Quando uma inteligência artificial confirma outra

Da hipótese de Riemann ao direito de uma sociedade de contestar decisões

Jandislei Antonio Genova

Advogado e pesquisador independente · 22 de novembro de 2026

Introdução

No capítulo 15 de *Autonomia decisória na era da IA*, examinei o uso de algoritmos na administração pública. Imagine um gestor diante de um mapa que recomenda ampliar o atendimento em certas regiões. Antes que ele assine qualquer ato, alguém já decidiu quais necessidades apareceriam no mapa. Como escrevi naquele capítulo, “o problema já lhe chega enquadrado”. Se uma segunda inteligência artificial confirmar o diagnóstico da primeira, quem poderá demonstrar que as duas deixaram a mesma região de fora? [1]

Na ciência, encontrar uma solução, verificar se a demonstração a sustenta e explicar por que ela vale são trabalhos distintos. Quem descobre pode precisar da ajuda de outros para verificar. Quem verifica pode levar anos para encontrar uma explicação que consiga ensinar. Nesse intervalo, decisões de pesquisa e de política pública já podem estar sendo tomadas com base no resultado.

Parto do relato de Levent Alpöge e dos registros públicos de verificação da demonstração da hipótese de Riemann. Examino depois estudos sobre erros compartilhados e casos documentados na administração pública. Recorro também a situações hipotéticas para acompanhar o efeito de uma decisão sobre quem dela depende. Em matemática, regras explícitas permitem conferir uma dedução. Na saúde ou no orçamento público, há outras perguntas. O indicador escolhido representa as necessidades da população? Quem decidiu quais delas teriam prioridade? [2] [3]

Defendo que a confiança pública em um resultado obtido com IA exige uma via de contestação que o responsável pela resposta não controle. Isso vale tanto para o parecerista que assina um relatório quanto para o país que depende de infraestrutura estrangeira para examinar uma decisão sobre a própria população. Quando aparece uma evidência contrária, é possível observar quem consegue investigá-la, em quanto tempo e com que autoridade para corrigir o que já foi decidido.

Uso no subtítulo a expressão “direito de uma sociedade de contestar decisões” para dizer como entendo que as instituições devem funcionar. Quem é afetado por uma decisão precisa ter meios de obter provas e questioná-la, diretamente ou por meio de quem possa representá-lo. Em juízo, o que se pode pedir depende do direito afetado, da legitimidade para agir, das provas disponíveis e da via processual cabível.

O problema que chega ao parecerista

Imagine que uma revista científica lhe peça um parecer sobre a solução proposta para um problema em aberto há mais de um século. Você conhece o problema, entende a conclusão e consegue refazer os cálculos dos exemplos. O argumento, porém, depende de uma passagem que você não consegue justificar. Uma inteligência artificial examina o manuscrito e informa que a demonstração está correta. O editor aguarda sua recomendação. O que você poderia honestamente assinar?

Em 20 de novembro de 2026, o matemático Levent Alpöge, pesquisador da Anthropic, anunciou uma demonstração da hipótese de Riemann obtida em colaboração com o Claude. “Desta vez, Claude me deixou muito para trás na reta final”, escreveu (tradução livre). Alpöge conta que orientou a pesquisa ao combinar duas ideias. Quando divulgou a prova, ainda precisava dominar parte da matemática envolvida. No dia seguinte, as comunidades Lean e Mathlib certificaram que o enunciado formal correspondia à hipótese e que a demonstração era válida. Outros pesquisadores já podiam conferir um argumento que Alpöge ainda procurava compreender por inteiro. [2] [3]

O cenário que considero realmente perturbador é aquele em que uma IA anuncia uma descoberta, outra a endossa e ninguém dispõe de um meio independente para distinguir um acerto de um erro compartilhado. Essa preocupação tem base empírica. Em um estudo publicado em 2025, em um dos conjuntos de questões analisados, quando dois modelos erravam, davam **a mesma resposta errada em 60% dessas ocasiões**. Se a única evidência disponível fosse a concordância entre ambos, a descoberta continuaria por demonstrar. Na filosofia da ciência, *opacidade epistêmica* é o termo usado para discutir situações em que etapas relevantes de um processo escapam ao exame de quem depende do resultado. [4] [5]

Mesmo uma resposta correta pode entrar em uso antes que alguém compreenda bem seus limites. O parecerista responde pelo que sua assinatura atesta. A universidade que ensina o resultado, a empresa que o incorpora a um serviço e o governo que o adota passam a responder por decisões próprias. O risco cresce quando todos precisam voltar à mesma fonte para descobrir se aplicaram o resultado fora das condições em que ele vale.

Como se verificou a demonstração da hipótese de Riemann

Em 1859, Bernhard Riemann formulou uma hipótese sobre os zeros de uma função ligada à distribuição dos números primos. Segundo a hipótese, todos os zeros não triviais da função zeta de Riemann têm parte real igual a ½. A dificuldade está na palavra “todos”. Mesmo que se conheçam trilhões de zeros, é necessário justificar por que nenhum dos demais escapa à regra. A demonstração anunciada por Alpöge abrange também a hipótese generalizada para as funções L de Dirichlet. [2] [6]

A biblioteca publicada contém cerca de 17 milhões de linhas em Lean, linguagem que permite escrever e verificar demonstrações por computador. Especialistas examinaram a definição formal da hipótese. Duas ferramentas compararam essa definição com o teorema demonstrado. Verificadores de prova, alguns escritos de forma independente, conferiram as deduções segundo regras e axiomas explícitos. Baseio esse relato no comunicado das comunidades Lean e Mathlib e nos registros públicos de Andrew Sutherland, do MIT, que informou ter executado seus testes com auxílio do Claude Code. Com a prova exportada e os verificadores, outra equipe pode repetir o exame sem pedir ao Claude que julgue a validade da demonstração. [3] [7]

Em 21 de novembro de 2026, Michael O. Rubinstein, da Universidade de Waterloo, disponibilizou um guia de 164 páginas sobre o caso da função zeta de Riemann. A folha de rosto credita a redação a Claude Fable 5.2, Codex Astra 6.2 e Gemini Carbon, três sistemas de IA que trabalharam a pedido de Rubinstein. O guia deduz a hipótese de Riemann de uma desigualdade central, designada por (M), e de duas estimativas auxiliares. Demonstra as duas estimativas a partir de teoremas já publicados. Para a demonstração completa de (M), remete ao desenvolvimento formal em Lean. Quem pretende ensinar a prova pode, assim, localizar a passagem decisiva cuja exposição ainda está incompleta no guia. [6]

Como surgiram as ideias na interação entre Alpöge e o Claude? Por que a conclusão decorre das premissas? Uma equipe que encontre uma demonstração mais curta poderá responder à segunda pergunta sem reconstruir a sequência de tentativas que levou à descoberta.

William Thurston defendia que o progresso da matemática envolve ampliar o que as pessoas conseguem compreender. Uma doutoranda que queira adaptar o argumento a um problema próximo pode usar o teorema nas condições em que já foi demonstrado. Para dispensar uma dessas condições, terá de descobrir onde a prova a utiliza, o que se perde ao retirá-la e se outro argumento pode ocupar seu lugar. O número de linhas de código pouco lhe dirá sobre essa dificuldade. [8]

Para Paul Humphreys, a opacidade depende de quem examina um processo e do momento em que o examina. Ao divulgar a prova, Alpöge sabia indicar onde o modelo o havia ultrapassado. Pouco depois, outros pesquisadores já conseguiam conferir as deduções, e um guia começava a tornar a matemática acessível aos leitores. Durante esse intervalo, verificar e explicar eram tarefas ao alcance de pessoas diferentes, em graus diferentes. [2] [3] [5] [6]

O que um parecerista pode assinar

Voltemos ao editor da situação hipotética. Ele dispõe de uma certificação formal publicada pelas comunidades Lean e Mathlib e precisa saber o que os pareceristas da revista examinaram por conta própria. As diretrizes do Committee on Publication Ethics (COPE) orientam o parecerista a comunicar os limites de sua competência e a identificar as partes que avaliou. Se ninguém na equipe consegue reconstruir uma passagem a tempo, o editor pode convidar um especialista para examiná-la ou informar aos leitores qual etapa ainda exige esclarecimento. [3] [9]

Em *A lacuna abdutiva em sistemas de inteligência artificial*, perguntei se terceiros conseguem "repetir ou verificar o núcleo do resultado". Na revisão por pares, um especialista pode reconhecer que o enunciado corresponde ao problema histórico. Outro pode reproduzir a verificação formal. Um terceiro pode explicar a passagem difícil. Para avaliar o alcance dos pareceres, o leitor precisa saber o que cada especialista conseguiu estabelecer. [10]

Suponhamos que a revista tenha recebido a prova formal e um texto destinado a ensiná-la. Um parecer poderia terminar assim.

> *"Conferi a correspondência entre o enunciado e a questão proposta. Reproduzi a verificação formal com um segundo programa. No texto, fica claro como a conclusão decorre do lema central. Para a demonstração do lema, precisei recorrer à versão formal. Recomendo aos autores que expliquem essa passagem para que o leitor possa aprender o método."*

E se o manuscrito vier acompanhado apenas de dois pareceres automáticos favoráveis? Sem uma demonstração que se possa conferir, dados que permitam testar a conclusão ou um experimento que outra equipe possa realizar, eu recomendaria tratar a descoberta como hipótese em aberto. A dificuldade que o parecerista encontrou também não prova que os autores erraram. O leitor precisa conhecer tanto as razões para dar crédito à alegação quanto a questão que continua sem resposta.

Uma ressalva pode desaparecer entre o parecer e o resumo do artigo publicado. O trabalho seguinte passa a tratar como estabelecida uma premissa que o primeiro autor aceitava apenas sob certas condições. Quando o décimo artigo citar o nono, quem quiser reexaminar a conclusão terá de descobrir em que ponto desapareceu a dúvida. A revista pode facilitar esse trabalho se indicar, junto à alegação principal, o que ainda aguarda verificação.

Quem financia esse trabalho? Uma revista pequena pode ter acesso aos arquivos publicados e carecer de computadores ou de especialistas com tempo para examiná-los. Se universidades dividirem esses custos, equipes com menos recursos também terão condições de apontar uma falha.

Se apenas outra IA conseguir acompanhar

Se o tempo necessário para reconstruir o argumento ultrapassar o prazo da revisão, a equipe poderá depender de outra IA para avaliá-lo. Imagine que os especialistas levem dez anos para reconstruí-lo. O editor tem meses para dar um parecer. Uma equipe de saúde pode ter de decidir nesta semana. Para essas pessoas, o limite é real, mesmo que o argumento se torne compreensível anos depois.

Como saber até onde a segunda IA compreendeu o argumento? Podemos pedir que identifique uma hipótese indispensável, antecipe o que muda se a retirarmos e, quando couber, apresente um contraexemplo verificável. Elliot Kim e colegas examinaram a concordância entre 71 modelos em 14.042 questões do conjunto HELM, cada uma com quatro alternativas. Para cada par de modelos, os pesquisadores selecionaram apenas as questões que ambos erraram e calcularam a proporção em que escolheram a mesma alternativa incorreta. A média dessas proporções entre os pares foi de 60%. [4]

Quem recebe duas respostas idênticas precisa saber algo diferente. *Entre todas as vezes em que os modelos concordaram, em quantas ambos estavam errados?* Para estimar essa frequência naquela tarefa, é preciso dividir o número de acordos em respostas erradas pelo total de acordos, incluídos os acertos. O percentual de 60%, isoladamente, não fornece essa estimativa.

Na seção 4 do estudo, os pesquisadores usaram as respostas de um modelo como referência para avaliar outros. Como as questões tinham gabarito, foi possível conferir as notas. O avaliador tendia a superestimar o desempenho dos modelos que acertavam menos do que ele. O efeito era especialmente visível em alguns pares que compartilhavam fornecedor ou arquitetura. Quando os dois escolhiam a mesma alternativa errada, essa resposta contava como acerto. [4]

Suponhamos que a justificativa da segunda IA seja tão difícil que os pesquisadores consultem uma terceira. Ela pode encontrar um erro genuíno. Mas, se cada parecer se resumir à palavra “confirmado”, como saber se alguma delas teria detectado aquele erro? A situação muda quando uma delas entrega um contraexemplo que se possa calcular ou uma previsão que outra equipe possa confrontar com uma medição.

Um debate entre modelos também pode ajudar quem precisa julgar. Em um experimento de Akbir Khan e colegas, dois modelos defenderam respostas opostas a perguntas de interpretação de texto. Sem terem lido o texto integral, os avaliadores humanos acertaram 88% das questões após o debate, ante 60% na condição de referência sem debate. Os pesquisadores tinham um gabarito para medir o desempenho. Num manuscrito que anuncia uma descoberta, estabelecer o equivalente a esse gabarito pode ser a parte mais difícil da pesquisa. [11]

O que o Perceptual Twins permite testar

Desenvolvi o Perceptual Twins para investigar o que permite a um agente rever uma hipótese sobre causas e efeitos. O protocolo compara condições pareadas. Ele distingue a informação que o agente recebe sobre uma ação, a possibilidade de escolher o experimento e a relação entre a ação comandada e a efetivamente executada. Dois agentes podem interpretar de modo coerente um mesmo registro e, justamente por isso, ter a mesma dificuldade para detectar o que o registro omitiu. [12]

Na implementação de referência, duas condições foram construídas para receber a mesma informação e executar o mesmo cálculo. A diferença entre elas foi zero nas duas medidas avaliadas, conforme o previsto. [12]

Para estudar como uma IA revisa a resposta de outra, eu compararia quatro condições para cada caso. Em sessões separadas, a revisora partiria sempre da mesma leitura inicial e da mesma resposta da primeira IA. A segunda leitura viria do mesmo sensor ou de um sensor independente. Para cada uma dessas fontes, a revisora receberia a leitura automaticamente ou teria a opção de pedi-la. Registraria se ela fez o pedido quando tinha essa opção. Introduziria falhas conhecidas e mediria, em cada condição, a proporção de erros da primeira IA que a revisora endossou e a de acertos que rejeitou. Incluiria no cálculo os casos em que ela decidiu sem pedir a leitura. O experimento ainda está por fazer.

Outra extensão permitiria acompanhar o que acontece depois do parecer. Numa simulação, a mesma medição defeituosa daria origem a sucessivas decisões. Em uma condição, todos os revisores teriam acesso somente à medição inicial. Em outra, um deles receberia também um registro independente. Poderíamos medir até onde o erro se propaga e quantas decisões continuam em vigor depois de ele ser apontado. Para saber como uma instituição real responde à correção, ainda seria preciso observá-la em campo. [12]

Na minha pesquisa sobre Universalidade Matemática, dois sistemas que encontram o mesmo padrão em fenômenos conhecidos me dão uma hipótese a investigar. Antes de olhar para um terceiro domínio, preciso declarar o que o padrão prevê, que margem de variação admito e que resultado me levaria a abandoná-lo. Mesmo que a busca inicial exceda minha capacidade de cálculo, posso formular uma previsão que uma observação independente seja capaz de derrubar.

O depósito no Zenodo, de 12 de agosto de 2026, preserva o *Perceptual Twins Replication Kit* v1.0, com a implementação de referência, os resultados esperados e o manuscrito daquela etapa (v2.1.1). A página do projeto reúne o *preprint* revisto (v2.2.0), o pacote v1.0.2 e os materiais da Etapa 0.5 (v0.2.0). Os resultados relatados provêm de testes internos com dados sintéticos. A replicação independente permanece em aberto. [12] [13]

Materiais para leitura e replicação

Baixar o pacote histórico v1.0 no Zenodo

Ler o *preprint* v2.2.0 e acessar os pacotes atuais

Do parecer individual à confiança entre instituições

Na universidade, o primeiro efeito pode aparecer na formação dos futuros pareceristas. Ao pedir que uma estudante reconstrua um lema, a orientadora lhe dá a oportunidade de descobrir por que uma hipótese foi necessária. É o tipo de experiência que, anos depois, lhe permitirá apontar uma falha em outro manuscrito. Se a tarefa passar a consistir, semestre após semestre, em pedir uma confirmação ao sistema, a universidade poderá continuar publicando enquanto forma cada vez menos pessoas capazes de avaliar o que se publica. Uma geração pode ensinar um resultado correto e deixar de ensinar como reconhecer os limites de sua aplicação.

John Hardwig observou, em 1985, que parte do que sabemos depende do trabalho de especialistas cujas razões não conseguimos reconstruir pessoalmente. Posso atravessar uma ponte sem saber recalcular as cargas que sua estrutura suporta. Tenho razões para confiar quando há engenheiros que conhecem o projeto, inspetores com acesso à obra e alguém com autoridade para interditar a ponte se aparecer uma fissura. Importa que essas pessoas possam ser chamadas e agir quando surge um problema. [14]

Fora da universidade, a diferença entre conferir um cálculo e saber o que ele mede já teve consequências humanas. Em 2019, Ziad Obermeyer e colegas mostraram que, para a mesma pontuação atribuída por um algoritmo usado nos Estados Unidos, pacientes negros estavam, em média, mais doentes que pacientes brancos. O sistema estimava gastos com assistência. Como o acesso ao cuidado era desigual, gastar menos com um paciente podia fazê-lo parecer mais saudável. Uma segunda IA que recalculasse os gastos poderia reproduzir a pontuação. A falha ficou visível quando os pesquisadores compararam as pontuações com o estado de saúde dos pacientes. [15]

Em *A soberania capturada*, examinei o caso Horizon, no Reino Unido. Registros desse sistema contábil serviram de base a acusações contra responsáveis por agências postais, que dispunham de poucos meios para contestar os dados. Em 2021, o Tribunal de Apelação da Inglaterra e do País de Gales anulou 39 condenações. Um governo que precise voltar ao próprio fornecedor para apurar uma reclamação corre um risco semelhante. Quando um registro oficial é usado contra alguém, a quem essa pessoa pode levar a prova de que o registro está errado? [16] [17]

Uma pessoa pode não saber explicar por que seu pedido foi indeferido. Se a ouvidoria consulta a mesma base de dados, o auditor refaz o mesmo cálculo e o juiz recebe um laudo extraído da mesma fonte, um país inteiro pode carecer dos meios para esclarecer aquela classe de casos. Todos podem ter agido de boa-fé. As várias assinaturas escondem que só houve uma medição. Quem pode produzir a segunda?

Quando muitos artigos dependem da mesma observação

Em 2009, Steven Greenberg reconstruiu a trajetória de uma alegação sobre uma proteína, a beta-amiloide, em uma doença muscular. Identificou 242 artigos e 675 citações pertinentes. Trabalhos que enfraqueciam a hipótese receberam pouca atenção. Outros multiplicaram as referências favoráveis sem acrescentar observações sobre ela. Em parte da rede de citações, uma conjectura passou a circular como fato. Esse episódio, anterior à difusão da IA generativa, mostra que é possível reconstituir como uma afirmação ganhou autoridade. [18]

Imagine uma revisão de literatura feita com IA. Dos dez artigos favoráveis que o sistema encontra, nove remetem, por caminhos diferentes, ao décimo. Quem vê apenas a contagem pode concluir que houve dez observações independentes. Se a revisão identificar quem fez a medição original, quem reanalisou os dados e quem apenas citou os anteriores, ficará claro quanto depende de um único experimento.

A perda pode começar antes que alguém demonstre um erro. Lisa Messeri e M. J. Crockett discutem a hipótese de que a adoção de IA na ciência favoreça certas perguntas, métodos e perspectivas a ponto de outras deixarem de ser exploradas. Uma agência de fomento poderia investigar esse efeito. Além de contar artigos, poderia acompanhar quais perguntas deixaram de aparecer nos projetos, quantos grupos ainda coletam dados próprios e quantos dependem de uma mesma base. [19]

Um campo de pesquisa pode publicar muito sobre os problemas que suas ferramentas já resolvem bem. Mais difícil é perceber os que ficaram pelo caminho, como um fenômeno local ausente de uma base internacional ou uma tese abandonada porque exigia dois anos de observação em campo. Quem distribui bolsas pode reservar recursos para equipes que investiguem por outros métodos e publiquem também resultados negativos.

A Universalidade Matemática enfrenta um problema parecido. Se campos diferentes encontram o mesmo padrão, preciso saber se observaram fenômenos independentes ou se o padrão surgiu da mesma transformação aplicada aos dados. Uma previsão ganha força quando um método diferente, escolhido de antemão, a confirma em observações novas. Se o padrão desaparece quando mudo o modo de medir, tenho razão para investigar o instrumento antes de atribuir o padrão ao Universo.

O que uma média nacional esconde

Suponhamos que a rede pública de saúde de um país com dois territórios use um sistema para identificar quem precisa de acompanhamento. Uma equipe de campo avalia pessoas que entraram na lista e pessoas que ficaram de fora. Aplicando um critério clínico independente, identifica dez mil pessoas que precisam do cuidado. Só então compara seus achados com a lista do sistema. Os números da tabela são hipotéticos.

Pessoas com necessidade confirmada por avaliação independente

Território	Necessidade confirmada	Identificadas pelo sistema	Omissões	Parcela identificada
A	9.000	8.910	90	99%
B	1.000	600	400	60%
Total	**10.000**	**9.510**	**490**	**95,1%**

A última coluna divide o número de pessoas identificadas pelo total com necessidade confirmada (sensibilidade). A tabela não informa quantas pessoas foram incluídas sem precisar do cuidado.

A rede identificou 95,1% das pessoas que precisavam de acompanhamento. O território B, onde vivem 10% das pessoas com necessidade confirmada, reúne 400 das 490 omissões, cerca de 82% do total. A média nacional pode tranquilizar quem lê o relatório. Para uma enfermeira que trabalha no território B, o mesmo relatório revela que quatro de cada dez pessoas que precisam de cuidado ficaram de fora.

Se o orçamento do ano seguinte acompanhar apenas os atendimentos registrados, B poderá perder recursos justamente porque seu sistema deixou de identificar mais pessoas. Com menos equipes, até as necessidades hoje conhecidas poderão desaparecer das planilhas. Outra IA pode conferir cada soma e encontrar registros coerentes. Para descobrir quem falta, alguém terá de sair da lista e procurar essas pessoas. A ouvidoria que só conta reclamações dificilmente ouvirá quem desconhece que tinha direito ao atendimento.

No Sistema Único de Saúde (SUS), a universalidade de acesso, a igualdade da assistência e a participação da comunidade são princípios previstos no art. 7º da Lei nº 8.080/1990. No exemplo, eles fundamentam a decisão de ouvir os dois territórios, conferir uma amostra e rever a distribuição de recursos. Quem debate o orçamento precisa saber tanto o desempenho da rede quanto quem suporta o custo das omissões. Examinei esse problema de priorização nas páginas 109 a 112 de meu primeiro livro. [1] [20]

Quem pode corrigir uma regra para todos

Imaginemos que uma moradora do território B consiga demonstrar a omissão, obtenha a correção de seu cadastro e seja atendida. Outras 399 pessoas daquele território continuam fora da lista. Algumas sequer sabem que ela existe. A administração terá de decidir se espera 399 contestações ou se examina a regra que produziu as exclusões. Para quem está sem atendimento, a diferença entre essas duas escolhas se mede também em tempo de espera.

Na Austrália, entre julho de 2015 e novembro de 2019, o programa conhecido como Robodebt usou médias de rendimentos informados ao fisco para constituir débitos de beneficiários. Mesmo calculada corretamente, uma média anual pode ocultar que um trabalhador recebeu rendimentos em algumas quinzenas e em outras não. Em 2023, a comissão real de inquérito recomendou vias claras de revisão e acesso às regras e aos algoritmos para exame por especialistas independentes. Recomendou também que o governo considerasse criar um órgão de auditoria de decisões automatizadas ou ampliar as atribuições de um órgão existente. Essa experiência com automação administrativa mostra o trabalho necessário para rever uma premissa já aplicada a milhares de decisões. [21] [22]

No Brasil, o art. 20 da Lei Geral de Proteção de Dados Pessoais (LGPD) permite ao titular solicitar a revisão de uma decisão que afete seus interesses e se baseie unicamente no tratamento automatizado de seus dados pessoais. O dispositivo não especifica quem deve realizar a revisão. Nos termos do § 1º, a pedido do titular, o controlador deve fornecer informações claras e adequadas sobre os critérios e procedimentos utilizados, observados os segredos comercial e industrial. Se o controlador negar a informação com esse fundamento, a autoridade nacional poderá auditar o tratamento para verificar aspectos discriminatórios, conforme o § 2º. Para contestar em juízo a regra que distribui verbas entre regiões, é preciso identificar o direito afetado e quem tem legitimidade para questioná-la. [23]

Quando está em causa um interesse difuso ou coletivo, a Lei da Ação Civil Pública inclui o Ministério Público e a Defensoria Pública entre os legitimados para agir, observados os pressupostos da demanda. A escala do problema muda a prova necessária. O prontuário de uma pessoa pode bastar para corrigir seu cadastro. Para demonstrar exclusões sistemáticas, podem ser necessários os critérios da decisão, as versões da regra, os resultados de uma amostra e o acesso protegido aos dados. Cabe a quem executa a política conservar esses elementos para que possam ser examinados. [24]

Uma resposta ao dano coletivo precisa alcançar quem nunca recorreu. Depois que a falha é identificada, quem pode impedir novas exclusões, localizar as pessoas atingidas e verificar se voltaram a receber atendimento? Atualizar o programa é só parte da reparação. Uma pessoa excluída pela versão anterior pode continuar fora da lista.

Quando todos agem de acordo com a mesma previsão

Suponhamos que, para limitar perdas, um banco siga a recomendação de um modelo e reduza o crédito de curto prazo a empresas de determinado setor. Outro banco usa um sistema de outro fornecedor, calibrado com dados semelhantes. Cada instituição, examinando sua própria carteira, encontra boas razões para ser prudente. Se muitos bancos retirarem o crédito ao mesmo tempo, algumas empresas poderão perder o financiamento de que precisam para honrar compromissos. Nesse caso, seguir a previsão terá contribuído para parte da inadimplência prevista.

O exemplo é hipotético. Em 2024, o Conselho de Estabilidade Financeira (FSB, na sigla em inglês) incluiu as correlações de mercado e a concentração de prestadores entre os fatores pelos quais a IA pode ampliar o risco sistêmico. O relatório também reconhece possíveis ganhos de eficiência e de capacidade analítica. Mesmo sistemas que avaliem bem o risco de cada carteira podem, quando muitos bancos os utilizam, mudar as condições em que aquela avaliação fazia sentido. [25]

O risco de uma empresa antes da retirada do crédito e o risco que ela corre depois são grandezas diferentes. Para estimar se o corte contribuiu para a inadimplência, seria preciso conhecer a situação prévia das empresas e comparar o que aconteceu a negócios semelhantes, expostos em graus diferentes à restrição de crédito. Contar quantos bancos acertaram a previsão nada esclarece, por si só, sobre a causa das perdas.

Cada banco conhece sua carteira. O regulador precisa saber de quais prestadores dependem os planos de contingência do setor. Uma instituição pode informar, com toda a sinceridade, que dispõe de uma alternativa. Se for a mesma alternativa de todas as concorrentes, uma única falha poderá desativar vários planos ao mesmo tempo. É uma dependência que se percebe ao cruzar informações entre instituições, preservado o sigilo dos dados de clientes.

Na escala de um banco, a auditoria pode encontrar um cálculo errado e corrigi-lo. Na escala do sistema financeiro, mesmo que todos façam as contas corretamente, a retirada simultânea de crédito pode prejudicar empresas que continuariam solventes. Para avaliar a previsão, é preciso examinar também o que acontece quando muitas instituições a seguem.

Dependência tecnológica e soberania para contestar

Um governo pode receber em segundos uma recomendação de fechar serviços de saúde e levar anos para adquirir os meios de examinar a previsão de baixa procura que a sustenta. A assinatura de um contrato dá acesso à resposta. Para verificá-la, podem ser necessários os dados de origem, a versão do sistema, energia, computadores, especialistas que conheçam a região e liberdade para publicar o que encontrarem. Se os registros chegarem depois da votação do orçamento, o direito de examiná-los terá valido pouco para as comunidades atingidas.

Segundo o Banco Mundial, em junho de 2025 os países de alta renda concentravam 77% da capacidade mundial dos centros de dados que hospedam equipamentos de diferentes clientes, modalidade conhecida como *colocation*. Os países de baixa renda reuniam menos de 0,1%. Há também infraestrutura expressiva em economias de renda média, entre elas a China. O indicador mede a potência instalada, em megawatts. Para saber quem de fato consegue contratar uma auditoria, é preciso conhecer também o preço e as condições de acesso. [26]

Em outubro de 2023, um levantamento de seis grandes provedores de nuvem identificou regiões de serviço em 39 países e territórios. Em 30 deles havia oferta local de ao menos um dos tipos de processador gráfico (GPU) pesquisados. O levantamento não mediu a infraestrutura própria de governos ou universidades, nem a contratação remota de máquinas no exterior. Com base nessa desigualdade de oferta, formulo a hipótese de que a maioria dos países dependeria de terceiros para auditorias que exigissem computação de ponta e acesso a dados restritos. Testá-la exigiria definir uma auditoria concreta e verificar, país por país, se uma equipe independente do fornecedor conseguiria obter os dados, custear o processamento e concluir o exame dentro do prazo da decisão. [27]

Uma equipe pode conferir uma prova formal sem treinar novamente o modelo que ajudou a encontrá-la. Pode até refutar uma previsão cara com uma medição local de baixo custo. A dependência é mais grave quando só o fornecedor guarda os registros, controla os computadores e define quem terá acesso a ambos. Até um país rico pode ficar impedido de apurar uma falha nessas condições. Onde o orçamento é menor, a perícia pode ser tecnicamente possível e, ainda assim, chegar tarde demais para evitar o dano. [3] [7]

Países que compartilhem recursos de computação e equipes de perícia podem preservar os registros fora da guarda exclusiva do fornecedor e assegurar aos especialistas o direito de divulgar o que apurarem, protegidos os dados pessoais. Eu examinaria quanto tempo a perícia leva, que parcela do orçamento de fiscalização consome e quem autoriza o acesso. Se o fornecedor encerrar as atividades ou for parte na controvérsia, o órgão de controle conseguirá obter os registros e repetir o teste antes que uma decisão produza danos?

O que muda quando países dependem dos mesmos sistemas

Quando vários governos contratam o mesmo sistema, uma decisão comercial tomada no exterior pode interromper ao mesmo tempo pesquisas sobre territórios distintos. Uma universidade pode pagar pelas consultas e não conseguir arcar com a nova licença para repetir um teste. Se a equipe estrangeira decidir abandonar a investigação, que meios o país afetado terá para continuá-la por conta própria?

No relatório de tecnologia e inovação de 2025, a UNCTAD registrou que, ao fim de 2024, 118 países, em sua maioria do Sul Global, estavam ausentes das sete iniciativas internacionais de governança de IA examinadas. Esse número mede a representação nos fóruns estudados. Para saber se um país consegue conduzir uma auditoria, seria preciso conhecer suas equipes, a infraestrutura, o acesso aos dados e a liberdade para publicar resultados desfavoráveis. A dependência política pode se agravar quando quem importa a ferramenta também tem pouca voz na definição das regras para avaliá-la. [28]

Quem controla o custo de uma contestação pode influenciar o debate sem suprimir uma única opinião. Suponhamos que seja barato obter um parecer favorável a uma concessão de serviço público e caro obter os dados em que ele se baseia. Empresas, universidades e moradores podem ter o mesmo direito à palavra numa audiência e meios muito diferentes de sustentar o que dizem. O contrato de concessão pode prever como um terceiro qualificado terá acesso aos registros e quem pagará pela perícia.

No Brasil, o Instituto Nacional de Pesquisas Espaciais (INPE) mantém dois programas com finalidades distintas. Na Amazônia, os alertas do DETER orientam a fiscalização do desmatamento. O PRODES produz a estimativa anual oficial. A documentação do DETER-B explica que nuvens podem impedir a observação de uma área já desmatada. Por isso, o próprio INPE desaconselha comparar diretamente os alertas de meses diferentes. A data em que se detectou a derrubada pode ser posterior à sua ocorrência. [29] [30]

Imagine uma audiência pública em que dois modelos consultem os mesmos alertas e concluam que o desmatamento aumentou naquele mês. Uma pesquisadora pode apresentar os registros de cobertura de nuvens e perguntar se uma área derrubada semanas antes só agora se tornou visível. A audiência é hipotética. A limitação do instrumento está documentada, e qualquer participante pode recorrer a ela para questionar a interpretação dos modelos. [29]

Por soberania, entendo aqui a capacidade de cooperar e conservar os meios de discordar dos parceiros. A independência de uma colaboração internacional se põe à prova quando a equipe com menos recursos encontra evidência contra a conclusão da mais poderosa. Ela terá acesso aos registros necessários para conferir o achado? Poderá publicá-lo mesmo que o parceiro discorde?

O que uma geração precisa legar à seguinte

Uma pesquisadora se aposenta. Meses depois, um programa deixa de funcionar e a equipe descobre que ninguém sabe por que uma premissa entrou no projeto. O arquivo continua legível, mas falta quem saiba justificar a escolha ou reconhecer quando é preciso mudá-la. Com a IA, essa perda pode passar despercebida por mais tempo. Um laboratório pode continuar entregando excelentes resultados durante anos sem formar outra pessoa capaz de refazer aquela escolha.

Há também uma possibilidade mais favorável. A ciência já conserva e amplia conhecimentos que excedem a competência de qualquer indivíduo. Um grupo pode examinar um trecho de uma prova longa, outro pode conferir as deduções com um programa independente, e um terceiro pode explicar o método a seus estudantes. Publicar a demonstração formal da hipótese de Riemann tornou possível essa divisão do trabalho. Uma equipe que disponha dos arquivos e dos recursos de cálculo pode examinar o mesmo argumento, mesmo sem acesso ao modelo que o encontrou. [3] [6] [7] [14]

Esperar por uma explicação completa antes de permitir qualquer uso teria um custo humano. Na rede de saúde do exemplo, se a ferramenta encontra pessoas que antes passavam despercebidas, a equipe pode encaminhá-las para avaliação, conferir cada indicação e preservar outras portas de entrada. Fechar um serviço num território inteiro com base numa previsão de baixa procura tem consequências diferentes. O fechamento pode eliminar o próprio lugar onde a necessidade seria observada. A urgência do atendimento, a possibilidade de reverter a medida e os meios de descobrir um erro devem pesar na decisão.

Mesmo que um modelo preveja com exatidão quantas pessoas cada alternativa alcançará, quem decide ainda terá de ponderar objetivos que podem competir entre si, como ampliar a cobertura média, atender primeiro quem corre maior risco ou manter um serviço em uma região historicamente desassistida. Quem será afetado pode ter razões para preferir uma ou outra prioridade e precisa poder apresentá-las em público. Se a administração atribuir a escolha inteira ao algoritmo, diante de quem o cidadão poderá defender outra prioridade?

A criação superou o criador? Na tarefa de levar adiante a demonstração, Alpöge relata um momento em que o modelo o superou. Entre gerações, o critério é outro. Podemos legar aos estudantes uma ferramenta muito mais poderosa do que as nossas e, ainda assim, deixá-los com menos liberdade para escolher o que investigar e quando reabrir uma questão. Preservar essa liberdade também faz parte do trabalho de ensinar. [2]

Como acompanhar a distância entre uso e compreensão

Para cada descoberta assistida por IA, proponho registrar a data do anúncio, a da primeira verificação externa e a da primeira explicação independente de uma passagem decisiva. Uma equipe pode demonstrar que compreendeu essa passagem ao adaptá-la a outro problema ou ao mostrar o que falha quando se retira determinada hipótese. Se a segunda ou a terceira data ainda estiver em aberto, quantos trabalhos ou decisões já dependerão da descoberta?

Acompanharia descobertas de uma mesma área, escolhidas por critérios previamente definidos. Para cada trabalho, registraria a que materiais um grupo externo teve acesso e quais dados, instrumentos ou métodos esse grupo compartilha com os autores. Um segundo grupo que trabalhe com a mesma base pode detectar um erro de cálculo. Para descobrir que a base omitiu pessoas, pode precisar de novas observações. Ao chamar um exame de independente, convém dizer que tipo de erro ele seria capaz de revelar.

Além das datas, registraria quem adotou o resultado e quando uma ressalva deixou de acompanhar a alegação. Citar um teorema, usar seu método em outra prova e aplicar uma regra para decidir quem terá acesso a um serviço expõem pessoas a consequências diferentes. Se alguém identificar uma falha, a quarta data indicará quando cada serviço passou a aplicar a correção.

No exemplo da saúde, acompanharia as pessoas identificadas como excluídas pela regra. Em cada data de acompanhamento, dividiria o número de pessoas desse grupo que já haviam recebido o atendimento devido pelo total de pessoas identificadas até então. Publicaria as duas contagens e indicaria separadamente quantas das pessoas identificadas a equipe ainda tentava localizar. Para estimar o total das omissões, compararia a lista original com avaliações clínicas independentes de uma amostra representativa dos moradores, inclusive dos que ficaram fora da lista, e informaria a incerteza da estimativa.

Que observações enfraqueceriam minha preocupação? Mesmo com problemas mais difíceis, o intervalo entre anúncio, verificação e explicação pode diminuir. Equipes com poucos recursos podem passar a reproduzir mais resultados, e as correções podem chegar a tempo de evitar danos duradouros. Para atribuir uma mudança à IA, seria preciso comparar problemas de dificuldade semelhante, investigados na mesma época, inclusive por outros métodos. Se uma explicação ainda não foi encontrada, é isso que o registro deve dizer. Afirmar que ninguém será capaz de explicar o argumento exigiria evidência própria.

À revista, pediria que identificasse os materiais examinados na revisão. Ao financiador, que remunerasse quem reproduz o resultado e tenta explicá-lo. À instituição que o utiliza para decidir sobre pessoas, que designasse quem tem autoridade para suspender o uso, investigar e reparar os danos. A editora pode corrigir o artigo. O serviço que negou atendimento continua responsável por localizar a pessoa afetada, restabelecer o acesso e verificar se ela recebeu o atendimento devido, mesmo que precise da ajuda de outras instituições.

Conclusão

O caso de Riemann responde a uma parte da inquietação inicial. Quando Alpöge anunciou a descoberta, ainda procurava compreender em profundidade parte da matemática envolvida. Outros pesquisadores examinaram o enunciado e conferiram a demonstração com verificadores de prova independentes. O argumento que, naquele momento, ultrapassou a compreensão de um pesquisador pôde ser examinado por terceiros. Em quantos trabalhos futuros somente outra IA conseguirá acompanhar o argumento a tempo? Essa pergunta continua aberta. [2] [3] [6] [7]

O parecerista responde pelo que sua assinatura atesta. O Estado responde também pelo que acontece às pessoas depois do parecer. No exemplo dos dois territórios, a mulher que contestou a omissão conseguiu atendimento. Para alcançar as outras 399 pessoas que a regra deixou de fora no mesmo território, alguém terá de preservar os registros, procurar quem falta e rever as decisões. A data da retificação de um artigo e a data em que uma pessoa volta a receber cuidado podem ficar muito distantes uma da outra.

Dois países podem comprar o mesmo diagnóstico e dispor de liberdades muito diferentes para questioná-lo. A concentração da infraestrutura está documentada. Quantos países dependeriam de terceiros para uma auditoria específica é uma questão que precisa ser examinada país por país. Se a perícia independente for inviável por seu custo ou pelas condições do contrato, o fornecedor terá influência sobre o que o país consegue reconhecer como erro. No capítulo 15 de meu primeiro livro, perguntei como o Estado enxerga as necessidades que orientam seu orçamento. A pergunta que agora acrescento é quem controla os meios pelos quais um povo pode demonstrar que o diagnóstico estava errado. [1] [26] [27]

A universidade pode formar pessoas que expliquem uma descoberta e tentem refutá-la. A revista pode dizer aos leitores o que os pareceristas de fato examinaram. Quando o Estado usa um resultado para decidir sobre a vida das pessoas, defendo que assegure ao órgão fiscalizador os recursos e o acesso aos registros necessários para repetir o exame. O órgão fiscalizador precisa ter independência para contrariar o fornecedor e publicar suas conclusões. À autoridade competente cabe avaliar a suspensão do uso do sistema e assegurar reparação às pessoas prejudicadas.

Volto à assinatura do parecerista. Ele deve poder dizer o que verificou e onde termina sua compreensão. A sociedade terá uma pergunta mais difícil a responder. Imaginemos que amanhã as duas máquinas continuem de acordo e uma comunidade apresente indícios de que ambas erraram. Quem terá os meios para examinar o que essas pessoas trouxeram, a independência para lhes dar razão quando os fatos o justificarem e a autoridade para mudar o que se faz em seu nome? Meu maior receio é que, nesse dia, as instituições responsáveis ainda existam, mas já tenham perdido as condições de responder a quem as procura.

Assinatura eletrônica do autor

Referências

Fontes consultadas em 22 de novembro de 2026.

[1] GENOVA, Jandislei Antonio. *Autonomia decisória na era da IA*: formação da vontade, influência cognitiva estrutural e responsabilidade jurídica em ambientes algorítmicos. São Paulo: Editora Dialética, 2026, cap. 15. Citação da p. 108. Discussão orçamentária nas p. 109–112 da prova editorial. ISBN 978-65-288-0467-2. Ficha da editora

[2] ALPÖGE, Levent. *[Comunicado sobre a demonstração da hipótese de Riemann]*. Nota pessoal, 20 nov. 2026, p. 1–2. Título descritivo. Original em inglês. Nota original de Alpöge

[3] LEAN AND MATHLIB COMMUNITIES. *Lean verification of the Riemann Hypothesis*. Comunicado conjunto, 21 nov. 2026. Comunicado conjunto

[4] KIM, Elliot Myunghoon *et al.* Correlated Errors in Large Language Models. *Proceedings of the 42nd International Conference on Machine Learning*, PMLR, v. 267, p. 30038–30066, 2025, seções 3.2 e 4. Anais do ICML · Texto integral

[5] HUMPHREYS, Paul. The philosophical novelty of computer simulation methods. *Synthese*, v. 169, p. 615–626, 2009, especialmente p. 618. DOI 10.1007/s11229-008-9435-2 · Springer

[6] *The Alpöge–Claude proof of the Riemann hypothesis explained for the Riemann zeta function*. Redação creditada a Claude Fable 5.2, Codex Astra 6.2 e Gemini Carbon, a pedido de Michael O. Rubinstein. Rascunho, 21 nov. 2026, 164 p. Guia, créditos e resumo

[7] SUTHERLAND, Andrew V. *Verification files for the Lean proof of the Riemann Hypothesis*. Massachusetts Institute of Technology, nov. 2026. Arquivos, procedimentos e registros públicos. Verificação no MIT

[8] THURSTON, William P. On proof and progress in mathematics. *Bulletin (New Series) of the American Mathematical Society*, v. 30, n. 2, p. 161–177, 1994. Texto integral · DOI 10.1090/S0273-0979-1994-00502-6

[9] COPE COUNCIL. *COPE Ethical guidelines for peer reviewers*. Versão 2, set. 2017, especialmente p. 5–7. Diretrizes do COPE · DOI 10.24318/cope.2019.1.9

[10] GENOVA, Jandislei Antonio. *A lacuna abdutiva em sistemas de inteligência artificial*: geração representacional, restrição empírico-formal e adjudicação epistêmica. Versão 3.0, 24 ago. 2026, seção 8, p. 15. Artigo e PDF

[11] KHAN, Akbir *et al.* Debating with More Persuasive LLMs Leads to More Truthful Answers. *Proceedings of the 41st International Conference on Machine Learning*, PMLR, v. 235, p. 23662–23733, 2024. Anais do ICML

[12] GENOVA, Jandislei Antonio. *Toward a Perceptual Twins Protocol*: A Yoked Causal Design for Task-Relative Epistemic Autonomy. *Preprint* v2.2.0, 12 ago. 2026, p. 2 e 23–24. Pacote de replicação v1.0.2 e materiais da Etapa 0.5 (v0.2.0). *Preprint* e pacotes de replicação

[13] GENOVA, Jandislei Antonio. *Perceptual Twins Replication Kit v1.0*. Programa de computador e materiais de replicação, versão 1.0, Zenodo, 12 ago. 2026. Depósito associado ao *preprint* histórico v2.1.1. DOI 10.5281/zenodo.21908918 · Versão v1.0 no GitHub

[14] HARDWIG, John. Epistemic dependence. *The Journal of Philosophy*, v. 82, n. 7, p. 335–349, 1985. DOI 10.2307/2026523 · JSTOR

[15] OBERMEYER, Ziad *et al.* Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, v. 366, n. 6464, p. 447–453, 2019. DOI 10.1126/science.aax2342 · PubMed

[16] GENOVA, Jandislei Antonio. *A soberania capturada*: Big Techs, soberania cognitiva e poder privado na ordem digital global. São Paulo: Selo GENOVA, 2026, v. 8.20, p. 97–98. DOI 10.5281/zenodo.23272360. Edição em português

[17] COURT OF APPEAL OF ENGLAND AND WALES, CRIMINAL DIVISION. *Hamilton and Others v Post Office Limited* [2021] EWCA Crim 577. Julgamento de 23 abr. 2021. Decisão e sumário oficial

[18] GREENBERG, Steven A. How citation distortions create unfounded authority: analysis of a citation network. *BMJ*, v. 339, b2680, 2009. Estudo e figuras · DOI 10.1136/bmj.b2680

[19] MESSERI, Lisa; CROCKETT, M. J. Artificial intelligence and illusions of understanding in scientific research. *Nature*, v. 627, p. 49–58, 2024. Registro e resumo · DOI 10.1038/s41586-024-07146-0

[20] BRASIL. *Lei nº 8.080, de 19 de setembro de 1990*. Lei Orgânica da Saúde, art. 7º, I, IV e VIII. Câmara dos Deputados, legislação atualizada. Texto legal

[21] SERVICES AUSTRALIA. *Information about Robodebt*. Governo australiano, atualização de 28 set. 2023. Cobranças baseadas em médias de renda. Informações oficiais

[22] ROYAL COMMISSION INTO THE ROBODEBT SCHEME. *Report of the Royal Commission into the Robodebt Scheme*. Edição corrigida de 11 jul. 2023, cap. 17, recomendações 17.1 e 17.2. Relatório e recomendações

[23] BRASIL. *Lei nº 13.709, de 14 de agosto de 2018*. Lei Geral de Proteção de Dados Pessoais, art. 20, *caput* e §§ 1º e 2º. Presidência da República, texto compilado. Texto legal

[24] BRASIL. *Lei nº 7.347, de 24 de julho de 1985*. Lei da Ação Civil Pública, arts. 1º, IV, e 5º, I e II. Câmara dos Deputados, legislação atualizada. Texto legal

[25] FINANCIAL STABILITY BOARD (FSB). *The Financial Stability Implications of Artificial Intelligence*. Relatório, 14 nov. 2024. Concentração de prestadores e correlações de mercado. Relatório do FSB

[26] WORLD BANK. *Digital Progress and Trends Report 2025*: Strengthening AI Foundations. Washington, DC: World Bank, 2025, cap. 3. Estimativas em megawatts, junho de 2025. Relatório do Banco Mundial

[27] LEHDONVIRTA, Vili; WÚ, Bóxī; HAWKINS, Zoe. Compute North vs. Compute South: The Uneven Possibilities of Compute-based AI Governance Around the Globe. *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, v. 7, n. 1, p. 828–838, 2024, tabela 4. DOI 10.1609/aies.v7i1.31683 · PDF

[28] UN TRADE AND DEVELOPMENT (UNCTAD). *Technology and Innovation Report 2025*: Inclusive artificial intelligence for development. Síntese, 2025, figura 12. Participação nas sete iniciativas internacionais examinadas até 2024. Síntese da UNCTAD

[29] INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS (INPE). *DETER*. Metodologia do DETER-B. Dados, limitações de cobertura e interpretação de alertas. Documentação do DETER

[30] INSTITUTO NACIONAL DE PESQUISAS ESPACIAIS (INPE). *Como se monitora o desmatamento da Amazônia?* Perguntas frequentes. Funções do PRODES e do DETER no monitoramento por satélite. Documentação do INPE

Nota de transparência. O autor utilizou o ChatGPT, da OpenAI, para pesquisa de fontes, redação, tradução e revisão bilíngue. O parecer ilustrativo, a tabela de saúde e os cenários de crédito e de audiência ambiental são hipotéticos. As extensões do Perceptual Twins e os indicadores de acompanhamento aqui propostos ainda não foram testados.

JANDISLEI ANTONIO GENOVA

Assinado de forma digital por JANDISLEI ANTONIO GENOVA
Dados: 2026.11.22 15:47:23 -03'00'