AI: looking beyond the hype

Nick Stobbs explores the realities of using Large Language Models such as ChatGPT for tax research

The complexity of the UK tax code, with its 10 million words and 18,500 pages, presents a challenge for accountants seeking efficient tax research. While AI, particularly Large Language Models (LLMs) like ChatGPT, have revolutionised various fields, its application to tax research has limitations.

LLMs lack explainability, specificity, real-time updates, context understanding, and interaction capability and recent legal cases highlight the risks, such as AI generating fictitious case citations. However, the emergence of Retrieval Augmented Generation (RAG) models offers a solution. Combining information retrieval and text generation, RAG technology improves context understanding, enables real-time updates and can efficiently handle complex queries. Tax on Demand, a UK-based AI company, has introduced Copilot, a tax research tool utilising LLM and RAG technologies, addressing the challenges and aiming to strike a balance between AI efficiency and human expertise.

Navigating the complexity of the UK tax code

The UK tax code is widely regarded as the most complex in the world: 10 million words across 18,500 pages of legislation; supplemented by 155 internal HMRC manuals, tribunal decisions and court judgements (in addition to the sheer volume, the language used, drafting style and diversity of taxes all contribute to the overall complexity).

The ability to have a full understanding of the tax code has become increasingly difficult over the past 20 years. Between 1965 and 1992 the average length of a Finance Act was 175 pages; since 1993 this has increased to 435 pages. This year alone there have been 797 new pages added to HMRC's internal manuals, alongside 3,623 page updates. That's before considering precedent-setting tax tribunal or court decisions and changes in the interpretation of tax laws (think of the recent changes to research and develop tax credits, albeit well publicised).

Even if you had all the knowledge, working time is almost entirely consumed by billable work, staff management, workflow planning, maintaining client relationships and safeguarding the future of the practice. Answering simple tax questions from clients can start to feel like a chore, not to mention detailed research on the tax treatment of a particular transaction or a complex tax planning project.

For many years now, accountants have turned to tax specialists, reference books and online resources to aid their tax research and help understand the practical application of the UK tax code to their client base. Traditionally, these approaches have been manually intensive, time consuming and economically inefficient – that is, until now…

The rise of AI

The history of artificial intelligence (AI) is a captivating journey that has evolved over decades, with one of the recent milestones being the rise of Generative AI. The roots of AI can be traced back to the mid-20th century when pioneers like Alan Turing and John McCarthy laid the conceptual groundwork and the term 'artificial intelligence' was officially coined in 1955 during the Dartmouth Conference.

Since 1955, the field of AI has experienced periods of rapid progress and also significant setbacks, including AI winters during the 1970s and 1980s. AI had a resurgence in the 1990s fuelled by the

availability of computational power and large datasets. Machine learning, particularly neural networks, gained prominence, setting the stage for the significant advancements we see today.

You'd be hard placed to find a single person that hasn't heard, read or experimented with the various Large Language Models (LLMs) currently available. ChatGPT, the most well-known LLM to date, is based on the GPT-4 engine and like all LLM models is an advanced artificial intelligence system that is designed to understand and generate human-like text based on patterns learned from extensive datasets.

The rise of Generative AI represents a new chapter in the ongoing narrative of artificial intelligence, showcasing the potential for machines to not only understand but also create and contribute in ways that were once considered the exclusive domain of human intelligence. As this trend continues, the impact on the accountancy profession as a whole is likely to become even more pronounced.

Implications for accountancy

For generations, accountants have embraced new technology to enhance the efficiency and effectiveness of their work. To date, the integration of artificial intelligence into day-to-day tools has enabled the automation of routine tasks, data analysis and process optimisation; empowering accountants to do what they do best and client's value the most: listening, building relationships and providing timely proactive advice.

In the coming decades, it's inevitable that intelligent systems will progressively assume more decision-making responsibilities from humans. In the short to medium term, AI presents numerous opportunities for accountants to enhance efficiency and scale expert tax knowledge throughout their practice to help protect their existing client base, maximise fee-income and win new clients.

In the long-term, the rise of AI offers the potential for positive and radical change, as systems increasingly take over decision-making tasks currently carried out by humans. Striking a balance between leveraging AI for efficiency gains and preserving the irreplaceable human touch in nuanced decision-making remains a critical but exciting challenge for the accounting profession.

Limitations of LLMs in tax research

LLMs are very powerful. They provide outputs that can be extremely accurate, replacing and, in some cases, far superseding human efforts. However, it should be cautioned that they do not replicate human intelligence. Despite their impressive ability to create content, they have significant limitations and the idea of delegating any critical tax research process to an LLM is severely misplaced and can lead to significant consequences. We need to recognise the strengths and limits of this different form of intelligence and understand the best ways for humans and computers to work together.

The limitations

Lack of explainability: all LLMs are based on machine learning which means they are inherently 'black box'. This means that it is challenging to interpret or explain how the model arrives at a specific output; they can be wrong and look right or be right and look wrong. The lack of explainability raises concerns in critical applications where understanding the reasoning behind decisions is crucial, such as tax technical research.

Lack of specificity: LLMs may generate responses that are overly general and lack the specificity required for tax technical research. Legislation, manuals and case law often include intricate details and nuances that may be overlooked or misunderstood by the model.

Inability to provide real-time updates: the interpretation of tax legislation is subject to frequent and subtle changes – think research and development tax credits. Staying up-to-date is crucial. LLMs do not always have real-time capabilities and may not reflect the most recent changes in tax regulations, potentially leading to outdated or incorrect information.

Limited context understanding: LLMs may struggle with understanding the broader context of specific tax scenarios. Tax technical research often involves considering various factors, such as the business context, industry-specific regulations and individual circumstances, which may be challenging for the model to grasp accurately.

Limited interaction and clarification: Tax research often involves a dynamic and iterative process of seeking clarification and refining queries based on initial findings. LLMs, as one-shot models, may not facilitate this interactive process as effectively as a human expert.

The consequences
In Felicity Harber v HMRC [2023] TC09101, artificial intelligence failed to assist a taxpayer in her appeal against penalties after it emerged that it had invented the case citations she had relied on in her defence:

Para 2: "However, none of those authorities were genuine; they had instead been generated by artificial intelligence (AI)."

Para 3: "We accepted that Mrs Harber had been unaware that the AI cases were not genuine and that she did not know how to check their validity by using the FTT website or other legal websites."

Para 24: "We acknowledge that providing fictitious cases in reasonable excuse tax appeals is likely to have less impact on the outcome than in many other types of litigation, both because the law on reasonable excuse is well-settled, and because the task of a Tribunal is to consider how that law applies to the particular facts of each appellant's case. But that does not mean that citing invented judgments is harmless. It causes the Tribunal and HMRC to waste time and public money, and this reduces the resources available to progress the cases of other court users who are waiting for their appeals to be determined."

Revolutionising tax research
Retrieval-augmented generation is an innovative approach in natural language processing that combines the strengths of information retrieval and text generation models. This methodology integrates pre-existing knowledge retrieval systems with advanced language models to enhance the context-awareness and relevance of generated content. By leveraging a dual mechanism of retrieving information from external sources and generating contextually appropriate responses, this technique addresses the limitations of traditional LLM models and can offer several benefits:

Improved specificity and accuracy: one of the primary benefits of using a retrieval-augmented generation model for tax research is the enhanced specificity and accuracy in responses. Retrieval-based components allow the model to pull relevant information from a knowledge base (for example HMRC's internal manuals), ensuring that the generated content is grounded in accurate and contextually relevant data. This helps address the intricate details and nuances of tax laws and regulations, providing users with more precise and reliable information compared to a standalone large language model like ChatGPT.

Real-time updates and dynamic content: interpretation of tax legislation is subject to frequent changes, and staying updated is critical for accurate tax research. A retrieval-augmented generation model can be designed to retrieve information in real-time from updated knowledge

bases. This ensures that the generated content reflects the most recent changes in tax regulations, offering users up-to-date information. In contrast, a standalone large language model may struggle to provide real-time updates, making it less suitable for tasks where accuracy is paramount.

Efficient handling of complex queries: tax research often involves complex queries that require consideration of various factors. The retrieval-based component of the model can efficiently handle these queries by retrieving relevant information from relevant knowledge bases. The generation component can then build upon this retrieved information to provide comprehensive and nuanced responses. This capability makes retrieval-augmented generation models well-suited for addressing intricate and detailed questions in tax research, where a standalone large language model might struggle to handle the complexity of queries and generate accurate, contextually rich responses.

Use Case 1: Tax document summarisation
Challenge: Tax professionals deal with an overwhelming amount of documentation and extracting key information from lengthy documents can be time-consuming.

Use Case: RAG technology can be employed to generate concise and accurate summaries of complex tax documents, aiding tax professionals in quickly understanding and applying the necessary information.

Use Case 2: Conversational 'tax chatbots'
Challenge: Taxpayers (clients) often seek clarification on tax-related queries and tax professionals may face challenges in providing real-time responses to a large volume of enquiries.

Use Case: Implementing RAG technology in tax advisory chatbots enables them to retrieve information from extensive tax databases and generate human-like responses, providing accurate and contextually relevant answers in seconds.

Use Case 3: Real-time updates and alerts
Challenge: Accountants and tax advisers need to invest time in continuing professional development in order to monitor legislative changes and keep up to date with the increasingly complex UK tax code.

Use case: RAG technology can monitor tax legislation, internal HMRC manuals and case law to swiftly retrieve and summarise the latest tax law changes, ensuring accountants stay current with evolving regulations.

In conclusion…
While large language models excel in general language understanding, their suitability for detailed tax research is limited. The intricate nature of tax laws and regulations demands a nuanced comprehension that surpasses the capabilities of traditional LLMs. Augmented generation models, however, present a promising solution by combining the strengths of large language models with specialised knowledge repositories.

These models can retrieve and generate contextually relevant tax information, offering a more targeted and accurate approach to tax research. By seamlessly integrating advanced language processing with domain-specific expertise, augmented generation models hold the key to enhancing precision and efficiency. In time, this will strike the balance between leveraging AI for efficiency gains and preserving the irreplaceable human touch in nuanced decision-making.

- *Nick Stobbs is CEO and Founder of Tax on Demand. Call 07581 561 370 or email nick@tod.tax. Tax on Demand is a UK-based artificial intelligence company. You can sign up at https://copilot.tod.tax/login*