Strategic Perspectives

Exploring the Use of AI in Tax Law at the ABA Tax Section's 2024 May Tax Meeting

By Anne H. Gibson, J.D., LL.M.

As people grow more comfortable with the concept of artificial intelligence ("AI"), tax lawyers are exploring the potential uses it may have in their work. The American Bar Association Tax Section, 2024 May Tax Meeting in Washington, D.C., May 2-4, included a total of four panels on AI and tax law. Panelists in two of those sessions explored potential opportunities for leveraging AI capabilities in both the private and public sectors.

The panelists did not just consider generative AI ("Gen AI"), such as OpenAI's ChatGPT, but also looked at other AI tools, such as those capable of ingesting and analyzing large data sets. In particular, the panelists noted that public and private sector actors in the tax field are already using machine learning, generative AI, large language models, and predictive analytics. Over the course of the panels, they explored the types of AI best suited for particular tasks, and the tasks that would be best completed by humans.

Public and Private Sector Uses of AI in Tax Law

Two panels at the conference explored current and potential AI uses in the tax law field in detail. These were "Making Tax Policy in an Era of Artificial Intelligence – Can It Survive the Robots?", which focused on tax policy and the regulations process, and "Cross-Examining AI: The Future of Tax Litigation?", which focused on tax controversies and litigation.

In the regulatory and policy area, panelists saw a potential for the use of AI in the notice-and-comment process for regulations. They discussed using AI to assist in writing comments on proposed regulations, for example, and explored how Treasury might use AI to address comments received, or to weed out AI-generated "spam" comments.

> *"AI could help to make the process more democratic and might actually eliminate some of the inherent biases in the notice-and-comment process."*

In the area of tax preparation and litigation-focused tasks, the panelists discussed how AI could be used to perform the data gathering and analysis that could support a position taken on audit. They also discussed how the IRS might use AI to gather and analyze data on relevant taxpayers, or to inform who was chosen for audit.

The panels also briefly touched on the many other ways that AI tools can be used, for example, in AI-augmented search for research, in writing first drafts of emails or memos, or in administrative tasks, such as keeping track of hourly billing or human resources issues.

Uses of AI in the Regulatory Area

The panel "Making Tax Policy in an Era of Artificial Intelligence – Can It Survive the Robots?", moderated by Jeremiah Coder, Deputy Director, International Tax Policy, Revenue Jersey, with panelists Macdonald "Mac" Norman, Todd Welty, P.C.; Adrienne E. Griffin, Branch Chief, Procedure & Administration Branch 2, IRS Office of Chief Counsel; Alexandra D. Minkovich, Baker & McKenzie LLP; and Varun Jain, Senior Counselor, Office of Information and Regulatory Affairs (OIRA), Office of Management and Budget (OMB), explored the possibilities for using AI in the regulations process. In the tax field, regulations are first issued as proposed regulations, and after a comment period, the regulations are revised and issued as final regulations. The panel explored how AI could help with that process.

Private sector. The panelists explored the possibility of using AI to help lawyers craft comments on proposed regulations. When the Department of Treasury and the IRS are reviewing comments made on proposed regulations prior to drafting final regulations, they are required to consider and respond to significant or substantive comments. Certain requirements set out in case law must be met in order for a comment to qualify as significant. These

©2024 CCH Incorporated and its affiliates. All rights reserved. Reprinted with permission from CCH Incorporated.

requirements could inform the prompt AI is given when asked to draft a comment.

In order to test whether AI could draft a comment meeting these requirements, Coder prepared a case study in which he asked ChatGPT to write a substantive comment after giving it a set of proposed regulations and the publicly facing website of a company that might be affected by those proposed regulations. The AI was instructed to write a comment addressing that company's potential concerns.

ChatGPT did produce a document, but it was not on par with what a law firm would produce, the panelists agreed. Despite that, Minkovich felt that the document was not useless and could potentially save time if used as a starting point. She noted that the AI spotted important issues fairly well and thus was able to provide a list of potential issues to consider when writing the comment. She noted, however, that its recommendations probably did not meet the requirements for a significant comment.

Jain also noted that, from the agency's point of view, specific details in comments are very helpful. The panelists agreed that ChatGPT does not yet seem able to provide that level of detail.

Additionally, the panelists discussed how using AI might open up the comment writing area to even more members of the public. Usually, proposed regulations are themselves so opaque to lay people that they are not able to comment on them. Generally it is only large companies or interest groups with capable lawyers that submit comments on proposed regulations. However, if a lay person could use an AI program to summarize the Notice of Proposed Rulemaking in plain language, that might allow people to respond to the proposed regulations who otherwise would not have.

AI could also help such taxpayers navigate the comment submission process, and even draft their comments (although, as noted above, there are currently limitations on ChatGPT's ability to write such comments). In this way, AI could help to make the process more democratic and might actually eliminate some of the inherent biases in the notice-and-comment process.

> *"AI can be like an overeager intern in terms of quality. It can answer questions quickly, but they will not always be the best answers."*

Public sector. In examining how AI could be used by government actors, Jain noted that the public sector is already subject to certain guardrails. One is the Executive Order on Safe, Secure, and Trustworthy Artificial Intelligence, issued by President Biden in October 2023. Another is the memo "Advancing Governance, Innovation, and Risk Management for Agency Use of Artificial Intelligence," OMB M-24-10, issued in March 2024. These documents provide the first guidance on AI for U.S. government agencies. Their goals include strengthening AI governance and determining how the use of AI might impact the rights of members of the public. They also address the coordination of the use of AI across agencies. A hope is that this guidance will provide greater transparency by informing the public about how the government is using AI currently and how it plans to use it in the future.

Jain emphasized that the Office of Information and Regulatory Affairs at OMB is tasked with facilitating interagency regulatory issues. This office is directed to improve and modernize the rulemaking process, and engaging with AI's potential could be a part of this. OIRA members are attempting to leverage AI to improve the comment process.

The panelists first looked at the possibility of using AI to summarize the comments submitted by the public, which would be especially useful in situations in which hundreds of lengthy comments

AI Risk Factors

1. **Accountability –** Know who is responsible if AI makes a mistake or someone uses it inappropriately, whether that is the lawyer or the provider of the tool.
2. **"Explain-ability" –** Understand and be able to explain how the AI has reached its results.
3. **Fairness –** Be aware that large language model AI tools potentially have biases built into them based on their underlying training material and work to address that.
4. **Accuracy –** Protect against inaccuracies and recognize hallucinations when they occur.

©2024 CCH Incorporated and its affiliates. All rights reserved. Reprinted with permission from CCH Incorporated.

are submitted. It could also be used to identify so-called "spam" in the form of large numbers of irrelevant AI-generated comments that are sometimes submitted by people trying to disrupt the system.

Another possibility suggested was that AI could be asked to determine which of the submitted comments reached the level of a significant comment and needed to be addressed in the final regulations. Coder provided a case study, using a real set of publicly available comments on a set of proposed regulations, and asked ChatGPT to summarize the comments, determine which were significant, and then draft a preamble to final regulations on that basis.

The panelists found many issues with the outcome of this case study. Griffin noted that the AI included the names of commenters in its draft preamble, which it should not do. Additionally, the AI did not seem to be able to identify significant comments and instead made changes to the proposed regulations based on every comment. Additionally, the preamble was not nearly as detailed as it should have been.

Jain noted that there are a lot of steps in the normal process of drafting final regulations and writing a preamble. Different government stakeholders exchange drafts and work on the various sections piece by piece, with a lot of back-and-forth between them. Consequently, a preamble would never be written in one shot, and a lot of deliberation and discussion is necessary. This raised doubts among the panelists that AI could be used to draft this kind of document.

Litigation-Focused Uses of AI

In the panel "Cross-Examining AI: The Future of Tax Litigation?", panelists Reza Rashidi, Director, Research, Applied Analytics and Statistics Division, IRS Data Management Division; Paul Butler, Associate Chief Counsel (Procedure & Administration), IRS Office of Chief Counsel; Tom Greenaway, Principal, Tax Controversy Services, KPMG LLP; and moderator Hon. Ronald L. Buch, Judge, United States Tax Court, discussed possibilities for the use of AI in litigation-focused tasks.

> *"[Quality and factual reviews] require careful reasoning and detailed understanding of the relevant context, things AI is not currently well-equipped for."*

Private sector. The panelists agreed that this area is moving incredibly fast. In the private sector, companies are making huge investments in AI and adoption is high. Greenway mentioned some of the findings of KPMG's report, "Tax Reimagined 2023: Perspectives from the C-suite," a survey of 500 executives from U.S. firms earning over $1 billion annually. The results showed that 59% of the business leaders surveyed said that their companies are using AI and the remaining 41% all said they were interested in doing so.

Companies are using different types of AI for various applications, including for the automation of basic tasks, data analysis and research, risk monitoring and analysis, and strategic planning and communications. However, Greenway noted that AI can sometimes be like an overeager intern in terms of quality. It can answer questions quickly, but they will not always be the best answers. Usually, a lot of revision and oversight is needed.

Greenway went on to discuss a case study regarding the research and experimentation tax credit. There are multiple phases to preparing for a client to take this credit: (1) gather source data, (2) conduct a business component analysis, (3) perform a qualification analysis, (4) conduct a subject matter review, (5) conduct KPMG's review, and finally (6) conduct the client's review.

The three initial stages of the process involve the collection and analysis of large amounts of documents. These are very time-consuming tasks and it can be difficult for humans to maintain their focus and give their best work over the entire course of the process. On the other hand, certain machine-learning AI tools are now capable of ingesting huge amounts of data and performing consistent analyses on them. Consequently, these types of tasks are good candidates for AI.

The last three steps, the quality and factual reviews, are best suited for humans to perform. They require careful reasoning and detailed understanding of the relevant context, things AI is not currently well-equipped for.

The panelists also considered how and even if AI-generated evidence could be used at trial. For example, if AI conducted the initial gathering and filtering of source data, how would that be presented as evidence? Judge Buch stated that practitioners can look to existing rules of evidence, for example, rules related to purpose, summaries, experts, and the rule against hearsay. The employee who gave the AI its prompts and provided it with the relevant documents could testify, for example, and explain how the AI was instructed to come to its conclusions . This would not need to be different from other instances in which an expert explains how information was obtained or provides an expert report to explain how conclusions were reached.

©2024 CCH Incorporated and its affiliates. All rights reserved. Reprinted with permission from CCH Incorporated.

Public sector. Rashidi explained some ways in which the IRS uses AI. He gave the example of the Large Partnership Compliance (LPC) program, which uses AI in its work. The Data Solutions team supports LPC by using machine learning to discover anomalies that indicate possible noncompliance in the data available on relevant large partnership taxpayers. Rashidi noted that not all anomalies result from noncompliance and that it is important to keep that in mind. The machine learning tool identifies possible noncompliance by looking for these anomalies and the cases are then passed on to classifications teams for human decision-making on whether to pursue an audit or other action. The results of their analyses are also passed back to the model, to help improve it.

Other IRS programs are also using various kinds of AI and this is being communicated to practitioners in IRS Announcements, which Rashidi urged attendees to look out for.

Risks

Both panels discussed the risks of using AI. In "Making Tax Policy," the panelists pointed out that AI has many constraints, including the basis for the underlying model as well as the sources from which the current analysis is drawn, the validity of the training materials, the potential for "hallucinations" (the phenomenon of a Gen AI tool answering a question with a false but realistic sounding answer), and the importance of addressing the inherent bias it might have.

Griffin pointed out that specific restrictions under I.R.C. §6103 related to confidentiality govern the IRS's use of tax return data and taxpayer information. Consequently, when the IRS and Treasury are working with taxpayer data, they must determine what can be used, and whether permission would be required for that use.

The panelists also considered whether a closed loop system, meaning an AI tool that was not open but rather was confined to use by a single organization, could satisfy such confidentiality requirements. The panelists pointed out that these I.R.C. §6103 restrictions are one of the reasons why the IRS always creates its own software, despite the time and expense, rather than using off the shelf software from private companies. Essentially, the IRS has already been limiting itself to closed loop systems due to confidentiality concerns, and this approach will probably be necessary in the use of AI as well.

> *"We do not necessarily have to invent a whole new set of rules to use AI-generated analyses in tax law."*

Griffin also discussed the ethical dilemmas presented by AI tools. The ABA Model Rules of Professional Conduct impose general requirements on attorneys in the practice of law. These general rules would require any attorney using an AI tool to check the AI's work, recognize and correct for bias, learn how to use AI properly, and ensure that confidentiality was maintained.

In "Cross-Examining AI," the panelists discussed four major risk factors of AI: (1) accountability, in other words, who is responsible if AI makes a mistake or someone uses it inappropriately, (2) "explain-ability," that is, the ability of the user to actually understand and explain how the AI has reached its result, (3) fairness, due to the fact that large language model AI tools potentially have biases built in to them based on their underlying training material, and (4) accuracy, in particular protecting against hallucinations and recognizing when they occur.

Moving Forward with AI

There were several key takeaways from the sessions. One was that, no matter one's feelings or misgivings about the use of AI, it is already being used in the tax field, and that usage will only grow. Rather than resisting it, determining how to use it safely and effectively is the only path forward.

Another takeaway was the importance of keeping the rules of evidence in mind. Judge Buch pointed out that these can actually answer many AI-related questions. We do not necessarily have to invent a whole new set of rules to use AI-generated analyses in tax law. Rather, we can apply existing rules in a sensible manner to the products of AI tools.

Key Takeaways

- AI tools can be helpful with tasks that require consistent analyses of large data sets, including document review and identifying patterns.
- Tasks that require detailed understanding of the facts and context are best suited for humans, not AI.
- Existing rules of evidence and rules of professional conduct can answer many of the issues around AI. A new set of AI-specific rules is likely not needed.
- AI is already being used widely in tax law. The best path forward is to learn how to use it safely and effectively.

©2024 CCH Incorporated and its affiliates. All rights reserved. Reprinted with permission from CCH Incorporated.