts).

13. Prompt Engineering 101 for Lawyers, N.C. Bar Ass'n (Aug. 20, 2024), <https://www.ncbar.org/nc-lawyer/2024-08/prompt-engineering-101-for-lawyers/> (introduction to prompt engineering skills for the legal sector).

14. Mastering AI Prompts for Legal Professionals in 2025, ContractPodAi (May 20, 2025), <https://contractpodai.com/news/ai-prompts-for-legal-professionals/> (use-case-driven prompt templates and compliance tips).

15. Creating Prompts for Legal GenAI: Legal Research Manual, Queen's Univ. Library (Jan. 4, 2024), <https://guides.library.queensu.ca/legal-research-manual/prompt-engineering-legal-research> (library guide on prompt design for GenAI legal research).

16. Prompt Engineering, its Application to the Law and a Recent Article, LinkedIn Pulse (May 10, 2024), <https://www.linkedin.com/pulse/prompt-engineering-its-application-law-recent-article-ralph-losey-k8dce> (discussion on lawyer adoption of prompt engineering through McKinsey lens).

17. The Role of Well-Designed Prompts in Applying AI to Legal Work, Thomson Reuters Legal Blog (Aug. 19, 2025), <https://legal.thomsonreuters.com/blog/the-role-of-well-designed-prompts-in-applying-ai-to-legal-work/> (analysis of prompt impact on accuracy in legal workflows).

18. Effective Prompt Writing – Generative AI & Legal Research, Widener Univ. Law Library (Aug. 24, 2023), <https://libguides.law.widener.edu/c.php?g=1342893&p=9902106> (law library guidance for prompt composition).

19. Introduction to Legal Prompt Engineering, CaseMark Blog (Oct. 9, 2025), <https://www.casemark.com/post/introduction-to-legal-prompt-engineering> (blog primer for legal professionals at all tech skill levels).

20. Mastering AI Legal Prompts: Best Practices & Examples, The L Suite (Oct. 1, 2025), <https://www.lsuite.co/blog/mastering-ai-legal-prompts> (collection of prompt techniques and compliance issues with real scenarios).

## Appendix B

### Settlement Deductibility Memo – Produced by Microsoft Copilot 10/23/2025 Based on Prompt Example

#### Memorandum

To: Attorney/CPA Name

From: Legal Research Assistant

Re: Deductibility of Settlement Payment for Wrongful Termination under I.R.C. §162(a)

Date: October 23, 2025

#### Issue

Is a one-time settlement payment made by a small real estate brokerage to a former employee to resolve a wrongful termination claim, where the payment is compensatory (not punitive) and made to avoid litigation, deductible as an ordinary and necessary business expense under I.R.C. §162(a)?

#### Rule / Authority

I.R.C. §162(a) allows a deduction for "all the ordinary and necessary expenses paid or incurred during the taxable year in carrying on any trade or business."

(<https://www.law.cornell.edu/uscode/text/26/162>).

- Ordinary: Customary or usual in the taxpayer's trade or business.
- Necessary: Appropriate and helpful for the business.

#### Key Cases:

- Commissioner v. Tellier, 383 U.S. 687 (1966): Legal fees incurred in defense of business-related claims may be deductible if the origin of the claim is business-related. (<https://casetext.com/case/commissioner-v-tellier>)
- Raytheon Prod. Corp. v. Commissioner, 144 F.2d 110 (1st Cir. 1944): The "origin of the claim" test determines the character of settlement payments. (<https://casetext.com/case/raytheon-prod-corp-v-commr-of-internal-revenue>)

#### IRS Guidance:

- IRS Publication 535: Payments must be substantiated as business-related and not personal. (<https://www.irs.gov/publications/p535>)
- IRS.gov: Settlements for compensatory damages are generally deductible if incurred in the ordinary course of business. (<https://www.irs.gov/government-entities/tax->

implications-of-settlements-and-judgments)

Distinction:

- Business vs. Personal: Only expenses directly related to the business are deductible.
- Ordinary vs. Capital: Ordinary expenses are immediately deductible; capital expenses must be capitalized and depreciated.

### Application

Facts:

- Taxpayer: Small real estate brokerage.
- Payment: One-time settlement to a former employee for wrongful termination.
- Nature: Compensatory, not punitive; made to avoid litigation.

Analysis:

The settlement payment arises from a business dispute and is compensatory, intended to resolve a claim that originated from the taxpayer's business operations. Under the "origin of the claim" test, such payments are generally considered business expenses if they are ordinary and necessary.

Courts and the IRS distinguish between business and personal expenses by examining the purpose and context of the payment. Here, the payment is not for personal benefit but to resolve a business-related claim. It is not a capital expenditure, as it does not create or improve a long-term asset, but rather addresses a current business liability.

IRS guidance and case law support the deductibility of settlement payments for wrongful termination when they are compensatory and incurred in the ordinary course of business.

### Conclusion

Based on I.R.C. §162(a), relevant case law, and IRS guidance, a one-time compensatory settlement payment made by a real estate brokerage to resolve a wrongful termination claim is generally deductible as an ordinary and necessary business expense, provided it is substantiated as business-related and not personal or capital in nature.

Caveat: If the payment includes punitive damages or is for personal benefit, those portions are not deductible.

Uncertainty: If the facts or allocation of the payment are unclear, further analysis of the settlement agreement and underlying claims may be required.



## Appendix C: Sample of AI Client Disclosure

### Bar & Ethics Guidance (contains sample language/excerpts)

1. State Bar of Texas — Sample AI Disclosures for Attorney–Client Contracts (PDF)

Source: <https://www.texasbarpractice.com/wp-content/uploads/2025/05/Sample-AI-Disclosures-for-Attorney-Client-Contracts-final.pdf> (May 2025)

2. New York State Bar Association — Task Force Report (Appendix C: Sample Engagement Letter Provision)

Source: <https://nysba.org/wp-content/uploads/2022/03/2024-April-Report-and-Recommendations-of-the-Task-Force-on-Artificial-Intelligence.pdf> (April 6, 2024)

3. California — State Bar Practical Guidance on Generative AI (PDF)

Source: <https://www.calbar.ca.gov/Portals/0/documents/ethics/Generative-AI-Practical-Guidance.pdf> (Nov. 23, 2023 (posted Jan. 31, 2024))

4. Florida Bar — Ethics Opinion 24-1 (Generative AI)

Source: <https://www.floridabar.org/etopinions/opinion-24-1/> (Jan. 19, 2024)

5. Oregon State Bar — Formal Ethics Opinion 2025-205 (Artificial Intelligence Tools)

Source: [https://www.osbar.org/\\_docs/ethics/2025-205.pdf](https://www.osbar.org/_docs/ethics/2025-205.pdf) (March 2025)

6. North Carolina State Bar — 2024 Formal Ethics Opinion 1

Source: <https://www.ncbar.gov/for-lawyers/ethics/adopted-opinions/2024-formal-ethics-opinion-1/> (Adopted Nov. 1, 2024)

### Law Firm Sample Engagement Letters / Terms

1. Babst Calland — Standard Terms of Engagement

Source: <https://www.babstcalland.com/standard-terms-of-engagement/> (Updated Mar. 25, 2025)

2. Poyner Spruill — Standard Terms of Representation

Source: <https://www.poynerspruill.com/standard-terms-of-representation/> (Effective July 10, 2025)

3. Berger Singerman — Standard Terms of Engagement (PDF)

Source: [https://www.bergersingerman.com/info/standard-terms-of-engagement/pdf?last\\_mod=yes](https://www.bergersingerman.com/info/standard-terms-of-engagement/pdf?last_mod=yes) (Retrieved Oct. 24, 2025)

4. Koley Jessen — Standard Terms of Engagement (PDF)

Source:

<https://www.koleyjessen.com/assets/htmldocuments/KJ%20Standard%20Terms%20of%20Engagement%20Updated%202.25.25.pdf> (Updated Feb. 25, 2025)

5. Conyers — Standard Terms and Conditions of Engagement

Source: <https://www.conyers.com/legal/legal-notices/standard-terms-and-conditions-of-engagement-multi-jurisdictional/> (Accessed Oct. 24, 2025)

6. Cassels — Standard Terms for Client Engagements (PDF)

Source: [https://cassels.com/wp-content/uploads/Cassels\\_StandardTermsForClientEngagement.pdf](https://cassels.com/wp-content/uploads/Cassels_StandardTermsForClientEngagement.pdf) (Accessed Oct. 24, 2025)

## Appendix D: Resources -- Policies on Use of AI

(Includes Ideas for Nonlegal Businesses as Well as Law Firm Specific)

Clio (Legal Practice Software) – Law Firm AI Policy Template, Tips & Examples URL:  
<https://www.clio.com/resources/ai-for-lawyers/law-firm-ai-policy/>

Texas Bar Practice – Law Firm Acceptable Use Policy for Artificial Intelligence (AI) URL:  
<https://www.texasbarpractice.com/wp-content/uploads/2025/05/AI-Acceptable-Use-Policy-for-Law-Firms-5.27.25.pdf>

Darrow AI – Free AI Policy Template for Law Firms (Plus Examples) URL:  
<https://www.darrow.ai/resources/law-firm-ai-policy>

Fisher Phillips – Acceptable Use of Generative AI Tools Sample Policy PDF URL:  
<https://www.fisherphillips.com/a/web/du6wach1kmRuPCgDcMLJ5Z/ai-policy.pdf> ]

Justice At Work – Responsible AI Use Policy Outline for Law Firms PDF URL:  
<https://www.justiceatwork.com/wp-content/uploads/2024/08/Responsible-AI-Use-Policy-Template-.pdf>

Traverse Legal – AI Acceptable Use Policy URL:  
<https://www.traverselegal.com/blog/corporate-ai-acceptable-usage-policy/>

Inns of Court URL: <https://inns.innsofcourt.org/media/199583/sample-2-law-firm-ai-gai-use-guidance.pdf>

DISCO – How to Build a Defensible AI Policy for Your Law Firm URL:  
<https://csdisco.com/blog/how-to-build-a-defensible-ai-policy-for-your-law-firm>

CaseMark – URL: <https://www.casemark.com/post/crafting-an-ai-policy-for-your-law-firm-a-step-by-step-guide>