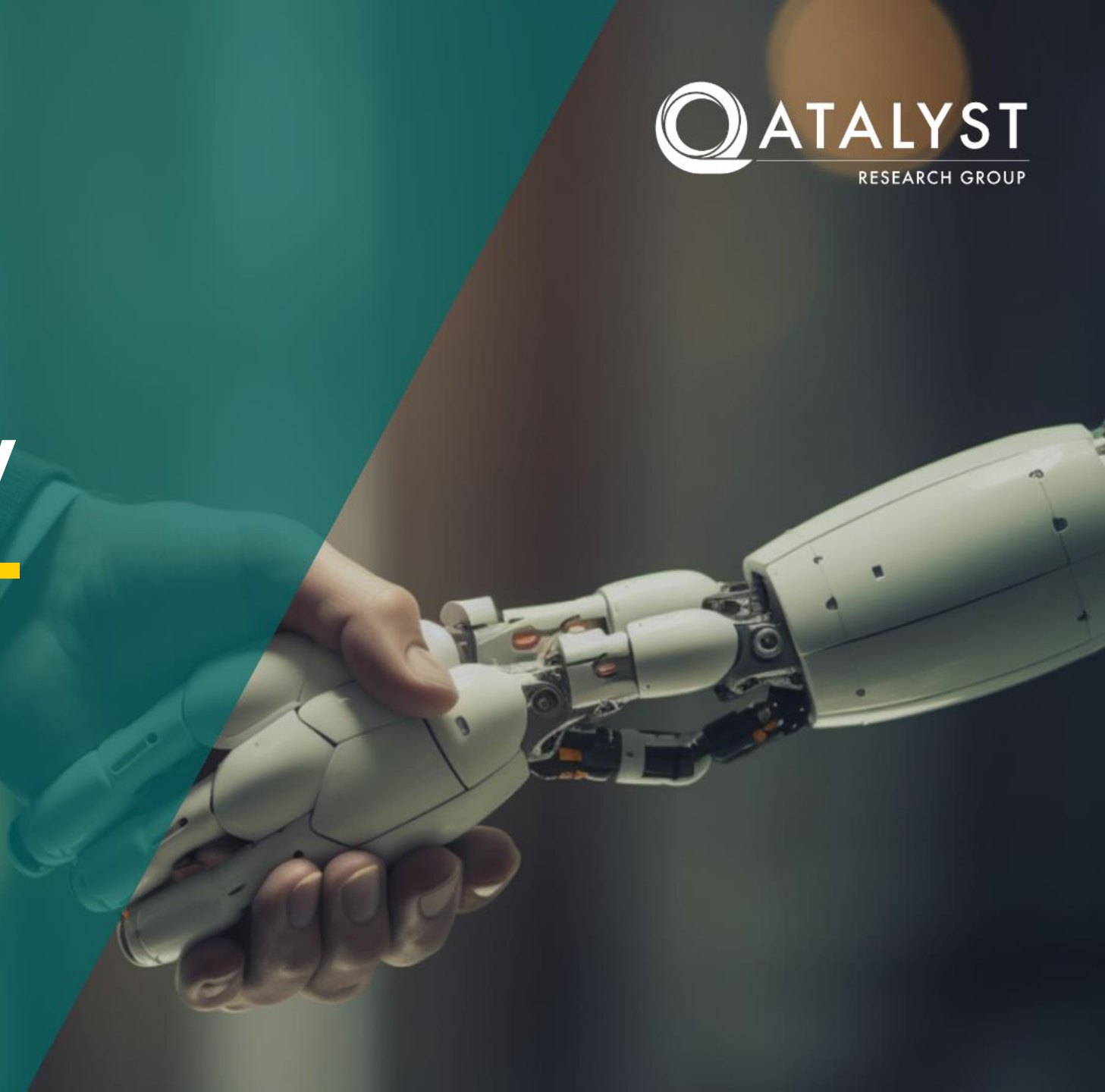
Generative AI & The Management Consulting Industry

Privacy, Intellectual Property, Safety, and Reliability Concerns



QATALYST
RESEARCH GROUP

Our previous posts in this series focused primarily on the potential impact that using AI as a productivity tool could have on the management consulting industry. This week's post is written by Kristen Gauthier, a Senior Indigenous Researcher with Qatalyst. She brings an important voice as someone not previously involved with AI but who, as both an artist and a researcher, rightfully has some significant concerns about the negative impacts that could result from the widespread adoption of AI.

My first impression of AI was that it was a technology that had the entirety of human knowledge at the ready, waiting for someone to ask a question. It was a tool that would continue to learn as time passed and more information was shared or asked of it. I had yet to learn that there were many large language models (LLMs), each with its own repository compiled specifically for it. I've spent the last few weeks educating myself about what generative AI is, what it can do, the limits, the possibilities, and the concerns. I hope that, with this article, we can open up some conversations about protecting the rights of creatives, scholars, scientists, and individuals.

Even within pro-AI communities, there are differing opinions about the need for regulations or legislation. Some believe that AI should be allowed to advance full throttle with no government limiters, while others believe that no further development of AI beyond ChatGPT4 should be allowed. Most believe that, in order for AI to properly reach its full potential, there have to be parameters laid out that will guide improvements, applications, and use.

The power and use of AI technology is growing exponentially. The many companies, countries, and organizations now involved in advancing AI projects raise questions and concerns about how we will protect privacy and intellectual property. How do we know that the information that's being generated is accurate, ethical, and unbiased? How can we ensure that private data remains private? How do we ensure that there is transparency in how data is collected and used?

UNESCO (United Nations Educational, Scientific and Cultural Organization) released the following statement commenting on its concerns with AI: “We see increased gender and ethnic bias, significant threats to privacy, dignity and agency, dangers of mass surveillance, and increased use of unreliable Artificial [Intelligence] technologies in law enforcement, to name a few. Until now, there were no universal standards to provide an answer to these issues.”

Privacy Concerns

While countries such as Canada have acts in place to protect privacy, such as the *Personal Information Protection and Electronic Documents Act* (PIPEDA), data breaches are not uncommon. On June 20th, 2023, Group-IB Threat Intelligence reported that over 100,000 ChatGPT accounts were compromised, which could lead to the sale of private user information and the sharing of confidential data. Companies have a shared responsibility with users to protect private information for the sake of the business and their clients. Going forward, there must be a collective responsibility between users, businesses, and LLM developers to implement robust encryption and information-sharing policies. Similarly, companies developing LLM databases must ensure that information is anonymized and that identifying information is removed when creating learning databases.

Transparency will be another component of ensuring privacy and anonymity. Users must be able to understand how the information they share with LLMs will be used; informed consent will enable users to make decisions about which products they choose to use, control how their data will be collected, and provide them with options for opting out of the sale of their information to third parties, the use of their data for research, and the use of their prompts and data to further train LLMs.

Creative Works and Intellectual Property (IP)

As an artist, when I heard about art-generating AIs, my immediate concern was about the impact they could have on those who make a living solely off their craft. If a computer could generate a logo with the same description a commissioner would provide to an artist, what would happen to that artist? After spending so much time researching AI, I have also become concerned about copyright infringement issues.

Most individuals have now heard about the lawsuit that Getty Images has filed against Stability AI (the creator of Stable Diffusion), one of the more popular image-generating AIs available. The staggering number of photographs (over 12 million, according to the lawsuit) on which the model was trained has resulted in Stable Diffusion generating images that still retain the Getty Images logo, in what Getty Images states is a "*brazen infringement of Getty Images' intellectual property*." Furthermore, a class-action lawsuit was brought against Stability AI by a collective of artists whose works were used without their consent to train the AI model, which has been used to imitate and nearly replicate their work. Their website further details how generative AI works and explains the lawsuit in depth.

Generative AI has been used by many companies, students, and writers to quickly create journals, essays, poems, articles, and other works that often pass as written by the individual submitting them. Similar to art-generating AI, these works are not original in that they are, in effect, simply an amalgamation of other material. A way to mitigate IP issues could be to use ethical generative AI models, which have been trained only on open-source material or copyrighted material with the consent of the owner.

In its current strike, the Writers Guild of America has expressed concerns that the Alliance of Motion Picture and Television Producers (AMPTP) won't even open a discussion about the significant impact that AI could have on all writers in the media. The impact of AI on writers from minority groups may be even more significant. A 2022 report about writers in television by the Think Tank for Inclusion & Equity noted that writers from minority groups are often at the lower levels of staff writing positions; these are the positions that AI would most easily replace since senior writers will be most likely retained to add to, edit, and proof what AI has created. This could remove significant and impactful opportunities for writers of colour, preventing them from advancing into more senior positions within the industry.

As more and more generative AI models are created, tweaked, and “perfected”, they have the ability to impact every industry that requires writing as a core task. This includes the world of management consulting. Generative AI can be used as a tool to improve productivity in the industry, which means that fewer workers are needed to produce the same level of work. If a company can largely rely on AI to find, research, and write a report, where does that leave us researchers? While I believe we are a way off from this reality, we should be asking how to address these concerns before they become an issue. How do we adapt what we do, and how do we do it to accommodate AI?

Safety Concerns

Since ChatGPT's release in November 2022, there has been an explosion in the number of existing products that incorporate generative AI, as well as the development of many new apps built on top of LLMs. Whether you are pro or anti-AI, you've likely seen many articles that share stories about the amazing and horrific things that AI has done. How do we ensure that AI is safe for everyone to use? Research has found that ChatGPT has provided accurate and helpful information to those who have asked it about suicide, addiction, mental, and physical health issues. However, an AI chatbot apparently drove a man to end his own life after six weeks of discussions about climate change.

Thankfully, there are thousands of people who are dedicating themselves to this ever-changing landscape, working to ensure that AI moves forward ethically and safely. For example:

- Dr. Rumman Chowdhury, who led ethical AI research at Twitter, cofounded Humane Intelligence with Jutta Williams. The focus of Humane Intelligence is on "*safety, ethics, and subject-specific expertise*". They believe that, through diverse staff contributing to solutions, issues with large-scale models can be solved.

- Daniela Amodei and other former senior members from OpenAI created a public-benefit corporation called Anthropic, whose chatbot Claude is guided by ethical principles from sources such as the United Nations Declaration on Human Rights.

- Hundreds of experts, professors, scientists, CEOs, and other notable figures signed an open letter calling for action to mitigate "the risk of global extinction from AI". The letter states that this should be given the same priority as mitigating other major risks, such as pandemics and nuclear war.

- Another letter was created in March of this year, calling for “all AI labs to immediately pause for at least 6 months the training of AI systems more powerful than GPT-4.” The letter indicated the pause should be used by AI labs and independent experts “to jointly develop and implement a set of shared safety protocols for advanced AI design and development that are rigorously audited and overseen by independent outside experts. These protocols should ensure that systems adhering to them are safe beyond a reasonable doubt. This does not mean a pause on AI development in general, merely a stepping back from the dangerous race to ever-larger unpredictable black-box models with emergent capabilities.” The letter, which can be read here, was signed by notable figures such as Andrew Yang (Presidential Ambassador of Global Entrepreneurship, and 2020 United States Presidential Candidate), Craig Peters (CEO of Getty Images), Meia Cita-Tegmark (Co-Founder of the Future of Life Institute), Steve Wozniak (Co-Founder of Apple), and Elon Musk (CEO of SpaceX, Tesla, and Twitter).

Reliability Concerns

The level of human bias and prejudice embedded into AI is largely a function of how it was trained. There have been examples of hiring algorithms which were biased against women because they were trained on resumes submitted over the years, most of which came from men. Another predicted that Black patients were less likely to require medical attention because of their lower spending for medical issues.

So how do we know when we can trust AI and when we should disregard the information that it provides us? Unfortunately, there's no easy answer. Just like anything on the internet, it's important to be able to think critically about information and determine if the sources are fair, ethical, and unbiased. While applications that are able to link AI's responses with specific data sources are a step ahead, having a source does not necessarily ensure that the content is unbiased and accurate. Furthermore, validation is still critical given that AI can hallucinate data sources (e.g., provide references to journal articles that do not exist). The onus will be on developers to continue to improve the validity and reliability of their data, improve transparency, and reduce bias.

So, what now?

The advent of LLMs and generative AI has the potential to create amazing opportunities, new fields of employment, and act as an extremely important tool for humanity. But ultimately, it should be used as a tool and not a replacement for human productivity or creativity.

Globally, there has been a rush to create legislation and regulations concerning AI. Documents from the United Nations' UNESCO, World Health Organization, European Union, Canada, and the United States, among others, are a small step forward into an ever-expanding world. Hopefully, we can encourage all key players to continue to expand on ensuring ethical advancement and a helpful future with artificial intelligence.