Using AI in your Tax Practice: Notes from IFA Canada's 2024 Seminar and Hands-on Workshop

By Anthony Tahan, Louis-Alexandre Cabanel, and Lucila Paglieri and with assistance from Laura Gheorghiu

This article presents a summary of the discussions held at the IFA Canada seminar and hands-on workshop title "Using AI in your Tax Practice" which took place in Montreal, Canada, on Friday August 16th, 2024. The goal of the seminar and workshop was to introduce tax professionals and in-house counsel to Artificial Intelligence ("AI") solutions applicable to their practices and explore the best methods for using the technology ethically to meet their own or their clients'needs. This paper summarizes the research and opinions presented by the speakers and does not reflect the authors' individual opinions

The seminar speakers were Abdi Aidid, Professor at the University of Toronto, Faculty of Law; Kyle Bunnell, former Director of Tax, Hootsuite; Shashi Kara, Principal and Chief Technologist, Sente Advisors; Denis Keseris, Partner and Patent agent, Gowling WLG (Canada) LLP ("Gowling WLG"); Ara Stepanyan, Principal, the Brattle Group; and Nicolas Vermeys, Director of the Centre de recherche en droit public (CRDP), Associate Director of the Cyberjustice Laboratory, and Professor at the Faculty of law, Université de Montréal.

The seminar was moderated by Laura Gheorghiu, partner at Gowling WLG and Second Vice-President IFA Canada, and Josephine L (Josie) Scalia, International President of the Tax Executives Institute, Vice President, Global Taxation, CAE.

The authors of this article are students from different universities across the Americas. Anthony Tahan is a Tax LL. M. student at HEC Montreal (Université de Montréal), Louis-Alexandre Cabanel is a J.D. student at Stetson University College of Law and Lucila Paglieri is a Tax LL. M. student at Torcuato Di Tella University.

The authors are grateful to the main sponsor of the event, IFA Canada, for making this seminar and subsequent article possible. They would also like to thank the seminar collaborators: Tax Executive Institute, HEC Montréal (Université de Montréal), Stetson University College of Law, and Gowling WLG.

Part I of this article describes the technology driving AI, focusing particularly on Large Language Models ("LLMs"). Part II of this article delves into the ethical considerations for tax practitioners using AI.

Parts III and IV of this article take a more practical approach. Part III aims to define "prompt engineering" — the process of crafting specific and effective queries or commands to guide AI models — and provides best practices for creating effective prompts. Part IV describes a case study that was conducted after the IFA seminar, analysing how tax practitioners can use AI in practice.

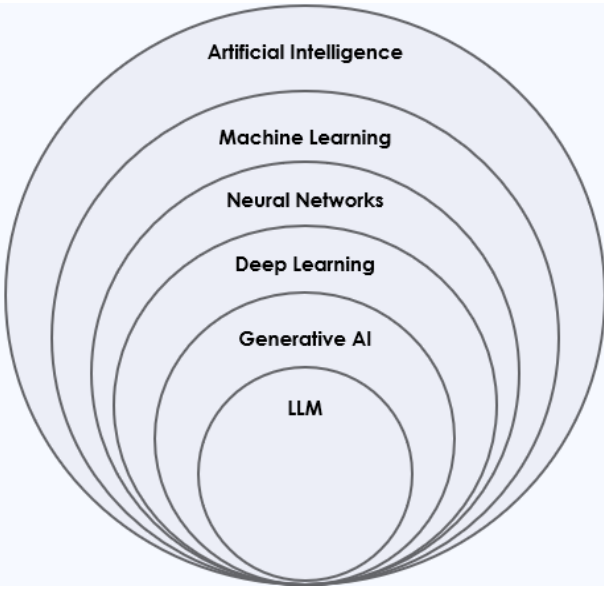
Part I: Understanding Current AI Applications for Tax Practitioners

This part aims first to describe what is AI, how it is structured, and how it works, focusing on LLMs. Second, we will discuss how LLMs apply specifically to tax, describing their strengths, weaknesses, and the potential use by tax partitioners of AI as an intelligence amplifier.

Understanding AI: the Technology and Function Behind LLMs

AI is a technology that enables computers to simulate intelligence and problem-solving capabilities. While there are a number of different AI technologies that serve a variety of functions, the most useful and prevalent for tax professionals are LLMs. To understand LLMs, it is important to first grasp the underlying AI technologies.

Under the umbrella of AI is machine learning; a type of AI that uses data and algorithms to imitate the way humans learn. Within machine learning are neural networks; mathematical models that mimic the way the human brain processes information, simulating the interaction of biological neurons to make decisions. A direct subset of neural networks is deep learning which uses multilayered neural networks to simulate the complex decision-making power of the human brain. Generative AI, in turn, is a collection of deep learning models specialized in generating text, images, and other content based on the data they were trained on. Generative AI does not imagine in the human sense; it generates new content based on patterns that it learned during training.



This brings us to LLMs, a type of generative AI that uses deep learning techniques and large data sets to understand, summarize, generate, and predict new text-based content. Within these models, neurons act as processing units organized into multiple layers to learn complex patterns and relationships in language. Each neuron assigns "weights" to different aspects of the input data, highlighting important features and reducing emphasis on others, while "biases"

adjust these weights to fine-tune the model's response. Together, weights and biases function as parameters that the LLM learned in training to enable it to generate the most probabilistic responses possible. The connections between neurons allow information to flow through the model, with each neuron in a layer linking to multiple neurons in the next layer, enabling the model to pass processed information forward until the most relevant output is generated. By refining these weights and biases across millions of parameters, LLMs capture nuanced language patterns, producing responses that are both syntactically accurate and contextually relevant to the input the user has submitted.

As complex as that sounds, the main function of LLMs is to generate text based on the vast data sets they were trained on. Simply stated, the mathematical models driving LLMs understand words as numbers, or tokens, that are connected probabilistically or statistically to other words, which are themselves represented as tokens. Once a prompt is entered by the user, the LLM will generate an answer where every word outputted is the most connected to the previous outputted word, until there is no next best word to generate and the LLM stops the text generation. Regardless of whether the probability of the next most related word is 95% or 0.35%, the LLM will pick the word that has the highest percentage of probability to be related to the previous word. Since they are driven by statistical analysis, LLMs are thus probabilistic rather than deterministic in their output generation.

In addition to generating answers from their internal datasets and parameters, LLMs can be enhanced with tools such as plugins and retrieval-augmented generation ("RAG") to fetch data from external sources. Plugins are external tools that expand an LLMs capabilities by allowing it to perform tasks or access external data sources beyond its inherent functions. This allows LLMs to interact with apps and outside sources, such as communications platforms (e.g. Microsoft Teams), to enhance their prediction capabilities. A RAG is similar in that it creates a framework that allows LLMs to get information at a particular source, parse through the data, and generate an answer based on that source. For example, a RAG system specializing in tax law could retrieve documents from legal databases. These enhancements are crucial because they allow the LLM to be directed to a particular source and then begin its generative and predictive functions based on the data contained therein. Narrowing the data set considered by an LLM through RAGs, plugins, and data engineering prompts (discussed in part III) is essential for improving the LLM's accuracy.

Other variables that are crucial in fine tuning the outputs LLMs generate are guardrails. Guardrails are mechanisms that are designed to ensure LLMs generates safe, accurate, and appropriate responses. They can operate internally through content filters, response constraints, and ethical or bias controls, or externally, through careful prompt engineering, user feedback loops, or keeping a "human-in-the-loop".

While LLMs have underlying technologies, parameters, and settings to assist with complex problem-solving, LLMs do not understand the meaning of words as humans do. Their ability to generate text is predicated solely on the probability of word connections learned during training, not an inherent understanding of language.

How to Use LLMs in a Tax Practice?

This article encourages a paradigm shift in how LLMs should be viewed — not as replacements for human professionals, but rather as tools to amplify human intelligence. Viewing LLMs as intelligence amplifiers enables us to better understand the tasks that the technology can help us with. Tasks which LLMs enhance particularly well are comprehension; simplification and summarization; synthesis; documentation and communication; analysis; problem-solving; decision-making support; and critical thinking support. While LLMs cannot think critically, they are tools for supporting problem-solving by synthesizing information, breaking down issues and brainstorming potential solutions. For intelligence tasks that previously required 70% effort doing the task and 30% effort for self review, LLMs will likely flip this time balance.

LLMs are less suited for complex critical thinking tasks such as decision-making, because they cannot make human like decisions or provide the nuanced understanding required. LLMs are also less suited for tasks requiring calculations, as they rely on pattern recognition rather than inherent mathematical understanding. Similarly, LLMs are less suited to paying attention to detail for numerical data or technically driven material, although some mathematical plugins are available. Lastly, LLMs are less good at retention and memory since they are generating answers based on datasets they were previously trained on. Interactions with a user will not allow them to recall personalized or newly generated information after the conversation ends. Only when a new generation of a certain LLM technology is released, and after it is trained on that new data set, will the AI be able to produce answers based on new data and information.

Critically, even when LLMs are used in their most useful domains, they are still capable of generating text that is incorrect, misleading, or fabricated, a phenomenon known as "hallucinations". Tax practitioners must be wary to always keep a "human-in-the-loop" to review and validate the output generated by LLMs and provide oversight, appropriate context, and personal perspective to LLM produced documents. Ultimately, rather than replacing tax practitioners, LLMs will likely enhance the capabilities of those who use these tools for intelligence tasks, enabling them to outperform those who do not leverage this technology.

Part II: Ethical and Professional Obligations in the Use of AI by Legal Professionals

1. Technological Competency Obligation

Overview of the Legal Requirement for Technological Competence

Under various rules of professional conduct, professionals are required to maintain a sufficient understanding of the technologies impacting their practice. The rapid integration of AI into tax practice may soon make technological competence not only a competitive advantage, but also an ethical obligation.

For example, in Quebec, the bar has formalized these requirements, mandating lawyers to use and master technologies, especially those impacting client confidentiality and privilege. According to the Quebec *Code of Professional Conduct of Lawyers*, lawyers must develop their

knowledge and skills related to technologies and keep them up to date[1]. In Ontario, while less explicit, similar guidelines suggest that lawyers should be adept at relevant technological tools to meet professional standards.[2] This means that understanding AI solutions, how they function, and the implications of using it may become a core component of a lawyer's professional duties.

Case Law Illustrating the Importance of Staying Updated on Technology

One landmark case that underscores the necessity of technological competence is *Mata* v. *Avianca Inc.* ("*Mata*")[3] In this case, an attorney used ChatGPT to generate legal research result, which unfortunately included non-existent case law. The attorney mistakenly believed that AI could not produce false information and, therefore, did not verify the AI's output. This led to severe repercussions, including sanctions and reputational damage.

This case highlights the ethical dangers of relying on AI tools without fully understanding their limitations. It also serves as a warning that technological competence is not just about knowing how to use these tools, but also about understanding their potential to produce incorrect or misleading outputs.

2. Mitigation Strategies for Ethical Risks

To mitigate these ethical risks, several strategies can be employed by tax practitioners:

1. **Human-in-the-Loop Approach**: AI outputs should always be subject to human review and oversight. This approach ensures that a tax practitioner is responsible for verifying the accuracy and appropriateness of AI-generated content before it is used in documents or provided to clients.

2. **Prompt Engineering:** Tax practitioners should learn how to effectively "engineer" prompts for AI systems. This involves crafting specific, clear, and well-structured inputs that guide the AI to generate more accurate and relevant outputs. By refining the way questions or instructions are presented to the AI, the likelihood of hallucinations or biased results can be reduced.

3. **Cross-Referencing AI Outputs**: tax practitioners should cross-reference AI outputs with traditional research methods. This is crucial in preventing reliance on incorrect or fabricated data, as seen in *Mata.* Firms can implement strict verification protocols to ensure that all AI-generated content is subjected to rigorous quality control.

4. **Education and Training**: Firms should invest in training programs that equip tax practitioners with the skills needed to understand and manage AI tools effectively.

---

[1] *Code of Professional Conduct of Lawyers*, RLRQ ch. B-1, r. 3.1, s. 21.

[2] *Law Society of Ontario, Rules of Professional Conduct*, art. 3.1-2, commentary 4.

[3] *Mata* v. *Avianca Inc*., 678 F. Supp. 3d 443 (S.D.N.Y. 2023).

This includes knowledge of AI's limitations, ethical risks, and the technical aspects of how these systems operate. It also means that being familiar with AI and how to use it must be included in continuing education.

3. Rethinking the Relationship with AI

Theoretical Perspectives on AI's Role in Legal Prediction and Decision-Making

As AI continues to evolve, its role in legal decision-making and prediction is expanding. The prediction theory of law offers one perspective on how AI can be used to predict legal outcomes based on historical case data and legal trends. By analyzing vast amounts of data, AI can assist in predicting the likely results of litigation, offering insights that may inform case strategy. However, while AI can provide valuable predictions, it lacks the human judgment needed to navigate the complexities of each individual case. Legal decisions often hinge on nuanced interpretations of facts and laws that AI cannot fully grasp.

Even in areas where AI is highly efficient in handling repetitive tasks such as document review, legal research, and even drafting initial versions of contracts or briefs, the human's role is indispensable in providing oversight, context, and ethical judgment. AI can support tax practice, but it cannot replace the critical thinking and decision-making abilities of experienced tax practitioners.

The Division of Tasks Between AI and Human, Emphasizing the Need for Human Oversight

In legal and tax departments, AI may begin to be employed for initial drafts of documents, leading, over time, to a reallocation of roles and responsibilities between human professionals and AI systems. This shift allows senior practitioners to focus on higher-level tasks such as strategic decision-making and client consultation, while junior associates and clerks review and refine AI-generated outputs for accuracy and relevance. This reallocation of time not only leverages legal and tax expertise more effectively but also underscores the need for human oversight.

This integration of AI encourages "just-in-time" learning, an approach that enables legal professionals to obtain knowledge or create documents at the precise moment they are needed. This type of learning is especially valuable in the legal field, where quick and timely access to up-to-date information can make a difference in case preparation or document accuracy. Instead of relying solely on prior knowledge or manuals that may not be current, tax practitioners can interact with AI systems to receive analyses, generate draft documents, or obtain specific summaries in real time.

As AI becomes more deeply integrated into tax practice, tax practitioners develop key skills, such as prompt engineering (to communicate effectively with AI) and critical oversight of generated results. These skills optimize the workflow by reducing the time spent on research

and drafting, and they ensure that professionals can adapt to an ever-evolving digital environment.

Part III: Prompt Engineering Basics

How To Prompt Engineer Efficiently

Prompt engineering is an AI user's most powerful tool. It refers to the process of crafting specific and effective queries or commands to guide LLMs. While it may seem daunting at first, much like learning to ride a bike, mastery comes with time and practice. The sheer volume of knowledge an AI system holds can feel overwhelming—like a vast library waiting to be explored. However, all that information is latent, an untapped resource.

The 'human-in-the-loop' plays a crucial role in unlocking this potential. To effectively collaborate with AI, tax practitioners should treat it as they would a junior colleague tasked with completing an assignment. How would you present the project to them? What facts are relevant? What is the current context? What specific tasks would you delegate? How would you frame your instructions? What expectations would you set for the deliverables? In what format should the results be presented? All these considerations shape how you engage with AI. It possesses knowledge but requires careful guidance.

In the following section, we will outline key tips to help you refine your approach to prompt engineering.

Choosing the Right AI

Different LLMs offer varying databases and specializations. When working within a specific field, it's recommended to choose an AI that is already tailored to that domain, ideally one with a tax-focused persona and knowledge base. Such models will be more familiar with the terminology, possess relevant expertise, and deliver more accurate results.

Be Clear and Precise

When providing information to an AI, clarity is key. The AI needs to fully understand the context and details of your request. However, LLMs are sometimes better at processing the beginning and end of a prompt, potentially overlooking the middle. To address this, breaking down complex queries into multiple prompts, should be considered, each refining or adding precision to the previous one. This layered approach ensures the AI does not miss any critical details.

Prompt engineering is like having a conversation. To ensure the AI fully grasps a request, providing as much relevant detail as possible is recommended. Adding extra context or background information helps the AI better understand the situation and produce a more accurate, tailored response. Precision is key—guiding the AI step by step to get the desired result is the way to go.

Specify the Target Audience

Specifying the intended audience helps the AI adapt its tone, complexity, and style. Whether the output is for a colleague or a client, the phrasing may need to change. For instance, a client memo might require a more accessible language, whereas an internal memo may need more technical terminology.

Provide Adequate Context

Giving the AI as much context as possible is always beneficial. It is like telling a friend a story—the more details provided, the better the interlocutor will understand the situation. Similarly, with AI, the more context offered, the more accurately it can respond. Clear background information helps the AI process requests in a more informed and meaningful way.

Attach a File

Attaching a file or an example can also enhance prompts. For instance, providing a sample response and asking the AI to model its reply in a similar style is feasible. It's also possible to upload documents and ask the AI to analyze or comment on them based on both the content provided and its built-in knowledge. This can be especially useful for tasks like reviewing drafts or offering feedback on complex documents.

Use Follow-Ups

An initial prompt might provide a satisfactory answer, but often, it will not be perfect right away. It is normal to need additional information or adjustments to refine the response. By amending the initial prompt or submitting follow-up prompts, the AI will be guided towards a more precise answer. As noted earlier, the more context given, the better the AI understands the situation. While the AI has access to vast amounts of knowledge, it requires direction from a prompter to pinpoint the exact current scenario.

Think Step By Step

Crafting an effective prompt is rarely simple, especially in fields like tax where the facts can be intricate and highly specific. As mentioned earlier, dumping all the information into a single, heavy prompt will often lead to suboptimal results. Instead, spacing out an input and highlighting key details can significantly improve the AI's understanding of a situation. One crucial strategy is to proceed step by step. After an initial prompt, following up with one-sentence prompts or prompts that contain only one idea at a time is recommended. This method helps guide the AI through various possibilities, filtering out irrelevant or incorrect answers until the user and the AI have narrowed it down to the correct one. It is like tightening a net gradually—each step helps refine the AI’s response, eventually leading to the precise result that is sought after.

Token Limitation

Most modern AI models have a limit on the number of tokens, or characters, they can handle as input and output. The more information provided in a prompt, the fewer tokens

remain available for the AI's response. Additionally, as long as a user continues prompting within the same conversation thread, the AI is working within the same token limit. The deeper the interaction goes, the more likely it is that earlier parts of the conversation will be truncated or lost. To avoid this, when adding significant new information or restarting a line of inquiry, it's best to begin a new thread.

PART IV: Integrating AI into Tax Practices: Insights from a Hands-On Workshop

Following the seminar, a hands-on workshop focused on AI in tax practice was hosted for the participants. Many participants were new to AI or had limited exposure to its application in tax practice. Some attendees already had access to tax-specialized LLMs at their workplaces but were hesitant to use them, largely due to a lack of understanding on how to do so effectively and productively. The LLM that was used during the workshop is AskBlueJ, an AI with a tax-focused persona, and pre-loaded with knowledge of Canadian and U.S. tax law.

The Test Case

The case presented to the participants centered around remote work. A Canadian multinational corporation, with subsidiaries worldwide, allows its senior management employees to work remotely from anywhere in Canada or the United States, as long as they remain available online. The participants were tasked with addressing several key questions, including personal tax considerations and risks related to permanent establishment.

The Task

The participants were instructed to use AskBlueJ to answer the questions and identify the tax issues. To assess their understanding and encourage a gradual learning process, the participants completed a survey while prompting. The survey was structured to be filled out after three distinct rounds of prompts. The questions were carefully crafted to capture their challenges, gauge their learning curve, and understand what adjustments they made to improve their prompts after each round.

The first set of questions focused on user experience and were distributed as follows:

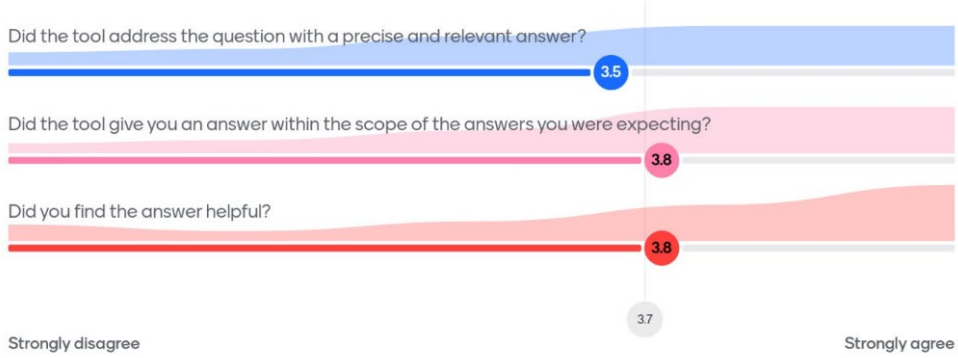


Figure 2: Graph Reflecting Round 1 Questions and Answers

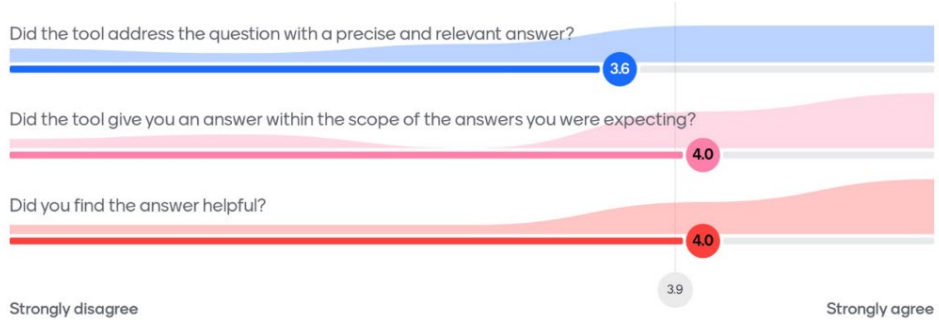


Figure 3: Graph Reflecting Round 2 Questions and Answers

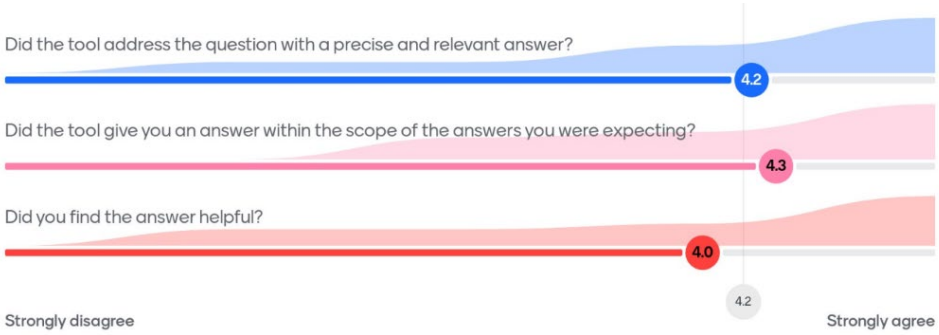


Figure 4: Graph Reflecting Round 3 Questions and Answers

The data suggests that the participants were becoming increasingly comfortable with the AI. The AI's responses became more precise, more aligned with the questions, and most importantly, more helpful to the user.

The second set of questions focused on how participants improved the AI's responses. A predefined list of options was provided, and participants could select as many as applied to their process. The most common methods for improving prompts included adding more context, providing additional information, and following up directly on the AI's initial response—whether by refining details, adding specific clarifications, or correcting any incorrect assumptions made by the AI.

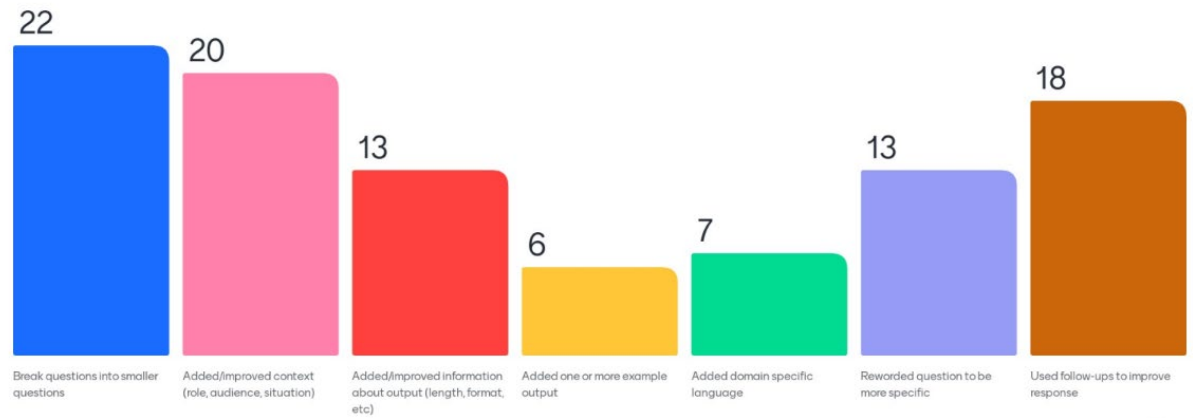


Figure 5: Bar Graph Reflecting Changes Users Made after 1st Round of AI Use

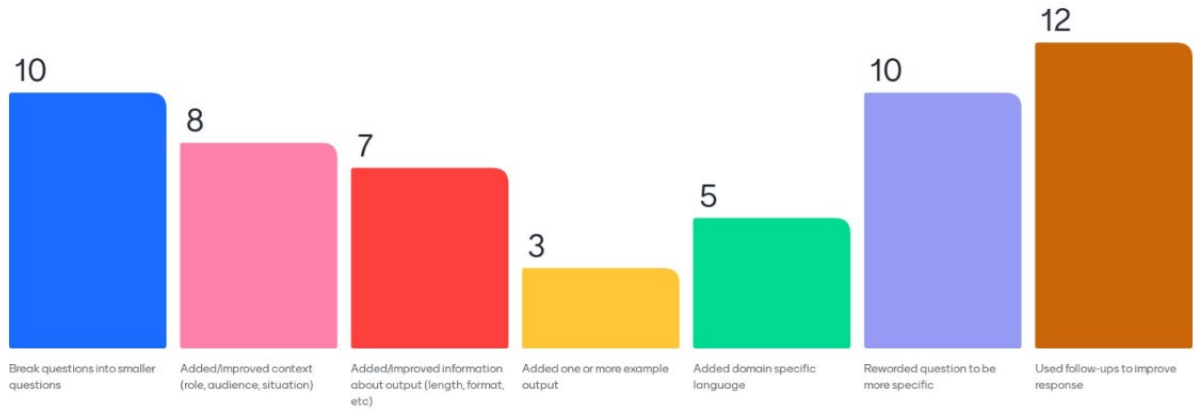


Figure 6: Bar Graph Reflecting Changes Users Made after 2nd Round of AI Use

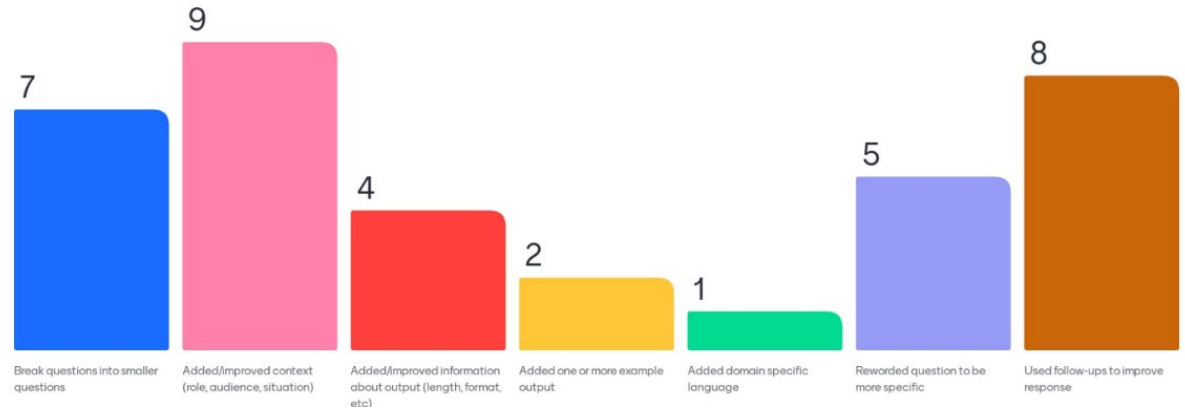


Figure 7: Bar Graph Reflecting Changes Users Made after 3rd Round of AI Use

The third set of questions explored the participants' thought process as they refined their prompts. This was an open-ended question, allowing participants to write freely about their experience. This personalized approach revealed the diverse progress levels among participants.

After the first round of prompting, most participants analyzed the AI's initial output. The AI's responses were often imprecise, vague, or incorrect (though not outright false). To improve the output, participants adjusted their prompts by being more specific, incorporating key terms, and breaking the task down into smaller steps. They learned that providing too much information in a single prompt caused the AI to lose focus, as it tends to process the beginning and end more effectively than the middle. By breaking the scenario into smaller, more focused prompts, participants achieved better results. If the AI gave vague responses, they asked follow-up questions to prompt deeper exploration.

After the second round, many participants were still seeking technical information from the AI. Key improvements included providing more context, being more precise, and breaking prompts down into smaller, manageable pieces. They also began adding specific questions, which helped guide the AI in its search and responses.

By the third round, some participants continued refining the technical precision of the AI's answers, while others explored its broader capabilities. They asked the AI to draft content in a format suited for specific audiences—whether a fellow tax professional or a client with

limited tax knowledge. This shift showcased the versatility of AI, not only as a technical assistant but also as a drafting tool for communication.

The participants received a variety of answers from the AI, some of which were accurate and validated by the tax professionals present. However, there were instances where the AI lacked the knowledge to provide an answer, such as when dealing with questions about Mexican tax law. Since the AI was only trained on Canadian and American tax laws, it was unable to respond to queries beyond its scope. Importantly, participants noticed that when the AI did not know an answer, it was honest about its limitations. It did not attempt to fabricate an answer or apply its local knowledge to unfamiliar territory. This had been programed into the parameters of the AI as a guardrail.

Reflecting On Their Abilities

After the workshop, participants were given a final set of exit questions. These questions aimed to assess whether they enjoyed using AI in the context of tax and if they were open to incorporating AI into their daily tax practice. The goal was to gauge their comfort level with the technology and their willingness to adopt it as a regular tool in their professional workflow.

The first set of exit questions focused on participants' self-assessment of their abilities and proficiency in using AI. Rather than imposing an objective scoring system, participants were encouraged to evaluate their own confidence and skill levels in utilizing the technology. This approach allowed for a more personal reflection on their experiences and growth during the workshop.

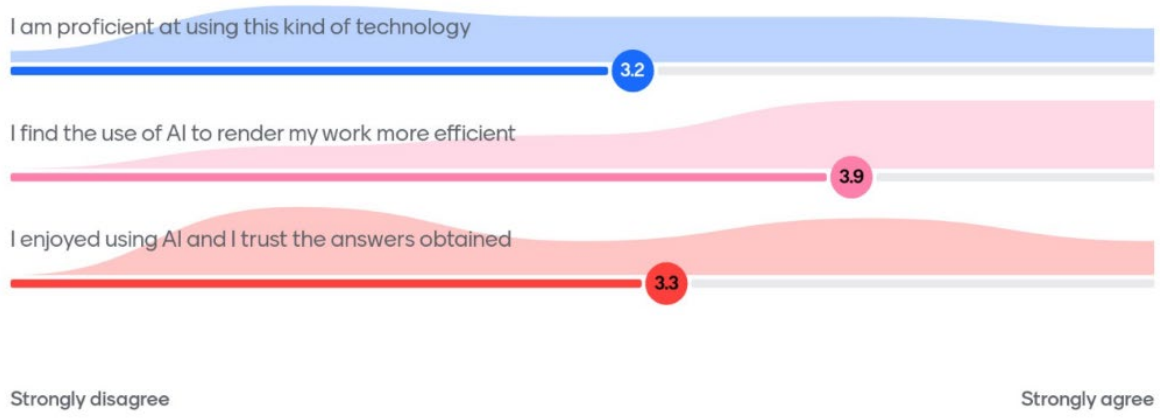


Figure 8: Graph Expressing User Ability Statements After Using AI in Workshop

While participants were generally modest about their skills in using AI, many acknowledged that it significantly enhanced their efficiency.. However, responses regarding participants' enjoyment and trust in AI were mixed, with opinions almost evenly divided between agreement and disagreement. As noted earlier, AI should be viewed as a tool, akin to a personal tax assistant or a technical colleague.. Users do not automatically trust AI; this trust develops over time as they engage with the technology and recognize its depth of knowledge across various topics.

Nevertheless, it is essential to remember our role as the "human-in -the-loop." We must rigorously double-check every response provided by the AI, as our professional liability is at stake, not that of the machine.

Participants were subsequently asked about their willingness to incorporate AI into their professional practice. Very few had already adopted AI, whether through standard LLMs like ChatGPT, Copilot, or Perplexity, or through tax-specific LLMs like AskBlueJ. However, the majority expressed a strong desire to integrate an AI tool into their work.

Participants were finally asked how they would validate an answer provided by an AI.. The most common response was to review the sources cited by the AI. When a tax-specialized AI like AskBlueJ generates an answer, it includes references to relevant laws, case law, and administrative positions. By examining these sources, users can safeguard their professional liability and ensure that the information is accurate and applicable to their question. Another popular method mentioned by participants was consulting with a colleague or advisor.

CONCLUSION

In conclusion, the seminar and workshop demonstrated practical ways for participants to incorporate AI into their tax practice. Through hands-on experience, attendees were able to explore how AI can be used to address tax-related issues and improve efficiency in their work. To work efficiently with AI, technological competence has become an essential requirement, compelling professionals to understand AI's capabilities, functions, and limitations to generate effective outputs.

By viewing AIs as intelligence amplifiers, tax practitioners can use them for decision-making support, brainstorming, and access a broad range of information efficiently. This perspective emphasizes AI as a tool that supports human expertise rather than replacing it. To navigate the complexities introduced by AI, tax practitioners must adopt proactive strategies, such as employing a human-in-the-loop approach, mastering prompt engineering, and rigorously cross-referencing AI outputs with traditional tax specific research methods. Education and training in these areas are crucial for using AI effectively. The most important things to keep in mind while prompt engineering are precision, thinking step by step and asking a clear question to the AI. While AI can enhance efficiency and streamline legal processes, it is the human’s role to provide the critical oversight, ethical judgment, and nuanced understanding required to navigate the intricacies of the law. By embracing these new skills and approaches, tax professionals can uphold their ethical obligations while harnessing the potential of AI to better serve their clients and the justice system. This ongoing evolution will require a commitment to lifelong learning and adaptation, ensuring that tax professionals remain at the forefront of ethical practice in an increasingly digital landscape.