



**IPSOS VIEWS**

# **O PODER DE TESTE DE PRODUTOS COM DADOS SINTÉTICOS**

**Série Humanizando a IA,  
parte dois**

**Colin Ho, Ph.D  
Dr. Nikolai Reynolds**

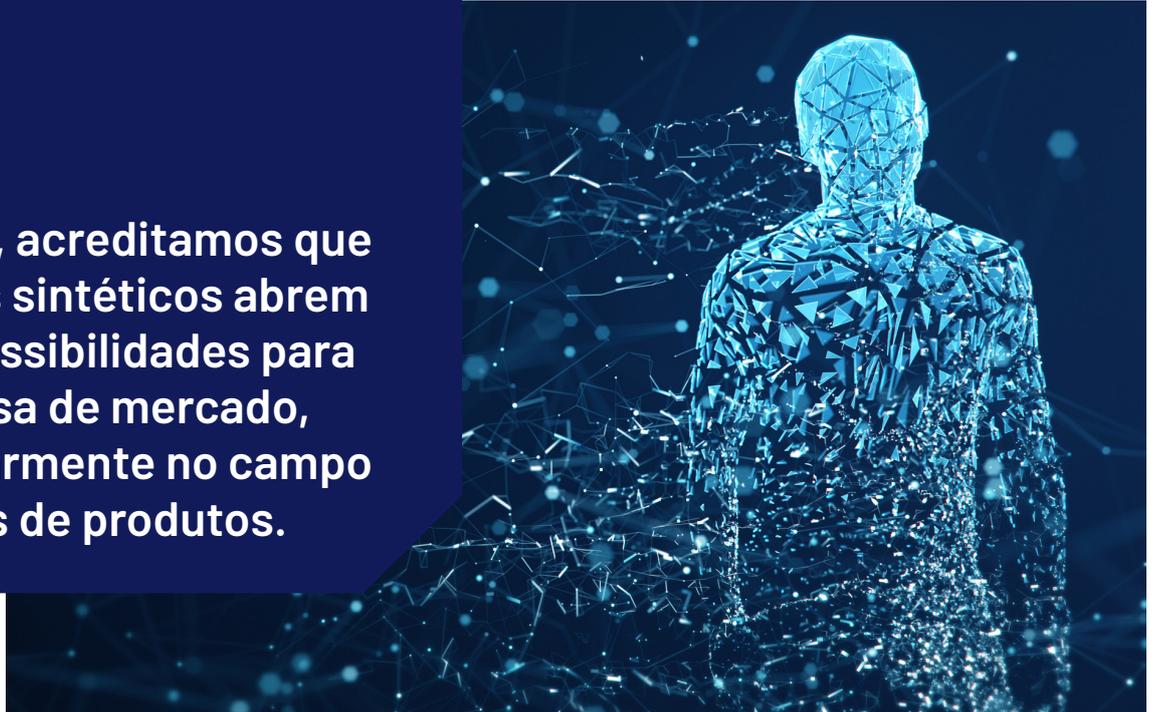


**Na Ipsos, acreditamos na combinação única entre Inteligência Humana (HI) e Inteligência Artificial (IA) para impulsionar a inovação e oferecer insights impactantes e centrados no ser humano aos nossos clientes.**

**Nossa Inteligência Humana vem da nossa expertise em engenharia rápida, ciência de dados e conjuntos de dados exclusivos e de alta qualidade – integrando criatividade, curiosidade, ética e rigor às nossas soluções de IA, impulsionadas pela plataforma Ipsos Facto Gen AI. Isso permite que nossos clientes obtenham insights mais confiáveis, ágeis e alinhados ao contexto humano.**



**Na Ipsos, acreditamos que os dados sintéticos abrem novas possibilidades para a pesquisa de mercado, particularmente no campo de testes de produtos.**



**Os dados sintéticos estão prestes a transformar o mundo.** Seja no desenvolvimento acelerado de medicamentos na área da saúde, na simulação de transações fraudulentas em serviços financeiros ou no aprimoramento de testes para veículos autônomos no setor automotivo, seu valor já é evidente em diversos contextos de negócios.

Na Ipsos, acreditamos que os dados sintéticos abrem novas possibilidades para o mercado de pesquisa, especialmente no campo dos testes de produtos. No entanto, muitas empresas ainda têm dúvidas sobre a qualidade desses dados ou como avaliá-los. Este artigo busca preencher essas lacunas.

Como a maior e líder consultora de testes de produtos do mundo, a Ipsos tem estado na vanguarda da alavancagem

de tecnologia de ponta para acelerar a inovação e o crescimento de empresas em todo o mundo. As páginas a seguir apresentam os insights da Ipsos sobre testes de produtos com dados sintéticos, fornecendo aos leitores:

- Recomendações para gerar e avaliar conjuntos de dados sintéticos de alta qualidade
- Aplicações específicas em testes de produtos para bens de consumo e serviços

Hoje, encontramos diferentes tipos de dados sintéticos no setor de pesquisa de mercado, cada um com seus pontos fortes e limitações. Neste artigo, focamos no aumento de dados, ou seja, no aprimoramento de conjuntos de dados por meio de dados sintéticos.

## Tipos de abordagens para usar dados sintéticos

### Aumento de dados

Aprimorar conjuntos de dados com dados sintéticos para criar amostras mais abrangentes, mantendo a integridade estatística.

### Input de dados & Fusão

Preenchendo dados ausentes usando informações existentes

### Agentes de IA generativa e persona bots

Assistentes digitais personalizados que imitam segmentos de consumidores, oferecendo insights de respostas sintetizadas

### Dados sintéticos completos

Utilizando amostras totalmente artificiais compostas por entrevistados sintéticos.



Começamos com uma visão geral do que é necessário para gerar dados sintéticos de qualidade e como avaliar essa qualidade. Como os efeitos e as nuances dos dados sintéticos só podem ser realmente compreendidos quando considerados em áreas de aplicação específicas, exploramos como os dados sintéticos podem ser aplicados aos testes de

produtos, especificamente. Este documento aborda apenas a geração de dados numéricos sintéticos, o formato mais utilizado por pesquisadores quantitativos. Ele não abrange a aplicação de imagens sintéticas, vídeos, input de dados ou personas sintéticas, todos os quais também fazem parte do amplo conceito de dados sintéticos <sup>1</sup>.

Conforme explicado no artigo da Ipsos Views, [Synthetic Data: From Hype to Reality](#) <sup>1</sup>, os dados sintéticos são dados artificiais gerados a partir de um modelo treinado para imitar as propriedades estatísticas e os padrões de dados do mundo real.





**Se uma IA não tiver sido treinada com dados do mundo real que sejam relevantes para sua empresa, ela não conseguirá gerar dados sintéticos que compartilhem as mesmas propriedades dos dados do mundo real. Simples assim!**



## Gerando e avaliando os dados sintéticos

Usamos dados para tomar melhores decisões de negócios no mundo real, e os dados sintéticos podem ser aplicados de diversas formas para apoiar esse processo. Embora não representem eventos ou pessoas reais, eles precisam refletir as propriedades estatísticas e os padrões dos dados do mundo real. Isso nos leva a duas questões fundamentais:

- 01** O que é necessário para gerar dados sintéticos que simulem os dados reais?
- 02** Como avaliar a semelhança dos dados sintéticos em relação aos dados do mundo real?

Antes que uma IA possa gerar dados sintéticos que reflitam os dados do mundo real, ela precisa ser treinada com esses dados. Como discutido no primeiro artigo da série Humanizando a IA da Ipsos, as IAs são, essencialmente, algoritmos — elas não possuem inteligência própria até que sejam treinadas. É por meio do aprendizado com dados de treinamento que as IAs adquirem a inteligência que associamos a elas. Este é o

ponto central deste artigo: se uma IA não for treinada com dados do mundo real relevantes para a sua empresa, ela não conseguirá gerar dados sintéticos com as mesmas propriedades dos dados reais. Simples assim!

O processo de avaliação também é direto. Os dados numéricos sintéticos devem, no mínimo, espelhar os dados do mundo real em métricas estatísticas comuns, como médias, distribuições, variações e relações entre variáveis (como correlações). Comparar diretamente os dados sintéticos com os dados reais nessas métricas nos dá uma noção de quão bem o conjunto sintético se aproxima dos dados humanos. Quanto mais próximos forem, menor o risco de utilizá-los. No entanto, sempre haverá algum nível de risco, pois os dados sintéticos nunca poderão reproduzir perfeitamente todas as características dos dados reais. Por isso, é essencial usá-los apenas quando estivermos dispostos a aceitar essa margem de incerteza.

## Geração de dados sintéticos usando LLMs

As abordagens para a geração de dados sintéticos podem ser divididas em duas categorias: LLMs (Large Language Models) e não-LLMs, diferenciadas por sua natureza – os primeiros são baseados em texto, enquanto os segundos são voltados para dados numéricos. LLMs prontos para uso ou de domínio público, pré-treinados em grandes volumes de dados, como sites, livros online e postagens em redes sociais, podem gerar dados sintéticos de qualidade para temas que estejam dentro do escopo do seu treinamento.

No entanto, os LLMs prontos para uso têm limitações na produção de dados sintéticos realistas (ver Figura 1)<sup>1,2</sup>. Primeiro, os dados de treinamento têm uma cobertura limitada – muitos temas são específicos demais ou privados para estarem amplamente disponíveis online. Segundo, os LLMs tendem a apresentar viés em favor de países ocidentais de língua inglesa, devido

à predominância desses dados em seus conjuntos de treinamento. Estudos mostram, por exemplo, que os valores culturais gerados por esses modelos estão mais alinhados com a anglosfera e a Europa protestante do que com outras regiões<sup>3</sup>. Terceiro, as informações podem se tornar rapidamente desatualizadas, impactando a precisão dos dados gerados.

Portanto, para gerar dados sintéticos de alta qualidade usando LLMs, é essencial treiná-los com dados atualizados, específicos do país e relevantes para o tema de interesse. Esse processo exige acesso a informações recentes e especializadas, além de conhecimento em estatística e ciência de dados. Também demanda um investimento significativo de tempo e esforço para garantir que os dados sintéticos reflitam com precisão as propriedades e padrões estatísticos do mundo real<sup>4</sup>.

**Figure 1:** Comparação entre um conjunto de dados totalmente humano e um conjunto de dados híbrido, composto por dados sintéticos e humanos



Fonte:  
Ipsos

## Geração de dados sintéticos usando abordagens não LLM

Muito antes de os LLMs roubarem os holofotes, os cientistas de dados usaram algoritmos de Deep Learning (DL)<sup>5</sup> para gerar dados numéricos sintéticos<sup>6</sup>. Os algoritmos DL, incluindo os tipos usados em LLMs, são ferramentas potentes para a geração de dados sintéticos, cada um possuindo vantagens exclusivas.

Os LLMs são particularmente eficazes na geração de dados de texto semelhantes aos humanos. Eles podem fornecer dados textuais detalhados e contextualmente ricos, tornando-os altamente valiosos em aplicações como criação de conteúdo, tradução de idiomas e chatbots<sup>7</sup>.

Os algoritmos DL não-LLM oferecem vantagens significativas na geração de dados numéricos sintéticos. Os

algoritmos de DL são especialmente eficazes na criação de dados que refletem fielmente as propriedades estatísticas dos conjuntos de dados do mundo real. Enquanto LLMs pré-treinados, como o ChatGPT, são voltados para tarefas de linguagem natural, os modelos de DL podem ser treinados especificamente para a síntese de dados numéricos, permitindo uma personalização mais precisa para aplicações específicas do setor ou mercado. Na Ipsos, temos uma trajetória consolidada no uso de técnicas de DL para a geração de dados sintéticos. Na próxima seção, apresentamos os resultados da aplicação desses modelos na criação de dados sintéticos para testes de produtos, analisados individualmente para cada caso.

## Por que testes de produtos com dados sintéticos?

Fora da pesquisa de mercado, muitas aplicações de dados sintéticos estão voltadas para o anonimato, como a anonimização de dados médicos para preservar a confidencialidade. No entanto, no setor de pesquisa de mercado, o principal benefício que muitas empresas buscam é a economia de tempo e custos ao reduzir a necessidade de coletar dados do mundo real.

Com a coleta de dados on-line se tornando cada vez mais rápida e acessível a cada ano, é importante avaliar cuidadosamente se a economia de tempo e dinheiro compensa a perda de precisão associada aos dados sintéticos. Por exemplo, utilizando a Ipsos.Digital, a plataforma de testes ágeis da Ipsos, um pesquisador nos Estados Unidos pode realizar uma pesquisa com 300 entrevistados por cerca de US\$ 2.000 e obter resultados em cerca de 24 horas. Hipoteticamente,

se o pesquisador optasse por usar dados sintéticos para responder às perguntas que os 300 entrevistados reais teriam fornecido, as respostas sintéticas poderiam ser geradas em 12 horas, custando US\$ 1.500, mas com menor precisão em comparação com os dados humanos reais. Nessa situação, pode fazer mais sentido coletar dados reais em vez de gerar dados sintéticos, pois a economia de custo e tempo não parece justificar a perda de precisão. Por que não investir os US\$ 500 adicionais e aguardar 12 horas a mais para obter os dados reais?

Devido à compensação inerente associada aos dados sintéticos, decidimos testar sua aplicação em cenários onde o custo de realização de pesquisas é geralmente elevado. O teste de produto se encaixa perfeitamente nesse contexto, devido aos diversos custos envolvidos:



**Fabricação:** fabricação ou aquisição de produtos e protótipos, ou ainda a modificação dos mesmos para os testes de produto



**Envio e devoluções:** Custos relacionados à entrega dos produtos aos participantes e à devolução ou destruição das embalagens vazias após os testes.



**Amostragem:** Custos de recrutamento, principalmente quando as empresas precisam ser seletivas sobre sua base de usuários

Devido aos custos envolvidos na fabricação, envio e amostragem, qualquer redução no número de participantes em um teste de produto pode resultar em economias significativas. Isso não significa que os dados sintéticos não possam ser aplicados em outras áreas da pesquisa de mercado, mas sim que, nesse caso específico, o risco pode superar os benefícios se a economia de custos for pequena.

## Ainda precisamos de humanos

A experiência com um produto é, por natureza, humana. A IA, por si só, não consegue capturar os cinco sentidos, as emoções, as expectativas ou o impacto do contexto que os consumidores vivenciam ao interagir com um produto. Por isso, nosso objetivo ao aplicar dados sintéticos nos testes de produtos não foi substituir completamente a participação

humana, mas sim complementá-la. O desafio foi determinar o número mínimo de respondentes humanos necessários para testar produtos em conjunto com dados sintéticos, garantindo resultados confiáveis. Para isso, as equipes de inovação da Ipsos conduziram duas linhas de pesquisa:

### Linha de pesquisa 1

Definir o número mínimo de respondentes humanos necessário para que os resultados dos testes de produtos se aproximem dos obtidos em amostras maiores (por exemplo, de 200 a 300 participantes) sem o uso de dados sintéticos.

Na **linha de pesquisa 1**, utilizamos dados do banco de testes de produtos da Ipsos, que inclui 40.000 entrevistados e 185 produtos de bens de consumo embalados (CPG) testados globalmente. O objetivo foi avaliar o número mínimo de respondentes humanos necessário para alcançar uma correlação suficientemente alta com os resultados de testes realizados em amostras maiores. Determinamos que, quando a diferença de desempenho entre

### Linha de pesquisa 2

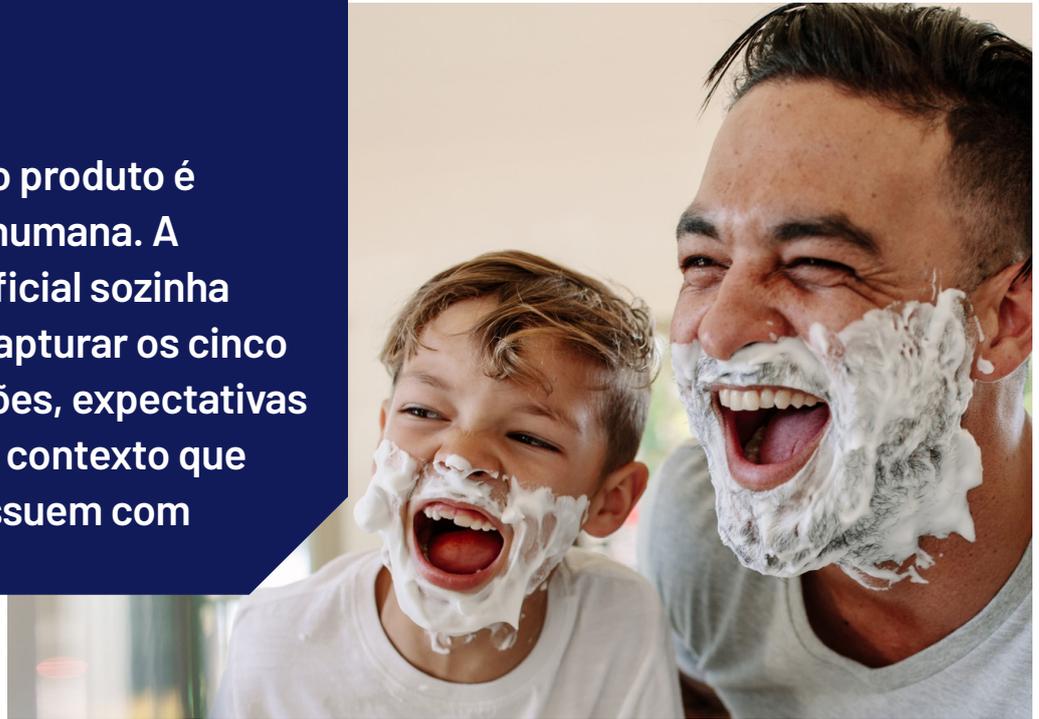
Validar se uma pequena amostra humana, quando complementada com dados sintéticos, gera resultados equivalentes aos de amostras compostas exclusivamente por participantes humanos.

o produto melhor avaliado e o pior avaliado é de pelo menos 8%, uma amostra de 50 respondentes humanos é suficiente para replicar as classificações de desempenho desses produtos, com um coeficiente de correlação  $r = 0,8$ .

O pequeno número de participantes necessários para replicar os resultados de amostras maiores provavelmente se deve ao fato de que a variação



A experiência do produto é inerentemente humana. A inteligência artificial sozinha não consegue capturar os cinco sentidos, emoções, expectativas ou o impacto do contexto que os humanos possuem com produtos.



nos dados dos testes de produtos é influenciada principalmente pelas diferenças na tecnologia do produto (como a quantidade de açúcar utilizada).

**Consequentemente, a variação observada nos testes de produtos tende a ser menor do que em outras áreas de pesquisa, como a avaliação das atitudes dos consumidores ou das percepções de marca.**

Idealmente, as empresas adotariam essas descobertas e realizariam testes de produtos com 50 participantes, especialmente nas fases iniciais, quando o risco é menor.

No entanto, a maioria das empresas não procede dessa maneira por dois motivos:

- 01 Uma amostra de 50 participantes não permite que as empresas obtenham insights detalhados sobre subgrupos. Em muitos casos, é essencial analisar segmentos específicos dentro da população consumidora.
- 02 Uma amostra de 50 participantes resulta em baixo poder estatístico para detectar diferenças significativas. Isso pode ser problemático, pois as empresas costumam basear suas decisões em

testes estatísticos. Na prática, essa limitação pode dificultar a adoção de ações com base nos resultados da pesquisa. Estatisticamente, amostras reduzidas aumentam o risco de erros do Tipo II, ou seja, a possibilidade de não identificar uma diferença real quando ela de fato existe.

Em vez de recrutar 200 pessoas para testar um produto, por exemplo, poderíamos recrutar 50 pessoas para testar o produto, gerar 150 respondentes sintéticos a partir dos dados desses 50 indivíduos, sem replicar ou reamostrar os mesmos, e depois combinar os entrevistados humanos e sintéticos para formar uma amostra mista de 200. Os 50 humanos seriam considerados uma amostra 'semente' para treinar a IA, permitindo que ela gere dados sintéticos que imitam com precisão as respostas humanas aos produtos.

É aqui que entra o fluxo de pesquisa 2, para validar se as pequenas amostras humanas, ampliadas com dados sintéticos, ainda gerariam os mesmos resultados que amostras totalmente humanas. Para garantir a generalização, em nossos testes com dados sintéticos, cobrimos um conjunto diversificado de

países e categorias. Também realizamos experimentos com amostras de sementes maiores (por exemplo, 75, 100), mas, por questões de brevidade, compartilharemos apenas as descobertas dos testes com 50 humanos.

Em seguida, os dados dos 50 humanos foram utilizados para treinar um algoritmo de aprendizado profundo (DL) para gerar dados sintéticos. Não utilizamos um modelo de linguagem grande (LLM) pronto para uso, pois os dados

públicos nos quais os LLMs foram pré-treinados não incluem as experiências multissensoriais das pessoas com produtos da categoria em questão (como protótipos ou novas formulações) e não fornecem dados numéricos robustos no nível do entrevistado. Além disso, não consideramos a ponderação dos dados como uma alternativa ao DL, pois a ponderação não ajuda a criar tamanhos de amostra adicionais para subgrupos e, geralmente, não é aceita em testes de produtos.



**[A Ipsos] validou os resultados comparando as descobertas de um conjunto de dados totalmente humano com as de um conjunto de dados aumentado com dados sintéticos.**



## Em geral, os dados sintéticos funcionam

Em nossos experimentos, validamos os resultados comparando os dados de um conjunto totalmente humano com os de um conjunto aumentado com dados sintéticos (ver Figura 1)<sup>8</sup>. Como lembrete, em nossa abordagem, não estamos apenas replicando ou copiando e colando dados de respondentes existentes.

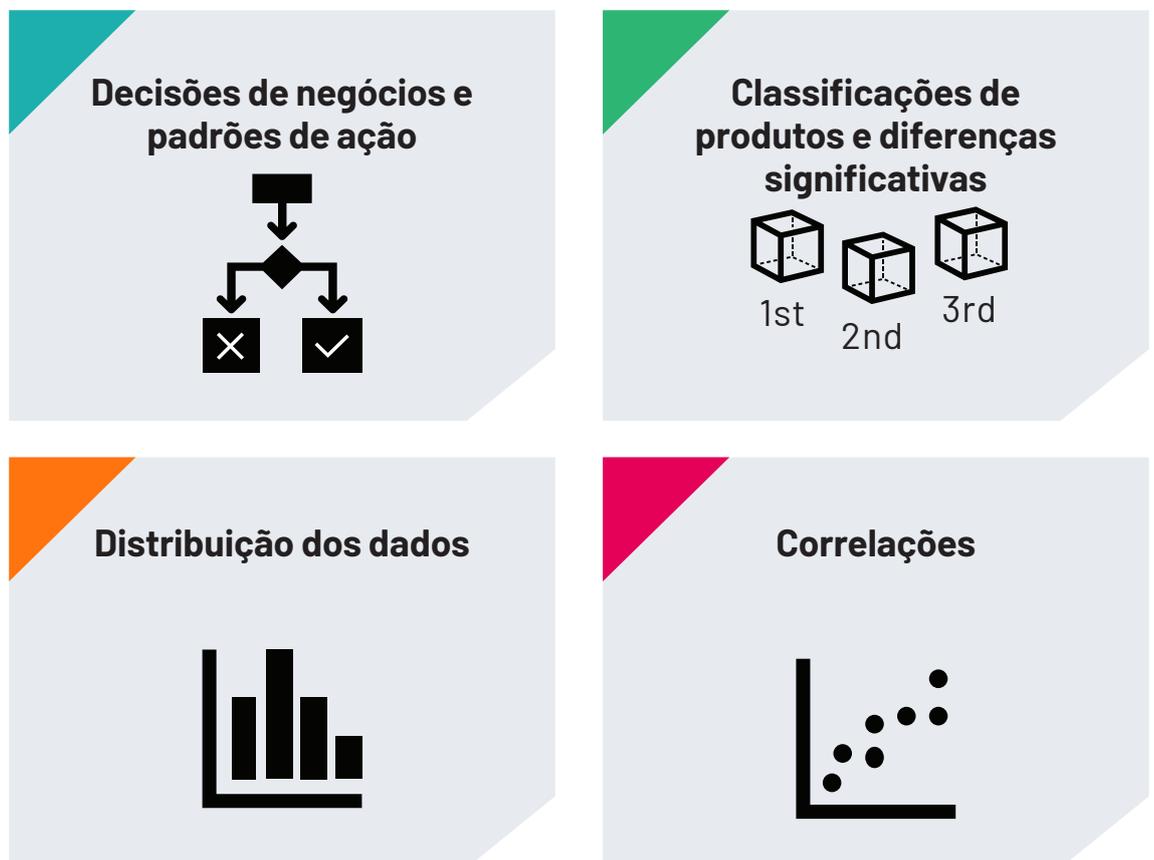
- Distribuição de dados (por exemplo, a distribuição das respostas das pessoas nas opções de resposta em perguntas individuais)
- As relações entre as variáveis nos dados (ou seja, correlações entre o gosto geral e os atributos do produto)

Em geral, descobrimos que os dois conjuntos de dados eram notavelmente semelhantes, em termos de:

- O desempenho relativo dos produtos (por exemplo, classificações, significância estatística)

Mais importante ainda, os dois conjuntos de dados apresentaram diferenças nas variações, mas levaram à mesma decisão de negócios em todos os conjuntos de dados que testamos (consulte a Figura 2).

**Figura 2:** Critérios utilizados para determinar a precisão da parte humana e sintética na replicação das descobertas do conjunto de dados totalmente humano



Fonte:  
Ipsos

Um dos principais benefícios do teste de produtos com dados sintéticos é a capacidade de ampliar os dados para populações de difícil acesso. Uma vez ampliados, as diferenças que antes não eram estatisticamente significativas podem se tornar significativas devido ao aumento no tamanho da amostra. Em um de nossos testes, por exemplo, geramos respostas sintéticas para aumentar o número de pessoas que usaram uma determinada marca de produto. Na nossa amostra totalmente humana, tivemos cerca de 100 usuários da marca

por produto. No conjunto de dados totalmente humanos, observamos algumas diferenças entre os três produtos testados, mas essas diferenças não atingiram significância estatística. No entanto, ao aumentar a amostra com 100 usuários de marcas sintéticas por produto, as diferenças entre os produtos se tornaram estatisticamente significativas devido ao aumento no tamanho da amostra (ver Figura 3).

**Figura 3:** Usuários humanos da marca aumentados com usuários sintéticos da marca





Quando buscamos feedback sobre produtos de um grupo-alvo específico, devemos definir cotas de recrutamento que estejam alinhadas aos objetivos comerciais.



## É preciso ter cautela

Apresentamos um cenário promissor para o uso de dados sintéticos. No entanto, como mencionado anteriormente, os dados sintéticos possuem uma limitação inerente, pois nunca podem corresponder completamente à precisão dos dados do mundo real. A capacidade de um conjunto de dados parcialmente humano e parcialmente sintético de replicar as descobertas de um conjunto totalmente humano depende de:

- **A representatividade dos dados humanos usados para treinar a**



**IA.** Em nossos primeiros testes, quando não controlamos a seleção

dos 50 participantes humanos para alinhá-los à estrutura da amostra original de 200 pessoas e geramos dados sintéticos a partir desses 50, notamos diferenças nos resultados. Sem esse controle, em 20 a 30% das amostras, os dados sintéticos gerados não refletiram com precisão o desempenho do produto observado nos 200 humanos originais. Isso demonstra que os dados sintéticos podem falhar caso a amostra de sementes humanas não seja

representativa. No entanto, quando estruturamos os 50 participantes humanos para corresponder à composição da amostra original de 200 pessoas, a decisão de negócios foi a mesma ao comparar os resultados da amostra combinada de 50 humanos e 150 sintéticos com os da amostra original de 200 humanos. Em resumo, ao buscar feedback sobre produtos de um grupo-alvo específico, é essencial definir cotas de recrutamento alinhadas aos objetivos comerciais. Assim como fizemos no passado com amostras maiores, esse cuidado garante que estamos segmentando corretamente o público-alvo.

- **Quão diferenciados os produtos são em primeiro lugar.** Quando há diferenças significativas entre os produtos no conjunto de dados totalmente humano, a IA consegue detectá-las e replicá-las facilmente nos dados sintéticos. Já quando as diferenças não são estatisticamente significativas, a IA, assim como a amostra totalmente humana, também não consegue identificá-las (veja a Figura 4). Nos diversos testes realizados em diferentes países e categorias, identificamos uma vantagem adicional no uso de amostras ampliadas em cenários



onde há pouca discriminação entre os produtos. Ao dividir a amostra original em subconjuntos menores, como usuários atuais da marca e não usuários, descobrimos que o uso de dados sintéticos gerados por IA para ampliar esses subgrupos melhorou a detecção de diferenças no desempenho do produto, com base nas preferências do público-alvo. Isso ficou evidente no nível dos subgrupos, onde os dados sintéticos ajudaram a aumentar o poder estatístico e revelar preferências que poderiam ter passado despercebidas na amostra original.

**Figure 4:** A precisão dos dados sintéticos depende das diferenças reais do produto



Como consideração final sobre a validação, é importante destacar que as descobertas apresentadas aqui são específicas para testes de produtos. Não queremos que o leitor tenha a impressão de que todas as pesquisas de mercado podem ser conduzidas com apenas 50 participantes humanos. As condições que permitiram o uso desse número reduzido em testes de produtos podem não se aplicar a outras áreas de pesquisa. Como líder mundial em consultoria de testes de

produtos, a Ipsos aproveitou aprendizados de seu extenso banco de dados nessa área. Além disso, os testes de produtos possuem características únicas que contribuem para a qualidade dos dados sintéticos, como atributos sensoriais que influenciam a percepção do consumidor. No entanto, os dados sintéticos devem sempre ser validados individualmente para cada aplicação específica, como segmentação, garantindo sua precisão e relevância.



Os dados são frequentemente vistos como a força vital das empresas, impulsionando decisões mais inteligentes que contribuem para seu crescimento e sucesso.



## A promessa dos dados sintéticos: do hype à realidade

Os dados são geralmente considerados a força vital das empresas, permitindo decisões mais inteligentes que impulsionam seu crescimento e sucesso. A possibilidade de gerar dados sintéticos sob demanda e em grande escala, portanto, é extremamente atraente. No entanto, as opiniões do público sobre esse tema são bastante polarizadas. Enquanto algumas empresas que oferecem serviços de dados sintéticos afirmam que eles são a única solução necessária, eliminando a necessidade de participação humana, pesquisadores mais cautelosos preferem adotar uma abordagem de espera e hesitação antes de incorporá-los. No centro desse debate está a tendência de enxergar o mundo de forma binária: bom ou ruim? Dados sintéticos ou do mundo real? Demos um primeiro passo para esclarecer essa questão, demonstrando que os dados sintéticos não são diferentes de outras ferramentas de pesquisa. Eles possuem pontos fortes e fracos e são mais adequados para determinadas situações.

Em nossos testes, demonstramos que a IA pode gerar dados sintéticos que imitam

dados do mundo real, mas, para isso, ela precisa de dados humanos de qualidade para treinamento. Portanto, a questão não é escolher entre dados sintéticos ou reais – ambos são necessários. A precisão dos dados sintéticos não é simplesmente boa ou ruim; na verdade, ela depende. Se as diferenças entre os produtos forem sutis entre os participantes humanos, será necessário analisar os subgrupos. Além disso, se os dados humanos usados para treinar a IA não forem representativos do público-alvo ou relevantes para o negócio, a qualidade dos dados sintéticos será comprometida. Para utilizar dados sintéticos de forma eficaz, é fundamental reconhecer que eles podem não funcionar em determinadas condições. **Como pesquisadores, nossa responsabilidade é garantir que seu uso ocorra apenas quando for apropriado, ou seja, em cenários que maximizem sua eficácia.**

A ampliação de dados sintéticos traz diversas vantagens em relação ao uso de amostras menores, como a possibilidade de realizar análises de subgrupos, manter o poder estatístico e conduzir investigações mais complexas.

No geral, não acreditamos que os dados sintéticos devam substituir completamente os humanos – pelo menos não nos testes de produtos. No filme de 1997 “Gênio Indomável”, o falecido ator, Robin Williams, interpretou um professor que orienta um jovem gênio, interpretado por Matt Damon. O prodígio detém uma vasta quantidade de conhecimento, devido à sua capacidade sobre-humana de absorver informações de livros. Em uma das cenas do filme, o professor aconselha o prodígio sobre a distinção entre o conhecimento do livro e a experiência do mundo real. O professor diz: “Mas aposto que você não pode me dizer como é o cheiro na Capela Sistina. Você nunca ficou lá e olhou para aquele lindo teto”. O verdadeiro conhecimento vem de viver a vida, não de livros, fotos, vídeos

ou quaisquer outras representações do mundo real.

Como o prodígio do filme, uma IA pode ser alimentada com todo o conhecimento do mundo que existe, mas uma IA nunca será capaz de experimentar o mundo como um ser humano. Há algo único e bonito em ser humano e ser capaz de sentir o calor do sol em nosso rosto, apreciar a melodia ou a batida da música ou a capacidade de contemplar um belo pôr do sol – experiências que a IA nunca será capaz de replicar totalmente, não importa o quão avançada ela se torne. A forma como os humanos reagem aos produtos, ou à vida em geral, não é capturada apenas no cérebro como conhecimento factual ou semântico, nossos corpos e experiências sensoriais também desempenham um papel significativo.



**O verdadeiro conhecimento vem de viver a vida, não de livros, fotos, vídeos ou quaisquer outras representações do mundo real.**

## Principais conclusões

### 01

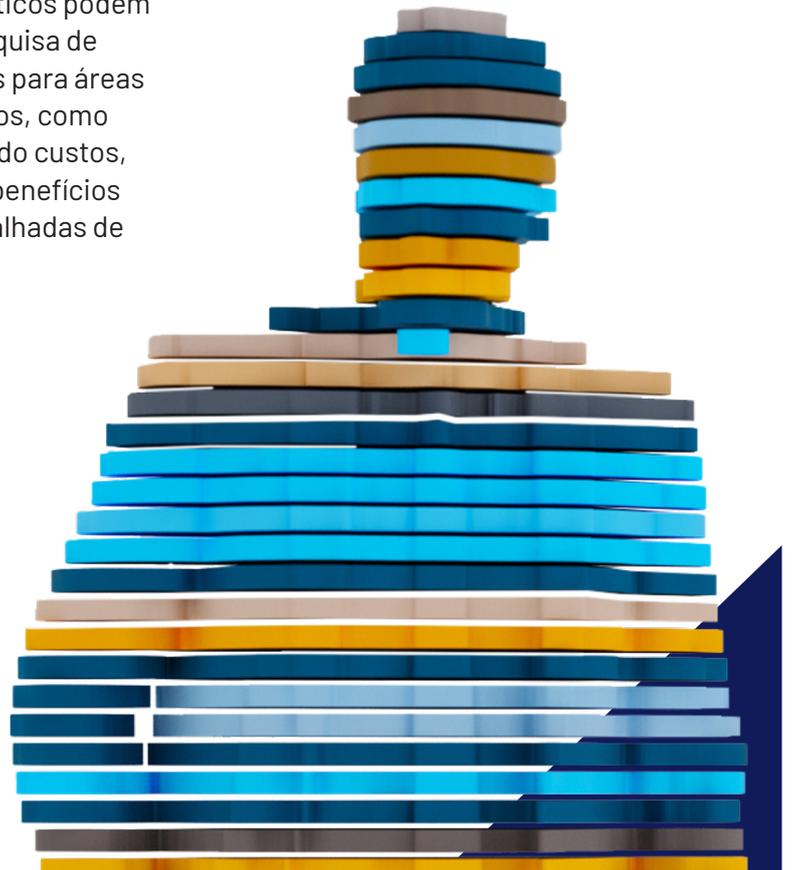
**Os dados sintéticos nunca serão humanos.** A IA, por si só, nunca poderá ecoar nossas experiências com produtos, que combinam os cinco sentidos, as emoções, as expectativas e o contexto. Portanto, nosso objetivo é aumentar a contribuição humana com dados sintéticos, não substituí-la.

### 03

**Quando preciso, pode impulsionar teste do produtos.** Os dados sintéticos podem aumentar a agilidade da pesquisa de mercado, tornando-os ideais para áreas com uso intensivo de recursos, como testes de produtos – reduzindo custos, economizando tempo, com benefícios adicionais para análises detalhadas de subgrupos.

### 02

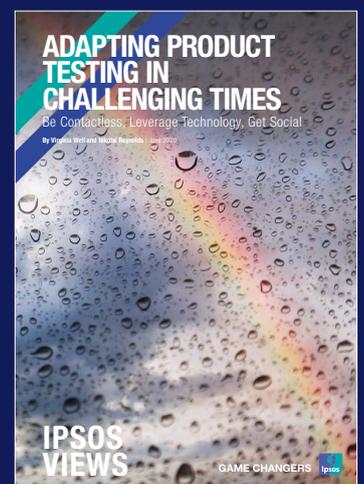
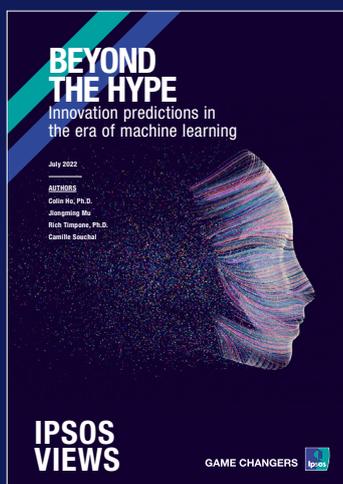
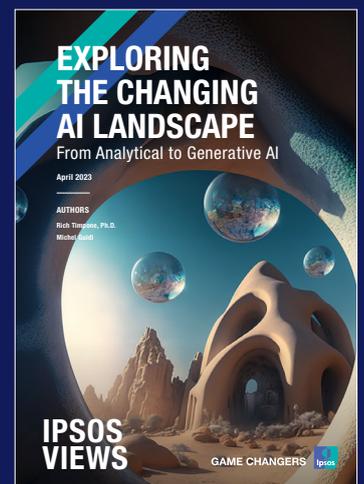
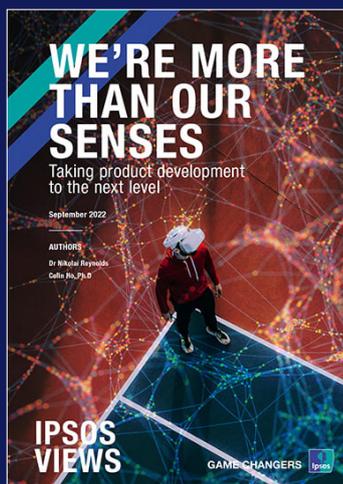
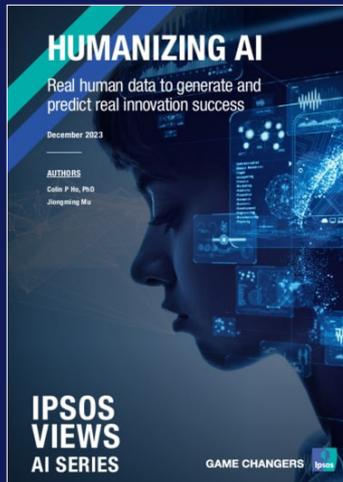
A precisão depende dos dados de treinamento. O valor dos dados sintéticos não é binário (bom ou ruim); A precisão dos dados sintéticos depende de muitos fatores, incluindo as diferenças nos dados que estamos tentando replicar e a representatividade dos dados do mundo real que estamos treinando uma IA para aprender. O uso de dados sintéticos deve ser estratégico, considerando os riscos e benefícios associados.



## Notas finais

- 1 Illic, Maya, Bangia, Ajay, Legg Jim (2024). [Conversations with AI Part V. Is there depth and empathy with AI twins?](#) Ipsos Views
- 2 Moore Chris, Stronge Cameron, Bhudiya, Manjula (2024). Judgment Day: The Machines Have Arrived – But how good are they at answering choice experiments? Sawtooth Conference.
- 3 Yan Tao, Olga Viberg, Ryan S. Baker and René F. Kizilcec (2024). Cultural bias and cultural alignment of large language models. PNAS Nexus, Vol. 3, No. 9
- 4 Lisa P. Argyle, Ethan C. Busby, Nancy Fulda, Joshua Gubler, Christopher Rytting, and David Wingate (2022). Out of One, Many: Using Language Models to Simulate Human Samples. arXiv.
- 5 O Deep Learning baseado em IA é uma maneira de os computadores aprenderem analisando grandes quantidades de dados e encontrando padrões, assim como os humanos aprendem com a experiência. Ele usa redes neurais inspiradas no cérebro humano para reconhecer informações e tomar decisões sem ser explicitamente programado..
- 6 Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, Yoshua Bengio (2014). Generative Adversarial Networks. arXiv.
- 7 Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., & Sutskever, I. (2019). Language Models are Unsupervised Multitask Learners.
- 8 Também comparamos os resultados dos dados totalmente sintéticos com os dados totalmente humanos; Os resultados são semelhantes aos relatados neste artigo

# Leitura complementar



JANUARY 2025

# O PODER DE TESTES DE PRODUTOS COM DADOS SINTÉTICOS

Série Humanizando a IA, parte dois

## AUTORES

### **Colin Ho, Ph.D**

Chief Research Officer,  
Innovation and Market Strategy  
& Understanding, Ipsos

### **Dr. Nikolai Reynolds**

Global Head of Product Testing,  
Ipsos

The **IPSOS VIEWS** white  
papers são produzidos pela  
**Ipsos Knowledge Centre.**

[www.ipsos.com](http://www.ipsos.com)

@ipsos

