This document was translated from English by machine and may contain mistakes. The English original is at https://www.novaracg.com/2026/09/14/the-ai-ceos-found-the-brakes-sign-language-ai-should-pay-attention/.

Os CEOs de IA Encontraram os Travões... A IA de Língua Gestual Deveria Prestar Atenção

Heather Grizzle · 14 de setembro de 2026

Frontier AI CEOs are now asking for testing, audits and incident reporting. Sign language AI should hold itself to the same brakes.

Conteúdo

1. A vanguarda está a pedir travões
2. A dívida de governação é real
3. O que a vanguarda está a aprender, a IA de língua gestual pode aprender agora
4. "We tested it" is not governance
5. Não espere pelo seu próprio momento de abrandamento
6. A garantia independente não é um ataque à inovação
7. Ainda há uma oportunidade de se antecipar a isto
8. O aviso

A vanguarda está a pedir travões

Durante anos, a narrativa dominante na inteligência artificial foi a velocidade. Construir o modelo mais capaz. Lançar o agente. Vencer o concorrente. Angariar capital. Expandir o padrão de referência. Colocar o produto no mercado. A governação podia recuperar o atraso mais tarde.

Parece que esse “mais tarde” chegou.

On September 9, OpenAI chief global affairs officer Chris Lehane published a call for mandatory, capability-based national AI safety regulation covering testing standards,

independent assessments, cybersecurity protections and incident reporting, alongside the company's endorsement of four California bills, including one establishing infrastructure for independent safety assessments and another setting standards for AI auditors.[^1] Three days later, Anthropic CEO Dario Amodei went considerably further. In an essay arguing that the industry must deliberately pace frontier development, he proposed embedding third-party evaluators inside frontier labs with near-employee access so they could independently examine safety measures, incidents and model behavior.

Depois aconteceu algo invulgar.

Within hours, OpenAI CEO Sam Altman publicly agreed that the frontier needed pacing, and told Fortune days later that his company would not pursue a public offering this year because it had too much safety work left to do. Elon Musk, who had pointedly declined to join earlier industry calls for restraint, quoted the essay approvingly. Demis Hassabis, now chair of Google DeepMind and chief scientist of Alphabet, wrote that the essay pointed toward the right path forward. Companies locked in one of the most consequential technology races in history converged, inside a single day, on the proposition that capability could no longer be allowed to outrun control. Axios ran the moment under a headline that needed no elaboration: "AI's most powerful CEOs hit the brakes."

Today, September 14, Microsoft added another piece. Microsoft AI CEO Mustafa Suleyman released a draft code of conduct that he describes as a constitution of sorts for the company's future models. The draft would require Microsoft's systems to communicate clearly with humans, respect human boundaries, remain controllable, accept correction and shutdown, and treat behavior that violates the code as a model failure rather than an acceptable expression of autonomous judgment. Microsoft assembled it after expert consultations and has opened it to roughly six weeks of public feedback.[^3] Suleyman called it a warning shot.

Essa sequência devia ser leitura obrigatória para todas as empresas que desenvolvem IA de língua gestual. Não porque um avatar de gestos esteja prestes a tomar conta da internet. Porque é assim que se parece a dívida de governação.

A dívida de governação é real

As empresas de tecnologia compreendem a dívida técnica. Construir depressa demais, adiar decisões de arquitetura suficientes, e eventualmente os atalhos tornam-se dispendiosos.

A dívida de governação funciona da mesma forma, e acumula-se através de decisões que, isoladamente, parecem cada uma razoável. Uma empresa lança um produto antes de definir quem é responsável quando o sistema falha. Publica uma alegação de precisão antes de estabelecer o que conta como erro. Constrói um padrão de referência antes de perguntar se esse padrão representa as pessoas que efetivamente utilizarão o produto. Valida internamente porque a avaliação independente é incómoda. Acrescenta supervisão humana sem especificar o que a pessoa deve notar, quando essa pessoa deve intervir, ou se possui a especialização para sequer reconhecer a falha. Nomeia um grupo consultivo depois de as decisões importantes do produto já terem sido tomadas. Chama acessível a um sistema porque o sistema produz sinais.

Cada uma dessas escolhas cria dívida de governação. E, ao contrário da dívida técnica, o custo não é suportado apenas pelo criador. Os utilizadores suportam-no. Os compradores suportam-no. As agências públicas suportam-no. Na tecnologia de língua gestual, as pessoas Deaf suportam-no.

O que a vanguarda está a aprender, a IA de língua gestual pode aprender agora

Há um facto extraordinário no centro do debate atual sobre IA. As empresas com algumas das maiores equipas de engenharia, orçamentos de investigação e organizações de segurança do mundo estão a pedir mais escrutínio externo. A Anthropic está a propor avaliação incorporada de terceiros. A OpenAI está a apoiar regulação federal obrigatória e avaliação independente. A Microsoft está a tentar formalizar os limites do controlo humano. Executivos rivais estão a discutir contenção coordenada em público, algo que até há pouco tempo seria impensável.

A lição não é que as empresas devem esperar até que um governo escreva uma norma perfeita. É quase o oposto. As organizações mais próximas da tecnologia estão a admitir que a confiança interna não é prova suficiente.

Isso importa para a IA de língua gestual porque o campo ainda é jovem o suficiente para escolher uma sequência diferente. A governação não tem de ser o trabalho de reparação. Pode fazer parte da arquitetura.

"We tested it" is not governance

A IA de língua gestual tem uma versão particularmente perigosa do problema da validação. Um sistema pode produzir algo reconhecivelmente semelhante a gestos e, ainda assim, falhar à pessoa que dele depende. Uma tradução pode ser fluente e errada. Um avatar pode estar visualmente polido enquanto introduz distorção linguística. Um padrão de referência pode reportar um número impressionante enquanto representa um conjunto de dados restrito, um ambiente de sinalização limitado, ou uma tarefa que pouco se assemelha à implementação real. Um estudo pode incluir validação humana enquanto se baseia numa amostra demasiado pequena para sustentar a alegação comercial eventualmente associada a ela. E uma demonstração de produto pode funcionar lindamente sem estabelecer que o sistema é adequado para cuidados de saúde, educação, emprego, comunicação de emergência ou serviços governamentais.

Estas não são simplesmente questões de aprendizagem automática. São questões de garantia. Quem o testou, e quem selecionou o teste? Que sinalizantes foram representados, e quais não foram? O que foi medido, e o que foi excluído? O que conta como erro material? Que contextos de implementação foram avaliados? Quem detém a autoridade para interromper a implementação? O que acontece quando o desempenho se degrada, e quem revê os incidentes quando isso acontece? E, para uma tecnologia construída em torno da comunicação Deaf, onde é que a autoridade Deaf efetivamente entra na decisão?

Estas perguntas tornam-se consideravelmente mais difíceis de responder depois de os contratos serem assinados, as alegações de marketing publicadas e os clientes informados de que a tecnologia está pronta.

Não espere pelo seu próprio momento de abrandamento

There is an understandable temptation for smaller AI companies to look at the frontier governance debate and file it under somebody else's problem. Different technology, different scale, different risks. All correct. But risk magnitude is not the relevant comparison. Governance timing is.

A indústria de vanguarda correu em frente sob uma pressão competitiva extraordinária, e os seus líderes enfrentam agora questões sobre avaliação independente, testes obrigatórios, comunicação de incidentes, controlo humano, supervisão governamental e se a indústria pode ser confiada para julgar quando os seus próprios sistemas são suficientemente seguros. Os criadores de IA de língua gestual podem assistir a isto a desenrolar-se em tempo real, e depois decidir se o repetem.

Uma empresa que desenvolve tecnologia de língua gestual hoje já deveria conseguir responder a cinco perguntas.

1. Que evidência nos levaria a não implementar o nosso próprio produto?
2. Quem, fora da empresa, tem permissão para contestar as nossas alegações?
3. Que autoridade detêm as pessoas Deaf para além de fornecer feedback ou dados de treino?
4. Que utilizações da nossa tecnologia consideramos inadequadas, mesmo que um cliente esteja disposto a pagar por elas?
5. O que acontece quando a evidência pós-implementação contradiz a evidência que usámos para vender o sistema?

Se essas respostas não existirem, a empresa não tem apenas documentação de governação inacabada. Tem dívida de governação.

A garantia independente não é um ataque à inovação

É aqui que o mercado de IA de língua gestual precisa de amadurecer rapidamente. O escrutínio independente não é hostilidade em relação à tecnologia. É um sinal de que a tecnologia é suficientemente importante para ser escrutinada.

Os fabricantes de aeronaves não estabelecem confiança convidando os compradores a admirar a aeronave. Os fabricantes de dispositivos médicos não estabelecem segurança publicando testemunhos. As auditorias financeiras não permitem que o entusiasmo da gestão substitua a evidência. À medida que a IA avança para a acessibilidade, aplica-se a mesma distinção. Um criador cria o sistema. A governação estabelece as regras à sua volta. A avaliação testa a evidência. A garantia pergunta se essa evidência é suficiente para sustentar a alegação que está a ser feita. São quatro funções diferentes, e

quando uma única empresa controla as quatro, os compradores devem fazer perguntas.

The frontier industry is beginning to acknowledge exactly that problem. Anthropic's proposal for embedded outside evaluators is significant precisely because it moves independent scrutiny closer to the development process rather than treating evaluation as a ceremonial checkpoint immediately before release.

A IA de língua gestual deveria ir ainda mais longe. A avaliação técnica independente é necessária, mas a legitimidade linguística, o contexto de implementação, o impacto na acessibilidade e a governação Deaf não podem ser reduzidos a testes de engenharia. Um modelo pode funcionar exatamente como foi concebido e, ainda assim, ser inadequado para a tarefa para a qual foi vendido.

Ainda há uma oportunidade de se antecipar a isto

É esta a parte que as empresas de IA de língua gestual não devem perder. Estão a receber um aviso invulgarmente valioso, e não têm de viver primeiro a crise de governação.

O trabalho é específico. Documente as limitações antes que o marketing escreva em torno delas. Separe o desempenho de demonstração da evidência de implantação. Use avaliadores independentes antes que um cliente os exija. Crie procedimentos de incidentes antes do primeiro incidente grave. Estabeleça os contextos em que o produto não será vendido. Dê aos especialistas Deaf autoridade de decisão real, e não apenas presença consultiva. Faça com que as alegações de validação sejam específicas o suficiente para que outra parte possa examiná-las e contestá-las. Preserve as evidências para que um comprador possa, mais tarde, determinar por que uma decisão de implantação foi justificada no momento em que foi tomada.

É assim que se parece tornar-se passível de garantia (assurable). É também consideravelmente mais barato do que reconstruir a confiança mais tarde.

O aviso

As empresas de IA mais poderosas do mundo estão descobrindo que a governança não pode permanecer permanentemente atrás da capacidade. Alguns de seus líderes agora pedem freios. Outros pedem leis. Outros estão redigindo constituições, estruturas de avaliação externa e novos mecanismos de controle. Elas comandam recursos

que a maioria das empresas de tecnologia de acessibilidade nunca possuirá, e mesmo elas estão descobrindo o quão difícil a governança se torna quando a tecnologia já está se movendo a toda velocidade.

Sign language AI companies should study that carefully. Do not wait until a failed deployment, a procurement challenge, a regulator, a lawsuit or the Deaf community forces the governance conversation. Do not build the evidence after making the claim. Do not invite independent scrutiny only after trust has been lost. And do not confuse being first to market with being ready for the responsibility that comes with entering someone else's language, communication and access.

A IA de fronteira está agora tentando encontrar seus freios. A IA de língua de sinais ainda tem a oportunidade de instalá-los antes de acelerar.

NOVARA CONSULTING GROUP

A Novara Consulting Group oferece garantia de IA, governança de acessibilidade e avaliação independente.

FOOTNOTES:
Chris Lehane, "The AI policy window is open. We need to act.," OpenAI, September 9, 2026; Reuters, "OpenAI pushes for mandatory national AI safety requirements," September 9, 2026.
Axios, "AI's most powerful CEOs hit the brakes," September 13, 2026.
Jeffrey Dastin, "Microsoft drafts code of conduct to keep its AI under human control," Reuters, September 14, 2026.

> **TIPS**
>
> Know something about a sign language AI system that the public should know? Send it to tips@novaracg.com. Every tip is read by me and kept confidential. If you work in the UK, writing to us is not by itself a protected disclosure under the Public Interest Disclosure Act 1998, so before you share employer information, check your options with Protect (protect-advice.org.uk), the free whistleblowing advice line.

SUPPORT NCG

NCG does this work without vendor money, so our findings stay ours. If it has been useful to you, you can support NCG's independent work here. NCG is a for-profit company, not a charity. Contributions are not tax-deductible and buy no service or influence.

Como citar

Grizzle, H. M. (2026, September 14). Os CEOs de IA Encontraram os Travões... A IA de Língua Gestual Deveria Prestar Atenção. Novara Consulting Group. https://doi.org/10.5281/zenodo.23003722