

The Future of ChatGPT in Market Research

Some Road Hazards We've Encountered and a Few Tips for Safe Driving

By Susan Schwartz McDonald, Ph.D. and Mike Kelly, Ph.D.

Since its debut last November, OpenAI's ChatGPT has dominated the news to an extent unmatched by any previous AI initiative. A wide variety of implications have been discussed, ranging from the *practical* (how do we grade students, which occupations and which tasks are most at risk of AI replacement?) to the *philosophical* (have computers finally achieved sentience?). Literally *everyone* is talking about it, in every walk of life.

The way ChatGPT dominates our collective conversation attests to the momentous leap it represents – specifically the versatility to deliver impressive performance on a variety of demanding intellectual tasks. AI scientists have been exploring and extending these capabilities for several years, earning only modest notice outside the field or attention in the popular press before now. What accounts for this year's Big Bang?

What suddenly changed was not so much the AI model itself – the “GPT” in “ChatGPT” – but the integration of the GPT engine into a hot race car with “Chat-power steering.” Prior to ChatGPT, programming skills were needed to work with GPT or similar models like Google's BERT. ChatGPT stands out as a key breakthrough because it democratized the underlying AI model, allowing anyone to drive it merely by typing questions or instructions into a Web browser. All that raw power in the hands of ordinary users promises to unleash vast potential in nearly limitless applications – and also, as anecdote suggests, the potential for embarrassing error when ChatGPT gets carried away and makes stuff up. So what can we expect from ChatGPT in the insights field?

After kicking the tires on ChatGPT over a period of months, we've repeatedly stubbed our toe on some fairly basic applications, signaling that we need to temper the excitement and drive with caution. ChatGPT is under-proficient for some tasks and for others, curiously over-powered. The version we're all road-testing is still in its early stages but *for now*, ChatGPT can be a bit of an idiot-savant. To use it effectively, we need to understand and work around all its current limitations. So far, we've encountered at least two snags in putting ChatGPT to work for market research applications: (1) inability to replicate its own answers to some questions and (2) failure to follow instructions consistently.

Application: Using ChatGPT to Gauge Brand Associations

Glitch Uncovered: Lack of Consistency

ChatGPT has been engineered to produce the sort of response variability needed to generate rich, human-like speech. But that capability, while critical to performing some tasks, can significantly handicap it in performing others.

Case in point ... In brand equity surveys, we routinely ask respondents to rate how closely they associate attributes with various competing brands. Curious to see what ChatGPT would do with a similar task, we asked it to rank specific digital brands based on degree of association with pre-specified words like “innovative,” “responsive,” etc. Given the tremendous amount of text used to train the AI engine under ChatGPT's hood – literally hundreds of billions of words – we'd expect it to have learned how various companies and brands are perceived in our culture.

And, indeed, ChatGPT readily ranked four major brands on all the dimensions we supplied. Critically, however, when ChatGPT was asked to repeat the ranking exercise again and again with the same prompt, it provided different answers each time. Since replicability is one of the most valued coins in the research realm, a pattern of inconsistency requires that we dig deeper into the source, consider it in the context of other forms of sampling error we routinely encounter, and look for ways to constrain or reduce it.

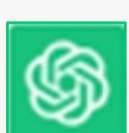
The ability to generate varied responses is deep in the DNA of GPT. By nature, large language-based AI models don't “know” things, they predict them. Without the freedom to improvise, Generative AI could not dazzle us with its endlessly creative and proliferative use of language. But improvisation has its downsides.

Mindful that some tasks require consistency rather than variability, the developers incorporated a parameter into the platform design called “temperature,” which users can adjust to dampen (but not fully eliminate) variability in responses to the same question. In theory, then, ChatGPT can be constrained by user-guided temperature settings, but experience shows again and again that it's not yet an obedient genie – it's more of a rebel. In fact, ChatGPT has a maddening tendency to ignore instructions that it has “agreed” to, and apologize when it doesn't, without necessarily guaranteeing to behave itself. Humans have programmed it to be polite but not always compliant, and for reasons unknown to all, it can misbehave even where no purpose is served. Here's an example.

User instruction



Please set temperature to 0 for this session unless I state otherwise.

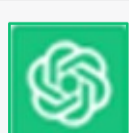


Understood. I will set the temperature to 0 for this session, unless otherwise stated. Using a temperature of 0 will result in more deterministic and focused responses.

Various Q&A interactions later...



Did you use a temperature of 0 for the last response?



Apologies for the oversight. No, I did not use a temperature of 0 for the previous response. I will ensure to set the temperature to 0 for the remaining responses in this session.

Application: Leaning on ChatGPT to Expedite Programming

Glitch Uncovered: Sloppy Work

NAXION data programmers have turned increasingly to ChatGPT for assistance writing code – and we anticipate that this will eventually be a huge time-saver. But here too, we have observed a maddening refusal to follow instructions and a tendency to be obtuse when explaining what it has, in fact, done.

Case in point ... ChatGPT will, for no apparent reason, feel free to rely on older, deprecated code that no longer works in the latest version of a programming language. When confronted with the error message we receive when trying to run outdated code, ChatGPT politely acknowledges its error and then does use the correct up-to-date code, but there is no offsetting utility that accounts for its failure to do what was asked in the first place. This sort of misbehavior puts us at risk and

makes us reluctant to use it when there are no built-in checks on its performance. The benefit of doing so is reduced if some of the time saved is spent checking ChatGPT's work.

NAXION data scientists are addressing these sorts of issues by writing code to control the GPT engine directly, bypassing Chat steering altogether. Our experiences to date have led us to conclude that the current Chat interface does not let us fully see and direct the engine under the hood. Neither does it adequately control GPT for applications where replicability is a high priority and gratuitous variability is unwelcome.

As a result, we've resorted to taking direct control from the Chat interface. For instance, we're tackling the replicability issue in much the same way we might address response variability through survey sampling. A standard way to address response variability in research surveys is to ask the same question of many participants and average the results. Similarly, we can ask GPT to rank brands for their similarity to attributes like "innovative" *hundreds of times*, using the average rankings in our analyses to stabilize the output. An important feature of this approach is that, since GPT has no "memory" of each individual trial, its responses across repeated queries are completely independent, though if the model is working properly, they are presumably similar. Data scientists can also constrain unwelcome variability by verifying that the temperature is set to zero each time, avoiding the hazards of Chat "disobedience."

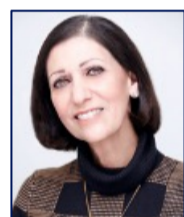
Going further, we are drawing inspiration from an emerging field called "prompt engineering" to find optimal ways of asking AI models to perform desired tasks as intended. GPT is still very much in training, and like any model, it tends to perform better when given examples of the type of output we're looking for. But because it comes fully pre-trained, unlike earlier generation AI models, we cannot teach it based on its mistakes. We can only give it better direction. We anticipate that our ability to nudge and tweak GPT will increase once newer models with expanded capacity are released.

ChatGPT deserves credit for creating an extraordinary appetite for AI applications and putting powerful models in the hands of the average "driver" – but for professional applications, driving them is harder than it looks. And mistakes can be dangerous. A more advanced interface than Chat is going to be needed for at least the foreseeable future as we learn what GPT can productively do, where greatest net gains can be achieved, and what it probably shouldn't be asked to do. *Meanwhile, for those in the insights field, we recommend bypassing Chat, checking GPT's work, and developing protocols that maximize consistency of results.* We also recommend a consistent program of experimentation to take measure of the challenges. Companies seeking research partners who promise to use AI should query them to ensure there are experienced drivers behind the wheel. Agencies, in turn, will need miles of test drives to run these models both productively and safely.

Here are a few things to keep in mind when using ChatGPT

- It's a language-based model – an "A-student" in English but unschooled in math. While it can be directed to make use of conventional calculation programs, its performance will not display the same kind of virtuosity using borrowed math skills.*
- As a language major, it displays impressive ability to develop codes for text analytics and accurately classify responses – but it currently has serious capacity limitations, making it ill-equipped to process a lot of complex language data (like full transcripts) or use verbal data to create segments.*
- GPT does not currently scrape the web looking for new "facts." The information it uses to make predictions reflects only what was known or published during its training period, making it ill-suited to track the evolution of attitudes since then.*
- While GPT outputs will, in some sense, reflect the many cultural voices captured in its training corpus, the approach we've taken of averaging its responses to the same question does not equate to survey sampling of a consumer population. It is just a form of repeated measures "within subject" to achieve answer convergence.*
- GPT needs a watchful eye. A stunning capacity to improvise and simulate human thinking gives it remarkable flexibility and vast potential – but like humans, it errs.*

About the Authors



Susan Schwartz McDonald, Ph.D.
CEO, NAXION
215.496.6850
smcdonald@naxionthinking.com

Susan's career focus has been on the development and protection of robust brands, and the research methodologies needed to support them. She has contributed to the evolution of many standard research techniques, and she writes frequently on industry topics and issues of broader interest. As someone who has dedicated much of her career to language, she is intrigued by the potential of large language models to enrich the mining of market insights and is spearheading the firm's ambitious initiatives in the development appropriate applications. Her commentary on language and culture can be found at www.smartmouth.blog. Susan holds MA and PhD degrees from UPenn's Annenberg School of Communication.



Michael Kelly, Ph.D.
VP, Information Technology Practice
215.496.6842
mkelly@naxionthinking.com

Mike designs and manages major engagements for clients seeking to develop B2B and B2C business strategies based on customer insight and advanced market analytics. Leveraging his skill in devising new modeling techniques and integrating multiple data streams, Mike has helped clients in Information Technology, Energy, Consumer Electronics, and Manufacturing optimize pricing strategies, guide product bundling, sharpen targeting, and prioritize service improvements that drive customer loyalty. Mike has notable subject matter expertise in the fields of cognition and computational linguistics, where his highly regarded academic work has been a platform for innovation on behalf of clients, while building the firm's intellectual capital in advanced methodologies.

About NAXION

NAXION is a nimble, broadly resourced boutique that relies on advanced research methods, data integration, and sector-focused experience to shape the destiny of brands. Our century-long history of innovation has helped to propel the insights discipline and continues to inspire contributions to the development and effective application of emerging data science techniques. For information on what's new at NAXION and how we might help you with your marketing challenges, please visit www.naxionthinking.com