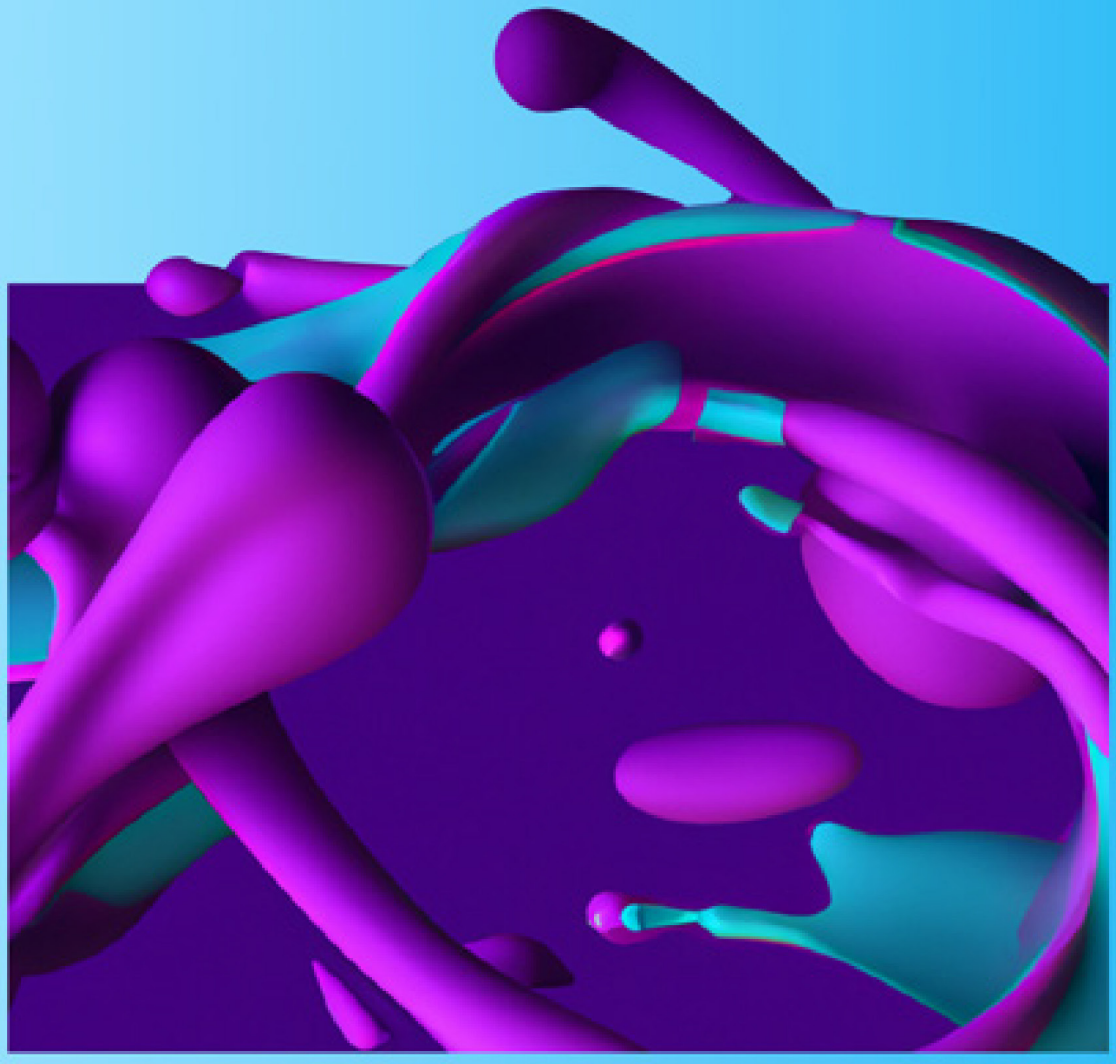
# Modelos de IA Generativa — Os riscos e as potenciais recompensas nos negócios

**O que a ascensão do Chat GPT, DALL.E2, Bard, e outros, pode significar para sua organização.**



KPMG

Julho de 2023

**Lisa Heneghan**
Sócia-líder global de Digital da KPMG International

“Os modelos de IA(Inteligência Artificial) generativa evidenciam o poder da tecnologia. Eles têm o potencial de nos tornar mais produtivos e, em alguns aspectos, facilitam nossas tarefas. No entanto, eles trazem consigo implicações de risco que as organizações e os indivíduos precisam levar em conta. Dito isso, não podemos ignorá-los. Eles estão se integrando rapidamente às nossas rotinas pessoais e profissionais. Precisamos determinar de que forma iremos abraçá-los - mas devemos fazer isso de maneira segura.”

As imagens desta publicação, incluindo a imagem da capa, foram elaboradas com o DALL• E 2, um gerador de arte de IA que cria imagens com base em descrições textuais.

Os *prompts* de imagem utilizados para a capa foram: abstrato fluido, coluna ondulada de azul e lilás, respingo, gotas, fundo púrpura.

Apesar do DALL• E 2 gerar conteúdo visual convincente, ele não foi treinado com base nas orientações de marca da KPMG. Além disso, não tem a expertise humana nem a habilidade necessárias para entender o posicionamento da marca da KPMG. Como resultado, essas imagens estão fora do padrão da marca. Nós as utilizamos especificamente nesta publicação para fins ilustrativos, com permissão especial da KPMG Global Brand.

# Conteúdo

# 1 Sumário executivo

Nós acreditamos que os modelos de inteligência artificial generativa (IA) têm o potencial de transformar os negócios por meio da automação e da execução de certas tarefas com velocidade e eficiência sem precedentes. Isso é especialmente verdadeiro quando a expertise e as habilidades humanas se aliam a um entendimento profundo acerca do uso desses programas e de como aproveitar efetivamente suas capacidades.

No entanto, para desbloquear todo o potencial da IA generativa de maneira responsável, confiável e segura, é importante estabelecer um conjunto de processos internos e controles que todos na organização devem conhecer e seguir.

Neste relatório, abordamos possíveis casos de usos e oportunidades e apontamos o que deve ser considerado caso você esteja pensando em adotar aplicações de IA generativa, como o ChatGPT, por exemplo, em sua organização.

**Aqui estão dez coisas que você deve saber sobre a IA generativa:**

1. As soluções de IA generativa mais comuns dividem-se em cinco categorias: geração de conteúdo; extração de informações; chatbots inteligentes; tradução de idiomas; e geração de códigos.

2. Os modelos de IA generativa podem resumir artigos, redigir *e-mails* e produzir imagens e vídeos. Treinados por humanos, alguns modelos desenvolvem habilidades de conversação para, por exemplo, responder perguntas, admitir erros, questionar suposições e filtrar ou rejeitar solicitações inadequadas.

3. O ChatGPT é um *chatbot* treinado com base em instruções humanas. Seu modelo de linguagem subjacente inicial, o GPT-3.5, dispunha de 175 bilhões de parâmetros e foi treinado com mais de um milhão de conjuntos de dados ou 500 bilhões de *tokens* (palavras ou fragmentos de palavras). O GPT-3.5 não esteva conectado à internet até setembro de 2021, data em que acabou de ser alimentado com dados. O GPT-4, novo grande modelo multimodal da OpenAI, evoluiu a partir do modelo de linguagem anterior.

4. Os modelos de IA generativa têm utilidade em diversas funções empresariais, desde TI, recursos humanos e operações, até finanças, auditoria, jurídico e marketing. Aplicações adequadas incluem redação de propostas, desenvolvimento e teste de código e extração e resumo de informações complexas.

5. A IA generativa adota *inputs* ou parâmetros de dados para aprender e construir conhecimento. A menos que você restrinja explicitamente o provedor de aplicativos de fazer isso, esses dados podem ser usados para responder a solicitações de outras pessoas, o que aumenta o risco de expor informações proprietárias da organização. Dependendo da aplicação, você também pode estar cedendo seus direitos autorais. Consultar os termos e as condições de uso da ferramenta é fundamental para saber o que acontece com os dados inseridos pelo usuário.

6. Dependendo do uso que você faz da IA generativa e de como a implementa, suas atividades podem expor propriedade intelectual ou segredos comerciais e tornar sua organização suscetível a riscos de fraude. É importante estar vigilante e garantir que sua organização não esteja usando a IA de maneira que viole leis aplicáveis (incluindo leis de privacidade), acordos com clientes ou normas profissionais.

7. Copiar informações ou código produzido por IA em qualquer entrega ou produto pode constituir violação de direitos autorais ou de propriedade intelectual. Isso pode acarretar danos legais e de reputação para sua organização.

8. Espera-se que tanto as versões de código aberto quanto as versões personalizadas de IA generativa continuem sendo integradas em muitos aplicativos, sistemas e processos comuns, desde navegadores de Internet até tecnologias conectadas à IA licenciadas pelas organizações, como os *softwares* baseados em nuvem e programas de mensagens instantâneas.

9. Criar diretrizes de uso seguro dentro da organização é fundamental para garantir a utilização adequada e eficaz de aplicativos de IA generativa. Também é necessário capacitar seus colaboradores, pois o envolvimento humano traz *insights* e compreensão únicos, que a IA generativa, por si só, não consegue replicar.

10. A KPMG adota uma abordagem responsável na concepção, construção e implementação de sistemas de IA de maneira segura, confiável e ética. Essa abordagem ajuda as empresas a acelerar o valor para os consumidores, as organizações e a sociedade.

# 2 Visão geral do mercado

De acordo com a empresa de pesquisa e consultoria Gartner, até 2025, 30% das mensagens enviadas por grandes organizações serão geradas por AI[1]. De acordo com a Pesquisa de Risco de IA, realizada pela KPMG nos Estados Unidos em setembro de 2022, 85% dos respondentes esperam um aumento no uso da IA e dos modelos de análise preditiva. Além disso, na edição de 2022 da Pesquisa de Tecnologia da KPMG nos Estados Unidos, metade dos respondentes afirma ter obtido retorno sobre o investimento em tecnologia de IA.

Os modelos de IA generativa chamaram a atenção no verão de 2022 quando uma imagem gerada pela IA ganhou um concurso de arte[2]. Em novembro eles estavam novamente sob os holofotes após o lançamento do ChatGPT. No entanto, foi durante uma sessão do Fórum Econômico Mundial em janeiro de 2023, quando a presidente e CEO da Microsoft, Satya Nadella, afirmou que a "era de ouro da IA" está em andamento[3], que o "barulho" em torno do ChatGPT realmente começou a intensificar-se, gerando muitas perguntas e conversas com os clientes das firmas-membro da KPMG.

O treinamento desses modelos exige um grande investimento de capital de risco, esforço humano e poder tecnológico. A OpenAI, criadora do ChatGPT, recebeu US$ 1 bilhão da Microsoft[4] e mais um investimento multibilionário da empresa no início de 2023[5], e o Google[6] e a Meta[7] criaram seus próprios modelos de IA generativa. Considerando a gama de aplicativos possíveis, toda uma indústria está sendo construída para tornar os modelos de IA generativa utilizáveis.

Os aplicativos de IA generativa podem ser divididos em cinco categorias: geradores de conteúdo, extratores de informações, chatbots inteligentes, tradutores de idiomas e geradores de códigos:

- **Geradores de conteúdo:** Onde as ferramentas transformadoras generativas pré-treinadas geram conteúdo como postagens de *blog*, *e-mails*, postagens de mídia social, imagens, *web copy* e anúncios.
- **Extratores de informações:** Esses aplicativos podem criar resumos curtos e longos de artigos, notícias, *posts* em *blogs*, documentos jurídicos e muito mais. Algumas empresas os utilizam para elaborar e analisar documentos jurídicos.
- ***Chatbots* inteligentes:** As empresas estão cada vez mais utilizando chatbots inteligentes como assistentes de atendimento. Os *chatbots* interagem de maneira conversacional e podem responder às perguntas de acompanhamento, admitir erros, contestar ideias incorretas e rejeitar pedidos inapropriados.
- **Tradutores de idiomas:** Ferramentas multilíngues que podem traduzir muitos idiomas. Elas têm o potencial de construir interfaces inteiras de *websites*, incluindo sites de tradução.
- **Geradores de códigos:** Os modelos de IA generativa podem converter *inputs* de texto naturais em trechos de código ou aplicativos. Com uma descrição básica ou uma pequena função de *input* de programação, esses modelos podem produzir códigos em várias linguagens de programação e identificar e corrigir *bugs*.

[1] GARTNER. 7 Technology Disruptions That Will Completely Change Sales. Disponível em: <https://www.gartner.com/en/articles/7-technology-disruptions-that-will-completely-change-sales>. Acesso em: jul. 2023.
[2] SMITHSONIANMAG. Art Made With Artificial Intelligence Wins at State Fair. Disponível em: <https://www.smithsonianmag.com/smart-news/artificial-intelligence-art-wins-colorado-state-fair-180980703/>. Acesso em: jul. 2023.
[3] WORLD ECONOMIC FORUM. Satya Nadella Says AI Golden Age Is Here and 'It's Good for Humanity'. Disponível em: <https://www.weforum.org/press/2023/01/satya-nadella-says-ai-golden-age-is-here-and-it-s-good-for-humanity>. Acesso em: jul. 2023.
[4] OPENAI. Microsoft invests in and partners with OpenAI to support us building beneficial AGI. Disponível em: <https://www.openai.com/blog/microsoft-invests-in-and-partners-with-openai>. Acesso em: jul. 2023.
[5] MICROSOFT. Disponível em: <https://www.blogs.microsoft.com/blog/2023/01/23/microsoftandopenaiextendpartnership/>. Acesso em: jul. 2023.
[6] META AI. We're unlocking the possibilities of AI, together. Disponível em: <https://ai.facebook.com/>. Acesso em: jul. 2023.

# 3 O que são modelos de IA generativa?

A inteligência artificial generativa se refere à inteligência artificial que pode gerar conteúdo em vez de simplesmente analisar ou agir com base em dados existentes.

Modelos de IA generativa, como o GPT-4, são construídos e treinados com um conjunto de dados coletados. Eles podem ser generalistas ou especialistas, construídos com base em coleções predefinidas de dados, e são projetados para produzir saídas que ajudam a atender a determinadas solicitações humanas. Alguns modelos podem, por exemplo, prever a próxima palavra com base em frases anteriores ou a próxima imagem com base em descrições de imagens anteriores.

Esse treinamento permite a rápida geração de conteúdo original, incluindo texto, imagens, vídeos e código. Com uma necessidade reduzida de recursos humanos, algumas empresas esperam ser capazes de produzir conteúdo mais rápido e a um custo menor, o que traz oportunidades para criar conteúdos que antes eram muito caros ou demorados. Isso muda fundamentalmente a interação entre humanos e máquinas e abre uma infinidade de casos de uso potenciais.

Essa capacidade de previsão permite que os modelos façam análises. Por exemplo, eles podem ser utilizados para identificar documentos que abordam tópicos descritos no texto inserido.

# 4 Como os modelos de IA generativa funcionam

Os modelos de IA generativa são projetados para produzir conteúdo com base em um conjunto claro de *inputs* e regras.

A aplicação mais comentada recentemente de um modelo de IA generativa é o ChatGPT, um *chatbot* treinado em instruções humanas, criado pelo laboratório de pesquisas OpenAI[8], sediado em São Francisco. A partir de 14 de março de 2023, os assinantes do ChatGPT Plus puderam utilizar o GPT-4, um modelo multimodal de grande porte (LMM) que aceita tanto imagens quanto texto como *input* e gera saídas de texto[9]. Em 23 de março de 2023, a OpenAI lançou *plugins* para o ChatGPT, incluindo seu próprio *plugin* de navegação na *web*. Isso significa que o ChatGPT pode agora acessar fontes e bases de dados de terceiros[10].

ChatGPT significa Chat (com base em conversas) G(generativo) P(pré-treinado) T(transformador). Ele foi refinado usando aprendizado reforçado a partir do *feedback* humano, no qual um modelo de recompensa que representa a preferência humana é treinado para ajudar a tornar as respostas mais humanas e tentar evitar distorções (inventar fatos).

O ChatGPT foi criado como um grande modelo de linguagem (LLM) e desde então evoluiu para uma ampla aplicação generativa multimodal de IA. Agora, a aplicação pode aceitar inputs de imagem e de texto, não somente de texto, como acontecia antes. Combinado com um modelo de rede neural que usa aprendizado não supervisionado para prever resultados, esse tipo de modelo generativo pode determinar os padrões linguísticos mais prováveis e as relações entre o conteúdo que já absorveu.

O adjetivo "grande" refere-se à quantidade de dados nos quais os modelos se baseiam, bem como ao tamanho dos modelos em si, e envolve o treinamento deles com uma enorme coleção de documentos eletrônicos publicamente acessíveis. Quando foi lançado em 2022, o ChatGPT tinha 175 bilhões de parâmetros (valor que controla o comportamento do modelo de aprendizado de máquina: quanto mais parâmetros, maior a capacidade do modelo efetuar análises).

Inicialmente, ele foi treinado com mais de um milhão de conjuntos de dados ou 500 bilhões de *tokens* (palavras ou fragmentos de palavras), incluindo dados do Wikipedia e do The New York Times. Para colocar isso em perspectiva, o ser humano médio fala 860,3 milhões de palavras em sua vida[11]

[8] https://www.forbes.com/sites/cindygordon/2023/02/02/chatgpt-is-the-fastest-growing-ap-in-the-history-of-web-applications/?sh=7510d916678c
[9] https://openai.com/research/gpt-4
[10] ChatGPT plugins (openai.com)
[11] https://openlibrary.org/books/OL3502128M/The_joy_of_lex

, o que torna essa coleção - ou *"corpus"*, na terminologia da área de linguística - equivalente a 300 anos de linguagem.

A versão básica do ChatGPT não está conectada à Internet e foi treinada com material *on-line* até setembro de 2021, o que significa que seu conhecimento não está atualizado. Novas implementações lançadas para um pequeno número de desenvolvedores premium, como um *plug-in* para o mecanismo de busca do Bing[12], estão conectadas à Internet e oferecem conteúdo mais recente.

O ChatGPT é um exemplo de inteligência artificial estreita (ANI). Sistemas ANI são adequados para realizar unicamente o tipo de tarefa para a qual foram treinados. Por exemplo: um sistema ANI projetado para gerar imagens provavelmente não será capaz de resolver problemas matemáticos.

De acordo com a OpenAI, o GPT-4 - embora ainda não seja totalmente confiável - é substancialmente mais confiável e capaz de lidar com instruções detalhadas em comparação com seu antecessor, o GPT-3.5. Mais significativamente, ele passou em um exame simulado da ordem dos advogados entre os 10% melhores participantes. Em comparação, o GPT-3.5 obteve pontuação entre os 10% inferiores no mesmo exame simulado. A OpenAI destaca que as limitações do GPT-4 são semelhantes às dos modelos anteriores, incluindo a possibilidade de inventar fatos e cometer erros de raciocínio[13].

Continue lendo para aprender sobre possíveis casos de uso para modelos de IA generativa.

# 5 Oportunidades potenciais e casos de uso

A ascensão meteórica do ChatGPT em popularidade se deve, em parte, ao fato de que qualquer pessoa pode usá-lo, mesmo aqueles que não têm conhecimentos técnicos. Seu rápido crescimento de usuários - 100 milhões até fevereiro de 2023[14] - é um sinal do entusiasmo das pessoas em usar essa tecnologia. E quanto mais usuários um *chatbot* tem, mais aprimorada a IA subjacente tende a ficar.

O ChatGPT tem o potencial de transformar negócios ao automatizar e executar tarefas baseadas em linguagem com velocidade e eficiência sem precedentes. Os LLMs podem ser implementados para ajudar em uma ampla gama de tarefas. Eles podem ser modificados para resumir e classificar documentos legais, responder a perguntas de consumidores, auxiliar consultores especializados e gerar desenhos de engenharia e arquitetura.

Eles podem atuar como ponto de partida para a inspiração humana, fornecendo ideias que podem ser transformadas em pensamentos novos e criativos. Isso os torna adequados para ajudar a gerar relatórios de negócios, apresentações de marketing e código para aplicativos de *softwares*.

Os modelos de IA generativa também podem ter aplicações em TI, auditoria, recursos humanos, operações e muitas outras funções empresariais. Ao explorar esses casos de uso, tenha em mente que, apesar das muitas oportunidades que a IA generativa oferece, elas não são imunes a riscos.

[12] BING. Introducing the new Bing. Disponível em: <https://www.bing.com/new>. Acesso em: jul. 23.
[13] OPENAI. GPT-4. Disponível em: <https://openai.com/research/gpt-4>. Acesso em: jul. 23.
[14] hTHE GUARDIAN. ChatGPT reaches 100 million users two months after launch. Disponível em: <https://www.theguardian.com/technology/2023/feb/02/chatgpt-100-million-users-open-ai-fastest-growing-app>. Acesso em: jul. 23.

Na área de **operações de TI**, os modelos de IA generativa poderiam ser utilizadas para:

### Sistemas de gerenciamento do conhecimento baseados em LMM
Recolher informações de várias fontes de dados em diferentes formatos. Essas informações podem ser consultadas para buscar itens específicos.

### Autosserviço de suporte de TI
Auxiliar os funcionários a lidar com erros de sistemas de TI por meio de instruções de suporte geradas por chatbots de IA conversacionais.

### Codificação ou teste de código
Converter código de uma função para outra - por exemplo, de SQL para Python - ou testar código para ter certeza de que ele funciona.

Em **auditoria/*compliance***, eles podem ajudar com:

### Automatização da revisão de auditoria
Automatizar a coleta de informações para a submissão de auditoria e as revisões detalhadas de auditoria com base em formatos de consulta.

### Avaliação dos requisitos de independência
Avaliar os requisitos de independência para o envolvimento da auditoria, de modo a simplificar os processos de aprovação para certificação de independência.

O potencial uso em **recursos humanos** inclui:

### Seleção de candidatos
Treinar modelos de IA generativa em descrições de empregos e dados de habilidades relevantes para ajudar a identificar candidatos adequados para o trabalho.

### Aplicativos *self-service*
Implementar *chatbots* que podem compartilhar conhecimento e resolver dúvidas de recursos humanos.

Na frente das **operações**, eles podem ajudar com:

### Relatórios de sustentabilidade e ESG
Contextualizar dados ESG e apoiar as operações de relatórios, incluindo a criação de declarações em linguagem simples que descrevem as iniciativas ESG.

### Gerenciamento de eventos virtuais
Coordenar o gerenciamento de eventos com o envio de convites, programação de sessões e resposta às perguntas dos participantes.

### Simplificação das operações de negócios
Desde redigir *e-mails* e preparar propostas até realizar uma análise competitiva e pesquisar para garantir o en-tendimento do mercado.

No campo **financeiro/logístico**, eles podem ajudar com:

### Categorização e validação de pagamentos
Ajudar as organizações a tornar públicas suas contribuições tributárias, por meio da classificação de grandes volumes de dados.

### Elaboração e revisão das condições contratuais
Revisar contratos e destacando cláusulas potenciais de conflito de interesse, além de redigir cláusulas e termos para agilizar o processo de contratos.

As opções de **governança jurídica e organizacional** incluem:

**Preparação de recomendações personalizadas e independentes para investimentos financeiros**
Permitir que as organizações forneçam respostas personalizadas a consultas relacionadas à independência por meio de *chatbots*.

**Apresentação de citações legais e *links* de fontes**
Pesquisar citações legais relevantes e exemplos de casos, auxiliando na identificação de fontes confiáveis.

Possíveis aplicações em **marketing** incluem:

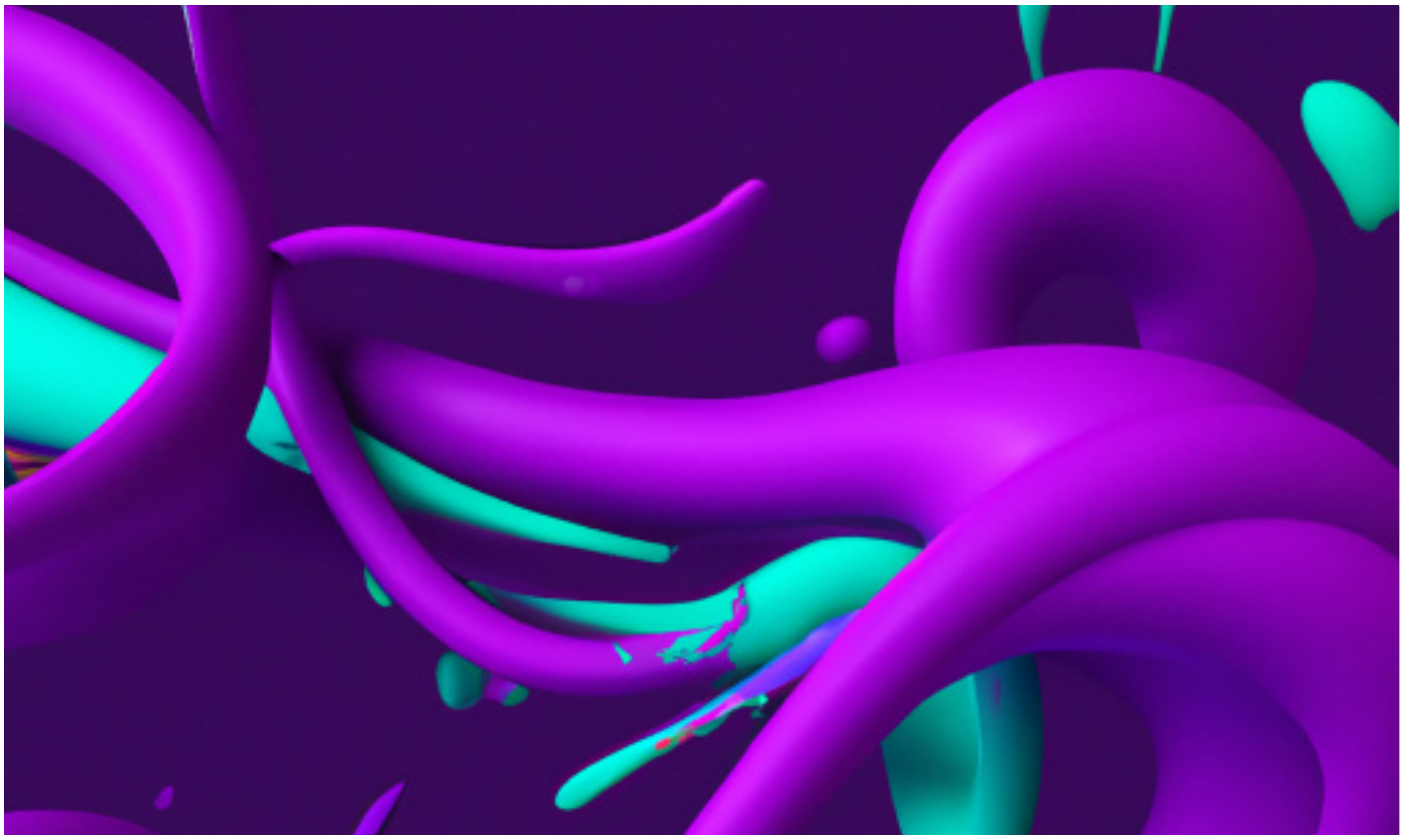
**Simplificação da linguagem de campanhas**
Encontrar opções de palavras que teriam uma tradução simples e adequada em uma variedade de idiomas.

**Localização de comunicações de marketing em grande escala**
Ajudar a localizar campanhas globais, por meio do compartilhamento de dados de conversas locais com o modelo.

**Simplificação de informações complexas**
Aprender o básico, por exemplo, de *due dilligence* financeira, para compreender e estruturar o conteúdo de modo a ajudar a construir uma campanha de marketing sólida.

Na KPMG, temos grande experiência em IA, aprendizado de máquina e análise de dados. Temos *expertise* em análises de risco relacionadas à tecnologia emergente. A KPMG pode ajudar a avaliar a ética, governança e segurança em torno das tecnologias de IA e aprendizado de máquina dos clientes.

As firmas-membro da KPMG têm estado na vanguarda da exploração e do aproveitamento de novas tecnologias e podem responder a perguntas sobre como o uso de IA generativa pode ajudar sua organização a crescer.



# 6 Considerações atuais

Os modelos de IA generativos podem ajudar os consumidores, otimizar processos organizacionais e liberar tempo para os profissionais se dedicarem a tarefas de maior valor. No entanto, o uso de IA generativa tem limitações e potenciais armadilhas.

Conforme mencionamos na parte 4, o GPT-3.5 (LLM subjacente do ChatGPT) foi treinado até setembro de 2021 e não está conectado à internet. Embora a OpenAI tenha possibilitado que o ChatGPT navegue na internet em certos casos, ainda é crucial recorrer à mão de obra humana para revisar o conteúdo gerado por IA. Os modelos de IA generativa podem ser o núcleo de um aplicativo de IA, mas exigem análises adicionais, tecnologia e processos humanos em torno deles para resolver problemas.

A seguir, vamos discutir os riscos associados ao uso de modelos e aplicativos de IA generativa e como gerenciá-los, incluindo a confidencialidade de clientes e empresas, o uso indevido por parte dos funcionários e *phishing*.

Modelos multimodais amplos, como o ChatGPT, geram respostas semelhantes às de humanos. No entanto, eles não possuem as mesmas habilidades de raciocínio de uma pessoa. Por isso, seu uso confiável demanda que os usuários apliquem as capacidades de IA somente a casos adequados. Além disso, cabe às empresas treinar e educar os profissionais para que eles façam uso responsável dessas tecnologias. É igualmente importante que os desenvolvedores usem conjuntos de dados confiáveis para treinar os modelos de IA e aplicar filtros relevantes de viés e conteúdo.

## Gerenciamento de riscos

Diante da crescente popularidade da IA generativa, é fundamental que sua organização se prepare para a implementação dessa tecnologia de uma maneira responsável e o mais imune possível a riscos. A seguir, discutiremos alguns desafios relacionados a essa questão.

## Riscos e considerações internas

### Quebra de confidencialidade e propriedade intelectual

Muitos modelos de IA generativa são projetados para absorver dados inseridos pelo usuário, de modo a aprimorar os modelos subjacentes ao longo do tempo – em essência, trata-se de ajudá-los a aprender e adquirir conhecimento. Esses dados, por sua vez, poderiam ser usados para responder a uma solicitação de outra pessoa, potencialmente expondo informações privadas ou proprietárias ao público. Quanto mais sua empresa utiliza essa tecnologia, maior a probabilidade de que outras pessoas possam acessar suas informações sensíveis ou confidenciais. Ou seja, sua organização deve buscar maneiras de proteger sua propriedade intelectual, sem renunciar aos benefícios do uso da IA generativa.

### Mau uso pelos funcionários e imprecisões

Mesmo o uso legítimo da IA generativa apresenta riscos. Os modelos geram respostas com base nos *inputs* recebidos, o que acarreta o risco de fornecer conteúdo falso ou malicioso. Para os profissionais de uma empresa usarem essa tecnologia, é fundamental que sejam cautelosos e revisem o conteúdo gerado pela IA com um olhar crítico e ênfase na garantia de qualidade.

Se o conteúdo gerado pela IA generativa contiver imprecisões e estas não forem detectadas, isso poderá afetar resultados ou criar problemas de responsabilidade. Por exemplo: o *bot* de IA generativa Galactica, da Meta, foi criado para resumir informações científicas e ajudar acadêmicos e pesquisadores a encontrar rapidamente artigos e estudos relevantes. Porém, ele produziu várias informações incorretas e citações erradas de cientistas respeitados[15]. Outro *bot* da Meta, o BlenderBot3, foi pego fazendo afirmações falsas e tendenciosas[16] logo após seu lançamento, em agosto de 2022. O *chatbot* Bard, do Google, causou à Alphabet, empresa criadora, uma perda de US$100 bilhões em valor de mercado depois de compartilhar informações incorretas durante sua primeira demonstração.[17] A ocorrência de fatos inventados pelo ChatGPT também é bem documentada[18,19], com a própria OpenAI reconhecendo suas limitações[20].

Outros riscos associados à IA generativa incluem a possibilidade de que a tecnologia possa disponibilizar informações sensíveis, como dados pessoais, que poderiam ser usados para roubo de identidade ou invasão de privacidade. Existem riscos como o de um profissional insatisfeito ou um cliente irritado criarem material falso para prejudicar a reputação de sua empresa ou de um de seus funcionários ou executivos, por exemplo.

### Evolução da IA generativa

À medida que o mundo entende melhor a IA generativa, cresce o número de regulamentações globais voltadas ao tema. É importante manter-se atualizado sobre essas regulamentações, mesmo que você não pretenda usar essa tecnologia.

A IA generativa deve continuar a ser integrada em aplicativos, sistemas e processos comuns, incluindo navegadores de internet e tecnologias conectadas à IA que sua organização poderá licenciar. Por isso, é fundamental estar atento para não fazer uso de IA no âmbito profissional de maneira que viole leis aplicáveis (incluindo leis de privacidade), acordos com clientes e/ou padrões profissionais.

[15] GIZMODO. Meta AI Bot Contributed to Fake Research and Nonsense Before Being Pulled Offline. Disponível em: <https://gizmodo.com/meta-ai-bot-galactica-1849813665>. Acesso em: jul. 2023.
[16] CNN. It didn't take long for Meta's new chatbot to say something offensive. Disponível em: <https://www.cnn.com/2022/08/11/tech/meta-chatbot-blenderbot/index.html>. Acesso em: jul. 2023.
[17] NPR. Google shares drop $100 billion after its new AI chatbot makes a mistake. Disponível em: <https://www.npr.org/2023/02/09/1155650909/google-chatbot–error-bard-shares>. Acesso em: jul. 2023.
[18] NPR. Here's what the latest version of ChatGPT gets right — and wrong. Disponível em: <https://www.npr.org/2023/03/17/1164383826/heres-what-the-latest-version-of-chatgpt-gets-right-and-wrong>. Acesso em: jul. 2023.
[19] YAHOO NEWS. From Wadsworth to Coventry to Green, school districts ask: Is ChatGPT cheating or a tool? Disponível em: <https://news.yahoo.com/factual-errors-inflated-bios-aside-100209244.html>. Acesso em: jul. 2023.
[20] OPENAI. GPT-4. Disponível em: <https://openai.com/research/gpt-4>. Acesso em: jul. 23.

## Questões a considerar:

1. Como garantir a confidencialidade e a precisão ao usar modelos de IA generativa?
2. Como garantir que seus modelos de IA generativa estejam em conformidade com as regulamentações globais que surgem continuamente?
3. Como automatizar a revisão e o gerenciamento das políticas de *compliance*?
4. O que sua equipe precisa saber sobre riscos e benefícios da IA generativa?

### Talentos: possíveis implicações

Resultados de excelência só podem ser alcançados com a performance de *experts* altamente capacitados. Por isso, é fundamental que as empresas capacitem sua força de trabalho e retenham conhecimento proprietário para fornecer os prompts corretos. Na KPMG, por exemplo, disponibilizamos treinamento em IA generativa para os nossos colaboradores por meio de nosso programa "Fundamentos Digitais e de Dados", que oferece conteúdo fundamental sobre a evolução da IA e mostra como construir, implementar e interagir com uma IA confiável.

Os profissionais precisam estar cientes de que não estão apenas usando uma solução - eles estão treinando e evoluindo com ela.

Em um futuro generativo, prevemos que o papel dos profissionais mudará de solução de problemas para definição de problemas, à medida que as equipes trabalham ao lado das máquinas para criar abordagens. As ferramentas de IA generativa são uma interface, não um oráculo.

O ser humano traz *insights* e compreensão únicos, que a IA não é capaz de replicar sozinha. As pessoas fornecem *feedback* crítico para refinar e melhorar o modelo ao longo do tempo e garantir que os outputs tenham acuracidade e atendam aos objetivos almejados.

Coisas grandiosas podem acontecer quando as pessoas e a tecnologia se harmonizam. Por isso, acreditamos que nenhuma mudança será duradoura sem a habilidade humana.

## Riscos externos e considerações

### Desinformação, viés e discriminação

Conforme discutido anteriormente, os LLMs e LMMs compartilharam informações falsas, desatualizadas e discriminatórias, apresentadas de tal forma que até mesmo o leitor mais cético poderia ser enganado.

A IA generativa pode – e tem sido usada – para criar imagens e vídeos *deepfake* (nos quais o conteúdo visual é alterado para dar a impressão de que alguém disse ou fez algo que não disse ou fez). Tais imagens e vídeos podem parecer extremamente realistas e não apresentam os indícios que normalmente encontramos em mídias digitais editadas. Por isso, é muito difícil identificar esses truques.[21]

### *Copyright*

Existem muitas dúvidas acerca da propriedade do conteúdo processado por meio de aplicações de IA generativa, e não há uma resposta que se aplique a todos os casos. Os termos e as condições variam de ferramenta para ferramenta. O modo como o conteúdo gerado será usado também importa.

Se um conteúdo for simplesmente copiado e colado, ou se ele derivar de um texto que pertença a outra pessoa, esteja protegido por direitos autorais e permaneça em grande parte inalterado, é possível que configure plágio. É difícil afirmar até que ponto as informações obtidas por meio de uma ferramenta de IA generativa precisariam ser modificadas para serem legitimamente consideradas como suas.

Reivindicar conteúdo gerado por IA como seu próprio pode levantar uma série de questões éticas. Em primeiro lugar, agir dessa maneira não seria responsável nem confiável; e, se a verdade viesse à tona, provavelmente seus clientes e consumidores duvidariam da sua integridade sob todos os aspectos. Além disso, se os clientes ou consumidores descobrirem que você está simplesmente repassando informações geradas por IA, o que os impediria de fazer o mesmo e eliminar completamente o intermediário (ou seja, sua organização)?

A seguir, aprofundaremos essa discussão sobre o risco reputacional associado à IA generativa.

### Riscos à marca, financeiros e reputacionais

Se você ou alguém em sua organização copiar informações ou código produzidos por IA em qualquer entrega ou produto, isso pode constituir violação de direitos autorais ou de propriedade intelectual, acarretando danos legais e reputacionais à sua organização.

Embora muitas dessas ferramentas instruam explicitamente os usuários a não inserir informações confidenciais de clientes, não é improvável que usuários sem treinamento e compreensão adequados exponham inadvertidamente propriedade intelectual ou segredos comerciais ao público e até à concorrência. Falhas dessa natureza poderiam levar a processos judiciais e afetar negativamente os resultados financeiros da sua empresa, na eventualidade de clientes e consumidores atuais ou em potencial questionarem se você e sua organização podem, de fato, ser considerados confiáveis.

## Questões a considerar:

1. Como garantir que as aplicações de IA generativa sejam gerenciadas de forma eficaz para evitar penalidades financeiras devido à falta de conformidade com as regulamentações?
2. É possível confiar nos aplicativos que você usa?
3. Como gerenciar proativamente seus aplicativos e estar ciente e vigilante em relação a possíveis vieses ou discriminação?
4. O uso de aplicativos de IA generativa está em conformidade com sua ética, valores e marca?

[21] PROPERTY CASUALTY. Deepfakes: An insurance industry threat. Disponível em: <https://www.propertycasualty360.com/2021/09/14/deepfakes-an-insurance-industry-threat/>. Acesso em: jul. 2023.
[22] GIZMODO. CNET Cops to Error Prone AI Writer, Doubles Down on Using It. Disponível em: <https://gizmodo.com/cnet-artificial-intelligence-writing-scandal-1850031292>. Acesso em: jul. 2023.

A falta de transparência ao usar conteúdo gerado por IA também pode causar problemas reputacionais. A editora de tecnologia CNET foi criticada por usar silenciosamente a tecnologia para escrever mais de 70 artigos desde novembro de 2022[22]. Muitos dos textos produzidos continham erros. Embora a editora afirmasse em seu site que uma equipe de editores estava envolvida no conteúdo "desde a concepção até a publicação", sua credibilidade ficou abalada.

## Segurança cibernética

Cibercriminosos podem usar a IA generativa para criar golpes de *phishing* mais realistas e sofisticados ou credenciais para *hackear* sistemas. Além disso, os algoritimos de IA não conseguem proteger seus conjuntos de dados de treinamento subjacentes (*datasets*).Estudos têm mostrado que algoritmos podem distinguir identidades individuais mesmo que os dados tenham sido "limpos" e deixados anônimos[23].

Outros riscos de segurança cibernética relacionados à IA generativa incluem "envenenamento" de dados, ou seja, a manipulação mal intencionada dos dados usados para treinar os modelos. Também constituem risco as tentativas de enganar os modelos de IA generativa, alimentando-os com *inputs* maliciosos.

À medida que sua organização explora os casos de uso do ChatGPT e outros aplicativos de IA generativa, convém que as equipes de cibersegurança e gerenciamento de risco estabeleçam diretrizes e regulamentações seguras de implementação. Isso pode incluir a definição das expectativas para o uso do ChatGPT e de outras soluções no contexto corporativo, a conscientização da equipe para um melhor entendimento dos benefícios e riscos associados ao uso de IA generativa e a implementação de controles de cibersegurança.

### Questões a considerar:

1. Quão seguros são seus aplicativos de IA contra a ação de *hackers*, agentes maliciosos e ameaças internas?
2. Seus controles de segurança estão funcionando? Como podem ser aperfeiçoados?
3. Os aplicativos que você utiliza violam a privacidade de alguém?

## Ataques adversários

Mesmo quando treinados para operar dentro de limites aceitáveis, os modelos de IA generativa têm se mostrado vulneráveis, assim como qualquer modelo analítico, à manipulação deliberada de usuários externos sofisticados. Se a sua organização planeja utilizar soluções de IA generativa, é preciso estar ciente de que esse risco existe quando a solução estiver exposta ao público.

### Questões a considerar:

1. Quais são as vulnerabilidades adversárias básicas das tecnologias que você está utilizando?
2. Como você pode testar os ataques prováveis e fortalecer as soluções existentes (ou implementar novas) para se preparer e enfrentá-los?
3. Que monitoramento você tem em vigor para identificar ataques adversários?

## Apoio ao uso adequado de IA generativa

Recomendamos criar diretrizes de uso seguro para a sua organização: isso é fundamental para garantir o uso adequado e eficaz de aplicativos de IA generativa. Tais diretrizes podem incluir treinamento obrigatório para qualquer pessoa que queira usar IA generativa e normas a respeito de como se deve (e não se deve) utilizá-la. Além disso, sua organização deve tratar a IA generativa como qualquer outra solução tecnológica e exigir que os funcionários sigam, por exemplo, as políticas de uso aceitável ou segurança da informação já existentes.

Ainda há muito trabalho a ser feito antes de podermos utilizar a última geração de IA em interações com consumidores, funcionários, cidadãos e negócios. Com um programa de IA responsável em vigor, as organizações podem avançar no desenvolvimento de processos e procedimentos em torno do uso de IA generativa.

[23] TECH CRUNCH. *Researchers spotlight the lie of 'anonymous' data*. Disponível em: <https://techcrunch.com/2019/07/24/researchers-spotlight-the-lie-of-anonymous-data/>. Acesso em: jul. 2023.

# 7 O que o futuro reserva?

Observando o que os *players* de tecnologia estão explorando na IA generativa, temos uma ideia de como esse campo tende a se desenvolver.

## Desenvolvimento e manutenção de *software*

A IA generativa está mostrando potencial para incrementar o processo de desenvolvimento de *software* e, assim, propiciar a entrega mais rápida de produtos e serviços cada vez mais confiáveis. Muitas empresas serão capazes de automatizar totalmente processos como geração de código, manutenção e correção de *bugs*.

## Criação de vídeos e realidade virtual

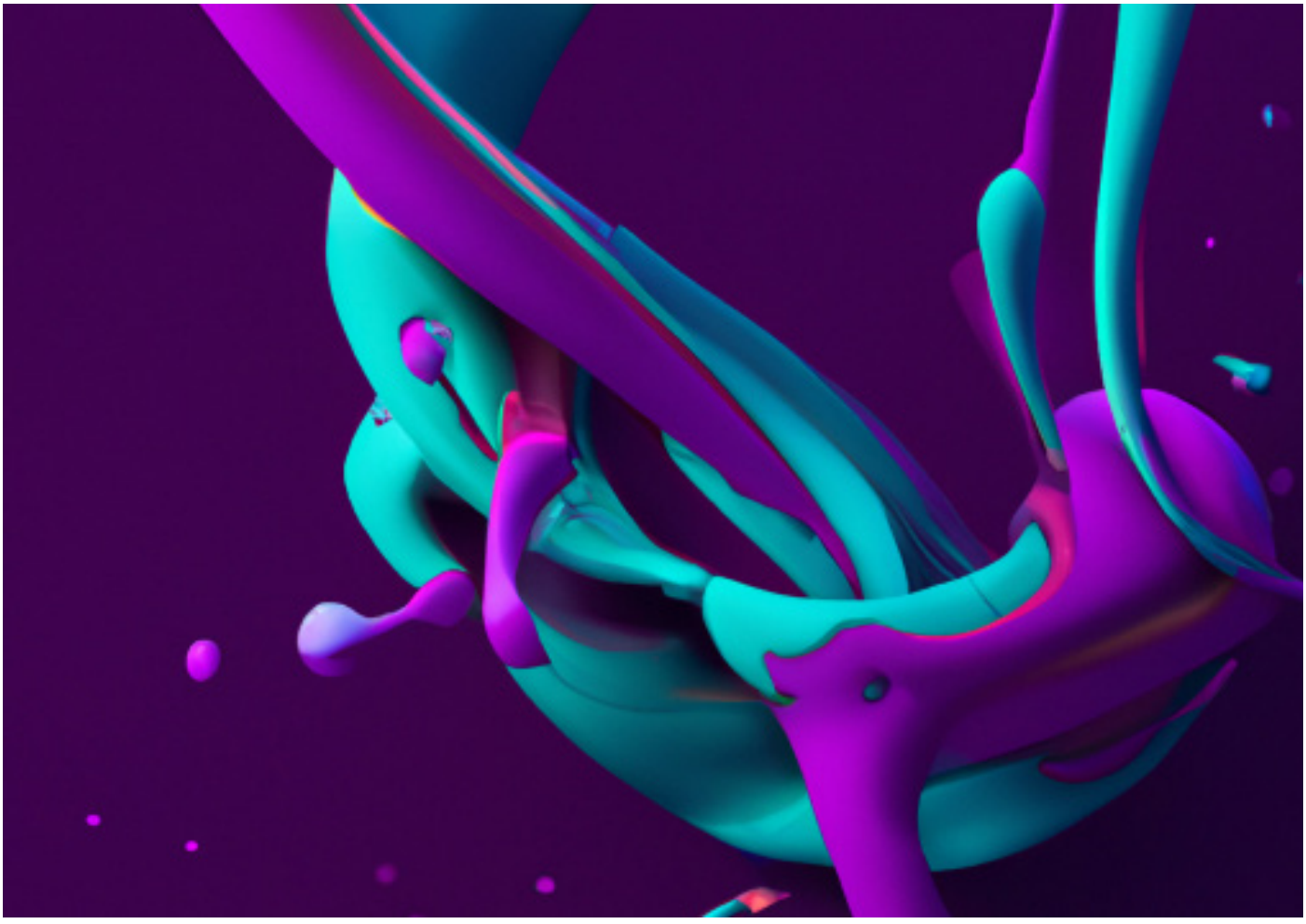
A IA generativa pode criar ambientes imersivos de videogames, projetar conteúdo visual e até personalizar vídeos de produtos para *sites* de comércio eletrônico. No futuro, as empresas poderão aproveitá-la como assistente virtual ou em aplicações de transmissão ao vivo – por exemplo, com a geração automática de legendas. Muitas empresas nesse setor estão agora dirigindo seu foco para clientes corporativo.

## O metaverso

Criar ativos 3D realistas no metaverso é caro e consome muito tempo. A IA generativa pode gerar ativos 3D por meio de texto para imagem ou voz, cenas 3D com base em imagens 2D e até produzir efeitos sonoros. Também pode gerar rostos humanos e dar características mais realistas aos avatares do metaverso.

## Melhoria na segurança da informação

A IA generativa pode ensinar as pessoas sobre os principais riscos associados a determinados pontos vulneráveis, ajudando-os a criar *scripts* apropriados e a entender métodos de ataque por parte de agentes maliciosos.



# 8 Como a KPMG pode ajudar

Por mais de 150 anos, as firmas-membro da KPMG têm desempenhado papel de destaque na exploração e no controle de novas tecnologias, como a IA generativa, oferecendo garantia e orientação para sua implementação.

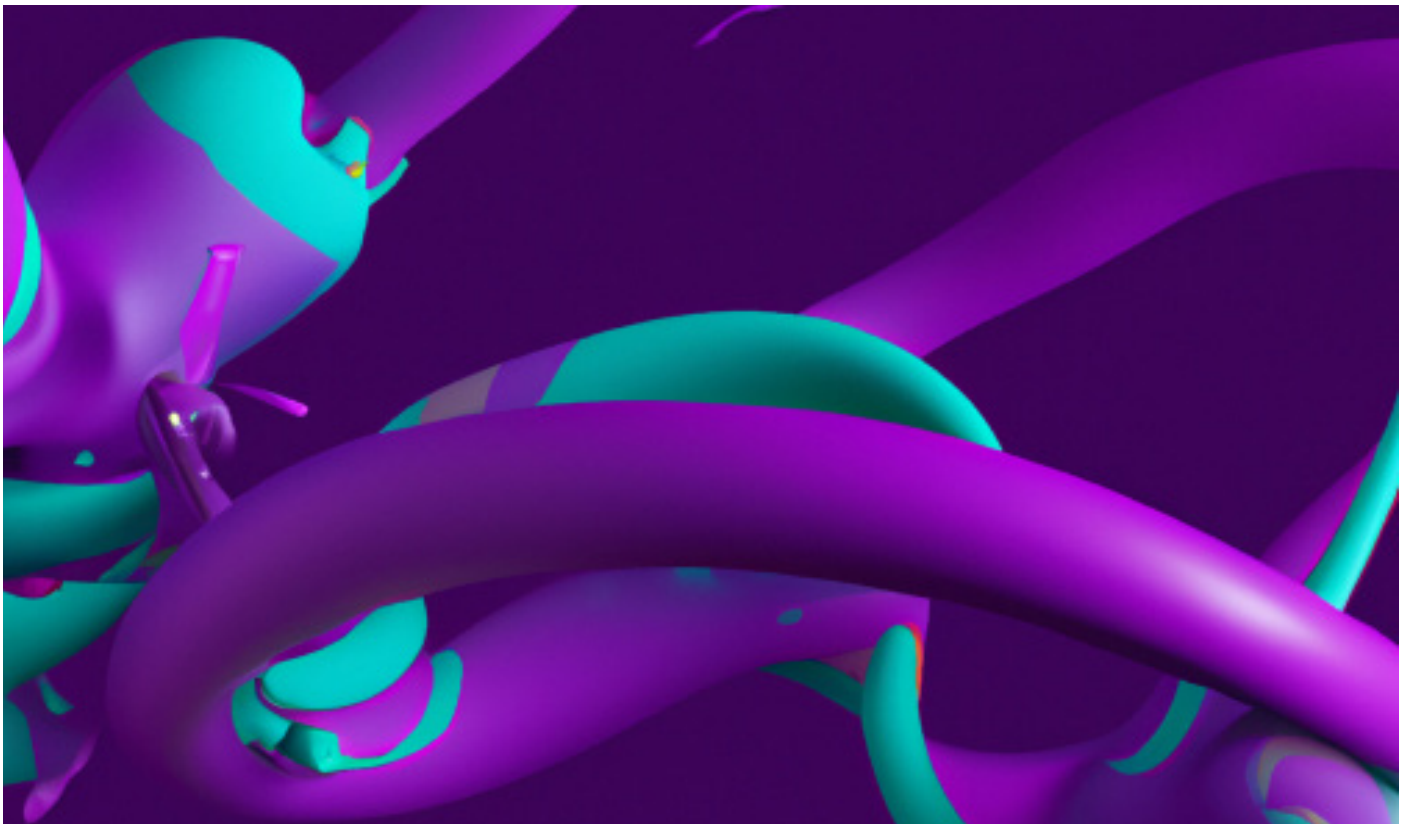
A Global Lighthouse é a rede global da KPMG, composta por mais de 15 mil especialistas em dados e análise, IA e tecnologias emergentes, com presença em 37 países nas Américas, Ásia-Pacífico e Europa. Compreendemos que a aplicação responsável de IA é um desafio complexo em termos de negócios, regulamentação e tecnologia. Por meio da Global Lighthouse e da rede de firmas-membro da KPMG, temos o compromisso de apoiar nossos clientes na implementação das soluções ideais a cada caso.

## Utilizando a IA generativa de forma responsável

A Global Lighthouse auxilia as organizações a desenvolver soluções de IA responsáveis, confiáveis e seguras. Além disso, a KPMG adota uma abordagem que envolve análises de natureza ética, de governança e de segurança em relação às tecnologias de IA e ao machine learning dos clientes. Nosso conjunto de estruturas, controles, processos e ferramentas respaldam o melhor aproveitamento de tudo que a IA oferece, projetando, construindo e implementando sistemas de maneira segura e confiável. Assim, nossos clientes podem acelerar a criação de valor que beneficiará os consumidores, as organizações e a sociedade.

Nossa abordagem responsável de IA inclui:

1. **Equidade:** garantia de que os modelos sejam equitativos e livres de viés.
2. **Clareza:** garantia de que a IA possa ser compreendida, documentada e aberta para revisão.
3. **Responsabilidade:** garantia da vigência de mecanismos que fomentem a responsabilidade em todo o ciclo de vida da IA.
4. **Integridade dos dados:** garantia de que as etapas de qualidade, governança enriquecimento dos dados fortaleçam a confiança.
5. **Confiabilidade:** garantia de que os sistemas de IA atingirão os patamares desejados de precisão e consistência.
6. **Proteção:** proteção contra acesso não autorizado, corrupção ou ataques.
7. **Privacidade:** garantia do cumprimento das normas regulatórias relativas à privacidade de dados e ao uso dos dados do consumidor.
8. **Segurança:** garantia de que a IA não afetará negativamente seres humanos, propriedades ou o meio ambiente.



# Fale com o nosso time


**David Kerry**
Sócio-diretor de Forensic &
Litigation da KPMG no Brasil
dkerry@kpmg.com.br

**Diogo Dias**
Sócio-líder de Risk Advisory Solutions
da KPMG no Brasil e na América do Sul
dsdias@kpmg.com.br

**Frank Meylan**
Sócio-líder de Tecnologia e Inovação da
KPMG no Brasil e na América do
fmeylan@kpmg.com.br

**Ricardo Santana**
Sócio-líder de Data & Analytics, Automação
e Inteligência Artificial da KPMG no Brasil
santana@kpmg.com.br

**Emerson Melo**
Sócio-líder de Forensic & Litigation da KPMG
no Brasil e Colíder para a América do Sul
emersonmelo@kpmg.com.br

**Fernando Lage**
Sócio-líder de Governança, Riscos
& Ccompliance da KPMG no Brasil
flage@kpmg.com.br

**Leandro Augusto Marco Antônio**
Sócio-líder de Cyber Security & Privacy
da KPMG no Brasil e na América do
lantonio@kpmg.com.br


kpmg.com.br

/kpmgbrasil