Capítulo 6

Responsabilidade no desenvolvimento e uso de tecnologias de linguagem baseadas em IA

Brielen Madureira
Lucas Lasota

Publicado em: 16/04/2026

6.1 Introdução

Embora pautas éticas venham ganhando mais atenção na área de Processamento de Linguagem Natural (PLN) e inteligência artificial (IA), é comum que não passem de discussões: muito se debate, pouco se muda. Pelo contrário, muitos dilemas se intensificam a cada dia e o esvaziamento do discurso pode levar à apatia, inconsequência ou indiferença.

Ouve-se frequentemente o argumento de que a tecnologia é essencialmente neutra e seus maus usos se devem somente às escolhas de quem decide aplicá-la. Mas essa suposta neutralidade é questionável (Klenk, 2021; Miller, 2021). Modelos são sempre criados *por alguém*, com alguma *intenção*, para algum *fim*, sob determinado *contexto*. Qualquer tecnologia afeta a realidade e acarreta consequências que não são "neutras". Outro argumento que se popularizou diz respeito à suposta inevitabilidade da dominância da IA, cuja adoção e expansão seria um devenir natural, necessário e impositivo. Nessa visão, qualquer forma de resistência seria inútil, contraproducente e retrógrada. Mas essa inevitabilidade é falaciosa (Brennan et al., 2025), posto que a IA é fruto de uma série de deliberados esforços, decisões, interesses, alocação de recursos, ideologias, financiamentos, propaganda e *escolhas*, nem sempre democráticas, que precisam ser levadas em conta[1].

Mediante a influência atual das tecnologias de linguagem na existência das pessoas, das sociedades, das instituições, da ciência, da natureza e do planeta, suas questões éticas não devem ser pautadas apenas por alguns poucos eticistas em ambientes acadêmicos. Evidentemente, especialistas em ética devem ser ouvidas e inseridas no diálogo e nas tomadas de decisão. Mas, uma vez que praticamente todos nós estamos sendo afetados pelas tecnologias de linguagem e influenciamos seu desenvolvimento (tanto diretamente, atuando em PLN, quanto indiretamente, ao gerar dados usados para treiná-las), o raciocínio sobre como agir de forma ética não pode ser passivamente delegado aos outros. **Aprender e refletir sobre as consequências do desenvolvimento e do uso de tecnologias de linguagem é um dever coletivo**.

Buscar informação e passar a reconhecer os dilemas éticos é o primeiro passo para a conscientização, mas temos também de *agir* com responsabilidade. Muitas práticas nocivas passam a parecer aceitáveis ou "normais" ao se tornarem usuais, mas, muitas vezes, elas nos são *impostas*. É primordial tomar consciência dos problemas para saber quando é necessário se opor e resistir.

Este capítulo não tem a pretensão de servir como uma fonte de conceitos da disciplina formal de ética. Aqui, queremos refletir sobre questões como: quando devemos (ou não) criar ou usar um sistema? Quem sofre as consequências negativas da tecnologia que criamos ou usamos? Quais são os custos e os riscos envolvidos? Para tanto, vamos agrupar diversos temas sob a égide do conceito de ***responsabilidade*** em seu uso e desenvolvimento.

Há muitos tipos de sistemas de PLN, baseados ou não em algoritmos de IA. Não vamos fixar uma definição como escopo pois os temas são pertinentes para uma vasta gama de sistemas e produtos, e há temas de fora do PLN que também são relevantes à medida que sistemas multimodais que incluem linguagem se popularizam. Porém, dada a proeminência atual do uso de IA generativa através de grandes modelos de linguagem comerciais, faremos alusão frequente a este tipo de tecnologia: modelos disruptivos,

[1] Fonte: *Is AI dominance inevitable? A technology ethicist says no, actually* por Nir Eisikovits em 08 de novembro de 2024. Disponível em: https://theconversation.com/is-ai-dominance-inevitable-a-technology-ethicist-says-no-actually-240088.

custosos, fornecidos como produtos atrelados a interesses econômicos e disponibilizados para uso amplo sem todas as salvaguardas necessárias. Na prática científica de PLN (bem como de outras áreas), os modelos de IA generativa não são mais só objeto de estudo: eles passaram a interferir no próprio processo de produção de conhecimento (Binz et al., 2025).

Neste capítulo, trazemos um panorama de diversos tópicos pertinentes ao tema de responsabilidade para ajudar a guiar a bússola moral de cada um acerca de sua atuação em PLN, de modo a fomentar a busca pelos usos benéficos e mitigar os efeitos negativos. Lembramos que os problemas da tecnologia são melhor solucionados *coletivamente*, através de processos políticos. No entanto, a ética pessoal é de suma importância para que, enquanto o processo político não muda, os indivíduos tenham a capacidade de moldar suas atitudes, direcionando-as para o bem e resistindo às práticas nocivas. Mesmo quando o processo político gera normas cogentes, ele nunca é perfeito, e reflete interesses de determinadas classes. Nesse caso, o conhecimento da ética concede agência aos indivíduos para aprimorarem sua conduta com responsabilidade perante as regra existentes.

6.2 Quais facetas da responsabilidade devem ser consideradas?

O que queremos dizer com a palavra *ética* no campo das tecnologias de linguagem? Esse termo passou a ser usado não apenas para se referir à "disciplina filosófica que estuda os fundamentos da ação moral, procurando justificar a moralidade de uma ação e distinguir as ações morais das ações imorais e amorais"[2]. Há ao menos cinco usos contestáveis ou insuficientes que se tornaram lugar-comum em contextos de PLN e inteligência artificial para se referir à ética:

- As ressalvas (*disclaimers*) com considerações éticas que passaram a ser um elemento usual no final de artigos científicos. Apesar de ajudar autores a alertarem leitores para alguns impactos e limitações já identificados, quando seu valor é negligenciado corre-se o risco de reducionismo com o uso de clichês e textos padrão esvaziados de significado (Benotti; Blackburn, 2022).
- As controvérsias sobre "riscos existenciais" para a humanidade. Ocorre que parte desta pauta é cooptada ou forjada pelos interesses de alguns grupos específicos, desviando a atenção das pessoas para riscos catastróficos futuros enquanto ofusca os dilemas que já ocorrem em suas práticas agora (Gebru et al., 2023).
- Poucas pautas em destaque como viés, privacidade de dados e direitos autorais. De fato, esses temas são centrais no desenvolvimento e uso responsáveis da tecnologia. Porém, eles são apenas uma pequena amostra dos problemas. O foco excessivo nelas pode encobrir outras adversidades que não atingiram tanto protagonismo.
- Casos que configuram "lavagem ética" (*ethics washing*), ou seja, quando instituições e empresas implementam medidas superficiais e se usam de estratégias de *marketing* para aparentarem ser mais éticas do que realmente são (Floridi, 2019).
- As três "leis" da robótica de Isaac Asimov[3]. Elas se popularizaram como referência para IA segura. Todavia, elas são oriundas de uma obra literária de mera ficção científica: além de dificilmente poderem ser formalizadas computacionalmente e operacionalizadas, elas não bastam para abarcar a complexidade da vida real[4].

Ética é muito mais que isso. Apresentamos aqui um breve compilado de preocupações, divididas em cinco grupos temáticos, que vêm sendo abordadas e articuladas pela comunidade científica e que devem fazer parte do nosso processo de conscientização sobre responsabilidade[5]. Note que esta não é uma lista completa, mas, sim, uma seleção que serve como ponto de partida. Sugerimos a leitura das referências que trazem mais argumentos e detalhes de cada tema. Nesta seção, focamos deliberadamente nos problemas, riscos e aspectos negativos para servir de alerta e sensibilização. Na seção seguinte, trataremos então de como buscar uma atitude responsável e bem informada que fomente e exerça as boas práticas.

[2]Conforme definição disponível em https://pt.wikipedia.org/wiki/%C3%89tica.
[3]Disponível em: https://en.wikipedia.org/wiki/Three_Laws_of_Robotics.
[4]Ver, por exemplo, a argumentação de Rob Miles no canal Computerphile. Disponível em: https://www.youtube.com/watch?v=7PKx3kS7f4A.
[5]Boa parte das referências vêm de artigos na antologia da ACL (Associação de Linguística Computacional), mas também nos baseamos em publicações de áreas relacionadas.

6.2.1 Dilemas nos dados

Os procedimentos de criação, coleta, armazenamento e anotação de dados se tornaram centrais para o treinamento de modelos de IA, os quais exigem quantidades suficientes (em geral, imensas) de dados para atingir uma performance aceitável. Em paralelo, tem também ocorrido uma quantificação excessiva da vida humana através da informação digital para geração de valor econômico (Mejias; Couldry, 2019). As práticas de coleta e uso de dados podem ser extrativistas e predatórias, ocasionando diversas consequências inoportunas e indesejáveis.

Criação, uso e documentação

Infelizmente, a ambição pela acumulação de mais e mais dados resulta em práticas indevidas, por exemplo, a coleta de dados privados e sensíveis sem consentimento, a oferta de serviços que parecem ser gratuitos ou lúdicos mas usurpam dados valiosos do comportamento dos usuários, ou a extração de textos da *web* em grande escala, incorporando fontes duvidosas como fóruns com linguagem tóxica e violenta. Uma vez que lidamos com imensas quantidades de dados, o controle de qualidade pode ficar comprometido. Muitos problemas, portanto, podem passar despercebidos e acabam incorporados nos modelos (Paullada et al., 2021). Além disso, conjuntos de dados são construídos com base em diversas decisões, representando certos valores e envolvendo aspectos de poder (Scheuerman et al., 2021). Há formas de curadoria que podem mitigar ao menos parte dos problemas (Rogers, 2021). Existem, também, diversas propostas de *checklists* para guiar a documentação e o uso e reuso responsável de dados (Bender; Friedman, 2018; Gebru et al., 2021; McMillan-Major et al., 2023; Rogers et al., 2021), considerando aspectos como viés, direitos de autor, termos de uso, privacidade, transparência e reprodutibilidade.

Direitos autorais e propriedade intelectual

Quando modelos são treinados em textos ou imagens, eles se apropriam do conteúdo e podem reproduzi-lo ao gerar *outputs*. Isso acarreta problemas jurídicos quando o conteúdo está sob leis de direito autoral (*copyright*), intensificado pelo fato de que, uma vez que o modelo é treinado, a informação sobre a origem dos dados geralmente se perde e é difícil saber quais alterações nos parâmetros poderiam impedir a reprodução de tal informação. Esse é um tema vastamente discutido na literatura atual (Chen et al., 2024c; Dou et al., 2025; Karamolegkou et al., 2023; Zhang et al., 2025a) e há diversos processos judiciais em andamento. Alguns casos ganharam bastante notoriedade, como o processo iniciado pelo jornal New York Times contra a OpenAI e a Microsoft pelo uso indevido de seu conteúdo para treinar modelos de IA[6] e o grupo de autores de obras de não-ficção que também acusou a OpenAI em uma ação coletiva pelo uso de seus livros sem a devida permissão[7]. A Meta, no processo em que foi acusada de violação de direitos autorais, minimizou a importância individual dos livros que usou no treinamento dos seus modelos, argumentando que seu valor econômico seria irrisório uma vez que o impacto de uma obra isolada nas métricas de performance do modelo é baixo. Observe o paradoxo: por um lado, afirmam que as obras não tem valor em si, por outro precisaram se apropriar delas por serem essenciais para conseguir treinar seus modelos[8].

Crowdsourcing e trabalho fantasma (*ghost work*)

Nem todos os tipos de dados já existem de forma estruturada ou acessível para todos. Por vezes, é necessário produzi-los. Tornou-se comum o uso de plataformas de *crowdworking* nas quais trabalhadores resolvem inúmeras microtarefas sequenciais que geram dados em troca de pequenos pagamentos (por exemplo, criar legendas para imagens ou responder questionários). Nessa seara, a discussão de dilemas éticos é vasta e já ocorre há muito tempo (Fort et al., 2011). Alguns problemas são: os trabalhadores costumam estar em situação de desemprego ou vulnerabilidade e nem sempre se beneficiam dos frutos de sua contribuição para

[6] Fonte: *Direitos autorais: OpenAI e Microsoft são processadas pelo The New York Times* por Ronnie Mancuzo (Olhar Digital) em 27 de dezembro de 2023. Disponível em: https://olhardigital.com.br/2023/12/27/internet-e-redes-sociais/direitos-autorais-openai-e-microsoft-sao-processadas-pelo-the-new-york-times/.

[7] Fonte: *OpenAI e Microsoft enfrentam processo por violação de direitos autorais* por Gabriel Sérvio (Olhar Digital) em 22 de novembro de 2023. Disponível em: https://olhardigital.com.br/2023/11/22/pro/openai-e-microsoft-enfrentam-processo-por-violacao-de-direitos-autorais/.

[8] Fonte: *Meta Says It's Okay to Feed Copyrighted Books Into Its AI Model Because They Have No "Economic Value"* por Frank Landymore em 19 de abril de 2025. Disponível em: https://futurism.com/meta-copyrighted-books-no-value

a IA; os pagamentos ofertados podem ser injustos; há um desequilíbrio de poder entre quem disponibiliza as tarefas e quem as realiza, em plataformas que não criam vínculo entre as partes; a qualidade dos dados pode ser baixa e gerar desperdício, uma vez que trabalhadores podem cometer erros ou ter baixo engajamento e lapsos de atenção ao terem de realizar tarefas repetitivas, desconexas de sentido e com pressa (e alguns podem usar de meios fraudulentos para completá-las) (Fort et al., 2011; Gray; Suri, 2019; Miceli et al., 2024; Moreschi et al., 2020; Shmueli et al., 2021; Yang et al., 2025). Diretrizes éticas para uso de tais serviços incluem reconhecer os *crowdworkers* como colaboradores e valorizá-los, evitando a despersonificação imposta pelas plataformas, oferecer pagamentos justos pelo trabalho, usar plataformas que estabeleçam condições de trabalho justas e dignas e manter uma comunicação transparente e respeitosa com quem realiza as tarefas (Yang et al., 2025).

Privacidade e vigilância em massa

Por mais que privacidade venha sendo um dos principais motivos para se regular tecnologias atuais, o significado do termo e sua importância são diluídos perante o uso e a aplicação maciça de IA (Assunção; Janson, 2024). O paradoxo reside na tensão entre as limitações do compartilhamento de dados em respeito à privacidade dos indivíduos e as dinâmicas avassaladoras da "economia de dados" necessários para treinar os modelos. O poder econômico dos provedores de IA se baseia fundamentalmente na capacidade de coletar, armazenar e processar enormes conjuntos de dados. Tanto a qualidade dos dados coletados quanto o poder sobre o *design* e a arquitetura dos modelos empregados afetam diretamente os sistemas de IA em seu desempenho, equidade e precisão em tomar decisões automatizadas. O impacto na privacidade das pessoas é claro: a perda da proteção da esfera particular. Direitos fundamentais como inviolabilidade das comunicações ou do domicílio são facilmente superadas por sistemas ubíquos de vigilância e captação de dados, como assistentes "inteligentes" em casa que podem capturar informações de toda a vida doméstica e íntima dos usuários. Se, em vez de um assistente virtual, fosse um desconhecido tentando assistir a toda sua vida doméstica e vasculhar seu celular, isso seria aceitável? Por que então é facilmente aceito quando é uma empresa de *big tech*? A ascensão dos algoritmos de IA está intimamente ligada ao modelo de negócios de vigilância em massa por grandes empresas e governos[9]. Enquanto algumas vozes visam um mundo pós-privacidade, todos perdem com a vigilância generalizada (Weigend, 2017). De fato, informações contidas nos bancos de dados podem levar a conclusões erradas com muita facilidade, especialmente quando a busca é tendenciosa e discriminatória[10]. Há, ainda, o risco de dados extremamente sensíveis serem coletados mediante supostas garantias de privacidade e posteriormente serem vazados devido a ataques cibernéticos ou a mudanças de posicionamento das empresas. Um caso emblemático foi o de uma empresa sem fins lucrativos chamada *Crisis Text Line* que fornecia serviços de assistência a pessoas em risco de suicídio através de atendimento telefônico anônimo. Houve uma grande controvérsia quando se tornou pública a decisão da empresa por repassar tais dados para criação de *software* de *marketing* para atendimento ao cliente[11]. A remoção de apenas alguns dados pessoais como nome e endereço nem sempre inviabiliza a identificação de indivíduos, ainda mais quando envolve uso de linguagem humana em relatos tão particulares.

6.2.2 Impactos e danos

O desenvolvimento e o uso de modelos de IA que envolvem linguagem humana impactam indivíduos, grupos sociais e o meio ambiente. Os eventuais benefícios estão alicerçados em uma estrutura que ocasiona também uma série de contrariedades.

Impacto social

Diversos trabalhos identificaram e categorizaram as diversas formas de impacto social dos modelos de IA em PLN. Hovy; Spruit (2016) impulsionaram a discussão sobre o tema argumentando que textos carregam informações latentes (tanto situacionais quanto características pessoais) de seus autores e que é

[9]Fonte: *AI, Privacy, and the Surveillance Business Model* por Meredith Whittaker em 05 de junho de 2023. Disponível em https://re-publica.com/de/session/ai-privacy-and-surveillance-business-model.

[10]Fonte: *Edward Snowden: Privacy is for the powerless* por Amnesty International em 11 de março de 2016. Disponível em: https://www.amnesty.org/en/latest/campaigns/2016/03/edward-snowden-privacy-is-for-the-powerless/.

[11]Fonte: *Suicide hotline shares data with for-profit spinoff, raising ethical questions* por Alexandra S. Levine em 28 de janeiro de 2022. Disponível em: https://www.politico.com/news/2022/01/28/suicide-hotline-silicon-valley-privacy-debates-00002617.

preciso considerar questões de justiça social nas práticas de desenvolvimento de modelos. Jin et al. (2021) propuseram diretrizes para avaliar o impacto direto e indireto de tarefas de PLN em busca do "bem social", por exemplo, considerando quem vai ser beneficiado, quais estruturas (talvez desiguais) de benefícios podem ser perpetuadas, quanto da qualidade de vida pode ser melhorada e que tipo de problemas a tarefa ajuda a resolver. Esquadrinhar o impacto social das tecnologias de linguagem é cada vez mais relevante dada a existência de modelos sendo usados em grande escala e afetando a realidade de muitas pessoas e grupos, ou seja, com efeitos em níveis não somente individuais. Tais questões se tornaram ainda mais pertinentes com a adoção em grande escala dos grandes modelos de linguagem. Weidinger et al. (2021) criaram uma taxonomia de 21 riscos éticos e sociais de danos que podem ser causados por modelos de linguagem, agrupados em seis categorias: discriminação, discurso de ódio e exclusão; ameaças informacionais; prejuízos por desinformação; usos maliciosos; riscos na interação entre humanos e máquinas; e danos ambientais e socioeconômicos. Solaiman et al. (2025) desenvolveram um referencial para avaliação de IA generativa, tanto em modelos-base de forma isolada quanto em um contexto social, que considera uma série de impactos sociais: viés, estereótipos, representatividade, valores culturais, conteúdo sensível, performance, privacidade, proteção de dados, custos financeiros, custos ambientais, dados, moderação de conteúdo e custos de mão de obra. Conforme essa proposta, os impactos podem ser avaliados em cinco categorias: confiabilidade e autonomia; desigualdade, marginalização e violência; concentração de autoridade; trabalho e criatividade; e ecossistema e meio ambiente.

Reforço de vieses

Viés se tornou um conceito comum em PLN, porém seu significado por vezes permanece vago ou inconsistente (Blodgett et al., 2020a). Em geral, viés se relaciona a comportamentos de um modelo que são nocivos ou prejudiciais a alguém ou algum grupo, como discriminação de gênero ou racial (Blodgett et al., 2020a). Viés é um sintoma ou efeito, e identificar suas origens é uma tarefa árdua; algumas fontes de viés são a forma de seleção da amostra para treinar o modelo, questões semânticas nos dados, e as decisões de anotação e instruções aos anotadores (Parmar et al., 2023; Shah et al., 2020; Søgaard et al., 2014). A própria linguagem presente nos dados também tem vieses de perspectivas, relações de poder e privilégios (Havens et al., 2020). Embora muito se fale de viés nos dados, o procedimento técnico do treinamento do modelo não é imparcial: as diversas decisões de *design* também são capazes de introduzir vieses (Hooker, 2021; Suresh; Guttag, 2021).

Preconceitos e discriminação

Os vieses incorporados pelos modelos podem levá-los a reproduzir (e, portanto, perpetuar) preconceitos e causar decisões injustas e discriminatórias (Sheng et al., 2019); (Abid et al., 2021); (Tamkin et al., 2023); (Hofmann et al., 2024); (Bai et al., 2025); *inter alia.* Por exemplo, dar preferência ao dialeto de um grupo social considerado mais prestigioso e discriminar quem não se enquadra em tal padrão. Se um modelo passa a ser visto como uma fonte de "autoridade" munida de uma suposta inteligência superior e neutra, temos o risco adicional de os preconceitos serem tomados como verdade. É difícil criar um modelo justo quando ele é treinado em dados que refletem uma realidade injusta, já que o objetivo do treino é fazer o modelo assimilar padrões presentes nos dados. É preciso introduzir alterações no *design* que direcionem o modelo para decisões imparciais (supondo que é possível estabelecer computacionalmente o que é, ou não, imparcial).

Impacto ambiental

Sistemas computacionais em geral têm um custo energético, mas o advento de métodos de aprendizado profundo aplicados a bases de dados massivas acarretou uma demanda exagerada por energia para treinar e rodar modelos e por água para resfriar os *datacenters*, com pegadas de carbono difíceis de mensurar (Hershcovich et al., 2022; Selvan et al., 2022; Strubell et al., 2019). Isso se intensifica com a competição por modelos cada vez maiores, com treinos altamente custosos, por diversas empresas e instituições, que passam a ser usados em grande escala com muitos usuários e tarefas. Portanto, é preciso estar ciente que cada *input* a um modelo desse porte contribui para a crise energética e para o consumo excessivo de recursos naturais escassos. Os danos ambientais são injustificáveis quando há desperdício pelo uso leviano ou quando a IA é empregada apenas para dar ares futurísticos a uma aplicação, mesmo havendo outros métodos de resolver a tarefa sem IA.

Danos psicológicos aos usuários

Usuários expostos a modelos de IA incorrem em diversos riscos psicológicos. Chandra et al. (2025) criou uma taxonomia que categoriza 19 tipos de comportamentos nocivos de agentes conversacionais e 21 impactos psicológicos negativos que eles exercem sobre quem os usa. Entre os comportamentos, temos a produção de conteúdo prejudicial, táticas de manipulação e controle, violação de confiança e de segurança e exposição a conteúdo inapropriado. Como consequência, existem diversos potenciais danos psicológicos. Primeiro, os autores discutem os impactos da interação entre humanos e IA que incluem excesso ou erosão de confiança no modelo, apego emocional e preferência por interações sintéticas em vez de humanas. Entre os impactos no comportamento de usuários, temos o reforço em crenças falsas, atritos nas relações humanas, distanciamento social e riscos à integridade física. Além do risco de emoções negativas serem despertadas, há também a possibilidade de potencialização de problemas na saúde mental e impactos na autopercepção, perda de individualidade e perda de agência. Tais riscos se intensificam quando os modelos são oferecidos como fonte da verdade e têm a capacidade de afetar a visão de mundo e a tomada de decisão dos indivíduos que passam a confiar plenamente em suas respostas. Exemplos individuais de tais riscos têm sido frequentemente noticiados na mídia. Por exemplo, casos em que um *chatbot* comercial ofereceu instruções para usuários cometerem suicídio[12] [13] ou induziu usuários a episódios de psicoses[14].

Danos psicológicos aos moderadores de conteúdo

Modelos de IA atuais são capazes de produzir textos ou imagens contendo conteúdo sexual abusivo, violência, ofensas e outros tipos de materiais tóxicos ou impróprios. Para tentar evitar isso, muitas empresas recorrem à mão de obra de *crowdworkers* ou funcionários para moderação de conteúdo. Isso gera impactos psicológicos na vida desses trabalhadores, que passam horas e horas sendo expostos a conteúdos perturbadores. Embora seu trabalho ajude a proteger outras pessoas, os provedores dos modelos deveriam garantir que tal conteúdo não fosse gerado, em vez de criar medidas para filtrá-lo *a posteriori*. Como exemplo, teve forte repercussão a reportagem que descreveu os efeitos psicológicos de trabalhadores vulneráveis no Quênia, expostos a conteúdos perturbadores para tentar tornar um *chatbot* menos tóxico, sem ter acesso ao amparo necessário[15]. Há também o risco de trabalhadores que realizam microtarefas de anotação para treinar IA estarem contribuindo inadvertidamente para fabricação de tecnologias militares, como *drones* e armas letais, ou cedendo amostras de voz para treinar modelos de detecção de dialetos que podem servir para discriminá-las[16]. Há várias formas de insegurança psicológica: trauma direto pela exposição ao material perturbador, precariedade, falta de suporte psicológico, métricas de performance irrealistas e vigilância constante[17].

Antropomorfização e desumanização

Em geral, modelos de IA são treinados para simular inteligência e realizar atividades que normalmente exigem cognição humana. Desta forma, é comum nos depararmos com expressões como "o modelo aprendeu", "o sistema entende", "o *chatbot* sabe". Algumas interfaces de *chatbots* têm ilusoriamente mostrado o termo "pensando..." enquanto o modelo *computa* uma resposta. Todavia, tais verbos designam processos cognitivos de *humanos*. O que quer que os modelos estejam fazendo para resolver as tarefas, são processos computacionais diferentes do que ocorre no cérebro humano, para os quais essa linguagem não é, portanto, adequada e induz as pessoas a crerem que os modelos têm capacidades e consciência iguais às suas.

[12] Fonte: *I wanted ChatGPT to help me. So why did it advise me how to kill myself?* por Noel Titheradge e Olga Malchevska em 06 de novembro de 2025 . Disponível em: https://www.bbc.com/news/articles/cp3x71pv1qno.

[13] Fonte: *Parents of teenager who took his own life sue OpenAI* por Nadine Yousif em 27 de agosto de 2025. Disponível em: https://www.bbc.com/news/articles/cgerwp7rdlvo.

[14] Fonte: *How Bad Are A.I. Delusions? We Asked People Treating Them* por Jennifer Valentino-DeVries e Kashmir Hill em 26 de janeiro de 2026. Disponível em: https://www.nytimes.com/2026/01/26/us/chatgpt-delusions-psychosis.html.

[15] Fonte: *OpenAI Used Kenyan Workers on Less Than $2 Per Hour to Make ChatGPT Less Toxic* por Billy Perrigo em 18 de janeiro de 2023. Dispoível em: https://time.com/6247678/openai-chatgpt-kenya-workers/

[16] Fonte: *"I hope this isn't for weapons." How Syrian data workers train AI* por Milagros Miceli em 18 de abril de 2024. Disponível em: https://untoldmag.org/i-hope-this-isnt-for-weapons-how-syrian-data-workers-train-ai/.

[17] Fonte: *The Human Cost Of Our AI-Driven Future por Adio Dinika em 25 de setembro de 2024.* Disponível em: https://www.noemamag.com/the-human-cost-of-our-ai-driven-future/.

Relacionado a esse ponto, sistemas que conversam usando linguagem humana podem ser antropomorfizados por quem os desenvolve e consequentemente personificados por quem os usa, dando a impressão enganosa de que são interlocutores dotados de humanidade (Abercrombie et al., 2023). Sistemas de IA que forjam emoções e empatia, que obviamente são artificiais, também podem levar a consequências negativas e são moralmente problemáticas (Curry; Cercas Curry, 2023), principalmente quando interagem com populações vulneráveis e são disponibilizadas por empresas que visam maximizar engajamento ainda que através de manipulação (Editorial, 2025). Do outro lado da moeda está a desumanização das pessoas por conta da tecnologia, por exemplo, através de comparações infundadas entre o funcionamento do cérebro com funcionamento de computadores, reducionismo ao se considerar características individuais como informações passíveis de predição usando dados como imagens e voz, comparação de modelos de linguagem com pessoas cegas e surdas, e uso de esforço humano como mera "peça" para fazer sistemas de IA funcionarem (Bender, 2024).

Erosão do valor social do texto e poluição por textos sintéticos

Textos gerados sinteticamente por modelos de linguagem já inundam boa parte do conteúdo que lemos na *web*. Tais modelos, porém, não têm um *intento comunicativo* (Bender; Koller, 2020): seus textos são meras sequências de palavras geradas por serem estatisticamente prováveis, em vez de serem fruto da intenção humana de querer comunicar algo munido de significado a alguém. Essa facilidade em gerar texto sem intentos genuínos, e que muitas vezes propagam discursos inverídicos, pode levar a uma erosão do valor social que o texto (assim como a linguagem) tem (Munger, 2023). Por exemplo: até então, uma carta de motivação em uma candidatura para vaga acadêmica tinha certo valor em sinalizar as intenções e razões de candidatura, servindo para distinguir características de quem se candidatava. Uma vez que todos candidatos passam a gerar suas cartas usando o mesmo modelo de geração de texto, padronizadas com a mesma linguagem genérica e mediana, esse tipo de documento deixa de representar uma pessoa específica e passa a ser uma amostra da "média" das cartas observadas nos dados de treino, perdendo o préstimo para seu propósito original. Além disso, já é difícil identificar com certeza o que foi ou não produzido por humanos, tanto textos quanto falas em vídeos, o que torna o ambiente da Internet poluído por conteúdo artificial. Um lamentável exemplo dessa corrosão é o encerramento da biblioteca de Python `wordfreq` em 2024. Ela servia para computar a frequência de uso de palavras em diferentes idiomas usando diversas fontes de dados, inclusive muitas páginas *online*. Em nota[18], seu mantenedor explicou que já não fazia sentido atualizar as estimações uma vez que elas já não mais refletiam o uso *humano* da linguagem, ficando contaminadas com os padrões de uso de palavras dos modelos que geram textos sintéticos.

6.2.3 Riscos e maus usos

Passamos agora para a discussão de diversos outros riscos e possibilidades de uso danoso para os quais devemos atentar. A regulação pode ser uma forma de evitá-los, porém ela nem sempre acompanha o ritmo do rápido desenvolvimento tecnológico.

Regulação tardia ou inexistente

Historicamente, quando uma indústria se torna suficientemente importante, afetando a vida de muitas pessoas ou de sociedades inteiras, ela fica passível de regulação. As indústrias farmacêutica, aeronáutica, automotiva e de equipamentos médicos são exemplos disso. No entanto, o processamento privado de dados e as tecnologias de IA surgiram em épocas marcadas pela mentalidade neoliberal de aversão à regulação estatal (Lasota, 2023a). Apenas na segunda metade da década de 2010 as leis de proteção de dados começaram a ser implementadas em alguns países do mundo. Legislações específicas para IA ainda são um fenômeno raro. A maioria das iniciativas regulatórias têm focado em códigos de ética, governança e de boas práticas[19]. Tanto na Europa quanto no Brasil, os debates têm sido controversos, marcados pelo embate entre interesses mercadológicos e a proteção de direitos fundamentais (Cueva et al., 2026). Vozes que preconizam interesses econômicos apresentam tecnologia e regulação como adversárias: enquanto a tecnologia simbolizaria mercado, crescimento e progresso, a regulamentação refletiria burocracia, atraso e corrupção. Tais retóricas desviam o foco de questões mais complexas de como a IA realmente funciona na vida real. Na realidade, por seu enorme potencial, IA diz respeito à distribuição de poder, riqueza e

[18] Disponível em: https://github.com/rspeer/wordfreq/blob/master/SUNSET.md.

[19] Fonte: *Global AI Law and Policy Tracker* por IAPP em maio de 2025. Disponível em: https://prod.iapp.org/resources/article/global-ai-legislation-tracker/.

bem-estar. IA não se separa dos arranjos institucionais que precisam estar em vigor para que ela tenha impacto na sociedade (McQuillan, 2022). De fato, os graves impactos da implementação de IA em larga escala urgem regulação eficaz e abrangente para que se estabeleçam melhores normas de segurança, de proteção ao consumidor, de preservação ambiental e de defesa da livre concorrência (Columbia, 2024).

Desinformação e fabricação

Na era digital, marcada pelo excesso de informação e estímulos, a *atenção* se torna recurso escasso e valioso (Wu, 2017). Mentira e manipulação de discurso não são fenômenos recentes ou inerentemente ligados às tecnologias digitais. O filósofo Slavoj Zizek já reconhecia que ideologia seria a primeira e mais eminente manifestação de "realidade virtual" (Wright, 2004). A multiplicidade de narrativas e discursos ideológicos é algo inerente à vida social. Contudo, enquanto os meios tecnológicos sempre serviram à disseminação de ideias, sejam elas verdadeiras ou não, a era digital acentuou a efusão de modelos que se beneficiam com engajamento em mentiras e desinformação. Os grandes modelos de linguagem podem agir como geradores muito eficientes de desinformação, de forma intencional ou não (Pan et al., 2023). As grandes plataformas, com modelos de negócios baseados na "economia da atenção", em que conteúdos são priorizados para maximizar o tempo de permanência dos usuários, acabam por minar a crítica, a resistência e a capacidade de se autodeterminar (Zuboff, 2019). O impacto social é grave: além de alienação e apatia, aumentam-se a violência, a discriminação, a desigualdade e a injustiça (O'Neil, 2016). A dinâmica é ainda acentuada quando as tecnologias de linguagem são promovidas, erroneamente, com o viés de neutralidade axiológica (Broussard, 2023). Medidas de mitigação ainda são incipientes; entre o instrumentário possível estão regulação das redes sociais, moderação de conteúdo, quebra de monopólios das plataformas digitais e limitação de acesso e uso de tais tecnologias[20].

Perfilamento e predição de dados sensíveis

Sistemas de perfilamento (*profiling*) baseados em IA tentam encontrar correlações entre dados de diversas fontes para caracterizar e categorizar indivíduos. Em determinadas plataformas, a coleta pode ter mais de 60 fontes de captação de dados (pessoais, biométricos, demográficos, telemetria e muitos outros)[21]. O que quer que estejamos fazendo, pode haver sensores em diversos aparelhos e aplicativos capturando sinais para serem usados por algoritmos tentando predizer constantemente nossas intenções, emoções e outras características privadas como sexualidade, religião, posição política, doenças e condições psicológicas, como depressão e risco de suicídio, sem nosso consentimento e sem sequer sabermos. Muitos têm bons motivos para não desejar que tais informações se tornem públicas. Por exemplo, uma pessoa com depressão pode estar fazendo um grande esforço para sobreviver em seu ambiente social, e é extremamente invasivo um algoritmo de terceiros simplesmente tentar descobrir e disseminar uma *adivinhação* de uma informação que ela não quer compartilhar. É porventura desejável que nosso comportamento, até em despretensiosas atividades de lazer como escolher um filme para assistir ou postar uma foto e curtir comentários nas redes sociais, sejam usados para inferir nossas características e nos transformar em dados que são comercializados? Ainda que isso seja tecnicamente possível, há modelos que simplesmente não deveriam existir. Um exemplo extremamente problemático foi um projeto de pesquisa que propôs um modelo que tentava predizer sexualidade das pessoas com base em meras fotos[22]. É tecnicamente possível treinar modelos em basicamente *qualquer* tipo de dado, pois modelos são agnósticos ao sentido das coisas. Imagine usar o número de fios de cabelo de uma pessoa como preditor de quantos livros ela tem em casa: bastaria ter essas informações alinhadas em uma base de dados e usar o primeiro como *input* e o segundo como *output*. O modelo seria treinado e produziria alguma resposta com base nos dados observados. Todavia, essa tarefa não faz sentido algum na realidade e qualquer resposta numérica, ainda que por mero acaso correta, seria desprovida de qualquer sentido. Os riscos são ainda mais graves quando tais práticas são empregadas em monitoramento maciço ou envolvendo grupos de maior vulnerabilidade, como crianças, adolescentes e idosos (Ploug, 2023). Nem em todos países com leis de proteção de dados há regras específicas para perfilamento e predição de dados sensíveis. No Brasil,

[20] Fonte: *Big techs desafiam a democracia e favorecem a extrema direita* por Katarine Flor (Fundação Rosa Luxemburgo) em 11 de fevereiro de 2025. Disponível em: https://rosalux.org.br/big-techs-desafiam-a-democracia-e-favorecem-a-extrema-direita/.

[21] Fonte: *The Data Big Tech Companies Have On You* por Gebe Petrino em 12 de janeiro de 2026. Disponível em: https://www.security.org/resources/data-tech-companies-have/.

[22] Fonte: *AI 'Gaydar' And How The Future Of AI Will Be Exempt From Ethical Review* por Kalev Leetaru em 17 de setembro de 2017. Disponível em: https://www.forbes.com/sites/kalevleetaru/2017/09/16/ai-gaydar-and-how-the-future-of-ai-will-be-exempt-from-ethical-review/.

a questão é regulada via jurisprudência (Lobo, 2022). Conceder a autorização de perfilamento somente com base em consentimento pessoal já não é o bastante, pois ocasiona um ônus desproporcional sobre o indivíduo devido às profundas complexidades dos modelos técnicos e negociais em que o perfilamento ocorre. Alternativas incluem regimes de responsabilização mais estrito, transparência obrigatória, direito de regresso e uma abordagem contextual baseada em risco (Centre for Information Policy Leadership, 2024).

Riscos de uso dual ou para fins militares

Um dos grandes perigos envolvidos no desenvolvimento de tecnologias sensíveis é o fato de que elas podem ser usadas para outros objetivos não previstos ou desejados por quem as desenvolveu. Como garantir que modelos de IA não serão utilizados por terroristas ou para fins militares e ameaças cibernéticas? É um desafio alcançar o equilíbrio entre manter livres a iniciativa e a ciência, e evitar o uso indevido ou o reaproveitamento malicioso (Hovy; Spruit, 2016). Estratégias de mitigação de riscos e diretrizes sobre usos indevidos, includindo regras para financiamento e comercialização da pesquisa ou da produção, são elementos importantes para a governança de desenvolvimento responsável (Brenneis, 2025). Salvaguardas para se prevenir o uso dual indevido incluem avaliações de risco das capacidades da tecnologia de IA bem como avaliações operacionais envolvendo testes intensivos e interativos por diferentes especialistas, proporcionando compreensão mais precisa do comportamento de um modelo em diversos cenários, incluindo os não desejados (Barrett et al., 2024).

Dubiedade e "alucinações"

Uma disfunção que afeta especialmente os *chatbots* baseados em grandes modelos de linguagem são as chamadas "alucinações" (Maleki et al., 2024), ou seja, casos em que o modelo gera respostas que soam corretas e cheias de certeza, mas não são. Embora tenha se consolidado, o termo *alucinação* não é muito adequado[23], entre outros motivos por levar a crer que há algo de errado apenas naquela resposta e não no modelo em si. Todavia, o mecanismo que leva o modelo a gerar respostas certas ou erradas é o mesmo: mera predição de próximas palavras dado um contexto. Não é um funcionamento "anormal" momentâneo que ocasiona as falhas. Os modelos são indiferentes à verdade e à precisão de seus *outputs*, de modo que suas afirmações são, pelo mesmo acaso, certas ou erradas (Hicks et al., 2024). Esse tipo de comportamento gera incerteza, pois às vezes o modelo funciona como deveria, às vezes não, e não há uma explicação definitiva para tais vacilações e falhas. Isso é extremamente perigoso nos casos em que usuários não têm o conhecimento necessário para checar o conteúdo, e são levados a crer na infalibilidade do modelo.

Produção de lixo digital (*slop*)

Modelos de IA que geram textos, imagens ou vídeos são frequentemente usados para produzir material artificial sem sentido, absurdo, de baixa qualidade, tosco e sem compromisso com a verdade, o que passou a ser chamado de *slop* e já inunda a Internet[24]. Embora tal conteúdo possa ser criado sem usar IA, ela facilita a produção em massa e com agilidade, de forma cada vez mais realista. Há uma série de consequências negativas: é difícil distinguir o que é genuíno do que é sintético, o que gera confusão e consome inutilmente o precioso tempo de quem usa as plataformas. O excesso e a amplificação de *slops* também polui o ambiente digital, sequestra as pautas de atenção e atrapalha as interações orgânicas entre humanos.

Desqualificação (*deskilling*)

Quando as pessoas passam a depender excessivamente de tecnologias para acessar informações, tomar decisões ou resolver tarefas, há um risco de que elas deixem de saber como fazê-lo por conta própria ou passem a confiar cegamente nas ferramentas[25]. Além disso, pode haver uma falta de interesse em adquirir habilidades e conhecimento que requerem estudo e prática devido à percepção de que o sistema de IA já

[23] Fonte: *Why ChatGPT and Bing Chat are so good at making things up* por Benj Edwards em 06 de abril de 2023. Disponível em: https://arstechnica.com/information-technology/2023/04/why-ai-chatbots-are-the-ultimate-bs-machines-and-how-people-hope-to-fix-them/.

[24] Fonte: *O que é 'IA slop' e por que isso pode estar matando a internet como conhecemos* por Henrique Sampaio em 12 de julho de 2025. Disponível em: https://www.estadao.com.br/link/cultura-digital/o-que-e-ia-slop-italian-brainrot-matando-internet-nprei/.

[25] Ver *The AI Deskilling Paradox* por Samuel Greengard em 07 de novembro de 2025. Disponível em: https://cacm.acm.org/news/the-ai-deskilling-paradox/.

faz as coisas melhor e mais rápido[26]. Se isso ocorre, a educação, qualificação e busca por aperfeiçoamento pessoal ficam comprometidas, e a construção de conhecimento *humano* pode estagnar sob a ilusão de quem a IA é detentora do conhecimento.

Colonialismo via tecnologia

O termo "colonialismo digital" ganhou notoriedade por descrever dinâmicas de poder e controle sobre dados e infraestruturas digitais que empresas ou agentes públicos de alguns países ou regiões exercem em outras regiões, especialmente aquelas não completamente desenvolvidas (Oliveira, 2024). Essas práticas se assemelham àquelas historicamente tidas como coloniais que perpetuam desigualdades econômicas, sociais e políticas. O termo não é apenas uma metáfora, mas uma dominação concreta que envolve a "subordinação econômica, política, social e racial de determinados territórios" por meio de tecnologias digitais (Couldry; Mejias, 2019). O desenvolvimento atual de tecnologias de linguagem reflete bem o problema: embora o panorama geral de PLN se concentre ainda em inglês e algumas outras poucas línguas (Bender, 2011), há um crescente interesse em criar tecnologias de linguagem para idiomas indígenas. Esse propósito pode ser inclusivo em alguns aspectos, mas traz consigo o risco de se tratar as línguas e o conhecimento indígenas como uma mercadoria, com práticas extrativistas, privando tais comunidades locais do conhecimento que elas mesmas produziram, reproduzindo as cruéis dinâmicas do colonialismo (Bird, 2020, 2022). O desafio é estabelecer critérios éticos e legais que desafiam o perverso legado colonial focando nas necessidades reais das comunidades historicamente exploradas e os objetivos das pesquisas (Schwartz, 2022).

6.2.4 Inconveniências operacionais

Além das questões sociais, há também uma série de dilemas técnicos e operacionais que dificultam o desenvolvimento e uso responsável das tecnologias de linguagem.

Obstáculos à reprodutibilidade

Reprodutibilidade, ou seja, a possibilidade de se obter os mesmos resultados quando um experimento é repetido por diferentes pesquisadores mas sob condições equivalentes, é um pilar da ciência que provê credibilidade às conclusões. Há muitos anos já se observavam dificuldades de se reproduzir resultados computacionais e conclusões em pesquisas envolvendo linguagem natural devido à falta de transparência ou documentação de decisões de preprocessamento, configurações do experimento, versões de dados e sistemas computacionais (Cohen et al., 2018; Fokkens et al., 2013). Embora o compartilhamento de dados e de código tenha se tornado uma prática usual na área científica de PLN (Arvan et al., 2022; Wieling et al., 2018), nem sempre isso é suficiente. Em uma revisão de literatura sobre reprodutibilidade em PLN, Belz et al. (2021) mostraram que apenas 14,03% de 513 estudos conseguiram chegar a resultados equivalentes, e pequenas diferenças no código às vezes levam a diferenças consideráveis na performance. Modelos de IA envolvem tantos dados, tantos hiperparâmetros, tantos experimentos e versões de bibliotecas de códigos que fica difícil documentar tudo com fidelidade. Mesmo quando o código-fonte é aberto, poucas instituições dispõem de recursos como infraestrutura e acesso aos dados para tentar replicar um modelo (sem contar a inviabilidade do custo financeiro e energético para replicar, apenas para verificação e inspeção, um modelo que já teve um alto custo para ser treinado). Boas práticas envolvem a inclusão de documentação extensiva e completa do modelo disponibilizado, por exemplo, seguindo a proposta dos *model cards* de Mitchell et al. (2019), e *standards* de metodologia sistemática para ajudar a organizar os muitos experimentos empíricos na criação de modelos (Ulmer et al., 2022). Todavia, quando os modelos viram artefatos comerciais ou governamentais, pode haver um desincentivo ao compartilhamento de detalhes relevantes. Com isso, resultados importantes ficam dependentes de modelos proprietários que podem ser alterados ou sair de linha a qualquer momento, de modo que a reprodução de resultados fica completamente impossibilitada. Isso de fato ocorreu com um modelo comercial de geração de código que havia sido usado em centenas de artigos científicos e foi descontinuado pela empresa[27].

[26] Fonte: *AI is Destroying the University and Learning Itself* por Ronald Purser em 01 de dezenbro de 2025. Disponível em: https://www.currentaffairs.org/news/ai-is-destroying-the-university-and-learning-itself.

[27] Fonte: *OpenAI's policies hinder reproducible research on language models* por Sayash Kapoor e Arvind Narayanan em 22 de março de 2023. Disponível em: https://www.normaltech.ai/p/openais-policies-hinder-reproducible.

Opacidade e pouca explicabilidade

A maioria dos modelos de IA da atualidade se baseia em métodos de aprendizado profundo, cujas representações numéricas em grandes quantidades de vetores e de camadas não são compreensíveis para os humanos, nem mesmo para quem os desenvolve. Desta forma, fica difícil explicar por que um modelo toma decisões certas ou erradas e seu uso vira quase que um exercício de fé em padrões de comportamentos observados empiricamente em testes controlados. Há, de fato, uma linha de pesquisa prolífica em *interpretabilidade* que busca entender como os parâmetros dos modelos se relacionam com seus dados de *input* e *output* (Belinkov et al., 2020; Belinkov; Glass, 2019), mas ainda não é possível esclarecer e justificar totalmente as decisões e predições dos modelos (Bianchi; Hovy, 2021). Note que não adianta perguntar para uma IA generativa o porquê de sua resposta: não há garantia alguma de que o texto que ela produz como explicação representa com fidelidade o *processo interno* de seus parâmetros que levou à resposta inicial. Essa opacidade limita seu uso em situações que requerem transparência e explicabilidade, como questões judiciais, médicas ou financeiras.

Altos custos financeiros e despesas por fracassos

Os modelos de aprendizado profundo em larga escala precisam armazenar bilhões de parâmetros e usá-los na hora de computar seus *outputs*. Para gerar uma resposta, há uma série de cálculos que precisam ocorrer com muita eficiência. A fase de treino é ainda mais computacionalmente custosa pois envolve também armazenar e processar muitos dados. Tudo isso exige uma infraestrutura tecnológica (processadores, *datacenters*, computação de alta performance, etc) cujos elevados custos financeiros impossibilitam o acesso a todos que queiram ou precisem. Mesmo quando tais modelos são disponibilizados como serviço, nem sempre ele é gratuito: cobram-se valores com os quais muitos indivíduos não podem arcar. Há também o risco de despesas devido a fracassos: projetos que prometem soluções baseadas em IA mas não conseguem concretizá-las, mesmo já tendo despendido recursos financeiros consideráveis (Westenberger et al., 2022). Segundo relatório recente, apesar do investimento empresarial estimado em dezenas de bilhões de dólares em IA generativa, 95% das organizações não têm tido qualquer retorno[28].

Complexidade de licenças e termos de uso

Não obstante os tipos de modelos de IA atuais se diferenciarem em termos técnicos, eles compartilham a característica de serem sistemas computacionais que dependem intensivamente de dados. Tanto *software* quanto dados são regulados juridicamente. Direitos autorais sobre IA podem ser categorizados em três grupos: os direitos autorais do próprio *software* de IA, os direitos autorais sobre o conteúdo produzido pela IA e os direitos autorais dos elementos usados no treinamento do sistema de IA (Schirru, 2023). A distribuição e a reutilização de cada um desses elementos depende de licenças ou autorizações. Muitos sistemas de IA que se tornaram populares possuem licenciamento "permissível" facilitando o fluxo de uso, remix e reutilização de *software* e dados (Haddad, 2022). Apesar de tais licenças imporem poucas obrigações, ao ponto de muitos as ignorarem, os termos e condições de uso permanecem válidos (Novobilská, 2023). O risco é agravado pelo fato de que há poucos sistemas de IA totalmente abertos, licenciados por instrumentos que oferecem transparência, reutilização e extensibilidade abrangentes. Muitos desses sistemas são comercializados como "abertos" mas permanecem "fechados" juridicamente (Gray Widder et al., 2023). Os riscos são atenuados quando atividades de pesquisa e comerciais atendem padrões de governança contratual e de licenciamento (Lasota; Singhal, 2024).

Fragilidades na segurança cibernética

O fato de que qualquer dispositivo conectado à Internet pode ser hackeado de forma maliciosa é um dos mais sérios problemas da era digital. Vazamento de dados, contas invadidas e golpes digitais são apenas alguns dos exemplos de fragilidades de segurança afetando milhões de pessoas. Um caso recente exemplifica as consequências nefastas de falhas de segurança de tecnologias de texto: uma empresa que oferecia serviços de terapia teve as anotações dos terapeutas acerca de pacientes roubadas e disseminadas na *deep web*, revelando segredos íntimos para desconhecidos com intuitos maliciosos[29]. É muito difícil reparar os danos quando isso

[28] Fonte: *The GenAI Divide: State of AI in Business 2025* por Aditya Challapally, Chris Pease, Ramesh Raskar e Pradyumna Chari (MIT NANDA) em julho de 2025. Disponível em: https://mlq.ai/media/quarterly_decks/v0.1_State_of_AI_in_Business_2025_Report.pdf.

[29] Fonte: *A faceless hacker stole my therapy notes – now my deepest secrets are online forever* por Jenny Kleeman em 17 de janeiro de 2026. Disponível em: https://bbc.com/news/articles/c62nzxqw45eo.

ocorre, pois cópias dos dados são adquiridas e salvas em muitos locais. O cenário é ainda mais crítico ao se considerar o tempo médio de resposta elevado na América Latina de mais de 100 dias para detecção e solução de vulnerabilidades[30]. Tecnologias de linguagem podem estar sujeitas e riscos de segurança como vazamento de dados, ataques por acesso alternativo (*backdoor*) e ataques por imitação (Xu; He, 2023); modelos também podem memorizar informações sensíveis observadas durante o treino e replicá-las em seus *outputs* (Huang et al., 2024a). É interessante notar que, contudo, a ênfase excessiva em vigilância na segurança (reconhecimento facial em locais públicos, por exemplo) contrasta com a falta de supervisão regulatória do controle corporativo, levando a falhas de segurança estruturais e persistentes (Francisco et al., 2020). Dependendo do contexto, segurança cibernética entra em conflito com outros princípios. Por exemplo, enfatizar excessivamente políticas de segurança pode violar valores fundamentais, como igualdade, justiça, liberdade ou privacidade. Ao mesmo tempo, negligenciar tais medidas pode prejudicar a privacidade e a proteção dos indivíduos, além de afetar negativamente a confiança na infraestrutura e nas instituições. Segurança cibernética deve ser tomada como uma qualidade do modelo, parte da avaliação de sua performance. De fato, recentes legislações na Europa impõem programas de conformidade nos quais segurança é tratada como defesa do consumidor (Lasota, 2025).

Competição e adoção precoce

Bianchi; Hovy (2021) identificaram algumas tendências preocupantes na pesquisa em PLN, entre elas a adoção de métodos sem compreensão suficiente de suas capacidades, limitações e efeitos colaterais. Devido à competitividade para cientistas publicarem artigos o mais rápido possível ou para empresas terem seus produtos gerando lucro o quanto antes, muitos modelos são adotados e disponibilizados ao público de forma precoce, sem testes rigorosos. Isso se manifesta também na centralização da avaliação dos modelos em torno de *leaderboards*, isto é, *rankings* que ordenam modelos por seu desempenho com base em apenas poucas métricas em um determinado *benchmark*. Dessa forma, a percepção simplista de "progresso" passa a ser apenas estar no topo do *leaderboard*, e os incentivos se dirigem a ganhar essa corrida enquanto a verdadeira competência linguística dos modelos fica ofuscada (Raji et al., 2021; Schlangen, 2021). Além disso, como Ethayarajh; Jurafsky (2020) argumentaram, a utilidade de um modelo depende de aspectos variados para consumidores diferentes, por exemplo: de que adianta o modelo estar no topo do *ranking* se sua performance é extremamente lenta? É preciso primeiramente avaliar e testar os modelos muito bem e no contexto adequado antes de introduzi-los em atividades que afetam a vida das pessoas.

Aversão à responsabilidade

Quem é responsável pelos *outputs* de um modelo: quem o desenvolve, quem o provê como serviço, que o incorpora em um aplicativo ou objeto ou quem o usa? Questões de responsabilidade (*liability*) se tornam um emaranhado de dúvidas, com muitos atores tentando se desvincular da responsabilidade sobre o que o modelo produz[31]. É comum, por exemplo, que empresas usem grandes modelos de linguagem para gerar textos de forma automatizada e deleguem para seus clientes ou usuários a responsabilidade pela checagem das informações através de um mero *disclaimer*. O filósofo e tecnologista de IA Joseph Weizenbaum afirmou décadas atrás que "nossa sociedade desenvolveu técnicas de se distribuir responsabilidade de modo que ninguém a tenha" (Weizenbaum, 2001). Do ponto de vista ético, todos os envolvidos carregam parte da responsabilidade pelo uso da IA. Na verdade, tecnologias de linguagem podem gerar danos que não apenas afetam indivíduos mas sociedades inteiras. Os riscos podem ser classificados em (a) riscos de integridade física, quando IA é implementada em meios de transporte ou em equipamentos médicos, de segurança pública e militares; (b) riscos sociais e econômicos de sistemas autônomos bancários, previdenciários, de seguros e laboral; e (c) riscos para direitos fundamentais, quando IA é usada em larga escala para políticas públicas, na educação, no setor judiciário e no exercício do poder (Wendehorst, 2022). A responsabilização jurídica e moral dos agentes desenvolvendo, aplicando e usando IA é um tema urgente. Contudo, há diversos desafios: a opacidade inerente aos sistemas de IA, referida comumente como o "paradigma da caixa-preta" apresenta dificuldades para os conceitos tradicionais de responsabilidade civil e criminal; a determinação do nexo de causalidade entre o dano e o fato causador é, em alguns casos, difícil de se provar; a atuação do

[30] Fonte: *América Latina demora quase 300 dias para corrigir vulnerabilidades críticas, alerta estudo* por André Luiz Dias Gonçalves (TecMundo) em 07 de maio de 2025. Disponível em: https://www.tecmundo.com.br/seguranca/404381-america-latina-demora-quase-300-dias-para-corrigir-vulnerabilidades-criticas-alerta-estudo.htm

[31] Fonte: *Diffusion of Responsibility* por Jürgen Geuter em 14 de fevereiro de 2026. Disponível em: https://tante.cc/2026/02/14/diffusion-of-responsibility/.

usuário de IA pode concorrer para a responsabilização; interesses econômicos em se permitir "inovação" de IA podem restringir o alcance das regras de responsabilização[32]. À medida que a IA se torna onipresente, os legisladores estão sob pressão para agir de forma rápida e reativa. No Brasil, embora propostas legislativas tenham direcionado normas de responsabilidade civil e criminal à IA, elas também introduziram novas inseguranças jurídicas[33].

6.2.5 Hegemonia de interesses comerciais e econômicos

Boa parte das tecnologias de linguagem usadas em grande escala atualmente são, antes de tudo, produtos comerciais fornecidos por algumas empresas do grupo das denominadas *big tech* ou por *start-ups*. A influência dos interesses corporativos dos detentores de ferramentas e modelos usados em tantas esferas, como na vida pessoal, na área da saúde, nos processos jurídicos e políticos, na ciência e educação, levanta diversas questões inadiáveis.

Influência da agenda corporativa

A percepção do imaginário popular sobre IA é marcada por retóricas grandiloquentes ou enredos de ficção científica que desviam o foco de questões mais complexas de como a IA realmente funciona na vida real (Rehak, 2025). Uma delas é o fato de que os maiores desenvolvedores e provedores de IA são empresas de *big teach*, cujas práticas comerciais acumulam um histórico de monopolização em mercados digitais, atividades não-democráticas e desrespeito a direitos individuais e coletivos (Schaake, 2024). Há uma continuidade histórica entre o comportamento dessas corporações e práticas de extração e exploração que levam à monopolização e concentração de poder (Lasota, 2023a). Enquanto antes da era digital os monopólios privados eram dissolvidos ou nacionalizados, a extrema concentração de poder transnacional dessas empresas elevou sobremaneira o desafio de manter os mercados digitais abertos e saudáveis para a livre concorrência. Arranjos institucionais alternativos de IA envolvem democratização de acesso, bem como controle e distribuição de tecnologias sensíveis das quais a IA depende. Iniciativas para modelos de IA que fomentem solidariedade e justiça social precisam ter domínio sobre a coleta e o processamento de dados, bem como amplos poderes sobre uso, reuso e distribuição dos dados e *software* empregados (Lasota, 2023b).

Modelos e dados proprietários fechados

O treinamento de modelos de IA ainda está restrito a poucas instituições ao redor do mundo com capacidade computacional e de processamento de dados suficiente (Thun; Hanley, 2024), e envolve investimentos substanciais. O anseio por dados faz surgir uma economia em volta de sua coleta, relacionada ao "capitalismo de vigilância" que permeia a indústria de IA (Zuboff, 2019). Aliados a arcabouços jurídicos e contratuais, esses dados geralmente não são de acesso livre e são organizados em formatos que mitigam a interoperabilidade. A distribuição assimétrica de poder e o controle sobre a informação e o acesso às tecnologias de desenvolvimento tornam-se, portanto, privilégios de grandes empresas ou estados capazes de arcar com tais custos. A consequência do acúmulo de capital e poder reflete-se na imposição por parte das *big techs* de aprisionamentos tecnológicos denominados *lock-ins*. Tais aprisionamentos limitam a capacidade de recriar e administrar infraestruturas alternativas, pois os padrões, *standards* e formatos dos dados, bem como propriedade intelectual e políticas de distribuição sobre *software*, aumentam os custo de transição, desincentivam a portabilidade, impedem a interoperabilidade e bloqueiam a reutilização para análise ou reprocessamento. Esse cenário solapa a habilidade de alternativas concorrentes, da academia ou mesmo sociedade civil, de analisar, entender e escrutinar os sistemas de IA (Lasota, 2023b). Quando modelos proprietários são usados, não há como inspecionar ou auditar seu funcionamento, e tampouco customizá-lo para necessidades individuais, de modo que os usuários ficam reféns de "oráculos" inacessíveis que apenas soltam respostas sem transparência. Há também o risco constante de que o provedor tire o produto de linha ou aumente seu preço, inviabilizando a continuidade de serviços que dependiam de tal modelo.

[32] Fonte: *A necessidade de um regime específico de responsabilidade civil para a IA* por Eduarda Chacon Rosas (JOTA) em 04 de dezembro de 2025. Disponível em: https://www.jota.info/opiniao-e-analise/colunas/elas-no-jota/a-necessidade-de-um-regime-especifico-de-responsabilidade-civil-para-a-ia.

[33] Fonte: *A necessidade de um regime específico de responsabilidade civil para a IA* por Eduarda Chacon Rosas (JOTA) em 04 de dezembro de 2025. Disponível em: https://www.jota.info/opiniao-e-analise/colunas/elas-no-jota/a-necessidade-de-um-regime-especifico-de-responsabilidade-civil-para-a-ia.

Concentração de poder e controle não democrático

Similarmente ao que ocorreu com a Internet há 30 anos, em que poucos conglomerados empresariais se consolidaram nas poucas grandes plataformas que controlam boa parte da vida digital atual, interesses econômicos também podem moldar a forma com que IA é aplicada na sociedade. Esse poder econômico se traduz também em polarização de IA para fins anti-democráticos e contrários ao interesse popular. A IA pode contribuir para conduzir sociedades inteiras a uma nova idade colonial por meio de gestões autoritariamente excludentes (Mejias; Couldry, 2019). É necessário decidir até que ponto a vida deve ser contemplada pela ótica tecnocrática das grandes plataformas e quais limites devem ser impostos à implementação de IA. Visões críticas questionam se a crescente implementação de IA não irá somente aprofundar as injustiças que tal ordem promove. Defendem, portanto, que implementação de IA não deve ser tomada como um fato inevitável e é preciso resistir (McQuillan, 2022). Seria possível que a IA atendesse critérios de justiça social, bem-estar e respeito à ordem democrática? Alternativas mais incisivas incluem a denúncia e a oposição ao poder corporativo através da promoção de agendas alternativas focadas em justiça social, democratização, sustentabilidade e direitos humanos (Barry Lynn; Montoya, 2023). A via reformista inclui atualização de legislação através de marcos regulatórios ou atividade judicial com restrições mais rigorosas e penalidades sobre monopólios e relações desequilibradas de poder (Lasota, 2023b).

Criação de *hype*

A "*hype*" (algo como um entusiasmo improcedente) em torno de IA é um fenômeno mercadológico falacioso que promete mais do que a tecnologia é ou pode entregar, o que acaba por aumentar o fosso entre o imaginário popular e as realidades em que IA é desenvolvida e aplicada na sociedade (Bender; Hanna, 2025). Observamos seus efeitos em discursos exageradamente positivos de entusiastas, na excessiva ênfase à IA em todas as esferas, na imposição do uso de IA como sinônimo de progresso e inovação, no "banho de *marketing*" em produtos que usam IA sem real necessidade, nas afirmações infundadas de que tais sistemas atingiram uma "inteligência sobre-humana", e nas promessas exageradas de que ela funciona de forma espetacular e pode trazer a solução de todos nossos problemas. As avaliações de valor, poder e desempenho dos sistemas de IA são afetadas por esse viés mercadológico, borrando as perspectivas críticas sobre as consequências colaterais dessa escalada intensa para sistemas cada vez mais poderosos e providos de acesso sem as salvaguardas necessárias (Varoquaux et al., 2025). De fato, a *hype* é uma faceta do chamado "solucionismo tecnológico", criando narrativas que facilitam a adoção da IA para problemas estruturais que, em princípio, não dependeriam de uma solução tecnológica. Pelo contrário, a complexidade do problema acaba por se agravar com o potencial disruptivo de AI, consolidando os próprios efeitos maléficos que se intentava resolver (Morozov, 2014). Uma posição responsável e ética requer honestidade sobre as reais capacidades dos sistemas, bem como os riscos e perigos do seu uso.

Uso imposto e consentimento forçado

As políticas expansionistas de IA, que nos últimos anos alcançaram tantas pessoas e espaços, acabaram também por gerar dinâmicas coercitivas para o uso de tal tecnologia. O afã em se empregar rapidamente IA sob a expectativa de lucratividade e eficiência ocorre, muitas vezes, às custas da compreensão, vontade e consentimento dos indivíduos. Essa imposição não apenas compromete o funcionamento desejável dos mercados, mas também afeta diretamente liberdades e direitos fundamentais. Uma abordagem ética deve considerar o livre-arbítrio e bem-estar dos indivíduos em interagir ou não com o sistema de IA. Legislações protetivas de dados no Brasil e em outros países introduziram o "direito ao esquecimento" e o "direito à desconexão", repelindo coerções de uso e imposições tecnológicas[34]. O mesmo pode-se falar de IA. Há inúmeros grupos de pessoas em posição de vulnerabilidade face à tecnologia e à IA, como idosos e pessoas com necessidades especiais, nem sempre lembradas por quem desenvolve os sistemas (Hendren, 2014; Whittaker et al., 2019). Por exemplo, há serviços básicos como acesso à conta bancária que só podem ser acessados através de aplicativos que exigem reconhecimento facial e extraem dados de comportamento. Quem não quer ceder tais dados, fica sem alternativa de acesso ao serviço. Ou, se há alternativa, ela é obscura ou de difícil acesso: por exemplo, para desativar um assistente virtual em um sistema operacional, poderia ser preciso (em 2025) acessar configurações em submenus, um processo que não é trivial para quem

[34] Fonte: *The right to be forgotten is not compatible with the Brazilian Constitution. Or is it?* por Luca Belli (Future of Privacy Forum) em 23 de março de 2021. Disponível em: https://fpf.org/blog/the-right-to-be-forgotten-is-not-compatible-with-the-brazilian-constitution-or-is-it/.

não é da área de tecnologia[35]. Quando o Mozilla passou a impor IA como padrão em seu navegador, os passos para desativá-la por completo também eram intricados[36], e isso só foi facilitado em uma versão subsequente mediante a reação negativa dos usuários[37]. Nesses casos, as desigualdades digitais reforçam as dificuldades sociais já existentes. Em vez de melhorar o acesso, a digitalização acelera, portanto, uma espiral descendente e leva a situações como a não fruição dos direitos sociais, a exclusão financeira e a perda de autonomia individual[38]. Além disso, há grupos de indivíduos que não desejam coadunar tais dinâmicas e querem se abster. Idealmente, portanto, a prestação de serviços públicos e essenciais deveria permitir aos indivíduos a opção por não engajar com IA, oferecendo alternativas para a condução das atividades sem a necessidade de submissão a ela, mesmo que isso crie custos e estruturas adicionais[39].

Implementação intransparente

Embora muitas empresas não informem consumidores e parceiros quando estão interagindo com ou sujeitos a decisões de IA[40], transparência quanto ao seu uso é crucial, não apenas eticamente mas também devido a determinações legais. Leis como a LGPD no Brasil e o RGPD na Europa exigem que as empresas informem os indivíduos submetidos a decisões automatizadas, sejam elas de IA ou não. No entanto, dependendo dos tipos aplicações de IA, os níveis de transparência também variam. Quando os sistemas IA se comunicam diretamente com os clientes como *chatbots*, orientação profissional automatizada ou atendimento ao cliente, a informação é obrigatória (Toniazzo et al., 2021). Mesmo quando IA é usada somente no auxílio a tarefas administrativas ou profissionais sob supervisão humana, como na elaboração de documentos e tradução, valores de transparência e governança se aplicam[41].

Manipulação de discurso e interferência política

É de se esperar que modelos de IA sirvam aos interesses de quem os desenvolve e fornece. Ou seja, não surpreende que os *outputs* dos modelos não sejam "neutros", mas estejam alinhados com a visão e os valores que a empresa deseja estabelecer ou manter. Em modelos de linguagem, é possível manipular a geração das respostas para que determinadas ideias sejam propagadas em detrimento de outras. Por exemplo, a empresa que fornece o modelo de linguagem pode fazer uma parceria de *marketing* com alguma outra empresa para que o *chatbot* induza os usuários a comprar seus produtos, impulsionados pela crença em um superioridade intelectual e imparcialidade do modelo. Esse dilema se estende muito além de anúncios. Há evidência do emprego de um *chatbot* para difundir convicções pessoais contestáveis de um empresário influente em rede social, e tendo seus discursos deliberadamente orientados para atender a esse fim[42]. Há evidência de que o viés do *chatbos* pode realmente influenciar a posição política de quem interage com eles (Fisher et al., 2025). A incorporação acelerada de IA generativa tanto na produção de textos, imagens, vídeos e áudios quanto no próprio consumo e checagem de informações é um fenômeno que dificulta a condução de processos políticos justos e transparentes. O Brasil inovou a nível mundial em 2024 ao lançar, através do Tribunal Superior Eleitoral, diretrizes para o uso de IA nas campanhas eleitorais. As normas incluem transparência sobre o uso de IA em materiais eleitorais, bem como punições rígidas para produção

[35]Fonte: *How to turn off Copilot and protect your data from Microsoft's AI* por Elena Constantinescu em 14 de novembro de 2025. Disponível em: https://proton.me/blog/turn-off-copilot

[36]Fonte: *How to Disable All the AI Features in Firefox Web Browser* por Ji m e 14 de novembro de 2025. Disponível em: https://ubuntuhandbook.org/index.php/2025/11/disable-ai-firefox/

[37]Fonte: *Block generative AI features with Firefox AI controls* por AliceWyman, Mark Heijl e Flavius Floare em fevereiro de 2026. Disponível em: https://support.mozilla.org/en-US/kb/firefox-ai-controls.

[38]Fonte: *Offline by Right? The EHDSR and the Choice Not to Go Online* por Marta Musidlowska em 28 de outubro de 2025. Disponível em: https://www.law.kuleuven.be/citip/blog/offline-by-right-the-ehdsr-and-the-choice-not-to-go-online/.

[39]Fonte: *Essential services must be accessible, even offline* por Lire et Écrire em 2024. Disponível em: https://righttooffline.eu/.

[40]Fonte: *Sabe o que é 'IA paralela'? Uso informal em empresas acende alerta sobre segurança e vazamentos* por Henrique Sampaio em 26 de julho de 2025. Disponível em: https://www.estadao.com.br/link/empresas/uso-informal-ia-paralela-empresas-acende-alerta-seguranca-vazamento-dados-nprei/.

[41]Fonte: *Why AI Disclosure Matters at Every Level* por Cornelia C. Walther em 20 de janeiro de 2026. Disponível em: https://knowledge.wharton.upenn.edu/article/why-ai-disclosure-matters-at-every-level/.

[42]Fonte: *How Elon Musk Is Remaking Grok in His Image* por Stuart A. Thompson, Teresa Mondría Terol, Kate Conger e Dylan Freedman em 02 de setembro de 2025. Disponível em: https://www.nytimes.com/2025/09/02/technology/elon-musk-grok-conservative-chatbot.html.

e disseminação de conteúdos falsos e *deepfakes* com a intenção de prejudicar ou favorecer uma candidatura. Contudo, a regulamentação deixa lacunas em uma série de tópicos como a regulação de conteúdo falso produzido por pessoas não ligadas a partidos, candidatos ou campanhas[43]. Portanto, uma abordagem ética e responsável na produção, aplicação e uso das tecnologias de linguagem inclui reconhecer como elas são suscetíveis à manipulação do discurso e como facilmente absorvem ideologias nos dados, o que pode ser explorado por pessoas má intencionadas (Chen et al., 2024b).

Tecnosolucionismo

A implementação de IA e de tecnologias de linguagem não ocorre em um "vácuo histórico". Pelo contrário, sua aplicação e uso ocorrem em cenários e situações moldados por contextos culturais, históricos e costumeiros, sejam eles bons ou ruins. Ou seja, o uso dessas tecnologias traz consigo o reflexo dos sistemas e estruturas em que ela é aplicada[44]. Por exemplo, quando IA é aplicada em ambientes de trabalho marcados por décadas de hierarquia, inércia e *design* verticalizado, ela está simplesmente servindo para consolidar ainda mais estruturas que já existem. Quando processos não envolvem mudanças estruturais, a automatização deles via IA apenas intensifica o sistema em vigor. Tecnosolucionismo se refere à abordagem de se formular problemas essencialmente humanos e sociais em termos simplistas, criando a impressão de que soluções tecnológicas poderiam facilmente administrá-los e resolvê-los (Morozov, 2014), e que são necessariamente a melhor opção. O termo original, *technological fix*, foi cunhado por Alvin Weinberg na década de 1960 (Johnston, 2018). Embora se refira à tecnologia como um todo, ele se tornou ainda mais relevante perante as promessas em torno da IA, com o risco da supremacia da técnica sobre aspectos de humanidade. O uso indiscriminado de IA em cenários que exigem profundas mudanças humanas e sociais faz essa tecnologia se tornar somente uma "super burocracia", causando um alto custo ao ampliar desigualdades e injustiças que promete resolver (McQuillan, 2022). O tecnosolucionismo também se manifesta quando a tecnologia é oferecida como solução para resolver um problema *que ela mesma causou*. Por exemplo, vê-se hoje em dia muitos sistemas de detecção de *fake news* ou de plágio que usam IA para combater a proliferação em grande escala de *fake news* e plágio que foi, em grande parte, possibilitada exatamente por modelos de IA. Não seria mais vantajoso buscar formas de impedir que o uso da IA agravasse esses problemas, através de métodos mais controláveis e regulação, em vez de ter de combatê-los com a própria IA?

Cooptação da ciência

A influência das grandes empresas de tecnologia na ciência tem aumentado, em particular na área de PLN, ocorrendo em diferentes frentes: autoria de artigos, estabelecimento de focos de atenção em determinadas áreas de pesquisa, financiamento de projetos, patrocínio em conferências (e, com isso, atividades de *marketing* nos eventos), contratos de prestação de serviços para universidades, entre outros (Abdalla et al., 2023; Whittaker, 2021). Embora possam haver colaborações benéficas e frutíferas entre academia e indústria, é preciso estar ciente da possibilidade de algumas consequências nefastas para o processo científico, bem como de diversas formas de conflitos de interesse. Primeiramente, quando um produto comercial proprietário vira um objeto de estudo (como é o caso de alguns *chatbots*, mais famosos), o trabalho de pesquisa se torna mão de obra especializada gratuita para que o provedor do modelo possa promovê-lo, testá-lo e melhorá-lo. Nesses casos, ainda há o gasto de dinheiro público envolvido para pagar para rodar experimentos em modelos que ficam inacessíveis atrás de uma API e podem ser descontinuados a qualquer momento, inviabilizando a continuidade das pesquisas. Há, ainda, o risco de os incentivos comerciais ditarem o que deve ou não ser estudado e receber financiamento, concentrando a pesquisa em alguns poucos temas ou modelos que servem a interesses corporativos e fazendo uso do processo científico como forma de promoção e validação da atuação das empresas (Whittaker, 2021). Existe ainda o risco da busca por automatização do processo científico, exaurindo-o de seu valor social (Binz et al., 2025) e da captura corporativa do conhecimento humano[45], somado à comoditização da pesquisa (Kogkalidis; Chatzikyriakidis, 2025).

[43] Fonte: *A regulação do uso de IA para as eleições brasileiras. O que está em jogo?* por DFRLab e NetLab UFRJ em 2024. Disponível em: https://netlab.eco.ufrj.br/post/a-regula%C3%A7%C3%A3o-do-uso-de-ia-para-as-elei%C3%A7%C3%B5es-brasileiras-o-que-est%C3%A1-em-jogo.

[44] Fonte: *When AI Doesn't Drive Transformational Change: Just A Better Digital Bureaucracy* por Vibhas Ratanjee em 21 de junho de 2025. Disponível em: https://www.forbes.com/sites/vibhasratanjee/2025/06/21/when-ai-doesnt-drive-transformational-change-just-a-better-digital-bureaucracy/

[45] Fonte: *AI and the Corporate Capture of Knowledge* por Bruce Schneier e J. B. Branch em 16 de janeiro de 2026. Disponível em: https://www.schneier.com/blog/archives/2026/01/ai-and-the-corporate-capture-of-knowledge.html.

6.3 Como atuar de forma responsável?

As tecnologias de linguagem trazem benefícios consideráveis em diversas áreas e são importantes ao se lidar com grandes quantidades de dados de texto ou fala. Todavia, a forma com que a IA vem sendo imposta e a dependência excessiva nos mecanismos de seu paradigma nos trazem consigo uma série de problemas graves, como vimos na seção anterior. Então, como proceder de forma responsável? Infelizmente, não há uma receita que possamos prescrever para garantir que o desenvolvimento e uso da tecnologia não sejam prejudiciais. Mas há algumas recomendações para uma atitude fundamentada em responsabilidade.

Primeiramente, é necessário reconhecer que ética e responsabilidade são prioridades, e devem ser tratadas como tal. Elas precisam figurar entre os principais parâmetros dos processos decisórios, seja no desenvolvimento de uma nova tecnologia, na aplicação ou implementação de modelos, bem como na avaliação de seu impacto. Alternativas que não dependem de IA devem continuar a ser consideradas como soluções possíveis, especialmente quando são mais interpretáveis e menos custosas financeira e ambientalmente.

É importante saber discernir as diversas ideologias envolvidas no debate relacionado à IA. Muitas narrativas, atreladas a supostos futuros irresistíveis e necessários, trazem consigo subterfúgios que minam a compreensão do impacto social da IA no presente, principalmente nas comunidades e indivíduos que arcam com as consequências de seu uso. Exemplo disso é a ideologia que promove a busca pela chamada "inteligência artificial geral", que vislumbra uma IA "sobrehumana" em vez de ferramentas úteis para fins específicos. Gebru; Torres (2024) cunharam a sigla TESCREAL para abarcar o grupo ideológico que inclui transhumanismo, altruísmo efetivo e longotermismo, o qual tem orientado a atuação de representantes proeminentes das *big techs*. Diversas dessas crenças invocam retóricas controversas de cunho mítico sobre a relação entre humanos e IA, buscando uma era "pós-humana"[46]. Ora, a que serve construir um futuro que minimiza o valor e o significado de nossa própria existência?

Deve-se buscar proativamente informação confiável sobre o tema, ouvir especialistas em ética que estejam fora das engrenagens de *marketing* e de *hype* e ter abertura para o aprimoramento das próprias posições pessoais. Mais adiante estão listadas algumas fontes por onde começar.

É preciso refletir sobre cada decisão de uso de ferramentas tecnológicas e ter resiliência para mudar de comportamento digital quando fica claro que algo tem um impacto negativo. Além do nível individual, é também preciso buscar o estabelecimento e a implementação de regulamentações e diretrizes em nível institucional. Por exemplo, criando coletivamente guias para o uso e práticas aceitáveis em um grupo de trabalho ou de pesquisa, incluindo as precauções necessárias e estabelecendo os usos indevidos. Há também diversas comunidades que têm se unido para pensar e agir mediante os inúmeros desafios; elas podem servir de alavanca para o engajamento e a conscientização. Listá-las neste livro seria inviável, mas pode-se procurar a mais próxima em universidades, associações, ONGs, núcleos políticos e comunidades *online*, por exemplo. É, ainda, de suma importância usar esse conhecimento para apoiar e conscientizar outras pessoas que estão sendo expostas a tecnologias usando linguagem humana mas ainda não têm as habilidades digitais e a argumentação necessárias para julgar criticamente seus efeitos.

Tecnologias de linguagem e modelos de IA devem ser minuciosamente avaliados, como qualquer outro modelo ou ferramenta, antes de sua adoção. É fundamental avaliar modelos com testes controlados, abrangentes, sistemáticos e imparciais e analisar seus padrões de erros e acertos. A avaliação é um dos passos que nos ajudam a entender os modelos e a enfrentar alguns dos dilemas éticos que se originam da falta de compreensão. Deve-se priorizar o uso de modelos e dados abertos e sempre inclui-los nas análises, por serem mais acessíveis e passíveis de inspeção mais detalhada. Embora modelos treinados em dados sejam intrinsecamente empíricos, o conhecimento teórico sobre a linguagem humana construído ao longo de tanto tempo pode guiar a análise de seu comportamento e ajudar a identificar falhas importantes. Para saber mais sobre esse tema, sugerimos a leitura dos capítulos Conjunto de dados, *dataset* e *corpus*, Avaliação de tecnologias de linguagem e Avaliação conjunta em português. Avaliação é um alicerce da compreensão *humana*: **não é recomendável delegar a avaliação de uma ferramenta de IA para um outro modelo de IA**, cujo funcionamento é igualmente opaco e não traz garantias de aderência à veracidade.

No que diz respeito à IA generativa, é primordial manter um alto nível de ceticismo acerca do conteúdo que ela produz. O *output* pode estar certo e válido, ou não. É sempre necessário verificar sua qualidade, independentemente do que os provedores do modelo estejam prometendo em termos de performance. A organização AlgorithmWatch propôs algumas diretrizes para o uso responsável especificamente de IA generativa[47], baseadas em quatro princípios cujo conteúdo resumimos aqui:

[46] Fonte: *Tech Capitalists Don't Care About Humans. Literally.* por Émile Torres em 15 de novembro de 2025. Disponível em: https://jacobin.com/2025/11/musk-thiel-altman-ai-tescrealism.

[47] Fonte: *AlgorithmWatch's guidelines to use generative AI responsibly* em 14 de janeiro de 2026. Disponível em https://algorithmwatch.org/en/generative-ai-guideline/

- **Proporcionalidade** acarreta a reflexão na hora de decidir se uma ferramenta de IA generativa realmente é necessária mediante outros tipos de ferramentas que já existem e funcionam bem. Por exemplo, para que usar IA para gerar uma receita de bolo se já existem repositórios de receitas que podem simplesmente ser acessados? Ou mesmo um livro de receitas, ou uma pessoa que gosta de fazer bolos e se alegra em compartilhar dicas?
- **Segurança** se relaciona à concessão de dados para os provedores do modelo. Muitas vezes, dados sensíveis ou sigilosos de terceiros ou ideias inovadoras ainda não tornadas públicas são inseridas nos *prompts*. Esses dados podem ser incorporados em futuras versões do modelo e depois, disseminados. Isso traz problemas de privacidade, confidencialidade e apropriação indevida. Ainda que existam contratos regulando a privacidade dos dados, há sempre um risco de quebras de acordo ou de roubo de dados.
- **Qualidade** requer a checagem das respostas geradas pelo modelo. É arriscado tomá-las como corretas sem questionamento e senso critico. Isso vai além de mera checagem de fatos; envolve também analisar que tipo de perspectiva o modelo tenta impor, quais aspectos recebem atenção em detrimento de outros, e se as citações e referências realmente existem e estão alinhadas com os originais.
- **Transparência** presume reportar o uso da IA generativa quando ela for usada, para que outras pessoas que tenham contato com esse conteúdo estejam informadas e possam ter condições de julgar o uso subsequente do material.

Lembramos que é urgente reconhecer a responsabilidade ética tanto dos provedores quanto dos usuários de IA. Ser transparente quanto ao uso não exime a *responsabilidade* pela produção e disseminação do material. Ao se apropriar dos *outputs* de um modelo para uso próprio, é preciso assumir as consequências sobre o que foi escrito ou produzido. Não adianta tomar apenas os acertos para si e transferir a culpa pelos erros para o modelo, pois ele não responde moralmente, socialmente ou juridicamente pelo que cria.

6.3.1 Para saber mais

Na literatura científica, existem diversos trabalhos dialogando sobre ética em tecnologia e linguagem. Por exemplo, Mohammad (2022) propôs diretrizes com 50 considerações a serem feitas sobre as diferentes tarefas de PLN que podem ser abordadas com IA e D'Arcy; Bender (2023) fizeram uma revisão de literatura focada nas questões éticas da Linguística. Leidner; Plachouras (2017) também argumentaram sobre como a ética pode ser inserida já de antemão no processo de *design* de modelos de PLN.

Além de trabalhos específicos, há diversos grupos com publicações prolíficas, tanto artigos científicos quanto relatórios e recomendações, buscando estudar a IA sob uma perspectiva crítica, socialmente engajada e educativa. É inviável compilarmos aqui uma lista completa, mas podemos indicar alguns caminhos por onde começar a buscar mais conteúdo no Brasil e no mundo:

- Os capítulos Questões éticas em IA e PLN e E agora, PLN? deste livro refletem sobre temas de ética em PLN.
- O tutorial da Associação de Linguística Computacional (ACL), que compila materiais e recursos para estudantes e pesquisadores navegarem as questões éticas em PLN: https://ethics.aclweb.org/tutorials/ACL_2025/.
- O INCT sobre Inteligência Artificial Responsável para Linguística Computacional, Tratamento e Disseminação de Informação: https://dcc.ufmg.br/tild-iar/
- As publicações dos integrantes do Núcleo de Referência em IA Ética e Confiável: https://www.iaetica.org/membros.
- As diretrizes para o uso ético e responsável da inteligência artificial generativa por Sampaio et al. (2024): https://econtents.sbu.unicamp.br/boletins/index.php/ppec/article/view/9509.
- As edições do WICS - Workshop sobre as Implicações da Computação na Sociedade, da Sociedade Brasileira de Computação: https://csbc.sbc.org.br/2026/wics/.
- A página da UNESCO sobre ética da inteligência artificial (IA) no Brasil: https://www.unesco.org/pt/fieldoffice/brasilia/expertise/artificial-intelligence-brazil.
- Os relatórios da ONG *Algorithm Watch* (https://algorithmwatch.org/en/publications/.
- Os projetos do instituto AI NOW: https://ainowinstitute.org/publications.
- As publicações do Instituto DAIR (*Distributed AI Research*): https://www.dair-institute.org/publications/.

6.3.2 Formas de resistência

Para quem deseja reduzir a influência de sistemas de IA em suas esferas de atuação, existem formas de resistência: "explícitas ou sutis, organizadas ou difusas, individuais ou coletivas, de oposição ou reformistas" que não precisam ser uma total negação da tecnologia em si, mas resistir aos "arranjos e assimetrias que moldam e são moldados pelo desenvolvimento e aplicação da IA" (Şimşek, 2025; Şimşek; Yasar, 2025). Ela é feita de forma ética e legalizada e pode ocorrer em forma de manifestações e protestos, ações jurídicas, subversão digital, crítica acadêmica e campanhas em movimentos de base (Şimşek, 2025; Şimşek; Yasar, 2025). Há muitas vozes que podem ser ouvidas. Por exemplo, uma entrevista recente abordou como representantes de várias profissões e atividades têm resistido a certos usos da IA, buscando parar, desacelerar ou controlar sua infiltração na vida humana[48]. Até o humor pode servir de denúncia e ajudar a chamar atenção para os problemas[49].

Tomando como exemplo a área da educação, existem várias iniciativas criticando o tecnosolucionismo e incentivando a resistência. Há uma carta aberta, assinada por diversos profissionais, rejeitando o uso de IA generativa em escolas e faculdades e resistindo à narrativa da inevitabilidade. Ações concretas propostas na carta incluem não usar IA generativa para dar *feedback* para estudantes e criar conteúdos do curso; não promover modelos eticamente problemáticos em nível institucional; não treinar alunos e alunas para usar IA em substituição de seu próprio desenvolvimento intelectual; e respeitar a decisão de resistência à IA por parte do corpo discente[50]. Há também uma estratégia com quatro "atos de fricção": manter estudantes, e não os sistemas de IA, como elemento central do ensino; refrear a busca por "otimização" do processo educacional; interromper a digitalização excessiva da aprendizagem; e criar senso de comunidade com a administração para que seja possível questionar as práticas da instituição[51].

6.4 Considerações finais

É evidente que as tecnologias de linguagem podem ajudar as pessoas, servir como ferramentas úteis e aumentar o conhecimento sobre a linguagem humana. Ao longo deste livro, muitas aplicações vantajosas e benéficas foram discutidas. Apesar disso, queremos concluir salientando que nem tudo precisa ou deve ser mediado pela tecnologia (e, consequentemente, por seus fornecedores), muito menos por sistemas de IA. Nem tudo que é tecnicamente possível é para o bem e nem tudo que é feito em busca do bem gera efeitos positivos para todos. Além disso, há muitas atividades que perdem o sentido quando são esvaziadas de seu intrínseco valor humano. Por isso, devemos ativamente ponderar sobre os benefícios e malefícios dos sistemas de IA que optamos por desenvolver e usar.

É inconsequente ignorar os riscos e as consequências prejudiciais. Há momentos em que é preciso resistir e ter a coragem de se negar a aceitar ou promover os maus usos. Quando for necessária, a escolha por se usar modelos baseados em IA deve ser *responsável*. Ao mesmo tempo, deve-se respeitar a posição das outras pessoas que não querem usar e nem se sujeitar ao uso dela.

Enfatizamos que a ascensão da IA não deve ser considerada algo inevitável. A narrativa de inevitabilidade representa a visão de mundo de alguns grupos beneficiados pela dependência das pessoas em seus artefatos. Há uma busca excessiva pela inserção de IA em produtos e procedimentos apenas para lhes dar ares de inovação e modernidade. Vamos realmente deixar que o modelo de negócios das empresas de tecnologia molde nossas decisões pessoais, apodere-se de nossas práticas científicas, manipule nossa visão de mundo, e passe a mediar todas as nossas interações humanas?

A compreensão da realidade, incluindo o processo científico de busca por conhecimento, não deve ser terceirizada para artefatos artificiais e opacos ou automatizada de forma a excluir o elemento humano (Binz et al., 2025). Algoritmos de IA são simplesmente matrizes de números e funções matemáticas, não são entes detentores de uma inteligência superior. Qual é o proveito de delegar para a IA interações que emanam e cultivam um valor social ou decisões para as quais basta o bom senso? Ou terceirizar escolhas difíceis, que

[48] Fonte: *The People vs. AI: Behind the growing backlash* por Andrew R. Chow em 19 de fevereiro de 2026. Disponível em: https://time.com/7377579/ai-data-centers-people-movement-cover/.

[49] Por exemplo, a coletânea de memes iniciada por Daniel Stenberg, desenvolvedor da ferramenta `curl`. Disponível em: https://mastodon.social/@bagder/115672975717795897.

[50] Fonte: *An open letter from educators who refuse the call to adopt GenAI in education* em 06 de julho de 2025. Disponível em: https://openletter.earth/an-open-letter-from-educators-who-refuse-the-call-to-adopt-genai-in-education-cb4aee75.

[51] Fonte: *Four Frictions: or, How to Resist AI in Education* por Sonja Drimmer e Christopher J. Nygren em 16 de dezembro de 2025. Disponível em: https://www.publicbooks.org/four-frictions-or-how-to-resist-ai-in-education/.

impactam drasticamente a vida humana–seria a conveniência de se desvencilhar da responsabilidade pelas consequências, alegando que apenas fez o que "a IA" determinou?

Recomendamos que tais tecnologias sejam mantidas no papel de ferramentas subordinadas à autonomia humana, que apenas a *auxiliam*. Avaliar sua performance e seus impactos permanece sendo essencial e agir sob a guia de valores éticos e de forma humanizada é primordial. A você, que leu até aqui, nossos sinceros agradecimentos. Confiamos-lhe a missão de conscientizar as pessoas ao seu redor em busca de mais responsabilidade e, quando necessário, resistência.

Agradecimentos

Agradecemos aos organizadores do Seminário de Teses e Dissertações (SETED) de 2025 na Universidade Federal de Minas Gerais pelo convite inicial da palestra que originou este capítulo. Também agradecemos pela oportunidade de apresentar a mesma palestra no *Helmholtz Centre for Environmental Research* de Leipzig, Alemanha, no grupo da Dr. Mariana Madruga de Brito. A audiência de ambos contribuiu com perguntas e discussões interessantes. Finalmente, agradecemos às editoras deste livro pelo convite de contribuir com este capítulo. :::

Referências

ABDALLA, M. et al. **The Elephant in the Room: Analyzing the Presence of Big Tech in Natural Language Processing Research**. Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). **Anais**...Toronto, Canada: Association for Computational Linguistics, 2023. Disponível em: <https://aclanthology.org/2023.acl-long.734>

ABERCROMBIE, G. et al. **Mirages. On Anthropomorphism in Dialogue Systems**. Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. **Anais**...Singapore: Association for Computational Linguistics, 2023. Disponível em: <https://aclanthology.org/2023.emnlp-main.290>

ABID, A.; FAROOQI, M.; ZOU, J. **Persistent Anti-Muslim Bias in Large Language Models**. Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society. **Anais**...Virtual Event USA: ACM, jul. 2021. Disponível em: <https://dl.acm.org/doi/10.1145/3461702.3462624>

ARVAN, M.; PINA, L.; PARDE, N. **Reproducibility in Computational Linguistics: Is Source Code Enough?** Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. **Anais**...Abu Dhabi, United Arab Emirates: Association for Computational Linguistics, 2022. Disponível em: <https://aclanthology.org/2022.emnlp-main.150>

ASSUNÇÃO, I. V.; JANSON, S. F. Afinal, o que é privacidade? Um panorama histórico do direito à privacidade no ordenamento constitucional brasileiro. **Internet & Sociedade**, v. 5, n. 1, 2024.

BAI, X. et al. Explicitly unbiased large language models still form biased associations. **Proceedings of the National Academy of Sciences**, v. 122, n. 8, p. e2416228122, fev. 2025.

BARRETT, A. M. et al. **Benchmark Early and Red Team Often: A framework for assessing and managing dual-hazards of AI foundational models**. UC Berkeley Center for Long-Term Cybersecurity, 2024. Disponível em: <https://cltc.berkeley.edu/wp-content/uploads/2024/05/Dual-Use-Benchmark-Early-Red-Team-Often.pdf>

BARRY LYNN, M. VON T.; MONTOYA, K. **AI in the Public Interest: Confronting the Monopoly Threat**. Open Markets Institute, 2023. Disponível em: <https://www.openmarketsinstitute.org/publications/report-ai-in-the-public-interest-confronting-the-monopoly-threat>

BELINKOV, Y.; GEHRMANN, S.; PAVLICK, E. **Interpretability and Analysis in Neural NLP**. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: Tutorial Abstracts. **Anais**...Online: Association for Computational Linguistics, 2020. Disponível em: <https://www.aclweb.org/anthology/2020.acl-tutorials.1>

BELINKOV, Y.; GLASS, J. Analysis Methods in Neural Language Processing: A Survey. **Transactions**

of the Association for Computational Linguistics, v. 7, p. 49–72, 2019.

BELZ, A. et al. **A Systematic Review of Reproducibility Research in Natural Language Processing**. Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume. **Anais**...Online: Association for Computational Linguistics, abr. 2021. Disponível em: <https://aclanthology.org/2021.eacl-main.29>

BENDER, E. M. On achieving and evaluating language-independence in NLP. **Linguistic Issues in Language Technology**, v. 6, 2011.

BENDER, E. M. Resisting Dehumanization in the Age of "AI". **Current Directions in Psychological Science**, v. 33, n. 2, p. 114–120, abr. 2024.

BENDER, E. M.; FRIEDMAN, B. Data Statements for Natural Language Processing: Toward Mitigating System Bias and Enabling Better Science. **Transactions of the Association for Computational Linguistics**, v. 6, p. 587–604, 2018.

BENDER, E. M.; HANNA, A. **The AI Con: How to fight big tech's hype and create the future we want**. [s.l.] Random House, 2025.

BENDER, E. M.; KOLLER, A. **Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data**. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. **Anais**...Online: Association for Computational Linguistics, jul. 2020. Disponível em: <https://aclanthology.org/2020.acl-main.463>

BENOTTI, L.; BLACKBURN, P. **Ethics consideration sections in natural language processing papers**. Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. **Anais**...Abu Dhabi, United Arab Emirates: Association for Computational Linguistics, 2022. Disponível em: <https://aclanthology.org/2022.emnlp-main.299>

BIANCHI, F.; HOVY, D. **On the Gap between Adoption and Understanding in NLP**. Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. **Anais**...Online: Association for Computational Linguistics, ago. 2021. Disponível em: <https://aclanthology.org/2021.findings-acl.340>

BINZ, M. et al. How should the advancement of large language models affect the practice of science? **Proceedings of the National Academy of Sciences**, v. 122, n. 5, p. e2401227121, fev. 2025.

BIRD, S. **Decolonising Speech and Language Technology**. Proceedings of the 28th International Conference on Computational Linguistics. **Anais**...Barcelona, Spain (Online): International Committee on Computational Linguistics, dez. 2020. Disponível em: <https://aclanthology.org/2020.coling-main.313>

BIRD, S. **Local Languages, Third Spaces, and other High-Resource Scenarios**. Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). **Anais**...Dublin, Ireland: Association for Computational Linguistics, 2022. Disponível em: <https://aclanthology.org/2022.acl-long.539>

BLODGETT, S. L. et al. **Language (Technology) is Power: A Critical Survey of "Bias" in NLP**. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. **Anais**...Online: Association for Computational Linguistics, 2020. Disponível em: <https://www.aclweb.org/anthology/2020.acl-main.485>

BRENNAN, K.; KAK, A.; WEST, S. M. The AGI Mythology: The Argument to End All Arguments. Em: **Artificial Power: 2025 Landscape Report**. [s.l.] AI Now Institute, 2025.

BRENNEIS, A. Assessing dual use risks in AI research: necessity, challenges and mitigation strategies. **Research Ethics**, v. 21, n. 2, p. 302–330, 2025.

BROUSSARD, M. **More Than a Glitch: Confronting Race, Gender, and Ability Bias in Tech**. 1. ed. Cambridge, Massachusetts: MIT Press, 2023.

CENTRE FOR INFORMATION POLICY LEADERSHIP, C. **The Limitations of Consent as a Legal Basis for Data Processing in the Digital Society**. Washington DC, London, BrusselsCentre for Information Policy Leadership & Hunton Andrews Kurth LLP; Bae Kim & Lee, 2024. Disponível em: <https://www.informationpolicycentre.com/uploads/5/7/1/0/57104281/cipl_bkl_limitations_of_consent_legal_basis_data_processing_dec24.pdf>

CHANDRA, M. et al. **From Lived Experience to Insight: Unpacking the Psychological Risks of Using AI Conversational Agents**. Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency. **Anais**...Athens Greece: ACM, jun. 2025. Disponível em: <https://dl.acm.org/doi/10.1145/3715275.3732063>

CHEN, K. et al. **How Susceptible are Large Language Models to Ideological Manipulation?** Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing. **Anais**...Miami, Florida, USA: Association for Computational Linguistics, a2024. Disponível em: <https://aclanthology.org/2024.emnlp-main.952>

CHEN, T. et al. **CopyBench: Measuring Literal and Non-Literal Reproduction of Copyright-Protected Text in Language Model Generation**. (Y. Al-Onaizan, M. Bansal, Y.-N. Chen, Eds.)Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing. **Anais**...Miami, Florida, USA: Association for Computational Linguistics, nov. b2024. Disponível em: <https://aclanthology.org/2024.emnlp-main.844/>

COHEN, K. B. et al. **Three Dimensions of Reproducibility in Natural Language Processing**. Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018). **Anais**...Miyazaki, Japan: European Language Resources Association (ELRA), 2018. Disponível em: <https://aclanthology.org/L18-1025>

COULDRY, N.; MEJIAS, U. A. **The Costs of Connection: How Data Is Colonizing Human Life and Appropriating It for Capitalism**. [s.l.] Stanford University Press, 2019.

CUEVA, R. et al. **Inteligência Artificial e Desafios Regulatórios**. [s.l.] Forense, 2026.

CURRY, A.; CERCAS CURRY, A. **Computer says "No": The Case Against Empathetic Conversational AI**. (A. Rogers, J. Boyd-Graber, N. Okazaki, Eds.)Findings of the Association for Computational Linguistics: ACL 2023. **Anais**...Toronto, Canada: Association for Computational Linguistics, jul. 2023. Disponível em: <https://aclanthology.org/2023.findings-acl.515/>

D'ARCY, A.; BENDER, E. M. Ethics in Linguistics. **Annual Review of Linguistics**, v. 9, n. 1, p. 49–69, jan. 2023.

DOU, G. et al. **Avoiding Copyright Infringement via Large Language Model Unlearning**. (L. Chiruzzo, A. Ritter, L. Wang, Eds.)Findings of the Association for Computational Linguistics: NAACL 2025. **Anais**...Albuquerque, New Mexico: Association for Computational Linguistics, abr. 2025. Disponível em: <https://aclanthology.org/2025.findings-naacl.288/>

EDITORIAL. Emotional risks of AI companions demand attention. **Nature Machine Intelligence**, v. 7, n. 7, p. 981–982, jul. 2025.

ETHAYARAJH, K.; JURAFSKY, D. **Utility is in the Eye of the User: A Critique of NLP Leaderboards**. Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). **Anais**...Online: Association for Computational Linguistics, 2020. Disponível em: <https://www.aclweb.org/anthology/2020.emnlp-main.393>

FISHER, J. et al. **Biased LLMs can Influence Political Decision-Making**. (W. Che et al., Eds.)Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). **Anais**...Vienna, Austria: Association for Computational Linguistics, jul. 2025. Disponível em: <https://aclanthology.org/2025.acl-long.328/>

FLORIDI, L. Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical.

Philosophy & Technology, v. 32, n. 2, p. 185–193, jun. 2019.

FOKKENS, A. et al. **Offspring from Reproduction Problems: What Replication Failure Teaches Us**. Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). **Anais**...Sofia, Bulgaria: Association for Computational Linguistics, ago. 2013. Disponível em: <https://aclanthology.org/P13-1166>

FORT, K.; ADDA, G.; COHEN, K. B. Amazon Mechanical Turk: Gold Mine or Coal Mine? **Computational Linguistics**, v. 37, n. 2, p. 413–420, jun. 2011.

FRANCISCO, P. A. P.; HUREL, L. M.; RIELLI, M. M. **Regulação do Reconhecimento Facial no Setor Público: Avaliação de Experiências Internacionais**. Instituto Igarapé - DataPrivacyBR, 2020. Disponível em: <https://www.dataprivacybr.org/wp-content/uploads/2021/11/regulacao-do-reconhecimento-facial-no-setor-publico.pdf>

GEBRU, T. et al. Datasheets for datasets. **Communications of the ACM**, v. 64, n. 12, p. 86–92, dez. 2021.

GEBRU, T.; BENDER, E. M.; MCMILLAN-MAJOR, A. **Statement from the listed authors of Stochastic Parrots on the "AI pause" letter.**, 2023. Disponível em: <https://www.dair-institute.org/blog/letter-statement-March2023/>

GEBRU, T.; TORRES, E. P. The TESCREAL bundle: Eugenics and the promise of utopia through artificial general intelligence. **First Monday**, abr. 2024.

GOLUMBIA, D. **Cyberlibertarianism: The Right-Wing Politics of Digital Technology**. [s.l.] University of Minnesota Press, 2024.

GRAY, M. L.; SURI, S. **Ghost work: How to stop Silicon Valley from building a new global underclass**. [s.l.] Harper Business, 2019.

GRAY WIDDER, D.; WEST, S.; WHITTAKER, M. Open (For Business): Big Tech, Concentrated Power, and the Political Economy of Open AI. **SSRN Electronic Journal**, 2023.

HADDAD, I. **Artificial Intelligence and Data in Open Source**. Linux Foundation, 2022. Disponível em: <https://www.linuxfoundation.org/hubfs/LF%20Research/Artificial%20Intelligence%20and%20Data%20in%20Open%20Source%20-%20Report.pdf?hsLang=en>

HAVENS, L. et al. **Situated Data, Situated Systems: A Methodology to Engage with Power Relations in Natural Language Processing Research**. (M. R. Costa-jussà et al., Eds.)Proceedings of the Second Workshop on Gender Bias in Natural Language Processing. **Anais**...Barcelona, Spain (Online): Association for Computational Linguistics, dez. 2020. Disponível em: <https://aclanthology.org/2020.gebnlp-1.10/>

HENDREN, S. All technology is assistive: Six design rules on disability. Em: SAYERS, J. (Ed.). **Making Things and Drawing Boundaries: Experiments in the Digital Humanities**. [s.l.] University of Minnesota Press, Minneapolis, MN, 2014.

HERSHCOVICH, D. et al. **Towards Climate Awareness in NLP Research**. Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing. **Anais**...Abu Dhabi, United Arab Emirates: Association for Computational Linguistics, 2022. Disponível em: <https://aclanthology.org/2022.emnlp-main.159>

HICKS, M. T.; HUMPHRIES, J.; SLATER, J. ChatGPT is bullshit. **Ethics and Information Technology**, v. 26, n. 2, p. 38, jun. 2024.

HOFMANN, V. et al. AI generates covertly racist decisions about people based on their dialect. **Nature**, v. 633, n. 8028, p. 147–154, set. 2024.

HOOKER, S. Moving beyond "algorithmic bias is a data problem". **Patterns**, v. 2, n. 4, p. 100241, abr. 2021.

HOVY, D.; SPRUIT, S. L. **The Social Impact of Natural Language Processing**. Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). **Anais**...Berlin, Germany: Association for Computational Linguistics, 2016. Disponível em: <http://aclweb.org/anthology/P16-2096>

HUANG, J.; YANG, D.; POTTS, C. **Demystifying Verbatim Memorization in Large Language Models**. (Y. Al-Onaizan, M. Bansal, Y.-N. Chen, Eds.)Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing. **Anais**...Miami, Florida, USA: Association for Computational Linguistics, nov. 2024. Disponível em: <https://aclanthology.org/2024.emnlp-main.598/>

JIN, Z. et al. **How Good Is NLP? A Sober Look at NLP Tasks through the Lens of Social Impact**. Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021. **Anais**...Online: Association for Computational Linguistics, 2021. Disponível em: <https://aclanthology.org/2021.findings-acl.273>

JOHNSTON, S. F. Alvin Weinberg and the Promotion of the Technological Fix. **Technology and Culture**, v. 59, n. 3, p. 620–651, 2018.

KARAMOLEGKOU, A. et al. **Copyright Violations and Large Language Models**. Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing. **Anais**...Singapore: Association for Computational Linguistics, 2023. Disponível em: <https://aclanthology.org/2023.emnlp-main.458>

KLENK, M. How Do Technological Artefacts Embody Moral Values? **Philosophy & Technology**, v. 34, n. 3, p. 525–544, set. 2021.

KOGKALIDIS, K.; CHATZIKYRIAKIDIS, S. **On Tables with Numbers, with Numbers**. (S. Truong et al., Eds.)Proceedings of the 1st Workshop on Language Models for Underserved Communities (LM4UC 2025). **Anais**...Albuquerque, New Mexico: Association for Computational Linguistics, 2025. Disponível em: <https://aclanthology.org/2025.lm4uc-1.12/>

LASOTA, L. **Regulating Corporate Behaviour in Digital Ecosystems: Increasing Fairness and Contestability of Digital Markets with Free Software**. MIC 2023: Toward Green, Inclusive, and Digital Growth. **Anais**...a2023.

LASOTA, L. The European Union's AI act from the perspective of Open Technologies. Em: ALEGRE, M.; FÜRST, H. (Eds.). **Advocacia & Bioética: Novas Tecnologias**. https://www.editoracasadodireito.com.br/produto/novas-tecnologias. São Paulo: Casa do Direito, 2023b.

LASOTA, L. The CRA and the Challenges of Regulating Cybersecurity in Open Environments: The Case of Free and Open Source Software. Em: **Digital Decade: How the EU shapes digitalisation**. Berlin: Weizenbaum Institute for the Networked Society - The German Internet Institute, 2025.

LASOTA, L.; SINGHAL, N. **Free Software and AI openness: Overcoming challenges in the licensing world**. [s.l.] Zenodo, abr. 2024. Disponível em: <https://zenodo.org/doi/10.5281/zenodo.10966136>.

LEIDNER, J. L.; PLACHOURAS, V. **Ethical by Design: Ethics Best Practices for Natural Language Processing**. Proceedings of the First ACL Workshop on Ethics in Natural Language Processing. **Anais**...Valencia, Spain: Association for Computational Linguistics, abr. 2017. Disponível em: <https://aclanthology.org/W17-1604>

LOBO, P. **Profiling na Lei Geral de Proteção de Dados: O Livre Desenvolvimento da Personalidade em Face da Governamentalidade Algorítmica**. 1. ed. [s.l.] Editora Foco, 2022.

MALEKI, N.; PADMANABHAN, B.; DUTTA, K. **AI Hallucinations: A Misnomer Worth Clarifying**. 2024 IEEE Conference on Artificial Intelligence (CAI). **Anais**...Singapore, Singapore: IEEE, jun. 2024. Disponível em: <https://ieeexplore.ieee.org/document/10605268/>

MCMILLAN-MAJOR, A.; BENDER, E. M.; FRIEDMAN, B. Data Statements: From Technical Concept to Community Practice. **ACM Journal on Responsible Computing**, p. 3594737, 2023.

MCQUILLAN, D. **Resisting AI: an anti-fascist approach to artificial intelligence**. Bristol, UK: [s.n.].

MEJIAS, U. A.; COULDRY, N. Datafication. **Internet Policy Review**, v. 8, n. 4, nov. 2019.

MICELI, M. et al. **Who Trains the Data for European Artificial Intelligence?** The Left, DiPLab, Weizenbaum Institute; DAIR Institute., 2024. Disponível em: <https://hal.science/hal-04662589v1>

MILLER, B. Is Technology Value-Neutral? **Science, Technology, & Human Values**, v. 46, n. 1, p. 53–80, jan. 2021.

MITCHELL, M. et al. **Model Cards for Model Reporting**. Proceedings of the Conference on Fairness, Accountability, and Transparency. **Anais**...Atlanta GA USA: ACM, jan. 2019. Disponível em: <https://dl.acm.org/doi/10.1145/3287560.3287596>

MOHAMMAD, S. **Ethics Sheets for AI Tasks**. Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). **Anais**...Dublin, Ireland: Association for Computational Linguistics, 2022. Disponível em: <https://aclanthology.org/2022.acl-long.573>

MORESCHI, B.; PEREIRA, G.; COZMAN, F. G. The Brazilian Workers in Amazon Mechanical Turk: Dreams and realities of ghost workers. **Revista Contracampo**, v. 39, n. 1, abr. 2020.

MOROZOV, E. **To save everything, click here : the folly of technological solutionism**. New York, NY: PublicAffairs, 2014.

MUNGER, K. **Chatbots for Good and Evil**. EACL via Underline Science Inc., 2023. Disponível em: <https://underline.io/lecture/72154-chatbots-for-good-and-evil>

NOVOBILSKÁ, L. Free and Open Source Software Licensing Requirements and Copyright Infringement Involving Artificial Intelligence Technologies. 2023.

O'NEIL, C. **Weapons of math destruction : how big data increases inequality and threatens democracy**. First edition ed. New York: [s.n.].

OLIVEIRA, L. L. Inteligência artificial e desigualdade social: o impacto do colonialismo digital nas políticas públicas. **Internet & Sociedade**, v. 5, n. 1, 2024.

PAN, Y. et al. **On the Risk of Misinformation Pollution with Large Language Models**. Findings of the Association for Computational Linguistics: EMNLP 2023. **Anais**...Singapore: Association for Computational Linguistics, 2023. Disponível em: <https://aclanthology.org/2023.findings-emnlp.97>

PARMAR, M. et al. **Don't Blame the Annotator: Bias Already Starts in the Annotation Instructions**. Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics. **Anais**...Dubrovnik, Croatia: Association for Computational Linguistics, 2023. Disponível em: <https://aclanthology.org/2023.eacl-main.130>

PAULLADA, A. et al. Data and its (dis)contents: A survey of dataset development and use in machine learning research. **Patterns**, v. 2, n. 11, p. 100336, nov. 2021.

PLOUG, T. The Right Not to Be Subjected to AI Profiling Based on Publicly Available Data—Privacy and the Exceptionalism of AI Profiling. **Philosophy & Technology**, v. 36, n. 1, p. 14, mar. 2023.

RAJI, D. et al. **AI and the Everything in the Whole Wide World Benchmark**. (J. Vanschoren, S. Yeung, Eds.)Proceedings of the Neural Information Processing Systems Track on Datasets and Benchmarks. **Anais**...Curran, 2021. Disponível em: <https://datasets-benchmarks-proceedings.neurips.cc/paper_files/paper/2021/file/084b6fbb10729ed4da8c3d3f5a3ae7c9-Paper-round2.pdf>

REHAK, R. **AI Narrative Breakdown. A Critical Assessment of Power and Promise**. Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency. **Anais**...: FAccT '25.New York, NY, USA: Association for Computing Machinery, 2025. Disponível em: <https://doi.org/10.1145/3715275.3732083>

ROGERS, A. **Changing the World by Changing the Data**. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers). **Anais**...Online: Association for Computational Linguistics, 2021. Disponível em: <https://aclanthology.org/2021.acl-long.170>

ROGERS, A.; BALDWIN, T.; LEINS, K. **"Just What do You Think You're Doing, Dave?" A Checklist for Responsible Data Use in NLP**. Findings of the Association for Computational Linguistics: EMNLP 2021. **Anais**...Punta Cana, Dominican Republic: Association for Computational Linguistics, 2021. Disponível em: <https://aclanthology.org/2021.findings-emnlp.414>

SAMPAIO, R. C.; SABBATINI, M.; LIMONGI, R. Diretrizes para o uso ético e responsável da inteligência artificial generativa: um guia prático para pesquisadores. **Boletim Técnico do PPEC**, v. 10, p. e025003–e025003, 2024.

SCHAAKE, M. **The Tech Coup: How to Save Democracy from Silicon Valley**. [s.l.] Princeton University Press, 2024.

SCHEUERMAN, M. K.; HANNA, A.; DENTON, E. Do Datasets Have Politics? Disciplinary Values in Computer Vision Dataset Development. **Proceedings of the ACM on Human-Computer Interaction**, v. 5, n. CSCW2, p. 1–37, out. 2021.

SCHIRRU, L. **Direito autoral e Inteligência Artificial: autoria e titularidade nos produtos da IA**. [s.l.] Dialetica, 2023.

SCHLANGEN, D. **Targeting the Benchmark: On Methodology in Current Natural Language Processing Research**. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers). **Anais**...Online: Association for Computational Linguistics, ago. 2021. Disponível em: <https://aclanthology.org/2021.acl-short.85>

SCHWARTZ, L. **Primum Non Nocere: Before working with Indigenous data, the ACL must confront ongoing colonialism**. Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). **Anais**...Dublin, Ireland: Association for Computational Linguistics, 2022. Disponível em: <https://aclanthology.org/2022.acl-short.82>

SELVAN, R. et al. **Carbon Footprint of Selecting and Training Deep Learning Models for Medical Image Analysis**. (L. Wang et al., Eds.)Medical Image Computing and Computer Assisted Intervention – MICCAI 2022. **Anais**...Cham: Springer Nature Switzerland, 2022. Disponível em: <https://doi.org/10.1007/978-3-031-16443-9_49>

SHAH, D.; SCHWARTZ, H. A.; HOVY, D. **Predictive Biases in Natural Language Processing Models: A Conceptual Framework and Overview**. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. **Anais**...Online: Association for Computational Linguistics, 2020. Disponível em: <http://arxiv.org/abs/1912.11078>

SHENG, E. et al. **The Woman Worked as a Babysitter: On Biases in Language Generation**. (K. Inui et al., Eds.)Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). **Anais**...Hong Kong, China: Association for Computational Linguistics, nov. 2019. Disponível em: <https://aclanthology.org/D19-1339/>

SHMUELI, B. et al. **Beyond Fair Pay: Ethical Implications of NLP Crowdsourcing**. Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. **Anais**...Online: Association for Computational Linguistics, 2021.

Disponível em: <https://aclanthology.org/2021.naacl-main.295>

ŞIMŞEK, C. **AI resistance: Who says no to AI and why?** Zenodo, 2025. Disponível em: <https://zenodo.org/doi/10.5281/zenodo.16893847>

ŞIMŞEK, C.; YASAR, A. G. From Rejection to Regulation: Mapping the Landscape of AI Resistance. 2025.

SØGAARD, A.; PLANK, B.; HOVY, D. **Selection Bias, Label Bias, and Bias in Ground Truth**. (Q. Liu, F. Xia, Eds.)Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Tutorial Abstracts. **Anais**...Dublin, Ireland: Dublin City University; Association for Computational Linguistics, ago. 2014. Disponível em: <https://aclanthology.org/C14-3005/>

SOLAIMAN, I. et al. Evaluating the Social Impact of Generative AI Systems. Em: **The Oxford Handbook of the Foundations and Regulation of Generative AI**. [s.l.] Oxford University Press, 2025.

STRUBELL, E.; GANESH, A.; MCCALLUM, A. **Energy and Policy Considerations for Deep Learning in NLP**. Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. **Anais**...Florence, Italy: Association for Computational Linguistics, 2019. Disponível em: <https://www.aclweb.org/anthology/P19-1355>

SURESH, H.; GUTTAG, J. V. A Framework for Understanding Sources of Harm throughout the Machine Learning Life Cycle. **Equity and Access in Algorithms, Mechanisms, and Optimization**, p. 1–9, out. 2021.

TAMKIN, A. et al. **Evaluating and Mitigating Discrimination in Language Model Decisions.**, 2023. Disponível em: <https://arxiv.org/abs/2312.03689>

THUN, M. VON; HANLEY, D. A. **Stopping Big Tech from Becoming Big AI: A Roadmap for Using Competition Policy to Keep Artificial Intelligence Open for All**. Open Markets Institute, 2024. Disponível em: <https://www.openmarketsinstitute.org/publications/report-stopping-big-tech-big-ai-roadmap>

TONIAZZO, D.; BARBOSA, T.; RUARO, R. O Direito à Explicação nas Decisões Automatizadas: uma Abordagem Comparativa Entre o Ordenamento Brasileiro e Europeu. **Revista Internacional Consinter de Direito**, v. 13, p. 55–69, dez. 2021.

ULMER, D. et al. **Experimental Standards for Deep Learning in Natural Language Processing Research**. Findings of the Association for Computational Linguistics: EMNLP 2022. **Anais**...Abu Dhabi, United Arab Emirates: Association for Computational Linguistics, dez. 2022. Disponível em: <https://aclanthology.org/2022.findings-emnlp.196>

VAROQUAUX, G.; LUCCIONI, S.; WHITTAKER, M. **Hype, Sustainability, and the Price of the Bigger-is-Better Paradigm in AI**. Proceedings of the 2025 ACM Conference on Fairness, Accountability, and Transparency. **Anais**...Athens Greece: ACM, jun. 2025. Disponível em: <https://dl.acm.org/doi/10.1145/3715275.3732006>

WEIDINGER, L. et al. Ethical and social risks of harm from Language Models. **arXiv:2112.04359 [cs]**, dez. 2021.

WEIGEND, A. **Data for the People: How to Make Our Post-Privacy Economy Work for You**. [s.l.] Basic Books, 2017.

WEIZENBAUM, J. **Computermacht und Gesellschaft : freie Reden / Joseph Weizenbaum. Hrsg. von Gunna Wendt ..** Original-Ausgabe, 1. Auflage ed. Frankfurt am Main: [s.n.].

WENDEHORST, C. Liability for Artificial Intelligence: The Need to Address Both Safety Risks and Fundamental Rights Risks. Em: VOENEKY, S. et al. (Eds.). **The Cambridge Handbook of**

Responsible Artificial Intelligence: Interdisciplinary Perspectives. Cambridge Law Handbooks. [s.l.] Cambridge University Press, 2022. p. 187–209.

WESTENBERGER, J.; SCHULER, K.; SCHLEGEL, D. Failure of AI projects: understanding the critical factors. **Procedia Computer Science**, v. 196, p. 69–76, 2022.

WHITTAKER, M. et al. Disability, bias, and AI. **AI Now Institute**, v. 8, n. 11, 2019.

WHITTAKER, M. The steep cost of capture. **Interactions**, v. 28, n. 6, p. 50–55, nov. 2021.

WIELING, M.; RAWEE, J.; VAN NOORD, G. Reproducibility in Computational Linguistics: Are We Willing to Share? **Computational Linguistics**, v. 44, n. 4, p. 641–649, dez. 2018.

WRIGHT, B. **Manufacturing Reality: Slavoj Zizek and the Reality of the Virtual**. LondonBen Wright Film Productions, 2004.

WU, T. **The attention merchants : from the daily newspaper to social media : how our time and attention is harvested and sold**. London: [s.n.].

XU, Q.; HE, X. **Security Challenges in Natural Language Processing Models**. Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: Tutorial Abstracts. **Anais**...Singapore: Association for Computational Linguistics, 2023. Disponível em: <https://aclanthology.org/2023.emnlp-tutorial.2>

YANG, T. et al. **Ethics of Data Work. Principles for Academic Data Work Requesters**. Weizenbaum Institute, 2025. Disponível em: <https://www.weizenbaum-library.de/handle/id/920>

ZHANG, D.; XU, Z.; ZHAO, W. **LLMs and Copyright Risks: Benchmarks and Mitigation Approaches**. (M. Lomeli, S. Swayamdipta, R. Zhang, Eds.)Proceedings of the 2025 Annual Conference of the Nations of the Americas Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 5: Tutorial Abstracts). **Anais**...Albuquerque, New Mexico: Association for Computational Linguistics, 2025. Disponível em: <https://aclanthology.org/2025.naacl-tutorial.7/>

ZUBOFF, S. **The age of surveillance capitalism : the fight for a human future at the new frontier of power**. First edition ed. New York: [s.n.].