Artificial Intelligence as a Partner for Retirement Professionals: What Are the Issues?

Anna M. Rappaport, FSA, MAAA

Any views and ideas expressed in the essay are the author's alone and may not reflect the views and ideas of the Society of Actuaries, the Society of Actuaries Research Institute, Society of Actuaries members, nor the author's employer.

Note: ChatGPT was the artificial intelligence/large language model chosen for illustrative purposes in this essay.

Managing retirement is a complex undertaking supported by financial service companies, actuaries, employee benefit plans, Social Security, and individual financial planners. The participants in managing retirement include the professionals who support these activities, the business organizations that participate in these activities include financial service companies that keep plan records, invest assets, and hold assets. Employers offering benefit plans must decide what benefits to offer, manage them, and interact with employees. The professional groups include actuaries, accountants, attorneys, financial planners, and more.

The entities that participate in the offering of services and products to support retirement plans are faced with a number of challenges. Most are regulated (but not necessarily by the same entities), some have fiduciary liability, and some are subject to professional standards. Their customers depend on them. Customers include employee benefit plans and other large organizations and/or individual consumers. Organizations serving employee benefit plans and other large organizations are often faced with competitive pressures and pressure on fees.

Over the last 60 years, there has been a tremendous growth in the technology available to support retirement professionals and in the way automation has been used as part of the process of managing retirement benefits and planning. Throughout this time, there have been a number of common threads related to the fundamentals facing professionals, but the way they apply has changed with the evolution of technology.

In 2023, there was tremendous interest in the potential impact of AI/large language models as implemented by ChatGPT and Bard. At the same time, concern was expressed about their potential misuse and whether they could create harm. I investigated ChatGPT's capability by reading about it, experimenting with using it, and listening to a variety of webcasts and discussions. The following are some general comments based on what I learned:

- This was not the beginning of AI, but rather the next step. The capability of large language models was growing rapidly. The availability of large data sets together with these models created massive potential capabilities.
- ChatGPT is conversational. The user can start a conversation with a prompt and then build the conversation with further questions.
- Public policy on the use of these models is evolving.
- The model is trained on a data set. The results produced by the model reflect the information in the data set rather than a process of thinking.
- If the world as reflected in the learning data set includes bias, the model's responses are likely to reflect that bias so that applications of the model will be biased and could be more biased.

- The model produces one word at a time, building sentences and paragraphs, and it usually produces what appears to be a rational text, but it sometimes moves in an incorrect direction. This is called hallucination.
- The information produced in a conversation may be correct, correct but incomplete, mostly correct, or completely off-base. The user needs to be aware of this.
- The user who is highly trained in the systems and writing prompts has a much better chance of getting good results.
- Organizations using these tools need to have an organizational policy about how they can be used and what quality controls are needed.
- Organizations using these tools need to be aware of legal requirements and consider such requirements in the development of policies. (The legal requirements affecting retirement administration are very complex because there are different agencies involved, there are different aspects of plan management being regulated, plans are tailored to fit the sponsoring organization, and individual situations are different.)
- These tools can support organizational decisions such as screening job or college applicants, determining claim payments, underwriting loans or insurance applications, choosing who will be promoted, etc. Bias is a big issue. The screening systems can build in bias, or they can eliminate bias. If the data used to train the model reflects bias, it can easily be preserved or even made worse.

One of the best resources that I found was "*The AI Revolution in Medicine: ChatGPT-4 and Beyond*".[1] This book explores the current and potential use of these models in medicine. I also attended a session at the 2023 Society of Actuaries ImpACT Meeting on the topic of AI and health care. From the book together with the discussions I listened to and my other research, I learned:

- These models do not have personal experience. They have been trained on specific data sets. They are limited by the data sets used in their training.
- These models should not be used as a substitute for professional opinion, but they can be a very valuable partner for the professional who understands their value and limitations. While providing medical care is different from managing retirement plans and retirement planning, both require a great deal of expertise and there are lessons about their use that can be transferred.
- These models can make a variety of mistakes, and so can humans who are working with individual situations. Challenges in building a good partnership are to improve the results and avoid mistakes. This is very important in medicine because some mistakes have dire consequences.
- AIs are not authorized to assume professional responsibility or make decisions that require professional responsibility.
- One of the immediately promising areas for improving the delivery of health care is the automation of the documentation of the patient visit and the patient record. Without such tools, the health care professional may spend much of the patient visit documenting rather than talking to and interacting with the patient. In the current environment, this is particularly important because medical professionals are often allocated a limited number of minutes to see a patient. Depending on the professional's skill, the documentation may not be very good. Special systems to support the interaction are under development and being used on an experimental basis by some health systems.

[1] Book published in 2023, by Peter Lee, Carey Goldberg, and Isaac Kohane. The book explores the use of AI in medicine and includes may conversations, showing how the prompts are developed.

- The tools for documenting the interaction could also be extended so that they could be used for ordering lab and other diagnostic tests and compiling the record. (Note that the same challenges regarding documenting interactions, follow-up steps, etc., exist in retirement planning and in interactions regarding plan management and administration.)
- AI tools can be used to review patient records and new lab tests and can provide information about possible diagnosis. This information should not be used to reach a diagnosis, but if used to create a starting point for the medical professional, it can help improve the diagnosis and reduce the time required to review the information. It may also go down a completely wrong track and distract the professional.
- In the case of rare diseases or unusual situations, it may be able to help the health care professional identify possible causes of a problem and lead them to information they would not have found otherwise. It can also completely mislead them and send them in the wrong direction.
- The book raised questions about individuals who do not have access to regular medical care and whether using AI to learn about their own situation is effective. (Worldwide there are many people without such access.) My understanding from the discussion is that using AI is definitely not a reliable way to get a diagnosis and treat a problem. However, the individual may be much better off than they would be without any information, and the chat may help them find a solution. In the last few years, many people have done internet searches to help them understand their own medical problems. These models produce faster answers than such searches, and the answers may be better or worse depending on the situation. It is very possible that more of them will be better. Either of the approaches to getting information supported by technology is better than simply talking to one's friends to see what they think the problem is and what to do about it.
- The potential exists to use AI as part of a strategy to expand access to care through a combination of personal research about the situation and about access to care, including telemedicine. Telemedicine can bring some care to individuals who do not have direct access to care.

My personal observations are that today with so many specialized medical professionals, they often focus on a particular problem rather than on the person. For example, when I experienced a wound on my leg, the wound care doctor focused on cleaning, scraping, and dressing the wound, but he did not seem to realize that there was a person attached to the wound. I was helped greatly by my primary care doctor and another doctor who looked at my overall situation and were able to make several recommendations to promote faster healing and better circulation by taking additional steps. With or without AI, it is critical to remember that health care providers should be dealing with the whole person and the conditions of their lives rather than with a single problem. This requires experience and personal judgement. This is also a key lesson for retirement planning.

MY CONVERSATION WITH ARTIFICIAL INTELLIGENCE

I decided to use ChatGPT to see what it would say about developing a policy for the use of ChatGPT.[2] I asked a series of three questions to help me understand the issues in general, see whether responses would change with the addition of fiduciary requirements, and to comments I provided on quality control over calculations. My conversation went as follows:

[2] I used ChatGPT3.5 for this conversation.

Question

"What should be considered in developing an organizational policy about the use of ChatGPT?"

ChatGPT Answer

"Developing an organizational policy about the use of ChatGPT involves careful consideration of various factors to ensure responsible and ethical use. Here are some key considerations:

Purpose and Scope

- *Clearly define the purpose of using ChatGPT within the organization.*
- *Specify the scope of its application and the departments or teams that can leverage it.*

Ethical Use

- *Establish guidelines for ethical use to prevent the creation or dissemination of inappropriate, offensive, or harmful content.*
- *Ensure that the use of ChatGPT aligns with the organization's values and ethical standards.*

Data Privacy and Security

- *Clearly outline how data generated or processed by ChatGPT will be handled, stored, and secured.*
- *Emphasize the importance of maintaining user privacy and complying with relevant data protection regulations.*

User Accountability

- *Define roles and responsibilities regarding the use of ChatGPT to ensure accountability.*
- *Provide guidelines for users to understand their responsibilities in generating and sharing content.*

Training and Awareness

- *Implement training programs to educate users on the proper use of ChatGPT.*
- *Raise awareness about potential biases in AI models and the importance of considering ethical implications.*

Transparency

- *Communicate transparently about the capabilities and limitations of ChatGPT to users and stakeholders.*
- *Disclose the use of AI-generated content when applicable to maintain transparency with users.*

Monitoring and Oversight

- *Establish mechanisms for monitoring the use of ChatGPT to identify any misuse or compliance issues.*
- *Implement oversight processes to regularly review and update the policy based on evolving best practices and ethical considerations.*

Legal Compliance

- *Ensure that the policy aligns with local and international laws and regulations related to AI and data privacy.*
- *Stay informed about legal developments that may impact the use of AI within the organization.*

Bias and Fairness

- *Implement measures to address and mitigate biases in AI models, including regular audits and assessments.*
- *Encourage a diverse and inclusive approach to data collection and model training to minimize bias.*

User Feedback Mechanism

- *Establish a feedback mechanism for users to report issues, concerns, or feedback related to the use of ChatGPT.*
- *Use feedback to continuously improve the system and update the policy as needed.*

Documentation and Versioning

- *Document the policy clearly and ensure that it is easily accessible to all relevant parties.*
- *Implement version control to track changes and updates to the policy over time.*

Collaboration with Legal and IT Teams

- *Collaborate with legal and IT teams to ensure that the policy is aligned with organizational goals and complies with legal requirements.*

By carefully considering these factors, organizations can develop a comprehensive policy that promotes responsible and ethical use of ChatGPT within their specific context. Regular reviews and updates to the policy are essential to adapt to evolving technologies and ethical standards."

Question

"Will the policy requirements change if the organization is subject to fiduciary liability?"

ChatGPT Answer

"Yes, the policy requirements may change if an organization is subject to fiduciary liability. Fiduciary liability involves the legal obligation to act in the best interest of others, often referring to the responsibility of managing assets or making decisions on behalf of others. In the context of using technologies like ChatGPT, where decisions and actions may have financial or legal implications, additional considerations may be necessary. Here are some specific aspects to consider:

Risk Management

- *Evaluate the potential financial and legal risks associated with the use of ChatGPT.*
- *Implement risk management strategies to mitigate and address fiduciary responsibilities, ensuring that the organization's interests align with those it serves.*

Compliance and Legal Obligations

- *Ensure that the policy is in compliance with fiduciary laws and regulations applicable to the organization's industry.*
- *Consider any specific legal obligations related to fiduciary duties and responsibilities when using AI technologies.*

Documentation and Record-Keeping

- *Maintain thorough documentation of the decision-making processes involving ChatGPT.*
- *Keep detailed records to demonstrate that fiduciary duties are being fulfilled in accordance with applicable laws and regulations.*

Due Diligence in Technology Adoption

- *Conduct a comprehensive due diligence process when adopting and implementing ChatGPT.*
- *Assess the technology's reliability, security, and ethical considerations to ensure that it aligns with fiduciary obligations.*

Transparency and Accountability

- *Emphasize transparency in AI-generated decisions and communicate the rationale behind decisions made with the assistance of ChatGPT.*
- *Establish clear lines of accountability to demonstrate that fiduciary duties are being upheld.*

Data Governance and Privacy

- *Strengthen data governance practices to safeguard sensitive information and ensure compliance with privacy regulations.*
- *Clearly define how AI-generated insights or decisions impact the handling of confidential or private data.*

Conflict of Interest

- *Address potential conflicts of interest that may arise in the use of ChatGPT.*
- *Implement measures to identify, disclose, and manage conflicts to prevent any compromise of fiduciary responsibilities.*

Continual Monitoring and Evaluation

- *Implement ongoing monitoring and evaluation processes to ensure that the use of ChatGPT aligns with fiduciary obligations.*
- *Regularly assess the impact of AI-generated decisions on the organization's financial well-being and compliance with fiduciary standards.*

Legal Counsel Involvement

- *Involve legal counsel specializing in fiduciary law to review and provide guidance on the organization's policies and practices related to ChatGPT.*

Organizations subject to fiduciary liability should integrate these considerations into their AI policies to meet their legal and ethical obligations. The specifics will depend on the nature of the organization, its industry, and the legal framework within which it operates. Collaboration between legal, compliance, and technology teams is crucial to ensure a comprehensive and compliant approach to the use of ChatGPT."

Question

"Retirement planning calculations include consideration of many factors and they extend the calculation over the lifetime of the individual. How can the user of ChatGPT to do a calculation confirm its accuracy?"

ChatGPT Answer

"When using ChatGPT or any AI tool for retirement planning calculations, it's crucial to confirm the accuracy of the results. Here are several steps and considerations to help users verify the accuracy of the calculations:

Understand the Model's Limitations

- *Be aware of the limitations of ChatGPT and the specific AI model being used. Understand the context in which it was trained and any biases it may have.*

Cross-Reference Information

- *Cross-reference the results generated by ChatGPT with information from reliable and authoritative sources. Use established financial calculators, retirement planning tools, or consult with financial professionals to validate the accuracy of the results.*

Use Multiple Tools

- *Don't rely solely on ChatGPT. Use multiple retirement planning tools, calculators, or financial planning software to compare results. Different tools may use different assumptions and methodologies, leading to varying outcomes.*

Check Assumptions

- *Verify the assumptions used by ChatGPT in the retirement planning calculations. Assumptions about inflation rates, investment returns, life expectancy, and other factors can significantly impact the accuracy of the results.*

Consult Financial Professionals

- *Seek advice from financial professionals such as financial planners, advisors, or accountants. They can provide personalized insights, verify calculations, and ensure that the retirement plan aligns with individual circumstances and goals.*

Review Input Data

- *Double-check the input data provided to ChatGPT. Garbage in, garbage out—ensuring accurate input is crucial for reliable output. Verify that all financial and personal information is entered correctly.*

Consider Sensitivity Analysis

- *Conduct sensitivity analysis by varying key assumptions to understand how changes in factors like investment returns, inflation rates, and retirement age affect the overall retirement plan. This can provide insights into the robustness of the calculations.*

Educate Yourself

- *Gain a basic understanding of retirement planning concepts and calculations. This knowledge will empower you to critically evaluate the results provided by ChatGPT and other tools.*

Evaluate Historical Performance

- *If the calculations involve historical financial data or market performance, verify the accuracy of the historical information used. Ensure that the AI model accurately incorporates past trends and events.*

Check for Updates

- *Ensure that the ChatGPT model is up-to-date. AI models can be periodically updated with new information, improvements, or training data. Using the latest version may result in more accurate calculations.*

User Feedback

- *Consider user feedback and reviews related to the specific implementation of ChatGPT for retirement planning. Feedback from other users may provide insights into the tool's accuracy and reliability.*

Continuous Monitoring

- *Regularly monitor and update your retirement plan. Financial circumstances and goals may change over time, and recalculating periodically helps ensure that the plan remains accurate and relevant.*

By combining these steps and being diligent in the verification process, users can enhance their confidence in the accuracy of retirement planning calculations conducted with ChatGPT or any AI tool. It's important to approach AI-generated results as one part of a broader financial planning process and not rely solely on automated tools for critical financial decisions."

GOING BEYOND THE CONVERSATION

I was impressed by the responses and felt that they were as helpful as summarizing the webinars I listened to or what I read.

I also learned that these systems could perform a wide variety of tasks, and that specialized training would be needed to do these tasks well. I decided to focus on two issues that I have focused on for many years:

- Who is responsible for the results of an actuarial calculation or the accuracy of a financial statement?
- Quality control and confirming whether results are accurate or not.

Both actuarial calculations and financial statements are subject to a variety of legal and professional guidance. For example, the actuarial profession has guides to professional conduct and standards that apply regarding who can sign an actuarial opinion and what must be considered in the preparation of that opinion and disclosed with that

opinion. The accounting profession is regulated in the U.S. by the FASB and GASB, and internationally by the international profession.

In setting guidelines and policy for the use of AI, it must be remembered that specialized functions require a professional certification or signature by a person who has the qualifications and/or licenses required, and that person must know which rules are applicable and follow them. Specialized software can be used to help with many of these functions, but it must be designed to match the requirements. Large language models are not currently built to do this, and information produced by such systems still requires human checking and sign-off.

Quality control is a very important part of professional work and of retirement management. I thought the responses about quality control made sense.

Quality control over actuarial calculations requires answering a number of questions:

- Was the right process or calculation used to solve the problem? How do we confirm what was done?
- Did the overall process provide for considering risk and variations due to changes in economic conditions, etc.?
- Was the correct data used as input?
- Were the assumptions used appropriate and did they follow the applicable guidelines, if any?
- Did the appropriate party sign off on the calculation?
- Will the user be able to understand the results?

I have been involved with the financial security system for more than 60 years and have seen the evolution of different methods of doing the calculations, and of more sophisticated calculations over time. The following exhibit provides some perspective on changes in quality control and its evolution over time.

Exhibit 1

EVOLUTION OF CALCULATION METHODS, TECHNOLOGY AND QUALITY CONTROL

Method of Calculation	Use of Technology	Quality Control
Calculations done manually one by one.	Hand calculators and slide rules.	Individual doing the calculation was expected to know what to do. Individual items in the calculation were done by two people – one who did the calculation and one who checked. Actuaries active at that time were very good at reviewing results for reasonableness. They also were skilled at comparing results for repeating calculations and determining the factors that were driving change using a gain and loss analysis.
Actuarial valuations were done individual by individual.	Computer programs were written to look at the characteristics of the individual and the benefit and calculate liabilities person by person.	Programs and calculations were checked. Test cases were used to test programs. Programs often were written for each situation. The responsible signing actuary was involved in doing or overseeing the calculations. Reasonable checks and year by year comparisons were also used.
Valuations were done with programs which could do many plans by coding parameters.	Companies secured programs that were pre-packaged and needed parameters coded to do the valuations. They were first used on main frame computers and then on both main frames and personal computers.	Organizations using a program needed to make sure it was appropriate. Test lives were used to check results. It was important to design a test sample. An actuary needed to confirm that the program matched the plan provisions or make suitable adjustments. Signing actuary needed to confirm that everything had been done correctly.
Personal retirement planning software focused primarily on period to retirement and on what percentage to save – it was deterministic.	Various software packages were available through financial service companies and in the public domain.	It was up to the user or professional assisting in planning to determine suitability and correct use. The formulas and method of entering assumptions were defined in the program. The user was responsible for assumptions.
Plan forecasts often were done using stochastic projections. Stochastic projections were used in personal planning software.	Software packages and higher speed computers were used. Stochastic results are more difficult to understand, particularly by unsophisticated users.	It was up to the user or professional assisting in planning to determine suitability and correct use. The formulas and method of entering assumptions were defined in the program. The user was responsible for assumptions
AI for various purposes.	The process depends on large data sets and fast technology and works differently. The process does not define a calculation in the same way that earlier processes did.	This is the responsibility of the professional user. How it will be done is to be determined. This is a major challenge. The AI response suggested using several different tools.This is an interesting (and time consuming) idea.

Source: Author's analysis

As stated earlier, quality control has always been an important issue. It has grown more complex with more sophisticated technology, larger data sets and new processes. At the same time, professionals today often have much less hands-on experience, may be less focused on reasonableness checks, and less able to have a sense about it when something is not correct. They may also be relying on others to choose software or decide on the software to be adopted by their organization. This area needs more research.

CONCLUSIONS

The introduction of the most recent version of AI created the potential for many tasks to be done differently than they were before. This will continue to change. They create the potential for skilled individuals to multiply what they can get done by using these tools. They also may lead to potential pitfalls and errors. Related legal issues and the potential for more regulation can complicate the situation.

Professionals and organizations doing management of retirement plans or offering retirement planning have opportunities to use these tools, but there are questions about what is reliable as well. It is essential that organizations and professionals examine the issues and develop policies and implement policies regarding the use of these tools in order to use them prudently. The policies must consider issues such as what customers need, what is reliable, legal issues, data privacy, etc.

Published research provides information about the way the use of these tools can improve the practice of medicine by operating in a partnership mode. It also identifies specific areas where there is the greatest chance for help in the short term. The research further provides information about the risks involved with AI.

In addition, research regarding health care provides a model that can be extended to apply to retirement management and planning. These are a few examples of how AI may fit into retirement planning and planning management:

- AI can be a very good partner to professionals involved in retirement planning and management tasks.
- As a partner, AI can help document discussions and improve the client situation.
- AI can help do research and literature searches on a variety of questions. The AI tool used needs to be up-to-date and have access to the internet if it is going to assist in these tasks.
- Planning involves many tasks. Research is needed to establish which tasks AI will be able to help with and what is needed for quality control.
- It is known that AI is sometimes wrong, and it provides incomplete answers at other times. It does not think but rather it reflects the information in the data set it learned from. It is essential that the user of AI be able to apply judgement to the situation and sort out what is helpful and what is not.
- While AI can be used to do computations, it is not well enough tested and developed at the time of this writing to be used for computations that affect plans and personal decisions. There are too many opportunities for errors. See my other essay in this essay collection for an example of a calculation gone very wrong with AI. Traditional software packages should be continued and enhanced. It should be noted that there are many packages and questions to be answered . It is a complex task to match the right calculation to the right situation. Further research is needed on whether AI can help with this.

We must not forget that AI cannot assume the responsibility of a signing professional or complete the tasks the professional does. Continued research as AI evolves will however likely enhance the support it can provide to these processes and help them become more efficient.

* * * * *

Anna M. Rappaport, FSA, MAAA is a phased retiree and president of Anna Rappaport Consulting. She can be reached at anna.rappaport@gmail.com.