NOVARA
CONSULTING GROUP

Este documento foi traduzido do inglês por máquina e pode conter erros. O original em inglês está em https://www.novaracg.com/2026/08/17/ai-can-translate-but-who-translates-the-risk/.

A IA Consegue Traduzir. Mas Quem Traduz o Risco?

Heather M. Grizzle · 17 de agosto de 2026

AI can translate, but who translates the risk? The AI accountability gap, procurement standards, and why verification falls on the person least able to check.

Conteúdo

1. O que é a lacuna de responsabilização da IA?
2. Capacidade, fiabilidade e responsabilização são três tipos diferentes de evidência
3. Por que motivo as leis de divulgação não fecham a lacuna
4. Assimetria de verificação: o problema estrutural subjacente
5. O direito da acessibilidade já respondeu em parte a isto, e estamos a desaprendê-lo
6. Por que motivo a demonstração do fornecedor não pode ser o registo de governação
7. O que as instituições devem exigir antes da implementação
8. A próxima etapa da IA é a verificação
9. Perguntas frequentes
10. Fontes

Todas as implementações de IA transferem risco para alguém. Na maioria dos casos, esse risco recai sobre a pessoa com menor capacidade de detetar um erro e menor autoridade para o contestar.
Heather M. Grizzle Novara Consulting Group 17 de agosto de 2026

No início deste mês, escrevi sobre a investigação da Google DeepMind em tradução de língua gestual e argumentei que uma demonstração não é o mesmo que evidência. A

resposta ensinou-me algo útil. Quase ninguém discordou de que a distinção importa. Em vez disso, muitas pessoas fizeram a pergunta óbvia seguinte: se uma demonstração não é suficiente, quem deve decidir quando um sistema está suficientemente bom para ser implementado, e quem responde por isso quando não está?

Essa pergunta não é específica da língua gestual. É o problema central por resolver na governação da IA neste momento, e tem um nome que vale a pena usar sem rodeios. Chamemos-lhe a lacuna de responsabilização.

O que é a lacuna de responsabilização da IA?

A lacuna de responsabilização da IA é a distância entre aquilo que um sistema é capaz de fazer e aquilo pelo qual qualquer parte identificável responde quando o faz mal. Abre-se sempre que uma organização implementa um sistema de IA mais depressa do que constrói os meios para verificar os resultados do sistema, ouvir as pessoas afetadas por esses resultados, e corrigir ou interromper a implementação quando o resultado está errado.

A lacuna não é, sobretudo, um problema técnico. Os modelos falham de formas previsíveis, e os engenheiros costumam ser francos sobre esses modos de falha. A lacuna é um problema institucional. As organizações estão a adquirir sistemas cujos resultados não conseguem avaliar, a implementá-los em contextos onde os erros têm consequências, e a não manter nenhum mecanismo prático para descobrir se o sistema está a funcionar.

Capacidade, fiabilidade e responsabilização são três tipos diferentes de evidência

A maioria das conversas de aquisição trata estas questões como se fossem uma só. Não são, e separá-las é o passo mais útil que um comprador institucional pode dar.

Capability evidence shows that a system can perform a task under conditions the vendor selected. Demonstrations, benchmark scores, and pilot footage are capability evidence. They answer the question "is this possible."

Reliability evidence shows how the system performs across the range of people, settings, and conditions the buyer will actually encounter, including the ones the vendor did not choose. Disaggregated performance data, independent testing, and error

rates from live deployment are reliability evidence. They answer the question "how often is this wrong, and for whom."

Accountability evidence shows what happens when the system fails. It names who monitors performance, who receives complaints, who has authority to suspend the deployment, what the affected person is owed, and how any of that is enforced. It answers the question "who is responsible."

Os fornecedores competem em evidência de capacidade porque é a mais barata de produzir e a mais persuasiva de observar. A evidência de fiabilidade é dispendiosa e pouco lisonjeira. A evidência de responsabilização não cabe verdadeiramente ao fornecedor produzir, porque descreve a própria governação do comprador, e a maioria dos compradores ainda não a construiu.

Uma organização que aceita a evidência de capacidade como se resolvesse as outras duas questões não tomou uma decisão de aquisição. Fez uma suposição e pagou por ela.

Por que motivo as leis de divulgação não fecham a lacuna

A resposta regulatória mais visível ao risco da IA neste momento é a transparência. Ao abrigo do Artigo 50.º do Regulamento de IA da UE (Regulamento sobre Inteligência Artificial da União Europeia), as obrigações que se tornaram aplicáveis a 2 de agosto de 2026 exigem que os fornecedores marquem o conteúdo sintético em formato legível por máquina e exigem que os implementadores divulguem deepfakes e determinado texto gerado por IA sobre assuntos de interesse público, sob pena de sanções que podem atingir 15 milhões de euros ou três por cento do volume de negócios anual global total. A Comissão Europeia concedeu um período de carência até 2 de dezembro de 2026 para os sistemas generativos já colocados no mercado.

Este é um trabalho significativo e apoio-o. Vale também a pena ser preciso quanto ao que ele realmente faz. O Artigo 50.º é um regime de proveniência. Diz a uma pessoa que uma máquina esteve envolvida. Não lhe diz se a máquina acertou.

Para uma grande parte dos conteúdos sintéticos, a proveniência é toda a questão. Se a preocupação é um vídeo fabricado de uma figura pública, saber que o vídeo é sintético resolve-a. Mas, para os sistemas de IA que medeiam a comunicação, fornecem informação ou influenciam uma decisão sobre uma pessoa, a proveniência é a metade fácil.

Um rótulo que diga "esta tradução foi gerada por IA" deixa o destinatário exatamente no mesmo ponto onde começou, ou seja, incapaz de saber se o conteúdo é exato.

A metade da responsabilidade civil da abordagem europeia moveu-se na direção oposta. A proposta de Diretiva sobre Responsabilidade em matéria de IA da Comissão, que teria aliviado o ónus da prova para as pessoas lesadas por sistemas de IA, foi formalmente retirada em outubro de 2025. A transparência avançou. A reparação não.

O padrão repete-se nos Estados Unidos ao nível estadual. O Colorado aprovou a primeira lei estadual abrangente sobre IA em 2024, tendo-a depois revogado e substituído pela SB 26-189, assinada em 14 de maio de 2026, cujas obrigações substantivas para desenvolvedores e implementadores se aplicam a partir de 1 de janeiro de 2027. A lei de substituição eliminou os programas obrigatórios de gestão de risco, as avaliações de impacto algorítmico e o dever autónomo de diligência razoável contra a discriminação algorítmica. O que sobreviveu foi o aviso prévio à utilização, uma divulgação em linguagem simples no prazo de trinta dias após uma decisão adversa, a retenção de registos por três anos e o direito de solicitar uma revisão humana significativa.

Não estou a argumentar que essas obrigações sobreviventes sejam inúteis. O aviso e a retenção de registos são reais. Mas o aviso diz-lhe que um sistema foi usado. Os registos dizem-lhe o que ele fez. Nenhum dos dois estabelece que o resultado foi exato, e nenhum atribui responsabilidade pelas consequências caso não tenha sido. A direção do percurso regulatório ao longo dos últimos dois anos tem sido no sentido de informar as pessoas de que a IA esteve envolvida, e no sentido oposto ao de obrigar alguém a responder pelo que produziu.

Assimetria de verificação: o problema estrutural subjacente

Aqui está a parte que considero pouco examinada, e é onde o caso da língua gestual deixa de ser um exemplo de nicho e passa a ser a ilustração mais clara disponível de uma condição geral.

Na maioria das implementações de IA, a pessoa mais bem posicionada para detetar um erro não tem autoridade para agir sobre ele, e a parte com autoridade para agir não tem meios de o detetar. Chamaria a isto assimetria de verificação, e é o mecanismo que produz a lacuna de responsabilização.

Considere o que uma instituição ouvinte vê quando implementa um sistema automatizado de língua gestual. Vê que o sistema produziu um resultado. O pessoal observa um avatar a gesticular ou legendas a aparecer. Desse ponto de vista, a transação parece completa.

Considere agora o que a pessoa Surda vivencia. Recebe um resultado que não consegue verificar em relação à fonte, entregue por um sistema que não escolheu, num contexto em que a instituição já concluiu que o acesso foi providenciado. Se a interpretação omitir uma negação, achatar um condicional, deslocar uma referência espacial ou omitir os marcadores não manuais que transportam significado gramatical, a pessoa poderá não saber. Se de facto notar que algo está errado, encontra-se na posição de contestar a própria conclusão da instituição de que a comunicação foi bem-sucedida.

Isto não é uma queixa sobre tecnologia. É uma distribuição de poder. A instituição detém a autoridade para determinar se o acesso ocorreu e carece da capacidade linguística para o avaliar. A pessoa Surda detém a capacidade linguística e carece da autoridade. O erro, a existir, não tem onde emergir.

Uma vez que se vê a forma disto, encontra-se-o praticamente em todo o lado. Um paciente não consegue avaliar uma recomendação de triagem assistida por IA. Um candidato não consegue inspecionar o modelo que o excluiu. Um requerente de benefícios não consegue auditar a determinação de elegibilidade. Uma pessoa que recebe instruções jurídicas traduzidas por máquina, seja em que língua for, não consegue verificar a tradução sem a fluência que a tradução existe precisamente para lhe fornecer. Em cada caso, a parte afetada absorve um risco que não consegue medir, e a instituição implementadora regista uma transação concluída.

O que torna a língua gestual um caso particularmente incisivo são as consequências associadas ao conteúdo. A linguagem em contextos institucionais transporta consentimento, testemunho, diagnóstico, instrução e efeito jurídico. Uma tradução incorreta não é uma experiência de utilizador degradada. É um formulário de consentimento defeituoso, um histórico médico impreciso, ou uma declaração deturpada num processo.

O direito da acessibilidade já respondeu em parte a isto, e estamos a desaprendê-lo

Já existe uma resposta viável na prática de acessibilidade, o que torna mais difícil desculpar o seu atual descuido.

Ao abrigo do regulamento do Departamento de Justiça constante do 28 CFR 35.160, uma entidade pública deve dar consideração primária ao meio ou serviço auxiliar solicitado pela pessoa com deficiência. A norma para os estabelecimentos privados de acesso público, prevista no 28 CFR 36.303, é mais fraca, exigindo consulta mas deixando a decisão final ao critério do prestador, e essa diferença é, em si mesma, esclarecedora quanto a onde a autoridade de verificação tende a erodir-se. As normas mínimas da National Association of the Deaf para interpretação por vídeo remoto em contextos médicos declaram diretamente o princípio subjacente: uma pessoa surda ou com dificuldades auditivas é quem melhor sabe qual o meio ou serviço auxiliar que alcançará uma comunicação eficaz. Essas mesmas normas vão mais longe e exigem que o intérprete por vídeo diga ao prestador para terminar a sessão remota e obter um intérprete presencial quando o intérprete determinar que a modalidade remota não está a produzir uma comunicação eficaz.

Lida como um desenho de governação, e não como uma regra de acessibilidade, esta abordagem está construída de forma invulgarmente sólida. Situa o juízo de verificação junto da pessoa que efetivamente o pode fazer. Dá a um profissional presente na sala a autoridade para interromper uma implementação a meio da utilização. Trata a eficácia como algo a determinar em contexto, e não como algo a presumir a partir da mera presença de equipamento.

Os sistemas automatizados tendem a eliminar ambas as características. Não existe um intérprete qualificado posicionado para pedir uma paragem, porque o intérprete é precisamente aquilo que o sistema substituiu. E o julgamento da pessoa afetada sobre se a comunicação funcionou é silenciosamente rebaixado de padrão legal a feedback de cliente.

A Federação Mundial de Surdos e a Associação Mundial de Intérpretes de Língua Gestual estabeleceram um limite para avatares gestuais em 2018. A sua posição era a de que os avatares são inadequados para informação em direto, complexa ou significativa, e que conteúdo estático pré-gravado pode ser aceitável quando pessoas Surdas aconselharam sobre o material e não é necessária interação em direto. Essa declara-

ção tem oito anos. A tecnologia avançou consideravelmente desde então. O limite de implementação que foi traçado não foi substituído por nada mais rigoroso, e continua a ser a linha mais clara disponível.

Por que motivo a demonstração do fornecedor não pode ser o registo de governação

A demonstração é concebida para ser persuasiva. Essa é a sua função, e não há nada de impróprio nisso. O problema é que, na ausência de uma avaliação independente, a demonstração torna-se, por defeito, a base probatória de uma decisão de aquisição, porque é o único documento presente na sala.

Uma demonstração estabelece a possibilidade em condições selecionadas. Não pode estabelecer o desempenho em toda a população que o comprador serve, o comportamento fora de condições controladas, a segurança em contextos de elevada consequência, nem a responsabilização quando ocorrem erros. Essas são propriedades de uma implementação, não propriedades de um modelo.

A política federal já reconhece isto em princípio. As orientações do OMB (Office of Management and Budget), emitidas em abril de 2025, definem IA de elevado impacto como sistemas cujo resultado serve como base principal para decisões ou ações com efeito legal, material, vinculativo ou significativo sobre direitos ou segurança, enumerando o memorando domínios que incluem direitos civis, acesso à educação, habitação, emprego, benefícios e serviços governamentais, e saúde. Para esses usos, exige testes prévios à implementação, avaliações de impacto de IA antes da implementação e periodicamente depois, supervisão humana adequada, e revisão humana atempada com oportunidade de recurso para as pessoas afetadas. As agências que não conseguirem cumprir esses mínimos devem descontinuar o uso em segurança. O memorando de aquisição complementar exige termos contratuais que dão à agência a capacidade de monitorizar e avaliar regularmente o desempenho, os riscos e a eficácia do sistema.

Seja qual for a opinião sobre os detalhes, a estrutura dessa orientação está correta. Trata a responsabilização como algo especificado no contrato, testado antes do lançamento, monitorizado após o lançamento, e aplicável através de suspensão. Isso é uma camada de verificação. A maioria das instituições que compram IA fora do sistema federal de aquisição não tem nada equivalente.

Acrescentaria uma advertência quanto a assumir que a capacidade chegará dentro do prazo e fechará a lacuna por si só. Em abril de 2026, o Departamento de Justiça adiou os prazos de conformidade da norma de acessibilidade web do Título II do ADA (Americans with Disabilities Act) em cerca de um ano para cada categoria de entidade, movendo as grandes entidades públicas para 26 de abril de 2027 e as mais pequenas para 26 de abril de 2028. Entre as razões apresentadas pelo Departamento estava o facto de ter sobrestimado as capacidades, tanto em pessoal como em tecnologia, das entidades abrangidas, e afirmou claramente que tecnologia avançada como a IA generativa ainda não automatiza de forma fiável a correção de conteúdo inacessível em larga escala. Trata-se de um regulador federal a registar que as alegações sobre a capacidade da IA ultrapassaram a entrega num contexto de acessibilidade. Vale a pena refletir sobre isso antes de tratar qualquer cronograma de fornecedor como um plano de governação.

O que as instituições devem exigir antes da implementação

As perguntas abaixo não são exóticas. São as perguntas que um comprador competente faria sobre qualquer sistema que acarrete consequências, e têm resposta.

Quem avaliou este sistema, e essas pessoas eram independentes do fornecedor e do comprador?

Que populações, dialetos, variantes regionais e contextos foram incluídos nos testes, e quais foram os resultados para cada um, em vez de agregados?

Com que dados foi treinado o sistema, sob que licença, e com que consentimento das pessoas cuja língua foi captada?

Qual é a taxa de erro medida em condições semelhantes às nossas, e o que acontece a essa taxa à medida que as condições se degradam?

Quem, na nossa organização, tem autoridade para suspender esta implementação, e o que aciona essa autoridade?

Como é que uma pessoa afetada comunica que o resultado estava errado, quem recebe essa comunicação, e o que muda em consequência disso?

A que informações sobre desempenho após o lançamento temos direito contratual, e qual é o nosso recurso se o desempenho baixar?

Uma organização que não consegue responder a estas perguntas não avaliou um sistema. Aceitou-o. A Novara Consulting Group construiu o Sign Language Access Trust Index para tornar estas perguntas respondíveis numa forma que a aquisição possa efetivamente utilizar, com padrões de evidência definidos, pontuação ao nível de indicadores, e uma distinção entre o que um fornecedor afirma e o que uma parte independente pode confirmar. O instrumento é específico para IA de língua gestual. A disciplina subjacente não é.

A próxima etapa da IA é a verificação

A conversa pública sobre IA tem sido organizada em torno da capacidade há três anos. Consegue escrever, consegue raciocinar, consegue ver, consegue agir, consegue interpretar comunicação humana. Essas perguntas estão a ser respondidas rapidamente, e frequentemente de forma afirmativa.

A pergunta que determina se algo disto é digno de confiança é diferente. É se as instituições que implementam estes sistemas conseguem perceber quando os sistemas estão errados, e se as pessoas por eles afetadas têm alguma legitimidade para o afirmar.

Neste momento, na maioria das implementações, a resposta a ambas é não. O risco não foi traduzido. Foi transferido para quem tem menos capacidade de o detetar.

Fechar essa lacuna não é uma questão de esperar por modelos melhores. Requer avaliação independente do fornecedor, padrões de evidência que resistam ao contacto com a aquisição, autoridade para interromper uma implementação atribuída a alguém presente, e um canal através do qual o julgamento da pessoa afetada conte para algo. Nada disso é tecnicamente difícil. É institucionalmente inconveniente, o que é um problema diferente e mais tratável.

Perguntas frequentes

O que é a lacuna de responsabilização da IA? A lacuna de responsabilização da IA é a distância entre aquilo que um sistema de IA é capaz de fazer e aquilo pelo qual qualquer parte identificável é responsável quando o faz mal. Surge quando as organizações implementam IA mais depressa do que constroem a capacidade de verificar o seu resultado, receber queixas das pessoas afetadas, e suspender a implementação quando esta falha.

A transparência da IA é o mesmo que a responsabilização da IA? Não. As regras de transparência, incluindo o Artigo 50 do EU AI Act, geralmente estabelecem a proveniência, ou seja, informam as pessoas de que o conteúdo foi gerado ou manipulado por IA. A responsabilização estabelece quem é responsável pela exatidão desse conteúdo e o que é devido à pessoa afetada quando este está errado. Um sistema pode estar totalmente rotulado e continuar a não ser responsabilizável.

Quem é legalmente responsável quando um sistema de IA causa dano? Na maioria das jurisdições, isto permanece por resolver. A Comissão Europeia retirou a sua proposta de Diretiva sobre Responsabilidade em matéria de IA em outubro de 2025, e vários quadros estaduais dos Estados Unidos estreitaram-se em direção a obrigações de notificação e de registo, em vez de deveres substantivos de cuidado. Na prática, a responsabilidade é determinada pelos termos contratuais, pela legislação setorial específica, como as leis de deficiência e de direitos civis, e pela doutrina geral de responsabilidade civil, e não por regras específicas para a IA.

Por que razão a IA de língua gestual é um caso de teste útil para a governação da IA? Porque torna visível a assimetria de verificação. A instituição que implementa o sistema geralmente não consegue avaliar o resultado, e a pessoa Surda que o consegue avaliar geralmente não tem autoridade para contestar a conclusão da instituição de que o acesso foi proporcionado. A mesma estrutura opera na IA médica, de contratação e de benefícios, mas é mais difícil de observar.

O que deve uma organização exigir antes de implementar um sistema de IA que afeta pessoas? Avaliação independente em vez de demonstração do fornecedor, dados de desempenho desagregados por população e contexto, proveniência divulgada dos dados de treino, uma pessoa designada com autoridade para suspender a implementação, um canal de queixas funcional para as pessoas afetadas, e direitos contratuais para monitorizar o desempenho após o lançamento.

O que é o SLAT Index? O Sign Language Access Trust Index é o instrumento de avaliação da Novara Consulting Group para sistemas de IA de língua gestual. Estabelece padrões de evidência definidos nos domínios de governação, linguístico, de transparência, de acessibilidade e de implementação, pontua os sistemas ao nível dos indicadores, e distingue a afirmação do fornecedor do facto independentemente confirmável, de modo a que as decisões de aquisição assentem em evidência e não em demonstração.

Heather M. Grizzle é a fundadora da Novara Consulting Group, que aconselha instituições sobre a governação e aquisição de tecnologia de acessibilidade de IA. É autora do SLAT Index Evaluation Standard.

Leitura relacionada: The Demo Is Not the Evidence: What Google DeepMind's Sign Language AI Must Prove Next

---

Fontes

- Comissão Europeia, Quick Facts: Transparency rules for AI systems (Artigo 50, aplicável a partir de 2 de agosto de 2026). https://digital-strategy.ec.europa.eu/en/factpages/quick-facts-transparency-rules-ai-systems
- Parlamento Europeu, Legislative Train Schedule, AI Liability Directive (retirada, outubro de 2025). https://www.europarl.europa.eu/legislative-train/theme-a-europe-fit-for-the-digital-age/file-ai-liability-directive
- Departamento de Justiça dos EUA, Extension of Compliance Dates for Nondiscrimination on the Basis of Disability: Accessibility of Web Information and Services of State and Local Government Entities, 20 de abril de 2026. https://www.federalregister.gov/documents/2026/04/20/2026-07663/extension-of-compliance-dates-for-nondiscrimination-on-the-basis-of-disability-accessibility-of-web
- Office of Management and Budget, M-25-21, Accelerating Federal Use of AI through Innovation, Governance, and Public Trust (abril de 2025). https://www.whitehouse.gov/wp-content/uploads/2025/02/M-25-21-Accelerating-Federal-Use-of-AI-through-Innovation-Governance-and-Public-Trust.pdf
- Office of Management and Budget, M-25-22, Driving Efficient Acquisition of Artificial Intelligence in Government (abril de 2025). https://www.whitehouse.gov/wp-content/uploads/2025/02/M-25-22-Driving-Efficient-Acquisition-of-Artificial-Intelligence-in-Government.pdf
- Colorado SB 26-189, que revoga e reformula o quadro da SB 24-205, sancionada em 14 de maio de 2026, com obrigações substantivas aplicáveis a partir de 1 de janeiro de 2027. https://leg.colorado.gov/bills/sb26-189
- 28 CFR 35.160, Communications (Título II, consideração primária). https://www.ecfr.gov/current/title-28/chapter-I/part-35/subpart-E/section-35.160

- 28 CFR 36.303, Auxiliary aids and services (Título III, padrão de consulta). https://www.ecfr.gov/current/title-28/chapter-I/part-36/subpart-C/section-36.303
- National Association of the Deaf, Minimum Standards for Video Remote Interpreting Services in Medical Settings. https://nad.org/minimum-standards-for-video-remote-interpreting-services-in-medical-settings/
- Federação Mundial de Surdos e Associação Mundial de Intérpretes de Língua Gestual, Statement on Use of Signing Avatars, 14 de março de 2018, atualizado a 14 de abril de 2018. https://wasli.org/wp-content/uploads/2023/07/WFD-and-WASLI-Statement-on-Avatar-FINAL-14032018-Updated-14042018-1.pdf

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, August 17). A IA Consegue Traduzir. Mas Quem Traduz o Risco?. Novara Consulting Group. https://doi.org/10.5281/zenodo.23003333