Inteligência Artificial e Desinformação: Riscos para a Saúde Pública e para os Processos Eleitorais

Gustavo Xavier Evangelista
FGA0221 — Inteligência Artificial
Universidade de Brasília — FCTE
Brasília, Brasil
Matrícula: 241025247
Professor: Prof. Dr. Fabiano Araujo Soares
Data: 3 de setembro de 2026

***Resumo*—A inteligência artificial reduziu o custo e o tempo necessários para produzir, personalizar e distribuir conteúdo sintético plausível. Esta resenha crítica examina como essas capacidades alteram a desinformação em dois contextos de alto impacto: vacinação e processos eleitorais. São confrontados os conteúdos das aulas com os relatórios AI100 de 2016 e 2021, a série AI Index de 2021 a 2025, documentos da Organização Mundial da Saúde, estudos empíricos e registros do AI Incident Database. A análise distingue produção, amplificação, exposição e impacto causal. Os casos estudados mostram que sistemas de recomendação podem aumentar a exposição a materiais enganosos sobre vacinas e que clonagem de voz já foi utilizada em tentativa de supressão eleitoral. Entretanto, a ocorrência desses mecanismos não prova, isoladamente, mudança efetiva em vacinação ou resultados eleitorais. Conclui-se que a IA não criou a desinformação, mas pode ampliar sua escala, personalização e credibilidade aparente; por isso, a resposta deve abranger modelos, plataformas, operadores humanos e instituições públicas.**

***Index Terms*—inteligência artificial, desinformação, vacinas, eleições, sistemas de recomendação, ética em IA**

***Abstract*—Artificial intelligence has reduced the cost and time required to produce, personalize, and distribute plausible synthetic content. This critical review examines how these capabilities affect misinformation in two high-impact contexts: vaccination and electoral processes. Course materials are compared with the 2016 and 2021 AI100 reports, the AI Index reports published from 2021 to 2025, documents from the World Health Organization, empirical studies, and records from the AI Incident Database. The analysis distinguishes content production, amplification, exposure, and causal impact. The examined cases indicate that recommender systems may increase exposure to misleading vaccine-related materials and that voice cloning has already been used in an attempt at voter suppression. However, the occurrence of these mechanisms alone does not demonstrate actual changes in vaccination behavior or electoral outcomes. The review concludes that AI did not create misinformation, but it can increase its scale, personalization, and apparent credibility. Therefore, responses must address models, platforms, human operators, and public institutions.**

***Index Terms*—artificial intelligence, misinformation, vaccines, elections, recommender systems, AI ethics**

I. Introdução

A inteligência artificial ampliou de maneira expressiva a capacidade de produzir textos, imagens, áudios e vídeos plausíveis. Essa capacidade oferece benefícios para educação, comunicação e criação, mas também reduz o custo e o tempo necessários para fabricar conteúdos enganosos. A desinformação não surgiu com a IA: rumores, propaganda e manipulação política são anteriores aos computadores. A novidade está na possibilidade de gerar material em grande volume, adaptá-lo a públicos distintos, imitar pessoas reais e distribuí-lo por sistemas digitais que selecionam conteúdo de acordo com métricas de atenção.

Nas aulas introdutórias da disciplina, a evolução da IA foi relacionada ao estado atual do campo, aos agentes inteligentes, à racionalidade e aos riscos e benefícios de suas aplicações. A retropropagação do erro, ao permitir o ajuste de redes neurais e o aprendizado de representações, constitui um marco técnico dessa evolução [7]. Entretanto, a capacidade de um modelo não determina sozinha seu impacto. Dados, objetivos, interfaces, plataformas, usuários e instituições compõem um sistema sociotécnico no qual decisões aparentemente locais podem produzir consequências coletivas.

Dois contextos tornam esse problema especialmente relevante. O primeiro é a vacinação, em que informações falsas ou enganosas podem afetar a confiança nas autoridades de saúde e a disposição de adotar medidas preventivas. A Organização Mundial da Saúde emprega o termo "infodemia" para caracterizar o excesso de informações — incluindo conteúdos falsos ou enganosos — que acompanha uma emergência sanitária e dificulta a identificação de orientações confiáveis [10]. Em experimento realizado no Reino Unido e nos Estados Unidos, a exposição a desinformação sobre vacinas contra a COVID-19 reduziu a intenção declarada de vacinação entre participantes que inicialmente afirmavam que certamente se vacinariam [5]. Esse resultado não significa que toda exposição produza o mesmo comportamento, mas demonstra que o dano não é apenas hipotético.

O segundo contexto são as eleições. Sistemas generativos podem produzir deepfakes, clonar vozes, operar perfis automatizados e oferecer interação política personalizada. Essas técnicas podem atribuir falas falsas a candidatos, segmentar narrativas e dificultar a distinção entre evidência autêntica e conteúdo fabricado. Ao mesmo tempo, é necessário evitar conclusões alarmistas. A existência de ferramentas persuasivas ou de casos conhecidos não prova que a IA tenha alterado o

resultado de uma eleição. Estudos recentes encontram tanto capacidade mensurável de persuasão em diálogos com modelos quanto baixa prevalência de conteúdo político gerado por IA em determinados conjuntos de mensagens. Essa tensão entre capacidade técnica, ocorrência observada e impacto efetivo será central para a análise.

Esta resenha investiga de que modo o uso da inteligência artificial altera a produção, a circulação e o impacto da desinformação em campanhas de vacinação e processos eleitorais. Para isso, confronta os conteúdos das aulas com os relatórios AI100 de 2016 e 2021 [4], [9], a série *AI Index* de 2021 a 2025 [6], documentos da OMS e estudos empíricos. Defende-se que a IA intensifica riscos ao tornar conteúdo persuasivo mais barato, escalável e personalizável, mas que seus efeitos dependem do alcance, da repetição, do contexto social e da resposta institucional. A mesma tecnologia também pode auxiliar a detecção de narrativas enganosas e a comunicação pública. Portanto, a responsabilidade deve considerar não apenas o modelo, mas toda a cadeia de produção, recomendação, divulgação e controle da informação.

A Fig. 1 sintetiza essa cadeia e explicita uma distinção essencial para a resenha: a capacidade de gerar uma mensagem não equivale à exposição do público, assim como a exposição não demonstra, por si só, mudança de comportamento. Entre essas etapas atuam escolhas técnicas, comerciais, políticas e institucionais.

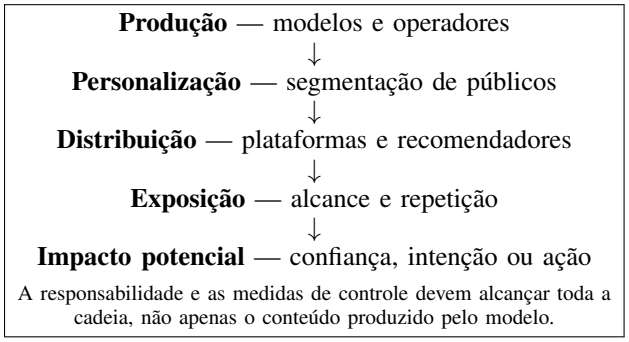

Figura 1. Cadeia sociotécnica da desinformação mediada por IA. Elaboração própria com base nas fontes analisadas.

A. Critérios de análise

As fontes foram selecionadas para combinar quatro níveis de evidência. Os relatórios AI100 e *AI Index* oferecem contexto histórico e indicadores gerais; os materiais da disciplina fornecem os conceitos técnicos e éticos empregados na interpretação; os estudos revisados por pares examinam mecanismos de exposição ou efeitos comportamentais; e os registros do AI Incident Database documentam ocorrências concretas. A comparação privilegia fontes que identificam método, contexto e limites. Relatos de incidentes são tratados como evidência de ocorrência e vulnerabilidade, não como prova automática de impacto causal. Do mesmo modo, resultados experimentais são usados dentro do público, da variável e das condições efetivamente estudadas.

II. Resumo do conteúdo

Os materiais da disciplina apresentam a inteligência artificial como um campo dedicado à construção de sistemas capazes de perceber ambientes, representar informações, aprender e selecionar ações. Sua história reúne antecedentes filosóficos, tentativas de formalizar o raciocínio, períodos de otimismo e retração e, mais recentemente, o crescimento do aprendizado de máquina e do aprendizado profundo. Essa trajetória não foi linear: expectativas elevadas alternaram-se com limitações de dados, processamento, métodos e capacidade de generalização.

No paradigma conexionista, redes neurais são compostas por unidades conectadas cujos pesos determinam como os sinais são transformados. A retropropagação calcula como o erro observado na saída se relaciona com os parâmetros das camadas anteriores, permitindo ajustar esses parâmetros durante o treinamento. Em conjunto com maior capacidade computacional, grandes volumes de dados e arquiteturas especializadas, esse processo contribuiu para avanços em visão computacional, processamento de linguagem e geração de conteúdo.

As aulas também definem agentes inteligentes por sua capacidade de perceber o ambiente por sensores e atuar por meio de atuadores. Um agente racional procura maximizar o desempenho esperado com base nas informações disponíveis, sem que racionalidade signifique perfeição. Diferentes estruturas — agentes reflexivos, baseados em modelos, objetivos, utilidade ou aprendizado — são adequadas a diferentes ambientes e níveis de incerteza.

Por fim, os materiais relacionam as aplicações da IA a benefícios e riscos. Na Aula 6, os benefícios da aplicação ética são organizados em seis grupos: personalização e inclusão; eficiência e avanço científico; apoio à tomada de decisão; segurança no trabalho; escalabilidade; e redução de custos com maior foco nas atividades principais [8]. Em contraste, também são discutidos riscos como discriminação, perda de privacidade, desinformação, desigualdade e impactos no trabalho. A ética em IA é apresentada, portanto, como parte do desenvolvimento e do uso dos sistemas, e não como uma preocupação posterior e separada da implementação.

III. Análise crítica

A. Pontos fortes e contribuições

A discussão sobre IA e desinformação não deve considerar a tecnologia apenas como fonte de ameaça. Os benefícios apresentados ajudam a identificar usos positivos diretamente relacionados ao problema. **Personalização e inclusão** permitem adaptar orientações de saúde e educação midiática a idiomas, necessidades de acessibilidade e diferentes níveis de conhecimento. **Eficiência e avanço científico** favorecem a análise de grandes conjuntos de dados e a atualização rápida de evidências. **Apoio à tomada de decisão** pode ajudar profissionais de saúde, jornalistas e autoridades eleitorais a priorizar narrativas que exigem resposta. **Escalabilidade** permite monitorar grandes volumes de conteúdo durante emergências, enquanto **redução de custos e foco** libera especialistas de

tarefas repetitivas para atividades que exigem julgamento humano [8].

As mesmas técnicas de processamento de linguagem e reconhecimento de padrões podem apoiar a identificação de narrativas emergentes, a comparação de alegações com fontes confiáveis, a tradução de orientações e a comunicação rápida com diferentes públicos. Durante uma emergência sanitária, por exemplo, o volume de publicações torna inviável a leitura manual de todo o conteúdo. Sistemas automatizados podem organizar mensagens por tema, detectar crescimento repentino de dúvidas e direcionar especialistas para os pontos que exigem esclarecimento. A OMS relata o emprego de uma plataforma de escuta digital capaz de identificar grandes volumes de documentos relacionados a eventos de vacinação, permitindo investigação e comunicação mais rápidas. A própria organização ressalta, contudo, que aprendizado de máquina e processamento de linguagem não substituem a avaliação de especialistas.

Em processos eleitorais, a IA também pode ampliar acessibilidade, traduzir debates, resumir propostas e auxiliar a verificação de materiais. Técnicas de detecção podem indicar manipulações em áudio ou vídeo, enquanto ferramentas de procedência podem registrar a origem e as alterações de um arquivo. Esses usos mostram que o problema não reside em uma propriedade moral da IA, mas nos objetivos, dados, incentivos e condições de implantação. Um modelo usado para gerar uma falsificação e um modelo usado para detectá-la podem compartilhar fundamentos técnicos semelhantes, embora participem de relações sociais opostas.

Outra contribuição importante é tornar visível a cadeia da informação. A análise não pode terminar no conteúdo falso. É necessário perguntar quem o produziu, quais sistemas permitiram sua criação, como foi distribuído, por que recebeu destaque e quais grupos foram expostos. Essa perspectiva evita atribuir toda a responsabilidade ao usuário final e permite identificar pontos de intervenção: limitações à clonagem não consentida, sinalização de conteúdo sintético, redução da recomendação de fontes reincidentes, acesso de pesquisadores aos dados das plataformas e respostas públicas rápidas.

B. Limitações, controvérsias e lacunas

O primeiro estudo de caso demonstra que a IA pode contribuir para a desinformação sem gerar uma única frase. Juneja e Mitra [3] auditaram a busca e as recomendações da Amazon em temas relacionados a vacinas. Na etapa sem personalização, 10,47% dos resultados analisados promoviam produtos com desinformação de saúde, e itens desinformativos apareciam acima de materiais de contestação em determinadas ordenações. Na etapa personalizada, contas que interagiam com esses produtos recebiam mais recomendações semelhantes. O Incidente 139 do *AI Incident Database* registra esse episódio como amplificação algorítmica de produtos contendo desinformação sobre vacinas [1].

Esse caso é relevante porque desloca o foco da geração para a recomendação. O sistema não precisava compreender ou defender a alegação antivacina. Bastava otimizar sinais como cliques, histórico e probabilidade de interação. Do ponto de vista restrito da plataforma, recomendar algo semelhante ao que o usuário consultou pode ser racional. Do ponto de vista da saúde pública, porém, repetir conteúdo enganoso pode reforçar uma visão inicial e reduzir o contato com fontes de correção. Surge, portanto, um conflito entre a medida de desempenho do agente e o bem-estar coletivo.

O estudo tem limites que precisam ser reconhecidos. Uma auditoria de plataforma identifica padrões de exposição, não mede diretamente se alguém passou a acreditar na alegação ou deixou de se vacinar. Além disso, o resultado se refere ao funcionamento observado naquele período, e plataformas alteram continuamente seus algoritmos e políticas. Para avaliar consequências comportamentais, é necessário recorrer a outro tipo de evidência. Loomba et al. [5] realizaram um experimento com participantes do Reino Unido e dos Estados Unidos e observaram que a exposição a desinformação sobre vacinas contra a COVID-19 reduziu a intenção declarada de vacinação entre pessoas inicialmente decididas a se vacinar. O experimento oferece evidência causal sobre intenção, mas intenção declarada ainda não equivale à vacinação efetivamente realizada. As duas fontes são, assim, complementares: a auditoria mostra um mecanismo de exposição; o experimento estima uma consequência possível dessa exposição.

O segundo estudo de caso envolve produção sintética e tentativa direta de interferir na participação eleitoral. Em janeiro de 2024, eleitores de New Hampshire receberam uma ligação automatizada que imitava a voz de Joe Biden e os orientava a não participar da eleição primária. O Incidente 628 do AIID reúne quinze relatos e registra que investigadores atribuíram a criação e distribuição da falsificação a agentes humanos identificados [2]. A classificação do próprio banco é instrutiva: a causa é humana, a intenção é deliberada e a IA aparece como meio utilizado depois da implantação. Isso sustenta a posição de que a tecnologia não deve absorver a responsabilidade de quem planeja uma ação de supressão eleitoral.

Ao mesmo tempo, o episódio mostra que afirmar apenas que "o ser humano é responsável" é insuficiente. A clonagem de voz reduziu a dificuldade de imitar uma figura pública, e a infraestrutura de chamadas permitiu distribuir a mensagem. O dano potencial nasceu da composição entre modelo, dados de voz, operador político, empresa de telecomunicações e ausência inicial de verificação para o destinatário. O caso também revela a importância da resposta institucional: investigação, rastreamento, divulgação pública e responsabilização podem limitar a propagação e desestimular repetições.

Não há, porém, evidência de que essa ligação tenha alterado o resultado da primária. O próprio conjunto de relatos do AIID inclui a avaliação de uma autoridade eleitoral de que a chamada provavelmente não afetaria o resultado. Essa distinção é essencial para evitar alarmismo. Um incidente pode demonstrar intenção, capacidade e vulnerabilidade sem demonstrar efeito eleitoral agregado. Da mesma forma, a presença de deepfakes em vários países não permite concluir automaticamente que eleitores foram enganados em massa.

Comportamento político depende de identidade partidária, confiança prévia, relações sociais, meios de comunicação e muitas outras variáveis.

Há ainda uma consequência menos mensurável: a erosão geral da confiança. Quando falsificações se tornam conhecidas, pessoas podem passar a desconfiar também de gravações autênticas. Um candidato pode alegar que uma evidência verdadeira foi fabricada, fenômeno frequentemente chamado de "dividendo do mentiroso". Mesmo uma falsificação pouco persuasiva pode, portanto, contribuir para um ambiente em que nenhuma prova audiovisual parece conclusiva. Medir esse efeito exige estudos longitudinais e acesso a dados que, muitas vezes, permanecem sob controle privado das plataformas.

A Tabela I coloca os dois casos sob os mesmos critérios. Ela evidencia que ambos documentam mecanismos relevantes, mas respondem a perguntas distintas e não autorizam inferências equivalentes sobre comportamento coletivo.

C. Comparação entre fontes

Os relatórios analisados convergem na ideia de que capacidade técnica não garante benefício social, mas diferem no momento histórico, no método e na intensidade atribuída aos riscos. O AI100 de 2016 [9] descreve sistemas especializados e examina aplicações em transporte, saúde, educação, segurança e trabalho. Sua perspectiva é predominantemente prospectiva: busca imaginar como a IA afetaria uma cidade norte-americana até 2030. A desinformação produzida por modelos generativos ainda não ocupa posição central, o que é coerente com as capacidades disponíveis naquele momento.

O AI100 de 2021 [4] representa uma mudança significativa. Além dos avanços em linguagem, visão e recomendação, o relatório discute falta de transparência, polarização, desigualdade e uso da IA a serviço da desinformação. Também critica o tecnossolucionismo, isto é, a tendência de tratar a IA como resposta suficiente para problemas que envolvem instituições, interesses e relações sociais. Essa crítica é diretamente aplicável aos dois casos. Classificar automaticamente conteúdo não resolve a desconfiança em autoridades sanitárias; detectar um deepfake não elimina os incentivos políticos para produzi-lo.

A série *AI Index* [6] acrescenta uma perspectiva quantitativa. As edições de 2021 a 2025 acompanham desempenho, investimento, adoção, incidentes, políticas e práticas de IA responsável. A transformação de capítulos chamados "ética técnica" em uma abordagem mais ampla de "IA responsável" sugere o reconhecimento de que riscos não podem ser reduzidos a propriedades internas do modelo. Entretanto, os indicadores do *AI Index* não substituem estudos de caso. Contagens de incidentes dependem de visibilidade pública, cobertura jornalística e critérios de inclusão; um aumento pode representar mais danos, melhor detecção ou ambos.

As fontes de saúde fornecem evidência mais próxima do comportamento individual. A OMS caracteriza a infodemia como um problema de excesso, qualidade e confiança na informação, enquanto Loomba et al. medem mudança na intenção de vacinação após exposição controlada. Juneja e Mitra, por sua vez, revelam como uma plataforma pode organizar a exposição. Há convergência entre elas: o dano não depende apenas da existência de uma alegação falsa, mas de sua repetição, apresentação e circulação. A divergência está no objeto medido — ambiente informacional, intenção comportamental ou recomendação algorítmica — e impede que os resultados sejam somados como se representassem a mesma variável.

No domínio eleitoral, a evidência é ainda mais incerta. O AI100 2021 alerta para ameaças à democracia; o AIID registra casos concretos de falsificação e influência; estudos experimentais recentes demonstram que diálogos com modelos podem alterar preferências declaradas. Em contraste, pesquisas de campo em ambientes específicos encontraram baixa prevalência de conteúdo político viral claramente gerado por IA. Em vez de invalidar uma das posições, essa divergência esclarece níveis distintos de análise. Os modelos possuem capacidade de persuadir em condições experimentais, atores maliciosos já os utilizaram e incidentes são documentados, mas a prevalência e o efeito agregado sobre eleições continuam difíceis de estimar.

Essa comparação evita dois extremos. O primeiro seria afirmar que a IA inevitavelmente destrói a democracia e a confiança nas vacinas. O segundo seria considerar o risco irrelevante enquanto não houver prova de uma eleição ou campanha de vacinação inteiramente determinada por IA. Uma análise proporcional reconhece capacidade, incidentes e efeitos experimentais, ao mesmo tempo que explicita limites metodológicos e evita transformar possibilidade em certeza.

D. Aplicações e avanços necessários

As respostas precisam atuar em diferentes pontos da cadeia. No nível da produção, serviços de geração e clonagem podem exigir consentimento para imitar pessoas, limitar usos de alto risco e manter registros adequados para investigação. Marcas-d'água e metadados de procedência podem ajudar, embora sejam removíveis e dependam de adoção ampla. Detectores não devem ser tratados como solução definitiva, pois podem errar, tornar-se obsoletos e produzir acusações injustas.

No nível da distribuição, plataformas devem considerar a qualidade e o risco da informação em seus objetivos, especialmente em saúde e eleições. Isso não significa que um sistema deva decidir sozinho o que é verdadeiro. Significa oferecer contexto, reduzir incentivos à repetição coordenada, tornar anúncios e conteúdo sintético identificáveis e criar canais rápidos para autoridades eleitorais, sanitárias e pesquisadores. Auditorias independentes são necessárias para verificar se políticas declaradas correspondem ao comportamento real dos algoritmos.

No nível social, letramento midiático e comunicação pública continuam indispensáveis. Orientar o público apenas a "olhar os dedos" de uma imagem ou procurar falhas de voz é uma estratégia frágil, porque a qualidade técnica evolui. É mais robusto verificar origem, data, contexto, confirmação por fontes independentes e possíveis interesses do emissor. Em vacinação, respostas devem reconhecer dúvidas legítimas e comunicar incertezas sem abrir espaço para equivalência

Tabela I
COMPARAÇÃO CRÍTICA DOS ESTUDOS DE CASO

Dimensão	Desinformação sobre vacinas	Desinformação eleitoral	Leitura crítica
Caso principal	Recomendações da Amazon associadas a produtos com desinformação sobre vacinas, registradas no Incidente 139 [1], [3].	Ligação automatizada com voz clonada de Joe Biden antes da primária de New Hampshire, registrada no Incidente 628 [2].	Um caso envolve amplificação algorítmica; o outro, produção sintética e distribuição deliberada.
Mecanismo de IA	Busca, ordenação, personalização e recomendação com base em sinais de interação.	Clonagem de voz combinada a chamadas automatizadas e imitação de autoridade política.	"Uso de IA" não é um mecanismo único; diferentes componentes exigem controles distintos.
Evidência disponível	Auditoria de plataforma identifica padrões de exposição; experimento separado encontra redução da intenção declarada de vacinação após exposição controlada [5].	Relatos e investigação documentam mensagem, intenção de desmobilização e agentes envolvidos.	Auditorias medem o sistema observado; experimentos estimam efeitos em condições delimitadas; incidentes comprovam ocorrência.
O que não foi demonstrado	Que cada recomendação mudou crenças ou que a exposição resultou necessariamente em recusa efetiva da vacina.	Que a chamada mudou votos, participação agregada ou resultado da eleição.	Capacidade, exposição e impacto causal devem permanecer analiticamente separados.
Resposta prioritária	Auditoria de recomendadores, acesso a dados, comunicação sanitária e redução de incentivos à repetição.	Proveniência de áudio, rastreamento, resposta eleitoral rápida e responsabilização dos operadores.	Prevenção eficaz combina medidas técnicas, transparência, instituições públicas e supervisão humana.

falsa entre evidências. Em eleições, desmentidos precisam ser rápidos, acessíveis e politicamente independentes.

Por fim, a governança deve preservar liberdade de expressão e pesquisa. Regras excessivamente vagas podem remover crítica, sátira ou conteúdo legítimo; regras inexistentes deixam o custo do dano para indivíduos e instituições públicas. A resposta adequada combina transparência, devido processo, avaliação proporcional ao risco e responsabilização humana. Não existe uma única solução técnica para um problema que também é econômico, político e cultural.

IV. INTEGRAÇÃO COM O CONTEÚDO DA AULA

O tema amplia quatro pontos trabalhados nas aulas. O primeiro é histórico. A retropropagação foi importante para o aprendizado de representações em redes neurais, mas não conduz inevitavelmente à desinformação. Entre o algoritmo de treinamento e uma chamada eleitoral falsa existe uma longa cadeia de escolhas técnicas e humanas. Essa constatação evita confundir fundamento computacional com destino social.

O segundo ponto é o conceito de agente. Um recomendador percebe ações como busca, clique e compra e atua ordenando novos itens. No caso da Amazon, essa interação forma um ciclo: o comportamento do usuário modifica as recomendações, e as recomendações modificam o ambiente informacional do usuário. Assim, o ambiente não é apenas observado pelo agente; ele é parcialmente produzido por suas ações.

O terceiro ponto é a racionalidade. Um agente pode maximizar engajamento ou probabilidade de consumo e ser racional em relação à métrica definida. Isso não torna o resultado socialmente desejável. Se confiabilidade, diversidade informacional e danos previsíveis não entram na medida de desempenho, a eficiência do sistema pode ampliar desinformação. A escolha da métrica é, portanto, também uma escolha de valores.

O quarto ponto é a ética. O slide da Aula 6 mostra que personalização, eficiência, apoio à decisão, escalabilidade e redução de custos são benefícios quando orientados à inclusão, à ciência e ao bem-estar. Os estudos de caso demonstram, porém, que essas propriedades não são positivas por si mesmas. A personalização pode adaptar uma orientação confiável ou aprofundar uma bolha de desinformação; a escalabilidade pode acelerar comunicação sanitária ou distribuir uma voz eleitoral falsa; a redução de custos pode beneficiar serviços públicos ou facilitar campanhas maliciosas. O caráter ético depende da finalidade, das restrições e da distribuição dos resultados.

Considero que a IA não deve ser responsabilizada isoladamente pela desinformação, pois pessoas e organizações escolhem como produzir, recomendar e divulgar conteúdos. Entretanto, sistemas que ampliam alcance ou persuasão não podem ser tratados como ferramentas neutras. Desenvolvedores, plataformas, usuários e instituições públicas compartilham responsabilidades proporcionais ao controle que exercem e aos riscos que poderiam prever. Essa posição modifica minha compreensão inicial: o risco não está apenas em uma resposta falsa gerada por um modelo, mas na arquitetura completa que transforma uma falsidade em influência social.

V. CONCLUSÃO

A inteligência artificial altera a desinformação em pelo menos duas dimensões: facilita a produção de conteúdo plausível e participa da seleção e amplificação do que circula. O caso das vacinas mostra que recomendadores podem aumentar a exposição a materiais enganosos e que essa exposição pode afetar intenção declarada. O caso eleitoral mostra que a clonagem de voz permite personificar candidatos e tentar influenciar a participação. Nenhum dos casos, isoladamente, prova mudança efetiva em vacinação ou resultado eleitoral, mas ambos documentam mecanismos, intenções e vulnerabilidades que justificam prevenção.

A comparação entre AI100, *AI Index*, OMS, estudos empíricos e AIID confirma que desempenho técnico e impacto social pertencem a níveis diferentes. A IA não age fora de

instituições: plataformas definem métricas, operadores escolhem objetivos, públicos interpretam mensagens e autoridades respondem. Por isso, responsabilizar apenas o modelo oculta decisões humanas; considerá-lo neutro ignora a escala e a influência que sua arquitetura possibilita.

Enfrentar o problema exige combinar procedência de conteúdo, auditoria de recomendadores, resposta institucional, educação midiática e responsabilização proporcional. Estudos futuros devem medir não somente a quantidade de conteúdo sintético, mas sua exposição, credibilidade, repetição e efeito sobre comportamentos reais. Também devem investigar quais intervenções reduzem danos sem concentrar poder de censura nas plataformas. A questão decisiva não é apenas se uma máquina consegue fabricar uma falsidade, mas como uma sociedade organiza os sistemas que determinam quem a verá, quantas vezes e com quais possibilidades de contestação.

Referências

[1] Artificial Intelligence Incident Database, "Incident 139: Amazon's search and recommendation algorithms found by auditors to have boosted products that contained vaccine misinformation," 2021. [Online]. Available: https://incidentdatabase.ai/cite/139. Accessed: Aug. 27, 2026.

[2] Artificial Intelligence Incident Database, "Incident 628: Fake Biden voice in robocall misleads New Hampshire Democratic voters in 2024 primary election," 2024. [Online]. Available: https://incidentdatabase.ai/cite/628. Accessed: Aug. 27, 2026.

[3] P. Juneja and T. Mitra, "Auditing e-commerce platforms for algorithmically curated vaccine misinformation," in *Proc. CHI Conf. Human Factors in Computing Systems*, 2021, doi: 10.1145/3411764.3445250.

[4] M. L. Littman et al., *Gathering Strength, Gathering Storms: The One Hundred Year Study on Artificial Intelligence (AI100) 2021 Study Panel Report*. Stanford, CA, USA: Stanford Univ., 2021.

[5] S. Loomba et al., "Measuring the impact of COVID-19 vaccine misinformation on vaccination intent in the UK and USA," *Nature Human Behaviour*, vol. 5, pp. 337–348, 2021, doi: 10.1038/s41562-021-01056-1.

[6] N. Maslej et al., *The AI Index 2025 Annual Report*. Stanford, CA, USA: Institute for Human-Centered AI, Stanford Univ., 2025.

[7] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning representations by back-propagating errors," *Nature*, vol. 323, pp. 533–536, 1986, doi: 10.1038/323533a0.

[8] F. A. Soares, *Inteligência Artificial: Ética, Riscos e Benefícios*. Brasília, DF, Brasil: Universidade de Brasília, material da disciplina FGA0221, 2026.

[9] P. Stone et al., *Artificial Intelligence and Life in 2030: One Hundred Year Study on Artificial Intelligence, Report of the 2015 Study Panel*. Stanford, CA, USA: Stanford Univ., 2016.

[10] World Health Organization, *5th Virtual WHO Infodemic Management Conference Meeting Report: Steps Towards Measuring the Burden of Infodemics*. Geneva, Switzerland: WHO, 2022.