

Future of AI: Perspectivas para Startups

2025





Índice

02 [Prefácio](#)

04 [Previsões de IA](#)

07 [Conselhos para fundadores](#)

11 [O que vem por aí em IA: perspectivas dos líderes do mercado](#)

↔ **Alison Wagonfeld**

CMO, Google Cloud

↔ **Izabelle Macedo**

Marketing Director, Google Cloud LATAM

↔ **Amin Vahdat**

VP/GM ML, Systems, e Cloud AI, Google Cloud

↔ **Eric Acher**

Co-Founder e Managing Partner, Monashees

↔ **Apoorv Agrawal**

Sócio, Altimeter Capital

↔ **Laura Constantini**

Founder, Nido e VC investor

↔ **Jill Greenberg Chase**

Sócia investidora, CapitalG

↔ **Mayada Gonimah**

CTO e Co-Founder, Thread AI

↔ **Rodrigo Scotti**

Founder e CEO, Nama

↔ **Yoav Shoham**

Professor emérito de ciência da computação,
Universidade de Stanford. Co-fundador, AI21 Labs

40 [Crie o futuro com o Google Cloud](#)



Prefácio



A IA está transformando todas as organizações ao redor do mundo e representa uma oportunidade sem precedentes para resolver problemas complexos, impulsionar o crescimento, aumentar a eficiência e abrir novas oportunidades de negócios. Isso vale sobretudo para as startups, que estão se movendo muito rapidamente para explorar novas oportunidades de mercado com IA. O Google Cloud está no centro da inovação em IA e temos orgulho da nossa liderança tecnológica, que continua ultrapassando os limites das possibilidades para os nossos clientes, incluindo mais de 60% de todas as startups de IA generativa financiadas globalmente. Estamos muito animados com as parcerias que firmamos com startups, com a comunidade de capital de risco e com líderes do setor para garantir que fundadores e suas equipes tenham acesso à tecnologia que os ajudará a redefinir o futuro.

Thomas Kurian
CEO, Google Cloud





O Google está construindo todos os componentes do stack de tecnologia de IA, desde chips personalizados e datacenters até modelos de ponta. Como resultado, nossos novos modelos do Gemini 2.0 são mais capazes, rápidos e eficientes do que as versões anteriores. São modelos nativamente multimodais – capazes de processar texto, imagens, áudio e vídeo. Eles também podem gerar imagens e fazer a conversão de texto em voz, criando áudios personalizáveis. Com janelas de contexto de até 2 milhões de tokens, o Gemini pode alimentar aplicativos avançados que exigem compreensão e memória profundas.

Além disso, o modelo de pensamento crítico do Gemini aprimora suas habilidades de raciocínio para resolver problemas complexos, o que é especialmente útil em matemática e ciências. O Gemini também pode usar nativamente ferramentas como a Pesquisa do Google para acessar informações em tempo real, e o Project Mariner da DeepMind demonstrou que um agente baseado no Gemini pode executar tarefas usando um navegador da web. As experiências conversacionais agora podem ser desenvolvidas com a API Multimodal Live do Gemini, que aceita entrada de streaming de áudio e vídeo. A combinação dessas capacidades possibilita uma nova classe de experiências baseadas em agentes, e estamos animados para ver o que as startups vão construir com o Gemini em 2025.

David Thacker

VP, Product, Google DeepMind





01

02

03

Previsões de IA



**Alison Wagonfeld**

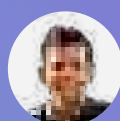
CMO, Google Cloud

O marketing transcenderá os tradicionais “quatro Ps” (Produto, Preço, Praça, Promoção) e haverá um 5º P para “Personalização”.

**Amin Vahdat**

VP/GM ML, Systems, e Cloud AI, Google Cloud

A sincronização precisa e os requisitos massivos de computação vão levar a infraestrutura a níveis de densidade e capacidade computacional nunca vistos antes.

**Apoorv Agrawal**

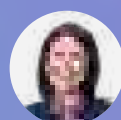
Sócio, Altimeter Capital

Ao combinar voz, visão e linguagem natural, a IA multimodal reduzirá a necessidade de dispositivos como computadores e celulares, tornando a interação com o mundo digital mais fluida.

**Eric Archer**

Co-Founder e Managing Partner, Monashees

Atualmente, estamos vivendo em duas realidades diferentes: uma delas está focada na tech tradicional, e a outra só quer saber do over appreciation da inteligência artificial.

**Izabelle Macedo**

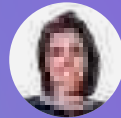
Marketing Director, Google Cloud LATAM

A IA vai continuar se expandindo e abrindo possibilidades de crescimento para os negócios. No entanto, também há espaço para reavaliar e melhorar as experiências que não deram certo antes.

**Jill Greenberg Chase**

Sócia investidora, CapitalG

Os modelos fundacionais permanecerão relativamente estáticos em capacidade nos próximos 18 meses, criando uma grande oportunidade para as startups desenvolverem soluções de IA especializadas que forneçam um claro retorno sobre o investimento.

**Laura Constantini**

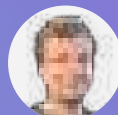
Founder, Nido e VC investor

A cada nova (r)evolução tecnológica e mudança de plataforma, a dinâmica de investimento/crescimento das startups muda, e é preciso um capital mais inovador e paciente. Uma das perguntas que precisamos fazer é: as novas empresas estão preparadas para enfrentar esse cenário?

**Mayada Gonimah**

CTO e Co-Founder, Thread AI

Estamos vendo uma explosão repentina de modelos, soluções e ferramentas que estão sendo lançados no mercado sem provar que realmente fazem a diferença, seja no ROI ou na mudança fundamental do ecossistema de desenvolvedores. Muitas dessas bibliotecas de dados vão morrer ou receber cada vez menos financiamento.

**Rodrigo Scotti**

Founder e CEO, Nama

É relevante considerar que organizações com amplo acesso a dados e recursos para desenvolver tecnologias de IA têm responsabilidades proporcionais às suas capacidades.

**Yoav Shoham**

Professor emérito de ciência da computação, Universidade de Stanford. Co-fundador, AI21 Labs

Os primeiros dias da IA – quando bastava fornecer um prompt e esperar por um bom resultado – acabaram. Para atender às demandas críticas das empresas, precisamos de “sistemas de IA” robustos que orchestrem vários modelos e ferramentas para obter resultados confiáveis.





01

02

03

Conselhos para fundadores



Crie campanhas de marketing personalizadas e direcionadas para evitar gastos desnecessários com campanhas genéricas, que atingem um público amplo demais.

Alison Wagonfeld
CMO, Google Cloud

Alinhe o preço ao valor entregue.

Não use apenas o modelo por usuário, mas considere preços baseados no uso ou no valor. Seu preço deve refletir o valor que seu produto oferece aos usuários.

Apoorv Agrawal
Sócio, Altimeter Capital

Cada ciclo é único. Sob este ponto de vista, a jornada do empreendedor pode ser transformada quando se cria uma empresa winner em seu meio. Portanto, se aquele negócio é pioneiro por algum motivo ou é destaque, ele será um modelo para os demais.

Eric Archer
Co-Founder e Managing Partner, Monashees

Com o tempo, a computação será muito mais rápida, barata e confiável.

Digamos que você tenha uma ideia incrível, e tudo o que precisa é reduzir os custos de computação em 10% ou até 90% para que ela se torne rentável. Isso agora está ao seu alcance.

Amin Vahdat
VP/GM ML, Systems, e Cloud AI, Google Cloud



A IA pode se tornar a porta de entrada para o desenvolvimento de modelos de negócios disruptivos:

Não como um modismo a ser incorporado a qualquer custo, mas como uma ferramenta poderosa para aprimorar processos, ganhar eficiência e impulsionar a inovação. Por isso, acredito que as startups que souberem explorar o seu potencial com uma abordagem estratégica e objetivos claros estarão mais bem preparadas para se destacar no mercado.

Izabelle Macedo
Marketing Director, Google Cloud LATAM



Nesse futuro de agentes, as startups terão a oportunidade de desenvolver agentes full-stack para mercados específicos e criar plataformas que ajudem as pessoas a desenvolver seus próprios agentes para atender a necessidades personalizadas.

Jill Greenberg Chase
Sócia investidora, CapitalG



Construa com uma infraestrutura “agnóstica” que permita que você aproveite os melhores modelos e bancos de dados da categoria, que estão em constante evolução.

Mayada Gonimah
CTO e Co-Founder, Thread AI

Domine seus dados. Entenda a origem, a qualidade e a estrutura de cada um deles. Não trate modelos de IA como caixas-pretas mágicas. Busque compreender o funcionamento delas. Sua vantagem competitiva virá de dados proprietários e da capacidade de ajustar modelos para suas necessidades específicas.

Rodrigo Scotti
Founder e CEO, Nama

Dica para startups aspirantes:

Foque na “adequação produto-algoritmo”. Entenda os pontos fortes e fracos da tecnologia atual e construa sua startup com base no que a tecnologia de IA pode fazer bem hoje, e não em uma suposta versão idealizada do futuro.

Yoav Shoham
Professor emérito de ciência da computação, Universidade de Stanford. Co-fundador, AI21 Labs



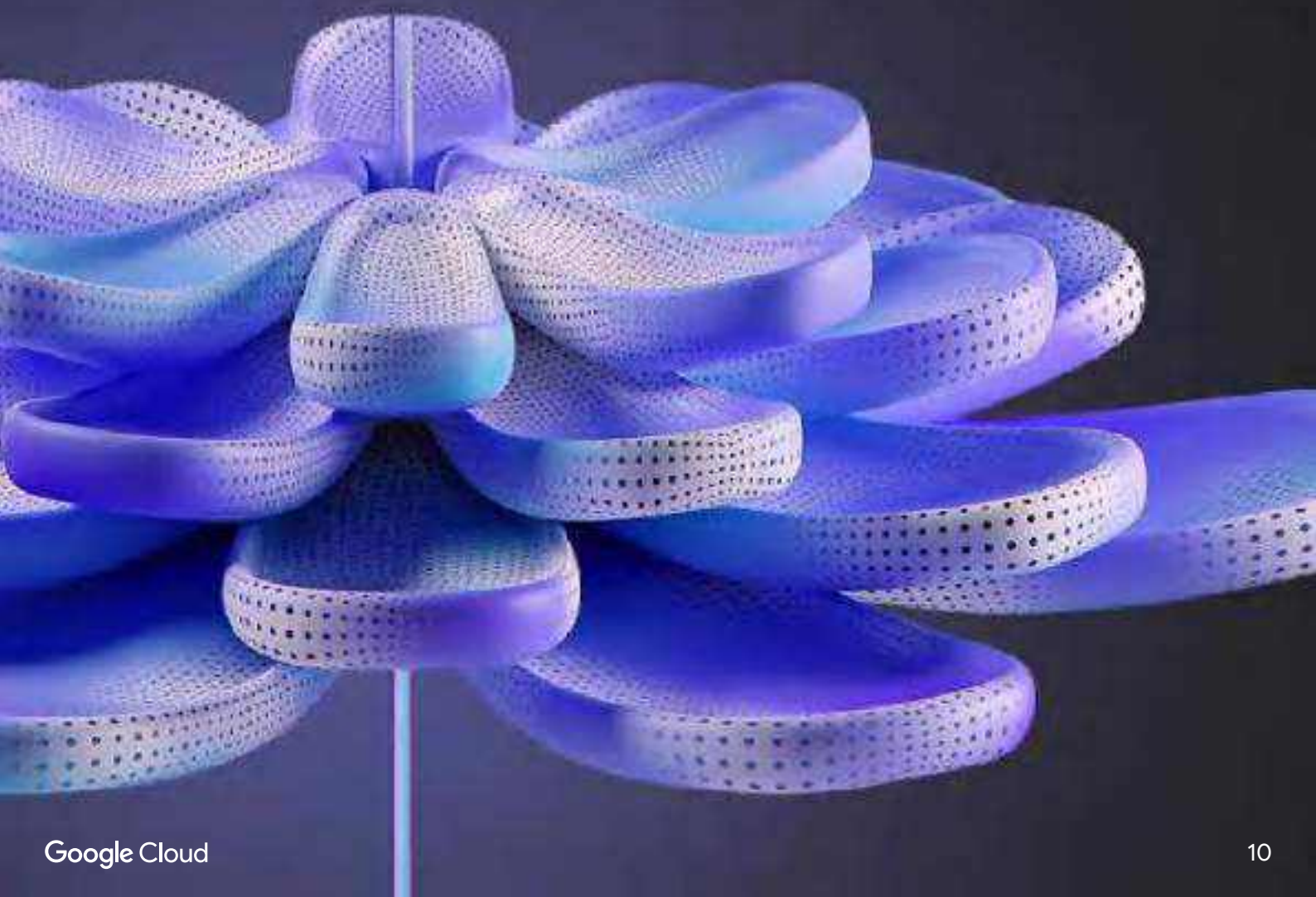
01

02

03

O que vem por aí em IA:

perspectivas dos líderes do mercado





Alison Wagonfeld



CMO, Google Cloud



Nomeada “Most Influential Women in the Bay Area” pelo The San Francisco Business Journal, Alison Wagonfeld é líder do Marketing Global do Google Cloud (Google Cloud Platform e Google Workspace) e possui experiência como Operating Partner na Emergence Capital (capital de risco em cloud empresarial).



Construiu dois marketplaces online de sucesso e é uma das fundadoras da QuickenLoans na Intuit.



Responsável por marketing e desenvolvimento de negócios em um site online de compra de carros (venture-backed), a profissional também é ex-Diretora Executiva do Harvard Business School California Research Center.

A era da hiperpersonalização: como a IA está transformando o marketing e a experiência dos clientes

Mais do que apenas uma ferramenta, a IA se tornou a força motriz na criação de jornadas de clientes personalizadas e intuitivas, capazes de antecipar necessidades e gerar um engajamento mais profundo.

A IA evoluiu para além de uma simples ferramenta, tornando-se um elemento transformador na interação entre empresas e clientes. Essa transformação é particularmente evidente quando se trata de marketing e experiência do cliente, áreas nas quais a integração da tecnologia tende a ser tão perfeita que se torna quase imperceptível.





As tendências que estou observando são todas voltadas para formas de tornar essas interações mais intuitivas e pessoais. A IA multimodal está permitindo que essa tecnologia compreenda o mundo de uma forma mais humana, processando texto, imagens, áudio e vídeo. Isso permite que as empresas criem experiências mais contextuais e intuitivas, analisem diferentes feedbacks dos clientes e gerem um conteúdo mais relevante.

Eu estou particularmente interessada nas seguintes áreas:

01 Personalização em larga escala

Os clientes esperam que as empresas os conheçam, prevejam suas necessidades e ofereçam experiências customizadas. A IA oferece um nível de personalização que nunca antes foi possível, permitindo que as empresas analisem o comportamento do usuário e forneçam recomendações de produtos e conteúdo de marketing em tempo real por e-mail, chat, online ou até mesmo em espaços físicos (dê uma olhada nesta [campanha](#) da PODS feita pela Tombras).

02 A análise do sentimento do cliente é essencial

Ela ajuda a entender como os clientes se sentem em relação à sua marca e aos seus produtos e serviços. A IA pode analisar uma variedade de dados, incluindo e-mails, redes sociais e interações de chat, só para avaliar as opiniões dos clientes em tempo real. Dessa forma, a empresa pode gerar respostas para qualquer problema e sanar as preocupações dos clientes de maneira proativa, falando até sobre os novos produtos.

03 A IA está possibilitando novos tipos de interação com os clientes

A IA vem sendo usada para criar novos canais de interação dos clientes com as marcas, como por exemplo os assistentes virtuais. É o que acontece nas ferramentas de planejamento de viagens, por exemplo. Em vez de os profissionais de marketing enviarem mensagens, estamos entrando na era das conversas que já estão embutidas no próprio marketing.



A IA nos ajudará a ajustar a criatividade em tempo real e a fazer um marketing mais personalizado em grande escala, conquistando o objetivo máximo da publicidade: alcançar o cliente certo, com a mensagem certa, no momento certo.

Para mim, a IA não é apenas uma tendência, mas uma ferramenta prática que aumenta a eficiência e impulsiona o crescimento, fazendo com que o marketing deixe de ser um processo linear e se transforme em um sistema dinâmico e interconectado. No futuro, a produção criativa será rápida; a ativação de mídia personalizada vai funcionar junto com a avaliação dos resultados em tempo real. É importante notar que, com a IA automatizando e personalizando a experiência do cliente, os profissionais de marketing terão mais tempo para se concentrarem na criatividade, na estratégia e na conexão autêntica com os clientes, que permanecem centrais para um marketing de sucesso.



Izabelle Macedo



Marketing Director, Google Cloud LATAM



Com passagem pela Universidade de Chicago e formação em Economia, além de um MBA concluído na Harvard Business School, Izabelle Macedo é Marketing Director no Google Cloud (América Latina).



Somando a habilidade em estratégia de negócios e marketing (B2C & B2B), Izabelle Macedo agrega a sua expertise ao time Google Cloud para alavancar projetos inovadores, alinhados com as atuais demandas de mercado.



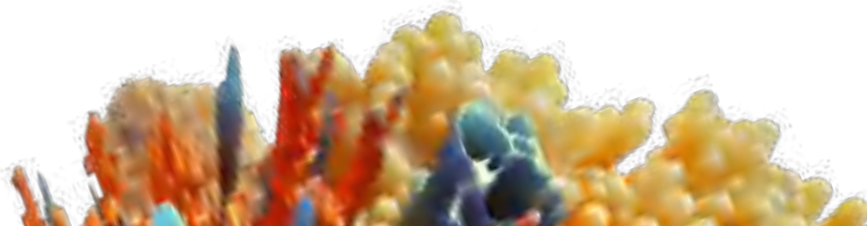
O seu background profissional conta com experiências anteriores na Deloitte Consulting e na Samsung Electronics, desempenhando posições de destaque nas áreas de consultoria, marketing e estratégia global.

A IA transforma o mercado e setores inteiros

Mais do que um recurso tecnológico, a inteligência artificial se tornou um aspecto fundamental para melhorar processos e criar conexões com os clientes.

Recentemente, testemunhamos uma evolução exponencial da tecnologia impulsionada pelo aumento do poder computacional, pelo grande volume de dados e pelo avanço de algoritmos cada vez mais sofisticados, como o Deep Learning. Essa crescente inovação tem permitido que a IA deixe sua marca, aperfeiçoando processos internos e elevando a experiência do cliente a patamares antes inalcançáveis.

O uso da IA se estende a áreas diversas e, para as empresas e indústrias, resulta em ações transformadoras, como a automação de tarefas repetitivas, a otimização da cadeia de suprimentos, a possibilidade de análise de dados para uma melhor tomada de decisão, a detecção de fraudes para garantir a segurança das informações etc.





Além disso, o atendimento inteligente com suporte 24 horas por dia em 7 dias por semana, a personalização de produtos e serviços, as recomendações e as experiências omnichannel oferecidas pela IA ajudam as organizações na tarefa de entregar produtos e serviços muito mais adequados às necessidades dos seus clientes.

Na América Latina, todas as vantagens que surgiram a partir dessa inovação já são uma realidade e estão acontecendo agora. Prova disso são os 90 casos de uso de IA em empresas, instituições e órgãos públicos da região, revelando avanços contínuos.



Aqui vão alguns exemplos:

01

No varejo, o Grupo Casas Bahia reforçou o seu e-commerce e a experiência de busca com IA.

02

Na saúde, a Clivi criou uma plataforma para monitorar as demandas de seus pacientes, fornecendo respostas personalizadas e reduzindo o número de complicações das doenças.

03

Nas telecomunicações, a TIM Brasil usa a IA generativa para transcrever os áudios das ligações realizadas no atendimento humano e classificar, sumarizar e qualificar com precisão a demanda do cliente.

04

Na área financeira, a KREDITO criou um modelo de avaliação de risco que melhorou a previsão de comportamentos de pagamento para ajudar os clientes a acessar capital de giro com mais agilidade.

05

Na educação, a YDUQS aplica a IA generativa para agilizar matrículas, trazendo mais precisão e segurança ao processo.

06

Na área de mídia, graças à IA, a Globo detecta, rastreia e combate a pirataria de conteúdos durante eventos ao vivo.

07

No setor público, o Governo de Minas Gerais está automatizando o processamento de autos de infração ambiental, aprimorando as análises.

08

No segmento de energia, a Casa dos Ventos tem na IA um apoio para acelerar a transição energética de seus clientes, com ganhos significativos em eficiência operacional e digitalização.

09

No setor de legaltech, a Fluna conta com a IA para realizar a avaliação, segmentação e sumarização de documentos.





Essas são apenas algumas das organizações latino-americanas que se beneficiam do impacto gerado pela inteligência artificial. Com base nesse cenário, podemos dizer que o futuro é otimista, por diversos motivos.

Um deles é a continuidade das pesquisas e do desenvolvimento das plataformas, gerando constantemente novas descobertas e algoritmos mais poderosos. Isso significa que as capacidades dessa tecnologia continuarão se expandindo e abrindo possibilidades de crescimento para os negócios (além disso, há também a oportunidade de reavaliação e melhoria das experiências que podem não ter dado certo anteriormente).

As ferramentas e plataformas ligadas à IA estão se tornando mais acessíveis e fáceis de usar, o que simplifica a implementação de soluções para organizações de todos os portes, sem a necessidade de grandes investimentos em infraestrutura ou equipes especializadas.

Em relação à experiência do cliente, o diferencial competitivo trazido pela inteligência artificial está diretamente relacionado aos recursos mais modernos para entender, engajar e encantar o público-alvo em cada ponto de contato.

Ao se fazer presente em diferentes cenários, a IA pode se tornar a porta de entrada para o desenvolvimento de modelos de negócios disruptivos — não como um modismo a ser incorporado a qualquer custo, mas como uma ferramenta poderosa para aprimorar processos, ganhar eficiência e impulsionar a inovação. Por isso, acredito que as startups que souberem explorar o seu potencial com uma abordagem estratégica e objetivos claros estarão mais bem preparadas para se destacar no mercado. Isso também envolve treinar os colaboradores, incentivar o uso responsável da tecnologia e criar um ambiente onde as pessoas se sintam seguras para aprender, compartilhar ideias e trazer inovações para dentro da empresa. Abraçar essa oportunidade sem medo é o primeiro passo para transformar potencial em resultado.



As ferramentas de IA estão se tornando mais acessíveis e fáceis de usar, simplificando sua adoção em companhias de todos os portes, sem a necessidade de grandes investimentos.





Amin Vahdat



VP/GM ML, Systems, e Cloud AI,
Google Cloud



Amin Vahdat é engenheiro, fellow e vice-presidente da equipe de Machine Learning, System e Cloud AI.



Antes de ingressar no Google há quase 15 anos, ele foi professor de Ciência da Computação e Engenharia da SAIC na Universidade da Califórnia em San Diego. Amin é Fellow da ACM e membro da Academia Nacional de Engenharia dos EUA.

Daqui a 10 anos, a nossa infraestrutura estará irreconhecível

Tanto em software quanto em hardware, estão ocorrendo mudanças fundamentais em computação, rede, armazenamento e outras áreas.

A arquitetura existente hoje é considerada um fato dado, pois representa a “sabedoria tradicional” de como os sistemas devem ser construídos. As mudanças em curso nos últimos cinco anos ou mais significam que estamos nos estágios iniciais ou intermediários de uma nova revolução, e até o fim da década ela transformará a emergente plataforma de computação em algo irreconhecível.

Um dos maiores responsáveis por essa transformação é a mudança na forma como acessamos e entregamos informações em tempo real para a geração proativa de insights e interpretações impulsionada pela IA generativa (GenAI).





Devido à enorme demanda de computação e comunicação que a GenAI exige, é necessário fazer algumas mudanças no design e na construção da infraestrutura subjacente de todo o stack, incluindo:

01 Computação específica para aplicativos

A computação da IA generativa envolve um volume imenso de multiplicações de matrizes e cálculos numéricos relacionados. Com essa especialização, a eficiência computacional de um hardware desenvolvido para um subconjunto muito específico de computação de propósito geral será, pelo menos, dez vezes maior em termos de desempenho normalizado pelo custo ou desempenho por watt.

02 Especialização em redes

Estamos observando uma grande transformação de hardware, protocolos e software de rede padronizados em redes especializadas, como a Inter Chip Interconnect (ICI) para TPUs e a NVLink para GPUs. Essas redes são projetadas para a tarefa muito específica de oferecer suporte a primitivas de computação de nível superior, como “all-reduce” em hardware com pouca ou nenhuma camada de software, sem gerenciamento de sistema operacional e, basicamente, fazendo transferência direta de memória para memória para poder executar essas operações.

03 Barreira da memória

A memória de alta largura de banda (HBM), que empilha a RAM em 3D no mesmo pacote do processador para reduzir drasticamente a latência e aumentar a largura de banda, está atingindo suas próprias limitações fundamentais: a barreira da memória. Aumentar a capacidade de processamento sem poder fornecer os dados necessários para isso, seja da memória local ou da rede, resultará não apenas em um grande desperdício de recursos computacionais e de energia, mas também na necessidade de inovações avançadas na arquitetura de processamento e memória para manter o equilíbrio do sistema.

04 Compactação computacional

A sincronização precisa e os enormes requisitos de energia levam a infraestrutura a níveis nunca antes vistos de densidade computacional. A computação deve ocorrer em elementos homogêneos, a comunicação deve ser planejada e coordenada, e a tolerância a falhas deve ser eficiente. Do ponto de vista da latência e do custo/energia, compactar quantidades cada vez maiores de poder computacional em ambientes menores e de alta densidade energética pode trazer benefícios significativos.

05 Resfriamento líquido

A taxa de computação intensa e contínua de machine learning (ML) muda a equação de tal forma que o ganho de eficiência no desempenho mais rápido dos chips justifica totalmente a maior densidade energética e o resfriamento líquido. Este último, por sua vez, exige a reconfiguração das instalações do datacenter e da infraestrutura de resfriamento, dos racks e de outros componentes, para oferecer, de ponta a ponta, o mais alto nível de eficiência — bem como tecnologias de geração de energia fora da rede (inclusive eólica, solar, hidrelétrica e conjuntos de baterias), que ajudem a entregar a computação de ML em vários locais do planeta.



Recursos prontos para uso com capacidade avançada de raciocínio, como orquestração de LLMs com geração aumentada de recuperação (RAG) e chamada de função, permitirão que as startups tenham acesso a modelos e à infraestrutura de IA sem grandes despesas de capital ou recursos de engenharia.



A evolução da infraestrutura de IA afetará as startups de diferentes maneiras

Este enorme crescimento na capacidade de processamento é acelerado graças às melhorias contínuas na confiabilidade do modelo e também por conta da redução dos custos de computação. Como resultado:

01 A qualidade, a confiança e a latência do modelo aumentarão

Se o foco da sua startup é a veracidade e seu objetivo é criar ofertas de alta qualidade, com pouca ou nenhuma alucinação, então você ainda deve se concentrar nesta questão. Mas se seu objetivo é fornecer outro serviço, aproveite o nível de qualidade cada vez maior que todos os modelos estão fornecendo agora.

02 Os custos de computação despencarão

O custo da computação está caindo rapidamente. Se eu fosse uma startup de IA, trabalharia com a crença de que a computação será muito mais rápida, barata e confiável no futuro. Vamos supor que você tenha uma ideia incrível e tudo que precisa é reduzir os custos de computação em 90% para que ela seja lucrativa. Para mim, essa é uma aposta fácil, mesmo que você precise de uma redução de 99%.

Como vimos em todo o setor, os desenvolvedores de modelos continuarão a otimizar a economia de custos, utilizando uma infraestrutura básica com uma margem de lucro muito pequena. A disponibilidade de um conjunto poderoso e diversificado de aceleradores (GPUs e TPUs) continuará sendo fundamental, assim como os recursos da rede de alta taxa de transferência destacados anteriormente.

A grande maioria das startups, focada na construção de aplicativos inovadores e diferenciados, passará a usar mais serviços baseados em software. Os recursos prontos para uso com capacidade avançada de raciocínio, como orquestração de LLMs com geração aumentada de recuperação (RAG) e chamada de função, terão maior demanda, pois dão acesso a modelos e infraestrutura sem grandes despesas de capital ou recursos de engenharia.





Eric Acher

Co-Founder e Managing Partner, Monashees



Eric sempre acreditou no poder do empreendedorismo, da tecnologia e do capital de risco para transformar a América Latina.



Em 2005, ele criou a Monashees junto com o cofundador Fabio Igel, como a primeira empresa de capital de risco latino-americana inspirada nas melhores práticas do Vale do Silício. Na Monashees, ele lidera a estratégia de investimento junto à gestão geral da empresa e também se concentra em investimentos em tecnologia emergente, IA, empresas de consumo e edtech.



Eric trabalhou em estreita colaboração com a Elo7 (adquirida pela Etsy), Enjoei (IPO), Jusbrasil, MadeiraMadeira, Moises (hoje Music AI), Neoway (adquirida pela B3), Nomad, PetLove e VivaReal/Grupo Zap (adquirida pela OLX), entre outras.

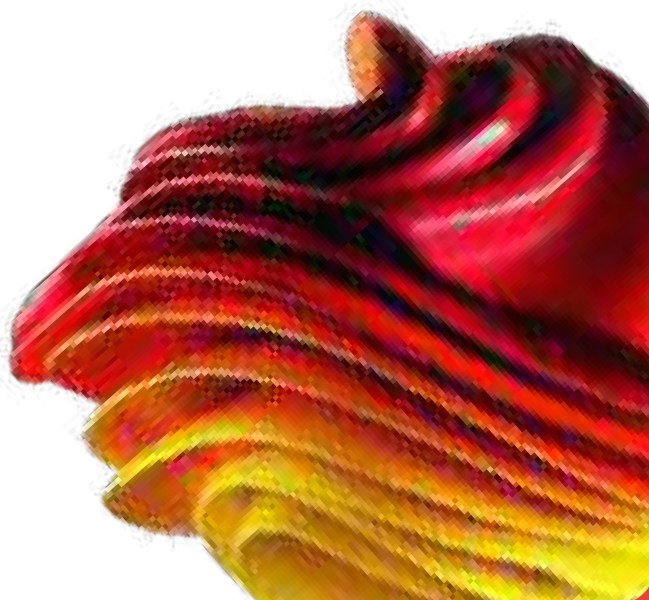
Planejar para vislumbrar crescimento

Na era da IA, a preparação das estratégias, a atenção aos anseios do mercado e a coragem para arriscar no momento ideal são os pontos altos.

Ousadia latino-americana no empreendedorismo

Uma das percepções que tenho é que principalmente no Brasil e na América Latina há um empreendedorismo latente e, às vezes, muito informal. Mas a verdade é que em outros países do mundo você não vê essa energia e essa resiliência para seguir em frente.

Em um território como o nosso, onde as pessoas são apaixonadas por tecnologia e são early adopters, existe um potencial absurdo a ser explorado, no que tange a área de inteligência artificial. Só que ainda temos muito chão pela frente, porque é preciso considerar o risco de execução dos projetos — isso envolve falta de infraestrutura, enfrentar burocracias, depender da qualidade de serviços, entre outras coisas.





Ainda que haja uma evolução a ser perseguida, ao acompanhar tudo o que está acontecendo em diversos segmentos, como por exemplo em FinTech ou LegalTech, a utilização de IA pode realmente causar um impacto incrível. Por isso, podemos afirmar que algumas empresas são candidatas a se transformarem no primeiro unicórnio de inteligência artificial da nossa região, porque crescem rápido e demonstram consistência.

Aqui, cabe uma curiosidade: hoje, investimos em empreendedores latino-americanos fora da América Latina também. Um exemplo é a Music AI, uma plataforma de inteligência artificial para música e áudio, situada em Salt Lake City (EUA) e encabeçada por brasileiros de João Pessoa, Paraíba. A plataforma já soma mais de 50 milhões de usuários no mundo, em cerca de 180 países. E são essas histórias inspiradoras que despertam cada vez mais a ambição e a confiança dos outros empreendedores.

Os ciclos de desenvolvimento

No entanto, sempre que refletimos sobre os avanços, é preciso entender os ciclos. No que se refere às plataformas tecnológicas, eles são semelhantes ao que vemos em Venture Capital. Há um período de 15 anos que vai da incubação da plataforma, passando pela sua aceleração até a fase de saturação. Depois da saturação, vem normalmente uma correção de mercado.

Atualmente, estamos vivendo em duas realidades diferentes: uma delas está focada na tech tradicional, e a outra só quer saber do over appreciation da inteligência artificial.

Navegar nesse mar é diferente, mas uma coisa segue definida: o momento de investir é agora. Afirmo isso porque se você investe na incubação, é muito cedo e perde o timing. Por isso, lá em 2018, quando começamos a ver os primeiros movimentos, ainda não era a hora certa. O Brasil oferece — e vai propiciar ainda mais — oportunidades de aplicação dessa tecnologia. A inovação virá do desenvolvimento de aplicações com barreiras de entrada, e isso vai depender muito do modelo de negócio.

Nós temos o passado como referência para reconhecer padrões e nos inspirar, mantendo a ideia de que cada ciclo é único. Sob este ponto de vista, a jornada do empreendedor pode ser transformada quando se cria uma empresa winner em seu meio. Portanto, se aquele negócio é pioneiro por algum motivo ou é destaque, ele será um modelo para os demais.



O Brasil oferece — e vai propiciar ainda mais — oportunidades de aplicação dessa tecnologia. A inovação virá do desenvolvimento de aplicações com barreiras de entrada, a depender do modelo de negócio.





O futuro é promissor, o que resta é saber quem está pronto para traçá-lo

No caminho de quem começa agora, é possível ver que as atualizações e as mudanças proporcionadas pela IA irão moldar cenários. Alguns ainda são céticos e negam essa realidade, enquanto os mais atentos se preparam para encarar a inovação. Eu tenho a tese de que, em ciclos próximos, as companhias puramente tecnológicas, que dependem apenas de software e tecnologia digital, devem se concentrar em construir negócios com mais eficiência de capital.

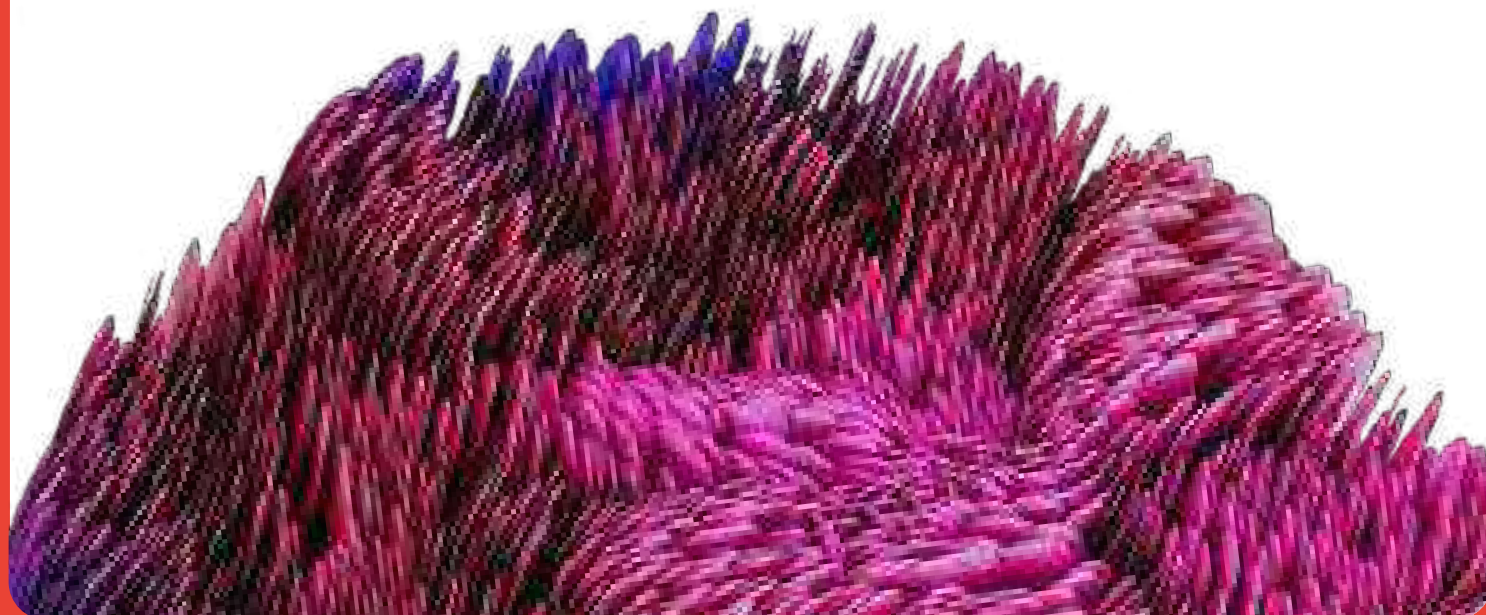
Dessa forma, o foco está na criação de aplicações, com empresas que crescem mais rápido e de forma mais eficiente. Por esta razão, quem trabalha com software precisará se reinventar e adotar a IA rapidamente para manter a sua relevância. Assim, empreendedores que atuam com modelos tradicionais terão mais dificuldades em encontrar espaço, pois soluções fáceis e superficiais serão facilmente commoditizadas.

Recentemente nós fizemos um exercício de planejamento de cenários, avaliando as dimensões e os gatilhos que podem ser originados de cada um desses cenários. Sabemos que o futuro é turbulento e incerto, mas também fascinante. Na Monashees buscamos o que pode funcionar e, nesse paradigma, vemos uma era de ouro para a utilização da IA nos próximos anos.

Nós acreditamos que a IA será utilizada para gerar produtividade e resolver problemas, exercendo um impacto incrível em áreas como educação, saúde, segurança e jurisprudência. O mais interessante disso é que os empreendedores irão liderar essa fase.

É claro que também existem ameaças, como a perda de alguns empregos e o desemprego localizado. A IA impulsionará as pessoas para novas profissões. Embora isso ocorra, vejo como um privilégio vivenciar essa era. Todos nós temos a responsabilidade de conduzi-la em uma direção positiva, com valores éticos para o uso da tecnologia.

Eu sou otimista por natureza e uma analogia que me vem à mente é o que houve com a pintura no momento em que nasceu a fotografia. A partir daí, o mundo presenciou uma explosão de criatividade no meio da pintura, vários artistas começaram a fazer parte de movimentos importantes, e a própria fotografia se tornou uma forma de arte. Seguindo essa linha, entendo que a IA fará coisas surpreendentes, ao passo que também haverá uma explosão de criatividade e inteligência humana.





Apoorv Agrawal



Sócio, Altimeter Capital



Lidera investimentos em startups de software e IA com investimentos notáveis em empresas como OpenAI, Glean, Parloa e Everest



Engenheiro por formação, iniciou sua carreira como Forward Deployed Engineer na Palantir e ainda programa

A automação computacional e a IA multimodal são o futuro da computação pessoal

A próxima etapa de crescimento da IA automatizará as tarefas rotineiras e humanizará a tecnologia

Vivemos um momento em que estamos viciados em computadores e telefones. Acredito que, no futuro, olharemos para isso como um período estranho na história humana. Daqui a cem anos, as pessoas não estarão mais grudadas em suas telas. A IA nos libertará. Particularmente, a ascensão da IA multimodal — ou IA que combina voz, visão e processamento de linguagem natural — tornará a interação com o mundo digital completamente fluida.

Imagine um mundo em que você pode controlar toda a sua vida digital de forma conversacional: fala, comandos visuais, toque e, quando tudo mais falhar, digitação. Esse é o futuro que estamos construindo, e ele desbloqueará grandes oportunidades para as startups nos próximos anos.





Apostando em automação, eficiência e experiências centradas no ser humano

Sou um investidor focado em IA que acredita em um futuro impulsionado pela IA multimodal, e por isso mesmo três temas centrais que guiam minha tomada de decisões:

01 Automatização de rotinas

O atendimento ao cliente, funções de back-office, qualquer coisa que envolva muitos dados e se baseie em regras está pronta para a IA. Quero investir em IA para áreas nas quais os humanos estão realizando tarefas que não trazem satisfação nem alegria, e que não são criativas. A IA se destaca em automatizar tarefas repetitivas e ricas em dados, liberando os humanos para tarefas mais significativas.

Da mesma forma, também estou interessado em fluxos de trabalho de agentes, nos quais agentes de IA gerenciam tarefas ou processos inteiros para nós. Não é surpresa para mim que algumas das mentes mais brilhantes estejam criando ferramentas para nos libertar de clicar em botões e ficar curvados diante de uma tela.



Quero investir em IA para áreas nas quais os humanos estão realizando tarefas que não trazem satisfação nem alegria, e não são criativas. A IA se destaca em automatizar tarefas repetitivas e ricas em dados, liberando os humanos para trabalhos mais significativos.

02 Aumento das capacidades humanas

A IA tem como objetivo nos tornar mais inteligentes e eficientes. Por exemplo, ferramentas capazes de sintetizar grandes quantidades de informações e fornecer insights acionáveis para desbloquear ganhos tangíveis de produtividade. Estou interessado em startups que desenvolvem soluções de IA que ajudam a melhorar o que os humanos podem fazer.

03 Humanização da tecnologia

Acho que a IA de voz, em particular, oferece uma maneira mais natural e intuitiva de interagir com a tecnologia. Tenho me concentrado em provedores de infraestrutura de IA de voz porque acredito que essa área está em plena expansão. As startups que podem criar experiências de IA verdadeiramente centradas no ser humano terão uma grande vantagem no mercado. E outra área interessante a explorar é como podemos usar a IA para nos entender melhor e nos conectar com os outros, melhorando os relacionamentos, e não só os ganhos de produtividade.





Mudando o foco para a camada de aplicação

À medida que o custo de infraestrutura e dos modelos despenca, a oportunidade está se deslocando para a camada de aplicação. Pense nisto: o stack de IA é atualmente um triângulo invertido, com a maior parte do valor concentrada em semicondutores e infraestrutura. Mas, conforme avançamos para os próximos estágios do crescimento da IA, esse triângulo vai se inverter. A camada de aplicação se tornará o local onde a inovação e a criação de valor acontecem.

Para as startups, isso significa focar em construir produtos e serviços baseados em IA que resolvam problemas do mundo real. Não se perca no complexo mundo dos modelos fundacionais. Em vez disso, aproveite a infraestrutura existente e concentre-se em criar aplicações inovadoras que entreguem um valor tangível aos usuários. E desenvolva já pensando na durabilidade, para que você não só sobreviva, mas também tenha sucesso com as próximas gerações de modelos e tecnologias de suporte. Acredito que essa é a verdadeira oportunidade e vejo duas grandes tendências emergindo:

01 Aplicativos horizontais

Ferramentas alimentadas por IA que melhoram a produtividade em todas as funções de trabalho.

02 Aplicativos verticais

Produtos especializados de IA criados para atender a funções específicas, como engenheiros ou representantes de atendimento ao cliente.

Um debate fundamental é se a IA horizontal se tornará tão boa a ponto de substituir a necessidade de soluções verticais. Outra questão-chave que as startups precisam considerar é: será melhor trabalhar com preços por usuário, baseados no uso ou baseados em valor para o seu produto? Não há uma resposta fácil, mas aconselho as startups a alinharem seus preços ao valor entregue.

Para finalizar, vou repetir o que costumo dizer aos fundadores que me pedem conselhos sobre como usar a IA de forma eficaz:



Dados

Não existe estratégia de IA sem uma estratégia de dados. Entenda quais dados você tem e como pode usá-los para treinar modelos de IA eficazes.



Integração

Considere como você pode integrar sua solução de IA a outras plataformas e fontes de dados para criar uma experiência de usuário fluida.



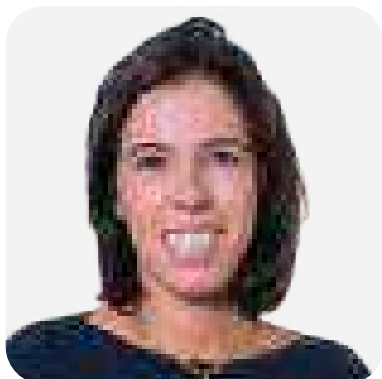
Fluxo de trabalho

Por fim, pense nas tarefas que as pessoas não querem ter que realizar, como reservar voos. A IA pode automatizar essas tarefas ou torná-las mais intuitivas?





Laura Constantini



Founder, Nido e VC investor



Contabilizando mais de 16 anos de atuação em Venture Capital, Laura Constantini investiu em diversas empresas no decorrer de sua jornada, como Jota, Cayena, Croct, Traive, Estoca, Destaxa e TáOn, a partir do seu portfólio construído durante os anos na Astella, empresa pela qual continua responsável.



A profissional é também fellow do Kauffman (class 23) e atua nos conselhos da LAVCA e da Endeavor Brasil, onde segue apoiando o crescimento de empreendedores.

A tecnologia como propulsora de negócios inovadores

Facilmente disponível e acessível a todos, a IA guarda o seu melhor para quem sabe exatamente como usá-la.

Humanização na era da IA

Com um papel relevante na Astella Investimentos e agora liderando a Nido, parte da minha missão profissional é preencher lacunas dentro do ecossistema de inovação do Brasil. Na Nido, fomentamos investimentos orientados por propósito, fundamentados na disciplina de Capital de Risco, direcionando portfólios familiares impactantes que incorporam seus valores e aspirações.

E é desta forma que um trabalho lado a lado com as famílias é feito, para definir as suas teses de investimento, alinhar expectativas e criar um portfólio adequado a cada realidade. Por meio de uma governança cuidadosa, exposição a diferentes ciclos de mercado e uma conexão profunda com a jornada empreendedora, a Nido apoia as famílias para que elas invistam em inovação, construindo um legado de orgulho e significado duradouros.





Nós usamos a tecnologia como aliada nesse processo. No entanto, ao longo da minha atuação no mercado, o que eu percebi é que a demanda atual no âmbito da gestão financeira é por uma humanização maior. Obviamente, muitas situações cotidianas poderão ser facilmente substituídas por robo-advisors, que oferecem serviço de planejamento de investimentos automatizado, com recomendações e calibração de portfólios. Por outro lado, o aconselhamento e a consultoria direcionada somente poderão ser alcançados se houver o fator humano como ponto de destaque.

O volume de informações com as quais lidamos diariamente é gigantesco. Os hedge funds, por exemplo, já operam com uma capacidade de processamento e muitos algoritmos enormes, cruzando flutuações de moedas, dívida pública, crescimento econômico etc., e conseguem chegar a conclusões de alocação de forma bem mais eficiente que os humanos. No entanto, organizar e adequar relatórios e dados para o contexto da realidade brasileira, em uma linguagem que o investidor não só entenda como tenha prazer em ler e interagir com o conteúdo, torna-se essencial sob esta perspectiva.



Tudo que é nativo de IA passa pelo preceito de trabalhar com bases nativas e proprietárias. Enriquecer essas informações custa caro, e é nesse sentido que eu enxergo a verdadeira barreira.

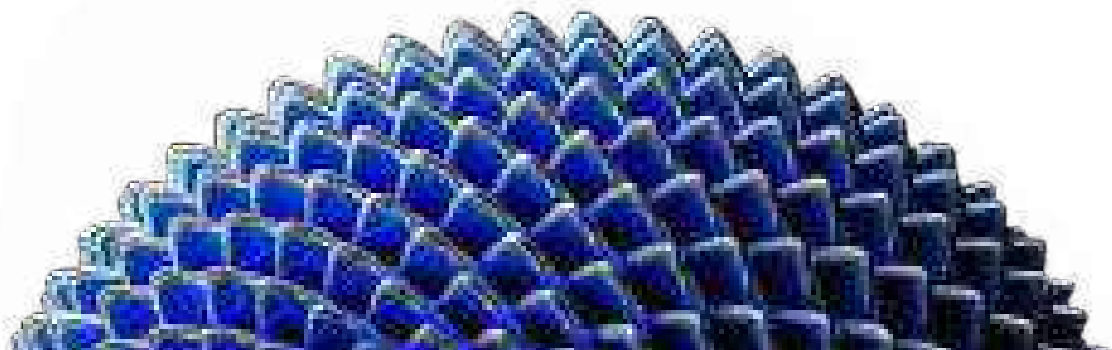
Brasil não é celeiro para LLMs, mas pode agregar valor de outras formas

Em relação a todas as mudanças tecnológicas que temos vivenciado, do mesmo modo que ocorreu com outras plataformas, chegamos à conclusão de que o Brasil não é um celeiro para os grandes LLMs (large language models). Os grandes transformers, LLMs e as iniciativas voltadas à inteligência de processamento desenvolvem-se fora do Brasil. Mas há, sim, essa camada de aplicações e de customização da inteligência artificial que abraça diferentes cadeias de valor e que, inclusive, já é feita no nosso país.

Aqui, nós contamos com questões muito particulares — e delas surgem soluções globais. O Pix foi gerado no contexto de uma demanda de mercados emergentes (e hoje sabemos que faz sentido para o mundo inteiro). O que eu quero dizer com isso é que carregamos a inovação no nosso DNA e podemos criar projetos muito interessantes para resolver problemas complexos.

Existem ações implementadas em solo brasileiro que podem ser a base de uma infraestrutura para suprir demandas nos setores dos mais variados, do agronegócio passando pela saúde e as fintechs. Neste paralelo, temos outro ponto de atenção. Simplesmente porque a dinâmica de investimento/crescimento muda e, muitas vezes, é preciso um capital mais inovador e mais paciente. Uma das perguntas que precisamos fazer é: as novas empresas estão preparadas para enfrentar esse cenário?

Ainda, junto a essa avalanche de novidades e modernização, vem a necessidade de uma capacidade de processamento de dados muito grande. A extração do conhecimento e da informação é massiva, combinada à acessibilidade, que segue distribuída para um contingente significativo de pessoas.





A diferenciação no mercado: o seu custo e tempo de maturação

Voltamos a falar, então, do surgimento de outras plataformas disruptivas para comparar com o momento em que vivemos. Na era mobile, precisávamos de construção de rede e de um celular nas mãos de cada indivíduo. Quando a IA surgiu, um dos motivos para a rápida adoção foi o fato de que a tecnologia necessária para usufruir de IA já era parte do dia a dia das pessoas.

Você conversa com a IA na sua casa, no tablet, no seu celular, e ativa essa interação em linguagem conversacional. Portanto, é uma interface facilitada. Se a plataforma, desse modo, está mais disponível e acessível, significa que a barreira de entrada é menor. Tudo que é nativo de IA passa pelo preceito de interagir com a tecnologia em linguagem natural. O diferencial está em trabalhar com bases de dados proprietárias. Agora, enriquecer e explorar essas informações custa caro, e é nesse sentido que eu enxergo a verdadeira barreira.

Você precisa se debruçar sobre os dados, higienizar, pesquisar, criar valor de uma forma única. Eu compreendo que, abordando as facilidades da IA, o que vemos é uma robustez maior e o que vem na sequência é uma inteligência focada em uma necessidade específica.

Dito isso, a minha reflexão para um empreendedor à frente da sua startup seria: Qual é a inteligência única que você está criando? Qual é o insight diferenciado que você traz e que pode ser aplicado em cima de uma base de dados?

Ao saber que é preciso ser proprietário desse fator único e inovador e, além disso, que ele deve ser duradouro o bastante no universo da inteligência artificial, um aspecto importante também diz respeito à capacidade de enriquecimento desses dados (pois é isso que vai catapultar o empreendedor para mais longe). Quem tem um olhar profundo para as cadeias de valor e para o que os clientes buscam alcançar, com certeza estará mais próximo do êxito.





Jill Greenberg Chase



Sócia investidora, CapitalG



Lidera os investimentos em IA para o fundo de crescimento independente da Alphabet, CapitalG. Os investimentos incluem Magic.dev e /dev/agents.



Ex-CEO de uma empresa financiada por capital privado e fundadora de uma startup financiada pela Y Combinator, além de palestrante convidada na Stanford University Graduate School of Business desde 2019.

Adeus, copilotos: olá, agentes

Acredito que estamos entrando rapidamente em uma era em que a IA é inteligente o suficiente para gerenciar fluxos de trabalho inteiros. Então, como as empresas podem se adaptar?

Embora muitos no campo da IA prevejam uma escalabilidade infinita que resultará na inteligência artificial geral (AGI), acredito que os modelos fundacionais alcançarão um ponto de estagnação nos próximos anos. Não que a escalabilidade seja ineficaz, mas as limitações técnicas nos impedem de avançar.

Imagine desenvolver um modelo com várias centenas de trilhões de parâmetros. Acho que seria um modelo extremamente atrativo que provavelmente alcançaria a AGI. Mas, na prática, é muito difícil treinar um modelo dessa magnitude. Estamos falando de enfrentar barreiras em todas as ordens de grandeza: 10 trilhões de tokens, 100 trilhões de tokens e assim por diante. Cada correção exige um novo treinamento, o que torna o processo demorado e com uso intensivo de recursos.

Essa limitação, no entanto, apresenta uma oportunidade interessante para as startups. Enquanto os grandes players sofrem com a escalabilidade, as startups podem se concentrar em oferecer um ROI tangível para os clientes. Isso pode acontecer no lado da infraestrutura, tornando o treinamento mais eficiente, ou no lado da aplicação, desenvolvendo soluções que não dependem de modelos massivos.





Como estou investindo na era dos agentes de IA

Olhando para o futuro, vejo duas grandes tendências emergirem e pretendo investir em startups que estejam explorando essas áreas.

01 A ascensão dos fluxos de trabalho agênticos

Acredito que estamos indo além dos copilotos de IA e entrando na era dos agentes autônomos. Essa mudança é alimentada por avanços nos modelos de base, especialmente na capacidade de raciocínio. As implicações orçamentárias para as empresas são enormes, permitindo que elas economizem não apenas nos custos de software, mas também na mão de obra.

Imagine agentes de IA que sejam capazes de fazer o atendimento ao cliente sozinhos, indo muito além de apenas auxiliar os agentes humanos ou atuar como uma parte da programação em par para que os desenvolvedores sejam mais produtivos. Isso significa um mercado de bilhões de dólares. As startups têm um papel fundamental para desempenhar nesse futuro de agentes autônomos. Elas podem desenvolver agentes full-stack para mercados específicos, como suporte ao cliente, saúde, finanças e jurídico, focando nos desafios complexos de controle e conectividade. Eles também podem criar plataformas que ajudem as pessoas a desenvolver seus próprios agentes, atendendo a necessidades personalizadas.



Eu acredito sinceramente que, quando toda essa tecnologia de agentes funcionar, os seres humanos poderão dedicar muito mais tempo às atividades de maior valor.

02 Agentes do tipo direct-to-consumer

É aqui que as coisas ficam realmente interessantes para os consumidores. Imagine um agente de IA que gerencia suas tarefas diárias. Eu poderia simplesmente dizer ao meu agente de IA: “Ei, me ajude a não me atrasar para a reunião”. Ele vai verificar minha agenda, calcular o tempo de viagem e chamar um Uber ou Lyft, tudo sem que eu precise navegar por diferentes aplicativos.

Isso tem o potencial de revolucionar o desenvolvimento de aplicativos. Se estivermos interagindo diretamente com um agente de IA, o valor de um app deixa de estar na sua interface e passa a estar na sua funcionalidade. Podemos até ver o surgimento de superapps de IA que atuam como um hub central para gerenciar todo o nosso dia, transformando ainda mais a forma como interagimos com a tecnologia.

Observe que o surgimento de agentes do tipo direct-to-consumer também levanta questões sobre privacidade e controle de dados. Quão confortável você se sente com diferentes aplicativos acessando e compartilhando suas informações? É aqui que as startups podem intervir, criando frameworks que permitam aos usuários definir suas preferências e controlar como seus dados são utilizados por esses agentes interconectados.

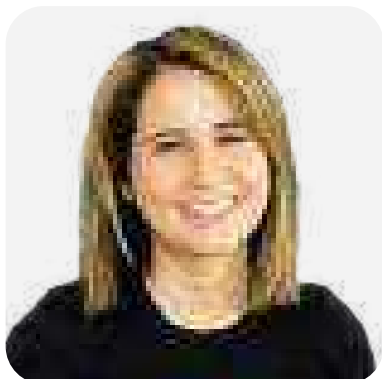
Construindo a estratégia da sua startup na era dos agentes

Os investidores estão em busca de startups que consigam demonstrar um bom product-market fit, uma vantagem competitiva distinta e um caminho claro para a lucratividade desde o início. Esses fundamentos continuam sendo essenciais para enfrentar o cenário em constante evolução da IA.

Eu acredito sinceramente que, quando toda essa tecnologia dos agentes funcionar, os seres humanos poderão dedicar muito mais tempo às atividades de maior valor. Imagine um mundo onde qualquer pessoa possa transformar suas ideias em realidade, e o único fator limitante é a própria imaginação. Esse é o tipo de futuro que me empolga, e estou ansiosa para colaborar com mais fundadores incríveis que estão tornando esse futuro possível.



Mayada Gonimah



CTO e Co-Founder, Thread AI



A Thread AI é uma plataforma de orquestração de IA modular para as organizações criarem, implementarem e gerenciarem seus fluxos de trabalho.



O foco e as principais responsabilidades de Mayada, como CTO, são o desenvolvimento, a arquitetura e a execução da estratégia técnica da Thread AI.



Mais recentemente, ela atuou como engenheira de sistemas distribuídos na Palantir, liderando o desenvolvimento de uma plataforma de inferência de IA em tempo real. Durante sua passagem pela Palantir, ela desenvolveu sistemas integrados de infraestrutura de IA, incluindo a criação de estruturas abrangentes de orquestração e avaliação de modelos, bem como sistemas complexos de containerização.

O fluxo de trabalho e a infraestrutura são mais importantes do que belas ferramentas

Foque nos fundamentos dos sistemas distribuídos e desenvolva pensando nas camadas de IA em evolução.

Estamos vendo uma explosão repentina de ferramentas que tentam resolver uma pequena parte da cadeia de valor. Embora seja ótimo que todos estejam entusiasmados com a IA e que ela esteja começando a resolver alguns problemas reais, alguns modelos de IA, soluções inacabadas e ferramentas frágeis estão sendo lançados no mercado sem provar que realmente fazem a diferença, seja no ROI ou na mudança fundamental do ecossistema de desenvolvedores.





Durabilidade, orquestração de fluxo de trabalho e observabilidade como elementos-chave

Há um retorno às origens, pois as empresas estão vendo que não precisam reinventar toda a sua estrutura para incorporar um fluxo de trabalho baseado em IA.

Minha tese é de que a IA não deve ser um fluxo de trabalho em si mesma, e sim ser parte de outro fluxo de trabalho. Portanto, as integrações e a infraestrutura para gerenciamento de fluxo de trabalho observável se tornarão partes ainda mais críticas do stack de IA. As empresas precisarão de interfaces que permitam inserir a IA nos fluxos de trabalho existentes e testá-los junto com os processos que já estão em funcionamento.

Acho que será fundamental sermos mais criteriosos em relação ao local em que incorporamos a IA, e também quanto às ferramentas necessárias para a observabilidade ao implementar esses fluxos de trabalho nativos de IA dignos de produção. Isso, trazendo de volta o foco nas bases de sistemas distribuídos e entendendo que, só porque algo é probabilístico, isso não significa que seja um paradigma completamente novo. O que precisamos é das verificações e controles certos para IA. Sem isso, não devemos colocá-la no meio de um processo crítico, onde não é possível lidar com resultados probabilísticos.



As empresas estão vendo que não precisam reinventar toda a sua estrutura para incorporar um fluxo de trabalho baseado em IA. A IA não deve ser um fluxo de trabalho em si mesma, e sim parte de outro fluxo de trabalho.

Desenvolva pensando em uma infraestrutura e um bancos de dados em evolução

Muitas camadas de IA — a infraestrutura e os bancos de dados — estão evoluindo. Então, ao elaborar seu roadmap e seu negócio como uma startup de IA, como você considera essas camadas? Como você constrói a camada de infraestrutura sabendo que esses modelos vão melhorar rapidamente ao longo do tempo?

Acho que você deve garantir que, se uma nova versão de um modelo surgir, não será necessário modificar toda a sua infraestrutura. Você precisa de um sistema de plug-ins onde as coisas sejam versionadas e, talvez, você integre os modelos no nível do protocolo, em vez de fazer isso com cada SDK. Você adiciona as ligações para gRPC, REST ou GraphQL, para poder simplesmente adicionar esse modelo mais recente e avançado. O mesmo está acontecendo com diferentes técnicas de banco de dados.





Adote uma abordagem modular para IA

Acredito sinceramente que será importante desenvolver modelos e uma infraestrutura de IA usando uma abordagem modular que empregue os melhores casos de uso e técnicas em toda a IA e, além disso, otimize processos, tais como servir e combinar diferentes modelos menores em diversas modalidades de dados. Portanto, eu aconselharia as startups de IA a adotarem essa abordagem ao desenvolver seus produtos e serviços. Também recomendaria investir na prontidão dos dados com limpeza, curadoria e hidratação. Isso ajudará você a estar mais preparado para adotar a IA rapidamente para seus processos de negócios e aproveitar ao máximo os modelos que usa.

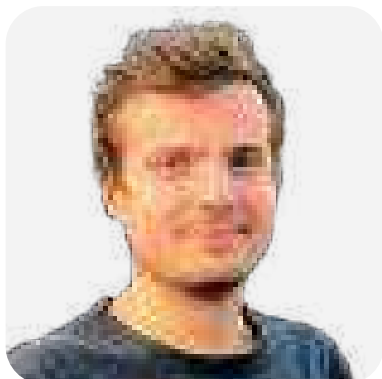
A necessidade de melhores ferramentas de explicabilidade

Gostaria que as ferramentas de explicabilidade estivessem mais avançadas. A forma como muitos desses modelos de IA generativa são treinados e depois operacionalizados é completamente desconectada. Uma vez que o modelo é treinado, ponderado e posteriormente hospedado para operacionalização, fica difícil voltar a alguns dados. Em certas empresas, os modelos foram treinados com dados que não deveriam ser usados, o que introduziu um risco em potencial. Seria ótimo ter uma maneira de rastrear parte da proveniência ou governança durante o treinamento. Há pesquisas sobre o uso de tecnologias semelhantes a livros-razão, e acho que tudo isso será importante, especialmente porque agora estamos vendo muitos casos de legalidade relacionados aos dados.





Rodrigo Scotti



Founder e CEO, Nama



Rodrigo Scotti é fundador e CEO da Nama, empresa pioneira em IA para Gestão do Conhecimento no Brasil. Com mais de uma década de experiência em IA, Scotti liderou a Nama na criação de uma plataforma no-code que combina busca semântica proprietária, RAG e modelos de linguagem para fornecer respostas precisas e personalizadas a partir de dados corporativos. Reconhecido pela Forbes 30 Under 30, Scotti é co-fundador da ABRIA (Associação Brasileira de IA) e um defensor do uso ético e responsável da IA.

Um olhar otimista sobre a IA e suas especificidades

Projetos podem favorecer a sociedade como um todo, com a certeza de que a responsabilidade é fundamental quando se lida com dados sensíveis

O começo de uma trajetória

Desde 2011, os estudos, as pesquisas e testes em torno de aplicações tecnológicas são parte crucial da minha jornada como empreendedor. Atualmente, o principal produto da Nama é focado em ser um buscador enterprise com algoritmos de retrieval, que são algoritmos de recuperação da informação. Por isso, costumamos comparar o que fazemos a “achar a agulha no palheiro”. A busca ocorre em planilhas, documentos, PDFs, apresentações, seja no Google Drive ou num documento específico.





Nesse caso, são milhões de palavras analisadas para gerar respostas que são referenciadas. Portanto, temos sistemas para mitigar alucinação de IA, conseguir referenciar informações e fazer um Deep Search, uma expansão de buscas dentro da base do cliente, para trazer retorno a partir disso.

Mas a minha atuação enquanto profissional na área de tecnologia também possui outra vertente. Sou um dos cofundadores da Abria, que é a Associação Brasileira de Inteligência Artificial. Desde 2019, o nosso trabalho passou a ser mais ativo, principalmente porque foi a primeira vez que se falou de regulação de IA.

O papel estratégico que esta grande ferramenta pode representar para as nações é, de fato, significativo, e eu percebo que o Brasil é um terreno extremamente fértil para progredirmos a partir disso.

Características únicas e desafios do Brasil abrem espaço para a inovação

A IA tem sido utilizada para resolver parte de alguns problemas estruturais do país. Diversos projetos vêm sendo encabeçados com base neste conhecimento amplo, a respeito das facilidades propiciadas pela solução tecnológica.

Um exemplo é a área da saúde, onde a Nama tem atuado em parceria com a Unimed e o Portal Telemedicina. Através da nossa plataforma, a Unimed conseguiu agilizar a busca por informações em contratos de saúde, um processo antes manual e demorado.

Além disso, o Portal Telemedicina utiliza a IA da Nama para analisar grandes volumes de dados de saúde, permitindo a identificação de padrões e a otimização do uso de recursos no sistema de saúde.

Por outro lado, no setor da educação, vejo que a IA pode ser uma ferramenta poderosa para democratizar o acesso ao conhecimento. Por exemplo, a tecnologia pode ser usada para criar plataformas de aprendizado personalizadas, que se adaptam às necessidades de cada aluno. E, mesmo em locais com conectividade limitada à internet, é possível disponibilizar conteúdo educacional relevante por meio de assistentes virtuais inteligentes, que funcionam como tutores personalizados, acessíveis via interfaces mais leves, como chatbots. Isso garante que o acesso à informação e ao aprendizado não seja restrito pela qualidade da conexão.

No fim das contas eu acho que, dessa forma, os desafios se entrelaçam e a IA chega para resolver problemas de diferentes dimensões. Sob esta perspectiva, o principal é que se criem os alicerces mais adequados, já que a necessidade de captação de informações é constante e contínua.





Parcerias no âmbito social e a sua importância no ecossistema de IA

Nós estamos vivendo um momento geopolítico geral, global, que é muito peculiar. Passamos por incertezas, uma certa pressão, que indica caminhos para a sociedade ou até mesmo rupturas muito fortes. Eu acredito que a ideia de “utopia tecnológica como salvadora de todas as coisas” não é uma saída inteligente.

Os governos, no que se refere ao uso da tecnologia avançada, precisam ser auxiliados por uma espécie de colegiado, multifacetado, de várias frentes. As universidades, os acadêmicos, as startups e empresas de tecnologia, todos devem cooperar em conjunto, na fronteira do conhecimento, colocando em prática os aprendizados adquiridos ao longo do tempo para beneficiar as populações de diversas maneiras.

Se esta visão de cooperação não estiver bem consolidada, diferentes prioridades podem surgir naturalmente. É relevante considerar que organizações com amplo acesso a dados e recursos para desenvolver tecnologias de IA têm responsabilidades proporcionais às suas capacidades. Por este motivo, uma abordagem cuidadosa parece prudente. A responsabilidade deve ser um princípio orientador neste contexto.

Insights valiosos para as startups

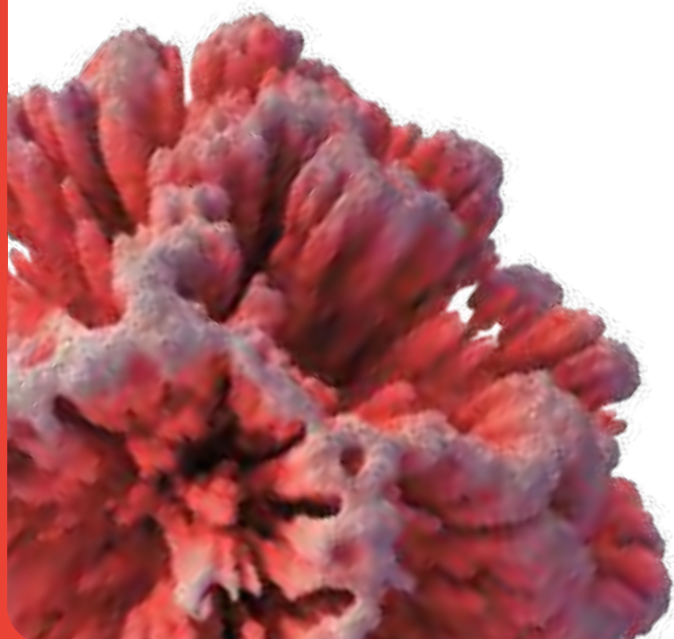
Assim, se eu pudesse dar um conselho para quem está começando e pretende expandir a sua atuação com inteligência artificial, seria algo em torno do entendimento dos dados. Compreender a origem das informações é fundamental — de onde vêm seus dados e o quão confiáveis eles são. Absorver o máximo de informações sobre o que está ao redor da caixa preta (um LLM, por exemplo) e, se possível, descobrir como funciona essa caixa preta, é ouro.

Se o intuito é ser uma referência em IA e construir soluções, somente com o máximo de conhecimento a pessoa saberá de que modo o modelo vai funcionar, e isso certamente irá direcionar melhor o output. Há muitas soluções no mercado. Contudo, se a sua empresa escalar e gerar dados únicos e próprios, ter a noção do melhor direcionamento da sua aplicação, compreendendo como ajustar e otimizar os modelos para sua necessidade específica, é algo que fará toda a diferença.

A verdadeira vantagem competitiva está nos seus dados proprietários e na aplicação inteligente deles. Identificar como as ferramentas operam permitirá a evolução do que você está construindo e a resolução de problemas reais de forma estratégica.



Compreender a origem das informações, saber ao máximo sobre o que está ao redor da caixa preta e, se possível, descobrir como funciona essa caixa preta, é ouro.





Yoav Shoham



**Professor emérito de ciência da computação, Universidade de Stanford.
Co-fundador, AI21 Labs**



Especialista em IA e fundador de três empresas de IA, adquiridas pela Ariba/SAP e Google (2x).



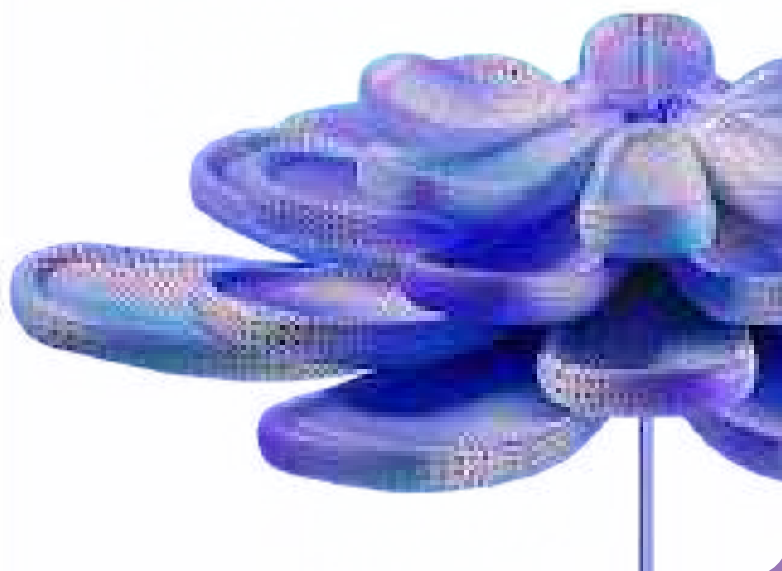
Ganhador de diversos prêmios acadêmicos e fellow da Associação para o Avanço da Inteligência Artificial, da Association for Computing Machinery e da Game Theory Society.

Chega de fornecer prompts e rezar pelo melhor: está na hora de construir sistemas completos de IA

Mas só você se importa com controlabilidade, confiabilidade e eficiência.

Para ser sincero, eu detesto o termo AGI, ou inteligência artificial geral. Isso não existe. Eu já vivi o bastante para saber que a inteligência é multifacetada. As máquinas, sem dúvida, serão capazes de automatizar cada vez mais funções, mas existe uma falsa noção de que essa mítica AGI será alcançada em um determinado momento. Acredito que esse pensamento vago e o hype em torno da AGI são uma distração.

Em vez disso, vamos falar sério sobre a tecnologia de IA, seus pontos fortes e fracos.





As empresas estão experimentando loucamente com IA, mas estão cautelosas na hora de implantá-la em escala.

É o que mais tem no mundo empresarial. Embora a adoção de IA pelo consumidor tenha se estabelecido um ritmo recorde, as empresas estão demorando para adotá-la. Com certeza, os CEOs e conselhos administrativos em todos os lugares estão apresentando suas empresas como “baseadas em IA” (ou planejando se tornar assim), e estão testando a tecnologia intensamente, às vezes com centenas de casos de uso.

Mas, apesar de toda a experimentação em massa que está acontecendo nas companhias, apenas uma fração dos projetos de IA realmente chega à implantação. Isso se resume a dois desafios principais:

01 Custo

O custo elevado para rodar LLMs dificulta a viabilidade econômica de softwares para empresas.

02 Os LLMs geram coisas incríveis, mas também fazem coisas sem sentido

Imagine que você está escrevendo um memorando de investimento ou respondendo a um cliente, e sua IA acerta 95% das vezes, mas erra feio nos outros 5%. Esse é um obstáculo intransponível na empresa.



Sistemas de IA vão superar amplamente os LLMs básicos em confiabilidade e eficiência, mas não serão perfeitos. Empreendedores inteligentes construirão produtos para aproveitar os pontos fortes desses sistemas, ao mesmo tempo em que levam em conta suas limitações.

O primeiro desafio é o custo inerente de fornecer (sem mencionar treinar) LLMs, e será resolvido por uma combinação de dois métodos. Um deles é usar LLMs menores (o termo Modelos de Linguagem Pequenos, ou SMLs, está ganhando popularidade), aqueles modelos “minúsculos” com menos de 7 bilhões e até menos de 3 bilhões de parâmetros. O outro método é usar arquiteturas diferentes, mais eficientes do que a arquitetura padrão de transformadores; a família de modelos Jamba da AI21 é um exemplo.

O segundo desafio é particularmente crítico e mais desafiador. O fenômeno é inerente à natureza probabilística dos LLMs e é ilusão acreditar que isso desaparecerá, não importa quanto esforço seja colocado em coisas como “alinhamento”, “proteções” e afins. Acredito que vamos nos afastar do que chamo de nosso atual modus operandi de “fornecer prompts e rezar pelo melhor”. O setor perceberá que os LLMs fazem parte de um sistema de IA mais abrangente, que pode integrar e aproveitar perfeitamente os pontos fortes de várias tecnologias de IA, incluindo LLMs, recuperação, ferramentas e outros códigos tradicionais. Os sistemas de IA oferecerão maior controle, eficiência e confiabilidade, especialmente ao lidar com tarefas que exigem raciocínio não trivial — algo que acontece com a maioria das tarefas.





Uma prática saudável para startups: use sistemas de IA, mas foque na adequação “produto-algoritmo”

Os sistemas de IA que combinam múltiplos LLMs e outras ferramentas oferecem uma solução atrativa. Eles permitem um melhor gerenciamento do custo e da computação, fazendo um roteamento inteligente das tarefas para os recursos mais adequados. Por exemplo, um LLM menor poderia atuar como um “roteador”, direcionando tarefas para LLMs ou soluções especializadas para otimizar a eficiência. Esses sistemas também podem aumentar a confiabilidade e a qualidade, incorporando mecanismos de verificação e balanço durante a computação.

Mas, ao aplicar esses sistemas de IA aos problemas do mundo real, você deve abordá-los com sabedoria. Meu conselho comum para líderes de startups de IA é buscar a “adequação produto-algoritmo”. O que quero dizer com isso é que, embora os sistemas de IA representem uma melhoria considerável em relação aos LLMs básicos em termos de confiabilidade e eficiência, eles ainda serão imperfeitos; a incerteza subjacente envolvida em chamadas de LLM, pesquisa e recuperação não desaparecerá completamente. Portanto, como um empreendedor criando um novo produto, entenda os pontos fortes e fracos da tecnologia e crie esse produto de uma forma que aproveite suas vantagens e compense suas imperfeições. É isso que chamo de “adequação produto-algoritmo”.





Crie o futuro com o Google Cloud



Hoje, mais de 60% de todas as startups de IA generativa financiadas globalmente e 90% dos unicórnios de IA generativa no mundo são clientes do Google Cloud. Temos orgulho de trazer nossa longa trajetória de liderança e inovação em IA para oferecer soluções avançadas de IA na nuvem que apoiam as oportunidades e desafios únicos de nossos clientes.

No Google Cloud, estamos colaborando com pesquisadores, fundadores, startups, empresas, parceiros e órgãos do setor público para refletir de forma crítica sobre como nossas soluções responsáveis de IA podem atender continuamente às necessidades de funcionários, clientes, pacientes e cidadãos. Isso inclui fornecer infraestrutura de excelência e capacidades full-stack que estão na vanguarda da inovação, firmar um engajamento forte com inventores de dados, agentes e aplicações para transformar novas ideias em realidade, e estabelecer parcerias com empresas para avaliar como os avanços da IA podem modernizar experiências dentro e fora das organizações. O ritmo vertiginoso de mudança impulsionado pela IA generativa significa que as startups estão enfrentando desafios sem precedentes – e estamos aqui para ajudar.

Darren Mowry

Managing Director, Global Startups,
Google Cloud





A IA permite um nível de personalização na estratégia de go-to-market que nunca foi possível antes — criando jornadas intuitivas e individualizadas para os clientes, analisando o comportamento dos usuários e oferecendo recomendações, mensagens e ofertas personalizadas em tempo real. O futuro trará uma produção criativa ágil, ativações sob medida e mensuração em tempo real, tudo funcionando em um sistema dinâmico e interconectado. E as startups terão mais tempo para focar na criatividade, na estratégia e na conexão autêntica com os clientes, que continuam sendo essenciais para impulsionar o engajamento.

Estamos animados para caminhar ao seu lado na vanguarda da IA generativa: este é um cenário que evolui constantemente, onde o que é verdade hoje pode não ser mais amanhã.

Alison Wagonfeld
CMO, Google Cloud



Não importa o ponto da sua jornada de adoção da IA: estamos aqui para ajudar.

Agende sua reunião sobre
IA generativa hoje mesmo.

→ Cadastre-se agora

Ganhe até US\$ 350 mil em créditos de nuvem
com o Google for Startups Cloud Program.

→ Inscreva-se agora

Entre em contato com nossa equipe
de vendas para Startups.

→ Fale conosco

