

O que é a IA Generativa e como funciona?

14 de Março de 2023

A inteligência artificial já concebe microchips e envia-nos spam – o que se segue? Eis como a IA Generativa funciona realmente e o que podemos esperar agora que chegou. Por Josh Fruhlinger

IA Generativa é um termo geral para qualquer tipo de processo automatizado que utiliza algoritmos para produzir, manipular ou sintetizar dados, muitas vezes sob a forma de imagens ou texto legível por humanos. **É chamada Generativa porque a IA cria algo que não existia anteriormente.** Isto é o que a diferencia da IA discriminatória, o que distingue entre diferentes tipos de dados. Por outras palavras, a IA discriminatória tenta responder a uma pergunta como “Esta imagem é uma imagem de um coelho ou de um leão?”, enquanto a IA Generativa responde a perguntas como “Desenha-me uma imagem de um leão e de um coelho sentados um ao lado do outro”.

Este artigo demonstra a IA Generativa e os seus usos em modelos populares, tais como ChatGPT e DALL-E. Analisaremos também as limitações desta tecnologia.

O surgimento da IA Generativa

A IA Generativa existe há anos, possivelmente desde que ELIZA, um chatbot que simula a conversa com um terapeuta, foi desenvolvido no MIT em 1966. Mas anos de trabalho na IA e na aprendizagem de máquinas compensaram recentemente com o lançamento de novos sistemas generativos de IA. Tenho a certeza que já ouviram falar do ChatGPT, um chatbot de IA baseado em texto que produz uma prosa notavelmente semelhante à humana. A DALL-E e a Stable Diffusion também atraíram a atenção pela sua capacidade de criar imagens vibrantes e realistas a partir de mensagens de texto. Referimo-nos frequentemente a estes e outros sistemas semelhantes como modelos, porque representam uma tentativa de simular ou modelar algum aspeto do mundo real a partir de um subconjunto (por vezes muito grande) de informação sobre o mesmo.

Os resultados destes sistemas são tão bizarros que muitos levantam questões filosóficas sobre a natureza da consciência e preocupam-se com o impacto económico da IA Generativa no emprego humano. Mas embora todas estas criações de inteligência

artificial sejam sem dúvida grandes notícias, o que se passa por debaixo da superfície é muito menos do que alguns supõem. Abordaremos abaixo algumas destas questões gerais. Primeiro, vejamos o que está subjacente a modelos como o ChatGPT e o DALL-E.

Como funciona a IA Generativa?

A IA Generativa utiliza a aprendizagem mecânica para processar uma enorme quantidade de dados visuais ou textuais, grande parte dos quais colhidos da Internet, e determinar quais as coisas que mais provavelmente aparecerão perto de outras coisas. Grande parte do trabalho de programação na IA gerativa envolve a criação de algoritmos capazes de distinguir as “coisas” que interessam aos criadores da IA: palavras e frases no caso de chatbots como ChatGPT, ou elementos visuais para DALL-E. Mas fundamentalmente, a IA Generativa cria os seus resultados através da avaliação de um enorme corpus de dados sobre os quais foi treinada, e depois responde a pedidos com algo que se enquadra na faixa de probabilidade determinada por esse corpus.

O preenchimento automático (quando o telemóvel ou o Gmail sugere o que pode ser o resto da palavra ou frase a ser dactilografada) é uma forma de IA Generativa de baixo nível. Modelos como o ChatGPT e DALL-E levam a ideia a níveis muito mais avançados.

Modelos de IA generativos de formação

O processo pelo qual os modelos são desenvolvidos para acomodar todos estes dados é chamado formação. Um par de técnicas subjacentes para diferentes tipos de modelos entram aqui em jogo. O ChatGPT utiliza o que se chama um transformador (daí o T). Um transformador deduz o significado de longas sequências de texto para compreender como diferentes palavras ou componentes semânticos podem relacionar-se entre si, e depois determinar a probabilidade de aparecerem próximos um do outro. Estes transformadores são executados sem supervisão num vasto corpus de texto de linguagem natural num processo chamado pré-formação, antes de serem afinados por humanos que interagem com o modelo.

Outra técnica utilizada para treinar modelos é a chamada rede de adversários generativos ou GAN. Nesta técnica, existem dois algoritmos concorrentes. Um gera texto ou imagens com base em probabilidades derivadas de um grande conjunto de dados; o outro é uma IA discriminatória, que foi treinada por humanos para avaliar se essa saída é real ou gerada pela IA. A IA Generativa tenta repetidamente “enganar” a IA discriminatória, adaptando-se automaticamente para favorecer resultados bem-sucedidos. Uma vez que a

IA Generativa “ganha” sistematicamente esta competição, os humanos ajustam a IA discriminatória e o processo recomeça.

Uma das coisas mais importantes a ter em conta é que, embora haja intervenção humana no processo de aprendizagem, a maior parte desta aprendizagem e adaptação acontece automaticamente. São necessários muitos milhões de iterações para conseguir que os modelos produzam resultados interessantes que a automatização é essencial. O processo é bastante intensivo do ponto de vista computacional.

A IA Generativa é consciente?

A matemática e a codificação necessárias para criar e treinar modelos generativos de IA é bastante complexa e muito além do âmbito deste artigo. Mas se interagir com os modelos que são o resultado final deste processo, a experiência pode ser decididamente estranha. Pode fazer com que a DALL-E produza coisas que se pareçam com verdadeiras obras de arte. Pode ter conversas com ChatGPT que se assemelham a conversas com outro ser humano. Será que os investigadores criaram realmente uma máquina pensante?

Chris Phipps, antigo chefe de processamento de linguagem natural da IBM que trabalhou nos produtos de IA da Watson, diz que não. Ele descreve o ChatGPT como uma “muito boa máquina de previsão”. É muito boa a prever o que os humanos vão achar coerente. Nem sempre é coerente (quase sempre é), mas isso não é porque o ChatGPT “compreende”. É o oposto: os humanos que consomem a produção são muito bons a fazer as suposições implícitas de que precisamos para dar sentido à produção.

Phipps, que também é comediante, faz uma comparação com um jogo de improvisação chamado **Mind Meld**. Duas pessoas pensam numa palavra e dizem-na em voz alta em simultâneo: você diz “bota” e eu digo “árvore”. Criámos estas palavras de forma completamente independente e, no início, elas não tinham nada a ver uma com a outra. Os dois participantes seguintes pegam nestas duas palavras e tentam inventar algo que têm em comum e dizem-no em voz alta ao mesmo tempo. O jogo continua até que dois participantes digam a mesma palavra. Talvez ambas as pessoas digam “lenhador”. Parece magia, mas trata-se realmente de usar o nosso cérebro humano para raciocinar sobre o input (“bota” e “árvore”) e encontrar uma ligação. Nós fazemos o trabalho de compreensão, não a máquina. Com ChatGPT e DALL-E, acontece muito mais do que as pessoas admitem. O ChatGPT pode escrever uma história, mas nós, humanos, fazemos muito desse trabalho para lhe dar sentido.

Testar os limites da inteligência informática

Certas instruções que podemos dar a estes modelos de inteligência artificial irão realçar o argumento de Phipps. Por exemplo, considere a questão: “O que pesa mais, um quilo de chumbo ou um quilo de penas?”. A resposta, claro, é que pesam o mesmo (um quilo), embora o nosso instinto ou bom senso nos diga que as penas são mais leves.

O ChatGPT responderá corretamente a este enigma, e pode assumir que o faz porque é um computador friamente lógico que não tem “senso comum” para tropeçar. Mas não é isso que acontece. O ChatGPT não está a raciocinar logicamente a resposta; está apenas a gerar resultados com base nas suas previsões do que deve seguir-se a uma pergunta sobre um quilo de penas e um quilo de chumbo. Porque o seu conjunto de treino inclui muito texto explicando o enigma, ele reúne uma versão dessa resposta correta. Mas se perguntar ao ChatGPT se dois quilos de penas pesam mais de um quilo de chumbo, dir-lhe-á com confiança que pesam o mesmo, porque essa continua a ser a resposta mais provável a uma pergunta sobre penas e chumbo, com base no seu conjunto de treino. Pode ser divertido dizer à IA que está errada e vê-la responder de forma hesitante; pedi-lhe que pedisse desculpa pelo seu erro e depois sugerisse que dois quilos de penas pesam quatro vezes mais do que um quilo de chumbo.

Porque é que a arte da IA tem demasiados dedos?

Uma peculiaridade notável da IA é que frequentemente retrata pessoas com mãos muito estranhas. Esta peculiaridade está a tornar-se um indicador comum de que a arte foi gerada artificialmente. Esta peculiaridade oferece uma maior perceção de como a IA Generativa funciona (e não funciona). Começamos com o corpus a partir do qual DALL-E e ferramentas semelhantes de IA Generativa visual: imagens de pessoas fornecem frequentemente uma boa visão dos seus rostos, mas as suas mãos são frequentemente parcialmente obscurecidas ou mostradas em ângulos estranhos, pelo que não se pode ver todos os seus dedos ao mesmo tempo. Acrescente-se a isto o facto de as mãos serem estruturalmente complexas, tornando-as muito difíceis de desenhar, mesmo para os artistas mais experientes. E uma coisa que a DALL-E não faz é montar um elaborado modelo 3D de mãos baseado nos vários renderings 2D do seu conjunto de treino. Não funciona dessa forma. A DALL-E nem sequer sabe necessariamente que “mãos” é uma categoria coerente para raciocinar. Tudo o que pode fazer é tentar prever, com base nas imagens que tem, como é que uma imagem semelhante pode parecer. Apesar da enorme

quantidade de dados de formação, essas previsões ficam muitas vezes aquém das expectativas.

Phipps acredita que um fator é a falta de dados negativos. Tanto quanto sei, ele treina principalmente com base em exemplos positivos. Não lhe foi dada uma imagem de uma mão com sete dedos e disse ‘NÃO! Mau exemplo de mão. Não faça isto’. Assim, ele prevê o espaço do possível, não o espaço do impossível. Basicamente, nunca lhe foi dito para não criar uma mão com sete dedos.

Há também o fato de que estes modelos não pensam nos desenhos que estão a fazer como um todo coerente, mas montam uma série de componentes que provavelmente estarão próximos uns dos outros, como mostram os dados da formação. A DALL-E pode não saber que uma mão deve ter cinco dedos, mas sabe que é provável que um dedo seja imediatamente adjacente a outro. (Os mesmos resultados podem ser obtidos com dentes.) De facto, mesmo esta descrição do processo do DALL-E está provavelmente a antropomorfizá-lo demasiado; como diz Phipps: “Duvido que ele tenha sequer uma compreensão de um dedo. É muito provável que preveja a cor dos pixels, e os pixels a cor de um dedo tendem a ser ao lado de outros pixels a cor de um dedo”.

Possíveis efeitos negativos da IA Generativa

Estes exemplos mostram uma das principais limitações da IA Generativa: aquilo a que os profissionais da indústria chamam alucinações, que é um termo talvez enganador para resultados que são, pelos padrões dos seres humanos que os utilizam, falsos ou incorretos. Todos os sistemas informáticos produzem erros de vez em quando, claro, mas estes erros são especialmente problemáticos porque é pouco provável que sejam facilmente detetados pelos utilizadores finais: Se fizer uma pergunta a um chatbot de IA, este normalmente não saberá a resposta. É também mais provável que aceite uma resposta dada na prosa confiante e totalmente idiomática que o ChatGPT e modelos semelhantes produzem, mesmo que a informação esteja incorreta.

Mesmo que uma IA Generativa possa produzir resultados sem alucinações, há vários impactos negativos potenciais:

Criação de conteúdos

Esperemos que por esta altura esteja claro que o ChatGPT e outras IA Generativas não são mentes reais capazes de produzir ideias criativas. Mas a verdade é que nem tudo o

que é escrito ou desenhado precisa de ser particularmente criativo. Muitos trabalhos de investigação a nível universitário ou de liceu visam apenas sintetizar dados disponíveis ao público, tornando-os um alvo perfeito para a IA Generativa. E o facto de a prosa sintética ou arte poder agora ser produzida automaticamente, a uma escala sobre-humana, pode ter resultados estranhos ou imprevistos. Os criadores de spam já utilizam o ChatGPT para escrever e-mails de phishing, por exemplo.

Propriedade intelectual

Quem possui uma imagem ou texto gerado por IA? Se um trabalho protegido por direitos autorais faz parte do conjunto de formação de uma IA, a IA “plagia” esse trabalho quando gera dados sintéticos, mesmo que não o copie palavra por palavra? Estas são questões legais complicadas e ainda não discutidas.

Viés

O conteúdo produzido pela IA Generativa é inteiramente determinado pelos dados subjacentes sobre os quais é treinada. Uma vez que esses dados são produzidos por seres humanos com todas as suas falhas e enviesamentos, os resultados gerados podem também ter falhas e enviesamentos, especialmente se operarem sem barreiras humanas. A OpenAI, a empresa que criou o ChatGPT, colocou proteções no modelo antes de o abrir para uso público que a impedem de fazer coisas como utilizar calúnias raciais; no entanto, outros afirmaram que este tipo de salvaguardas representam o seu próprio tipo de preconceitos.

Consumo de energia

Para além das questões filosóficas vexatórias, a IA Generativa coloca alguns problemas muito práticos: por um lado, a formação de um modelo de IA Generativa requer um enorme consumo de energia. Isto pode levar a grandes contas de computação em cloud para empresas que tentam entrar neste espaço, e em última análise levanta a questão de saber se o aumento do consumo de energia (e em última análise das emissões de gases com efeito de estufa) vale a pena (esta questão também se levanta em relação às moedas criptográficas e à tecnologia da cadeia de bloqueio).

Casos de uso generativo de IA

Apesar destes potenciais problemas, é difícil ignorar a promessa da IA Generativa. A capacidade do ChatGPT de extrair informação útil de enormes conjuntos de dados em resposta a consultas em linguagem natural tem vindo a desafiar os gigantes da pesquisa. A Microsoft está a testar o seu próprio chatbot de IA, chamado de “Sydney”, embora ainda esteja em beta e os resultados tenham sido misturados.

Mas Phipps acredita que tipos mais especializados de pesquisa são um bom ajuste para a tecnologia. “Um dos meus últimos clientes na IBM foi uma grande empresa de navegação internacional que também tinha um negócio multimilionário de consultoria de cadeia de fornecimento”, explica ele. O seu problema era que não conseguiam contratar e formar consultores de nível básico da cadeia de abastecimento com rapidez suficiente e estavam a perder negócios porque não conseguiam responder rapidamente a simples perguntas dos clientes. Criámos um chatbot para ajudar os consultores de nível básico a pesquisar a extensa biblioteca da empresa de manuais e apresentações da cadeia de abastecimento para que pudessem responder às perguntas dos clientes.

Se eu criasse hoje uma solução para esse mesmo cliente, apenas um ano após a criação do primeiro, utilizaria o ChatGPT a 100% e provavelmente seria muito superior ao que criei. O bom neste caso de utilização é que ainda há um perito humano no loop a verificar a resposta. Isso atenua muitas das questões éticas. Existe um enorme mercado para este tipo de ferramenta de pesquisa inteligente para peritos.

Outros casos de utilização possível são:

Geração de código

A ideia de que a IA Generativa pode escrever código informático para nós tem vindo a borbulhar há anos. Acontece que grandes modelos linguísticos como ChatGPT podem compreender linguagens de programação bem como linguagens faladas naturais, e embora a IA Generativa não vá provavelmente substituir os programadores num futuro previsível, ela pode ajudar a aumentar a sua produtividade.

Criação de conteúdos

Embora esta seja uma preocupação (mencionada acima), é também uma oportunidade. A mesma IA que escreve e-mails de spam pode escrever e-mails legítimos de marketing, e tem havido uma explosão de iniciados de copywriting de IA. A IA Generativa prospera

quando se trata de formas de escrita altamente estruturadas que não requerem muita criatividade, tais como currículos e cartas de apresentação.

Desenho de engenharia

As artes visuais e a linguagem natural têm recebido muita atenção no espaço generativo da IA porque são fáceis de compreender para as pessoas comuns. Mas técnicas semelhantes estão a ser utilizadas para conceber tudo, desde microchips a novos medicamentos, e quase de certeza que em breve entrarão no reino da arquitetura informática.

Conclusão

A **IA Generativa** irá sem dúvida perturbar algumas indústrias e mudar (ou eliminar) muitos empregos. No entanto, artigos como este continuarão a ser escritos por humanos, pelo menos por agora. A CNET tentou recentemente conseguir que a IA Generativa escrevesse artigos, mas o esforço fundiu-se com uma onda de alucinações. Se achar interessante, esta poderá ser uma profissão de futuro: engenharia da IA.

Retirado de www.computerworld.com.pt em junho 2023