taxnotes federal

Volume 179, Number 9 ■ May 29, 2023

The Rise of Generative AI in Tax Research

by Benjamin Alarie, Kim Condon, Susan Massey, and Christopher Yan

Reprinted from *Tax Notes Federal*, May 29, 2023, p. 1509

The Rise of Generative AI in Tax Research

by Benjamin Alarie, Kim Condon, Susan Massey, and Christopher Yan

Benjamin Alarie

Kim Condon

Susan Massey

Christopher Yan

Benjamin Alarie is the Osler Chair in Business Law at the University of Toronto and the CEO of Blue J Legal Inc., Kim Condon is a senior legal research associate at Blue J Legal, Susan Massey is the vice president of legal research at Blue J Legal, and Christopher Yan is a senior legal research associate at Blue J Legal.

In this article, the authors consider the benefits, concerns, and areas of improvement that come to light when using large language model artificial intelligence chatbots to answer complex tax law questions.

I. Introduction

ChatGPT has exploded into the popular consciousness. Generative artificial intelligence, in the form of natural language chatbots,[1] is newly poised to significantly shift how tax professionals conduct research. These emerging advanced AI systems can analyze vast amounts of text and generate coherent, contextually relevant responses.

As various industries continue to adopt AI-powered tools, lawyers need to understand and adapt to these new technologies. As Jason Chen wrote last month in *Tax Notes*,[2] "It is only by incorporating these types of technological advances into daily business practices that modern tax professionals can gain valuable advantages over others in the increasingly competitive world of taxation."

In this article, we will consider the benefits of using generative AI, examine the concerns, and discuss ways that it can be improved. To illustrate these considerations, we submitted a series of tax law questions to three different generative AI models: ChatGPT (based on GPT-3.5), ChatGPT

[1] Another common term for the type of foundational models that support AI chatbots is "large language models," or LLMs.

[2] Jason Chen, "Is ChatGPT Harmful or Helpful to Tax Professionals?" *Tax Notes Federal*, Apr. 10, 2023, p. 189.

Table 1. 'Which Code section provides for a step-up in basis at death?'

GPT-3.5 (Default)	The Internal Revenue Code (IRC) section that provides for a step-up in basis at death is section 1014. This section provides that the basis of property acquired from a decedent is the fair market value of the property at the decedent's date of death. This means that if a person inherits property from a decedent, the person's basis in the property is the fair market value of the property at the time of the decedent's death, rather than the decedent's original cost or basis in the property. The IRS provides guidance on how to determine the fair market value of property at the decedent's date of death in Publication 551, Basis of Assets. This publication explains that the fair market value of property is the price at which the property would change hands between a willing buyer and a willing seller, neither being under any compulsion to buy or sell and both having reasonable knowledge of relevant facts. Here is a link to section 1014 of the Internal Revenue Code: https://www.law.cornell.edu/uscode/text/26/1014 Here is a link to IRS Publication 551, Basis of Assets: https://www.irs.gov/publications/p551

Plus (based on GPT-4), and Ask Blue J. The questions were selected to represent a variety of tax law areas and to present the types of research questions that a typical tax law practitioner might encounter in their day-to-day practice.

For this experiment we posed identical questions to ChatGPT and Ask Blue J and recorded their responses. We tested the questions against the GPT-3.5 and GPT-4 versions of ChatGPT. We primed each conversation with ChatGPT using the following instruction: "I want you to answer all questions as they apply in the context of U.S. federal income tax law. Please provide IRS resources along with your answers." This additional instruction was provided to ChatGPT only because Ask Blue J is already fine-tuned to focus on tax and to provide sources with its responses.[3] Providing the priming instruction to the GPT models ensures a fair comparison. Other than this introductory prompt, our questions to each chatbot were the same. The results highlight the importance of developers and users engaging with this technology thoughtfully when harnessing it for tax research purposes.

II. What Is Generative AI?

Put simply, generative AI is an algorithmic model that can create novel content based on the input it receives. Generative AI chatbots, such as Open AI's ChatGPT,[4] generate textual responses to user queries in a conversational, human-like manner. Development of a powerful generative AI algorithm requires a massive amount of training data to allow the model to learn the underlying structure and context of language. The largest generative AI models are trained on hundreds of billions of words of text.[5] Once trained, the chatbot can generate new content by mimicking the patterns it has learned. For example, when asked a question, it analyzes the words and phrases it has seen before and combines them to produce a coherent and contextually relevant answer. This ability to understand and generate original content potentially makes generative AI a game-changing technology for conducting research.

When used effectively, AI chatbots have the potential to alleviate many challenges faced by tax professionals, such as keeping up to date with changing rules and regulations, parsing complex legal documents, and effectively researching statutes and case law. The result is a streamlined research process and enhanced efficiency.

For example, Table 1 shows a response from ChatGPT to a simple question regarding the basis of property received from a decedent.

This answer provides a useful response for someone unfamiliar with the law. In this case, the answer, while basic, is accurate and provides links to the relevant code provision and IRS Publication 551.

[3] For this article, we have provided Ask Blue J's sources as footnotes.

[4] Other examples include Microsoft's Bing and Google's Bard.

[5] Aatish Bhatia, "Let Us Show You How GPT Works — Using Jane Austen," *The New York Times*, Apr. 27, 2023.

Table 2. 'What is section 1292?'

GPT-3.5 (Default)	Section 1292 of the Internal Revenue Code (IRC) pertains to the taxation of certain passive foreign investment companies (PFICs). Under this section, if a U.S. person owns shares in a PFIC, they may be subject to additional taxes and reporting requirements. Specifically, Section 1292 provides rules for determining the tax treatment of PFICs, which are defined as foreign corporations with at least 75% of their income derived from passive sources, such as investment income or rents. The rules are intended to prevent taxpayers from deferring taxation on income earned through PFICs. The tax treatment of PFICs can be complex, and taxpayers who own shares in a PFIC may be subject to additional taxes, such as the excess distribution tax, as well as reporting requirements, such as the annual filing of Form 8621. It is recommended that taxpayers consult with a qualified tax professional for guidance on the tax treatment of PFICs. Here is a link to Section 1292 of the Internal Revenue Code: https://www.law.cornell.edu/uscode/text/26/1292

Early-career lawyers and tax accountants can benefit from the use of AI research tools, enabling them to rapidly achieve a high level of competency. AI can also assist with issue-spotting, helping to catch potential mistakes or traps that might have gone unnoticed and acting as an additional set of eyes on a research problem. Further, the user-friendly natural language interface makes AI chatbots easy for most users to adopt, allowing legal professionals to conduct research without the need for specialized training in the platform. This intuitive interaction ensures that lawyers can seamlessly integrate AI chatbots into their daily workflows, ultimately enhancing the quality and speed of their legal research and decision-making.

III. Traps and Pitfalls

While undoubtedly a powerful tool, ChatGPT has significant limitations when used for tax law research. Some of the risks have been described by others.[6] The most common problems include ChatGPT's presenting inaccurate information, fabricating information, and not being transparent with the information's origin — all of which prevents its confident use in tax research.

A. False Information

ChatGPT's tendency to provide false information renders it untrustworthy.[7] Table 2 shows an example of an incorrect answer, which it offered confidently. For those unfamiliar with this area, section 1292 is not a real code provision.

Although section 1292 does not exist, ChatGPT has dutifully provided the user with a lengthy description of a different provision, which appears near section 1292 in the code. It also purports to provide a link to the text of section 1292 in a syntax that is consistent with how Cornell Law School provides access to provisions of the IRC. Unsurprisingly, because the provision does not exist, that link is broken.

We also presented the same question to Open AI's GPT-4 model, which is available with ChatGPT Plus. GPT-4 has stronger analytical abilities than ChatGPT, which runs on GPT-3.5. Table 3 shows its response.

As with the earlier version, GPT-4 provides a description of a nearby code provision, failing to recognize that section 1292 does not exist. In this case, it has provided a functional web link to Form 8621, which could be helpful to a professional researching passive foreign investment companies. These results demonstrate why it is important for practitioners to use these new AI tools with caution.

B. Hallucinations

Another concern with generative AI is that it is prone to fabricate results or "hallucinate" when it doesn't know how to answer a question. This presents a serious risk for the unwary researcher who may be led to believe, because of a chatbot's large database of information and confident tone, that its answers are accurate.

[6] *See* Chen, *supra* note 2; and Daniel Mayo, "Can I Replace My Tax Advisor With ChatGPT?" *Forbes*, Feb. 18, 2023.

[7] Blue J Legal, "GPT-4 Needs Tax Help: Better Call the Experts," Blue J Blog, Mar. 17, 2023.

Table 3. 'What is section 1292?'

GPT-4	Section 1292 of the U.S. Internal Revenue Code (IRC) pertains to the taxation of investments in Passive Foreign Investment Companies (PFICs). PFICs are foreign corporations with passive income that meets certain thresholds. U.S. shareholders of PFICs are subject to specific tax treatment and reporting requirements. Section 1292 specifically deals with the Qualified Electing Fund (QEF) election. A QEF election allows a U.S. shareholder to elect to include their pro-rata share of the PFIC's earnings as ordinary income and their pro-rata share of the PFIC's net capital gains as long-term capital gains. The election can help mitigate the potentially adverse tax consequences associated with PFIC investments. To make a QEF election, a U.S. shareholder must file Form 8621, "Information Return by a Shareholder of a Passive Foreign Investment Company or Qualified Electing Fund," along with their federal income tax return. Please note that tax laws and regulations are subject to change, so it is essential to consult a tax professional or the most recent version of the IRC for the latest information. Here is a resource to learn more about PFICs and Section 1292: U.S. Internal Revenue Service (IRS) — Form 8621 Instructions: https://www.irs.gov/pub/irs-pdf/i8621.pdf

Table 4. 'Examples of REIT prohibited transactions'

GPT-3.5 (Default)	A real estate investment trust (REIT) is a company that owns or operates income-generating real estate assets and passes on at least 90% of its taxable income to its shareholders. One of the requirements for maintaining REIT status is that the company must not engage in certain types of prohibited transactions. These include sales of property held primarily for sale to customers, sales of property that results in the REIT having more than 10% of its assets as non-qualifying securities or other non-real estate assets, and certain types of transactions with related parties. Private Letter Rulings (PLRs) are written decisions issued by the IRS in response to a taxpayer's request for guidance on a specific issue. While PLRs are only binding on the taxpayer who requested them, they can provide helpful insight into the IRS's thinking on a particular issue. Here are a few examples of REIT prohibited transactions with associated PLRs: 1. Sale of property held primarily for sale to customers: In PLR 201619009, the IRS considered whether a REIT's proposed sale of a parcel of land would be a prohibited transaction. The REIT had purchased the land with the intention of developing it into a residential subdivision for sale to customers. However, due to a change in market conditions, the REIT decided to sell the undeveloped land to a third party instead. The IRS determined that the sale would not be a prohibited transaction, as the REIT had held the land for more than a year and the decision to sell was made in the ordinary course of business. 2. Transactions with related parties: In PLR 201744002, the IRS considered whether a REIT's proposed lease of a property to a related party would be a prohibited transaction. The related party was a taxable REIT subsidiary (TRS) of the REIT, meaning it was a subsidiary that did not meet the requirements to be a REIT but was subject to certain limitations and restrictions. The IRS determined that the lease would not be a prohibited transaction, as it was entered into at arm's length and provided fair market value rent. 3. Sales of property resulting in more than 10% non-qualifying assets: In PLR 201840005, the IRS considered whether a REIT's proposed sale of a hotel would be a prohibited transaction. The REIT had determined that the sale would result in its non-qualifying assets exceeding the 10% limit. However, the IRS determined that the sale would not be a prohibited transaction, as the REIT had a reasonable belief that the sale would not result in its non-qualifying assets exceeding the limit. Please note that PLRs are specific to the individual taxpayer and should not be relied upon as a general rule or precedent. It is always important to consult with a qualified tax professional for guidance on your specific situation.

Table 5. 'Examples of REIT prohibited transactions'

<table>
<tr><td>GPT-4</td><td>
A Real Estate Investment Trust (REIT) is a company that owns, operates, or finances income-producing real estate. REITs are subject to specific tax rules and regulations, including restrictions on engaging in prohibited transactions. Prohibited transactions are those that involve the sale or disposition of property held primarily for sale to customers in the ordinary course of a trade or business. Engaging in prohibited transactions can result in a 100% tax penalty on the net income from such transactions.
Here are some examples of REIT prohibited transactions with corresponding Private Letter Rulings (PLRs) issued by the IRS as guidance:
1. PLR 201504008: This PLR dealt with a situation where a REIT planned to acquire and develop land for sale to a third-party builder. The IRS concluded that the proposed transaction would be considered a prohibited transaction, as the REIT intended to sell the property primarily to customers in the ordinary course of a trade or business.
Source: https://www.irs.gov/pub/irs-wd/201504008.pdf
2. PLR 201214003: In this PLR, a REIT acquired properties through foreclosure and intended to sell them as soon as reasonably possible. The IRS concluded that the REIT's intent to sell the properties primarily to customers in the ordinary course of a trade or business would constitute a prohibited transaction.
Source: https://www.irs.gov/pub/irs-wd/1214003.pdf
3. PLR 201423009: A REIT proposed to contribute certain properties to a taxable REIT subsidiary (TRS) and then sell the TRS's stock to a third party. The IRS determined that the transaction would be considered a prohibited transaction because the TRS was effectively a customer, and the property was sold in the ordinary course of a trade or business.
Source: https://www.irs.gov/pub/irs-wd/1423009.pdf
Please note that Private Letter Rulings (PLRs) are issued by the IRS in response to specific taxpayer requests and are based on the facts presented in those requests. PLRs may not be used or cited as precedent, but they can provide insight into the IRS's position on certain issues. Consult a tax professional to discuss specific situations and how they relate to REITs and prohibited transactions.
</td></tr>
</table>

Table 4 is an example of a hallucination. We received this answer from GPT-3.5 when looking for examples that would inform a user's understanding of prohibited transactions for real estate investment trusts.

After some lead-in language regarding REITs, ChatGPT provides the user with three private letter ruling citations, along with substantially detailed descriptions for each of them. While this response appears helpful at first glance, the examples are not what they seem. The first and third letter ruling citations offered are completely fabricated. Although the second example is a citation for a real private letter ruling (and even involves a REIT), it has no connection to the description provided by the chatbot. In short, these private letter rulings do not exist, and the descriptions merely reflect an attempt by the AI to provide a response that best matches the user's inquiry — all without regard to reality.

We also gave GPT-4 the opportunity to respond to the same inquiry (see Table 5).

GPT-4's examples are bogus as well, including the descriptions. The links are also hallucinations. The first two links direct the reader to unrelated private letter rulings, and the third link is broken.

By contrast, an AI model that is properly directed to retrieve only relevant information can provide a far more responsive and insightful answer (see Table 6).

Here, Ask Blue J can provide a response that includes references to relevant sources and a description of factors to consider when analyzing the application of this code provision. It also offers a substantial list of private letter rulings directly addressing inquiries about REIT prohibited transactions. These materials offer pertinent information to the researcher and will also provide inspiration for additional research questions to pose to the tool. Access to real, primary source material is crucial if AI is to serve as a reliable resource tool.

Table 6. 'Examples of REIT prohibited transactions'

Ask Blue J	Examples of REIT prohibited tax transactions for REITs include the sale or other disposition of property held primarily for sale to customers in the ordinary course of its trade or business, also known as "dealer property," as defined in Section 1221(1) of the Code. This is because Section 857(b)(6)(A) of the Code imposes a 100-percent tax on a REIT's net income from prohibited transactions, which are defined in Section 857(b)(6)(B)(iii) as the sale or other disposition of property described in Section 1221(a)(1) which is not foreclosure property. Other examples of prohibited tax transactions for REITs that are not excluded under section 857(b)(6)(C) include sales of property held primarily for sale to customers in the ordinary course of the REIT's trade or business, such as sales of condominium units or subdivided lots in a development project. Factors that may be considered in determining whether a taxpayer holds property primarily for sale to customers in the ordinary course of its trade or business include the nature and purpose of the acquisition of the property, the extent and nature of the taxpayer's efforts to sell the property, the number, extent, continuity, and substantiality of the sales, the extent of subdividing, developing, and advertising to increase sales, and the time and effort the taxpayer habitually devoted to the sales.[a]
[a]LTR 200824018; LTR 9041047; LTR 200701008; LTR 200702021; LTR 200728037; LTR 9123042; LTR 201844003; and LTR 201346005.	

C. Opacity Regarding Sources

Compounding these issues is the fact that ChatGPT does not reliably provide sources to back up its responses, making it difficult for users to verify the accuracy of the information provided. Thus, even when the answer is accurate, users may be unable to verify it. Further, although it has an extremely broad knowledge base, ChatGPT lacks sufficient in-depth knowledge of tax law to be useful to tax professionals for anything but the most basic research problems.[8]

In the example regarding REIT prohibited transactions, ChatGPT provided a basic definition of a REIT and explained that REITs are prohibited from engaging in specific types of activity. However, neither model provided any documentation that would lend credibility to its answer. The practitioner is left with a rudimentary and tentative overview and a phony list of examples in response to her question, which contributes little to the advancement of her research.

IV. Genesis of Ask Blue J

Understanding tax law can be challenging, especially when relying on language models such as ChatGPT, which have been trained on a wide range of sources. As noted, these language models can produce information that appears correct but is inaccurate because of limitations with how existing language models are trained.

To tackle these challenges, Ask Blue J incorporates tax expertise during the development of the tool by engaging in a thorough testing and feedback process incorporating close collaboration with in-house experts and industry leaders. The development process involves training the model on legal terms of art and ensuring that the model understands the language specific to the tax domain. Moreover, Ask Blue J continuously curates the information that is used to inform the answers, keeping up to date with the latest changes in tax law and ensuring that the information generated remains relevant and accurate.

While it may be tempting to input the entire body of tax law into a language model like ChatGPT before asking a question, ChatGPT and other language models have constraints that limit the amount of information that can be processed in a single query.[9] To overcome this limitation, Ask Blue J combs through the entire body of tax law in

[8]Even then, any decent tax professional will conduct separate and complete research on the topic, which necessarily involves identifying and analyzing primary sources.

[9]LLMs have token limits in a query because of computational constraints and efficiency. Tokens are a way for language systems to break down words, numbers, or symbols information into simpler, more manageable pieces that the system can understand and manipulate. Limiting the number of tokens ensures that the model processes and generates responses in a reasonable time, while managing hardware resources effectively.

Table 7. 'Which Code section provides for a step-up in basis at death?'

Ask Blue J	Section 1014 of the Internal Revenue Code provides for a step-up in basis at death. This means that the basis of property in the hands of a person acquiring the property from a decedent or to whom the property passed from a decedent shall be the fair market value of the property at the date of the decedent's death. This applies to property acquired by bequest, devise, or inheritance, or by the decedent's estate from the decedent, as well as property transferred by the decedent during his or her lifetime in trust to pay the income for life to or on the order or direction of the decedent, with the right reserved to the decedent at all times before his or her death to revoke the trust. However, there are certain circumstances in which this may not apply, such as appreciated property acquired by the decedent through gift within one year of the decedent's death, if such property passes, either directly or indirectly, from the donee-decedent to the original donor or the donor's spouse.[a]
[a]LTR 9308002; Rev. Rul. 84-139, 1984-2 C.B. 168; IRS Publication 4895, "Tax Treatment of Property Acquired From a Decedent Dying in 2010"; LTR 9026036; LTR 9321050; LTR 201544002; LTR 9218019; LTR 199939038; Rev. Rul. 78-213, 1978-1 C.B. 255.	

its carefully curated database of tax documents and surfaces the sources of information that are most relevant to a user's query. Ask Blue J then uses only that information to answer the question. This approach results in enhanced accuracy because the model's responses are based on information that has already been validated as primary sources of law or as reliable explanations.[10] By combining extractive AI, which identifies the most relevant information, with generative AI, which processes the information to formulate a response, this approach strikes a balance between the resource-intensive process of language models and the relevance of the results.

Table 7 shows the result when the basis step-up question from earlier is asked of this specially trained model.

Not only is this response more detailed than those produced by the other models, but it also highlights an important exception regarding appreciated property acquired by the decedent within one year of death and provides the user with helpful and accurate source information.[11] Each of the models was able to quickly locate the correct code provision (section 1014) even though the phrase "step-up in basis" can't be found in the section itself. AIs can identify the right section because their training data include texts that use similar language to refer to section 1014.

Also, by providing references to relevant sources and displaying crucial IRS documents and rulings directly in its platform, Ask Blue J assists tax practitioners in understanding how the step-up in basis rule has been interpreted by the IRS in various scenarios. The speed with which these resources are gathered for the user — mere seconds — significantly improves tax research efficiency.

As noted earlier, another important deficit in existing large language models is the prevalence of hallucination. This tendency can (and should) be curbed to ensure the accuracy of results being provided to users. Once that has been accomplished, this is what the response looks like to the query about nonexistent section 1292.

Table 8. 'What is section 1292?'

Ask Blue J	Unfortunately I am not able to provide a confident answer to your tax question in its current form. Please try again. Here are some suggestions: • Break your question down into smaller parts. • Rephrase your question. • Check our FAQ for prompting suggestions.

In contrast to GPT-3.5 and 4, Ask Blue J has been trained not to hallucinate (that is, fabricate) information when unable to provide a correct answer to a question. As the process of training and refining the model continues, the expectation is that when Ask Blue J is prompted with a question like this it will respond with additional information explaining why it can't provide an answer — because the code provision does not exist — and offer additional suggestions to enable the user to reframe the question for a better response.

[10]Such as legislative history and preambles to IRS regulations.

[11]Among the many relevant sources, Ask Blue J even supplies an interesting historical tidbit regarding special tax treatment that applied in 2010 (IRS Publication 4895, "Tax Treatment of Property Acquired From a Decedent Dying in 2010").

Table 9. 'What is a Granite Trust transaction?'

GPT-3.5 (Default)	I'm not aware of a tax term or transaction called a "Granite Trust transaction" in US Federal income tax law. It's possible that this term may refer to a specific tax strategy or arrangement that is not widely recognized or accepted by the IRS or the tax community. If you can provide more information or context about what you mean by "Granite Trust transaction," I may be able to provide a more specific answer. Alternatively, you may want to consult with a licensed tax professional or conduct further research to determine if this is a legitimate tax term or strategy. In general, it's important to be cautious of tax schemes or strategies that promise to significantly reduce or eliminate your tax liability, as these may be illegal or fraudulent. The IRS maintains a list of tax scams and schemes on its website, which you can review for more information: https://www.irs.gov/privacy-disclosure/report-phishing.
GPT-4	A Granite Trust transaction is not a term commonly used in the context of US Federal income tax law or IRS resources. It might be a term related to another area of law, finance, or a specific type of transaction. If you could provide more context or clarify your question, I'd be happy to help answer it or find relevant information for you. If you are looking for information about trusts in general in relation to tax law, you can refer to the following IRS resources: 1. Trusts — IRS Publication 559 (Survivors, Executors, and Administrators): https://www.irs.gov/pub/irs-pdf/p559.pdf 2. Filing requirements for trusts — IRS Instructions for Form 1041 (U.S. Income Tax Return for Estates and Trusts): https://www.irs.gov/pub/irs-pdf/i1041.pdf Please provide more context or specific questions, and I will do my best to provide helpful information.
Ask Blue J	A Granite Trust transaction refers to the sale of stock by a parent corporation to reduce its ownership in its subsidiary's stock below 80 percent prior to the liquidation of the subsidiary in order to recognize a loss. The court found that the sale of stock was a bona fide transaction and allowed the parent corporation to avoid the nonrecognition provisions of I.R.C. section 332. The case involved Granite Trust Company and its subsidiary, Granite Trust Building Corporation.

An additional advantage of this approach is that Ask Blue J will show its work by identifying the primary sources used in generating the response. This allows users to validate the accuracy of the information. As tax professionals, we should verify work and remain cautious when using emerging technologies that, while impressive, have significant limitations.

Finally, because Ask Blue J's answers are limited to being based on specific sources, it is more transparent about its capabilities. If the model does not have the information to provide an accurate answer to a question, Ask Blue J will simply inform you of this limitation. This transparency further helps tax professionals to better understand the model's boundaries, making it a valuable and trustworthy tool in navigating the complexities of tax law.

V. Additional Specialized Questions

The benefits of applying tax expertise when developing a specialized AI chatbot become even more apparent when questions are posed that require a deeper understanding of the underlying material. Some terminology and legal definitions have particular meanings in one professional context as compared with others. An AI model that is uniquely designed for a particular area, such as tax law, will have little difficulty identifying the appropriate meaning of concepts in that field, whereas a general model is likely to produce results reflecting the wide range of inputs it received from all over the internet. The following representative examples illustrate the difference.

A. Subchapter C Example

First, we asked all three chatbots about a type of transaction that subchapter C practitioners will recognize but is likely unfamiliar to others (see Table 9).

Surprisingly, both versions of ChatGPT lack enough tax knowledge to recognize the term "*Granite Trust*." It is a common term[12] used to refer to the type of transaction described by Ask Blue J — one in which ownership in a subsidiary is reduced to avoid application of section 332 nonrecognition treatment. GPT-3.5 provides no sources other than the link to the IRS phishing page. GPT-4 provides links to IRS publications for trusts and estates.

In addition to the case citation, Ask Blue J provides the user with FSA 1995-26 as a source. This field service advice sets forth a detailed discussion of *Granite Trust*,[13] related case law, related code provisions, and potential applications of the step transaction doctrine to this type of transaction.

B. Tax-Exempt Organizations Example

In this round, we put the AI's understanding of exempt organizations to the test, focusing on the boundaries of their political involvement. Confusion in this area causes some section 501(c) organizations to shy away from all forms of advocacy, even permissible ones, for fear of losing their exempt status (see Table 10).[14]

None of the chatbots provided an answer that was completely relevant and useful in this scenario. Of the three, GPT-3.5's answer was the least satisfactory. Although GPT-3.5 correctly stated that EOs are allowed to engage in limited lobbying activities, it then incorrectly stated that churches may determine their lobbying limits under the expenditure test. Compounding this error, GPT-3.5 cited paragraph 501(c)(3) for the expenditure test. The correct section is 501(h).

GPT-4's answer is slightly better. It correctly referred to subsection 501(h). However, GPT-4 neglected to point out that churches are classified as disqualified organizations and therefore are not eligible to elect to use the expenditure test.

Out of the three chatbots, Ask Blue J was the only one that mentioned that churches are obligated to use the substantial part test to measure their lobbying activities because they are disqualified from electing to use the expenditure test in subsection 501(h).[15] This is a critical error because a church that attempted to rely on the safe harbor of the expenditure test could risk losing its exempt status if the IRS later determined that a substantial part of its activities constituted attempts to influence legislation.

Ask Blue J's answer fell short, however, in its failure to recognize that the section 4911 tax on excess lobbying expenditures applies only to organizations that have made a 501(h) election. Because churches are disqualified organizations under 501(h)(5)(A), this section has no applicability to churches. This demonstrated difficulty in making the appropriate inference in this context underscores the importance of tax professionals' examining the relevant sources relied upon by an AI research tool before rendering advice to clients.

C. Reverse Acquisitions Example

For our final question to the chatbots, we chose the famously difficult-to-parse reverse acquisition provision from the consolidated return regulations (Table 11).

Both versions of ChatGPT describe a merger involving a public and private company, which seems to be a reference to an accounting concept, rather than a tax law provision.[16] After describing this accounting concept, the GPT-3.5 model gives a generic caution that the transaction may be complex and to check with a tax professional. This is not helpful. It offers as its references the name of an IRS page that the authors could not find, along with an Investopedia page describing reverse takeovers (also not the same as a reverse acquisition).

The GPT-4 model makes a basic reference to tax-free reorganizations as described in section 368, a link to the text of section 368, and a link to Publication 542, "Corporations." None of these materials give the user any hint of consolidated returns, let alone a framework for understanding the convoluted language of reg. section 1.1502-75(d)(3).

[12] A simple Google search provides links to various practitioner websites discussing the transaction.

[13] *Granite Trust Co. v. United States*, 238 F.2d 670 (1st Cir. 1956).

[14] *See* Fraser Nelson, David W. Brady, and Alana Conner Snibbe, "Learn to Love Lobbying," 5(2) *Stan. Soc. Innovation Rev.* 57-63 (2007).

[15] *See* section 501(h)(5)(A).

[16] International Financial Reporting Standards, "IFRS 3 Business Combinations" (Mar. 2013).

Table 10. 'Is a tax-exempt church allowed to engage in direct lobbying activities?'

GPT-3.5 (Default)	Yes, a tax-exempt church is allowed to engage in direct lobbying activities, but the amount of lobbying must be within certain limits to maintain its tax-exempt status. The limits on lobbying by a tax-exempt church are determined by either the "substantial part" test or the "expenditure test" under section 501(c)(3) of the Internal Revenue Code. Under the substantial part test, a church may not engage in substantial lobbying activities, which is generally defined as an activity that makes up a substantial part of the organization's overall activities. There is no specific percentage or dollar amount that constitutes a "substantial part," and each situation is determined based on the facts and circumstances. Under the expenditure test, a tax-exempt church may engage in direct lobbying activities, but it may not spend a substantial part of its funds on such activities. The expenditure test allows a church to spend up to a certain amount on lobbying activities each year, based on a sliding scale formula that takes into account the size of the organization's annual budget. It is important to note that if a church exceeds the limits on lobbying activities under either the substantial part test or the expenditure test, it may risk losing its tax-exempt status. Here is a link to IRS Publication 1828, Tax Guide for Churches and Religious Organizations, which provides more information on lobbying by churches: https://www.irs.gov/pub/irs-pdf/p1828.pdf
GPT-4	A tax-exempt church or religious organization, typically classified under Section 501(c)(3) of the Internal Revenue Code, is allowed to engage in some lobbying activities, but there are limits. Lobbying refers to attempting to influence legislation, either by directly contacting legislators and government officials or by urging members or the public to take action on a specific piece of legislation. According to the IRS, a 501(c)(3) organization may engage in lobbying activities as long as they do not constitute a "substantial part" of its overall activities. However, there is no clear-cut definition of "substantial part," and it may depend on various factors, such as the amount of time and resources dedicated to lobbying activities. If a church or religious organization engages in excessive lobbying activities, it risks losing its tax-exempt status. To provide some guidance and a safe harbor for lobbying activities, 501(c)(3) organizations may elect to follow the expenditure test under Section 501(h) by filing Form 5768, Election/Revocation of Election by an Eligible Section 501(c)(3) Organization to Make Expenditures to Influence Legislation. It is important to note that 501(c)(3) organizations, including churches, are absolutely prohibited from engaging in political campaign activities, such as supporting or opposing candidates for public office. For more information on lobbying and political activities by 501(c)(3) organizations, you can refer to the following IRS resources: Lobbying: https://www.irs.gov/charities-non-profits/lobbying The Restriction of Political Campaign Intervention by Section 501(c)(3) Tax-Exempt Organizations: https://www.irs.gov/charities-non-profits/charitable-organizations/the-restriction-of-political-campaign-intervention-by-section-501c3-tax-exempt-organizations Form 5768, Election/Revocation of Election by an Eligible Section 501(c)(3) Organization to Make Expenditures to Influence Legislation: https://www.irs.gov/forms-pubs/about-form-5768
Ask Blue J	Yes, a tax exempt church is allowed to engage in direct lobbying activities to a limited extent. However, if the church is an electing public charity, it may be subject to the section 4911 tax on excessive lobbying expenditures. The substantial part test is used to determine if a church's or religious organization's attempts to influence legislation constitute a substantial part of its overall activities. Churches must use the substantial part test since they aren't eligible to use the expenditure test. Churches can engage in limited lobbying and take positions on political issues, but cannot participate or interfere in a political campaign on behalf or in opposition to any candidate for public office. If religious believers feel that their faith calls them to do more than that, then they have the option of organizing a section 501(c)(4) or organizing a section 527.[a]

[a]T.D. 8308; Publication 1828, "Tax Guide for Churches and Religious Organizations"; edited transcript of the July 31, 1998, American Bar Association Exempt Organizations Committee Meeting.

Table 11. 'What is a reverse acquisition?'

GPT-3.5 (Default)	A reverse acquisition, also known as a reverse merger or reverse takeover, is a type of business transaction where a private company acquires a public company in order to become publicly traded without going through the traditional initial public offering (IPO) process. In a reverse acquisition, the private company typically buys enough shares of the public company to gain control, and then merges its operations into the public company. As a result of the merger, the private company becomes a subsidiary of the public company, and the public company's shareholders become the new owners of the combined entity. Reverse acquisitions can be complex transactions and may have significant tax implications. For example, the tax treatment of the transaction may depend on whether the merger is structured as a stock acquisition or an asset acquisition. Additionally, the tax consequences may differ depending on whether the public company is a C corporation or an S corporation. It is important for companies considering a reverse acquisition to consult with a tax professional to fully understand the tax implications of the transaction. Sources: — IRS: Tax Consequences of a Reverse Acquisition — Investopedia: Reverse Takeover (RTO)
GPT-4	A reverse acquisition, also known as a reverse merger or reverse takeover, is a transaction in which a private company acquires a public company, effectively becoming a public company without going through the traditional initial public offering (IPO) process. In this transaction, the private company's shareholders exchange their shares for a controlling interest in the public company. As a result, the private company's management takes control of the public company, and the private company becomes the operating entity of the combined companies. In the context of US federal income tax law, the tax implications of a reverse acquisition may vary depending on the structure of the transaction and the companies involved. The Internal Revenue Service (IRS) might treat the transaction as a tax-free reorganization under certain conditions, as specified in the Internal Revenue Code (IRC) Section 368. If the transaction doesn't meet the requirements for a tax-free reorganization, it may be treated as a taxable transaction, with potential tax consequences for the shareholders. For more information on reverse acquisitions and their tax implications, you may want to consult the following resources: 1. Internal Revenue Code Section 368: https://www.law.cornell.edu/uscode/text/26/368 2. IRS Publication 542: Corporations: https://www.irs.gov/pub/irs-pdf/p542.pdf Please note that my knowledge is based on information available up to September 2021, and tax laws may have changed since then. It's essential to consult with a tax professional to obtain the most current and accurate information related to your specific situation.
Ask Blue J	A reverse acquisition is a transaction in which a corporation acquires substantially all the assets of another corporation in exchange for stock, and the stockholders of the second corporation subsequently own more than 50 percent of the fair market value of the outstanding stock of the first corporation. This definition is provided in Section 1.1502-75(d)(3) of the regulations. The reverse acquisition rule treats the acquired larger group as continuing and the smaller acquiring group as terminating. The tax consequences of a reverse acquisition may vary depending on the specific circumstances of the transaction. Rev. Rul. 72-322 holds that the reverse acquisition rule may be applicable even though the corporations in a reorganization were not members of an affiliated group prior to the transaction.

Ask Blue J provides an overview of the concept of a reverse acquisition, along with the references to reg. section 1.1502-75(d)(3) and Rev. Rul. 72-322, 1972-1 C.B. 287, as well as the documents for FSA 1995-90 and several letter rulings and technical advice memoranda raising issues concerning reverse acquisitions.[17] All these documents provide additional, relevant discussion of reverse

[17]LTR 8630033; LTR 8829063; FSA 1995-90; LTR 8534027; LTR 8619004; LTR 8628006; LTR 8522014; LTR 9351002; LTR 9036011; LTR 8510096.

acquisitions for a practitioner gaining familiarity with the provision.

VI. Limitations of Current Approaches

In the pursuit of leveraging advanced language models for tax law applications, it is important to acknowledge some limitations of the current approach. One limitation is that Ask Blue J's responses are confined to the tax-specific data set it has been trained on, rather than drawing on the broader general knowledge base inherent to large language models. This presents several key limitations, which create exciting challenges and opportunities to continue pushing the boundaries.

A. Complex Interplay of Tax Law Sections

The relationship between different sections of the code can be inherently complex. At the same time, the interaction of the different sections is not always readily apparent, with provisions sometimes working in harmony or conflict. Some provisions may expand or provide definitions and context to other sections, while other provisions will invalidate or override a general rule.

The team behind Ask Blue J is actively addressing this challenge by exploring the following approaches and considerations:

- *Improving relevant document retrieval:* Enhancing the capabilities of large language models to retrieve relevant documents more accurately, ensuring that the most pertinent information is considered when generating responses.
- *Advancing data processing and deduction abilities:* Developing and combining different models' abilities to process data more efficiently and make better deductions, facilitating a deeper understanding of the complex relationships between different tax law provisions.
- *Incorporating curated secondary sources:* Introducing carefully curated secondary sources into the knowledge base, allowing Ask Blue J to learn from a wider range of tax law applications and strategies.
- *Learning from examples in secondary sources:* Analyzing how examples in literature published by practitioners have combined separate provisions and assessed their consequences, allowing our models to eventually make similar deductions and suggest novel strategies.

As Ask Blue J's understanding of the practical application of different provisions expands, practitioners may be encouraged to rethink the way secondary sources are used in tax law, ultimately leading to improved outcomes.

B. The Evolving Nature of Tax Law

The second limitation stems from the nonstatic nature of tax law. Unlike fixed documents such as contracts or reports, the corpus of tax law evolves over time.

The following approaches present possible solutions to this challenge:

- *Obsolete document detection:* Implementing mechanisms to identify and exclude obsolete documents from the knowledge base, ensuring that outdated information does not inform the answers generated.
- *Timeline tracking and anchoring:* Keeping track of the timeline and anchoring each document to a point in time, allowing the model to account for changes in the law when generating responses.
- *Continuous curation and maintenance:* Regularly updating the knowledge base with new information and curating the content to ensure accuracy and relevance.
- *Automated updating processes:* Developing processes to automate the updating of the knowledge base, minimizing manual intervention and ensuring that the information remains up to date and reliable.
- *Tailored user experience:* Creating a personalized experience for users by considering their specific situations and the relevant points in time. For example, a document that references law that is not current may still be useful for someone filing for a previous tax year. By anchoring the relevance of the law to specific points in time, Ask Blue J can help users understand what the law was at that point and customize the information to their unique timeline.

By maintaining an ongoing commitment to thoughtful design, analysis, tagging, and processing of documents, the Ask Blue J team strives to ensure the model's responses are relevant and accurate despite the continuously changing nature of tax law.

C. Varied Nature of Tax Law Documents

The third drawback concerns the differential treatment of various types of tax law documents. Text from the code, rulings, cases, and secondary sources each possess unique characteristics and should be treated accordingly. To address this issue, Ask Blue J is exploring several approaches and considerations:

- *Source-specific weighting:* Assigning priorities to different types of sources based on their authority and relevance. This approach ensures that the most authoritative sources are prioritized, while still considering other relevant information.
- *Contextual understanding:* Enhancing the model's ability to recognize and understand the context in which a particular source is relevant. This includes understanding the specific legal issue being addressed and the hierarchy of the sources of law, enabling the model to generate more nuanced and accurate responses.
- *Source limitations:* Clearly identifying and communicating the limitations of each source, allowing users to make informed decisions about the reliability and applicability of the information provided.
- *Source integration:* Developing methods to intelligently combine information from various sources, such as rulings and cases, to provide a comprehensive understanding of the tax law issue at hand.

By diligently addressing each type of document with special attention and dealing with how the sources of law interact while prioritizing them and drawing attention to each source's limitations, Ask Blue J continues to refine and enhance its capabilities. As we progress, these challenges present valuable opportunities to advance the capabilities of language models in the complex domain of tax law.

D. Now Limited to 'Ask' Functionality

The current version of Ask Blue J primarily focuses on its "Ask" functionality; that is, the model provides typed responses to questions. However, the development team envisions expanding its capabilities in future iterations to further leverage the generative abilities of the latest technology. Some potential features of future versions include:

- *Memo generation:* Using the generative capabilities of language models to automatically produce memos that summarize and analyze relevant sources for specific tax law issues.
- *Assisted form completion:* Identifying the issues in a given tax scenario and guiding users through the process of completing the appropriate forms with the correct information, making the filing process more efficient.
- *Document drafting:* Streamlining the drafting of tax-related documents, such as contracts or agreements, by generating content based on the user's input and the applicable tax law provisions.
- *Conversational history:* Retaining a history of user interactions, thus allowing for a more seamless and personalized experience when addressing follow-up questions or providing additional context.
- *Client document information retrieval:* Enhancing the model's capability to quickly and accurately extract relevant information hidden within a client's set of documents. This feature would help users uncover necessary information to support tax positions.

By incorporating these features in future versions, Ask Blue J intends to become an even more powerful and comprehensive tool for tax professionals, further demonstrating the potential of advanced language models.

VII. Conclusion

AI has the potential to revolutionize tax research when employed responsibly. However, the inclusion of inaccurate or incomplete data when training AI models directly implicates the accuracy of the information they generate, which in turn could lead to faulty legal conclusions. This danger underscores the importance of using high-quality input when designing and building these tools.

Moreover, to provide tax practitioners with the power to confirm the accuracy of the information provided and to gain confidence in the model, it is crucial that the AI tools reveal the source documents used in drafting the response. While developers of AI tools can implement effective safeguards to control and refine the tools' output, users must recognize their integral role in the development process, whereby they can validate the accuracy of the information or reject it.

New AI models offer the promise of cutting-edge research tools to practitioners, which will allow them to rapidly absorb a new area of tax law, identify relevant resources, and formulate the structure and legal descriptions for various written communications. As this groundbreaking technology is still in its early stages, leaders in the field continue to study its limitations and devise ways to address them to harness its capacity to the fullest in the realm of tax law. ■