AI Research Brief

#1 • 5 de novembro de 2026

Nubank reduz de três semanas para menos de cinco dias o ciclo de atualização de agentes de IA ao testá-los primeiro com clientes sintéticos

Pesquisa mostra que clientes sintéticos podem ajudar o Nubank a testar mais cenários, tomar decisões mais rapidamente e melhorar a experiência dos clientes antes de ampliar o uso de agentes de IA, incluindo aqueles que utilizam modelos de pesos abertos como o Qwen.

ARTIGO · Screen Before You Serve: Simulation for Production Customer Experience AI Agents at 140M Scale
arXiv:2609.30137

21,2 → 4,4
dias por nova versão do agente, sem vs. com simulação

+4,9 p.p.
taxa de autoatendimento no teste A/B real

+36,7 pts
transactional NPS (tNPS) no teste A/B real

São Paulo, 5 de novembro de 2026 – O time de AI Research do Nubank constatou que testar agentes de IA com clientes sintéticos em simulações de larga escala pode acelerar o desenvolvimento e ajudar as equipes a escolher versões melhores antes de disponibilizá-las para clientes reais. No estudo, cada nova versão do agente levou, em média, 4,4 dias para ser desenvolvida com o uso de simulação, ante 21,2 dias em um fluxo de trabalho anterior sem esse recurso. Em um teste A/B real, o agente desenvolvido com apoio da simulação registrou uma taxa de autoatendimento 4,9 pontos percentuais maior e um transactional Net Promoter Score (tNPS), índice que mede a recomendação do cliente após uma interação, 36,7 pontos maior do que o agente usado para comparação.

Descrita em um novo artigo de pesquisa, a abordagem usa um ambiente de teste controlado no qual clientes sintéticos, com diferentes perfis e necessidades, conversam em várias etapas com agentes de IA. Assim, as mudanças podem ser avaliadas antes de chegarem aos clientes reais.

O estudo acompanhou dois agentes de IA de atendimento ao cliente em diferentes etapas do ciclo do cartão de crédito: o Card Delivery, que trata do status de entrega e da emissão de segunda via, e o Card Management, que permite atualizar o endereço e solicitar uma nova via do cartão. A simulação orientou as mudanças no Card Management, enquanto o Card Delivery foi usado como agente de comparação no teste A/B real.

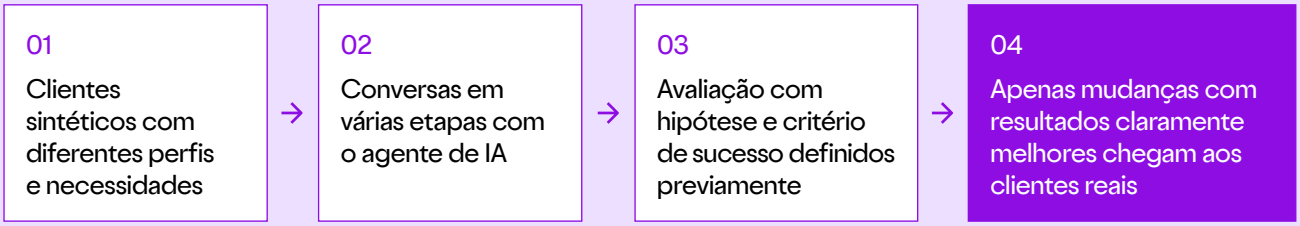
> “Cada mudança em um agente de IA é uma aposta na experiência do cliente. A simulação nos permite testar essas apostas primeiro com clientes sintéticos, para que apenas as mudanças com resultados claramente melhores cheguem às pessoas”

Rohan Ramanath
Gerente-geral de IA do Nubank

NESTA EDIÇÃO

01 Aprendizado mais rápido, impacto para os clientes p. 2
02 Uma forma mais rápida e abrangente de avaliar agentes de IA p. 3
03 O que a pesquisa possibilita p. 4

Fig. 1 · Como a simulação filtra mudanças antes de chegarem aos clientes



01

Aprendizado mais rápido, impacto para os clientes

A simulação ajudou a equipe a decidir quais mudanças no agente Card Management valia a pena levar aos clientes reais. Na avaliação offline, uma versão do agente reduziu os encaminhamentos desnecessários para o atendimento humano de 22,4% para 16,0% — uma queda de 6,4 pontos percentuais em comparação com a versão anterior.

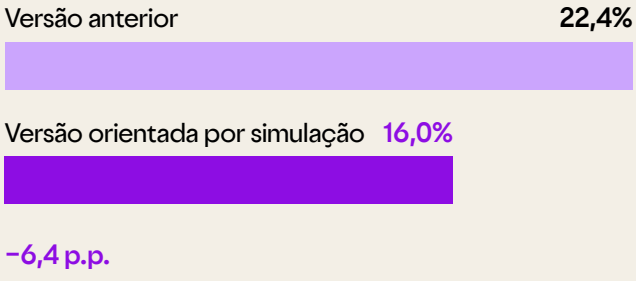
O teste A/B ao vivo realizado em seguida mediu o impacto com clientes reais. Em comparação com o Card Delivery, o Card Management apresentou uma taxa de autoatendimento 4,9 pontos percentuais maior e um tNPS 36,7 pontos maior. Como o Card Management atende a um conjunto mais amplo de tarefas do que o Card Delivery, esses ganhos refletem tanto o novo agente quanto o processo de desenvolvimento orientado por simulação. Após o teste, o Card Management foi disponibilizado para a base de clientes do Nubank no Brasil.

A taxa de autoatendimento representa a parcela das sessões concluídas com o agente sem que o cliente peça atendimento humano. Isoladamente, ela não comprova que todos os problemas foram resolvidos com sucesso. Por isso, a análise também considera métricas de satisfação e avaliações específicas de qualidade.

+4,9 p.p.
taxa de autoatendimento maior vs. agente de comparação

+36,7 pts
tNPS maior vs. agente de comparação

Fig. 2 · Encaminhamentos desnecessários para o atendimento humano, avaliação offline



GLOSSÁRIO

Taxa de autoatendimento: parcela das sessões concluídas com o agente sem que o cliente peça atendimento humano.

tNPS: transactional Net Promoter Score, índice que mede a recomendação do cliente após uma interação.



02

Uma forma mais rápida e abrangente de avaliar agentes de IA

29 configurações de modelos avaliadas

16 mil+ conversas simuladas (mais de 16 mil)

+8,8 p.p. na taxa de autoatendimento com a configuração aprimorada do Qwen

−25% na latência de resposta no p95

DETECTADO ANTES DO LANÇAMENTO

A simulação revelou um ponto fraco, corrigido com otimização de prompts antes do teste com clientes reais.

Sem uma camada de simulação, o teste de novas versões de um agente pode depender muito de cenários escritos manualmente, de amostras limitadas e de experimentos com clientes reais. Simulações em larga escala permitem comparar mais combinações de instruções, modelos e arquiteturas de agentes em um ambiente controlado, antes de decidir quais mudanças devem avançar.

A mesma abordagem ajudou o Nubank a avaliar modelos de pesos abertos para atendimento ao cliente. A equipe avaliou 29 configurações de modelos em mais de 16 mil conversas simuladas e selecionou o Qwen, com raciocínio habilitado, para aprimoramento. A simulação revelou um ponto fraco que a equipe tratou por meio da otimização de prompts antes do teste com clientes reais.

Em um teste A/B separado com clientes reais, comparando o modelo então utilizado pelo Card Management com a configuração aprimorada do Qwen, a configuração aprimorada aumentou a taxa de autoatendimento em 8,8 pontos percentuais e reduziu a latência de resposta no percentil 95 (p95) em 25%, sem alteração estatisticamente significativa no tNPS.



03

O que a pesquisa possibilita

A contribuição mais ampla do artigo é mostrar que avaliar agentes de IA exige mais do que verificar se eles produzem a resposta certa. O estudo também indica que o papel de um simulador é classificar corretamente as alternativas — e não reproduzir perfeitamente o comportamento humano — e que os próprios simuladores precisam ser validados antes que as equipes dependam deles. Essas ideias abrem novas frentes de pesquisa: testes padronizados para avaliar a confiabilidade dos simuladores; ciclos de melhoria nos quais agentes são revisados, avaliados e aprimorados com menos esforço manual; e avaliações de agentes mais proativos, capazes de orientar clientes durante a realização de tarefas.

A abordagem atual termina nos sistemas chamados pelo agente. Por isso, ampliar a simulação para incluir o comportamento e os tempos de resposta dos sistemas de backend é um próximo passo natural. Como cada mudança está vinculada a uma hipótese e a um critério de sucesso definidos previamente, o fluxo de trabalho também registra com clareza por que cada alteração foi aprovada, algo valioso em setores regulados.

Para o Nubank, essa abordagem conecta a pesquisa aplicada em IA a um desafio concreto de negócio: como desenvolver experiências mais úteis e confiáveis à medida que a IA conversacional passa a fazer parte de mais interações.

IDEIA-CHAVE DO ARTIGO

O papel de um simulador é classificar corretamente as alternativas, e não reproduzir perfeitamente o comportamento humano.

PRÓXIMAS FRENTES DE PESQUISA

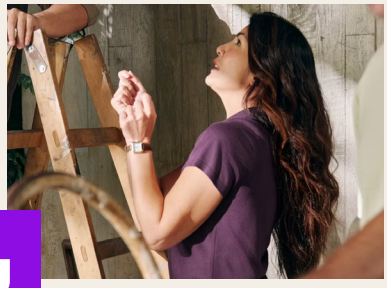
01
Testes padronizados de confiabilidade dos simuladores

02
Ciclos de melhoria com menos esforço manual

03
Avaliação de agentes mais proativos

04
Simulação do comportamento e dos tempos de resposta do backend







Relações com Investidores

Guilherme Souto
investors@nubank.com.br

Relações com a Imprensa

Simone Iwasso
press@nubank.com.br

Nu Holdings LTD.

NYSE: NU
https://nu.com/en/newsroom

LEIA O ARTIGO COMPLETO

Screen Before You Serve: Simulation for Production Customer Experience AI Agents at 140M Scale

Nubank AI Research
arxiv.org/abs/2609.30137

Sobre o Nu

O Nu é uma das maiores instituições financeiras digitais do mundo e o maior banco digital da América Latina. Com a missão de combater a complexidade e empoderar as pessoas, atende a mais de 140 milhões de clientes no Brasil, no México e na Colômbia. Recentemente, o Nu iniciou operações nos Estados Unidos e lançou a Nu Global, sua conta multimoeda que permite a movimentação de dinheiro entre mais de 35 países.

O Nu vem liderando uma transformação no setor por meio de produtos inovadores e serviços focados nas pessoas. Tecnologia proprietária e dados em escala impulsionam a plataforma digital da companhia, desenvolvida para promover acesso, evolução e transparência financeira. Seu modelo de negócios se fortalece graças à satisfação dos clientes e a custos mais baixos, alimentando um ciclo contínuo de crescimento e rentabilidade.