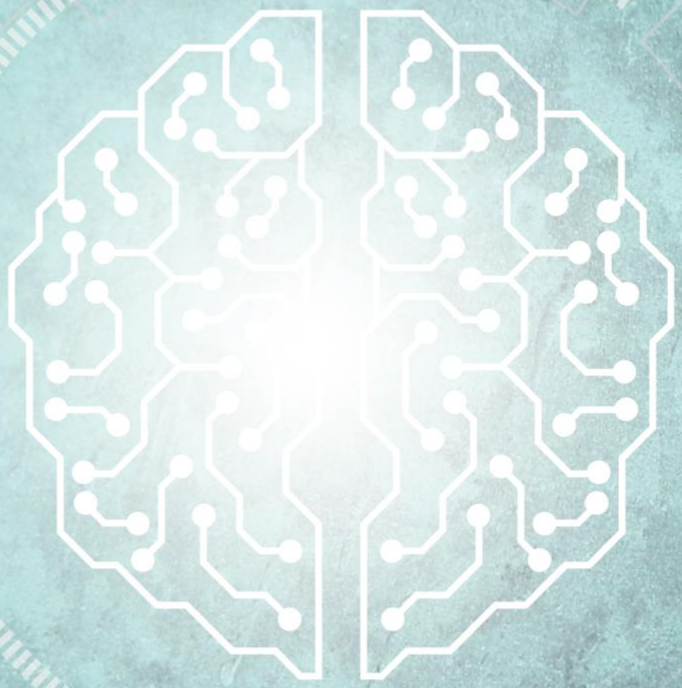
White Paper

INSIGHTS ON GENERATIVE AI

June 2023



ELDER RESEARCH

DATA SCIENCE · AI · MACHINE LEARNING

Table of Contents

1.0 Introduction

Generative AI has recently broken through into the public consciousness with applications involving Large Language Models[1] (LLMs) and image-generation models.[2] These technologies have been developed in the open over the past several years, but in the past year they have achieved new levels of capability. With tools like DALL-E and ChatGPT providing new, user-friendly interfaces to generative AI, these applications have captured the public's attention and are even beginning to be deployed at scale.[3]

As these new generative AI applications shift the AI landscape—and impact the public's acceptance of AI-powered tools—organizations are beginning to investigate the benefits and risks of these innovative applications in business settings. This document highlights some of the current benefits and risks associated with business applications of generative AI. Because it is impossible to list all the potential benefits of a new technology, we instead concentrate here on a few early applications that enhance or speed up existing workflows. Following that, we outline four potential risks to business: legal risks, privacy and data-governance risks, factuality and accuracy risks, and model-ownership concerns.

We encourage organizations to actively explore these technologies. Not only are there long-term benefits to keeping up with this new set of capabilities, but current-generation generative AI applications can provide value to business *right now*. Of course, at the same time it is crucial to evaluate and apply these tools judiciously to protect personnel, businesses, and brands from unanticipated consequences.

As models proliferate and new interfaces or applications enable new uses, the generative AI landscape should continue to expand rapidly. The next generation of generative AI technologies will likely be even more fascinating than the ones we are so excited about right now. Investigating the benefits and risks of these technologies now—especially seeking out "convex" opportunities, where potential harms are limited but potential gains are not—will win organizations a practical understanding of generative AI's practical benefits and risks and lay the groundwork, too, for understanding and applying the next wave of generative AI, whatever it may be.

---

[1] Wikipedia contributors. Large language model — Wikipedia, the free encyclopedia. (2023).
[2] Wikipedia contributors. Text-to-image model — Wikipedia, the free encyclopedia. (2023).
[3] Hu, K. ChatGPT sets record for fastest-growing user base - analyst note. (2023).

2.0 Potential Benefits of Generative AI

Because generative AI applications like ChatGPT have arrived so recently, and because these tools and the models that underpin them are still changing rapidly, it is difficult to construct a comprehensive list of the technology's potential benefits. (We find it easier to reason about the potential risks of generative AI applications; many such risks are tangible and can be extrapolated from previous machine learning and internet-scale applications.) And, because leaders like OpenAI are providing general-purpose platforms instead of single-purpose models, it is fascinating to watch the invention of diverse applications and tools for interacting with these platforms. These new applications might depend on generative AI models in the background but only to provide some other tangentially related service. In this section we highlight a few applications; cases where large language models are being used either directly or as part of a larger workflow to perform useful work.

2.1 Select Generative AI Categories

We might categorize these applications as:

(1) general workflow augmentation

(2) knowledge retrieval

(3) image generation

(4) coding assistants

2.1.1 General Workflow Augmentation

First, we are seeing a proliferation of applications in the "workflow augmentation" domain as people experiment with generative AI in the course of their work. One instructive example combines language models (ChatGPT) and image models (Stable Diffusion) to create marketing emails, social media campaigns, and even build websites.[4]

2.1.2 Knowledge Retrieval

Second, LLMs provide a natural way for users to search databases and interact with document collections.[5] By providing the contents of relevant documents in language-model queries, or by updating LLMs incorporate relevant company information, we can request that these applications to query and connect internal knowledge sets in novel ways. Combining company information with these

4 Mollick, E. Superhuman: What can AI do in 30 minutes? (2023).
5 Boegner, M. Knowledge retrieval architecture for LLM's (2023). (2023).

models' trained abilities to retrieve conceptually similar information provides new interfaces to often-siloed company data, and this type of fine tuning makes the models' intrinsic querying capabilities accessible to businesses' own data sets.

2.1.3 Image Generation

Third, image-focused applications are rapidly becoming capable of producing convincing images and videos starting from only a text prompt or other contextual information. As one example, there are already multiple applications purporting to generate business headshots, and some of these are at the very least competitive with the real thing. Many are also finding that general-purpose tools like DALL-E,[6] Stable Diffusion,[7] and Midjourney[8] can open-up new avenues of creativity, and recent work is blurring the lines between text and image models[9] to build hybrid applications.

2.1.4 Coding Assistants

Fourth, because software code is almost exclusively expressed in text, LLMs are ideally suited to serve as "coding assistants" that perform varying degrees of programming tasks. GitHub and others already offer powerful assistants [10] capable of writing large blocks of mostly (or entirely) correct code in seconds,[11] even for niche languages, and these technologies are advancing rapidly as newer, more powerful language models are developed. [12] These models promise significant increases in efficiency by 'handling' some routine or repetitive programming tasks and allowing programmers to concentrate more on the critical design and details of their programs.

These examples show how a variety of generative AI applications already provide novel capabilities or more efficient workflows. We are only scratching the surface of what these generative models are and what kinds of applications they could make possible, but still these early applications are more than simple prototypes: they offer real value to their users.

---

[6] OpenAI. DALL·E 2.
[7] Stable Diffusion. Stable Diffusion online.
[8] Midjourney. Midjourney.
[9] OpenAI. GPT-4 technical report. (2023).
[10] GitHub. GitHub Copilot.
[11] Yegge, S. Cheating is all you need. (2023).
[12] GitHub. Introducing GitHub Copilot X.

2.2 Looking Ahead

Beyond the applications themselves, we are just beginning to understand the *kinds* of language-based generative AI we can build. For all its promise, it is not obvious that current LLM methods are the "best" ones. Recent work from DeepMind,[13] for example, found that current language models are far less powerful than they could be simply because current data sets (gigantic as they are, being made up of large chunks of the internet) *are too small* to keep pace with the size of the models being trained and deployed. The same research also found that data sizes must increase at the same rate as model sizes to keep parity: doubling the size of an LLM will require doubling the size of the training data. This makes training ever-larger LLMs in the current mold infeasible. New architectures will be needed to continue improving these applications—and to make self-hosting generative AI models feasible for business.

In the short term, the length of time involved in LLM query-response workflows might also limit these models' real-world usefulness for business. Each query to an LLM can require seconds to produce a response depending on the number of tokens requested in return, how much additional contextual data is provided in the query, etc. This is a long time to wait for a response, especially as compared to the millisecond response times to which people and systems have grown accustomed. As industry further incorporates LLMs, connecting multiple LLM-based agents to determine which actions to take and in what order (i.e., multiple sequential requests to one or more models), tools could easily request multiple outputs in response to a single prompt, which would further increase application latencies. Response delays per request will likely decrease over time as technology improves, but how quickly this happens will depend on how large LLMs grow or the architectures used to build these applications.

[13] Hoffmann, J. *et al.* Training compute-optimal large language models. (2022).

3.0 Generative AI Risks and Mitigations

The introduction of a new technology carries both potential and risk, particularly with rapid adoption like we have seen in the early stages of this generative AI boom. Some of the risks we might associate with these new generative AI tools can be seen as straightforward extensions of considerations businesses must already make when training machine learning models, storing large amounts of data, and making decisions with the help of model outputs. Other risks will be inherent to generative AI as a new *kind* of tool. Here we focus primarily on risks specific to generative AI.

We also acknowledge that many AI observers and practitioners are concerned about big-picture issues, including the potential development of an "artificial general intelligence" (AGI)[14] and the societal costs of pervasive machine learning and artificial intelligence algorithms. Addressing these concerns is important, especially as machine learning and AI are already deeply embedded in society at the levels of government, business, and individuals. As a guide for business, however, we focus here on four areas of practical interest: legal risks, data privacy risks, factuality or correctness risks, and risks around the control and ownership of these generative AI models.

3.1 Legal Risk

There are already potential legal risks to incautious adoption of present forms of generative AI, and more regulations are emerging. Although large corporations are actively building and applying generative AI, including GitHub,[15] Meta (previously Facebook), OpenAI, Google, and others, many are not convinced of these technologies' legality—see ongoing lawsuits related both to coding assistants[16] (GitHub Copilot[4]) and to image-generation applications[17, 18] (Stable Diffusion[10]). These lawsuits target the data used to train these models, and the lack of consent given by the data's creators or owners, rather than the application of the technology itself, and GitHub (and the other large players) apparently believe they are in the right: as one example, Microsoft/GitHub currently offer Copilot as a paid, AI-powered coding assistant. Nonetheless, these and other lawsuits will take time to sort out, and it is rare that legal cases involving technology are neatly decided.

[14] Wikipedia contributors. Artificial general intelligence — Wikipedia, the free encyclopedia. (2023).
[15] National Conference of State Legislation, Legislation Related to Artificial Intelligence. (2002).
[16] Butterick, M. GitHub Copilot litigation. (2022).
[17] Butterick, M. Stable Diffusion litigation. (2023).
[18] Brittain, B. Getty Images lawsuit says Stability AI misused photos to train AI. (2023).

Thinking more broadly about accountability and legal responsibility, LLMs are and always will be imperfect like any other machine-learning technology. Individuals and companies are legally accountable for their decisions, so each will—and must—choose how to apply generative AI tools. In this sense the situation is the same as it has always been but with new, and in some ways more powerful, tools available to us.

3.2 Corporate Data Risk

Second, we note an increased risk to corporate data simply because of how generative AI tools are designed. Useful AI applications involve interaction and contextual information, usually via documents, free text, images, etc., provided by the user. This ability (or requirement) to contribute information to a model creates a risk of proprietary or sensitive information being shared with an outside party—the owner of the AI tool.[19] As a first line of defense, users ought not be providing sensitive information to tools like ChatGPT without corporate guidance, especially given the potential for user-provided data to be used in model training and then to be retrieved later, in some form, by downstream users of the models. (The risk is probably less for tools where users are not providing their own contextual information.)

A simple, though perhaps costly, mitigation strategy for text or chat-based tools is to fine-tune an in-house, proprietary language model. Relatively small "base" models are already proliferating, and given the necessary expertise, it seems plausible to tune a custom question-answering AI for relatively low cost.[20, 21, 22] Even these early, commoditized models demonstrate similarly useful properties as the larger models like ChatGPT.

3.3 Factuality and Correctness Risk

Third, text models risk getting facts wrong, "hallucinating," or reporting incorrect information.[23] (Image-generation models can likewise produce unrealistic output, though that might be less unwelcome.) There are plenty of examples of ChatGPT hallucinating incorrect proofs, facts, and the like. And there is no guarantee that, as the world changes, the models will keep up to date or will not begin to provide outdated information. LLMs do

[19] DeGeurin, M. Oops: Samsung employees leaked confidential data to ChatGPT. (2023).
[20] Chase, H. LangChain. (2022).
[21] Conover, M. *et al.* Hello dolly: Democratizing the magic of ChatGPT with open models. (2023).
[22] AI, M. Introducing LLaMA: A foundational, 65-billion-parameter large language model. (2023).
[23] Wikipedia contributors. Hallucination (artificial intelligence) — Wikipedia, the free encyclopedia. (2023).

not "think" in the sense that we usually mean; instead, they sample over distributions of words or related information given some context—often with the objective of producing the next word in a sequence. In this light, it is easy to see how a model might find its way to a (very) wrong answer to a user prompt via linguistically plausible chains of inference. We find it impressive that these models perform as well as they do.

Perhaps the most robust way to address this issue is to build a discipline of critically evaluating LLM responses—or any other output from generative AI. As these models gain capabilities, it might become even more difficult to tell falsehood from fact, so independent critical thinking will and vetting of results will continue to be essential. For AI coding assistants we already have some templates for how to approach this problem: testing, code reviews, etc.[5] We ought to consider similar practices for other, more general LLM applications.

From another point of view, we also already have examples of mitigating this risk through middleware in AI technologies like the personal assistants Siri, Alexa, etc. These applications serve as a buffer between users and the "raw" output from their various models operating behind the scenes and allow the insertion of other technologies (e.g., a profanity filter to filter out potentially offensive language) to help control and guide the models and mitigate the risks of an incorrect response.

3.4 Model Ownership Risk

Finally, we consider the risks of consolidated ownership of these generative models, which leaves businesses' use of these technologies' dependent on the decisions of a few owners' preferences around how to shape and train their models. In this situation, we have neither control of the model (e.g., operating our own, mostly equivalent technology) nor choice in selecting a platform (choosing from multiple options, depending on our requirements). This ownership problem interacts with data governance issues as well: If a few key players own the technologies most businesses use, they are provided a highly privileged position in terms of the data that businesses are contributing simply to use the models.

With only a few players controlling generative AI, their models and processes will likely be opaque. Without incentive (e.g., competition), model owners are unlikely to provide helpful views into either the model training data or the kind of biases, norms, etc. the models encode.

A related but more active concern is that model owners have the unique ability to shape models' notion of the 'world.' For example, OpenAI already spends a large amount of effort shaping their LLMs to mitigate a wide range of potential harms before releasing the technology.[12] These same techniques, though, could of course be used with malintent. And, even without any explicit intent at all, every LLM will express some kind of "worldview"—

generally a mixture of the worldview its owners encode, and a worldview derived from its training data.

Diversification of models, much like diversification of investments, is not a threat but rather a benefit and a protection. The greater the variety of technological approaches that we can encourage and employ, the better. LLMs represent the potential to further distribute knowledge because they offer new interfaces to data that are both convenient and transformative. Our natural tendency, though, seems to drift toward convenient centralization (Google, Wikipedia, etc.). When incorporating generative AI into their workflows, users and businesses should be sure to consider the issues of diversity and potential lock-in.

4.0 Summary

As a collection of data science, machine learning, and AI practitioners, Elder Research naturally keeps an eye on developments in the fields of deep learning and generative AI. We also work in these areas with our clients and use these new generative applications ourselves: We are actively developing tools for applying these capabilities to our own business processes.

Generative AI is a fascinating technology, and it is exciting to work in the field during this wave of adoption. It is also remarkable that such simple models, trained (using a tremendous volume of data) to solve simple problems, can contribute in interesting ways to such a variety of tasks. It is early days, and there are likely many important developments lying ahead, but these technologies also hold value for business right now—applied wisely and with discretion.

Have questions? We'd love to chat!

Reach out to us at **contact@elderresearch.com**

About the Author

Tom Shafer, Ph.D.

Principal Data Scientist

Tom (PhD, Physics) has spent most of his career building computational tools and applying them to complex problems. As a Principal Scientist at Elder Research, Tom contributes to a diverse collection of projects and clients across the company. Highlights include Bayesian multifactor modeling with Stan, object detection with PyTorch, and network science and graphs research. He most enjoys the opportunity to work with many other Elder Research scientists to learn, experiment, and solve interesting problems.

ELDER RESEARCH

DATA SCIENCE · AI · MACHINE LEARNING

ElderResearch.com