PALCO

CENTRE STAGE

DIA	LISBOA
10 NOV 2026	10:20

NO PROGRAMA The open source advantage

NEM TODA TAREFA PRECISA DA IA MAIS CARA

Dois laboratórios chineses de modelos abertos e um investidor mostram quando a IA aberta sai mais barata e onde o custo evitado reaparece.

Alan Shen
MINIMAX
MD Global, MiniMax

Dahai Li
MODELBEST
Cofundador e CEO da ModelBest (Mianbi, em chinês), empresa de Pequim criada em agosto de 2022 como spin-off do laboratório de...

Rohit Sharma
TRUE VENTURES
Sócio da True Ventures, gestora americana de venture capital de estágio inicial, onde entrou em 2012.

Reya El-Salahi
CGTN EUROPE
Jornalista e apresentadora britânica baseada em Londres, mediadora da sessão.

EM 30 SEGUNDOS

1. Alan Shen, da MiniMax: o trabalho comum de IA roda em quase qualquer modelo, e falta uma triagem que mande cada tarefa ao modelo certo.
2. Dahai Li, da ModelBest: o modelo pequeno vence quando cabe na máquina do cliente, e rodar em casa compensa com volume estável ou tarefa estreita.
3. Rohit Sharma, da True Ventures: a abertura deve vencer, como nas redes e nos celulares, mas pode faltar computação para quem não comprou antes.

Cerca de 95% das tarefas de uma empresa não precisam do modelo de IA mais avançado: escrever e-mails e apoiar o trabalho do dia a dia cabem em quase qualquer modelo. A estimativa é de Alan Shen, diretor-geral global da MiniMax. Se uma triagem mandar cada tarefa ao modelo certo, as empresas podem economizar milhões de dólares, diz ele. Shen dividiu o painel com Dahai Li, cofundador e CEO da ModelBest, e Rohit Sharma, sócio da True Ventures, e cada um mostrou um lado da mesma conta: quando a IA aberta sai mais barata e o que custa operá-la.

01

CADA TAREFA PEDE UM MODELO DIFERENTE

O que falta às empresas, segundo Shen, é essa camada de triagem: um sistema que reconheça o tipo de tarefa antes de chamar a IA. Ele vê aí espaço para novas empresas já no próximo ano.

Rohit Sharma desenhou a divisão para uma empresa de 500 a 1.000 pessoas. O modelo mais avançado, chamado de fronteira, fica com as tarefas mais difíceis e valiosas. No meio entram modelos abertos, ajustados sem parar com os dados da própria empresa. E, em três a cinco anos, modelos muito pequenos vão rodar no laptop, no relógio e no celular.

Os modelos fechados não saem de cena, avisou Sharma: um Walmart ou um JP Morgan sempre terá trabalho que exige o topo de linha. A novidade está em outro ponto. Na semana anterior ao painel, contou, ficou possível gerar de uma vez um modelo pequeno a partir de um grande, feito para uma tarefa só, e descartá-lo depois de algumas semanas de uso.

A TRIAGEM MANDA CADA TAREFA AO MODELO CERTO

CAMADA DE TRIAGEM

- **Modelo de fronteira** — Só o trabalho difícil e valioso
- **Modelos abertos** — Ajustados sempre com dados da casa
- **Modelos minúsculos** — Laptop, relógio e celular em 3 a 5 anos
- **Modelos descartáveis** — Gerados de um grande para uma tarefa só
- **Tarefas comuns** — E-mails e rotina cabem em quase qualquer modelo

02

RODAR EM CASA COMPENSA EM DOIS CASOS

Dahai Li explicou quando fecha a conta de rodar o modelo em casa. O primeiro caso é o uso diário grande e estável: a nuvem cobra cada consulta, e o servidor próprio só compensa se ficar ocupado o tempo todo. O segundo é a tarefa interna e estreita, que um modelo aberto pequeno, ajustado com os dados da empresa, faz por uma fração do custo.

Muitas vezes, porém, quem decide é a privacidade, não o preço. Se a regra proíbe mandar dado sensível para fora, rodar em casa é a única opção, custe o que custar. Li recomenda misturar os dois: o modelo local cuida das tarefas frequentes e sensíveis, e a nuvem, dos problemas que pedem conhecimento amplo.

Para Li, maior nem sempre é melhor: as empresas olham qualidade, velocidade, custo e controle dos dados. O caso que ele contou é de 2024. Um cliente testou quatro projetos de IA ao mesmo tempo e, para proteger os dados, não quis usar a nuvem. Só o projeto da ModelBest chegou à produção. As outras equipes escolheram modelos grandes demais para as máquinas do cliente; o modelo de 7 bilhões de parâmetros da ModelBest coube nelas.

NÚMEROS DA PALESTRA

95%
das tarefas de uma empresa não precisam do modelo de IA mais avançado

6 de 10
modelos do ranking de vídeo derivam do modelo aberto da MiniMax

1 de 4
projetos de IA de um cliente chegou à produção em 2024

300 mi
usuários da MiniMax no mundo

03

ABRIR O MODELO TRANSFORMA A COMUNIDADE EM LABORATÓRIO

Para Shen, publicar o modelo para que qualquer um o baixe e ajuste (o chamado modelo de pesos abertos) cria o melhor retorno possível. Desenvolvedores de fora treinam em cima dele e **fazem coisas que os engenheiros da casa não tinham imaginado**.

A prova que ele deu: seis dos dez modelos do ranking de geração de vídeo da Artificial Analysis foram treinados sobre o modelo de vídeo que a MiniMax abriu. E se um concorrente copiar tudo e cobrar menos? Para Shen, a defesa não é o segredo: é inovar mais rápido que os outros.

Dahai Li comparou o risco de uso indevido ao de um carro, que a montadora protege com cinto, airbag e freio. Do mesmo jeito, a ModelBest limpa os dados de treino e ensina o modelo a recusar pedidos perigosos. E milhares de pesquisadores de fora encontram falhas mais rápido do que uma empresa sozinha.

> **Nem todo token é igual, e precisamos reconhecer isso.**
>
> ROHIT SHARMA

04

O CUSTO VOLTA EM OUTRA LINHA, E A COMPUTAÇÃO PODE FALTAR

Sharma foi o contrapeso do painel. A queda inicial de custo convence as empresas a adotar modelos abertos, mas **a conta pode voltar como gasto com engenharia e segurança**: testar se o modelo faz o trabalho, instalá-lo dentro de casa e garantir que as regras de dados valem onde ele roda.

Mesmo assim, a história joga a favor da abertura, segundo ele. Ericsson, Nokia e BlackBerry começaram com sistemas fechados, e Cisco e Juniper chegaram a ter 99% das redes corporativas. Hoje há opção aberta em cada parte desses mercados, e ele espera o mesmo na IA.

O medo dele para os próximos três a cinco anos é faltar máquina. OpenAI, Anthropic e os grandes provedores de nuvem compraram boa parte da capacidade de computação, com encomendas de milhares de chips à Nvidia. A empresa que quiser rodar modelo aberto em grande escala pode não ter onde.

Dahai Li fechou com uma aposta. A ModelBest previa a onda da inteligência artificial geral para por volta de 2030, e ela parece vir antes. Em dois ou três anos, disse, modelos pequenos, abertos ou fechados, serão muito poderosos. "É fácil prever o próximo token, mas é difícil prever o que acontece em três ou cinco anos", ressalvou.

O MODELO ABERTO ECONOMIZA NA ENTRADA E COBRA DEPOIS

O QUE ATRAI	O QUE REAPARECE
Custo inicial menor — A queda de preço puxa a adoção	**Engenharia** — Testar e instalar o modelo dentro de casa
Dados sob controle — Nada sensível sai de casa	**Segurança** — Garantir as regras de dados onde ele roda
Mercado mais aberto — Como em redes e telecom, surgem opções em cada parte	**Falta de máquina** — Chips já reservados pelos grandes compradores

PARA LEVAR

1. **Antes de contratar IA, separe as tarefas: só as difíceis e valiosas justificam o modelo mais caro.**
2. **Servidor próprio só sai barato se trabalhar o dia todo; parado, custa mais que pagar por uso.**
3. **Comece pelo dado que não pode sair da empresa: ele decide onde o modelo roda antes do preço.**

TRÊS ÂNGULOS PARA AS EMPRESAS

ESTRATÉGIA
Contrato longo com um único fornecedor de modelo parte da ideia de que toda tarefa pede o mesmo nível de inteligência. O painel descreveu o contrário: um conjunto de modelos, revisto com frequência, em que alguns duram poucas semanas.

OPERAÇÃO
O modelo aberto passa trabalho do fornecedor para dentro da empresa: testar, ajustar, proteger e manter. Sem equipe técnica para isso, a economia fica no papel.

TECNOLOGIA
A máquina que a empresa já tem limita o modelo que ela pode usar. A capacidade de computação disponível, dentro de casa ou alugada, entra como critério de escolha desde o primeiro dia.

CONTRAPONTO

Dois dos três palestrantes vendem modelos abertos, e nenhum mostrou a conta completa de um cliente.

O caso da ModelBest não tem nome nem valor. Os "milhões de dólares" de economia vieram de manchetes citadas por Alan Shen, e os 95% são estimativa dele, sem fonte.

A liderança dos modelos abertos em uso na OpenRouter também veio sem número. O alerta de Rohit Sharma sobre falta de computação é preocupação de investidor, não medição, e o painel não calculou o custo de engenharia e segurança que, segundo ele, acompanha a adoção.

LEITURA FFX

O entusiasmo do palco pede um desconto.

A Menlo Ventures (dezembro de 2025) mediu que a fatia dos modelos abertos no uso das empresas caiu de 19% para 11%. Isso confirma a ressalva do próprio Alan Shen: o ranking da OpenRouter, onde os abertos lideram, mede sobretudo consumidores e profissionais avulsos, não as grandes empresas. No Brasil, o iFood mostra quando a conta de ter modelo próprio fecha: o modelo especializado em comércio que ele criou com a Prosus custa 60 vezes menos que referências internacionais e já responde por 90% das interações de IA generativa da plataforma interna da empresa (TI Inside, março de 2026).

É o perfil que Dahai Li descreveu: volume enorme, estável e tarefa específica. Para quem não tem esse volume nem equipe para operar o modelo, pagar por uso na nuvem continua sendo o caminho mais barato.

RETRATO DA EMPRESA

MINIMAX, MODELBEST E TRUE VENTURES

As três empresas do palco: dois laboratórios chineses que publicam modelos abertos e um fundo que investe em startups desde o primeiro cheque.

MiniMax: laboratório chinês de modelos de IA, fundado no fim de 2021, aberto e com vários formatos (como texto e vídeo) desde o início; a maior parte da receita vem de empresas (Alan Shen, no palco)

ModelBest: nascida na Universidade Tsinghua, em Pequim, faz a família aberta MiniCPM de modelos pequenos que rodam no próprio aparelho; captou mais de US$ 737 milhões no primeiro semestre de 2026, com avaliação acima de US$ 3 bilhões (KrASIA, julho de 2026)

True Ventures: fundo que investe em startups em estágio inicial, com primeiro cheque de US$ 1 milhão a US$ 6 milhões (Rohit Sharma, no palco)

PROVOCAÇÃO FFX

SE NEM TODA TAREFA VALE O MESMO, POR QUE A SUA EMPRESA PAGARIA O MESMO PREÇO DE IA POR TODAS?

FONTES

1. KrASIA: ModelBest's on-device AI model is coming to Samsung smartphones (21/07/2026)
2. Gasgoo: WAIC 2026, ModelBest unveils embodied AI model series MiniCPM-Robot
3. 36Kr: cobertura da Densing Law na Nature Machine Intelligence
4. Menlo Ventures: 2025, The State of Generative AI in the Enterprise (09/12/2025)
5. TI Inside: iFood alcança 12 milhões de pedidos em fevereiro impulsionados por LLM proprietário (19/03/2026)
6. Prosus: Prosus reinvents ecommerce with launch of a specialist Large Commerce Model (10/10/2025)
7. True Ventures: Rohit Sharma (bio)