Resolva seu problema com um toque de humor

A IA forneceu as fontes. Uma delas até existe

Como verificar citações e fontes geradas por ferramentas de inteligência artificial



Max Paradox

A IA forneceu as fontes. Uma delas até existe

Como verificar citações e fontes geradas por ferramentas de inteligência artificial

Max Paradox

Rozwiąż swój problem z przymrużeniem oka

ISBN: 9788384906293

ALL-IN Rafał Płonkowski



ÍNDICE

INTRO 5

Capítulo 1 - Bibliografia do reino da fantasia, ou por que a IA inventa fontes 11

Capítulo 2 - A citação está entre aspas, mas a verdade saiu de férias 21

Capítulo 3 - O link funciona, mas para onde leva? Anatomia de uma prova falsa 31

Capítulo 4 - "Mas a IA já verificou", ou as desculpas mais caras do mundo 40

Capítulo 5 - Não procure uma agulha na biblioteca, ou como encontrar uma fonte sem perder a sanidade 52

Capítulo 6 - Quando a resposta combina perfeitamente, ou a armadilha das próprias expectativas 63

Capítulo 7 - Telefone sem fio de terno, ou como as fontes se perdem entre pessoas e documentos 74

Capítulo 8 - O PDF aceita tudo, ou quando a IA interpreta mal documentos verdadeiros 85

Capítulo 9 - Cinco minutos pela verdade, ou como fazer a primeira auditoria de uma resposta da IA 97

Capítulo 10 - Da citação ao original, ou uma investigação com um final de verdade 107

Capítulo 11 - Os números não mentem, mas suas descrições têm uma imaginação fértil 118

Capítulo 12 - Construa uma bibliografia da qual você não precise ter vergonha 129

Capítulo 13 - Ensine a IA a mostrar evidências, não apenas a causar boa impressão.....140

Capítulo 14 - Da resposta da IA a uma decisão que você consegue defender..... 151

Capítulo 15 - Duas fontes, três verdades e uma pessoa à beira do desespero 163

Capítulo 16 - A fonte que começa a dar ordens, ou por que não confiar em todas as instruções de um documento.....174

Capítulo 17 - Quando não é possível verificar tudo, ou a arte de aceitar honestamente o que ainda falta 186

Capítulo 18 - A verificação que não termina quando você clica em “Enviar”.....197

Capítulo 19 - Seu próprio radar para detectar bobagens, ou como não voltar aos velhos hábitos 209

Capítulo 20 - A IA pode escrever com segurança. Você pode saber de onde vem essa certeza 220

INTRO

Imagine que você está preparando um relatório importante. O prazo se aproxima na velocidade de um entregador que resolve aparecer justamente quando você acabou de entrar no chuveiro. Você tem os dados, tem os argumentos, só faltam algumas fontes confiáveis para sustentar suas afirmações. Então, abre uma ferramenta de inteligência artificial e pede publicações que confirmem determinada informação. Instantes depois, recebe cinco referências bibliográficas impecavelmente apresentadas. Os autores têm nomes, os artigos têm títulos, os periódicos têm volumes e números, e algumas publicações vêm até com identificadores DOI. Tudo parece tão profissional que dá vontade de imprimir a resposta, encaderná-la em couro e colocá-la na estante, ao lado das enciclopédias.

Existe apenas um pequeno problema. O primeiro artigo não existe. O segundo existe, mas trata de outro assunto. O terceiro apresenta um autor verdadeiro e um título inventado. O quarto leva a uma página que deixou de funcionar há muito tempo. O quinto realmente existe, mas seus autores defendem algo quase oposto ao que a IA afirmou.

Parabéns. Você acaba de receber uma bibliografia em que a realidade faz apenas uma participação especial.

O mais preocupante, porém, não é o fato de a inteligência artificial conseguir inventar uma fonte. Muito mais interessante é a capacidade que ela tem de apresentar essa invenção de maneira convincente. Não aparece nenhum aviso dizendo: “Atenção: os nomes apresentados a seguir podem ser fruto da imaginação literária do processador”.

Você não ouve uma tossida nervosa. Não percebe nenhum rubor digital. Recebe uma resposta tranquila, organizada e segura, que parece ter sido escrita por alguém que acabou de voltar da biblioteca depois de passar três dias estudando publicações científicas.

Enquanto isso, um modelo de linguagem pode produzir um texto semelhante a uma bibliografia correta sem dispor de fundamentos suficientes para confirmar a veracidade de cada elemento. Essa é uma diferença importante. Saber escrever uma referência bibliográfica com aparência confiável não é a mesma coisa que localizar uma publicação, lê-la e verificar se ela realmente sustenta determinada afirmação. Da mesma forma, saber desenhar um mapa do tesouro não significa que exista um baú cheio de ouro enterrado debaixo da palmeira indicada. Às vezes, o que está enterrado ali é apenas a conta das férias.

E, convenhamos, usar inteligência artificial é realmente prático. Ela ajuda a organizar ideias rapidamente, compreender assuntos complicados, preparar rascunhos e encontrar caminhos para novas pesquisas. O problema começa quando confundimos conveniência com comprovação. A resposta parece convincente, então deixamos de perguntar de onde veio a informação. Recebemos um link e presumimos que ele leva a um conteúdo confiável. Vemos aspas e acreditamos que aquelas palavras foram realmente pronunciadas por alguém.

É assim que uma pessoa pode publicar uma citação que ninguém jamais disse, mencionar uma pesquisa inexistente ou compartilhar com colegas uma informação incorreta acompanhada da expressão “segundo especialistas”. Mais tarde, os especialistas descobrem que participaram de um estudo que nunca chegaram a conhecer. Talvez fiquem tão

surpresos quanto um unicórnio ao descobrir que foi incluído no cadastro oficial de animais de criação.

Este livro não foi escrito para convencer você de que a inteligência artificial é inútil. Isso seria tão sensato quanto jogar uma calculadora fora porque alguém digitou os números errados. A proposta é aprender a distinguir uma ferramenta que ajuda a procurar informações de um processo que realmente confirma sua veracidade. Você pode utilizar a IA intensamente, economizar tempo e, ainda assim, manter o controle sobre aquilo que decide aceitar como fato.

Primeiro, vale compreender uma verdade um tanto inconveniente: nem todas as fontes incorretas são completamente inventadas. Se toda referência falsa levasse a um artigo intitulado "Estudo Extremamente Importante Que Acabei de Inventar", a vida seria muito mais simples. O que você encontrará com maior frequência, porém, são erros sutis. Uma publicação verdadeira é atribuída ao autor errado. Um relatório autêntico aparece com o ano de publicação incorreto. Uma paráfrase é apresentada como se fosse uma citação literal. Um link direciona para uma página legítima, mas o conteúdo disponível nela não confirma a conclusão apresentada.

Por isso, simplesmente perguntar "Essa fonte existe?" não é suficiente. Também precisamos descobrir se o documento indicado é o correto, se seu conteúdo foi reproduzido com fidelidade, se o autor realmente aborda o assunto em questão e se a conclusão apresentada não ultrapassa aquilo que as evidências permitem afirmar. São verificações diferentes, embora, na correria, seja tentador jogar todas elas na mesma gaveta com a etiqueta

“Verificado”. A gaveta parece organizada. O conteúdo, nem sempre.

Na prática, vamos tratar as respostas da IA como pontos de partida para a verificação, e não como certificados automáticos de verdade. Você aprenderá a reconhecer sinais de alerta em bibliografias, localizar documentos originais, conferir nomes de autores e datas, verificar identificadores de publicações e encontrar citações dentro de seus contextos corretos. Também descobrirá por que diferenças aparentemente pequenas entre “o estudo demonstrou”, “os autores sugerem” e “em determinadas condições, observou-se” podem transformar completamente o significado de uma informação.

Isso não significa que você precise se tornar um detetive bibliográfico profissional. Não é necessário ter um escritório revestido de livros, uma lupa ou uma trilha sonora dramática tocando sempre que abrir outra aba do navegador. O que você precisa é de um método que permita decidir o que verificar primeiro, onde procurar confirmação e quando as evidências disponíveis são insuficientes. Caso contrário, acabará dedicando o mesmo tempo à conferência do título de um artigo pouco relevante e à verificação de uma informação da qual depende uma decisão importante.

Afinal, nem todas as situações exigem o mesmo nível de rigor. Se você está preparando uma anotação informal para uso pessoal, pode aceitar um grau maior de incerteza, desde que não tente escondê-la. Mas, se está produzindo uma publicação científica, um material educativo, um documento profissional ou um texto relacionado à saúde, ao direito ou às finanças, as consequências de um erro podem ser muito mais sérias. Nesses casos, verificar as fontes não

é excesso de cautela. É parte da responsabilidade por aquilo que você comunica.

Outra habilidade importante será aprender a admitir quando não foi possível confirmar determinada informação. Parece banal, mas as ferramentas modernas nos acostumaram a receber respostas imediatas. Um espaço vazio na bibliografia começa a parecer quase uma derrota pessoal. Entretanto, afirmar honestamente "Não encontrei uma confirmação confiável para esta citação" tem muito mais valor do que apresentar uma referência perfeitamente formatada que não leva a lugar nenhum. A ausência de evidências nem sempre comprova que uma afirmação seja falsa, mas certamente não autoriza apresentá-la como verificada.

Com o tempo, você perceberá que a habilidade mais importante não é flagrar a inteligência artificial cometendo erros. É saber fazer as perguntas certas e distinguir uma resposta de uma prova. Um modelo pode ajudar a encontrar pistas, indicar possíveis publicações, sugerir termos de pesquisa e organizar resultados. A confirmação definitiva, entretanto, exige consultar os materiais relevantes, preferencialmente junto ao editor original ou em outro local confiável que disponibilize o documento.

Portanto, se você já copiou uma referência bibliográfica sem abrir a fonte, acreditou em uma citação simplesmente porque ela parecia inteligente ou considerou a presença de um link uma prova suficiente, está no lugar certo. Não vamos organizar um julgamento público do seu navegador. Vamos aprender algo muito mais útil: aproveitar as possibilidades da inteligência artificial sem entregar a ela o poder de decidir o que é verdade.

Porque a inteligência artificial consegue preparar uma lista de dez fontes em questão de segundos. Sua tarefa não é ficar impressionado com o tamanho da lista. Sua tarefa é saber quantas dessas dez referências realmente merecem aparecer no seu texto.

E, se você descobrir que apenas uma delas existe, pelo menos saberá qual é.

Capítulo 1 - Bibliografia do reino da fantasia, ou por que a IA inventa fontes

Segunda-feira, 9h17. Na sala de reuniões, você apresenta os resultados de uma análise preparada na noite anterior. Tudo está indo maravilhosamente bem. Os gráficos são claros, as conclusões fazem sentido e a apresentação tem até um slide dedicado às fontes, algo que, no universo das apresentações corporativas, às vezes é tratado quase como uma demonstração de educação aristocrática. Em determinado momento, alguém pergunta de onde veio a informação sobre a eficácia da solução apresentada. Sem hesitar, você aponta para uma publicação científica que a inteligência artificial havia sugerido no dia anterior.

Seu interlocutor digita o título do artigo no mecanismo de busca. Nada. Pesquisa o nome do autor. Encontra um pesquisador, mas que trabalha em uma área completamente diferente. Tenta localizar a revista científica. Ela realmente existe, embora o número indicado não contenha nenhuma publicação semelhante. O silêncio toma conta da sala. Até o projetor, que até então produzia ruídos parecidos com a decolagem de um pequeno avião, parece começar a funcionar de maneira mais discreta.

Você acaba de descobrir que sua bibliografia tem mais imaginação do que o departamento de criação.

É uma situação especialmente desagradável porque o erro não tinha aparência de erro. Não havia palavras digitadas incorretamente, expressões aleatórias nem frases suspeitas. A referência apresentava autor, data, título, nome da revista e número do volume. Todos os elementos

estavam nos lugares certos. Faltava apenas um pequeno detalhe: a publicação real que alguém pudesse consultar.

Vale a pena entender por que situações assim acontecem. Sem essa compreensão, é fácil adotar uma de duas posições extremas. A primeira afirma que a inteligência artificial é uma bibliotecária genial em quem devemos confiar. A segunda a apresenta como uma golpista digital que tenta deliberadamente nos colocar em situações constrangedoras. As duas imagens são impressionantes, mas nenhuma ajuda a utilizar bem a ferramenta.

Um modelo de linguagem não precisa ter a intenção real de enganar você para gerar uma informação falsa. Ele pode simplesmente produzir uma resposta compatível com os padrões aprendidos durante o treinamento, mesmo que essa resposta não esteja suficientemente vinculada a um documento verdadeiro. Essa é a diferença fundamental entre uma resposta com aparência correta e uma resposta cujo conteúdo corresponde à realidade.

Quando uma referência elegante finge ser uma prova

Imagine uma tarefa que consiste em completar a frase: “Segundo pesquisas publicadas na revista...”. Uma pessoa que conhece a língua e as regras básicas da escrita acadêmica consegue sugerir uma continuação natural. Isso não significa, porém, que ela acabou de consultar os arquivos da editora responsável pela revista. Da mesma maneira, um modelo pode produzir referências bibliográficas extremamente convincentes utilizando relações reconhecidas entre sobrenomes, títulos, áreas de conhecimento e formatos de apresentação de publicações.

Dependendo de sua arquitetura e dos recursos disponíveis, a ferramenta também pode pesquisar na

internet, ler documentos ou consultar bases de dados externas. Essas funcionalidades oferecem melhores condições para realizar verificações, mas, por si só, não garantem que a resposta final esteja correta. O modelo pode interpretar incorretamente o material encontrado, combinar informações de duas publicações diferentes ou atribuir a uma fonte verdadeira uma conclusão que seus autores jamais formularam.

O problema, portanto, apresenta vários níveis. Às vezes, a fonte simplesmente não existe. Em outras ocasiões, ela existe, mas foi descrita incorretamente. Há ainda situações em que todos os dados bibliográficos estão corretos, mas o erro está na interpretação do texto. Esse último caso é especialmente traiçoeiro, porque uma simples busca para confirmar se a publicação aparece nos resultados não revela o equívoco.

Vamos considerar um exemplo fictício. Você pede à IA pesquisas sobre os efeitos de pequenas pausas na concentração durante o trabalho em escritórios. Recebe o título “The Impact of Microbreaks on Cognitive Performance in Office Environments”, os nomes dos autores, o ano de publicação e o nome de uma revista de psicologia. Tudo parece extraordinariamente plausível. O assunto combina com a área de conhecimento, o título transmite seriedade acadêmica e os sobrenomes parecem pertencer a pessoas que participam regularmente de congressos científicos.

Isso não significa que o artigo indicado exista. O título deste exemplo é apenas uma ilustração, não uma referência bibliográfica confirmada. É exatamente assim que devemos tratar qualquer sugestão de fonte ainda não verificada: como uma informação que precisa ser conferida, mesmo quando apresenta uma aparência perfeitamente profissional.

Curiosamente, quanto mais detalhada é uma resposta, mais forte pode ser a impressão de confiabilidade que ela provoca. Número de página, data de publicação e identificador DOI parecem carimbos oficiais. Ao vê-los, a pessoa presume, quase inconscientemente, que alguém já realizou as verificações necessárias. Entretanto, uma resposta detalhada pode ser apenas isso: detalhada. Um erro extremamente preciso continua sendo um erro, apenas mais bem-vestido.

Aliás, nem é necessário inventar o nome de uma revista científica para criar problemas. Basta combinar o nome verdadeiro de uma editora respeitada com um artigo inexistente. O elemento autêntico transmite uma aparência de solidez à descrição inteira. É como se alguém fornecesse o endereço fictício de um apartamento localizado em um prédio que realmente existe. O edifício está lá, a rua está correta, mas, atrás da porta indicada, existe apenas uma parede.

Por que somos convencidos com tanta facilidade

Nossos problemas não resultam exclusivamente do funcionamento da tecnologia. A outra metade da questão está na maneira como avaliamos as informações. Quando um texto é fluente, organizado e escrito em tom seguro, temos maior tendência a considerá-lo competente. Isso acontece especialmente quando ele trata de um assunto que não dominamos suficientemente.

Também não ajuda o fato de geralmente recorrermos à IA justamente quando queremos economizar tempo. Precisamos de material rapidamente, e a bibliografia recebida corresponde perfeitamente às nossas expectativas. Verificar as referências significa trabalho adicional. Abrir publicações, conferir autores e comparar

citações parece muito menos atraente do que apertar o botão “Copiar”.

Além disso, existe uma tentação sutil de confirmar aquilo em que já acreditamos. Se a ferramenta apresenta cinco fontes que sustentam nossa tese, sentimos satisfação. Se indicasse cinco publicações que contradizem a mesma tese, provavelmente começaríamos a fazer perguntas adicionais. É justamente por isso que precisamos ter cuidado especial com fontes que dizem exatamente aquilo que gostaríamos de ouvir.

Imagine, por exemplo, que você esteja preparando argumentos a favor da adoção de determinada forma de organização do trabalho. Pergunta: “Quais pesquisas comprovam que esse método aumenta a produtividade?”. A própria pergunta já pressupõe que existem evidências adequadas. O modelo pode responder de acordo com essa expectativa, em vez de questionar a premissa.

Seria mais sensato perguntar: “Quais publicações confiáveis investigam a eficácia desse método? Apresente resultados favoráveis e desfavoráveis e indique claramente as informações que você não consegue confirmar”. Essa formulação não garante uma resposta livre de erros, mas muda a direção da pesquisa. Você deixa de encomendar provas para uma conclusão previamente escolhida.

Isso também é importante quando você utiliza uma ferramenta que realmente tem acesso a mecanismos de busca. A internet contém uma enorme quantidade de materiais de qualidade muito variada. Encontrar uma página não significa que a afirmação apresentada nela seja verdadeira. Pesquisar e avaliar a confiabilidade são duas atividades distintas.

As quatro formas de uma fonte falsa

Na prática, é útil reconhecer quatro situações fundamentais, porque cada uma exige uma reação diferente.

A primeira é a ficção bibliográfica completa. A publicação, os autores ou a revista não correspondem a nenhum documento real. Você pode passar meia hora procurando um artigo que existe exclusivamente na resposta gerada. Se, depois de uma verificação razoável em catálogos e bases de dados, não encontrar confirmação, não inclua a referência na bibliografia como se fosse verdadeira. Você não precisa demonstrar que o documento jamais existiu em toda a história da humanidade. Basta reconhecer honestamente que não foi possível verificá-lo.

A segunda situação é a mistura de dados. O artigo existe, mas o ano, o título ou os autores estão incorretos. Às vezes, o modelo monta uma única referência bibliográfica utilizando fragmentos de várias publicações verdadeiras. Identificar esse erro exige comparar a descrição completa com o registro bibliográfico original. Não é recomendável corrigir os elementos com base em suposições, porque você pode acabar criando acidentalmente mais uma referência híbrida.

A terceira forma envolve citações distorcidas. O autor realmente escreveu determinada frase, mas a IA modifica algumas palavras, omite uma ressalva importante ou combina dois fragmentos em uma única declaração. O sentido pode permanecer semelhante, mas o texto colocado entre aspas deve corresponder ao original. Se você quiser apresentar o significado da declaração com suas próprias palavras, utilize uma paráfrase e indique adequadamente a fonte. Não vista uma paráfrase com a fantasia de uma citação literal.

A quarta situação é a mais traiçoeira: a fonte existe, a citação está correta, mas não sustenta a afirmação apresentada. A publicação descreve uma associação entre duas variáveis, e a IA apresenta essa relação como prova de causalidade. O estudo envolve um grupo pequeno e específico, mas a resposta formula uma conclusão aplicável a todas as pessoas. Os autores discutem uma possibilidade, enquanto o resumo gerado anuncia uma certeza.

Portanto, você pode ter um artigo verdadeiro, um link correto e, ainda assim, um argumento equivocado. A bibliografia é absolvida, mas a conclusão acabou de fugir da cena do crime.

O primeiro filtro, antes de começar a pesquisar

Você não precisa realizar imediatamente uma análise completa de cada publicação. No trabalho cotidiano, é útil ter um filtro preliminar que ajude a decidir quais referências exigem uma verificação mais detalhada. Comece lendo a resposta da IA como uma proposta, e não como uma pesquisa já concluída.

Preste atenção às fontes apresentadas sem dados completos, àquelas que parecem suspeitosamente perfeitas para sua pergunta e às que contêm afirmações muito categóricas sem indicar trechos específicos do documento. Essas características não demonstram que a informação seja falsa, mas justificam uma conferência. Da mesma maneira, a ausência de autor ou data nem sempre significa falta de confiabilidade, porque nem todos os materiais valiosos seguem a estrutura de uma publicação científica.

Para cada referência importante, tente responder a três perguntas. Consigo localizar o documento independentemente da resposta da IA? A descrição

bibliográfica corresponde ao documento real? Sei exatamente qual trecho sustenta minha afirmação? Se você não consegue responder à primeira pergunta, não avance automaticamente para a terceira fingindo que está tudo certo.

Suponha que você tenha recebido cinco fontes para uma apresentação. Encontra rapidamente duas delas em catálogos confiáveis. Uma apresenta o ano incorreto, mas é possível identificar o documento original. As outras duas não aparecem nas pesquisas. Isso não significa que você precise descartar todo o trabalho. Pode manter as referências confirmadas, corrigir os dados da terceira e eliminar as duas que não foram verificadas. O importante é não tratar as cinco como se tivessem o mesmo nível de confiabilidade.

Se uma fonte for essencial para uma decisão importante, confirmar sua existência será apenas o começo. Também será necessário avaliar o conteúdo e as limitações da publicação. Por outro lado, quando se trata de material complementar, você pode simplesmente abrir mão de uma referência incerta, em vez de passar horas procurando por ela.

Também vale a pena guardar um registro das verificações realizadas. Não precisa ser um sistema complicado. Uma anotação breve como “confirmado no site da editora”, “texto completo não encontrado” ou “a fonte não sustenta a conclusão” frequentemente é suficiente para que, uma semana depois, você saiba por que determinada publicação foi utilizada ou descartada.

A IA como assistente, não como testemunha principal

A mudança mais útil consiste em separar dois papéis. A inteligência artificial pode ser excelente para ajudar a formular perguntas, organizar conceitos e desenvolver estratégias de pesquisa. Entretanto, não deve se transformar automaticamente na testemunha definitiva da veracidade das informações que ela própria gerou.

Quando você pergunta “As fontes que você apresentou são verdadeiras?”, pode receber outra resposta convincente que também precisa ser verificada. Isso não constitui uma conferência independente. A ferramenta pode corrigir o erro, mas também pode manter a imprecisão anterior. A simples declaração “verifiquei” não substitui a possibilidade de identificar o que realmente foi conferido e onde essa conferência ocorreu.

É melhor utilizar a IA para preparar uma lista dos elementos que precisam ser confirmados e, em seguida, compará-los com os documentos originais. Você também pode pedir que a ferramenta indique as limitações da própria resposta. Assim, recebe ajuda para organizar o trabalho sem transferir toda a responsabilidade para o sistema.

Não é necessário tratar cada frase gerada como um suspeito de investigação criminal. Basta ajustar o rigor da verificação às possíveis consequências de um erro. Uma informação utilizada como inspiração informal tem um peso diferente de uma citação em uma dissertação ou de uma referência que fundamenta uma decisão empresarial. É justamente nesses casos mais importantes que você precisa de algo além de uma referência elegante.

Portanto, da próxima vez que receber uma bibliografia que parece ser resultado de um mês inteiro de trabalho de uma equipe de pesquisadores, não a descarte automaticamente. Mas também não a aceite sem conferir.

Trate-a como uma lista de endereços possíveis que merecem uma visita.

Porque a bibliografia mais profissionalmente formatada do mundo continua sendo inútil se metade de seus autores mora exclusivamente na imaginação do modelo.

Capítulo 2 - A citação está entre aspas, mas a verdade saiu de férias

As aspas estão entre os menores sinais tipográficos aos quais as pessoas concedem um poder surpreendentemente grande. Basta colocar algumas palavras entre elas, acrescentar o nome de uma pessoa famosa e, de repente, uma frase comum começa a parecer uma declaração histórica. Se ao lado aparecer uma data, o título de um livro ou o nome de uma conferência, a credibilidade aumenta ainda mais. A pessoa deixa de perguntar se alguém realmente pronunciou aquelas palavras e começa a pensar na melhor maneira de utilizá-las em sua apresentação.

Certo dia, você está preparando um artigo sobre tomada de decisões. Pede à IA uma citação curta de um pensador conhecido que destaque a importância do pensamento crítico. Recebe uma frase tão perfeita que dá vontade de colocá-la em um cartaz acima da mesa de trabalho. A declaração tem um ritmo adequadamente solene, o autor é reconhecido e a mensagem combina perfeitamente com o texto.

Entretanto, quando você tenta localizar a citação na publicação original, surge um problema. O mecanismo de busca mostra centenas de páginas repetindo a mesma frase, mas nenhuma indica o local específico em que ela apareceu pela primeira vez. Um portal menciona um livro, outro aponta um discurso e um terceiro apresenta uma data que não pode ser associada a nenhum acontecimento confirmado. O quarto cita o primeiro, o quinto cita o quarto e o sexto garante que a frase é “amplamente conhecida”.

Acaba de surgir uma orquestra digital em que todos os músicos estão tocando a partir da mesma partitura errada.

O problema das citações falsas é mais antigo que a inteligência artificial. Há muito tempo, as pessoas atribuem a figuras conhecidas declarações que parecem suficientemente inteligentes, engraçadas ou inspiradoras. A internet facilitou sua disseminação, e as ferramentas de IA podem acelerar ainda mais a criação de novas versões. Isso não significa, porém, que toda declaração atribuída incorretamente tenha sido produzida por um modelo de linguagem. Parte desses erros já estava presente nos materiais disponíveis antes mesmo do treinamento dos modelos.

É justamente por isso que verificar citações exige mais precisão do que simplesmente perguntar quem é o autor. Precisamos determinar quais palavras foram realmente utilizadas, em que contexto apareceram, onde foram registradas e se a pessoa indicada é de fato responsável por elas. Só então as aspas começam a cumprir sua função.

Quando a paráfrase veste o terno de outra pessoa

O problema mais comum não precisa envolver uma declaração completamente inventada. Às vezes, a IA reconhece o sentido geral do pensamento de alguém, reformula esse pensamento com suas próprias palavras e depois o apresenta como uma citação literal. O resultado é uma frase fiel a determinada ideia, mas infiel à formulação original do texto.

Suponha que o autor de uma publicação escreva: “No grupo de participantes analisado, as pausas regulares estavam associadas a níveis menores de fadiga relatada”. A IA pode transformar esse trecho em algo muito mais

impactante: “Pausas regulares reduzem o cansaço dos trabalhadores”. À primeira vista, as duas frases tratam de assuntos semelhantes. A diferença, entretanto, é enorme.

A primeira descreve uma associação observada em um grupo específico. A segunda sugere uma relação geral de causa e efeito. A limitação referente às pessoas estudadas foi eliminada, desapareceu a informação de que o cansaço havia sido relatado pelos próprios participantes, e uma formulação cautelosa foi substituída por uma afirmação categórica.

Se você colocar a segunda frase entre aspas e atribuí-la ao autor, estará colocando na boca dele palavras que, nesse exemplo, nunca foram pronunciadas. Além disso, modificará o significado da declaração original. Isso já não é uma simples alteração estética do texto.

Lembre-se de que citações literais e paráfrases desempenham funções diferentes. A citação literal serve para reproduzir fielmente determinadas palavras. A paráfrase permite apresentar o pensamento de outra pessoa utilizando sua própria linguagem, preservando o significado e indicando adequadamente a fonte. Nos dois casos, é necessário compreender o original, mas somente no primeiro se espera correspondência com a formulação exata.

A inteligência artificial pode ajudar a preparar uma paráfrase, desde que você tenha acesso ao texto original e consiga avaliar se o resultado está correto. Quando solicita uma citação exata, por outro lado, você precisa de um trecho verdadeiro do documento. O modelo não deve substituir palavras ausentes por sua própria criatividade, mesmo quando faz isso com uma elegância literária impressionante.

Nas citações traduzidas, surge uma dificuldade adicional. Uma tradução para o português de uma declaração originalmente feita em inglês não é uma reprodução literal das palavras no mesmo idioma. Ela pode ser apresentada como uma tradução, mas seu caráter precisa ser identificado adequadamente. Se você utilizar uma tradução publicada, vale a pena indicar a fonte correspondente. Caso faça a tradução por conta própria, não apresente sua versão como se fosse a única formulação autorizada.

Essa diferença é especialmente importante em textos acadêmicos e profissionais, nos quais uma única palavra pode modificar o alcance de uma afirmação. O inglês “may” não significa a mesma coisa que “must”, assim como “associated with” não equivale automaticamente a “caused by”. Ao tornar a linguagem mais fluida, podemos transformar discretamente uma conclusão cautelosa em uma garantia categórica.

Setecentas páginas repetem a mesma coisa. Então, quem verificou?

Uma das características mais enganosas da internet é a possibilidade de repetir uma informação inúmeras vezes sem aumentar a quantidade de evidências independentes. Uma citação pode aparecer em centenas de sites, milhares de publicações nas redes sociais e dezenas de apresentações, embora todos esses materiais tenham surgido, em última análise, de uma única postagem não verificada.

Imagine que alguém publique uma frase impactante atribuída a um cientista famoso. Outro autor a copia para um artigo, uma terceira pessoa a coloca em uma imagem e alguém mais a acrescenta a uma coleção on-line de citações inspiradoras. Depois de alguns anos, o mecanismo de busca