CHATGPT: ENTRE O MITO E A REALIDADE

Indiscutivelmente que as novas ferramentas de IA como DALL-E e mais enfaticamente o chatGPT criaram um verdadeiro frenesi, principalmente porque muitas pessoas estão tendo contato com sistemas de IA pela primeira vez. E ficam extasiadas. Isso, naturalmente leva a previsões extremamente emocionais que transformarão a sociedade, acabarão com muitas profissões e quem não pular no trem, vai, inevitavelmente, ficar alijado do mercado. É o efeito FOMO na sua essência. Mas, à medida que entendemos o potencial e limitações dessas ferramentas, fica mais claro onde e como serem usadas.

Claro, muitas profissões terão muitas das suas tarefas repetitivas feitas por essas ferramentas e com o tempo, os sistemas "generative AI" como chatGPT tem o potencial de mudar a forma como vivemos e trabalhamos, apesar das atuais falhas claras da tecnologia. E, como sempre acontece com mudanças tecnológicas, desde tempos imemoriais, muitas pessoas fazem transições para atividades que ainda não existem, mas que são possibilitados por essa nova tecnologia. Um desses trabalhos, é "prompt engineer", que é a pessoa que insere textos em uma IA generativa para criar imagens ou textos. E provavelmente será uma tarefa de curta duração, uma vez que cada um de nós começaremos a criar mais e mais prompts, sem necessidade de ajuda externa. Também devemos estar alertas para os sinais de espuma que lembra o boom das criptomoedas, que fracassou recentemente.

Esses sistemas popularizam o uso da IA e inclui essa a tecnologia mais profundamente na maneira como vivemos e trabalhamos, mais do que vemos atualmente nos algoritmos de recomendação por exemplo, que já nos direciona o que ler, ver e ouvir, mas agora também respondendo às nossas perguntas, escrevendo nossos memorandos e discursos, e até mesmo produzindo poesia e arte. E como os recursos financeiros, intelectuais e computacionais necessários para desenvolver e executar a tecnologia são enormes, as empresas que controlam esses sistemas de IA, as BigTechs, se tornarão as maiores e mais ricas. O poder vai se concentrar ainda mais nelas.

Grande parte das discussões e entusiasmo pelos "generative AI" se concentra nas atividades pessoais, no mundo B2C. E quanto ao mundo corporativo?

Sistemas de IA não são novidade nas empresas, e diversas iniciativas já aconteceram, até mesmo com a criação de "AI Labs" por parte de algumas. Mas, uma grande parcela dos projetos e desses labs acabou de forma muito frustrante. Os fatores inibidores geralmente são o tempo e os custos associados à construção, treinamento e implantação de modelos de próprios de IA.

Algoritmos, que podem conter grande número de variáveis e restrições, são demorados e caros para serem construídos. Muitas vezes cada algoritmo é único e requer quantidades

variáveis de tempo e capital, dependendo de seu escopo, escala e número de variáveis. Geralmente, a maior parte do custo são os salários dos cientistas de dados e engenheiros de ML, que escrevem esses algoritmos, os treinam com os dados coorporativos, os testam, ajustam e os monitoram após a implantação.

Mas a popularidade viral do ChatGPT está estimulando o interesse pelos conselhos e pelos CEOs, com novos aplicativos de IA, explorando a tecnologia "generative AI", como na sua capacidade de gerar relatórios de negócios, propostas de marketing e código para programação de software, entre outras coisas. No entanto, ao contrário do entusiasmo de quem não está na gestão das empresas, existem, por parte dos CIOs, muito mais preocupações em integrar esses sistemas à pilha de sistemas corporativos, além dos questionamentos sobre o uso de dados online e riscos de segurança, que podem afetar as regras de compliance e eventualmente colocar em debate os procedimentos atuais de segurança e privacidade.

Mas acima de tudo, eles estão preocupados com o controle do ChatGPT sobre a realidade. Assim as "alucinações" que ele produz como explicar com detalhes muito

convincentes porque os ovos de vaca são maiores que os de galinha e porque a lua é maior que o sol, não são tão engraçadas quanto algo similar é colocado em um relatório para o conselho ou acionistas.

O ChatGPT é um tremendo passo à frente para a IA generativa, mas no momento, o ChatGPT deve ser usado com cautela em um ambiente de negócios corporativos. Além de seus problemas de precisão, o ChatGPT requer uma série de outras melhorias antes de poderem ser usados em aplicativos corporativos, como atualizações necessárias incluam resultados mais rápidos, recursos de segurança avançada e melhores habilidades de linguagem.

À medida que essas e outras melhorias de desempenho forem lançadas, começaremos a ver o surgimento de aplicativos corporativos como pesquisa corporativa, integração com plataformas de comunicação, ferramentas de vendas e assim por diante. Mas, na fase atual, em beta, fica restrito ao mundo B2C, menos exigente. O problema básico com o ChatGPT e sistemas similares é que eles escrevem extremamente bem, mas podem estar errados e não fundamentados em fatos. E não esqueça que a Internet nunca esquece (AI Tools Bug Out Because the Internet Can Never Forget)! Assim, a tecnologia ainda deve ser limitada para escrever clichês corporativos, como anúncios de produtos ou serviços ou outros materiais promocionais. Mas, mesmo assim, nenhuma empresa emitiria tais coisas sem revisão humana. Em termos de setores nas empresas, podemos pensar em vendas e marketing, call centers ou para resumir relatórios de ganhos, estudos e outros documentos de negócios, onde os discursos de vendas ou e-mails gerados podem ser facilmente revisados por humanos antes de serem enviados. O artigo mostra alguns exemplos de marketing (ChatGPT Use Cases For Marketing Professional Services).

A IA generativa é capaz de coisas incríveis, mas como um todo, precisa amadurecer. O próprio Sam Altman, executivo-chefe da OpenAI, alertou quanto a confiar no ChatGPT "para

qualquer coisa importante no momento". Em um tuíte ele disse "ChatGPT is incredibly limited, but good enough at some things to create a misleading impression of greatness."

Mas, quanto a geração de código de programação? O chatGPT é um chatbot generativo que foi treinado em um enorme volume de artigos, sites e postagens de mídia social coletados de data sets, bem como de entrevistas transcritas que capturam as nuances da vida humana. Ao detectar padrões linguísticos e frases familiares, o algoritmo aprendeu a prever qual palavra provavel

mente seguirá uma sequência de palavras. A partir daí, foi capaz de prever a próxima frase e o próximo parágrafo, eventualmente criando um texto coerente.

Essa abordagem também pode ser aplicada para escrever código de computador, permitindo que o ChatGPT preveja grandes blocos de código que os desenvolvedores precisariam inserir para executar uma determinada tarefa em um programa de software, um recurso que os CIOs se entusiasmam, pois pode acelerar o desenvolvimento de aplicativos.

Mas, na prática, o chatGPT, no seu estágio atual ainda não tem condições de fazer os desenvolvedores de software profissionais economizarem muito tempo, uma vez que produz longas cadeias de comandos que precisaram ser verificadas e muitas vezes reescritas linha por linha. No mundo corporativo, os casos de uso do ChatGPT são ainda um universo menor do que as pessoas estão imaginando. Até agora o chatGPT e outras ferramentas "generative AI" estão muito longe do core business das empresas. Integrar o chatGPT ao Office não é, definitivamente, algo que afeta o core business de nenhuma empresa.

Por enquanto, os CIOs devem experimentar o ChatGPT para determinar como ele pode ser usado, principalmente por tentativa e erro. Antes de integrar o ChatGPT aos negócios, devemos lembrar que ainda estamos no início do ciclo de vida da tecnologia. Precisamos ser prudentes e nos perguntar o que ela pode fazer hoje. Em sua forma atual, o ChatGPT responde muitas vezes, de forma imprecisa, embora com confiança, sua matemática está atrasada e o conjunto de dados está atualizado apenas até uma determinada data.

Indústrias altamente regulamentadas como medicina, precisarão ser especialmente cautelosas ao usar o ChatGPT , sua variante treinada em textos médicos, BioGPT, ou qualquer forma de IA generativa. E até agora os resultados não tem sido muito positivos (An early evaluation of ChatGPT on common medical NLP tasks). A conclusão do estudo: "We do not recommend these models for production use today. They are impressive research advances, and we use them internally to bootstrap smaller and more accurate models, but they are not fit for the vast majority of real-world use cases". Além disso, sem proteção, os dados podem facilmente ser mal utilizados, ou pior, servir de base para multiplicar resultados ruins.

Em contraste com o ChatGPT, cuja utilidade na empresa ainda é questionada, ferramentas de codificação de IA como o Copilot, parecem ser mais capazes de resolverem alguns

problemas de negócios do mundo real. O modelo de IA por trás do Copilot é treinado em dados do GitHub, que abriga uma comunidade de código aberto onde os desenvolvedores contribuem e compartilham códigos. O objetivo dessas ferramentas não é substituir os desenvolvedores, mas ajudá-los a melhorar sua produtividade, da mesma forma que ferramentas como verificação ortográfica e preenchimento automático de frases ajudam as pessoas a escrever documentos. Eles trabalham sugerindo novos trechos de código e testes e fornecendo recomendações técnicas dentro dos programas de escrita de código que os desenvolvedores já usam.

O Copilot e ferramentas semelhantes como CodeWhisperer da Amazon não substituirão os desenvolvedores e criarão softwares ou aplicativos corporativos por conta própria. Mas, auxiliam na documentação do código, geram pedaços de código mais simples e facilitam planejar os testes. Seu principal apelo é a promessa de um processo de desenvolvimento de software mais rápido e menos manual, que pode ajudar as empresas a aliviar a pressão causada pela escassez de talentos em software.

Alguns cuidados são necessários. Existem possíveis riscos de segurança cibernética e propriedade intelectual. Com o objetivo de resolver algumas dessas preocupações, o GitHub lançou recentemente uma nova versão do Copilot for Business, que inclui a capacidade de gerenciar usuários e executar em VPN. Outro recurso impede que a ferramenta sugira códigos que possam estar sob licença de outra empresa, ajudando a mitigar os riscos legais. O GitHub é objeto de uma proposta de ação, movida em novembro, que argumenta que o Copilot não dá crédito aos autores originais cujo código é usado para gerar seus resultados. Um artigo interessante sobre o tema é “The lawsuit that could rewrite the rules of AI copyright”.

A versão corporativa do Copilot também inclui um bloqueio mais amplo de sugestões de código que podem incluir vulnerabilidades conhecidas de segurança cibernética, algo com o qual os CIOs estão mais preocupados à medida que cresce a conscientização sobre a integração da segurança no desenvolvimento de aplicativos corporativos.

Em resumo, as tecnologias “generative AI” ainda são incipientes e não devemos, no âmbito corporativo, mergulhar fundo nelas, antes que as questões que nos preocupam hoje sejam resolvidas ou mitigadas. E, claro, o ambiente corporativo é muito mais sério que o ambiente pessoal. O uso individual é uma coisa bem diferente do uso corporativo!