Artificial Intelligence for Analysis: The Road Ahead

Dennis J. Gleeson, Jr.

Today's chatbots are not intelligent, but they are innovative, exciting, and full of potential in the context of the volumes and varieties of information the IC collects, processes, triages, and uses in support of its global mission.

Chatbots like OpenAI's ChatGPT, Google's Bard, and Anthropic's Claude provide us with interesting and exciting new ways to interact with information. These products respond to users queries by transforming a statistical analysis of patterns existing in a large amount of information—a large language model (LLM)—into a natural language response that mimics human intelligence. *Mimic* is the key word here: these platforms do not understand the data they are analyzing and interpreting in the same ways that people do.

The problem that these products represent for sophisticated consumers of information, such as analysts, academics, and journalists, lies in their design: to date, LLMs preclude insight into or an understanding of the basis for the answers they generate. Users are being asked to trust the technology, but they are not given the opportunity to verify the way the underlying algorithms weigh information (or even what information is, or is not, being used in the formulation of answers). In short, both the "dots" and the connections between those dots exist within a black box at a time when organizations like the IC continue to work toward greater transparency about the underpinnings of their judgments and actions.

The opportunity in front of us lies beyond the words often used to describe these technologies. The idea of developing artificial intelligence dates back to the 1940s and 1950s. Today's chatbots are not intelligent, but they are innovative, exciting, and full of potential in the context of the volumes and varieties of information the IC collects, processes, triages, and uses in support of its global mission. The challenges and opportunities for organizations looking to implement generative AI (GenAI) start with the breadth, depth, richness, and cleanliness of the data itself.

Why GenAI Isn't Enough

GenAIs appear to be most successful when the scope of user requests, the rules around requests, and users' expectations roughly align. If I ask ChatGPT to write a poem in French about frogs wearing hats, it does so. My objective is loosely defined but is specific enough for the algorithm to produce a passable response because the rules around the request, while not explicit, can be inferred. In other words, frogs wearing hats are the subject of an undefined narrative and, while there are several styles of poetry, a rhyming scheme is a reasonable place to start; specifying the type of poem—ode, haiku, limerick—would further help the AI understand the task and rules around the task.

At the other end of the spectrum is a task that is as likely to be defined by the rules as it is by the objective.

AI Terminology

When prompted, Google Bard offered these definitions:

Artificial general intelligence is a hypothetical type of AI that would be as intelligent as a human or even more intelligent. It would be able to learn and adapt to new situations and to perform any intellectual task that a human can.

Generative AI is a type of AI that can create new content, such as text, images, and music. It does this by learning from existing data and then using that knowledge to generate new outputs that are similar to the data it has seen.

The main difference between AGI and GenAI is their scope of capabilities. AGI is designed to be intelligent in a general sense, while GenAI is designed to be creative and generate new content.

If I ask an AI to write the Python code to perform an analytic task using a dataset I am familiar with, AI is almost certain to accomplish the task far better than I—a nondeveloper. That said, despite my lack of technical coding expertise, I would assess the AI's efficacy against what I expected or suspected the answer should be, based on my understanding of the data.

The expectations of analytic users are the admittedly imperfect bars against which experts are likely to judge AIs: does the AI they are using generate responses consistent with their understandings of the issues? This is the fundamental challenge with trying to apply GenAI to (qualitative) analyses: analytic users have much more than general understanding of the domain in which they work. Based on my experiences with an earlier generation of commercial AI, the first questions analysts are likely to ask an AI are ones they know the answers to. If the AI fails to answer in a manner consistent with their expectations, they'll judge the AI—rightfully or wrongfully—as not being ready to support them in their work.

A simple test for this is to take an issue that you are aware of and ask an AI about it. Is the AI's answer consistent with your understanding of the issue? Is the AI's answer at least as informative and detailed as the Wikipedia page on the topic? If the answer is no, legacy approaches to search and discovery—such as Boolean queries—will persist, even as we know they are not up to the challenges and opportunities presented by big data.

Experts: A User's Guide

When I left government service, a friend who had never worked in the IC asked, "How do analysts think?" My response was a snarky but honest "idiosyncratically." The style of their thinking is a function of factors like temperament, education, and experience. In short, the idiosyncrasies in how they think contribute to the collective insights of an analytic cadre.

As individuals, analysts tend to be intelligent, inquisitive, insightful, and tenacious. They are likely to hold graduate-level degrees in fields related to their work. As communities of experts, they hold themselves and their peers to incredibly high standards because their work has real-world implications: During their training, analysts are taught not just about good analytic tradecraft, but about intelligence failures, their causes—notably information gaps, issues with sourcing and information quality, and logical fallacies—and the human—and reputational—costs of failure.

Analysts will hold AIs to the same standards they are held to and to the same standards they apply to their colleagues across the IC, government, think tanks, academia, and media. The basis for and the sourcing behind an assertion, and the confidence they have in that assertion, are every bit as important as the answer itself. For this reason, AI is going to have to be sufficiently transparent and explainable for analysts to take it seriously.

To understand analysts, it's worth starting with Philip E. Ross, contributing editor for *Scientific American.* In a 2006 article on expert minds, Ross observed:

> [K. Anders Erricson of Florida State University] *also cites studies of physicians who clearly put information into long-term memory and take it out again in ways that enable them to make diagnoses…. The researchers explained these findings by recourse to a structure they called long-term working memory, an almost oxymoronic coinage because it assigns to long-term memory the one thing that had always been defined as incompatible with it: thinking. But brain-imaging studies done in 2001 at the University of Konstanz in Germany provide support for the theory by showing that expert chess players ac-*

tivate long-term memory much more than novices do.[a]

This description of long-term working memory struck me because it reflects my experiences as an analyst and as part of a community of analysts: the analyst's power of recall often is uncanny.

As a result, this is why GenAIs are likely to be somewhat limited in intelligence applications: users will be experts, not generalists. They will be up-to-date and have historical understanding of the issues they and their colleagues work on. To work alongside and in support of analysts, GenAIs will need to not just use the information that informs analysts' understanding of current events, but also be aware of the evolution of analytic lines. As a result, the first step for analytic organizations is focusing less on the AI and more on knowledge and insights they and their organizations have generated.

Toward Responsible AI

In 2019, Alejandro Barredo Arrieta and his colleagues published "Explainable Artificial Intelligence (XAI): Concepts, Taxonomies, Opportunities and Challenges toward Responsible AI," which explained that one of the dangers of machine learning systems is "creating and using decisions that are not justifiable, legitimate, or that simply do not allow obtaining detailed explanations of their behaviour." [*Information Fusion* 58 (June 2020): 82–115] Echoing Arrieta, et al., the Office of the Director of National Intelligence in its *Artificial Intelligence Ethics Framework for the Intelligence Community* (June 2020) sees explainability as an element of transparency. Unfortunately, the Center for Strategic and International Studies in its Tech Recs database notes there has been no progress on creating standards for explainable AI (https://techrecs.csis.org/, accessed September 27, 2023).

Institutional Knowledge

Today's AIs are still in their infancy; the volume and variety of research, development, and refinement around the technology are staggering. Wanting to adopt AI makes sense for organizations like the IC's all-source analytic elements because they have been awash in information since the advent of the internet. The problem is, at this stage, the best AI that exists at the time of acquisition and implementation might be embarrassingly unsophisticated one year later. From an enterprise perspective, chasing "the best" AI is likely to be a Sisyphean task. There are several steps the IC can take.

Clean and enrich the data

AI is a tool. Insights and information are assets. The quality of an organization's information is all but certainly going to affect the sophistication and success of its efforts to implement and use artificial intelligence. Cleaning and enriching the data isn't glamorous, but it is essential. For the software engineer, data cleanup and enrichment are nowhere near as shiny and exciting as working on user-facing products and services. For analysts, data clean-up lies far outside of the work on which their performance is evaluated. The job of data engineer is coming into vogue but if that person does not have a substantive understanding of the domains in which analysts work or of the value of the work of analysts to their organizations' customers, it is unlikely they will be able to unlock anything approaching the full potential of an organization's information. Why? Tacit knowledge.

Leverage tacit knowledge

Tacit knowledge—defined as "skills, ideas and experiences that are possessed by people but are not codified and may not necessarily be easily expressed"—cannot completely elude capture in an enterprise setting: metadata, audit data, and knowledge management tools (e.g., filing, tagging) can serve as the basis for novel weighting systems.[b] Analysts weigh information intuitively as part of their workflow: they open, read, tag, file, and use only the documents they think contain insights that are the most relevant to their work.

The documents analysts produce and the talks they give are not just explicit knowledge. This output should also be used to train AIs on how experts think and how that thinking has evolved over time: an unexplained assertion today masks the evolution of the thinking that underpins that assertion. To intellectually curious analysts (or to the policymakers and decisionmakers they serve), an unexplained assertion—even if seemingly reasonable—might as well have come from a fortune cookie. Experts need to be able to understand the bases on which AIs make their assertions because their reputations and the reputations of their organizations lie

a. Philip E. Ross, "The Expert Mind," *Scientific American*, August 1, 2006.
b. Ritesh Chugh, "Do Australian Universities Encourage Tacit Knowledge Transfer?" *Proceedings of the 7th International Joint Conference on Discovery, Knowledge Engineering and Knowledge Management* (2015): 128–35.

AI is not a quick fix to the challenge of the big data that is part and parcel of our professional and personal lives. Rather, it is a strategic shift in how we think about interacting with massive volumes of data.

in their ability to explain not just the what but also the why.[a]

Realign corporate structures

In the last decade, we saw the rise of data scientists and chief data officers (CDOs). With AI, we're seeing the rise of prompt engineers—specialists in creating text that generative AI can interpret and understand—but we also might need to reconsider the idea of a chief knowledge officer. Knowledge is more than the documents in which insights and information are contained. Information technologists and engineers are unlikely to be familiar with the substantively meaningful contours of the information analysts use, the tradecraft analysts employ, or the prompts that are most likely to produce the responses and user experiences that will resonate with analysts and their customers. The chief knowledge officer could be the voice of both the user and customer by working with the CDO and the chief technology officer to develop the strategies most likely to transform enterprise data holdings—raw information through finished intelligence—into knowledge that can be accessed by and interacted with through AIs as appropriate.

Rethink analyst responsibilities and reimagine their training

CIA often looks back to Sherman Kent as the "father of intelligence analysis." Kent died in 1986, years before the internet first became commercially available, decades before the language of big data and data science entered our vernacular. What is as true now as it was when Kent was alive is that analysis is more an art than a science: Uncertainties, variables, and information gaps still persist, now compounded by misinformation, disinformation, and oceans of low-value information. The practice of intelligence analysis evolves as customer needs and producer capabilities evolve (or fail).

It is worth noting that AI is not new to the IC: In-Q-Tel made a strategic investment in Primer in 2017. AI today, however, seems different: the breadth, depth, and pace of change is dizzying. Despite this, the Pew Research Center in May 2023 reported that only 14 percent of Americans had used ChatGPT.[b]

As a result, the first step toward adoption might simply be to expose analysts to AIs in training classes. There they can begin to understand, at a conversational level at the very least:

- the design considerations and limitations of the systems they are expected to use or interact with;
- large language models (or the successors to LLMs); and
- design and prompt engineering as functions of analytic methodologists.

In the classroom they might begin to learn how to map out their tacit knowledge and determine how it might be made explicit, and they might discuss the broader ramifications of AI for their profession.

I expect there will be a tsunami of open-source information, disinformation, and low-quality information that might, at first blush, seem passable. While the demand for open-source information is voracious and rarely satisfied, sober assessments about the nature of sources will be more important than ever. In this, the work of public benefit corporations and nonprofits like Ad Fontes Media or Truth in Media is critical to ensuring that experts have a deep understanding of the reliability and biases of the sources that they use to craft their analysis.

Insights and Knowledge First, then Artificial Intelligence

There's good reason to be excited about AI and its potential applications in the IC and in support of national security. Satisfying a global coverage mission across multiple mission areas and numerous analytic disciplines areas requires triaging and making sense of volumes of information that scale well beyond human capacity. GenAIs like ChatGPT represent promising and exciting alternatives to search even as they fall well short of

a. For an examination of tacit knowledge in a historical setting, see Michael Aaron Dennis "The Less Apparent Component Tacit Knowledge as a Factor in the Proliferation of WMD: The Example of Nuclear Weapons" in *Studies in Intelligence* 57, no. 3 (September 2013). At the time of publication, Dennis was adjunct lecturer in Security Studies at Georgetown University's Edmund Walsh School of Foreign Service. His article would be selected as a Studies Annual Award winner in 2013.
b. Emily A. Vogels, "A Majority of Americans Have Heard of ChatGPT, but Few Have Tried It Themselves," Pew Research Center, May 24, 2023, www.pewresearch.org/short-reads/2023/05/24/a-majority-of-americans-have-heard-of-chatgpt-but-few-have-tried-it-themselves/

the professional standards to which members of the IC hold themselves. AI is not a quick fix to the challenge of the big data that is part and parcel of our professional and personal lives. Rather, it is a strategic shift in how we think about interacting with massive volumes of data.

If we start with that task—finding relevant information and insights in what are currently overwhelming volumes of information—the implementation of any given type or brand of AI is not nearly as important as preparing the data. Vendors will rise and fall; enterprise data holdings are a constant of sorts. They represent both the explicit knowledge of the organization (as captured in their written products) and the source material that informs their sense of the world and the trends most likely to affect the world's trajectory.

Recalling the old saw about data scientists spending 60 percent of their time cleaning data, we should be asking if data cleanup and enrichment are getting 60 percent of the resources being devoted to acquiring, implementing, operating, and maintaining AI. The answer probably falls between a deafening silence and a resounding no.

The quality of AIs will improve. Brands will come and go. What an organization—be it CIA, the broader IC, or any one of the thousands of other knowledge- or information-based organizations around the world—knows is the constant.

Data cleanup and enrichment are grueling and thankless tasks even if they are the most likely means of enabling better outcomes using AI. Working on user-facing applications is usually more appealing to technologists; transforming massive amounts of information into clear and compelling narratives is the job of analysts. Clean, enriched data are essential to the success of both parties but, all too often, it is the primary responsibility of neither.

Rather than try to apply a brand of AI to all enterprise data, it may be best first to go through a period of A/B testing of various AI offerings, using as the key variable the quality of data each AI is asked to consider. For example, in Test A the data examined should be data as it exists in enterprise data holdings at the time of the testing. In Test B, the tested data would first have been cleaned and enriched in ways that seem to allow for maximum analytic flexibility (irrespective of the brand of AI being tested or used). Each data set would be queried by the same brand of GenAI. This could help tell us which data set more accurate, less hallucinatory results; how the various AI tools compare; and where we need to enhance data quality and richness.

Today's AI is in its infancy. It is exciting and promising, yes, but it is immature. There is an incredible amount of research, development, testing, and evaluation being done to improve the capability and quality of AI. The quality of AIs will improve. Brands will come and go. What an organization—be it CIA, the broader IC, or any one of the thousands of other knowledge- or information-based organizations around the world—knows is the constant. How any organization processes, structures, cleans, and enriches its knowledge is likely to be the key determinant in its success with using AI in support of its missions both today and tomorrow.

❖ ❖ ❖

The author: Dennis Gleeson is a former director of strategy for analytic systems at CIA. He frequently writes about artificial intelligence on Substack.