

Desvendando a GenAI: conhecendo as complexidades e oportunidades

A GenAI está pronta para redefinir a natureza do trabalho, fazendo parte de nossas vidas cotidianas. No entanto, é importante reconhecer que essa tecnologia ainda está em estágio inicial. Compartilhamos três armadilhas comuns que podem comprometer os esforços até mesmo dos pioneiros mais ousados que buscam obter o sucesso e a inovação prometidos pela GenAI - e como lidar com essas armadilhas.



As expectativas são reais



O ChatGPT atingiu 100 milhões de usuários em apenas dois meses (Facebook: 4,5 anos; TikTok: 9 meses).¹

No final de 2022, “GPT” e “ChatGPT” se tornaram palavras da moda nas áreas de tecnologia e negócios, representando o que há de mais avançado em inteligência artificial generativa (GenAI). Essa tecnologia existia em segundo plano há pouco mais de cinco anos², antes de chamar a atenção de todos e se tornar uma das inovações adotadas mais rapidamente na história recente.



Em 20 anos acompanhando a evolução da Internet, não recordamos de um avanço tão rápido de uma aplicação de Internet para o consumidor.³

A GenAI alcançou um público global, desencadeando uma onda de aplicações que despertaram a criatividade e a inovação de maneira jamais vista. Desde a criação de arte e melodias até a produção de textos que se assemelham à escrita humana, a GenAI está reformulando nossa interação com a tecnologia e abrindo uma nova dimensão de possibilidades.

Com sua capacidade de simular conversas humanas e ajudar na tomada de decisões críticas, a GenAI tem um potencial de disrupção significativo para as empresas e os setores.

Conforme entramos em uma era em que humanos e máquinas inteligentes trabalham juntos para alcançar mais do que qualquer um deles poderia alcançar sozinho, a GenAI está pronta para redefinir a natureza do trabalho, tornando-se parte de nossas vidas cotidianas, muitas vezes operando invisivelmente em segundo plano.

No entanto, é importante reconhecer que essa tecnologia ainda está em seus estágios iniciais. Neste guia, apresentamos três armadilhas comuns que podem comprometer os esforços até mesmo dos pioneiros mais ousados no mundo da GenAI.



A IA generativa (GenAI) já superou o pico das expectativas infladas, já que o foco dos negócios está deixando de ser a empolgação com os modelos fundamentais e passando a ser os casos de uso que geram ROI.⁴

¹AIPRM. 100+ ChatGPT Statistics 2024. | Acessado em 9 de setembro de 2024.

²NTT DATA Corporation. Languages, models and business applications | Acessado em 9 de setembro de 2024.

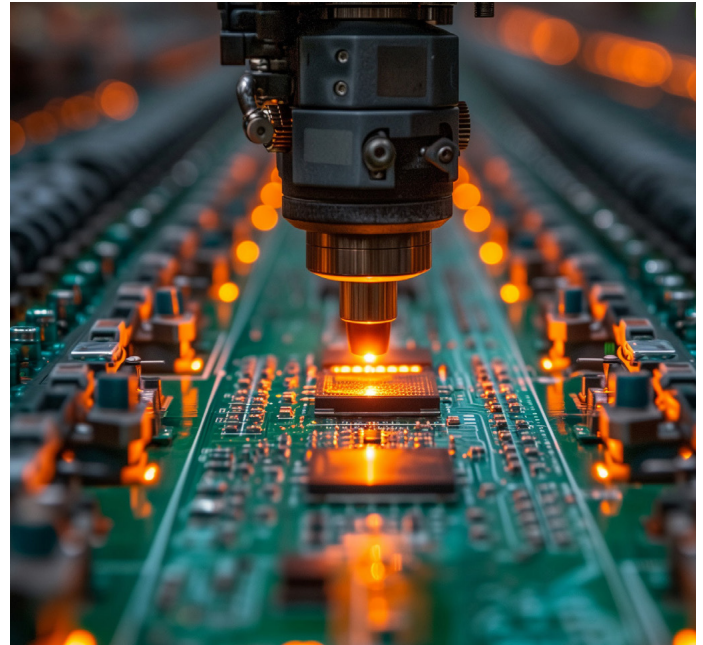
³Analistas da USB, citados em: Krystal Hu, Reuters (2 de fevereiro de 2023). ChatGPT sets record for fastest-growing user base – nota de analista. | Acessado em 9 de setembro de 2024.

⁴Gartner (17 de junho de 2024). Hype Cycle for Artificial Intelligence | Acessado em 9 de setembro de 2024.

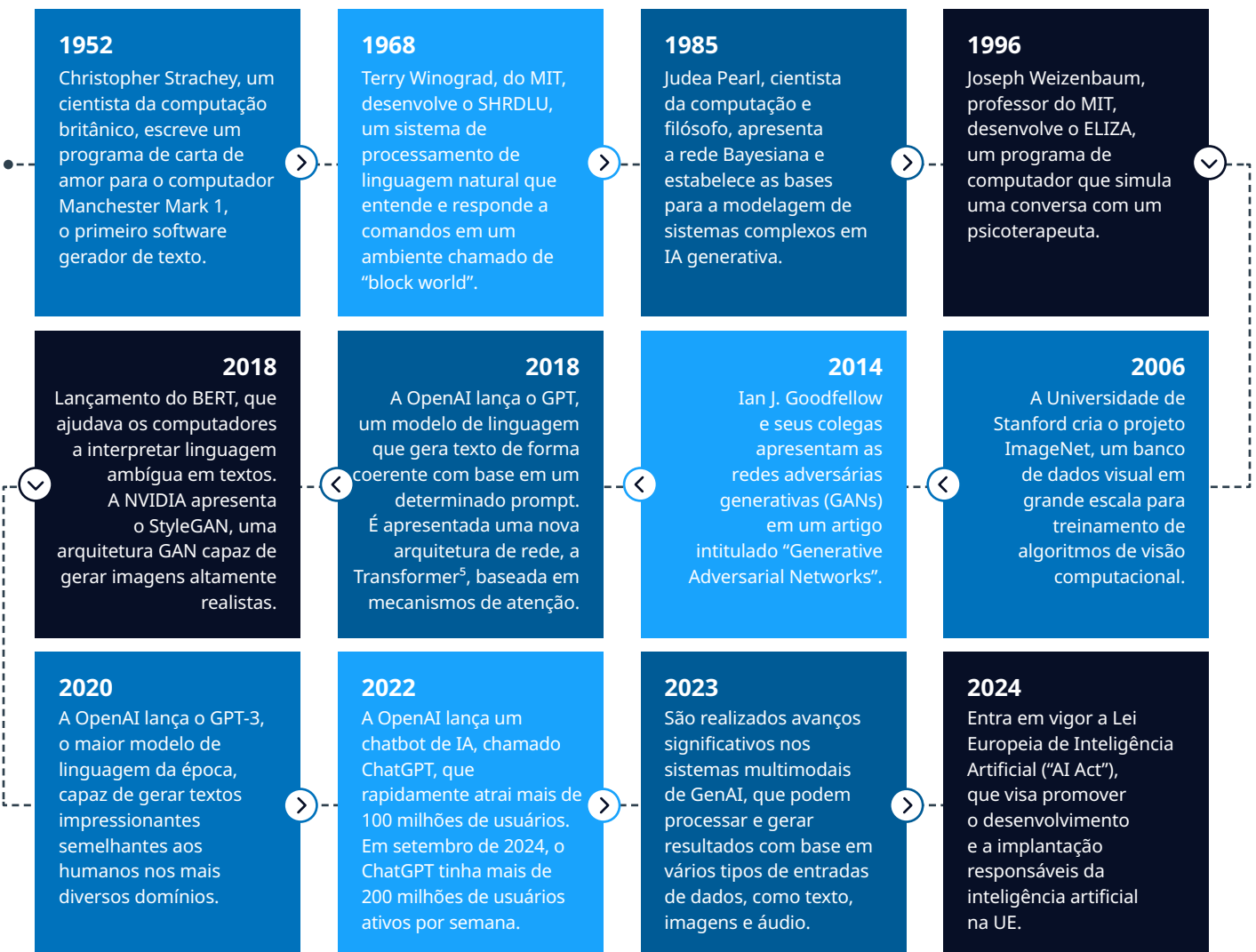
A evolução da GenAI

A evolução da GenAI foi simplesmente impressionante nos últimos anos. A qualidade e os recursos desses modelos de IA passaram por uma profunda transformação, desde os primeiros experimentos básicos até as aplicações no mundo real. Desde a área de saúde e análise de dados até a criação de conteúdo e entretenimento, não há praticamente nenhum campo ou setor em que a GenAI não tenha deixado sua marca.

Com os avanços contínuos em machine learning e redes neurais profundas, a GenAI promete redefinir a forma como interagimos com a tecnologia.



GenAI: 50 anos em desenvolvimento



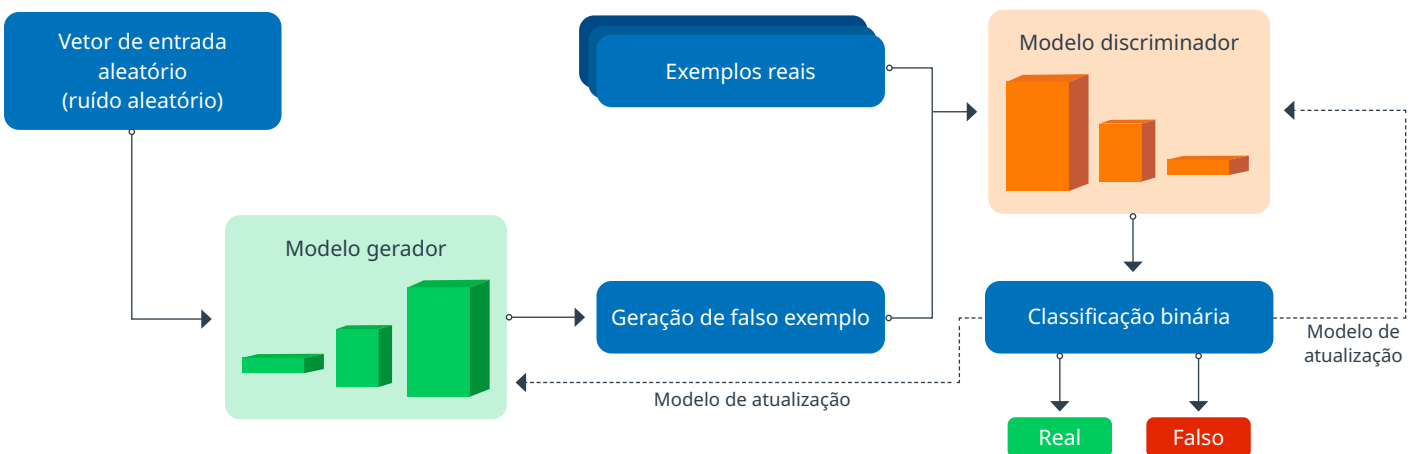
⁵Ashish Vaswani et al. (2 de agosto de 2023). Attention Is All You Need. | Acessado em 9 de setembro de 2024.

Desmistificando a GenAI

A GenAI, em sua essência, é um subcampo da inteligência artificial com foco na criação de algoritmos e modelos capazes de gerar dados semelhantes ao conteúdo criado por humanos. Esses algoritmos usam redes neurais, com abordagens como autoencoders variacionais (VAEs, na sigla em inglês) para dados estruturados, redes adversárias generativas (GANs, na sigla em inglês) para geração de conteúdo altamente realista e redes neurais recorrentes (RNNs, na sigla em inglês) ou transformadores para dados sequenciais, como texto ou fala. O processo de treinamento envolve a exposição desses modelos a extensos conjuntos de dados, permitindo que aprendam padrões e correlações dentro desses dados.

O colapso do modo em GANs e a garantia de dados de treinamento de alta qualidade são algumas das considerações cruciais no desenvolvimento e na implantação ética da GenAI. A GenAI encontra aplicações em diversos campos, desde a geração de linguagem natural para chatbots e criação de conteúdo até a síntese de imagens para arte e design. Ao desmistificar a GenAI e compreender seus princípios técnicos fundamentais, as empresas e os pesquisadores podem utilizar a GenAI para automatizar a geração de conteúdo, aprimorar os processos criativos e inovar em uma ampla gama de áreas, sem deixar de lado as considerações éticas e a qualidade dos dados.

Arquitetura de rede adversária generativa (GAN)



Os modelos de GenAI utilizam VAEs como um “livro de receitas” para entender a essência de vários tipos de dados, combinando-os para criar conteúdo novo e exclusivo. Os GANs agem como um crítico gastronômico (discriminador) que dá feedback sobre se os resultados são bons ou ruins. Essa interação contínua ajuda os modelos a refinar suas técnicas, resultando na criação de resultados realistas e convincentes.



Onde, como e por quê?



How do I hyperlink an image on a website?



To hyperlink an image on a website, you can use the HTML `<a>` (anchor) tag along with the `` (image) tag. This allows you to create a clickable image that will direct users to another web page when clicked. Here's the basic structure:

```
html Copy code  
  
<a href="URL_OF_THE_TARGET_PAGE">  
    
</a>
```

Um simples prompt pode gerar código. Fonte: OpenAI

Principais áreas em que a GenAI está redefinindo tarefas e abrindo oportunidades

Conversas e interações com clientes

Imagine ter a GenAI como sua assistente nas interações com os clientes. Esse bot inteligente e cognitivo compreende a intenção de seus clientes e cria respostas que melhoram a qualidade da experiência deles. O sistema é tão avançado que as conversas geradas são praticamente impossíveis de distinguir de uma interação humana. Na área da saúde, por exemplo, a GenAI pode ajudar os profissionais da área médica fornecendo dados do paciente em tempo real, propondo alternativas de tratamento e respondendo a perguntas médicas em uma linguagem de fácil compreensão.

Essas trocas podem melhorar a compreensão, o atendimento e a eficiência dos pacientes nos ambientes de assistência médica.

Criação de conteúdo

A GenAI traz recursos revolucionários para capacitar editores, escritores e criadores de conteúdo. Ao explorar tendências e conhecimentos globais em tempo real, a GenAI ajuda os criadores a gerar conteúdo envolvente e tem o potencial de transformar a forma como o conteúdo é criado em diferentes domínios.

Em vez de substituir os colaboradores do conhecimento e os redatores de conteúdo, a GenAI deve ser aproveitada para aprimorar e impulsionar seus recursos. A GenAI pode gerar textos e rascunhos de artigos, e-mails e até mesmo poemas.

Essa capacidade se estende ao conteúdo visual e à música, oferecendo uma abordagem moderna e criativa à expressão artística. O DALL-E, por exemplo, pode gerar obras de arte digitais exclusivas e visualmente impressionantes, enquanto o Midjourney ajuda os artistas a gerar novas ideias, refinar sua arte e criar trabalhos mais impactantes.

“ A NTT DATA oferece serviços especializados para administradores públicos e clientes do setor financeiro, ajudando-os a entender regulamentos complexos e questões financeiras.

Gestão do conhecimento

A gestão do conhecimento tinha suas limitações para lidar com extensos conjuntos de dados e extrair insights relevantes. Os métodos tradicionais dependiam de curadoria manual e pesquisa de palavras-chave, o que geralmente resultava em falhas na obtenção de informações.

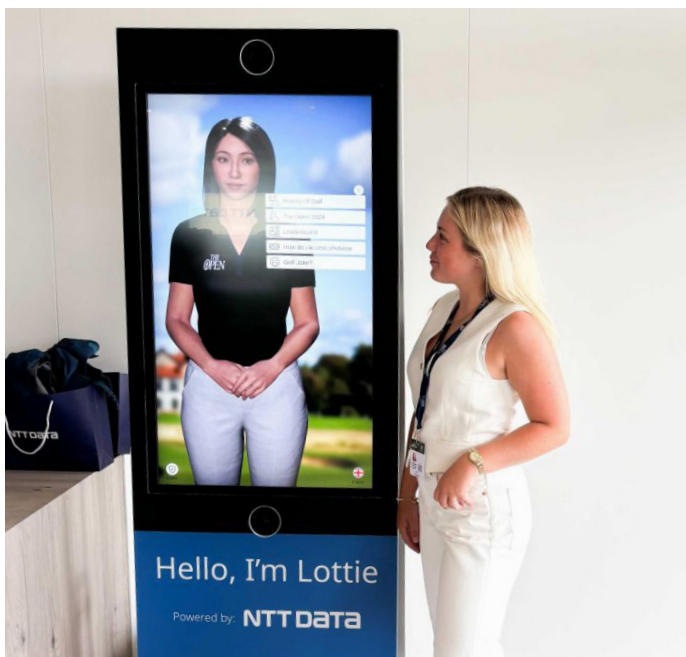
Com a capacidade de lidar com grandes conjuntos de dados, a GenAI pode desempenhar um papel fundamental na gestão do conhecimento e nos ajudar a obter e aproveitar as informações certas de forma mais eficaz. Os modelos baseados em transformadores podem captar nuances contextuais e resumir rapidamente conteúdos extensos.

Essa tecnologia ajuda a aprimorar os recursos dos mecanismos de pesquisa e fornece resultados mais precisos e conscientes do contexto.

Descoberta de conteúdo

Os modelos de GenAI aprimoram a descoberta de conteúdo oferecendo uma experiência personalizada e aprofundada em várias plataformas.

A GenAI pode definir a busca de conteúdo com recomendações personalizadas, fornecendo conteúdo sob medida com base no comportamento do usuário. Com o processamento avançado de linguagem natural, o GenAI refina a compreensão do conteúdo, melhorando a categorização e a marcação para obter algoritmos de pesquisa mais eficientes.



O “humano digital” da NTT DATA, chamado de Lottie, foi apresentado no Open Championship 2023, que aconteceu no Royal Liverpool.

“

A NTT DATA fez uma parceria com a L'Oréal para criar a Lore, uma assistente virtual baseada em nossa plataforma de IA conversacional para reimaginar a experiência do cliente e aumentar as vendas on-line.

Automação de processos

Com um vasto conhecimento contextual e dados históricos, a GenAI pode simplificar e acelerar as interações com os clientes, analisando automaticamente os processos que podem ser automatizados, reduzindo o esforço manual.

As soluções de automação de processos da NTT DATA incluem:

- Dolffia, uma solução GenAI multicloud que otimiza o processamento de documentos, melhora a criação de conteúdo e fornece insights orientados por dados para casos de uso personalizados e escaláveis, específicos por setor e multifuncionais.
- LITRON®, é uma IA de compreensão de documentos que pode ler e compreender textos em japonês com rapidez e precisão. Inclui o LITRON® Generative Assistant, que combina um grande modelo de linguagem (LLM) com a tecnologia LITRON® Generative Assistant para pesquisar documentos internos e fornecer respostas em um formato de chat. Essa tecnologia avançada permite que as organizações aproveitem ao máximo o valor de seus documentos em japonês.

Desenvolvimento e manutenção de software

A GenAI é proficiente em várias linguagens e estruturas de programação conhecidas e pode se converter perfeitamente entre elas. Também pode compreender instruções de problemas escritas por humanos e gerar instantaneamente o código necessário, acompanhando os últimos avanços e considerações de segurança.

O GitHub Copilot é um exemplo de GenAI utilizado em aplicações de programação. Esse assistente de programação funciona como uma ferramenta inteligente de preenchimento de código. Ajuda os desenvolvedores a entender o contexto e a gerar código em tempo real, o que aumenta significativamente a produtividade.

Quatro etapas para integrar a GenAI nas organizações

Incorporar a GenAI às operações empresariais é muito mais do que um trabalho técnico. É uma etapa estratégica que requer precisão, alinhamento com as metas organizacionais e monitoramento e refinamento contínuos.



Etapa 1: Definição de objetivos de negócios e avaliação de casos de uso de impacto

Comece articulando metas de negócios bem definidas que se alinhem com a visão organizacional geral. Isso ajudará você a se concentrar em áreas específicas nas quais a GenAI pode ser aplicada de forma eficaz.

Avalie minuciosamente os possíveis casos de uso da GenAI, levando em consideração fatores como a facilidade de implementação, o retorno sobre o investimento (ROI, na sigla em inglês) projetado e como esses casos de uso estarão alinhados aos objetivos da empresa. Considere a conformidade com a regulamentação do setor e as leis de privacidade de dados em cada etapa.

Isso garante que a adoção da GenAI não seja apenas um avanço tecnológico, mas um processo estratégico com impacto significativo nos resultados organizacionais.



Etapa 2: Descoberta do projeto e planejamento detalhado

Esta etapa envolve a identificação de desafios técnicos (distintos de questões comerciais mais amplas) e soluções. Isso inclui a definição de modelos, estratégias de ajuste fino e integração com bases de conhecimento externas, assim como o conjunto de tecnologias adequado, abrangendo serviços de cloud, estruturas, bibliotecas e bancos de dados vetoriais.

Também é importante estabelecer métricas de sucesso técnicas e não técnicas e realizar uma avaliação abrangente dos custos.

A engenharia imediata desempenha um papel fundamental na obtenção de sistemas de IA eficientes e deve ser monitorada ativamente para reduzir computações desnecessárias e otimizar a utilização de recursos.



Etapa 3: Prova de conceito e implementação piloto

Antes de realizar uma implantação em larga escala, é essencial conduzir uma prova de conceito minuciosa em pequena escala, em um ambiente controlado, para validar a viabilidade de sua abordagem. Isso permite que você identifique os desafios antecipadamente e refine sua estratégia conforme for necessário.

Uma interface fácil de usar e integrada aos sistemas existentes ajudará a obter feedback para melhoria contínua, ajuste fino e otimização.



Etapa 4: Implementação completa - concretização da visão; otimização e manutenção contínua

A partir da validação do piloto e das informações recebidas nas etapas anteriores, podemos executar uma implantação em escala total da GenAI.

Essa fase envolve tanto a implementação tecnológica quanto a incorporação estratégica do modelo refinado na estrutura organizacional.

Para continuar refinando a ferramenta para atender às necessidades organizacionais em constante mudança, é necessário um ciclo de otimização, manutenção e governança permanentes após a implementação.

Defina um modelo de referência e uma arquitetura ou estrutura empresarial padronizada que forneça um projeto para os resultados desejados. Isso ajudará você a atender os fatores críticos para obter sucesso ao ir da prova de conceito à produção e, ao mesmo tempo, habilitar a escalabilidade, o desempenho e o monitoramento da segurança, da privacidade e da governança.

Evite as armadilhas mais comuns na implementação da GenAI

Apesar das inúmeras vantagens que a GenAI oferece, muitas empresas hesitam em realizar uma implantação completa. Um dos principais motivos para isso é a complexidade dos sistemas da GenAI. Houveram alguns problemas de “caixa preta” com a IA – falta de capacidade de visualização de como os sistemas de aprendizagem profunda tomam suas decisões. Além disso, as preocupações relacionadas à escalabilidade, à compatibilidade com os sistemas existentes e à necessidade de talentos especializados.

Considerações éticas estão na linha de frente dessas preocupações. É fundamental garantir a transparência, a responsabilidade, a privacidade dos dados e o potencial de uma colaboração harmoniosa entre humanos e IA.



Perda da experiência humana

A dependência excessiva da IA pode levar a um declínio das habilidades e do conhecimento humano, assim como a uma possível perda de criatividade e pensamento crítico.



Falta de comunicação

Funções e expectativas incertas na colaboração entre humanos e IA podem levar a equívocos e resultados abaixo do esperado.




Disrupção no ambiente de trabalho

Sistemas de IA mal integrados podem causar disrupção nos fluxos de trabalho existentes e exigir uma adaptação significativa, levando à resistência dos colaboradores.


É fundamental abordar os problemas que podem ser negligenciados involuntariamente durante a implementação da GenAI.

Privacidade e segurança dos dados

A privacidade e a segurança dos dados são fundamentais na implementação da GenAI, pois os modelos de IA geralmente exigem acesso a informações confidenciais. Deixar de priorizar a privacidade e a segurança dos dados pode levar a sérias consequências.

 Implemente uma criptografia de dados robusta, controles de acesso e anonimização de dados.

Para estar em conformidade com as estruturas regulatórias, como o Regulamento Geral de Proteção de Dados (GDPR) e a Lei de Privacidade do Consumidor da Califórnia (CCPA), é essencial estabelecer quem tem controle sobre os dados (soberania dos dados) e onde os dados são armazenados (residência de dados), o que acrescenta uma camada adicional de complexidade.

 Realize auditorias de segurança regulares para garantir a conformidade com as leis de proteção.

Colaboração entre humanos e IA

A colaboração eficiente entre humanos e a GenAI é essencial para fluxos de trabalho produtivos e eficazes. Deixar de encontrar esse ponto de equilíbrio pode levar a frustrações e falta de eficiência.



Projete cuidadosamente fluxos de trabalho que incorporem a GenAI como uma ferramenta de apoio e não como uma ferramenta de substituição.



Treine e capacite os colaboradores em habilidades relacionadas à IA para promover uma colaboração mais eficaz entre humanos e a inteligência artificial.

Medição da eficiência e do ROAI

Embora a GenAI prometa uma produtividade incomparável, é essencial medir a eficiência e o impacto da GenAI.

Calcular o retorno sobre o investimento em IA (ROAI, Return on Artificial Intelligence) ajuda você a entender o valor econômico gerado pela GenAI e a avaliar os benefícios de longo prazo de seu investimento em IA.



Estabeleça indicadores-chave de desempenho (KPIs) e valide-os em relação às metas de negócios para determinar seu ROAI.

Da prova de conceito à produção: seleção de amostras e framework

A transição da prova de conceito para a produção requer uma amostra que permita testar aplicações críticas com as soluções de GenAI.



Escolha amostras que sejam relevantes em cenários do mundo real.



Após a transição, adote o framework correto para integrar a GenAI aos sistemas de TI existentes.

Entendendo as limitações da GenAI



Os modelos de GenAI são suscetíveis a gerar conteúdo que pode parecer real, mas não tem base na realidade.



Os modelos de GenAI podem promover vieses, levando a resultados discriminatórios ou injustos.



Os sistemas GenAI são baseados em recursos estatísticos. Entretanto, não possuem a capacidade de um verdadeiro raciocínio humano.




Ter um contexto reduzido com alguns milhares de palavras significa que o modelo terá dificuldades com o reconhecimento abrangente do contexto.

Embora a GenAI seja uma grande promessa, ainda tem suas limitações. Essas limitações nos lembram da necessidade de discernimento e análise criteriosa.


Alucinações

Assim como os sonhos podem tomar rumos inesperados e sem sentido, os modelos de GenAI são suscetíveis a gerar conteúdo que não tem base na realidade. Essas “alucinações” da IA podem resultar em resultados que, embora criativos, são factualmente incorretos ou sem lógica.

 A validação humana e o reconhecimento do contexto são necessários para interpretar o conteúdo da GenAI.


Viés

A GenAI aprende com vastos conjuntos de dados, que podem inadvertidamente conter preconceitos presentes na sociedade. Consequentemente, os modelos de IA podem perpetuar esses vieses, levando a resultados discriminatórios ou injustos.

 A supervisão rigorosa e a compreensão das considerações éticas são fundamentais para reconhecer e atenuar os preconceitos no desenvolvimento e na implantação de sistemas de GenAI.

Falta de raciocínio humano

Embora a GenAI seja excelente em reconhecimento de padrões e geração de conteúdo, não tem capacidade de raciocínio humano verdadeiro. A GenAI não é capaz de compreender conceitos, fazer julgamentos com base em princípios éticos ou morais ou tomar decisões complexas como os humanos costumam fazer.

 A supervisão e a orientação humanas são necessárias ao usar a GenAI em aplicações críticas.

Visão geral do contexto limitado

Uma abordagem de geração aumentada de recuperação (RAG, Retrieval Augmented Generation) pode analisar informações relevantes de um grande volume de texto para gerar um texto de alta qualidade que seja coerente e diversificado. Isso leva a respostas mais precisas e informativas, tornando a GenAI ideal para aplicações como resumo de texto, tradução automática, chatbots e criação de conteúdo.

No entanto, devemos ter em mente que a GenAI opera dentro das restrições dos dados com os quais foi treinada. GenAI pode ter dificuldades para fornecer uma percepção abrangente do contexto, especialmente em situações em que as informações necessárias vão além dos dados de treinamento. Isso pode resultar em respostas sem nuances e que não levam em consideração fatores contextuais mais amplos.

Compreender e navegar por essas limitações é essencial para usar a GenAI de forma eficaz. Ao reconhecer esses limites, as organizações podem tomar decisões informadas sobre o uso da GenAI e garantir que suas aplicações estejam alinhadas com as considerações éticas e práticas.



Regulamentação - riscos e complexidades

A regulamentação da GenAI apresenta um desafio complexo. Por um lado, a regulamentação pode mitigar riscos como a geração de conteúdo prejudicial e a manipulação de deepfake. Por outro lado, a regulamentação excessiva pode sufocar a inovação e impedir o desenvolvimento de aplicações úteis. Alcançar o ponto de equilíbrio é fundamental para aproveitar o potencial da GenAI e, ao mesmo tempo, proteger-se contra possíveis usos indevidos. Encontrar um meio-termo que incentive a pesquisa e as aplicações responsáveis é essencial para enfrentar esse risco.

Impacto econômico da GenAI⁶

- O impacto da GenAI na produtividade pode gerar trilhões de dólares em valor para a economia global.
- Cerca de 75% do valor gerado pelos casos de uso da GenAI se concentra em quatro áreas: operações com clientes, marketing e vendas, engenharia de software e P&D.
- A GenAI terá um impacto significativo em todos os setores da indústria e poderá gerar um valor equivalente entre US\$ 200 bilhões e US\$ 340 bilhões adicionais por ano.
- A GenAI tem o potencial de mudar a forma como trabalhamos, aumentando as capacidades de cada colaborador com a automação de algumas de suas atividades pessoais.
- A GenAI pode aumentar substancialmente a produtividade da força de trabalho em toda a economia, mas isso exigirá investimentos para oferecer suporte aos colaboradores à medida que mudam de atividade ou de emprego.

⁶McKinsey (14 de junho de 2023). *The economic potential of generative AI: The next productivity frontier*. | Acessado em 10 de setembro de 2024.

IA generativa na NTT DATA

Desde 2018, a NTT DATA realiza pesquisas e desenvolve serviços e produtos que utilizam tecnologias voltadas para o processamento de linguagem natural e modelos inovadores, como no caso do GPT, onde foram publicados diversos white papers sobre LLMs (incluindo GPT-2 e GPT-3)⁷.

Continuaremos nossa jornada de desenvolvimento com modelos fundamentais do idioma japonês, que são a base da nossa inovação contínua. O tsuzumi, LLM japonês de propriedade da NTT, possui recursos de processamento de língua japonesa leves, de excelência mundial, reduzindo significativamente os custos necessários para treinamento, inferência e ajuste.

Estamos convencidos de que a GenAI, especialmente os LLMs, continuará definindo o futuro cenário da IA. Nosso histórico compromisso com o processamento de linguagem natural nos posicionou para alavancar nossa experiência tecnológica e atender à crescente demanda por GenAI.

Para tornar nossas ofertas de IA escaláveis e eficientes e reduzir o tempo de adoção, adotamos uma abordagem de consultoria baseada em ativos (ABC, Asset Based Consulting). Isso nos habilita a oferecer os melhores serviços escaláveis e uma rápida implementação aos nossos clientes, combinando com os nossos melhores talentos em diversas soluções:

Processamento inteligente de documentos e gestão do conhecimento

Atendemos às necessidades dos clientes em áreas como processamento inteligente de documentos (IDP, Intelligent Document Processing), gestão do conhecimento e automação de processos.

IA Conversacional

Nossas soluções aproveitam os recursos de geração de código de modelos fundamentais para automatizar inúmeras tarefas de ponta a ponta, desde o desenvolvimento até a manutenção de software.

Geração e transformação de código

Aproveitando a GenAI, modernizamos com eficiência aplicações legadas, melhorando o desempenho, a escalabilidade e a capacidade de manutenção.

Serviços gerenciados inteligentes

Aproveitamos a GenAI para automatizar os resultados empresariais, revolucionando a gestão do ambiente de TI com o provisionamento simplificado de recursos, a otimização do fluxo de trabalho e a eficiência da infraestrutura.

Nossa missão é reforçar as aplicações práticas da IA baseada em linguagem, aprimorando seus atributos fundamentais com eficiência, confiabilidade, padrões éticos e personalização. Temos o compromisso de eliminar a lacuna entre a IA baseada em linguagem e o mundo real, e buscamos criar sistemas de IA que vão além dos algoritmos computacionais para produzir resultados no mundo real. Ao capacitar a IA para promover mudanças positivas em nosso mundo, podemos concretizar nossa visão de um futuro em que a tecnologia contribua perfeitamente para melhorar nossa vida cotidiana.

Vantagens de nossas soluções

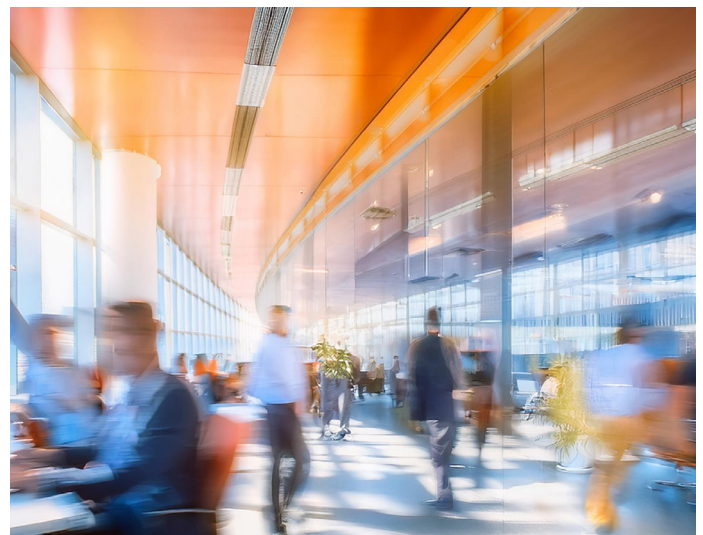
Nossos ativos são reutilizáveis.

Abordamos desafios e pontos fracos.

Soluções alinhadas com o feedback dos clientes.

Percepção do valor investido mais rapidamente.

Oferecemos um roteiro de desenvolvimento bem definido.



⁷NTT DATA Corporation. [Languages, models and business applications](#). | Acessado em 9 de setembro de 2024.

Panorama geral do uso da GenAI

Com o potencial de remodelar os setores e explorar novas possibilidades, a GenAI é uma ferramenta extraordinária. No entanto, não é a única solução em nosso conjunto de ferramentas. Nós consideramos a GenAI como um instrumento versátil que complementa o portfólio de soluções que oferecemos aos clientes para ajudá-los a enfrentar desafios complexos.

Antes de implementar a GenAI, é importante realizar uma análise de custo-benefício. A capacidade da GenAI tem um preço, tanto em termos de capacidade da unidade de processamento gráfico (GPU, Graphics Processing Unit) quanto de consumo de energia elétrica. É importante avaliar se o ROI justifica seu uso, principalmente quando uma tarefa pode ser tratada com mais eficiência por especialistas humanos ou outras técnicas de IA. As considerações de infraestrutura, como a necessidade de recursos computacionais robustos, devem ser ponderadas em relação aos possíveis benefícios.

Portanto, um reforço inteligente seria utilizar LLMs menores para atingir um equilíbrio entre desempenho e praticidade. Essa maneira de melhorar o custo-benefício da eficiência e da transparência torna esses modelos alternativos atraentes aos grandes LLMs para uma ampla gama de aplicações.

Por exemplo, uma empresa de mídia que usa a GenAI para criar conteúdo pode descobrir que consegue produzir um grande volume de artigos com eficiência. No entanto, se não for gerenciado com cautela, o custo dos clusters de

GPU e de energia pode diminuir a possível redução de custos dessa abordagem.

Nesse cenário, a adoção de LLMs menores seria mais econômica, eficiente e fácil de implantar.

Por outro lado, uma instituição financeira que usa a GenAI para análise de dados pode descobrir que a velocidade e a precisão dos insights orientados por IA justificam as despesas computacionais.



“ O tsuzumi da NTT está disponível em duas versões: uma versão leve, com 7 bilhões de parâmetros, e uma versão ultraleve, com 600 milhões de parâmetros. Com capacidades avançadas de processamento do idioma japonês, o tsuzumi é cerca de 300 vezes mais leve do que o GPT-3 da OpenAI e é considerado capaz de superar o GPT-3.5 e os principais LLMs internos de acordo com a Rakuda, um benchmark para a GenAI.⁸

⁸NTT. (25 de abril de 2024). [NTT's Large Language Models 'tsuzumi'](#). | Acessado em 10 de setembro de 2024.

A GenAI, com suas habilidades criativas e insights orientados por dados, pode, sem dúvida, impulsionar a inovação e a eficiência. No entanto, deve ser vista como um complemento valioso ao nosso portfólio de soluções, e não como uma solução única para todos os casos.

Toda tecnologia, inclusive a GenAI, é um meio para alcançar um fim, não o fim em si. A verdadeira dimensão do sucesso está no uso eficaz dessas ferramentas para resolver problemas do mundo real, sem deixar de lado as considerações éticas e a relação custo-benefício de nossas decisões. Ao adotar a mudança de forma ponderada e holística, podemos navegar pelo cenário em evolução da IA com confiança e inovar para um futuro mais brilhante.

Vamos começar

Veja o que a NTT DATA pode fazer por você



Profundo conhecimento do setor e tecnologias líderes de mercado



Recursos personalizados de acordo com seus objetivos



Parcerias para ajudar você a construir e concretizar sua visão

Todas as imagens deste relatório foram geradas com o uso do Azure Open AI, Adobe Firefly e Microsoft Copilot.



Para mais informações, acesse nttdata.com.

Sobre a NTT DATA A NTT DATA, parte do Grupo NTT, é uma prestadora de serviços de TI e parceira de inovação global líder de mercado, sediada em Tóquio. Ajudamos os clientes em todo o mundo a se transformarem por meio de consultoria, soluções industriais, serviços de processos de negócios, modernização de TI e serviços gerenciados. A NTT DATA permite aos clientes, assim como à sociedade, avançar com confiança para o futuro digital. Nosso foco está no compromisso de longo prazo, combinando alcance global com uma visão local para fornecer serviços profissionais de ponta em mais de 50 países em que estamos presentes.



