Can Artificial Intelligence Become an Accounting Expert?

By Yigal Rechtman

Artificial Intelligence (AI) traces its origin to antiquity, when humans attempted to use formal reasoning to predict the future of events. Today, AI has made great progress in perceiving, synthesizing, and inferring data in contexts by a machine-learning mechanism. The applications of AI are many and varied—from self-driving cars to human speech recognition to predictive decision-making investment tools. Most recently, the general public has become aware of natural language learning models (LLM) due to the public release of ChatGPT and BARD, automatic web-based learning machines (Pedro Domingos, *The Master Algorithm: How the Quest for the Ultimate Learning Machine Will Remake Our World*, Basic Books, 2015).

Ethical debates about the proper application of AI have arisen from the very "natural" sounding responses generated by this newest generation of LLMs (e.g., *Asimov's Three Laws of Robotics and Machine Metaethics*, Association for the Advancement of Artificial Intelligence (AAAI), 2005). Some have claimed that, even without much teaching, an AI-based LLM could become a lawyer, doctor, or accountant. The author decided to put this claim to the test by convening a focus group of humans with expertise in the areas of auditing, tax, internal auditing, risk management, and forensic accounting to evaluate the responses of an AI to professional queries. We decided to use the Google-sponsored BARD system because it was more readily available to our team of professionals (ChatGPT requires an invitation).

Our purpose was to assess the current state of the BARD LLM as an example of LLMs overall. It is not a specific comparative assessment, but rather more of an experiment to answer the question: can an AI LLM be an "expert" in accounting?

The basis of our expert evaluation is both practical and theoretical. In theory, the panel utilized the tenets of Bloom's Taxonomy, which employs a hierarchy of learning, from basic remembering, through to creativity, as shown in *Exhibit 1*. The panel's goal was to establish if BARD, the candidate for expert AI, could rise to the level of "evaluating" facts and circumstances. Based on this standard, some panelists assigned a grade level to the AI, from "A" to "F."

The complete transcript of the panel's experimental session can be found at https://tinyurl.com/3a9ab9vb. The panel's critical analysis is described below.

Exhibit Bloom's Taxonomy
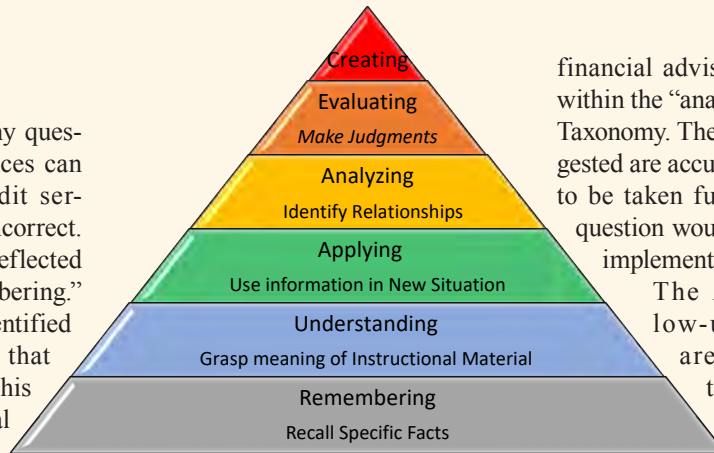
Internal Audit

By Doug Bennett

The AI LLM response to my question "What types of services can a governmental internal audit service provide?" was factually incorrect. Within Bloom's Taxonomy, it reflected the most basic level of "remembering."

Specifically, the AI LLM identified financial audits as a service that internal auditors can provide. This is incorrect—the Chief Financial Officer's Act of 1990 and subsequent revisions require financial audits of federal agencies be performed by the agency Inspector General (IG) or an independent external auditor determined by the IG. All federal agencies use private sector audit firms for their financial statements, except for the SEC and the IRS, both of which are audited by the Governmental Accountability Office (GAO).

(e) Each financial statement prepared under section 3515 by an agency shall be audited in accordance with applicable generally accepted government auditing standards—
(1) in the case of an agency having an Inspector General appointed under the Inspector General Act of 1978 (5 U.S.C. App.), by the Inspector General or by an independent external auditor, as determined by the Inspector General of the agency; and
(2) in any other case, by an independent external auditor, as determined by the head of the agency. (31 USC 3515)

Additionally, the AI LLM's response listed audits for compliance, efficiency, and effectiveness; however, these are really the objectives of performance audits as defined by the GAO and not a discrete service. This AI LLM failed to identify other attestation services, including examinations, reviews, and agreed upon procedures, all of which are relevant and unique tools available to internal audits in the government sector.

Conclusion. Although the responses from the AI LMM to the internal auditing questions were disappointing, it is possible that a more informed answer could be possible with additional follow-on questions. Furthermore, although the Military Department audit services are robust, there are presumably many more government auditors at the state and local level who may follow different standards than those specified by the GAO and AICPA. ∎

Doug Bennett is an auditor general with the U.S. Department of the Air Force. The views expressed are his own and do not reflect those of the U.S. Air Force.

CFO Perspective

By Lori Edo

High-level financial professionals are entrusted with managing the financial responsibilities of an organization. Cashflow is a critical process to manage and evaluate and high on the list of topics that keeps CEOs and CFOs up at night. This prompted my inquiry to BARD: "What steps should be implemented to improve cash flow of a private organization?"

The response provided by the AI LLM listed the following: create a cashflow forecast; track your spending; negotiate better terms with vendors; collect payments on time; pay your bills on time; get a line of credit; invest in your business; and consult with a financial advisor. This response falls within the "analyzing" level of Bloom's Taxonomy. The ideas the AI LMM suggested are accurate; however, they need to be taken further. The logical next question would be how to achieve or implement these suggestions.

The AI was given a follow-up question: "What are the most beneficial terms to have with a vendor?" BARD's response included: price; payment terms; discounts; warranties; and service levels. In my opinion, more follow-up questions were needed. The next follow-up question was engineered to determine if BARD would give an opinion ("evaluating," on Bloom's Taxonomy). We asked, "Hypothetically, if an organization has $50K free cash, should it invest the funds or hold the funds in today's market fluctuations?" True to form, the AI LLM gave general comments on factors to be considered when investing, such as risk tolerance, the organization's goals, and market conditions.

Conclusion. I would give BARD a B+. The AI LLM can provide general responses and ideas but that is where I believe it reaches its limit. The responses are too general for what a CFO needs to provide to the CEO. The concepts may not fit with every organization's structure or operations. Without institutional knowledge or specific data points about the organization, these suggestions may be irrelevant. It was clear from the last follow-up question that BARD will not give an opinion, and certainly cannot speak from prior experiences, but rather prefers to give definitions and general guidelines. ∎

Lori Edo, CPA, is the global director of accounting and finance at K2 Integrity.

Fraud Detection and GAAS Topics

By Smadar Rinat

The AI LLM was initially posed with the question: "What are the responsibilities of the auditor conducting an audit in accordance with GAAS for the detection of fraud?" The AI's response, "I'm just a language model, so I can't help you with that," with a link to the AICPA-CIMA.com, was unhelpful; it is unclear as to why this question failed to elicit a useful response, while other follow-up questions succeeded.

The follow-up question was, "Is a GAAS auditor responsible for the detection of fraud?" The essence of the AI LLM's response to this question was: "Yes, GAAS auditors are responsible for the detection of fraud. However, it is important to note that they are not responsible for detecting all fraud, only material fraud." This response is an out-of-context oversimplification of the GAAS guidance, trying to fit the answer into one simple sentence. According to GAAS: "The primary responsibility for the prevention and detection of fraud rests with both those charged with governance of the entity and management." This AI LLM correctly went on to list the auditor's actual responsibilities related to fraud in performing an audit in accordance with GAAS and cited the performance of procedures to identify the risk of material misstatement due to fraud, as well as the auditor's responsibility to respond to that risk by designing and performing procedures to address the risk. Nevertheless, its opening sentence demonstrates an inaccurate conclusion or inference.

Trying to drill down further, the AI LLM was asked: "Who is primarily responsible for the detection of fraud in financial statements?" This time, the LLM gave a partially correct response: "The primary responsibility for the detection of fraud in financial statements rests with management," but left out "those charged with governance."

It also continued with the imprecise assertion that "External auditors are also responsible for detecting fraud in financial statements."

To be fair, perhaps the above questions did not provide an opportunity for the AI LLM to demonstrate deeper application and analytical capabilities. In terms of Bloom's Taxonomy, BARD didn't seem to rise above the first two levels of remembering and understanding—basically using a search engine capability to provide a simplified summarized response. (In fact, a Google search of the question "Is a GAAS auditor responsible for the detection of fraud?" yielded a response that was more precise and consistent with the standards.)

Conclusion. Although it appears that BARD has the potential to be a valuable tool for a wide range of applications, it may require further "teaching" in order to extract and synthesize complex and nuanced data into more reliable arguments. This will allow BARD to achieve a higher level of precision when reasoning about questions that cannot be simply answered with a "yes" or "no." ∎

Smadar Rinat, CPA, CFE, is a principal at Prager Metis CPAs.

Tax Topics

By Daniel J. Belfiore

I challenged the AI LLM about the taxation of Restrictive Stock Units (RSU). Generally, RSUs are a type of equity compensation, in the form of corporate stock that is granted to employees of a company. RSUs are typically granted with a vesting schedule subject to terms and conditions set forth by the company. On the date of vesting, tax is paid on the fair market value of the shares. This income is generally considered wage income subject to income and payroll taxes. However, a taxpayer may make an alternative election under IRC section 83(b) to pay taxes on the RSUs when they are granted, rather than when they vest.

When asked how RSUs were taxed in the United States, the AI LLM's response was thorough and accurate, outlining the general points in an understandable manner. In response to a follow-up question of whether it would be beneficial to make an 83(b) election assuming the fair market value of the company was to increase over time, the AI LLM gave an answer of 'it depends,' identifying some risks and urging the user to ask their tax advisor. The answer to this question is seemingly a simple "yes"; however, the taxpayer making the election needs to consider their ability to satisfy that accelerated obligation. I concluded by asking directly for a recommendation, assuming the taxpayer's income was sufficient and paying the tax would not be a lifestyle burden. In response, the AI LLM recommended making the election and proceeded to outline the benefits and emphasize the risks, while also urging the user to contact a tax advisor.

Conclusion. Based on this limited exchange, the utilization of AI LLM for tax topics seems to have some value and usefulness. The AI LLM can give thorough, relevant responses to specific questions, more useful than a traditional search engine. Although the AI LLM has access to vast amounts of tax data and can analyze complex tax codes and regulations, it cannot replace the expertise and experience of a human CPA, who has years of training and experience in interpreting these regulations and can provide personalized advice tailored to a specific individual's circumstances. The main issue is the user: a user may fail to ask the correct questions, which can lead to improper action and adverse tax consequences. In addition, users are putting their personal data at risk by asking BARD. ∎

Daniel J. Belfiore, CPA, is a tax manager in the private wealth department of Baker Tilly.

Forensic Accounting Topics
By Yigal M. Rechtman

A common misconception of the public, and of some CPAs, is that forensic accounting is a more robust form of an "audit." Judges, attorneys, insurance companies, and the public often times refer to a "forensic audit." This term is an oxymoron, because an audit's objective is to express an *opinion* about a set of financial statements, while forensic accounting only expresses an expert's opinion; in most cases, forensic accountants are precluded from expressing an opinion but must rather communicate the result of their procedures and analyses. "The ultimate decision regarding the occurrence of fraud is determined by a trier of fact; therefore, a member performing forensic services is prohibited from opining regarding the ultimate conclusion of fraud," an AICPA-CIMA report read. "This does not apply when the member is the trier of fact. A member may provide expert opinions relating to whether evidence is consistent with certain elements of fraud or other laws based on objective evaluation." (AICPA-CIMA, *Statements on Standards for Forensic Services*, para. 10)

For a forensic expert to be effective, this distinction should be something an AI LLM can analyze in order to arrive at a conclusion. It did not. The AI LLM's response to the question, "Can you please explain the difference between forensic accounting and auditing?" was vague at best and focused on a definition of both doctrines. It went further into misinformation and concluded by stating that the difference is not in the objective, but rather that forensic accounting requires a "specialized knowledge" while implicitly auditing does not. The truth is that both fields require specialization, but the application of such specialization is different, and they have different objectives.

I sharpened the questioning about opinion, and asked, "Would you say that forensic accountants are equipped to express an opinion like financial auditors, or how would you categorize the forensic accountants' results?" This was a compound question, and the AI LLM mostly got it right when it responded, "Forensic accountants' opinions are not binding on the court, but they can be very persuasive." But then it gave an example that was a faux pas: "Opinion: The forensic accountant opined that the company's CEO committed fraud." This is an opinion that is expressly prohibited under various professional standards. ∎

Yigal M. Rechtman, CPA, CFE, CITP, CISM, is a partner of RSZ Forensic Associates.

Risk Management
By David Lafond

The LLM's ability to function in the area of risk management appears to depend on the following factors:

■ Questions must be clear for the AI to have an opinion. Questions must be designed to challenge the AI to meet this criterion. During the question period, the LLM did not offer an opinion if not asked directly.

■ Questions cannot infer a relationship, but must clearly request the LLM to provide an outcome. Questions must be designed to challenge the LLM to evaluate the criteria. The LMM does not infer a relationship if not directly asked about it.

For example, when asked to define the most important business risk, the LLM responded by ranking cybersecurity highest. While cybersecurity is an important risk, and could possibly be the greatest for many organizations, risk managers will likely rank financial-risk or people risk as more important. According to the Allianz Risk Barometer report (https://bit.ly/3DDoYyB), cybersecurity and business interruption are noted as the largest concerns in 2023. While the Allianz report helps identify top concerns, this does not always translate into the highest risk for any specific organization because risks to organizations differ. Some organizations are well prepared to meet cybersecurity challenges, and others are not. I expected the LLM to ask some clarifying questions prior to responding with statistical averages.

During the interaction, the AI LLM's responses did not appear to have much depth. Although this could be considered a failure on its part, it is also possible that the type of question requires tailoring to engage the analyze and evaluate functions of the model. Risk questions happened to be earlier stage questions. Later in the group interaction, it was discovered that other types of questions could have solicited more capable responses from the LLM.

Conclusion. It is possible that the LLM can develop evaluation and analysis traits, but it seems to require overcoming an inability to understand when to require more input. ∎

David Lafond is the chief information security officer of the United Teamsters.

AI Is Not There Yet

Can today's AI pass for an accounting expert? The conclusion, in the view of the principal author of this article, is that AI is still a work in progress. The prevailing thoughts of the subject-matter experts above are that the AI is not yet ready for prime time when it comes to presenting professional expertise.

This is of course a developing paradigm, and much can be learned in the interim by what our LMM did not do well, and why it did not perform as well as an individual with expertise, experience, and human intuition. ∎

Yigal Rechtman, CPA, CFE, CITP, CISM, is a partner at RSZ Forensic Associates, a litigation consulting, valuation, and forensic accounting firm. He is also a member of The CPA Journal *Editorial Advisory Board.*