

# Uma revisão de escopo sobre importância da neutralidade de viés e transparência dos dados em sistemas de recomendações



Eduardo Alves de Oliveira Freitas\* e Humberto Lidio Antonelli

Universidade Federal de Mato Grosso do Sul, Campus de Três Lagoas, Três Lagoas, MS, Brasil

\*Informação de contato: <[alves.freitas@ufms.br](mailto:alves.freitas@ufms.br)>

*Trabalho de Conclusão de Curso entregue em 28 de novembro de 2024.*

## Resumo

Os sistemas de recomendação são fundamentais para personalizar experiências em plataformas digitais, influenciando desde compras online até o consumo de conteúdo. Apesar de seus benefícios, como eficiência e conveniência, enfrentam críticas sobre vieses, falta de transparência e impactos sociais negativos, levantando questões éticas sobre equidade e confiabilidade. Este trabalho revisou como esses aspectos têm sido abordados na literatura, focando na neutralidade de viés e transparência dos dados. A revisão identificou 14 estudos relevantes e revelou que os principais tipos de viés incluem popularidade, seleção e homogeneização, enquanto a falta de transparência prejudica a confiança e limita a diversidade das recomendações. Para mitigar esses problemas, destacaram-se a diversificação das sugestões, explicações claras aos usuários e técnicas para regulação de vieses. Os resultados também apontam para desafios como a opacidade algorítmica e a necessidade de regulamentação adaptável. Conclui-se que a criação de sistemas éticos depende de combinar neutralidade, transparência e responsabilidade social, orientando melhorias nos algoritmos e políticas mais inclusivas.

**Palavras-chaves:** sistemas de recomendação. neutralidade de viés. transparência de dados. ética. inteligência artificial.

## 1. Introdução

Nos últimos anos, a Inteligência Artificial (IA) tornou-se uma parte essencial do dia a dia, sendo utilizada em diversas áreas, como assistentes virtuais e automação de processos. Ela oferece benefícios como a personalização e a eficiência no tratamento de grandes volumes de dados, impactando setores como saúde, educação e entretenimento (STINSON, 2022b).

Os sistemas de recomendação surgiram com o objetivo de facilitar a descoberta de conteúdos e produtos, sendo popularizados por plataformas como Amazon e Netflix. Hoje, são amplamente usados em sites, aplicativos e serviços para oferecer uma experiência personalizada ao usuário, aumentando o engajamento e as vendas (STINSON, 2022a; STINSON, 2022b).

No entanto, esses sistemas também podem reforçar vieses e discriminações, promovendo conteúdos populares em detrimento de itens de nicho. A falta de transparência nos algoritmos gera desconfiança e frustrações, além de poder marginalizar grupos específicos (STINSON, 2022b; STINSON, 2022a). Esses problemas serão o foco deste trabalho, com ênfase na neutralidade de viés e na transparência dos dados.

Assim, considerando a problemática sobre o emprego de sistemas de recomendação, neste artigo é apresentada uma revisão de escopo que visa investigar como o impacto ético e social dos sistemas de recomendação tem sido abordados na literatura. O foco da investigação concentra-se, especialmente, nas questões de neutralidade de viés e transparência dos dados, com objetivo de identificar implicações práticas e diretrizes discutidas na área para promover o desenvolvimento ético dessas tecnologias.

O restante deste trabalho está organizado da seguinte maneira: na [Seção 2](#) é introduzido o referencial teórico sobre os principais conceitos sobre o tema da pesquisa. Na [Seção 3](#) são descritos os procedi-

mentos metodológicos utilizados nesta pesquisa, explicando todas as etapas seguidas no protocolo de revisão. Na **Seção 4** são discutidos os resultados obtidos, respondendo às questões de pesquisa. Por fim, na **Seção 5** são apresentadas as conclusões levantadas por intermédio desta pesquisa.

## 2. Referencial teórico

Esta sessão apresenta uma fundamentação teórica sobre sistemas de recomendação, inteligência artificial e suas tecnologias. Além disso, explana-se sobre a questão de injustiça relacionada a esses conceitos.

### 2.1 Inteligência Artificial

A Inteligência Artificial, pode ser definida como uma ferramenta que auxilia na tomada de decisões diante dos problemas apresentados (BAIA, 2024). Esta área de pesquisa abrange uma variedade de abordagens e técnicas, com destaque para o “Aprendizado de Máquina” como um dos principais tipos de IA no contexto de sistemas de recomendação. Segundo McCarthy (2007), a inteligência artificial é definida como:

[...] a ciência e engenharia de fazer máquinas inteligentes, especialmente programas de computador inteligentes. Está relacionada à tarefa semelhante de usar computadores para entender a inteligência humana, mas a IA não precisa se limitar a métodos que são biologicamente observáveis.

O termo “Inteligência Artificial” foi cunhado por John McCarthy, embora seja importante ressaltar que Alan Turing já havia concebido um conceito similar anteriormente (ZENDESK, 2024). Turing postulou a ideia de que máquinas poderiam utilizar informações disponíveis para resolver problemas e tomar decisões, assim como seres humanos. Suas contribuições pioneiras nesse campo ajudaram a estabelecer os fundamentos para o desenvolvimento posterior da IA.

O aprendizado de máquina, uma subárea da IA, foca no desenvolvimento de algoritmos que permitem que computadores aprendam a partir de dados. Durante os anos 1980, técnicas como o “*deep learning*” foram popularizadas por pesquisadores como John Hopfield e David Rumelhart, permitindo que computadores aprendessem usando experiência. Na mesma época, sistemas especialistas foram desenvolvidos para automatizar processos de tomada de decisão em diversas indústrias, demonstrando aplicações práticas da IA (ANYOHA, 2017).

Esses avanços permitiram o desenvolvimento de técnicas mais sofisticadas para processamento de linguagem natural e visão computacional nos anos 1990, impulsionados pelo aumento do poder computacional e pela disponibilidade de grandes volumes de dados. O uso de métodos estatísticos e redes neurais convolucionais possibilitou modelos mais precisos e flexíveis para interpretar dados não estruturados, como texto e imagens, expandindo ainda mais as aplicações da IA (COPELAND, 2023).

Os algoritmos baseados em IA vêm ganhando cada vez mais espaço no mercado de tecnologia. No ano de 2022, por exemplo, um dos assuntos mais comentados foi sobre o “Chat GPT”<sup>1</sup>, desenvolvido pela OpenAI. Diversos trabalhos estão sendo facilitados pelo uso de IA, como o próprio metrô da cidade de São Paulo, que está passando por mudanças de segurança com o auxílio dessas tecnologias (CAMARGO, 2023).

### 2.2 Injustiça

Os algoritmos de recomendação possuem a capacidade, mesmo que involuntariamente, de perpetuar preconceitos, sejam eles de ordem socioeconômica, de gênero ou racial (PASCUAL, 2019). Essas

<sup>1</sup><<https://openai.com/blog/chatgpt>>

ferramentas automatizadas, ao analisarem e processarem grandes volumes de dados, podem gerar resultados enviesados que refletem desigualdades existentes na sociedade.

A preocupação com a injustiça social nos algoritmos ganhou destaque à medida que a IA começou a ser utilizada em decisões críticas, como na justiça criminal, educação e saúde. Esses algoritmos, treinados com dados históricos, muitas vezes perpetuam preconceitos sociais preexistentes (ANGWIN *et al.*, 2016). Por exemplo, uma investigação da ProPublica em 2016 revelou que algoritmos de avaliação de risco usados no sistema de justiça criminal dos EUA tendiam a classificar desproporcionalmente pessoas negras como de alto risco de reincidência (ANGWIN *et al.*, 2016). Isso levantou questões éticas sobre a necessidade de transparência e revisão contínua para evitar a perpetuação de injustiças sociais (LEDFORD, 2019).

A definição de “injustiça social” tende a ser múltipla, a depender do aspecto e das condições em que é analisada. De modo simples e sucinto, o padrão de injustiça ocorre quando dois indivíduos semelhantes e em iguais condições recebem tratamento desigual. (CAMARGO, 2023)

Um dos principais desafios decorrentes do viés e da imparcialidade nos sistemas de recomendação é a injustiça social, na qual se criam bolhas sociais que marginalizam determinados grupos de pessoas. Como um algoritmo é somente um programa com inteligência artificial, ele carece de conceitos humanos como ética, podendo assim assimilar e propagar padrões discriminatórios, prejudicando a imparcialidade da sua atuação para todos os indivíduos.

A injustiça se manifesta de diversas maneiras na sociedade, e um exemplo disso pode ser observado no caso da Amazon. Em 2018, a empresa tomou a decisão de deixar de utilizar algoritmos de recomendação em seus processos de contratação de recursos humanos. Isso ocorreu após uma série de experiências com o software em produção, que revelaram uma preocupante tendência de favorecer candidatos do sexo masculino em detrimento das mulheres, mesmo quando estas possuíam qualificações idênticas ou até superiores. O algoritmo foi criado com base em mais de uma década de dados de contratação, mas, mesmo assim, acabou contribuindo para a perpetuação desse problema dentro da empresa (CRISTINA, 2021). Mesmo ao tentar isolar certos termos, a Amazon constatou que os problemas persistiam.

O problema da injustiça gerada pela inteligência artificial não se restringe apenas ao ambiente corporativo. Um exemplo disso ocorreu na rede social Twitter em 2015, quando foi acusada de usar um algoritmo de reconhecimento de imagem que propagava o racismo. Esse software tinha como objetivo destacar objetos ou pessoas nas fotos para facilitar a experiência dos usuários. Alguns usuários do Twitter notaram que, ao publicar fotos de diferentes tamanhos e formatos com pessoas negras e brancas, as pessoas brancas sempre recebiam destaque na imagem (MORAIS, 2020). A empresa emitiu uma nota oficial se desculpando pelo incidente e afirmou que haviam sido realizados diversos testes no algoritmo antes de sua implementação, mas que esses problemas não foram identificados. No entanto, eles anunciaram que iriam revisar o software e se concentrar na resolução desse problema.

### 2.3 Sistemas de recomendação

Os sistemas de recomendação são concebidos com a finalidade de viabilizar a facilitação na tomada de decisões. Eles desempenham um papel de dupla natureza, na medida em que são capazes de gerar simultaneamente informações relevantes tanto para os desenvolvedores/empresas envolvidos quanto para os usuários do sistema, por meio de recomendações personalizadas.

Sistemas de recomendação são ferramentas e técnicas de software que proveem sugestões de itens para ser de utilidade para um usuário. As sugestões se relacionam a vários processos de tomada de decisão, tais como quais itens comprar, que música ouvir ou que notícias *online* para ler. (RICCI; ROKACH; SHAPIRA, 2015)

Os sistemas de recomendação podem ser classificados em diferentes categorias, dependendo da forma como processam os dados e geram sugestões personalizadas (RICCI; ROKACH; SHAPIRA, 2015). Esses sistemas são amplamente utilizados em diversos setores, como comércio eletrônico, entretenimento e educação, com o objetivo de oferecer produtos, serviços ou conteúdos de maneira direcionada, promovendo experiências mais relevantes e satisfatórias para os usuários. Por meio da análise de grandes volumes de dados, eles auxiliam na tomada de decisão, contribuindo para aumentar o engajamento e otimizar os resultados para empresas e consumidores.

Entre as principais abordagens, destacam-se os **sistemas baseados em conteúdo**, que utilizam informações sobre as características dos itens para fazer recomendações personalizadas. Por exemplo, em uma plataforma de filmes, o sistema analisa atributos como gênero, diretor e elenco dos filmes que o usuário já assistiu para sugerir títulos similares. A principal vantagem desse tipo de abordagem é a capacidade de recomendar itens relevantes com base no histórico do usuário, mesmo para novos usuários, caso existam informações sobre suas preferências iniciais (RICCI; ROKACH; SHAPIRA, 2015). Entretanto, esses sistemas tendem a criar “bolhas de conteúdo”, limitando a diversidade das sugestões ao focarem excessivamente nas preferências previamente demonstradas pelo usuário (STINSON, 2022b).

Outra categoria importante são os **sistemas de filtro colaborativo**, nos quais as recomendações são feitas com base em similaridades entre usuários. O sistema identifica grupos de usuários com preferências semelhantes e sugere itens populares dentro desses grupos. A principal vantagem desse tipo de sistema é que ele não depende de atributos explícitos dos itens, sendo eficaz em cenários onde o conteúdo dos itens não está bem categorizado (RICCI; ROKACH; SHAPIRA, 2015). Contudo, esses sistemas enfrentam dois desafios principais: o problema de “cold start”, que ocorre quando novos usuários ou itens possuem poucas interações registradas, e o viés de popularidade, que favorece itens amplamente consumidos em detrimento de opções mais específicas ou de nicho (BEUTEL; CHEN; COVINGTON, 2019).

Por fim, há os **sistemas híbridos**, que integram técnicas de sistemas baseados em conteúdo e de filtro colaborativo, buscando combinar os pontos fortes de ambas as abordagens. Por exemplo, a Netflix utiliza um modelo híbrido que considera as características dos filmes e os padrões de consumo de grupos de usuários semelhantes (SIGLIANO; FAUSTINO, 2016). As principais vantagens desses sistemas incluem a redução dos problemas de “cold start” e a maior flexibilidade na geração de recomendações diversificadas (RICCI; ROKACH; SHAPIRA, 2015). Por outro lado, eles exigem maior capacidade computacional e enfrentam desafios similares às abordagens individuais, como o viés presente nos dados de treinamento (STINSON, 2022b).

Os algoritmos de recomendação desempenham um papel cada vez mais significativo na vida cotidiana, influenciando decisões tanto no âmbito profissional quanto pessoal. Seja em análises de crédito, processos de recrutamento ou sugestões de entretenimento em plataformas como Netflix e TikTok, essas tecnologias moldam experiências e interações de forma personalizada. As abordagens baseadas em conteúdo, filtros colaborativos e sistemas híbridos demonstram a evolução contínua desses mecanismos, cada qual com suas vantagens e desafios. Apesar dos avanços, questões como o viés de dados, o problema do “cold start” e a necessidade de maior capacidade computacional ainda representam obstáculos a serem superados. Assim, a busca por soluções que equilibrem eficiência, diversidade e ética nas recomendações permanece essencial para o desenvolvimento de sistemas cada vez mais inclusivos e eficazes.

## 2.4 Neutralidade de viés

A neutralidade de viés em sistemas de recomendação refere-se à capacidade de algoritmos fornecerem recomendações imparciais, evitando favorecimentos ou desvantagens a determinados grupos de

usuários ou tipos de conteúdo. Isso envolve minimizar padrões discriminatórios e promover uma representatividade equitativa nas sugestões apresentadas. A ausência de neutralidade pode levar à perpetuação de preconceitos existentes nos dados de treinamento, reforçando desigualdades sociais e limitando a diversidade de conteúdos acessados (STINSON, 2022a).

A busca pela neutralidade de viés é um desafio ético e técnico, exigindo a adoção de práticas como amostragens balanceadas, regularizações específicas e ajustes pós-processamento para assegurar que as recomendações sejam justas e representativas (DELDJOO; BEIGI; BERTI-EQUILLE, 2020). Nesse contexto, a neutralidade torna-se um elemento essencial para aumentar a confiabilidade e equidade nos sistemas.

## 2.5 Transparência e opacidade algorítmica

A transparência algorítmica refere-se à capacidade dos usuários e desenvolvedores de compreenderem como um sistema de recomendação opera, incluindo os critérios utilizados para gerar sugestões personalizadas e as fontes de dados empregadas no processo. A transparência é essencial para aumentar a confiança dos usuários, pois permite que eles entendam as decisões tomadas pelo sistema e identifiquem possíveis vieses ou falhas (TINTAREV; MASTHOFF, 2007).

Por outro lado, a opacidade algorítmica caracteriza-se pela dificuldade de interpretar e explicar o funcionamento interno desses sistemas, especialmente devido à complexidade de modelos avançados, como redes neurais profundas, e à falta de explicações claras. Essa opacidade pode levar a uma sensação de manipulação, reduzindo a percepção de controle dos usuários sobre as interações com o sistema (STINSON, 2022b).

A interação entre transparência e opacidade levanta questões éticas sobre a necessidade de explicações claras e detalhadas para mitigar a desconfiança e promover a responsabilidade nos sistemas de recomendação. Ambos os conceitos desempenham papéis cruciais na discussão sobre a neutralidade de viés e a confiabilidade dos algoritmos.

## 3. Metodologia

Este estudo fundamenta-se na classificação metodológica proposta por Gil (2022), que organiza a pesquisa científica com base nos objetivos e nos procedimentos técnicos. De acordo com essa abordagem, trata-se de uma pesquisa descritiva, de corte transversal e com abordagem qualitativa, voltada para mapear e analisar informações de maneira sistemática. No que diz respeito aos procedimentos técnicos, a pesquisa enquadra-se como documental, uma vez que utiliza dados secundários obtidos por meio de uma Revisão de Escopo (RE).

A RE é uma metodologia que tem sido considerada eficaz para identificar pesquisas dentro da literatura, oferecendo uma abordagem sistemática e abrangente para mapear rapidamente conceitos-chave relevantes dentro de um campo específico de estudo (MUNN *et al.*, 2018). As revisões de escopo são úteis para examinar evidências emergentes em áreas onde ainda não é claro quais questões específicas devem ser investigadas, oferecendo uma visão geral do conhecimento disponível e identificando lacunas para pesquisas futuras (PETERS *et al.*, 2020).

Neste trabalho, a RE foi conduzida seguindo as cinco etapas sequenciais recomendadas (PETERS *et al.*, 2020), conforme pode ser visto na Figura 1. Esse processo foi adotado para assegurar a organização e clareza na coleta e análise dos dados, proporcionando uma compreensão abrangente do estado atual da pesquisa sobre o tema.

Na primeira etapa, com o objetivo de fornecer uma visão geral das pesquisas que vêm sendo conduzidas a respeito da neutralidade de viés e transparência dos dados em sistemas de recomendações, foram

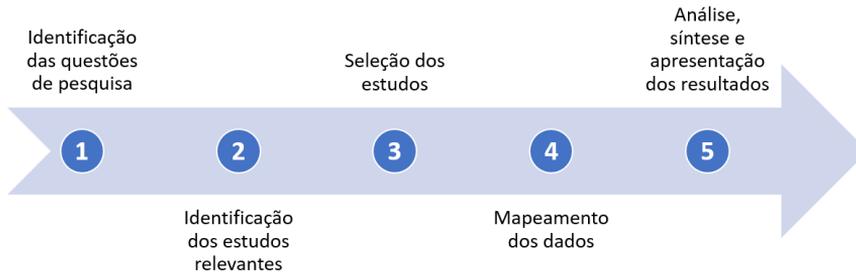


Figura 1. Etapas metodológicas empregadas na RE

formuladas as seguintes Questões de Pesquisa:

- QP1: Quais são os principais tipos de viés que podem afetar sistemas de recomendações?
- QP2: Como a transparência dos dados pode influenciar a percepção dos usuários sobre a confiabilidade dos sistemas de recomendações?
- QP3: Quais são as melhores práticas para garantir a neutralidade de viés em sistemas de recomendações?
- QP4: De que maneira a falta de transparência nos algoritmos de recomendação pode impactar a experiência do usuário?
- QP5: Quais são os principais desafios quando falamos em fiscalizar a justiça de algoritmos de recomendação?
- QP6: Quais são as implicações sociais e econômicas da falta de neutralidade de viés e transparência em sistemas de recomendações?

Para a etapa seguinte, na qual é feita a identificação dos estudos, foram selecionadas as seguintes fontes de dados: ACM Digital Library, IEEE Digital Library, ScienceDirect, Scopus e SpringerLink. Essas bases foram escolhidas por sua relevância temática em Ciência da Computação, Engenharia e áreas afins, além de oferecerem mecanismos eficientes de filtragem, maximizando a recuperação de estudos relevantes. Além disso, são amplamente reconhecidas e frequentemente utilizadas por pesquisadores com interesse no tema desta revisão.

A estratégia de busca fundamentou-se na combinação de palavras-chave e no emprego de operadores lógicos, adaptados às especificidades de cada base de dados consultada. O operador “OR” foi utilizado para ampliar a busca, abrangendo termos alternativos, enquanto o operador “AND” serviu para refinar os resultados ao estabelecer conexões lógicas entre os termos. A construção da *string* de busca considerou critérios relevantes ao contexto do estudo, sendo a seguinte:

```
( Recommender system OR AI Recommender OR Algorithmic Recommender OR
Automated Recommender OR Content Recommender OR Recommendation
System ) AND ( Bias neutrality OR Fairness OR Impartiality OR
Neutrality OR Non-discriminatory ) AND Data transparency
```

Na terceira etapa, o objetivo é identificar e remover da revisão os trabalhos que não são relevantes para as questões de pesquisa. Assim, para garantir a relevância e a qualidade dos estudos selecionados, foram estabelecidos os seguintes critérios de inclusão e exclusão:

- **Critérios de inclusão:**

- CI-1 - Estudos que abordam a neutralidade de viés em sistemas de recomendação
- CI-2 - Estudos que abordam considerações éticas e impactos sociais em sistemas de recomendação
- CI-3 - Estudos que abordam ética e sistemas de recomendação

**CI-4** - Estudos que avaliam as injustiças que podem ser geradas por sistemas de recomendação

**CI-5** - Estudos que avaliam os problemas que sistemas de recomendação podem gerar

• **Critérios de exclusão:**

**CE-1** - Estudos que não abordam sobre ética em sistemas de recomendação

**CE-2** - O acesso ao artigo completo não está disponível

**CE-3** - Publicações duplicadas em diferentes bases de busca

**CE-4** - Estudos que não abordam sobre injustiça em sistemas de recomendação

**CE-5** - Trabalhos cujo idioma não seja português ou inglês

**CE-6** - Índices, documentos que estão disponíveis na forma de resumos (menos de 4 páginas) ou apresentações (*gray literature*) e estudos secundários

A RE foi conduzida entre 10/04/2024 e 07/11/2024, período em que foram realizadas as etapas de busca, seleção e análise dos artigos, seguindo rigorosamente os critérios previamente estabelecidos. A revisão abrangeu publicações no intervalo de 2004 a 2024. Inicialmente, foram recuperados 1.157 trabalhos, que passaram por um processo estruturado de triagem, incluindo a seleção preliminar (título e resumo) e a avaliação final (texto completo), resultando em 14 estudos considerados relevantes para análise e extração de dados.

Na quarta etapa, os artigos selecionados foram submetidos à extração de dados-chave, com o objetivo de responder às questões de pesquisa e atender aos objetivos propostos pelo estudo. Esses dados subsidiaram a etapa final, dedicada à sumarização e análise dos resultados, cuja apresentação detalhada encontra-se na seção seguinte.

## 4. Discussão e análise dos resultados

Nesta seção, são discutidas as respostas para as Questões de Pesquisa. Os 14 artigos selecionados foram analisados em profundidade, com as informações extraídas de forma padronizada para assegurar consistência e comparabilidade. As subseções apresentam os principais achados relacionados a cada questão, destacando as contribuições dos estudos analisados e suas implicações no contexto do tema investigado.

### QP1: Quais são os principais tipos de viés que podem afetar sistemas de recomendações?

A partir desta revisão de escopo, observou-se que diversos estudos têm investigado o impacto do viés em sistemas de recomendação. Dentre esses estudos, os principais tipos de viés que afetam sistemas de recomendação incluem:

- **viés de popularidade**, em que itens populares são excessivamente recomendados em detrimento de itens menos conhecidos, o que reduz a diversidade das sugestões (STINSON, 2022a; STINSON, 2022b);
- **viés de seleção**, que ocorre ao treinar o sistema apenas com interações registradas, ignorando itens não visualizados e limitando a adaptabilidade do sistema (RICCI; ROKACH; SHAPIRA, 2015; PRIMO *et al.*, 2007);
- **viés de filtro colaborativo**, que reforça padrões demográficos e culturais preexistentes, restringindo a diversidade (SANTINI; SALLES, 2020; SIGLIANO; FAUSTINO, 2016); e,
- **viés de homogeneização**, que diminui a variedade dos itens recomendados, criando bolhas de filtro e limitando a exposição dos usuários a novos conteúdos (RICCI; ROKACH; SHAPIRA, 2015; SANTINI; SALLES, 2020).

Por exemplo, há evidências de que sistemas de recomendação podem apresentar diferenças significativas na utilidade de suas sugestões para diferentes grupos demográficos, como gênero e idade,

afetando a experiência de usuários de maneira desigual (SANTINI; SALLES, 2020). Além disso, estudos destacam como bolhas de filtro e a homogeneização das recomendações podem aprisionar usuários em círculos viciosos de consumo, perpetuando desigualdades e estereótipos (STINSON, 2022a; SIGLIANO; FAUSTINO, 2016).

Diante deste contexto, destaca-se a importância de considerar a ocorrência desses tipos de vies no desenvolvimento de sistemas de recomendação, buscando uma maior promoção de equidade e representatividade das recomendações aos usuários (RICCI; ROKACH; SHAPIRA, 2015).

### **QP2: Como a transparência dos dados pode influenciar a percepção dos usuários sobre a confiabilidade dos sistemas de recomendações?**

A transparência e explicabilidade dos algoritmos de recomendação tem sido um tema recorrente de estudo na literatura. Tintarev e Masthoff (2007) revisaram métodos de explicação em sistemas de recomendação, destacando a importância de fornecer aos usuários justificativas compreensíveis para as recomendações recebidas. Isso não apenas aumenta a confiança dos usuários, mas também pode ajudar a identificar e corrigir vieses inadvertidos.

Mais recentemente, Abdollahi e Nasraoui (2018) investigaram a aplicação de técnicas de explicabilidade para melhorar a transparência dos sistemas de recomendação. O estudo demonstrou que explicações claras e detalhadas podem ajudar a mitigar a percepção de vies e aumentar a aceitação do sistema pelos usuários.

A transparência dos dados em sistemas de recomendação pode influenciar positivamente a percepção de confiabilidade dos usuários, ao permitir que compreendam como suas informações são utilizadas para gerar sugestões personalizadas. Conforme discutido por Ricci, Rokach e Shapira (2015) e Santini e Salles (2020), usuários tendem a confiar mais em recomendações quando conseguem visualizar critérios e dados que fundamentam essas indicações. Isso ocorre porque explicações claras sobre o processo de recomendação reduzem a opacidade algorítmica, permitindo maior controle e compreensão por parte dos usuários. No entanto, a literatura também destaca que a transparência pode ser limitada por dificuldades técnicas, como o alto custo computacional para fornecer explicações detalhadas e a complexidade em balancear a transparência com a proteção de privacidade (PRIMO *et al.*, 2007; SIGLIANO; FAUSTINO, 2016). Assim, a transparência dos dados atua não apenas como um mecanismo para aumentar a confiança, mas também como um elemento crucial na responsabilização dos sistemas de recomendação.

### **QP3: Quais são as melhores práticas para garantir a neutralidade de vies em sistemas de recomendações?**

A neutralidade de vies em sistemas de recomendação refere-se à capacidade dos algoritmos de oferecer recomendações sem favorecer ou prejudicar grupos de usuários. A falta de neutralidade pode levar à discriminação e amplificação de preconceitos sociais, limitando a pluralidade de perspectivas (STINSON, 2022a; STINSON, 2022b).

Pesquisas, como a de Deldjoo, Beigi e Berti-Equille (2020), focam em técnicas para mitigar vieses em sistemas de recomendação. Os autores propuseram um *framework* que integra técnicas de pré-processamento dos dados, ajustes nos algoritmos de recomendação e pós-processamento dos resultados para reduzir vieses e aumentar a equidade. Este trabalho é fundamental para entender as diferentes camadas onde o vies pode ser introduzido e como pode ser tratado em cada etapa do *pipeline* de recomendação.

Estudos de caso específicos, como o de Beutel, Chen e Covington (2019), que analisaram o sistema de recomendações do YouTube, ilustram os desafios práticos de implementar neutralidade de vies e transparência. O estudo mostrou como a plataforma lida com vieses algorítmicos e os esforços para

tornar suas recomendações mais justas e transparentes, incluindo ajustes nos algoritmos e maior escrutínio sobre os dados de treinamento utilizados.

As melhores práticas para garantir a neutralidade de viés em sistemas de recomendação incluem:

- minimizar viés nos dados de treinamento, utilizando amostragens balanceadas e correções algorítmicas (RICCI; ROKACH; SHAPIRA, 2015; PRIMO *et al.*, 2007);
- aumentar a transparência e explicabilidade, permitindo aos usuários compreenderem como são feitas as recomendações (SIGILIANO; FAUSTINO, 2016; SANTINI; SALLES, 2020);
- aplicar regularizações para penalizar o uso de informações demográficas, promovendo a neutralidade (STINSON, 2022a); e,
- diversificar ativamente as recomendações, evitando o foco excessivo em itens populares e ampliando a exposição do usuário a novos conteúdos (RICCI; ROKACH; SHAPIRA, 2015).

#### **QP4: De que maneira a falta de transparência nos algoritmos de recomendação pode impactar a experiência do usuário?**

A falta de transparência nos algoritmos de recomendação impacta negativamente a experiência do usuário, reduzindo sua confiança e compreensão das sugestões apresentadas. Isso pode levar a uma sensação de manipulação e formar “bolhas de filtro” que limitam a diversidade de informações, afetando a percepção de controle e satisfação do usuário (RICCI; ROKACH; SHAPIRA, 2015; PRIMO *et al.*, 2007; SIGILIANO; FAUSTINO, 2016; SANTINI; SALLES, 2020). Além disso, a opacidade dos algoritmos pode gerar desconfiança, como destacado por Santini e Salles (SANTINI; SALLES, 2020), já que os usuários têm dificuldade em entender como e por que determinadas recomendações são feitas. Isso pode resultar em uma percepção negativa sobre a equidade do sistema e limitar o engajamento do usuário. Sigiliano e Faustino (SIGILIANO; FAUSTINO, 2016) também mencionam que a falta de transparência em sistemas como o Netflix afeta a experiência do usuário, levando a uma percepção de falta de controle sobre as recomendações recebidas, o que contribui para uma menor satisfação com o serviço.

#### **QP5: Quais são os principais desafios quando falamos em fiscalizar a justiça de algoritmos de recomendação?**

A fiscalização da justiça em algoritmos de recomendação enfrenta desafios significativos, como a dificuldade de garantir a transparência e interpretabilidade dos modelos, essencial para detectar vieses prejudiciais (MAGRANI; SILVA, 2024). Além disso, a complexidade dos algoritmos e a opacidade algorítmica dificultam a avaliação de vieses implícitos, criando barreiras para assegurar decisões justas e responsáveis (RICCI; ROKACH; SHAPIRA, 2015; PRIMO *et al.*, 2007; SIGILIANO; FAUSTINO, 2016). Outro desafio inclui a necessidade de padronização e regulamentação, uma vez que políticas uniformes são raras, aumentando o risco de injustiças e limitações na capacidade de monitorar o impacto desses sistemas em diferentes contextos (SANTINI; SALLES, 2020).

#### **QP6: Quais são as implicações sociais e econômicas da falta de neutralidade de viés e transparência em sistemas de recomendações?**

A falta de neutralidade de viés e transparência em sistemas de recomendação pode ter implicações sociais e econômicas profundas, como o reforço de desigualdades e a perpetuação de estereótipos, impactando negativamente a autonomia e a privacidade dos usuários (MAGRANI; SILVA, 2024). No âmbito econômico, a opacidade desses sistemas pode favorecer práticas comerciais injustas e limitar a concorrência ao criar “bolhas de filtro” que restringem o acesso dos usuários a opções diversificadas, perpetuando vieses preexistentes (RICCI; ROKACH; SHAPIRA, 2015; SANTINI; SALLES, 2020; SIGILIANO; FAUSTINO, 2016). Além disso, a manipulação de dados por parte das plataformas pode reduzir a confiança dos consumidores e afetar o funcionamento saudável dos mercados ao direcionar

escolhas e comportamentos de forma não transparente (STINSON, 2022a).

## 5. Considerações finais

Neutralidade de viés e transparência dos dados em sistemas de recomendações são temas que têm sido amplamente discutidos na comunidade científica, destacando-se como tópicos cruciais no desenvolvimento de sistemas justos e éticos (RICCI; ROKACH; SHAPIRA, 2015; SANTINI; SALLES, 2020). Garantir que os sistemas de recomendação sejam transparentes e livres de vieses é essencial para promover a confiança dos usuários e para assegurar que as recomendações oferecidas sejam representativas e imparciais, minimizando impactos negativos sobre grupos sub-representados.

Este trabalho apresentou os resultados de uma revisão escopo que analisou as abordagens existentes para lidar com a neutralidade de viés e a transparência nos sistemas de recomendação. A principal tendência observada foi a crescente preocupação com o desenvolvimento de métodos de mitigação de viés e técnicas que aumentem a explicabilidade dos algoritmos, permitindo que os usuários compreendam melhor as recomendações recebidas.

A revisão de escopo apresentada neste trabalho representa uma contribuição teórica significativa, a qual pode ser utilizada como base para projetos mais abrangentes e aprofundados na área. Por exemplo, os resultados obtidos podem ser aplicados no desenvolvimento de novos algoritmos de recomendação que priorizem a neutralidade e a explicabilidade, bem como no aprimoramento das políticas de transparência adotadas por plataformas que utilizam sistemas de recomendação.

Outro desafio significativo identificado após este estudo é a criação de leis e regulamentações sobre sistemas de recomendação, visto que eles estão em constante desenvolvimento e atualização, o que pode fazer com que regras fiquem defasadas antes mesmo de serem estabelecidas. Isso exige que as regulamentações sejam flexíveis e adaptáveis, permitindo ajustes rápidos conforme os sistemas evoluem, para garantir que permaneçam eficazes na proteção dos direitos dos usuários e na promoção de um ambiente justo.

Para trabalhos futuros, sugere-se a avaliação empírica de técnicas de mitigação de viés, o desenvolvimento de métricas de transparência para comparação de sistemas e a análise de casos reais em que recomendações transparentes foram implementadas, investigando seu impacto sobre o engajamento e a satisfação dos usuários. Além disso, recomenda-se a exploração da relação entre transparência e desempenho dos sistemas, buscando um equilíbrio entre a clareza dos processos e a eficiência das recomendações.

## Referências

ABDOLLAHI, B.; NASRAOUI, O. Transparency in fair machine learning: The case of explainable recommender systems. **Proceedings of the 2018 Conference on Fairness, Accountability and Transparency**, p. 237–248, 2018.

ANGWIN, J.; LARSON, J.; MATTU, S.; KIRCHNER, L. Machine bias. 2016. Disponível em: <<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>>. Acesso em: 3 jun. 2024.

ANYOHA, R. The history of artificial intelligence. **Science in the News**, Harvard University, 28 ago. 2017. Disponível em: <<https://sitn.hms.harvard.edu/flash/2017/history-artificial-intelligence/>>. Acesso em: 3 jun. 2024.

BAIA, C. **Inteligência Artificial no Recrutamento e Seleção**. 2024. Disponível em: <<https://www.gupy.io/blog/inteligencia-artificial-no-recrutamento-e-selecao#:~:text=A%20Intelig%>>

C3%Aancia%20Artificial%20pode%20ser,e%20decidindo%20de%20maneira%20ordenada>. Acesso em: 3 jun. 2024.

BEUTEL, A.; CHEN, J.; COVINGTON, P. Fairness in recommendation ranking through pairwise comparisons. **Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining**, p. 2212–2220, 2019.

CAMARGO, O. **Injustiça Social**. 2023. Disponível em: <<https://brasilecola.uol.com.br/sociologia/fome-miseria-altos-impostos.htm>>. Acesso em: 3 jun. 2024.

COPELAND, B. **Artificial Intelligence - Alan Turing, AI Beginnings**. Britannica, 2023. Disponível em: <<https://www.britannica.com/technology/artificial-intelligence>>. Acesso em: 3 jun. 2024.

CRISTINA, C. **Como o uso da IA para selecionar currículos pode ampliar a desigualdade**. Tilt Uol, 2021. Disponível em: <<https://www.uol.com.br/tilt/noticias/redacao/2021/06/14/como-o-uso-da-ia-para-selecionar-curriculos-pode-ampliar-a-desigualdade.htm>>. Acesso em: 3 jun. 2024.

DELDJOO, Y.; BEIGI, G.; BERTI-EQUILLE, L. A survey on fairness-aware recommender systems. **Proceedings of the 2020 Conference on Fairness, Accountability and Transparency**, p. 239–249, 2020.

GIL, A. C. **Como elaborar projetos de pesquisa**. São Paulo: Atlas, 2022. ISBN 9786559771653.

LEDFORD, H. Millions of black people affected by racial bias in health-care algorithms. **Nature**, v. 574, n. 7780, p. 608–609, out. 2019. ISSN 0028-0836, 1476-4687. Disponível em: <<https://www.nature.com/articles/d41586-019-03228-6>>.

MAGRANI, E.; SILVA, P. G. F. d. The ethical and legal challenges of recommender systems driven by artificial intelligence. **AI & Society**, v. 35, p. 957–967, 2024.

MCCARTHY, J. What is artificial intelligence. Stanford University, 2007.

MORAIS, M. L. e M. **Viés da tecnologia: o que acontece quando o algoritmo é racista**. 2020. Disponível em: <<https://portal.unicap.br/-/vi%C3%A9s-da-tecnologia-o-que-acontece-quando-o-algoritmo-%C3%A9-racista>>. Acesso em: 3 jun. 2024.

MUNN, Z.; PETERS, M. D. J.; STERN, C.; TUFANARU, C.; MCARTHUR, A.; AROMATARIS, E. Systematic review or scoping review? guidance for authors when choosing between a systematic or scoping review approach. **BMC Medical Research Methodology**, v. 18, p. 143, Nov 2018. ISSN 1471-2288. Disponível em: <<https://doi.org/10.1186/s12874-018-0611-x>>.

PASCUAL, M. **Quem vigia os algoritmos para que não sejam racistas ou sexistas?** 2019. Disponível em: <[https://brasil.elpais.com/brasil/2019/03/18/tecnologia/1552863873\\_720561.html](https://brasil.elpais.com/brasil/2019/03/18/tecnologia/1552863873_720561.html)>. Acesso em: 3 jun. 2024.

PETERS, M.; GODFREY, C.; MCINERNEY, P.; MUNN, Z.; TRICCO, A.; KHALIL, H. Chapter 11: Scoping reviews. In: \_\_\_\_\_. **JBI Reviewer's Manual**. JBI, 2020. Disponível em: <<http://dx.doi.org/10.46658/JBIRM-20-01>>.

PRIMO, T. T.; VICARI, R. M.; SILVA, J. M. C. D.; SILVA, J. M. C. D. Rumo ao uso de metadados educacionais em sistemas de recomendação. **portalobaa.org**, v. 1, 2007. ISSN 2316-6533.

RICCI, F.; ROKACH, L.; SHAPIRA, B. Recommender systems: Introduction and challenges. In: \_\_\_\_\_. **Recommender Systems Handbook**. Boston, MA: Springer US, 2015. p. 1–34. ISBN 978-1-4899-7637-6. Disponível em: <[https://doi.org/10.1007/978-1-4899-7637-6\\_1](https://doi.org/10.1007/978-1-4899-7637-6_1)>.

SANTINI, R. M.; SALLES, D. O impacto dos algoritmos no consumo de música. **Signos do Consumo**, v. 12, 2020.

SIGILIANO, D.; FAUSTINO, E. Netflix: Sistemas de recomendação inteligentes. **Revista Tecer**, v. 9, 2016. Disponível em: <[https://www.academia.edu/26751836/NETFLIX\\_Sistemas\\_de\\_Recomendac\\_a\\_o\\_Inteligentes](https://www.academia.edu/26751836/NETFLIX_Sistemas_de_Recomendac_a_o_Inteligentes)>.

STINSON, C. Algorithms are not neutral. **AI and Ethics**, v. 2, n. 4, p. 763–770, nov. 2022. ISSN 2730-5961. Disponível em: <<https://doi.org/10.1007/s43681-022-00136-w>>.

STINSON, L. Bias mitigation in recommender systems to improve diversity. **Journal of Artificial Intelligence**, v. 34, 2022. Disponível em: <<https://ceur-ws.org/Vol-3318/short6.pdf>>.

TINTAREV, N.; MASTHOFF, J. A survey of explanations in recommender systems. **Proceedings of the 2007 IEEE 23rd International Conference on Data Engineering Workshop**, p. 801–810, 2007.

ZENDESK. **Qual é a origem da inteligência artificial? Onde tudo começou?** 2024. Disponível em: <<https://www.zendesk.com.br/blog/qual-e-a-origem-da-inteligencia-artificial/>>. Acesso em: 3 jun. 2024.