MESTRADO ACADÊMICO
EM COMUNICAÇÃO

ASSOCIAÇÕES AUTOMÁTICAS ENTRE INTELIGÊNCIA ARTIFICIAL E CONFIANÇA: ANÁLISE DAS ATITUDES IMPLÍCITAS DOS INTERNAUTAS FRENTE A IMAGENS PRODUZIDAS POR IA

DIEGO CAMPOS SALGADO BRAGA

Brasília-DF, 2025

DIEGO CAMPOS SALGADO BRAGA

ASSOCIAÇÕES AUTOMÁTICAS ENTRE INTELIGÊNCIA ARTIFICIAL E CONFIANÇA: ANÁLISE DAS ATITUDES IMPLÍCITAS DOS INTERNAUTAS FRENTE A IMAGENS PRODUZIDAS POR IA

Dissertação apresentada ao Programa de Pós Graduação em Comunicação, do Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa, como requisito parcial para obtenção do grau de Mestre.

Orientadora
Professora Doutora Ébida Rosa dos Santos

Brasília-DF 2025

DIEGO CAMPOS SALGADO BRAGA

ASSOCIAÇÕES AUTOMÁTICAS ENTRE INTELIGÊNCIA ARTIFICIAL E CONFIANÇA: ANÁLISE DAS ATITUDES IMPLÍCITAS DOS INTERNAUTAS FRENTE A IMAGENS PRODUZIDAS POR IA

Dissertação apresentada ao Programa de Pós Graduação em Comunicação, do Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa, como requisito parcial para obtenção do grau de Mestre.

Aprovado em 04 / 12 / 2025

Banca Examinadora

Profa. Dra. Ébida Rosa dos Santos- Orientadora

Profa. Dra. Vanessa Clarizia Marchesin

Prof. Dr. Bruno Saboya

Código de catalogação na publicação – CIP

B813a Braga, Diego Campos Salgado
Associações automáticas entre inteligência artificial e confiança: análise das atitutes implícitas dos internautas frente a imagens produzidas por IA / Diego Campos Salgado Braga. — Brasília: Instituto Brasileiro Ensino, Desenvolvimento e Pesquisa, 2026.
151 f. : il.; color.

Orientadora: Profa. Dra. Ébida Rosa dos Santos

Dissertação (Mestrado Acadêmico em Comunicação) — Instituto Brasileiro Ensino, Desenvolvimento e Pesquisa – IDP, 2025.

1. Inteligência artificial. 2. Mídia digital. 3. Comunicação digital. 4. Confiabilidade (computador). I.Título

CDD 006.3

Elaborada pela Biblioteca Ministro Moreira Alves

DEDICATÓRIA

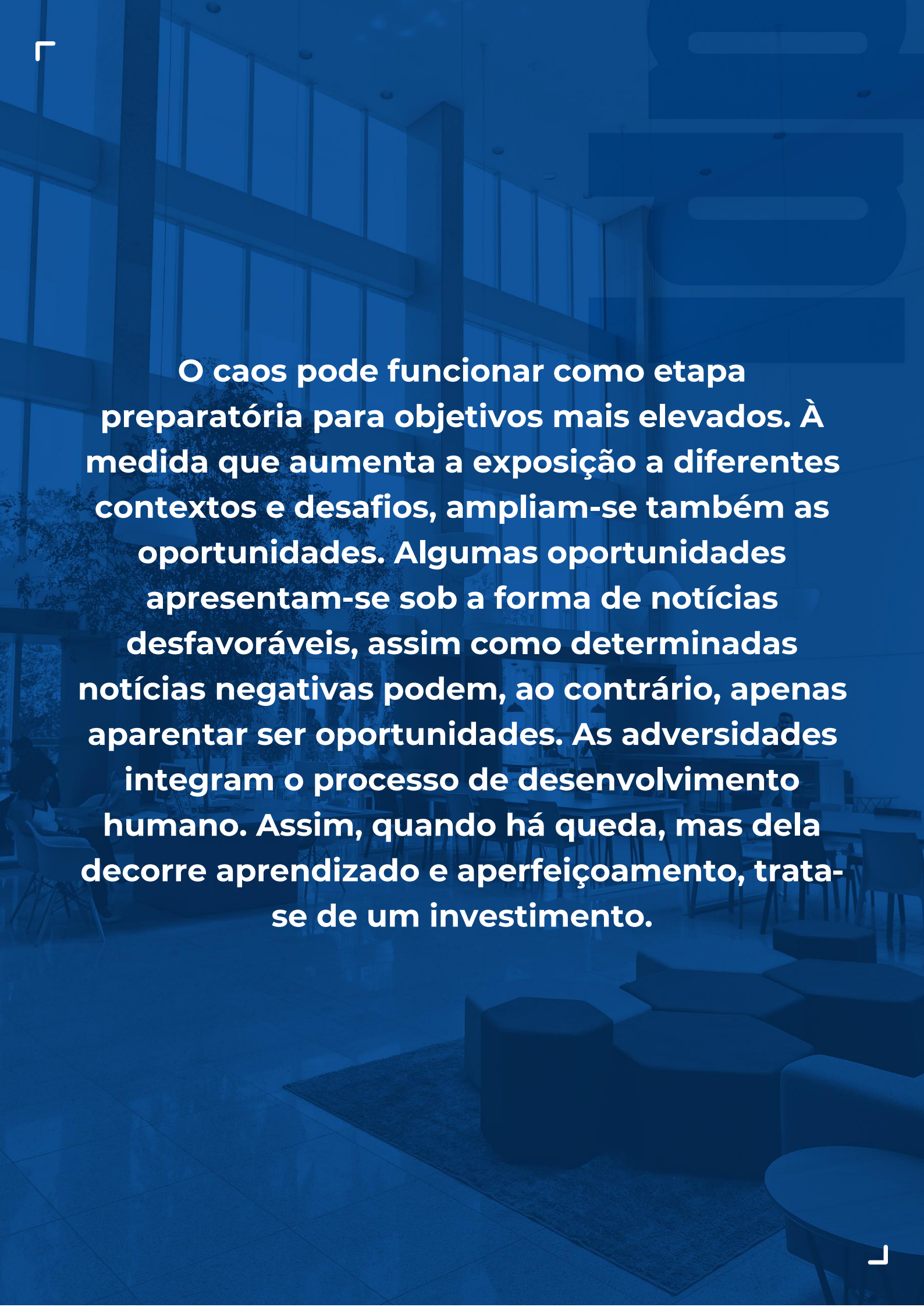
Agradeço a Deus, fonte de força e direção nos momentos em que os limites humanos se impuseram.

Uma alegria dividida com a pessoa amada são duas; uma tristeza dividida é só a metade.

Por isso, dedico este trabalho à minha esposa Maria Rita, companheira de todas as jornadas, cuja presença tornou cada conquista mais significativa.

AGRADECIMENTOS

A todos que, de alguma forma, contribuíram para este percurso – em especial à minha orientadora Prof.ª Dr.ª Ébida Rosa dos Santos, pela sabedoria, paciência e inspiração constante – o meu sincero agradecimento.



O caos pode funcionar como etapa preparatória para objetivos mais elevados. À medida que aumenta a exposição a diferentes contextos e desafios, ampliam-se também as oportunidades. Algumas oportunidades apresentam-se sob a forma de notícias desfavoráveis, assim como determinadas notícias negativas podem, ao contrário, apenas aparentar ser oportunidades. As adversidades integram o processo de desenvolvimento humano. Assim, quando há queda, mas dela decorre aprendizado e aperfeiçoamento, trata-se de um investimento.

RESUMO

A presente pesquisa investigou como internautas percebem, consciente e inconscientemente, a confiabilidade de imagens produzidas por Inteligência Artificial Generativa (IAG), em comparação com imagens reais, especialmente no contexto do sistema de justiça. A pesquisa foi conduzida com abordagem quantitativa, utilizando o Teste de Associação Implícita (IAT) e o Teste de Atribuição Explícita de Atributos (EAAT), os quais permitiram avaliar atributos como competência, integridade e benevolência. A dissertação se apoia em uma sólida fundamentação teórica que abarca as contribuições de Mayer, Davis e Schoorman (1995), Greenwald et al. (1998) e Zaltman (2003), articulando os conceitos de confiança, atitudes implícitas e credibilidade em ambientes digitais mediados por Inteligência Artificial (IA). Os resultados da pesquisa revelam discrepâncias entre as percepções conscientes e inconscientes dos participantes, evidenciando que conteúdos gerados por IA podem ativar vieses implícitos que influenciam o julgamento de confiabilidade, mesmo quando conscientemente identificados como artificiais. A dissertação contribui para o debate acadêmico e ético sobre o uso de tecnologias de IA na comunicação institucional, jurídica e comercial, apontando a importância da transparência algorítmica e da literacia digital como fatores determinantes para o fortalecimento da confiança em ambientes digitais.

Palavras chave: Inteligência Artificial; Confiança; Teste de Associação Implícita; Comunicação Digital; Credibilidade *on-line*.

ABSTRACT

This research investigates how internet users perceive the trustworthiness of images generated by Generative Artificial Intelligence (GAI), in comparison to real images, especially within the context of the justice system. The study employed a quantitative methodology, utilizing the Implicit Association Test (IAT) and the Explicit Attribute Assignment Test (EAAT) to assess key trust attributes such as competence, integrity, and benevolence. The theoretical framework draws on the works of Mayer, Davis, and Schoorman (1995), Greenwald et al. (1998), and Zaltman (2003), integrating concepts of trust, implicit attitudes, and digital credibility mediated by Artificial Intelligence (AI) technologies. The results indicate a significant dissociation between conscious and unconscious perceptions, suggesting that AI-generated content may activate implicit biases that influence users' judgments of trustworthiness—even when the content is explicitly identified as artificial. The dissertation provides valuable insights for the academic and ethical discourse surrounding the adoption of AI in institutional, legal, and marketing communication, highlighting the importance of algorithmic transparency and digital literacy to foster trust in digital environments.

Keywords: Artificial Intelligence; Trust; Implicit Association Test; Digital Communication; Online Credibility.

LISTA DE ABREVIATURAS E SIGLAS

AAs	Artificial Agents
AGI	Artificial General Intelligence
APE	Avaliação Associativa-Proposicional
CGI.br	Comitê Gestor da Internet no Brasil
EAAT	Teste de Atribuição Explícita de Atributos (Explicit Attribute Assignment Test)
EC	Experiência do usuário
fMRI	Ressonância magnética funcional
IA	Inteligência Artificial
IAG	Inteligência Artificial Generativa
IAT	Teste de Associação Implícita (Implicit Association Test)
IBGE	Instituto Brasileiro de Geografia e Estatística
IDP	Instituto Brasileiro de Ensino, Desenvolvimento e Pesquisa
PKD	Philip K. Dick
QI	Quociente de Inteligência
ROI	Return of Investment
SaaS	Software as a Service
SDSCA	Summary of Diabetes Self-Care Activities Measure
SSL	Secure Sockets Layer
SVMs	Máquinas de vetor de suporte
TCLE	Termo de Consentimento Livre e Esclarecido
TR	Tempos de reação
UTAUT	Unified Theory of Acceptance and Use of Technology
XAI	Inteligência Artificial Explicável (eXplainable Artificial Intelligence)

LISTA DE ILUSTRAÇÕES

LISTA DE QUADROS

LISTA DE TABELAS

SUMÁRIO


1

1 INTRODUÇÃO

O avanço crescente da tecnologia tem contribuído para o desenvolvimento de diversos setores e segmentos da sociedade, com o objetivo de encurtar distâncias, aumentar a acessibilidade, permitir alcance global e oferecer respostas cada vez mais ágeis frente às constantes mudanças nos negócios e nas preferências dos usuários. Nos últimos anos, a *internet* se consolidou como a principal fonte de informação para grande parte da população mundial. De acordo com um estudo do *Pew Research Center* (2021), cerca de 93% dos adultos nos Estados Unidos utilizam a rede mundial de computadores regularmente, e, entre os jovens de 18 a 29 anos, esse número chega a 99%. O Comitê Gestor da Internet no Brasil (CGI.br), por meio de sua "Pesquisa sobre o uso das tecnologias de informação e comunicação nos domicílios brasileiros: TIC Domicílios 2023", aponta que 84% da população tem acesso à *internet*, sendo que os jovens são os usuários mais ativos (CGI.br, 2023).

Nesse contexto, as estratégias do *marketing* digital, que buscam atingir e engajar usuários no ambiente *on-line*, atuam como um processo de otimização dos canais de comunicação nos quais a informação é consumida e compartilhada, quer por rede sociais, *e-mails*, *websites* ou aplicativos móveis. Isso tem transformado a maneira como as pessoas se comunicam, estimulando a interação com novos mercados que oferecem uma vasta diversidade de produtos e serviços. Assim, os usuários da *internet* podem tomar decisões mais conscientes e embasadas, de acordo com as suas necessidades, sejam elas de entretenimento, compra, aprendizado, interação com instituições oficiais, entre outras.

A inteligência artificial (IA) é um ponto de inflexão nesse mundo digital. Para Russell e Norvig (2016), trata-se de uma área da ciência da computação que visa a criar sistemas capazes de realizar tarefas que normalmente requerem inteligência humana, como reconhecimento de fala, tomada de decisão e tradução de idiomas. Os referidos autores definem a IA como sistemas que simulam habilidades humanas, incluindo resolução de problemas e aprendizado. Apesar das expectativas iniciais de que a IA poderia superar a inteligência humana em todos os domínios, seu desenvolvimento atual é classificado

principalmente como “inteligência estreita”, segundo a qual a IA é excelente em tarefas específicas, mas carece de generalização e intuição humanas.

Nesse cenário, em um ambiente cada vez mais concorrente e veloz, a IA apresenta-se como um conjunto de ferramentas com potenciais competitivos que agregam novas dinâmicas à construção de informações e podem ser amplamente utilizadas nas estratégias de interação na *internet* em geral. Sua participação tornou-se uma realidade indelével na forma como os conteúdos são criados e consumidos na *web*, remodelando suas estratégias em função dessa personalização baseada no comportamento do usuário. Isso aumenta a relevância das mensagens e intensifica as interações com os internautas, enriquecendo suas experiências e tornando as avaliações de negócios pelas empresas mais vanguardistas. Segundo Madeira *et al.* (2020, p. 98),

> Graças ao alcance das mídias sociais e ‘massas’ de dados deixados para trás, conscientemente e inconscientemente, durante navegações na Internet, a IA possui um grande potencial na área do marketing digital. Usar a inteligência artificial para uma melhor experiência do usuário-consumidor, análises preditivas e marketing direcionado, certamente irá fornecer um grande ROI (Return of Investment) para os negócios.

Como resultado, observa-se uma melhoria na eficiência das taxas de engajamento e conversão no mercado digital, haja vista que os sistemas alimentados por inteligência artificial generativa (IAG) e *machine learning* são aplicações poderosas que “[...] aprendem enquanto existem, aperfeiçoando suas respostas a partir da experiência acumulada em interações anteriores com usuários, fazendo uso do aprendizado de máquina para a detecção de padrões e elaboração de novas respostas” (Madeira *et al.*, 2020, p. 95).

No escopo desta dissertação, é importante distinguir entre os conceitos de Inteligência Artificial (IA) e Inteligência Artificial Generativa (IAG). A IA é um campo amplo da ciência da computação que busca desenvolver sistemas capazes de simular capacidades humanas, como reconhecimento de padrões, tomada de decisão e aprendizado adaptativo (Russell; Norvig, 2016). Já a IAG representa um subcampo específico da IA, responsável pela criação autônoma de novos conteúdos – como textos, imagens, sons e vídeos – a partir de dados previamente treinados. Essa capacidade criativa da IAG a diferencia dos

sistemas tradicionais de IA, que apenas classificam ou predizem com base em entradas existentes. De acordo com Bommasani *et al*. (2021), os chamados "modelos fundacionais", como *DALL·E* e *ChatGPT*, inauguram uma nova era na qual sistemas computacionais participam da produção simbólica, desafiando fronteiras entre o real e o artificial. Assim, sempre que este trabalho se referir à criação de imagens sintéticas ou conteúdo visual automatizado, o termo IAG será utilizado, reservando-se o uso de IA para referências ao campo mais geral da inteligência computacional.

A IA e a Neurociência se interrelacionam, pois, no campo de estudo, analisam como os processos cerebrais humanos resultam em padrões comportamentais. Esse ponto de conexão pode ser avaliado por meio da análise da reação emocional e cognitiva dos internautas, que têm seu comportamento e sua tomada de decisão influenciados pela nova tecnologia. O que se procurou investigar com a presente pesquisa é que, muitas vezes, essas reações são demonstradas implicitamente. Conforme apontam Goodfellow, Bengio e Courville (2016), os primeiros algoritmos de aprendizado profundo reconhecidos hodiernamente foram pensados como modelos computacionais de aprendizado biológico, ou seja, modelos de como o aprendizado acontece ou pode acontecer no cérebro. O aprendizado profundo está intimamente associado à arquitetura das redes neurais artificiais. Kriegeskorte e Douglas (2018, p. 1148) observam esse mesmo raciocínio, afirmando que "Modelos computacionais que imitam o processamento de informações cerebrais durante tarefas perceptivas, cognitivas e de controle estão começando a ser desenvolvidos e testados com dados cerebrais e comportamentais"[1]. Ainda nessa perspectiva, os autores acrescentam que a "IA, e em particular o aprendizado de máquina, é, portanto, uma disciplina fundamental que fornece a base teórica e tecnológica para a neurociência computacional cognitiva"[2] (Kriegeskorte; Douglas, 2018, p. 1151).

Outro ponto essencial para esta dissertação e precisa ser introduzido ao leitor é o que destaca Gerald Zaltman (2003) ao afirmar que 95% dos pensamentos e decisões dos usuários ocorrem em nível inconsciente, por conseguinte, no contexto da IAG; isso significa que as

[1] Tradução livre do original: "Computational models that mimic brain information processing during perceptual, cognitive and control tasks are beginning to be developed and tested with brain and behavioral data".
[2] Tradução livre do original: "AI, and in particular machine learning, is therefore a key discipline that provides the theoretical and technological foundation for cognitive computational neuroscience".

percepções sobre credibilidade e autenticidade de conteúdos desenvolvidos artificialmente podem ser influenciadas por associações implícitas que os usuários não conseguem expressar conscientemente. No mesmo sentido, Lemmers-Jansen *et al.* (2023, p. 1) salientam a relação entre neurociência e IAG, especialmente no que diz respeito ao processamento inconsciente das informações pelos internautas e sua influência na tomada de decisão, entendendo que "A pesquisa de neuroimagem pode fornecer mais informações sobre os mecanismos subjacentes às dificuldades sociais (cognitivas)"[3]. Os autores ponderam, ainda, que

> Nas últimas duas décadas, deficiências na cognição social (CS), referindo-se aos processos psicológicos que permitem às pessoas compreender o comportamento social dos outros, surgiram como alguns dos possíveis fatores que podem estar na base das dificuldades no funcionamento social (Lemmers-Jansen *et al.*, 2023, p. 1)[4].

Dados do Instituto Brasileiro de Geografia e Estatística (IBGE) evidenciam que, no Brasil, os efeitos da IAG no processo de geração de conteúdo e imagem têm se mostrado significativo, conforme apresentado em um estudo recente intitulado "O impacto transformador da inteligência artificial na geração de conteúdo e imagem: uma jornada evolutiva" (IBGE, 2025). O referido Instituto destacou a transformação evolutiva da IAG em diversos setores, incluindo sua aplicação em redes sociais e plataformas digitais, remodelando, profundamente, as interações humanas e a dinâmica de engajamento entre usuários, marcas, instituições públicas e privadas (IBGE, 2025). Esse cenário reforça a necessidade de explorar as implicações emocionais e cognitivas do uso de IAG, especialmente no que diz respeito à confiança em ambientes digitais, tema central deste trabalho.

Em um cenário de crescente exposição digital, a forma como as pessoas constroem julgamentos sociais, como confiança e credibilidade, passa cada vez mais por representações visuais mediadas por algoritmos. O Brasil, em particular, figura entre os países que mais consomem conteúdo audiovisual no mundo, segundo o Comitê Gestor

[3] Tradução livre do original: "Neuroimaging research can give further insights into the underlying mechanisms of social (cognitive) difficulties".
[4] Tradução livre do original: "In the past two decades, impairments in social cognition (SC), referring to the psychological processes that enable people to understand other's social behaviour, have emerged as some of the possible factors that may underlie difficulties in social functioning".

da Internet no Brasil (CGI.br, 2024), com mais de 94% dos usuários de *internet* assistindo a vídeos regularmente e um tempo médio diário de mais de três horas em redes sociais, majoritariamente consumindo imagens e vídeos. Esse comportamento reforça a escolha metodológica desta pesquisa por estímulos visuais e justifica a análise da percepção de imagens como ferramenta eficaz para mensuração de julgamentos implícitos em ambientes digitais mediados por IAG.

Justifica-se a relevância do tema pelo contexto atual de transformação digital e pela crescente utilização da IAG nas estratégias de crescimento das redes sociais, *marketing* e instituições públicas, bem como em quase todos os setores produtivos que utilizam o poder computacional na cadeia produtiva. De certo, a IAG tornou-se uma ferramenta essencial para as organizações, que buscam melhorar suas estratégias produtivas e a fidelização de clientes no ambiente digital. Essa realidade é corroborada por pesquisa do IBGE (2025), que revelou que 84,9% das indústrias de médio e grande porte no Brasil já utilizam tecnologias digitais avançadas, destacando a importância dessas ferramentas para a competitividade e eficiência no cenário atual.

A comunicação digital evoluiu significativamente nas últimas décadas, transformando a maneira como as informações são criadas, distribuídas e consumidas, razão pela qual há um impulsionamento do uso de tecnologias como algoritmos avançados e IAG, os quais permitem uma personalização mais precisa e relevante das interações digitais. Desde os primórdios da *internet*, a busca por eficiência e engajamento levou ao desenvolvimento de sistemas inteligentes capazes de interpretar dados e moldar conteúdos de acordo com o perfil do usuário (Lanier, 2018). O papel dos algoritmos na comunicação digital é essencial e disruptivo, uma vez que eles são responsáveis por organizar, ranquear e recomendar conteúdos, moldando a experiência do usuário em plataformas digitais como redes sociais e serviços de *streaming* (Gillespie, 2018). A IAG, por sua vez, tem ampliado essa capacidade ao permitir a criação automatizada de conteúdos, desde textos e imagens até vídeos personalizados, que são frequentemente indistinguíveis dos produzidos por humanos (Wirth, 2018). Essa tendência está em linha com a ideia original proposta por Turing (1950), ao sugerir que a inteligência poderia ser avaliada pela capacidade de uma máquina imitar o comportamento humano de modo indistinguível.

Essa integração entre IAG e comunicação digital aumenta a eficiência e levanta desafios relacionados à confiança dos usuários, conforme pesquisas de Paletta e Costa do Lago (2022), segundo os quais o uso de algoritmos e sistemas automatizados requer maior transparência e ética, haja vista que as interações mediadas por IA trazem consequências diretamente nas percepções e comportamentos dos indivíduos.

Em relação especificamente ao programa de mestrado pioneiro no Brasil do Instituto Brasileiro de Ensino Desenvolvimento e Pesquisa (IDP) na área de Comunicação, observa-se que o tema da presente dissertação é extremamente relevante, uma vez que compila o estudo da comunicação digital e a tecnologia de IA, entendendo que ambos têm movimentado a economia global e a comunicação *lato sensu*, bem como as mídias sociais.

O avanço da IAG tem modificado significativamente a produção de conteúdo digital, incluindo a criação de imagens hiper-realistas que podem ser utilizadas em diversas áreas, como *marketing*, redes sociais e até mesmo no sistema de justiça. Nesse último caso, pode-se imaginar, por exemplo, a utilização por profissionais do Direito de uma reprodução simulada de um crime, utilizando IAG para criar as cenas. No entanto, a capacidade das IAG de criar imagens convincentes levanta questões sobre a percepção de confiança dos usuários em relação a esses conteúdos.

O objetivo geral deste trabalho visou a investigar como a percepção de confiança em profissionais do sistema de justiça representados por imagens geradas por IAG se diferencia da percepção de confiança em imagens reais, considerando os atributos de competência, integridade e benevolência[5].

Para responder ao objetivo geral, foram estabelecidos como objetivos específicos:

[5] Para os fins desta pesquisa, o termo “profissionais do sistema de justiça” abrange os principais atores que compõem o funcionamento cotidiano da justiça no Brasil, incluindo magistrados, membros do Ministério Público, advogados e agentes da segurança pública. A inclusão dessas categorias se justifica pela sua relevância institucional e simbólica, bem como pela frequência com que são representadas em imagens públicas associadas à legalidade, à autoridade e à confiança. Além disso, a expressiva presença de advogados na estrutura jurídica nacional amplia a aderência empírica da pesquisa, refletindo a diversidade de papéis no ecossistema jurídico brasileiro.

1. **examinar as associações implícitas entre imagens geradas por IAG e atributos de confiança (competência, integridade e benevolência) por meio do Teste de Associação Implícita (IAT);**
2. **avaliar a percepção explícita dos usuários em relação à confiabilidade das imagens utilizando o Teste de Atribuição Explícita de Atributos (EAAT);**
3. **comparar as medidas implícitas e explícitas de confiança, identificando possíveis dissociações entre percepção consciente e inconsciente; e**
4. **analisar as repercussões da percepção de confiança em IA no sistema de justiça, explorando possíveis implicações para o uso de imagens artificiais em tribunais, *marketing* jurídico e comunicações institucionais.**

O Quadro 1, em prosseguimento, sintetiza os objetivos que nortearam este trabalho, o geral e os específicos, além de indicar os instrumentos utilizados para geração de dados e os seus participantes:

Quadro 1 – Objetivos, instrumentos e participantes da pesquisa

Objetivo Geral		
Investigar como a percepção de confiança em profissionais do sistema de justiça representados por imagens geradas por IAG se diferencia da percepção de confiança em imagens reais, considerando os atributos de competência, integridade e benevolência.		
Objetivos Específicos da pesquisa	**Geração de dados**	**Participantes**
Examinar as associações implícitas entre imagens geradas por IAG e atributos de confiança (competência, integridade e benevolência).	Teste de Associação Implícita (IAT)	Pesquisador Participantes
Avaliar a percepção explícita dos usuários em relação à confiabilidade das imagens.	Teste de Atribuição Explícita de Atributos (EAAT)	Pesquisador Participantes
Comparar as medidas implícitas e explícitas de	Teste de Associação Implícita (IAT)	Pesquisador

confiança, identificando possíveis dissociações entre percepção consciente e inconsciente.		
Teste de Atribuição Explícita de Atributos (EAAT)		Pesquisador

Fonte: Elaborado pelo pesquisador (2025).

Com a adoção do IAT como instrumento principal, as questões de investigação foram reformuladas para explorar associações inconscientes entre conteúdos desenvolvidos por IAG e atributos da confiança, especialmente quanto a três dimensões: competência, integridade e benevolência.

Além disso, para complementar a análise implícita, utilizou-se o Explicit Attribute Assignment Test (EAAT), cuja metodologia é baseada em escolha forçada, o que permite capturar percepções conscientes dos usuários sobre os conteúdos analisados. O EAAT, assim, possibilitou verificar se há discrepâncias entre as associações implícitas e as respostas explícitas dos participantes, aprofundando a compreensão sobre como conteúdos gerados por IAG são avaliados em termos de confiabilidade.

A escolha do sistema de justiça como eixo simbólico das imagens utilizadas nos testes está relacionada à atuação profissional do autor como Promotor de Justiça no Ministério Público do Estado de Goiás, o que permite observar, de forma empírica e cotidiana, as implicações da confiança institucional na mediação de informações, na interpretação de evidências e na legitimação de discursos. Assim, a representação visual de figuras associadas à justiça busca explorar as percepções relacionadas à confiança institucional, especialmente no que tange aos atributos de competência, integridade e benevolência – aspectos que se tornam ainda mais sensíveis diante da mediação algorítmica e da produção de conteúdos por IAG.

As questões norteadoras deste trabalho são:

1. **Quais são as associações implícitas feitas pelos usuários entre conteúdos identificados como gerados por IAG e atributos de confiança?**
2. **Como os usuários associam conteúdos criados por IAG à confiabilidade (confiável x duvidoso, honesto x desonesto,**

competente x incompetente, qualificado x desqualificado, capaz x incapaz, seguro x inseguro e simpático x antipático)?

3. **Há diferenças significativas nas percepções de qualidade percebida (valioso x sem valor, profundo x superficial, substancial x vazio) entre conteúdos gerados por IAG e conteúdos criados por humanos?**
4. **Quanto à Percepção Explícita de Confiabilidade, ao serem informados de que determinado conteúdo foi criado por IAG, os usuários o percebem como menos autêntico ou confiável?**
5. **No tocante à Aceitação Consciente da IAG, quando forçados a julgar conscientemente um conteúdo, os usuários demonstram maior resistência ou aceitação em relação a conteúdos de IAG comparativamente aos conteúdos humanos?**

A incorporação do EAAT permitiu a comparação entre os julgamentos automáticos (medidos pelo IAT) e as respostas conscientes (coletadas pelo EAAT). Com isso, foi possível compreender se os usuários apresentam tendências inconscientes de aceitação ou rejeição de conteúdos gerados por IAG, ao mesmo tempo em que racionalmente expressam posicionamentos distintos quando forçados a tomar decisões explícitas. Essas questões guiaram o delineamento experimental, permitindo analisar tanto as associações cognitivas implícitas quanto as percepções conscientes dos usuários em relação ao output desenvolvido por IAG.

Feitas a apresentação do problema de pesquisa e a explanação dos objetivos – geral e específicos – que conduziram este trabalho, segue a estrutura da dissertação.

1.1 ESTRUTURA DA DISSERTAÇÃO

Para melhor compreensão do estudo, estruturou-se e dividiu-se este documento em oito seções, além das referências bibliográficas. A primeira seção introduz o tema da pesquisa, contextualizando a relevância do estudo sobre a confiança em imagens desenvolvidas por IAG, especialmente no sistema de justiça. A partir dessa introdução, apresenta-se o problema de pesquisa, formulado com base nos desafios contemporâneos do uso de IAG na comunicação visual e na percepção pública. São, também, delineados os objetivos gerais e

específicos da investigação, que foram reformulados para considerar a influência dos atributos de competência, integridade e benevolência na construção da confiança, seguindo o modelo de Mayer, Davis e Schoorman (1995). Na segunda seção, inicia-se a revisão de literatura, aprofundando a fundamentação teórica da pesquisa e abordando conceitos essenciais para a compreensão do tema e – ao final – confrontar com os resultados obtidos por meio dos testes. Aborda-se a revolução da IA, na qual é apresentada uma visão histórica e conceitual da evolução da IA, diferenciando a IA estreita (Narrow AI) da IA geral (AGI) e discutindo a importância dos algoritmos no contexto contemporâneo, inclusive no ambiente das redes sociais digitais.

Já na terceira seção são discutidas as emoções nos seres humanos no contexto do tema proposto pela pesquisa. Aborda-se o papel das emoções na tomada de decisão humana, analisando a interação entre emoções e percepções em ambientes digitais mediados por IA.

A quarta seção tem o seguinte título: "Inteligência Artificial e os Novos Parâmetros da Confiança e Credibilidade no Ambiente Digital". Introduz-se o fenômeno da ambivalência cognitiva diante de conteúdos criados por IAG. Discorre-se sobre o conceito de confiança, sua importância nas interações sociais e institucionais e os fatores que a influenciam, destacando-se as dimensões de competência, integridade e benevolência como determinantes da percepção de confiabilidade. Também são exploradas as bases cognitivas da percepção de confiança, considerando as teorias sobre atitudes implícitas e explícitas, bem como os estudos sobre e-trust, que ajudam a compreender como as emoções e a credibilidade se relacionam com conteúdos digitais.

A percepção do usuário sobre *outputs* criados por IAG é o tema da quinta seção, a qual examina as pesquisas existentes sobre dissociação entre atitudes explícitas e implícitas em relação à IAG, além de explorar a teoria do vale da estranheza (*uncanny valley*) e suas implicações na percepção de confiança em imagens produzidas artificialmente, fornecendo um referencial teórico para embasar este estudo.

Em seguida, na sexta seção, discute-se a metodologia adotada, detalhando a escolha do IAT e do EAAT como instrumentos essenciais

para compreender como os participantes percebem imagens de IAG, tanto em nível consciente quanto inconsciente.

A sétima seção é dedicada à análise dos dados encontrados nos testes. Apresenta-se o delineamento experimental do estudo, detalhando a composição da amostra e os procedimentos adotados para aplicação dos testes. Explica-se a estrutura do IAT, descrevendo os blocos de associação de imagens e palavras que permitem medir a rapidez e precisão das respostas dos participantes, identificando padrões de associações inconscientes entre imagens criadas por IAG e atributos de confiança. Complementarmente, discute-se o EAAT, instrumento que viabiliza avaliar a percepção consciente dos participantes sobre as imagens, permitindo uma análise comparativa entre atitudes explícitas e implícitas. Além da descrição dos testes, são abordados os critérios de validação estatística, garantindo a confiabilidade das medidas obtidas.

A título de considerações finais, na última seção, retoma-se a discussão sobre os resultados, tecendo-se considerações conclusivas acerca da investigação, bem como exposição das limitações da investigação e as recomendações sugeridas tendo em vista potenciais investigações futuras. Analisa-se como a percepção de confiança nas imagens geradas por IAG se diferencia da confiança atribuída às imagens reais a partir dos resultados obtidos nos testes IAT e EAAT. Além disso, avaliam-se as implicações desses achados para o uso da IAG no sistema de justiça, refletindo sobre as possíveis consequências da adoção dessas tecnologias para a comunicação institucional.


2

2 A REVOLUÇÃO DA IA

A tecnologia de Inteligência Artificial começou a ser desenvolvida na década de 50. Os primeiros estudos que foram publicizados para o grande público foram realizados por Alan Turing (2012), no chamado teste de *turing*, segundo o qual um jogador humano entrou numa conversa, em linguagem natural, com outro humano e uma máquina programada para produzir respostas indistinguíveis de outro ser humano. Todos os participantes estavam separados um dos outros. Se o juiz não fosse capaz de distinguir, com segurança, o computador do humano, diz-se que a máquina passou no teste. O teste não verificava a capacidade de dar respostas corretas para as perguntas, mas, sim, o quão próximas as respostas são daquelas dadas por um ser humano típico. Desde 1950, o teste provou ser, ao mesmo tempo, altamente influente e criticado, e é um conceito fundamental da filosofia da inteligência artificial.

Turing (2012) definiu a IA como sendo um sistema com a capacidade e competência de alcançar um desempenho em nível de um ser humano em todas as tarefas cognitivas, de maneira a realizar a simulação de uma conversa humana com uma pessoa. Já Russel e Norvig (2016) acrescentaram à discussão o fato de que a inteligência está mais direcionada à aquisição de conhecimento, à planificação e à solução de obstáculos. Noutro giro, na perspectiva dos autores Legg e Hutter (2007), esse conceito também envolve o conhecimento, a autoconsciência, o conhecimento emocional, a razão, a criatividade, a lógica e o pensamento crítico.

Segundo Sheth e Thirunarayan (2021), pode-se dividir o ciclo de desenvolvimento da IA em três grandes períodos. O primeiro deles é aquele compreendido entre as décadas de 1940 e 1980, comumente denominado de paradigma simbolista, o qual a IA era concebida como um sistema lógico-formal capaz de manipular símbolos segundo regras sintáticas previamente definidas, a partir da premissa de que o pensamento humano poderia ser reproduzido por meio da computação simbólica. O marco teórico inaugural dessa abordagem remonta aos trabalhos de Alan Turing, especialmente sua formulação do conceito de máquina universal (Turing, 2012). Na prática, esse paradigma se consolidou com o desenvolvimento de sistemas

baseados em regras, como os sistemas especialistas, que buscavam replicar decisões humanas por meio de uma base de conhecimento e um motor de inferência lógico, ou seja, lógica subjacente era de que o conhecimento podia ser explicitado em proposições formais e operado computacionalmente, o que implicava uma forte confiança na racionalidade cognitiva como estrutura formalizável (Sheth; Thirunarayan, 2021). No entanto, os limites dessa abordagem tornaram-se evidentes com a crescente complexidade das tarefas e a dificuldade de lidar com incertezas, ambiguidades e contextos abertos – características centrais da cognição humana que não podiam ser plenamente modeladas em regras fixas. Todavia, o paradigma simbolista lançou os alicerces epistemológicos da IA e influenciou decisivamente a forma como o campo seria entendido por décadas, contribuindo também para os primeiros embates sobre o que significa inteligência quando dissociada de intencionalidade, subjetividade e contexto vivencial.

A partir da década de 1980, o campo da IA passou por um deslocamento paradigmático com o advento das abordagens conexionistas e estatísticas, marcando o segundo grande ciclo de sua evolução (Sheth; Thirunarayan, 2021). Diferentemente do modelo simbolista, que operava com regras explícitas e manipulação lógica de símbolos, o paradigma conexionista inspirou-se na estrutura dos sistemas neurais biológicos, utilizando redes neurais artificiais para modelar processos cognitivos por meio do aprendizado a partir de dados. Essa transição esteve alinhada ao crescente interesse pelas chamadas *soft computing techniques*, capazes de lidar com incertezas, ruído e imprecisão, em oposição à rigidez lógica da IA simbólica. Sistemas conexionistas aprendem padrões a partir de exemplos, o que permitiu sua aplicação em tarefas complexas como reconhecimento de fala, escrita e imagens. No plano institucional, esse período foi também influenciado pelo desenvolvimento de algoritmos como o *backpropagation*, que viabilizou o treinamento eficaz de redes multicamadas (Rumelhart; Hinton; Williams, 1986), e pelas bases teóricas do aprendizado de máquina, que deram origem a métodos como máquinas de vetor de suporte (SVMs), árvores de decisão e *ensembles*. Essa nova lógica estatística tornou-se dominante nos anos 1990 e 2000, principalmente com a ampliação das capacidades de processamento computacional e da disponibilidade de grandes volumes de dados digitais. No entanto, o paradigma conexionista foi criticado por sua opacidade, isto é, pela dificuldade de interpretar os critérios usados por redes neurais para chegar a uma decisão –

fenômeno que motivaria, anos mais tarde, o surgimento das abordagens explicáveis na terceira onda da IA. Não obstante, essa fase preparou o terreno técnico e epistemológico para os avanços que culminariam na IAG contemporânea.

À luz dos ensinamentos de Sheth e Thirunarayan (2021), a IA atravessa, a partir de 2010, sua terceira grande onda de desenvolvimento, caracterizada pela convergência entre métodos conexionistas de *deep learning*, estruturas simbólicas interpretáveis e uma crescente demanda por explicabilidade. Essa fase é marcada por avanços significativos em modelos de aprendizado profundo, especialmente após a introdução da arquitetura *transformer*, que permitiu o surgimento de modelos fundacionais (*foundation models*) como *BERT*, *GPT* e *DALL·E* (Bommasani *et al*., 2021). Tais sistemas passaram a ser capazes de desenvolver textos, imagens e outras produções com alto grau de verossimilhança, superando marcos anteriores de desempenho e ampliando exponencialmente a capacidade de simulação algorítmica de linguagem natural e raciocínio multimodal. Entretanto, esse salto técnico trouxe novos desafios epistemológicos e éticos, sobretudo no que diz respeito à opacidade das decisões algorítmicas, à confiabilidade dos *outputs* e à atribuição de responsabilidade pelos eventuais erros cometidos (Sheth; Thirunarayan, 2021). Em resposta, consolida-se o movimento da IA explicável (XAI), que busca construir sistemas capazes de justificar suas decisões de forma auditável e compreensível para seres humanos. Simultaneamente, cresce o interesse pela IA neuro-simbólica, que visa a integrar o poder de generalização das redes neurais com a precisão lógica e interpretativa das estruturas simbólicas.

Diante dessas transformações tecnológicas, que ampliam tanto a capacidade de processamento quanto a inteligibilidade das máquinas, a doutrina mais recente tem se empenhado em organizar conceitualmente o campo da IA. Uma das classificações recorrentes distingue entre dois grandes tipos de IA, quais sejam:

a) ***Narrow AI*: *Narrow AI* está destinada a um problema ou tarefas específicas e, por isso, não consegue lidar com outros desafios sem ser novamente treinada ou adaptada. Sistemas de *Narrow AI* ficam para além da flexibilidade da inteligência humana, mas podem ser bastante importantes no seu domínio. A maior parte da IA que está atualmente operacional**

enquadra-se nessa categoria. Alguns exemplos conhecidos são a *Siri*, *Google Assistant* e *Alexa*. (Wirth, 2018).

b) *Strong AI*: *Strong AI* ou *Artificial General Intelligence* (AGI) é visto como um sistema tão poderoso e flexível quanto a inteligência humana e não está adaptado apenas a uma tarefa ou a um problema específico, ao contrário do *Narrow AI*. Uma AGI consegue adaptar-se a novos contextos para além daqueles em que foi treinada (Alves, 2023). Normalmente, esse tipo de IA é adaptável e retratado nos filmes, e, até a presente data, é possível afirmar que não passa de ficção científica, pois, ainda, não foi alcançada (Wirth, 2018).

A consolidação da IAG a partir da década de 2020 representou um salto qualitativo na história dos sistemas artificiais, alterando a capacidade técnica dessas tecnologias, bem como a percepção humana em relação ao conteúdo digital. A partir dos avanços em *deep learning*, emergiu um novo paradigma no qual sistemas de IA classificam ou predizem informações e têm a capacidade de criar autonomamente imagens, textos e sons com grau de realismo (Alves, 2023). De acordo com Kate Crawford (2021), a IAG altera o ecossistema cognitivo ao introduzir incerteza ontológica sobre a origem das informações. Essa incerteza desafia os critérios tradicionais de veracidade, haja vista que afeta emocionalmente os usuários, induzindo sensações de estranhamento, dúvida e, por vezes, desconfiança sistemática em relação ao ambiente digital.

Embora a IA tenha evoluído significativamente em termos de capacidade de processamento, geração de linguagem natural e produção de *outputs* sintéticos, permanece latente uma tensão entre o avanço técnico e a compreensão filosófica daquilo que se convencionou chamar de inteligência (Yigael, 2011). A euforia tecnológica muitas vezes obscurece os limites epistemológicos da IA, ao pressupor que funções cognitivas humanas podem ser integralmente modeladas em linguagem computacional, razão pela qual essa perspectiva, embora funcionalista, é criticada por diversos pensadores contemporâneos, que alertam para o risco de se confundir desempenho com compreensão, e processamento com consciência. Nesse contexto, vozes dissonantes como a de Yigael (2011) tornam-se centrais para uma reflexão crítica sobre os fundamentos da IA, especialmente ao confrontar a suposição dominante de que o comportamento inteligente pode ser simulado de forma mecanicista,

sem consideração pela intencionalidade, semântica e pela estrutura fenomenológica da cognição humana.

A crítica epistemológica elaborada por Yigael (2011), ao distinguir entre a ação automatizada e a cognição dotada de significado, reforça a tese de que a simulação de comportamentos humanos pela IA não equivale à reprodução autêntica da inteligência. Como aponta Yigael (2011), a máquina pode executar ações semelhantes às humanas, mas permanece incapaz de experienciar ou atribuir significância às suas operações, por conseguinte, essa ausência de sentido interno torna a IA um simulacro funcional, mas não uma entidade cognitiva. Em outras palavras, a ação da máquina carece de intencionalidade, pois não se ordena à sua própria preservação, tampouco à construção de um mundo simbólico que sustente sua existência – diferentemente do organismo humano, cuja cognição emerge de necessidades biológicas e interações sociais complexas. Ao adotar essa perspectiva, este trabalho assume como premissa que a confiança do usuário em imagens criadas por IAG não se apoia na suposição de que esses *outputs* visuais têm uma origem inteligente no sentido humano, mas, sim, na forma como essas imagens ativam representações mentais e respostas emocionais implícitas. Daí a relevância de se recorrer a instrumentos metodológicos como o IAT e o EAAT, que buscam captar reações que não são plenamente acessíveis à consciência reflexiva, mas que operam no nível da associação automática entre percepção e julgamento.

2.2 ALGORITMO

O termo algoritmo possui uma longa história associada à matemática e ao raciocínio lógico. A origem etimológica remonta ao matemático persa Abu Abdullah Mohammad Ibn Musa al-Khawarizmi, que viveu no século IX e é amplamente reconhecido por suas contribuições fundamentais à álgebra e à matemática como um todo. Um algoritmo é, essencialmente, um conjunto de passos detalhados que visam à solução de um problema específico e – em termos gerais – pode ser entendido como uma sequência lógica e finita de instruções (Medina; Fertig, 2022).

A literacia algorítmica – entendida como o nível de compreensão dos usuários sobre o funcionamento e as limitações dos algoritmos – emerge como um fator crítico na percepção da credibilidade do *output* produzido por IA (Cotter, 2019). Por conseguinte, a confiança nos

algoritmos é um elemento determinante para que os usuários aceitem ou rejeitem as recomendações feitas por esses sistemas.

A despeito da crescente presença de algoritmos e sistemas de IA no cotidiano dos brasileiros, a compreensão sobre seu funcionamento ainda é limitada, o que acentua os riscos de percepção enviesada ou emocionalmente ambivalente. A pesquisa "Uso de Tecnologias Digitais" (Cetic.br, 2023) mostrou que 57% dos brasileiros afirmam já ter ouvido falar em IA, mas apenas 18% conseguem explicar o que é, e 40% declararam não confiar que algoritmos sejam justos ou imparciais. Esses dados revelam o desconhecimento técnico generalizado sobre a IA, bem como a existência de desconfiança social latente, que pode se manifestar tanto de forma explícita quanto implícita.

O estudo de Shin *et al*. (2020) demonstrou que a interação entre literacia algorítmica e confiança desempenha um papel crucial na forma como os usuários atribuem credibilidade aos sistemas algorítmicos. Os resultados mostraram que: 1) usuários com alta literacia e alta confiança apresentaram a maior percepção de credibilidade dos *chatbots* e buscaram mais informações a partir de suas recomendações; 2) aqueles com alta literacia, mas baixa confiança, foram mais críticos e demonstraram menor propensão a aceitar as recomendações algorítmicas sem questionamento; 3) usuários com baixa literacia e alta confiança demonstraram um risco maior de aceitar informações sem avaliar sua veracidade; 4) já os com baixa literacia e baixa confiança tenderam a rejeitar as recomendações algorítmicas, mesmo quando precisas e relevantes.

Por oportuno, cabe ressaltar a relação entre IA, confiança e credibilidade *on-line*, a qual é elemento-chave deste estudo. Se, por um lado, algoritmos possibilitam maior eficiência na personalização e automação de processos, por outro, podem reforçar viéses cognitivos, bolhas informacionais e distorções na percepção da autenticidade de conteúdos (Gillespie, 2018). Assim, é imperioso aprofundar a discussão sobre a transparência algorítmica e os efeitos da IAG na construção da confiança em espaços digitais.

A IA, particularmente por meio de algoritmos de *machine learning* e *deep learning*, exerce papel fundamental nas redes sociais ao personalizar as informações que os usuários recebem. Essa personalização é baseada na coleta massiva de dados sobre preferências, localização e comportamentos *on-line* (Kaufman, 2019).

A tecnologia de *deep learning* permite que sistemas algorítmicos reconheçam padrões em grandes volumes de dados e façam predições, consequentemente, essa capacidade é o que torna possível a criação de *feeds* personalizados, como o *News Feed* do *Facebook*, que ordena conteúdos de acordo com a relevância percebida para cada usuário (Medina; Fertig, 2022). Segundo Mosseri (2018), o algoritmo de ranqueamento do *Facebook* considera múltiplos fatores, como inventário de histórias disponíveis, predições de comportamento e pontuações de relevância para decidir quais histórias aparecem no topo da página inicial. O sistema de pontuação, conhecido como *Relevance Score*, estabelece afinidades entre o usuário e os objetos de classificação (postagens, usuários, grupos), gerando uma hierarquia de conteúdos (Moseri, 2018).

Esse mecanismo de personalização, embora eficiente, também levanta questionamentos sobre os critérios que definem a visibilidade da informação e os efeitos dessa filtragem algorítmica na formação da opinião pública. Autores como Rouvroy e Berns (2015) discutem o conceito de governamentalidade algorítmica, segundo o qual os algoritmos controlam e regulam as interações humanas, tomando decisões que afetam diretamente a vida cotidiana dos indivíduos. Esses algoritmos não seriam neutros; eles incorporariam valores e prioridades que são definidos por aqueles que os programam e controlam. Como resultado, a governamentalidade algorítmica criaria uma nova forma de poder que se manifesta por meio das plataformas digitais, moldando o comportamento dos indivíduos de maneiras, muitas vezes, invisíveis (Rouvroy; Berns, 2015). Ocorre que esse fenômeno traz um problema amplamente discutido quando se fala de redes sociais, qual seja: a falta de transparência ou opacidade.

Acerca dessa opacidade, Tarleton Gillespie (2018) chama esse sistema de caixa-preta técnica, moldando decisões políticas e culturais da sociedade. A analogia com a caixa preta, elaborada por Flusser (2002), refere-se ao fato de que esses sistemas operam com uma lógica interna inacessível à maioria dos usuários e, até mesmo, aos programadores, dada sua complexidade.

Gillespie (2018) argumenta que a personalização algorítmica pode criar bolhas de informação, nas quais os indivíduos são expostos a conteúdos que reforçam suas visões de mundo existentes e limitando o acesso a perspectivas alternativas. Segundo o autor, os algoritmos são máquinas inertes até que sejam programados para trabalhar com

bancos de dados, nos quais há decisões significativas sobre as informações que serão incluídas ou excluídas. Gillespie (2018) entende que a indexação não é um processo neutro, haja vista que há uma escolha deliberada que reflete interesses comerciais, culturais e políticos.

Outra dimensão central na análise de Gillespie (2018) são os ciclos de antecipação, que se referem às tentativas dos provedores de algoritmos de prever o comportamento e os interesses dos usuários. Com base nas atividades registradas, algoritmos de plataformas, como *Instagram*, *Tiktok*, *Facebook* e *YouTube*, são programados para antecipar preferências, oferecendo *outputs* personalizados e anúncios direcionados. Essa prática criaria um ciclo de retroalimentação e os usuários seriam encorajados a fornecer mais dados para melhorar a capacidade de antecipação do algoritmo, em troca de uma experiência de usuário mais personalizada.

Compreendida a evolução histórica, conceitual e tecnológica da IA, e suas consequências sobre os modos de produção e circulação de informações na sociedade contemporânea, faz-se necessário avançar na análise de outro elemento fundamental para a compreensão da interação humano-máquina: as emoções. Se, por um lado, os algoritmos e sistemas de IA moldam comportamentos e decisões de forma automatizada e preditiva, por outro, as emoções humanas continuam a exercer papel decisivo na percepção de conteúdos/imagens, na construção da confiança e no engajamento em ambientes digitais. Por conseguinte, a próxima seção examina a natureza das emoções, sua função adaptativa e a forma como os ambientes virtuais, mediados por IAG, podem amplificar, modular ou explorar essas respostas afetivas, influenciando, muitas vezes de maneira inconsciente, as escolhas dos usuários.


3

3 INTELIGÊNCIA ARTIFICIAL E OS NOVOS PARÂMETROS DA CONFIANÇA E CREDIBILIDADE NO AMBIENTE DIGITAL

À medida que a IAG avança na reprodução de aspectos antes restritos à cognição humana – como a escrita, a criação visual e sonora, a conversão de fala em texto e a simulação de emoções –, observa-se uma mudança tecnológica, bem como uma verdadeira inflexão paradigmática nas formas de produção, circulação e recepção das mensagens no ambiente comunicacional contemporâneo. As IAGs respondem a comandos e participam da construção ativa da experiência sensível dos sujeitos, tornando-se mediadoras de relações simbólicas, sociais e afetivas, muitas vezes de forma invisível. Essa transformação marca uma nova etapa do que pode ser chamado de revolução cognitiva digital, implicando uma reconfiguração da maneira como humanos percebem, interagem e atribuem sentido ao mundo ao seu redor, conforme corroboram os apontamentos de autores como Zaltman (2003) e Floridi (2021).

No campo da comunicação, a emergência da IAG pode ser compreendida como a continuidade – e ao mesmo tempo o rompimento – de um processo histórico de mediatização. Desde a invenção da imprensa, passando pelo rádio, pela televisão e pela *internet*, os dispositivos técnicos sempre desempenharam o papel de estender os limites da linguagem humana, conectando emissores e receptores, organizando fluxos simbólicos e moldando práticas sociais (Martín-Barbero, 2003). No entanto, o que diferencia a atual etapa tecnológica das anteriores é a capacidade da máquina de simular intenção comunicativa, algo que, até poucos anos atrás, era prerrogativa exclusiva do ser humano. Se antes a mediação era técnica, hoje ela é cognitiva e emocional (Floridi, 2021). Tarleton Gillespie (2018) afirma que avatares, assistentes virtuais e sistemas de recomendação baseados em IAG já não apenas veiculam mensagens: eles as produzem, escolhem, priorizam e modulam, razão pela qual intervém diretamente na construção da realidade percebida pelos sujeitos.

Essa mudança está diretamente associada à explosão recente das tecnologias de IAG, sobretudo a partir de 2022, com o lançamento de ferramentas como o *ChatGPT*, *DALL·E*, *Midjourney* e outras

plataformas de linguagem e imagem natural (Bommasani *et al*., 2021). Em questão de meses, essas tecnologias tornaram-se acessíveis ao público geral, produzindo efeitos sem precedentes em múltiplas esferas da vida social: do consumo à educação, da produção de conteúdo à vida íntima. A comunicação mediada por IAG deixou de ser um experimento de laboratório e passou a ser um elemento onipresente, muitas vezes naturalizado, da experiência digital cotidiana. Como observa Floridi (2021), a IAG tem a capacidade de mudar o que os seres humanos sabem e como sabem – e, sobretudo, como se sentem.

Nesse novo ecossistema digital, a IAG atua como uma instância mediadora simbólica entre emissores, mensagens e receptores. O conteúdo que as pessoas veem, leem ou escutam é uma representação de uma intenção humana direta, além de ser o resultado de um processo automatizado de produção ou curadoria algorítmica. Algoritmos selecionam o que é "relevante", o que "agrada", o que "prende atenção", operando com base em parâmetros estatísticos e perfis comportamentais, muitas vezes sem qualquer transparência ou possibilidade de contestação. Essa curadoria automatizada reconfigura os processos tradicionais de construção de sentido e autoridade discursiva, influenciando os julgamentos que são feitos sobre credibilidade, autenticidade e confiança (Gillespie, 2018).

Ao ocupar esse lugar privilegiado de mediação, a IAG também se insere nas esferas mais sensíveis da vida humana: nossas emoções, memórias, afetos, hábitos e valores. Segundo Zaltman (2003), a maior parte das decisões humanas é governada por processos inconscientes e simbólicos, frequentemente mediados por metáforas profundas que organizam o pensamento e o sentimento. Os sistemas de IAG aprendem com as escolhas humanas que são feitas, antecipam preferências, ajustam respostas e modulam afetos, tornando-se, em alguma medida, extensões do *self* digitalizado, conforme sugere Sherry Turkle (2011), ao refletir sobre como as tecnologias moldam a subjetividade na era da hiperconectividade. Esse processo é ambivalente: de um lado, há o encantamento com a eficiência, a personalização e a ilusão de companhia que essas tecnologias oferecem; de outro, emerge o estranhamento, a desconfiança e a incerteza ontológica sobre a natureza do que é "real" ou "simulado", especialmente quando os sistemas de IAG são projetados para promover familiaridade emocional e comportamentos responsivos, como discutem Montag *et al*. (2024) e Floridi (2021).

Zaltman (2003) argumenta que cerca de 95% das decisões humanas são tomadas de forma inconsciente, sendo influenciadas por emoções, metáforas e experiências implícitas. Essa perspectiva é especialmente relevante para o campo da comunicação digital e dos escopos da presente pesquisa, pois ajuda a compreender como se constrói a confiança em ambientes mediados por tecnologia. Quando a IAG produz a imagem de uma pessoa, redige um texto em primeira pessoa ou responde com aparente empatia a uma dúvida, ela ultrapassa a função de transmitir informação: estabelece uma relação simbólica. Essa relação é capaz de acionar mecanismos emocionais e cognitivos profundos, despertando sensações de conforto, familiaridade, confiança ou autoridade. Por outro lado, também pode ativar reações de estranhamento, receio e rejeição. Ao simular traços humanos, a IAG atua sobre camadas sensíveis da cognição, influenciando os julgamentos dos usuários de maneira sutil e automatizada.

3.1 CONFIANÇA

De maneira geral, a confiança pode ser definida como a expectativa positiva sobre o comportamento de outra parte em um relacionamento baseado na incerteza e na interdependência (Morgan; Hunt, 1994). Rotter (1971), por sua vez, conceitua a confiança como a expectativa generalizada de que a palavra, promessa ou compromisso de um indivíduo ou grupo seja confiável. Já Gambetta (2000) afirma que a confiança pode ser definida como a expectativa de que uma entidade agirá de maneira benéfica ou, pelo menos, não prejudicial em um determinado contexto.

Mayer, Davis e Schoorman (1995) apresentam um modelo integrativo que propõe três fatores essenciais para a construção da confiança em um ambiente organizacional: habilidade (ou competência), benevolência e integridade. Esse modelo será adotado como base conceitual desta dissertação para a análise das percepções de confiança em imagens geradas por IAG, orientando tanto a fundamentação teórica quanto a estrutura dos atributos avaliados no IAT.

O modelo proposto por Mayer, Davis e Schoorman (1995) é considerado integrativo porque reúne, de maneira articulada, diversas abordagens teóricas sobre a confiança que até então estavam dispersas na literatura. Ao combinar dimensões cognitivas (como a

avaliação da competência), afetivas (como a percepção de benevolência) e normativas (como o julgamento de integridade), o modelo fornece uma estrutura coesa para compreender como a confiança é formada e mantida em ambientes organizacionais e relacionais. Por conseguinte, essa integração teórica permite compreender a disposição do indivíduo em confiar, além de fundamentar como essa confiança é modulada por percepções sobre o outro e pelo contexto de risco inerente à relação.

Ainda em relação ao conceito proposto por Mayer, Davis e Schoorman (1995), quando eles discorrem sobre a confiança organizacional, diz respeito à expectativa de que instituições, lideranças e sistemas internos de uma organização agirão de forma competente, íntegra e benevolente em relação a seus membros e ao público externo. No contexto de ambientes corporativos ou institucionais, como o sistema de justiça, essa forma de confiança é construída com base na consistência das ações, transparência das decisões, cumprimento de promessas e percepção de justiça procedimental. A literatura organizacional reconhece que níveis elevados de confiança dentro de uma organização contribuem para a cooperação, a inovação e a redução de conflitos, além de impactarem diretamente na legitimidade percebida por agentes externos. No presente estudo, a confiança organizacional é central para entender como os internautas avaliam a confiabilidade de imagens de profissionais da justiça desenvolvidas por IAG, uma vez que tais imagens ativam representações simbólicas ligadas à autoridade institucional.

No que tange à habilidade, refere-se ao conhecimento, competência e capacidade técnica de um indivíduo ou grupo para realizar determinada tarefa ou função (Mayer; Davis; Schoorman, 1995). Para que a confiança seja estabelecida, o *trustor* (aquele que confia) deve perceber que o *trustee* (o confiado) possui as qualificações necessárias para desempenhar sua função de maneira eficaz, razão pela qual essa percepção está diretamente ligada à especialização e experiência da pessoa ou grupo em um domínio específico.

Em relação à benevolência, esta representa o desejo genuíno de um indivíduo em agir no melhor interesse do outro, sem buscar apenas benefícios próprios (Mayer; Davis; Schoorman, 1995). Esse fator diferencia-se da habilidade, pois está relacionado às intenções e ao grau de cuidado que um indivíduo demonstra para com o outro dentro da relação de confiança. Além disso, a benevolência pode ser

especialmente relevante em ambientes caracterizados por alta incerteza e risco.

O terceiro fator essencial, trazido por Roger Mayer *et al.* (1995), é a integridade, definida como a adesão a princípios éticos e morais aceitos pela parte que confia (Mayer; Davis; Schoorman, 1995). Esse fator envolve consistência entre palavras e ações, transparência e justiça nas relações interpessoais entre os indivíduos, por conseguinte, a integridade percebida em um líder – por exemplo – está diretamente associada ao engajamento dos funcionários e à satisfação no trabalho. Segundo os referidos autores, quando há percepções de falta de integridade, a confiança tende a ser prejudicada, resultando em menor comprometimento e maior intenção de rotatividade.

Dirks e Ferrin (2002) demonstram que a confiança se fortalece à medida que os indivíduos acumulam experiências positivas e confirmam suas expectativas sobre o comportamento dos outros. No mesmo contexto, Coleman (1990) afirma que a confiança surge em situações nas quais um ator assume o risco de depender do outro para alcançar um objetivo, ou seja, trata-se de uma escolha racional, na qual os indivíduos calculam os benefícios e custos de confiar ou não em determinada pessoa ou instituição. Nesse sentido e nas palavras de Colquitt *et al.* (2007, p. 910): "Confiabilidade refere-se às características de um administrador que inspiram expectativas positivas por parte de outros indivíduos, incluindo capacidade, benevolência e integridade"[6]. Essas abordagens tradicionais ajudam a compreender os fundamentos cognitivos da confiança, no entanto precisam ser recontextualizadas diante dos desafios impostos pelos ambientes digitais mediados por IA, nos quais as interações ocorrem entre humanos e sistemas automatizados.

No contexto da tecnologia, a confiança pode ser entendida como a disposição de um indivíduo em depender de um sistema tecnológico, atribuindo-lhe confiabilidade suficiente para que sua adoção seja viável (Taddeo; Floridi, 2011), ou seja, esse instituto (confiança) é um fator essencial nas interações digitais, especialmente quando os usuários precisam avaliar a credibilidade de um conteúdo *on-line*.

[6] Tradução livre do inglês: "Trustworthiness refers to the characteristics of a trustee that inspire positive expectations on the part of other individuals, including ability, benevolence, and integrity".

A confiança na IA e nas plataformas que a utilizam influencia diretamente as decisões de engajamento dos usuários, haja vista que aqueles que confiam na tecnologia e implicitamente na plataforma utilizada estão mais propensos a interagir ativamente com a IAG e a consumir conteúdo gerado por ela. Conforme já explicitado nas seções anteriores, a transparência sobre o uso de IAG e a garantia de privacidade e segurança dos dados são cruciais para fomentar essa confiança.

Contudo, essa confiança não pode ser compreendida isoladamente como uma disposição individual ou uma consequência de *design* técnico; ela está inserida em um ecossistema informacional marcado por assimetrias de poder, práticas extrativas e regimes de vigilância. A forma como os dados dos usuários é coletada, processada e monetizada afeta diretamente sua percepção de segurança e autenticidade das interações. Nesse sentido, a questão da confiança na interação com as IAGs tem sido tema recorrente na lição de diversos acadêmicos. Na obra de Paletta e Costa do Lago (2022, p. 6), fica evidente essa importância:

> Nas ações do cotidiano o indivíduo, na maior parte das vezes, sequer questiona o risco de compartilhar os seus dados com inúmeras plataformas. Boa parte desses sistemas também permite a vinculação de uma conta de e-mail ou rede social durante o acesso a aplicativos e sites. Trata-se de uma maneira cômoda de suprir a necessidade contemporânea de estar conectado a tudo, da forma mais rápida e que exija menos ações.
> Assim como em sua versão tradicional, a lógica do capitalismo digital também visa essencialmente o lucro e é imperativo questionarmos se nossas informações pessoais estariam realmente seguras nas mãos dessas empresas.
> Vivemos a era do Data-Colonialismo em que os dados de acesso às redes são retratos fiéis do que vivemos, gostamos ou estamos propensos a consumir:
> "O Colonialismo de Dados é, em essência, a ordem emergente para a apropriação da vida humana, de forma que esses dados podem ser continuamente extraídos dela para o lucro. Essa extração é operacionalizada por meio de relações de dados, formas de interação um com o outro e com o mundo, viabilizadas por ferramentas digitais. Através das relações de dados, a vida humana não é somente anexada ao capitalismo, mas também se torna sujeita a monitoramento e vigilância contínuos" (Couldry; Mejias, 2019, p. 13).

A confiança está diretamente relacionada ao conceito de risco percebido, que pode ser definido como a incerteza das pessoas em relação aos resultados negativos que podem surgir ao adotar uma nova tecnologia (Bauer, 1960). O risco percebido é particularmente alto em contextos de inovação, no qual a falta de experiência prévia com o produto, por exemplo, pode promover insegurança (Pavlou, 2003). Segundo Molm, Takahashi e Peterson (2000), o risco é um fator necessário para o desenvolvimento da confiança, pois a tomada de decisão em um ambiente incerto exige que o consumidor confie em elementos externos, como a reputação da marca, por exemplo. Esse fenômeno pode ser observado no mercado de dispositivos eletrônicos, no qual marcas estabelecidas, como *Apple* e *Samsung*, conseguem lançar novos produtos com alta taxa de adoção, pois sua confiabilidade já foi consolidada ao longo dos anos (Erdem; Swait, 2006).

Para minimizar o risco percebido, as empresas utilizam estratégias de sinalização de qualidade, sendo a marca um dos principais critérios utilizados (Akerlof, 1970). Quando uma marca já tem uma reputação positiva, os consumidores tendem a associar essa credibilidade ao lançamento de novos produtos, reduzindo a incerteza e aumentando a intenção de adoção (Barone; Taylor; Urbany, 2005). Outro fator que os *websites* e empresas utilizam para minimizar esse risco percebido é relativo ao *layout*. A literatura aponta que fatores visuais têm um papel determinante na percepção de confiança em ambientes digitais. Fogg *et al*. (2003) demonstram que o *design* de um *site* – incluindo um *layout* organizado, selos de qualidade e transparência nas informações – influencia diretamente a credibilidade percebida pelos usuários. Esse princípio pode ser estendido para a análise de imagens elaboradas por sistemas generativos de IA, uma vez que aspectos como realismo, coerência estética e familiaridade visual podem desencadear efeitos na forma como os usuários interpretam e avaliam esses *outputs*. É nesse sentido que esta pesquisa busca investigar se representações visuais produzidas por IAG evocam maior ou menor confiança quando comparadas a imagens reais.

Com o avanço da digitalização, surge o conceito de *e-trust*[7] (confiança eletrônica), que se refere à confiança depositada em agentes tecnológicos e digitais. Taddeo e Floridi (2011) argumentam que a confiança eletrônica é distinta da confiança interpessoal, pois envolve a interação com sistemas automatizados que operam de forma

[7] Esse conceito será mais explorado numa seção específica.

independente. No entanto, pesquisas indicam que a desconfiança nos sistemas digitais é maior quando há incerteza sobre a decisão tomada pela máquina (Dietvorst; Bharti, 2020).

Essa forma de confiança "tecnomediada" – o *e-trust* – não anula os fundamentos psicossociais tradicionais da confiança, ao contrário, reconfigura-os em novos moldes, nos quais a previsibilidade, a competência percebida e as intenções atribuídas aos sistemas automatizados continuam sendo elementos centrais. A confiança, como fenômeno psicossocial, é construída a partir da expectativa de coerência e estabilidade nas ações do outro, bem como da percepção de que há boas intenções e competência envolvidas na relação. No campo da comunicação digital, essa expectativa torna-se mais instável e dependente de signos e pistas inferenciais, já que o contato direto entre sujeitos é, muitas vezes, mediado por interfaces algorítmicas e Ias (Silva; Paletta, 2025).

Nesse contexto, a confiança passa a operar como um princípio organizador da credibilidade, ou seja, da disposição subjetiva para aceitar determinado conteúdo como válido, legítimo ou verdadeiro. Como afirmam Fogg e Tseng (1999), a credibilidade é um julgamento que resulta da combinação entre confiabilidade (*trustworthiness*) e expertise percebida, estando, portanto, diretamente atrelada à confiança prévia que o receptor deposita no emissor – mesmo quando esse emissor é um sistema automatizado. No presente estudo, esse modelo é particularmente útil para compreender como os internautas avaliam imagens de autoridades do sistema de justiça quando são informados de que essas imagens foram geradas por IA: a aparência visual ativa a percepção de expertise, mas o rótulo (IA ou real) influencia a confiabilidade, compondo, assim, a percepção final de credibilidade. Dessa feita, compreender os mecanismos que sustentam a confiança torna-se essencial para investigar como os usuários atribuem ou negam confiança a conteúdos elaborados por IAG, especialmente em cenários que exigem julgamento crítico, como o jurídico, o científico e o informacional.

3.2 CREDIBILIDADE *ON-LINE*

Com a *internet*, o acesso ao conhecimento tornou-se democrático, embora também mais caótico, haja vista que qualquer indivíduo pode produzir e compartilhar conteúdo sem necessariamente atender a critérios tradicionais de validação. De

maneira clássica, a credibilidade pode ser definida como a percepção da confiabilidade e expertise de uma fonte ou informação (Fogg, 2003).

Uma abordagem amplamente utilizada para compreender como os internautas avaliam a credibilidade da informação *on-line* é a teoria da conversação de Gordon Pask (1976). Segundo essa teoria, o conhecimento não se adquire de forma passiva, sendo uma construção que se dá por meio da interação entre diferentes agentes. No contexto digital, essa interação ocorre principalmente por redes sociais, fóruns, comentários, sistemas de recomendação coletiva e – mais recentemente – inteligência artificial.

Essa questão da credibilidade, especificamente, foi objeto de outras abordagens teóricas importantes para o presente trabalho. Oh *et al*. (2020) realizou um experimento no qual participantes avaliaram notícias criadas por humanos e por algoritmos de IAG. Os resultados mostraram que, embora as notícias automatizadas fossem percebidas como objetivas, eram também consideradas mais artificiais e menos confiáveis. Por conseguinte, os usuários tendem a confiar mais em conteúdos que incluem explicações sobre como foram desenvolvidas e quais critérios foram usados para sua seleção. Embora o referido experimento não tenha sido realizado diretamente com representações visuais produzidas por IAG, pode-se inferir que o raciocínio seja o mesmo a ser investigado nesta dissertação.

O crescimento das plataformas colaborativas, como a *Wikipedia* e o *Reddit*, exemplifica esse processo. Diferentemente das enciclopédias tradicionais, em que a credibilidade da informação é garantida pela autoridade editorial, a *Wikipedia* adota um modelo de credibilidade emergente, no qual os próprios usuários corrigem, refinam e verificam as informações. A literatura especializada indica que, em muitas áreas, a qualidade da informação na *Wikipedia* se equipara ou até supera fontes tradicionais, como a *Encyclopaedia Britannica* (Giles, 2005).

Além disso, a ascensão de *sites* de *fact-checking*, como *Snopes* e Aos Fatos, demonstra que a credibilidade na era digital não depende somente da origem da informação, mas, sim, da capacidade de 45orna45er-la coletivamente. Esse novo paradigma reforça a ideia de que a confiabilidade não é um atributo fixo, ao contrário, trata-se de um processo dinâmico, construído e validado por meio da interação social (Flanagin; Metzger, 2008). A pesquisa de Flanagin e Metzger (2000)

demonstrou, portanto, que a credibilidade da informação *on-line* é afetada por diversos fatores, como a percepção da expertise do emissor, a qualidade da argumentação e a consistência da informação em diferentes fontes.

Esse entendimento processual da credibilidade é particularmente relevante quando se trata de conteúdos produzidos por IAG, cujas formas de verificação e validação ainda estão em construção. As mesmas variáveis identificadas por Flanagin e Metzger (2000) – como expertise percebida, consistência intertextual e qualidade argumentativa – tornam-se ainda mais desafiadoras quando aplicadas a conteúdos cujo emissor não é humano, mas algorítmico. Diante da complexidade do ambiente digital, os usuários frequentemente tomam decisões baseadas em heurísticas cognitivas. As heurísticas são "[...] mecanismos que reduzem a carga cognitiva necessária para a tomada de decisões e facilitam a interpretação de grandes quantidades de informações" (Taraborelli, 2008, p. 195). Flanagin e Metzger (2000) discutem como os usuários da *internet* empregam heurísticas cognitivas na avaliação da credibilidade da informação *online*, por conseguinte, em vez de um exame detalhado e sistemático dos *outputs*, os indivíduos costumam recorrer a atalhos mentais que facilitam e agilizam a tomada de decisão, embora possam introduzir vieses e erros na avaliação da confiabilidade das informações.

A heurística da autoconfirmação ocorre quando os indivíduos acreditam mais facilmente em informações que confirmam suas crenças pré-existentes, rejeitando ou questionando informações que as contradizem. Esse fenômeno está ligado ao viés de confirmação, bem documentado na literatura da psicologia cognitiva (Klayman; Ha, 1987).

Por outro lado, há, também, a heurística da violação das expectativas, a qual pode ser melhor explicada com um exemplo: quando um *site* ou fonte apresenta erros gramaticais, *design* amador ou informações contraditórias, os usuários tendem a 46orna46era-lo menos confiável. O fator estético e a usabilidade influenciam fortemente as percepções de credibilidade (Fogg, 2003). Estudos demonstram que *sites* bem estruturados e visualmente profissionais tendem a ser julgados como mais confiáveis, independentemente da qualidade da informação apresentada (Flanagin; Metzger, 2008). Esse ponto é especialmente relevante para ajudar a pensar a influência de uma imagem na credibilidade transmitida ao internauta.

Em ambientes digitais, nos quais algoritmos personalizados, inclusive utilizando a IAG, promovem conteúdos alinhados às preferências dos usuários, essa heurística pode reforçar bolhas informacionais e contribuir para a polarização ideológica (Flanagin; Metzger, 2008). O referido estudo mostra que os usuários frequentemente utilizam heurísticas como reputação, consistência e expectativa na avaliação da credibilidade de conteúdos *on-line*, razão pela qual se pode concluir que imagens identificadas como geradas por IAG podem acionar heurísticas específicas, como a da violação de expectativas (se a IA for percebida como menos autêntica) ou da reputação (se a IA for associada a grandes plataformas confiáveis). Assim, a credibilidade é influenciada por processos automáticos e não totalmente conscientes, reforçando a pertinência do uso do IAT para investigar as associações implícitas entre conteúdo gerado por IAG e os atributos de confiança (competência, integridade e benevolência).

No ecossistema digital contemporâneo, a credibilidade deixa de ser atribuída exclusivamente ao conteúdo em si ou à reputação explícita da fonte, e passa a ser fortemente influenciada pelos mecanismos invisíveis que organizam e filtram o acesso à informação. Nesse cenário, a atuação dos algoritmos se torna central, uma vez que são eles os responsáveis por definir o que é visível, relevante ou confiável para cada usuário. Como apontam Bucher (2018) e Gillespie (2014), os algoritmos operam como curadores automatizados de conteúdo, estabelecendo hierarquias simbólicas e afetando diretamente a percepção de autoridade e legitimidade discursiva. Essa influência algorítmica, muitas vezes opaca e difícil de ser compreendida pelos usuários, acaba por tensionar ainda mais os critérios de julgamento da credibilidade em ambientes digitais. Diante disso, é imperioso compreender os algoritmos como ferramentas técnicas e atores sociotécnicos que modulam o fluxo informacional, e que se torna essencial para aprofundar a análise da confiança e da percepção de veracidade na comunicação mediada por IAG – tema que será abordado na próxima seção.

3.3 AMBIVALÊNCIA COGNITIVA DIANTE DA IAG

A forma como os usuários avaliam a confiabilidade de *outputs* gerados por IAG não é um processo puramente racional, mas, sim, influenciado por processos cognitivos automáticos e implícitos. A Teoria da Mediação Tecnológica é particularmente relevante para ajudar a

entender como as tecnologias de comunicação, a exemplo de imagens criadas por IAG, mediam as relações entre os usuários e os conteúdos nas plataformas digitais. A interação com a tecnologia transformou profundamente a comunicação humana, sobretudo com a integração da IAG em plataformas digitais.

Nesse cenário, a Teoria da Mediação Tecnológica oferece uma abordagem crítica para entender como as ferramentas digitais influenciam as percepções e comportamentos dos usuários. Essa teoria, proposta por Marshall McLuhan (1964) e, posteriormente expandida e sistematizada por autores como Don Ihde (1990) e Haddon e Silverstone (2000), sugere que as tecnologias são atores ativos que moldam a informação e a interação.

A aplicação dessa teoria auxilia o presente trabalho a investigar como as representações visuais produzidas por IAG reconfiguram as relações de confiança e credibilidade entre os usuários e as representações simbólicas de autoridade institucional na *internet*. McLuhan (1964, p. 7) afirma que "[...] o meio é a mensagem", sugerindo que a forma como a informação é entregue pode ser tão ou mais significativa que o conteúdo da mensagem, por conseguinte as representações visuais produzidas por IAG, enquanto mediações tecnológicas, moldam percepções, afetos e julgamentos de confiança, especialmente no ambiente simbólico e visual das interações digitais.

A crescente presença da IAG na produção de conteúdo digital tem levantado questões sobre a confiabilidade, autenticidade e qualidade percebida desses materiais. O modo como os usuários avaliam esses *outputs* é influenciado por processos cognitivos automáticos e deliberativos, conforme descrito na Teoria do Processo Dual de Kahneman e Frederick (2005). Esses estudos demonstram que a tomada de decisão humana ocorre por meio de dois sistemas distintos: o primeiro seria "Sistema 1", caracterizado por ser rápido, intuitivo e baseado em processos automáticos, emocionais e heurísticos, que responde de forma instantânea a estímulos sem necessidade de reflexão profunda; o segundo seria o "Sistema 2", mais lento e analítico, que requer maior esforço cognitivo e é ativado quando há necessidade de análise detalhada das informações.

Nos dizeres de Kahneman e Frederick (2005), a interação entre esses sistemas influencia diretamente a percepção do usuário em relação à confiabilidade do conteúdo, visto que, diante de uma situação

em que um usuário se depara com um *post* rotulado como “gerado por IA”, o Sistema 1 pode ativar respostas automáticas de desconfiança, noutro giro, o Sistema 2, ao ser acionado, pode reavaliar a informação de maneira mais racional. Porquanto, heurísticas cognitivas desempenham um papel essencial na avaliação de credibilidade desses conteúdos, de forma que a heurística da representatividade faz com que um *post* que não se encaixa no padrão típico de uma comunicação humana legítima (ou uma “representação institucional autêntica”) seja automaticamente percebido como menos autêntico.

A literatura cognitiva reforça que tais sistemas não operam de modo isolado, mas em um fluxo contínuo no qual o Sistema 1 gera uma resposta inicial e o Sistema 2, quando acionado, monitora, corrige ou substitui esse julgamento preliminar. Kahneman e Frederick (2005) destacam que muitos erros de avaliação surgem justamente quando o Sistema 2 não intervém de forma eficaz, permitindo que heurísticas rápidas governem o processo decisório. No contexto da IAG, isso significa que elementos visuais minimamente discrepantes — como textura artificial, proporções corporais ligeiramente incomuns ou ausência de microexpressões — podem ativar automaticamente percepções de baixa confiabilidade, mesmo quando o observador, em nível racional, reconhece que a geração sintética não implica necessariamente má-fé ou manipulação.

A distinção entre respostas automáticas e deliberadas é fundamental para interpretar os resultados desta pesquisa, uma vez que o IAT acessa predominantemente respostas intuitivas do Sistema 1, enquanto o EAAT mobiliza avaliações conscientes mediadas pelo Sistema 2. A dissociação observada entre as duas medidas coaduna-se com a teoria do processo dual, sugerindo que a artificialidade visual continua a operar como um gatilho perceptivo para julgamentos automáticos de desconfiança, mesmo quando, em nível explícito, os participantes reconhecem a legitimidade da IAG como ferramenta comunicacional. Essa dinâmica demonstra que o componente afetivo-intuitivo permanece relevante na formação da confiança, sobretudo em domínios institucionais como o sistema de justiça, nos quais a credibilidade visual é especialmente sensível.

A fluência cognitiva, ainda, exerce impacto significativo, pois conteúdos de IAG que apresentam erros de coerência ou que parecem “perfeitos demais” podem promover desconfiança nos usuários. O viés de ancoragem, também, se faz presente, uma vez que a primeira

impressão sobre um conteúdo (por exemplo, saber que foi gerado por IAG) pode influenciar de forma desproporcional sua credibilidade percebida.

Produções científicas recentes, como os de Fortunati e O'Sullivan (2020), apontam para a necessidade de compreender as nuances de como os usuários percebem a presença e a atuação da IAG em suas interações cotidianas. Ao explorar essa teoria, o presente trabalho investiga diferentes dimensões, como a autenticidade percebida do conteúdo (imagens) criado por IAG e sua influência na confiança e no engajamento do usuário. Segundo Dal Verne (2023), as tecnologias de IAG influenciam diretamente as emoções dos consumidores digitais, o que dialoga com a Teoria da Mediação Tecnológica ao mostrar que as plataformas digitais não são neutras, ao contrário, moldam ativamente as interações humanas.

> O papel das emoções na tomada de decisões humanas e, portanto, da parte irracional e inconsciente do cérebro, desafiou a suposição do consumidor como o chamado *homo economicus* que age racionalmente no mercado movido apenas por considerações utilitárias (Dal Verne, 2023, p. 86)[8].

A Teoria do Processo Dual reforça o entendimento de que os usuários avaliam a credibilidade de representações geradas por IAG na *internet*, consequentemente, a prevalência de processos intuitivos pode desencadear vieses negativos em relação à IA, enquanto a ativação do pensamento analítico pode mitigar essas percepções. O desafio para empresas e plataformas digitais é encontrar um equilíbrio entre transparência, engajamento e confiança, garantindo que a IA seja utilizada de forma eficaz e responsável na mediação de conteúdo *on-line*. Esse aspecto se torna ainda mais sensível no contexto do sistema de justiça (como apresentado no IAT realizado na presente pesquisa), no qual a confiança institucional é um pilar fundamental e qualquer ambiguidade na mediação tecnológica pode comprometer a legitimidade percebida, especialmente quando associados a símbolos de autoridade, como os do poder judiciário.

A discussão sobre confiança e credibilidade em ambientes digitais revelou a complexidade das avaliações feitas pelos usuários

[8] Tradução livre do inglês: "The role of emotions in human decision-making, and thus of the irrational and unconscious part of the brain, has challenged the assumption of the consumer as the so-called homo economicus who acts rationally in the market driven only by utilitarian considerations".

diante de conteúdos mediados por algoritmos e IAG. No entanto, para compreender de maneira mais aprofundada os mecanismos cognitivos e afetivos que sustentam essas avaliações, é necessário analisar como os indivíduos percebem concretamente os conteúdos produzidos por IAG. Nesse sentido, a próxima seção explora a percepção dos usuários, abordando a dissociação entre atitudes explícitas e implícitas, bem como fenômenos como a teoria do vale da estranheza, que contribuem para a compreensão das ambiguidades emocionais e cognitivas que permeiam a interação com agentes artificiais.


4

4 PERCEPÇÃO DO USUÁRIO SOBRE CONTEÚDOS GERADOS POR IAG

Por oportuno, é essencial compreender como os usuários identificam e percebem o conteúdo gerado por IAG, mesmo quando personalizados. Esta seção apresenta um conjunto de pesquisas (recentes e antigas) que investigam como os usuários identificam e percebem *outputs* criados por IAG em diferentes contextos – incluindo redes sociais, *e-commerce*, comunicação institucional e, especialmente, o sistema de justiça. Essa revisão busca compreender as dinâmicas cognitivas e emocionais envolvidas na recepção de conteúdos artificialmente produzidos, com destaque para fatores como transparência, autenticidade e credibilidade. Investigações científicas, como os de Smith (2018), sugerem que a transparência a respeito do uso de IAG na promoção de conteúdo é crucial para a confiança do internauta, que só pode ser alcançada por meio de uma maior abertura sobre o funcionamento desses sistemas, incluindo como os dados e as informações são utilizados. Isso garante que essas ferramentas operem de maneira justa e sem preconceitos, haja vista que a credibilidade que as plataformas digitais e os sistemas automatizados conferem aos usuários torna-se importante para a aceitação e adoção dessas tecnologias.

Estudos empíricos, como o de Moriuchi (2019), investigam como a transparência no uso de IA afeta a percepção do usuário. O referido pesquisador descobriu que a revelação de que um material ou serviço é desenvolvido por IAG, quando comunicada de forma transparente, pode aumentar a aceitação do usuário, uma vez que esse atributo (a transparência) ajuda a mitigar a sensação de engano e a construir uma relação de confiança com a tecnologia.

Com efeito, a capacidade de discernimento entre conteúdos genuinamente humanos e aqueles fabricados por máquinas pode variar de maneira significativa entre indivíduos, influenciando diretamente a receptividade e engajamento com o conteúdo. Ocorre que os usuários com frequência não conseguem explicar racionalmente suas decisões, pois grande parte do processamento cognitivo ocorre abaixo do nível consciente (Zaltman, 2003, p. 35). Isso sugere que a aceitação desse material desenvolvido por IAG pode

depender de fatores sutis e subjetivos, como as emoções despertadas pelo *design* e narrativa do conteúdo.

Quando os usuários se tornam cientes da utilização de uma IAG, a compreensão de que suas emoções estão sendo manipuladas por algoritmos pode causar reações emocionais adversas, incluindo sentimentos de manipulação e invasão de privacidade. Do ponto de vista da neurociência, saber que suas respostas emocionais e comportamentais são intencionalmente moduladas por IA pode desencadear reações emocionais complexas, como desconforto e ambivalência, à medida que os usuários compreendem os mecanismos de recompensa e validação social utilizados pelas plataformas. Sérgio Ferreira (2021, p. 51-52), em seu artigo intitulado "O que é (ou o que estamos chamando de) 'Colonialismo de Dados'?", conclui:

> Jaron Lanier, em seu livro de título contundente *Dez argumentos para você deletar agora suas redes sociais*, traz um depoimento de Sean Parker, primeiro presidente do Facebook, que descreve essa intencionalidade de controle e modulação comportamental das plataformas de redes sociais: 'Precisamos lhe dar uma pequena dose de dopamina de vez em quando, porque alguém deu like ou comentou em uma foto ou uma postagem, ou seja lá o que for [...]. Isso é um circuito de feedback de validação social [...] exatamente o tipo de coisa que um hacker como eu inventaria, porque explora uma vulnerabilidade na psicologia humana.
> [...]
> A argumentação de Lanier a respeito dos riscos do uso das plataformas de redes sociais vai justamente pensar essa exploração das pequenas doses de dopamina – substância que funciona como neurotransmissor e é associada às sensações de prazer e ao sistema de recompensa do cérebro – por meio da validação social presente nos mecanismos de respostas de outros usuários sobre o conteúdo publicado (botões de reações ou comentários, por exemplo).

No mesmo sentido do estudo de Ferreira (2021) apontado acima, Kang e Lou (2022) afirmam que os internautas têm reações emocionais variadas ao interagir com conteúdo personalizado por IAG. Os referidos autores fazem uma análise específica da plataforma do *TikTok* e concluem que a personalização baseada em IAG é geralmente bem recebida, pois proporciona uma experiência mais relevante e envolvente. No entanto, os usuários também podem experimentar uma mistura de emoções positivas e negativas. Emoções positivas incluem prazer e satisfação ao ver conteúdo que se alinha com seus

interesses, enquanto emoções negativas podem surgir de preocupações com privacidade e a sensação de manipulação algorítmica. *In verbis*:

> Os usuários têm reações emocionais variadas ao interagir com conteúdo personalizado por IA. No caso do *TikTok*, a personalização baseada em IA é geralmente bem recebida, pois fornece uma experiência mais relevante e envolvente. No entanto, os usuários também podem experimentar uma mistura de emoções positivas e negativas. As emoções positivas incluem prazer e satisfação ao visualizar conteúdo que se alinha com seus interesses, enquanto as emoções negativas podem surgir de preocupações com a privacidade e da sensação de manipulação algorítmica (Kang; Lou, 2022, p. 6)[9].

Ainda nesse contexto de possível ativação de uma sensação emocional de desconforto com a utilização da tecnologia de IAG na *internet*, Wiederhold (2020) entende que a percepção dos usuários de que suas preferências e comportamentos estão sendo analisados e influenciados sem seu total conhecimento pode ter efeitos negativos, especialmente em relação à transparência, confiança e autonomia. Nos dizeres da referida autora: "A ideia de que as empresas podem ter acesso e manipular as preferências dos consumidores a um nível neurológico levanta preocupações sobre a transparência, a confiança e a potencial perda de autonomia individual na tomada de decisões" (Wiederhold, 2020, p. 9)[10]. Essa percepção de manipulação silenciosa e invasiva que pode interferir no senso de autonomia do indivíduo é que se insere a proposta deste estudo, ao investigar de que maneira atributos presentes em imagens geradas por IAG – sobretudo aqueles relacionados à confiança – influenciam julgamentos implícitos sobre símbolos relacionados à justiça, por exemplo.

---

[9] Tradução livre do inglês: "Users have varied emotional reactions when interacting with AI-personalized content. In the case of TikTok, AI-based personalization is generally well received, as it provides a more relevant and engaging experience. However, users can also experience a mix of positive and negative emotions. Positive emotions include pleasure and satisfaction when viewing content that aligns with their interests, while negative emotions can arise from privacy concerns and the feeling of algorithmic manipulation".

[10] Tradução livre do inglês: "The idea that businesses can access and manipulate consumer preferences at a neurological level raises concerns about transparency, trust, and the potential loss of individual agency in decision-making".

A literatura especializada sobre o sistema de recompensa, como a de Wise (2005), destacam que regiões, como o núcleo *accumbens*[11], desempenham um papel central na percepção de recompensas, sendo ativadas por estímulos positivos. No contexto das interações com imagens desenvolvidas por IAG, respostas personalizadas ou reconhecimento social podem desencadear a liberação de dopamina, ativando o núcleo *accumbens* e promovendo emoções positivas. Esse processo reforça comportamentos de engajamento, demonstrando como os mecanismos de recompensa podem ser explorados por plataformas digitais para aumentar a interação do usuário. Essa relação evidencia o impacto dos algoritmos de personalização na modulação do comportamento humano (Wise, 2005). Assim, estímulos visuais com alta carga simbólica – como os vinculados à autoridade institucional – podem ativar respostas automáticas de confiança ou rejeição, mesmo sem consciência plena desse processamento.

De fato, a personalização extrema pode levar à criação de "bolhas de filtro", nas quais os usuários são expostos apenas a informações que reforçam suas crenças e interesses existentes, potencialmente prejudicando a confiança na plataforma como um meio equilibrado de informação. Além disso, Kang e Lou (2022, p. 8) entendem que, "[...] além disso, se os utilizadores perceberem que os *bots* de IA estão a recolher e a utilizar dados de forma invasiva, isso pode levar a uma diminuição da confiança e, consequentemente, do envolvimento"[12].

Nessa mesma linha de raciocínio, Wolton (2010) observa que a verdadeira comunicação (incluindo aquela realizada com uma IA) requer a negociação contínua e a convivência pacífica, destacando a importância de se afastar da tecnicidade para focar na essência humana das interações. A revelação de que o conteúdo é criado por IAG pode levar a uma diminuição da confiança, haja vista que os usuários podem questionar a autenticidade das mensagens e a intenção por trás das comunicações automatizadas, à medida que a percepção de

[11] Núcleo *accumbens*: Estrutura localizada no prosencéfalo basal, parte integrante do sistema mesolímbico, está fortemente associada à regulação da motivação, recompensa e reforço comportamental. Estudos neurocientíficos indicam que a ativação dessa região está relacionada à liberação de dopamina, que é crucial para a sensação de prazer e para o aprendizado de comportamentos recompensadores (WISE, 2005).

[12] Tradução livre do inglês: "[...] moreover, if users perceive that AI bots are collecting and using data invasively, it can lead to a decrease in trust and consequently in engagement".

que estão sendo manipulados por algoritmos pode reduzir a confiança nas plataformas e, consequentemente, diminuir o engajamento:

> Os raros pensadores que se ocuparam dessas questões teóricas tiveram de enfrentar nos anos 1970, no que se refere à televisão, as mesmas resistências que existiram nos anos 1950 em relação ao rádio. Misturam-se nisso o temor ao grande número, a desconfiança em relação à cultura e à democracia de massa e o medo da imagem, do face a face e da alteridade, assim como a fraqueza da mentalidade crítica em relação à tecnologia, que é tão importante quanto a reflexão sobre a ciência, e a dificuldade em admitir a inteligência do receptor (Wolton, 2010, p. 61-62).

No que tange ao sentimento do usuário ao descobrir que o conteúdo foi produzido exclusivamente por IAG, é necessário fazer uma distinção quanto à finalidade do *post* na rede social. Quando este tem o escopo de *e-commerce*, pode-se dizer que os *feedbacks* são, em sua maioria, positivos. Essas são as conclusões dos estudos de Muhammad Bilal, Yunfeng Zhang, Shukai Cai, Umair Akram, Alrence Halibas (2024) sobre essa abordagem específica.

A satisfação dos internautas também é influenciada pelo apego afetivo, que se refere ao vínculo emocional que um usuário desenvolve com uma marca ou tecnologia. Esse apego pode moderar a relação entre satisfação e intenção de compra, aumentando a probabilidade de que usuários satisfeitos façam compras repetidas (Hsiao; Chen, 2016). A IAG pode fortalecer esse apego ao criar interações personalizadas e significativas, que fazem os usuários se sentirem valorizados e compreendidos (Hsiao; Chen, 2016).

> O apego afetivo pode desempenhar um papel na intenção de compra *on-line*, pois envolve sentimentos de confiança e segurança, que são importantes em qualquer tipo de relacionamento. No comércio eletrônico, o apego afetivo se refere à conexão emocional de um consumidor com uma marca, *site* ou vendedor específico (Hsiao e Chen, 2016). Wang *et al.* (2020) descobriram que os consumidores que se sentiam confortáveis e confiantes comprando em uma loja ou marca *on-line* específica eram mais propensos a fazê-lo no futuro. No entanto, se os consumidores têm um apego afetivo negativo a uma loja ou marca *on-line* específica, eles podem ter menos probabilidade de fazer compras desse vendedor, porque podem não se sentir confiantes ou confortáveis com a transação (Yang *et al.*, 2020). Portanto, construir e manter um apego afetivo positivo com os clientes é essencial para os negócios *on-line*. A imagem da marca pode ser apoiada

> oferecendo uma experiência pessoal, fornecendo excelente atendimento ao cliente e mantendo uma imagem consistente e confiável. Ao fazer isso, as empresas podem aumentar a fidelidade do cliente e impulsionar a intenção de compra *on-line* (Bilal *et al*., 2024, p. 5)[13].

Imperioso traçar um paralelo entre a experiência do usuário (EC) e a utilização da IAG, haja vista que aquela é composta por elementos cognitivos, sensoriais, emocionais e físicos, por conseguinte pode contribuir sobremaneira como esta poderá interagir com as pessoas. No contexto da IAG, o elemento cognitivo envolve funções mentais superiores, como percepção, memória e resolução de problemas. A IAG melhora esses aspectos ao oferecer recomendações personalizadas baseadas em grandes volumes de dados, proporcionando uma experiência de compra mais fluida e satisfatória (Potdar *et al*., 2018). Veja-se:

> A experiência do consumidor descreve como os clientes interagem e se sentem sobre uma empresa ou marca com base em suas interações e reações. A experiência de um consumidor compreende uma combinação de elementos cognitivos, sensoriais, emocionais e físicos. O elemento cognitivo consiste em funções mentais de nível superior, incluindo pensamento abstrato, percepção, linguagem, memória e resolução de problemas (American Psychological Association, 2016, s.n.)[14].

A personalização habilitada por IAG também pode ser crucial durante a fase de ativação e retenção, uma vez que os sistemas de IA podem monitorar o comportamento dos usuários dentro de um

---

[13] Tradução livre do inglês: “Affective attachment can play a role in online purchase intention, as it involves feelings of trust and security, which are important in any type of relationship. In e-commerce, affective attachment refers to a consumer’s emotional connection with a particular brand, website, or seller (Hsiao and Chen, 2016). Wang et al. (2020) find that consumers who felt comfortable and confident purchasing from a particular online store or brand were likelier to do so in the future. However, if consumers have a negative affective attachment to a particular online store or brand, they may be less likely to make purchases from that seller, because they might not feel confident or comfortable with the transaction (Yang et al., 2020). Therefore, building and maintaining a positive affective attachment with customers is essential for online businesses. Brand image can be supported by offering a personal experience, providing excellent customer service, and maintaining a consistent and trustworthy image. By doing so, businesses can increase customer loyalty and boost online purchase intention”.

[14] Tradução livre do inglês: “Consumer experience describes how customers interact with and feel about a company or brand based on their interactions and reactions. A consumer’s experience comprises a combination of cognitive, sensory, emotional, and physical elements. The cognitive element consists of higher-level mental functions, including abstract thinking, perception, language, memory, and problem solving”.

aplicativo ou serviço e fornecer recomendações personalizadas ou intervenções automatizadas que aumentem o engajamento e a satisfação do usuário. Análise preditiva pode identificar usuários em risco de *churn*[15] e acionar ações específicas para retê-los, como ofertas especiais ou melhorias no produto que atendam às suas necessidades específicas. Ellis e Brown (2017, p. 10) destacam:

> Assim que o programa de indicação foi ao ar, imediatamente vimos convites sendo enviados por *e-mail* e mídias sociais, resultando em um aumento de 60% nas inscrições de indicação. O plano estava funcionando, sem dúvida, mas não paramos por aí; determinados a aproveitar ao máximo a oportunidade, nossa equipe trabalhou furiosamente por semanas para otimizar cada elemento do programa, desde as mensagens, até as especificidades da oferta, os convites por *e-mail*, a experiência do usuário e os elementos da interface[16].

Segundo Potdar *et al.* (2018), os elementos emocionais são igualmente importantes, uma vez que a IAG pode evocar sentimentos de prazer e satisfação por meio de interações personalizadas e proativas. Essas interações podem incluir desde recomendações por imagens de produtos até o atendimento ao cliente via *chatbots*, que resolvem problemas de maneira eficiente e amigável (Hsiao; Chen, 2016). Além disso, os elementos físicos e sensoriais, como a facilidade de navegação em interfaces *on-line*, também são aprimorados com o uso de IAG, tornando a experiência do consumidor digital mais intuitiva e agradável (Hsiao; Chen, 2016).

> A IA afeta positivamente a experiência do consumidor e o engajamento do consumidor nas mídias sociais. Da mesma forma, existe uma relação positiva entre o engajamento nas mídias sociais e a experiência do consumidor, levando a um consumidor mais satisfeito e intenções de compra

[15] O risco de *churn*, também conhecido como *churn rate*, refere-se à probabilidade de um cliente deixar de utilizar os produtos ou serviços de uma empresa. O termo *churn* é utilizado principalmente em setores em que a retenção de clientes é fundamental, como telecomunicações, serviços de assinatura, SaaS (*Software as a Service*) e bancos. A compreensão e gestão do risco de *churn* são essenciais para a sustentabilidade e crescimento das empresas, pois a perda de clientes pode impactar significativamente a receita e a reputação. Resposta gerada pelo modelo de linguagem *ChatGPT*, da *OpenAI*. Disponível em: https://www.openai.com/chatgpt. Acesso em: 2 jul. 2024.

[16] Tradução livre do inglês: "Once the referral program went live, we immediately saw invites flooding out via email and social media, resulting in a 60 percent increase in referral sign-ups. The plan was working, no doubt about it, but we didn't stop there; determined to make the most of the opportunity, our team worked furiously for weeks to optimize every element of the program, from the messaging, to the specifics of the offer, to the email invites, to the user experience and interface elements".

> amplificadas. Além disso, o apego afetivo modera a relação entre a satisfação do consumidor e a intenção de compra. Os resultados revelam que a IA pode ser usada nas mídias sociais para melhorar a experiência do consumidor e aumentar os níveis de satisfação do cliente e a intenção de compra (Bilal *et al*., 2024, p. 2)[17].

Estudos como o de Israfilzade e Sadili (2024) demonstram que se vivencia uma evolução no *marketing*, haja vista que se passa de um monólogo para o diálogo dinâmico impulsionado pela IAG, que personaliza as experiências do usuário por meio de diálogos sob medida em tempo real, potencializando o engajamento e a satisfação do cliente. Veja-se:

> O campo do *marketing* mudou drasticamente nos últimos anos, particularmente devido à rápida evolução das tecnologias digitais e ao surgimento da Inteligência Artificial (IA). Este artigo se estende ao reino do *marketing* conversacional, uma mudança de paradigma do *marketing* de transmissão tradicional para uma abordagem mais interativa e centrada no cliente. A integração da IA generativa, que redefiniu os limites do engajamento e interação do cliente, é central para essa transformação.
>
> A IA generativa está definida para se tornar uma pedra angular do *marketing* conversacional, fornecendo recursos avançados para simular conversas semelhantes às humanas e gerar conteúdo cada vez mais indistinguível daquele produzido por humanos. Olhando para o futuro, podemos antecipar que a IA generativa criará não apenas texto, mas também conteúdo de voz e vídeo que pode envolver os consumidores interativamente em um diálogo bidirecional, adaptando-se às respostas do consumidor em tempo real e aprendendo com cada interação para refinar sua abordagem de comunicação (Israfilzade; Sadili, 2024, p. 14)[18].

---

[17] Tradução livre do inglês: “AI positively affects consumer experience and consumer engagement on social media. Similarly, a positive relationship exists between social media engagement and consumer experience, leading to a more satisfied consumer and amplified purchase intentions. Additionally, affective attachment moderates the relationship between consumer satisfaction and purchase intention. The results reveal that AI can be used on social media to improve consumer experience and increase customer satisfaction levels and purchase intention”.

[18] Tradução livre do inglês: “The field of marketing has dramatically shifted in recent years, particularly due to the rapid evolution of digital technologies and the emergence of Artificial Intelligence (AI). This paper extends into the realm of conversational marketing, a paradigm shift from traditional broadcast marketing to a more interactive, customer-centric approach. The integration of Generative AI, which has redefined the boundaries of customer engagement and interaction, is central to this transformation.

Tecnologias como a geração de imagens personalizadas auxiliam no aprimoramento do processo da experiência do cliente, fazendo com que os usuários tenham um atendimento muito mais exclusivo. Veja as conclusões de Maura Martins (2024, n.p.)[19] sobre as vantagens de usar a IA nesse tipo de interação:

> Personalização: a inteligência artificial analisa informações individuais de cada consumidor (como seu histórico de compras e navegação) e consegue oferecer interações mais personalizadas;
> Disponibilidade: a tecnologia possibilita que o atendimento inicial ocorra em horário integral, acelerando o processo;
> Velocidade: a IA recebe os estímulos do cliente e os processa de maneira muito ágil, garantindo o atendimento rápido e diminuindo o tempo de espera;
> Aprendizagem permanente: a IA segue aprendendo a cada interação mantida com o cliente, o que melhora a sua eficiência constantemente;
> Menos uso de recursos: a IA generativa ajuda na agilização das operações e na automação de processos, gerando mais tempo e menos custos para a empresa, liberando mais espaço para questões são estratégicas.

Por outro lado, quando uma imagem visa a promover uma relação emocional mais profunda, seja ela pessoal, espiritual ou religiosa, a utilização da IAG pode resultar num afastamento do usuário. Isso se deve ao fato de que a tecnologia atualmente conhecida pelo público não conseguir entregar esse tipo de sintonia. Essa questão é observada no trabalho de Montag *et al.* (2024, p. 165):

> Pessoas com pontuação alta em espiritualidade relatam, entre outros, serem "tocadas pela beleza da criação" e a espiritualidade "sendo uma fonte primária de paz interior" (ver página 165: Montag *et al.* 2021). Ser uma pessoa espiritual, portanto, anda junto com ser mais cético em relação à IA. Esse resultado não foi hipotetizado e, portanto, é apenas um ponto de partida a ser mais explorado. Observe que a descoberta no presente trabalho também não se manteria para ajustes para muitas correlações, como as correções de Bonferroni.

Generative AI is set to become a cornerstone of conversational marketing by providing advanced capabilities to simulate human-like conversations and generate content that is increasingly indistinguishable from that produced by humans. Looking forward, we can anticipate Generative AI to craft not just text but also voice and video content that can interactively engage consumers in a two-way dialogue, adapting to the consumer's responses in real-time and learning from each interaction to refine its communication approach".

[19] Disponível em https://www.tecmundo.com.br/mercado/282248-conversando-robos-ia-ajuda-relacionamento-com-clientes.htm. Acesso em: 8 mai. 2024.

> [...]
> Em suma, o presente trabalho é, até onde sabemos, o primeiro a usar os sistemas emocionais primários de Panksepp e a Teoria da Neurociência Afetiva para lançar luz sobre atitudes em relação à inteligência artificial. Esse trabalho mostra que tal estrutura evolucionária pode ajudar a entender um pouco melhor por que algumas pessoas têm visões mais negativas ou positivas sobre a inteligência artificial. Indo além do nosso trabalho, mais integração da pesquisa neurocientífica poderia estimular ainda mais os desenvolvimentos da IA (Bain e McCay 2023), mas os aspectos éticos devem ser fortemente considerados quando o conhecimento da neurociência e da IA é aplicado, por exemplo, conforme descrito em um artigo recente sobre diagnósticos de reincidência criminal em adolescentes (Munoz e Marinaro 2022)[20].

Esse afastamento emocional diante de imagens artificialmente criadas, especialmente em contextos que envolvem espiritualidade, fé ou vínculos afetivos mais profundos, revela um limite significativo da mediação algorítmica. Segundo Montag *et al*. (2024), embora a IAG seja capaz de simular traços expressivos e ativar reações emocionais básicas, ela ainda carece da densidade semântica e experiencial necessária para evocar emoções autênticas e relacionais complexas. Tais situações envolvem códigos simbólicos que ultrapassam a superfície visual e demandam referências culturais, éticas e subjetivas construídas ao longo da vida. No campo da comunicação, isso significa que, embora a IAG possa operar como mediadora eficiente da informação, seu desempenho como mediadora do afeto permanece limitado. Essa dissonância contribui para a ambivalência na percepção de confiança e credibilidade em conteúdos digitais, especialmente quando os usuários se deparam com produções que pretendem

[20] Tradução livre do inglês: "Persons scoring high on spirituality report among others to be "touched by the beauty of creation" and spirituality "being a primary source of inner peace" (see page 165: Montag et al. 2021). Being a spiritual person thereby goes along with being more skeptical towards AI. This result was not hypothesized and therefore it is just a starting point to be further explored. Please note that the finding in the present work would also not hold for adjustments for many correlations such as Bonferroni corrections.
[...]
In sum, the present work is to our knowledge the first to use Panksepp's primary emotional systems and Affective Neuroscience Theory to shed light on attitudes towards artificial intelligence. This work shows that such an evolutionary framework might help to understand a bit better why some persons have more negative or positive views on artificial intelligence. Going beyond our work, more integration of neuroscientific research could further stimulate AI developments (Bain and McCay 2023), but ethical aspects should be strongly considered when knowledge from neuroscience and AI is applied, e.g. as outlined in a recent paper on diagnostics of criminal recidivism in teenagers (Munoz and Marinaro 2022)".

representar o humano em sua totalidade simbólica, sem, no entanto, alcançar a profundidade emocional que o vínculo humano requer.

Deve-se levar em consideração que as tecnologias que envolvem a IA estão cada vez mais surpreendentes do ponto de vista da aproximação com a realidade, haja vista que a interação com os usuários das redes sociais, por exemplo, está extremamente humanizada. Com isso, emergem questões éticas e de transparência com o uso das inovações. Para o presente trabalho, é relevante destacar que a capacidade desses sistemas de imitar comportamentos humanos pode levar a uma zona cinzenta, as quais os internautas podem não estar cientes de que estão interagindo com uma máquina. Isso coloca em destaque a importância de estratégias de comunicação claras que informem os usuários sobre a natureza das interações com a IA. Pode-se observar essas conclusões no artigo de Israfilzade e Sadili (2024, p. 24) novamente:

> No entanto, também há desafios associados à humanização da IA. Um desafio significativo são as implicações éticas e psicológicas da IA que se assemelha muito ao comportamento humano. Há uma linha tênue entre criar uma IA que seja relacionável e criar uma IA que seja enganosa em seu nível de semelhança humana. Isso levanta questões sobre o consentimento do usuário e a transparência das interações da IA. Outro desafio é a complexidade técnica envolvida na criação de IA antropomórfica sofisticada. O desenvolvimento de sistemas de IA que interpretem e respondam com precisão às emoções e nuances humanas requer tecnologia avançada e refinamento contínuo, o que pode exigir muitos recursos[21].

De acordo com Ebben e Bull (n.d.), "[...] avanços na inteligência artificial generativa que produzem conteúdo cultural cuja veracidade é difícil de determinar"[22] complicam ainda mais a autenticidade nas mídias sociais, levantando questões éticas sobre a manipulação da informação. Essas tecnologias têm a capacidade de produzir conteúdos

[21] Tradução livre do inglês: "However, there are also challenges associated with humanizing AI. One significant challenge is the ethical and psychological implications of AI that too closely resembles human behavior. There is a fine line between creating AI that is relatable and creating AI that is deceptive in its level of human resemblance. This raises questions about user consent and the transparency of AI interactions. Another challenge is the technical complexity involved in creating sophisticated anthropomorphic AI. Developing AI systems that accurately interpret and respond to human emotions and nuances requires advanced technology and ongoing refinement, which can be resource-intensive".

[22] Tradução livre do inglês: "advancements in generative artificial intelligence that produce cultural content whose veracity is difficult to determine".

que podem ser indistinguíveis dos criados por humanos, levantando preocupações sobre a autenticidade e a manipulação da informação. Veja-se:

> A autenticidade nas mídias sociais enfrenta novos desafios. Na esteira dos avanços na inteligência artificial generativa (IA), como *ChatGPT*, *Bard*, *DALI*, *Midjourney*, *Stable Diffusion*, *MusicLM* e outros, as noções sobre autenticidade são ainda mais complicadas. A IA generativa produz conteúdo cultural (por exemplo, imagens, texto, áudio) cuja veracidade e autoria são difíceis de determinar (Ebben; Bull, n.d.)[23].

A percepção de conteúdo gerado por IAG pode afetar significativamente a confiança das pessoas, conforme abordado, portanto a autenticidade é crucial para a confiança na *internet*. Se os usuários identificarem ou suspeitarem que o *output* é elaborado por IAG, sem que isso tenha sido explicitamente informado, isso pode diminuir sua percepção de autenticidade, impactando negativamente a confiança e, por consequência, as decisões de engajamento[24]. Essa são as conclusões de pesquisas de Mulholland e Frajhof (2019, p. 15), bem como de Nicolau (2024), haja vista que a reação emocional das pessoas e a percepção de representações geradas por IAG são influenciadas pela transparência, ética e uso responsável da tecnologia. O texto "Artificial Intelligence – Friend or Foe" sugere que, para maximizar os benefícios e minimizar os riscos, é crucial que as plataformas de redes sociais adotem práticas de IA que priorizem a confiança, autenticidade e segurança dos usuários.

> O problema do Controle de I.A. requer uma solução para evitar conflitos existenciais com a humanidade. Os desenvolvedores e distribuidores de I.A. devem prometer aos Órgãos Reguladores e aos usuários finais que as ferramentas de desenvolvimento de I.A. criadas e implantadas por eles conterão e respeitarão as regras éticas da humanidade, prevenindo a destruição mútua (Nicolau, 2024, p. 35)[25].

---

[23] Tradução livre do inglês: "Authenticity in social media faces new challenges. In the wake of advancements in generative artificial intelligence (AI) such as ChatGPT, Bard, DALI, Midjourney, Stable Diffusion, MusicLM, and others, notions about authenticity are further complicated. Generative AI produces cultural content (e.g., images, text, audio) whose veracity and authorship are difficult to determine".

[24] SCIRE, Sarah. Most readers want publishers to label AI-generated articles — but trust outlets less when they do. Nieman Journalism Lab. 5 dez. 2023. Disponível em: https://www.niemanlab.org/2023/12/most-readers-want-publishers-to-label-ai-generated-articles-but-trust-outlets-less-when-they-do/. Acesso em: 26 abr. 2024.

[25] Tradução livre do inglês: "The A.I. Control problem requires a solution to prevent existential conflicts with humanity. The developers and distributors of A.I. must pledge

Nas pesquisas realizadas por Chang *et al*. (2024), pode-se observar a influência das características de personalidade dos usuários quando são utilizadas pela IAG para personalizar as imagens apresentadas, um fator essencial para melhorar as experiencias das pessoas com as máquinas. Isso é relevante para a presente dissertação, haja vista que ajuda a entender as nuances emocionais que podem afetar a confiança e as decisões de engajamento dos usuários em plataformas *on-line*. *In verbis*:

> Além disso, estudos anteriores sobre *chatbot* para atendimento ao cliente geralmente se concentram na geração de respostas gramaticalmente corretas, ignorando vários fatores que podem potencialmente impactar a experiência do usuário [10]. Para obter uma compreensão mais profunda de como diferentes traços de personalidade influenciam a interação, este experimento manipulará três traços de personalidade distintos (introvertido, ambivertido e extrovertido) para robôs virtuais (telepresentes) e físicos (copresentes). Isso nos ajudará a entender como os traços de personalidade afetam as experiências de interação dos usuários com diferentes robôs, enriquecendo nossa compreensão do comportamento do usuário em diferentes cenários. Em relação à coleta de dados, este estudo usará pesquisas para entender as percepções e preferências subjetivas dos participantes. No entanto, as pesquisas têm limitações em capturar ou explicar os processos de tomada de decisão do consumidor em tempo real. Os participantes podem não expressar com precisão seus pensamentos ou emoções devido a preconceitos pessoais. Portanto, este estudo incorpora medições de EEG para detectar mudanças contínuas na atividade cerebral para avaliar as reações inconscientes e respostas sensoriais dos participantes. A funcionalidade do EEG visa a analisar objetivamente as mudanças nas ondas cerebrais dos participantes durante as interações com diferentes robôs, explorando mudanças emocionais e escolhas de preferência. O objetivo é alinhar e complementar os resultados subjetivos da pesquisa, abordando a lacuna de pesquisa existente entre os padrões de ondas cerebrais e os processos cognitivos específicos ou traços de personalidade (Chang *et al*., 2024, p. 312)[26].

---

to Regulatory Bodies and the end-users that the A.I. development tools created and deployed by them will contain and respect the ethical rules of humanity preventing mutual destruction".

[26] Tradução livre do inglês: "Furthermore, previous studies on chatbot for customer care usually focuses on the generation of grammatically correct responses, overlooking various factors that could potentially impact user experience [10]. To gain a deeper understanding of how different personality traits influence interaction, this

Em sua obra “Marketing 5.0”, Kotler *et al*. (2021) lembram o caso do hotel Henn-na, no Japão, segundo o qual há sérios desafios e limitações da automação extrema. No referido exemplo, os proprietários do hotel utilizaram robôs para diversas funções inicialmente prometendo reduzir custos operacionais, mas acabou provocando frustração entre os hóspedes devido à falta de flexibilidade e compreensão das necessidades humanas. Esse caso ilustra que a tecnologia, por si só, não é suficiente para proporcionar uma experiência satisfatória ao cliente. A interação humana continua sendo essencial, especialmente em setores que exigem um “toque pessoal”.

Um dos fatores que influencia a percepção de confiança em conteúdos desenvolvidos por IAG, especialmente quando vinculados a instituições, é a transparência sobre sua origem e funcionamento. Quando os usuários sabem que um material foi produzido artificialmente, a reação pode variar de desconfiança automática (“Sistema 1” da teoria do processo dual) à racionalização mais deliberada (“Sistema 2” da teoria do processo dual), conforme discutido anteriormente. Nessa toada, a transparência algorítmica desponta como elemento central na mitigação de ambivalência cognitiva. O estudo de Oh *et al*. (2020) sugere que a transparência algorítmica é uma estratégia fundamental para mitigar a desconfiança dos internautas, razão pela qual incluir informações sobre o funcionamento dos algoritmos, bem como permitir a intervenção do usuário na personalização do conteúdo (seja notícia ou imagem), pode fortalecer a credibilidade percebida dos sistemas de informações. Nesse sentido, Silva e Paletta (2025) enfatizam que a ética da informação exige transparência e justificabilidade nos processos automatizados, de forma que os usuários possam compreender e confiar na forma como os dados são processados e utilizados. Segundo os autores, a

---

experiment will manipulate three distinct personality traits (introvert, ambivert, and extrovert) for both virtual (telepresent) and physical (copresent) robots. This will help us understand how personality traits affect users' interaction experiences with different robots, enriching our understanding of user behavior in different scenarios. Regarding data collection, this study will use surveys to understand participants' subjective perceptions and preferences. However, surveys have limitations in capturing or explaining real-time consumer decision-making processes. Participants may not accurately express their thoughts or emotions due to personal biases. Therefore, this study incorporates EEG measurements to detect continuous changes in brain activity to evaluate participants' unconscious reactions and sensory responses. EEG functionality aims to objectively analyze changes in participants' brainwaves during interactions with different robots, exploring emotional changes and preference choices. This aims to align with and complement subjective survey results, addressing the existing research gap between brainwave patterns and specific cognitive processes or personality traits”.

governança da IA deve ser centrada no ser humano e alinhada a princípios como responsabilidade e justiça, garantindo que os algoritmos produzam conteúdo confiável, além de que operem sob regras claras de explicabilidade e supervisão ética (Silva; Paletta, 2025).

Compreender a percepção dos usuários sobre conteúdos criados por IAG exige, inicialmente, uma análise do que já foi produzido pela literatura científica a respeito da dissociação entre atitudes explícitas e implícitas. Antes de abordar diretamente os dados da pesquisa empírica, faz-se necessário examinar os trabalhos acadêmicos que revelam como esses dois níveis de avaliação podem divergir no contexto das interações com sistemas artificiais.

4.1 PESQUISAS JÁ CONHECIDAS ACERCA DA DISSOCIAÇÃO ENTRE ATITUDES EXPLÍCITAS E IMPLÍCITAS DOS USUÁRIOS EM RELAÇÃO À INTELIGÊNCIA ARTIFICIAL

Convém ressaltar algumas pesquisas que já analisaram atitudes explícitas (conscientes e declaradas) e implícitas (automáticas e inconscientes) das pessoas em relação à IA. Fietta *et al.* (2021) constataram uma discrepância significativa entre as atitudes explicitamente expressas pelos participantes e suas respostas implícitas. Os dados do referido estudo indicaram que, enquanto 70,45% dos participantes relataram uma atitude explicitamente positiva em relação à IA, apenas 6,15% demonstraram atitudes implicitamente favoráveis. Por outro lado, 77,1% apresentaram uma atitude implícita negativa, sugerindo que, apesar do discurso positivo sobre IA, há um viés inconsciente contrário à tecnologia. Esse fenômeno pode ser explicado pelo fato de que as atitudes explícitas são influenciadas por normas sociais e expectativas culturais, enquanto as implícitas refletem associações automáticas construídas ao longo da vida (Rydell; Mackie, 2008).

A literatura sobre vieses cognitivos corrobora esse achado, indicando que, frequentemente, as pessoas mantêm crenças inconscientes que contradizem suas declarações conscientes. Produções científicas na área de Psicologia Social demonstram que indivíduos podem expressar apoio a políticas de igualdade racial, por exemplo, ao mesmo tempo em que apresentam vieses implícitos contra determinados grupos (Greenwald *et al.*, 1998). Esse mesmo

mecanismo pode estar operando na aceitação da IA: os participantes reconhecem publicamente os benefícios da tecnologia, mas inconscientemente a associam a riscos e ameaças.

Diferentes modelos psicológicos tentam explicar a dissociação entre atitudes explícitas e implícitas (Fietta *et al*., 2021). O modelo da Avaliação Associativa-Proposicional (APE), proposto por Gawronski e Bodenhausen (2006), sugere que as atitudes explícitas são moldadas por processos racionais e deliberados, enquanto as implícitas são formadas por associações automáticas estabelecidas ao longo da experiência. Assim, a exposição a narrativas de que a IA pode substituir empregos ou comprometer a privacidade pode ter criado associações negativas inconscientes, mesmo entre aqueles que verbalizam opiniões positivas.

Outro fator relevante é o efeito do desconhecido, descrito por Zajonc (1980), que indica que os seres humanos tendem a ter preferências por aquilo que lhes é familiar e relutância em aceitar o que é percebido como novo ou desconhecido. Esse fenômeno foi identificado em trabalhos acadêmicos sobre preconceitos contra minorias sociais e pode ser estendido à IA, uma vez que seu desenvolvimento e adoção em larga escala ainda são recentes.

A análise demográfica realizada por Fietta *et al*. (2021) revelou que fatores como idade, gênero, nível educacional e familiaridade com IA afetam tanto as atitudes explícitas quanto as implícitas. No que tange ao gênero, algumas pesquisas sugerem que mulheres utilizam e confiam mais nas informações de saúde *on-line* do que homens (Cotten; Gupta, 2004). No entanto, outras investigações científicas apontam que não há diferenças significativas entre os gêneros na avaliação da credibilidade da informação (Harris; Sillence; Briggs, 2011). Especificamente quanto à IA, as mulheres demonstraram uma atitude mais negativa, tanto em avaliações explícitas quanto implícitas. Essa discrepância pode estar relacionada a fatores históricos e sociais, como a menor representatividade feminina na área de tecnologia e engenharia (Chien *et al*., 2019). Ademais, pesquisas indicam que mulheres tendem a ser mais céticas em relação a tecnologias que possam comprometer aspectos éticos e sociais, como segurança de dados e tomada de decisão automatizada (Fast; Horvitz, 2017).

Usuários mais jovens (25-55 anos) tendem a confiar mais na informação digital do que idosos, devido à sua maior familiaridade com

a tecnologia (Mcmillan; Macias, 2008). No entanto, idosos que usam a *internet* regularmente desenvolvem maior confiança nos conteúdos *on-line* (Medlock *et al*., 2015). No mesmo sentido, Fietta *et al*. (2021) também identificou que pessoas mais idosas apresentam uma atitude mais negativa em relação à IA. Esses achados são coerentes com pesquisas sobre adoção de novas tecnologias, que mostram que pessoas mais velhas tendem a ser menos receptivas a mudanças tecnológicas, seja por menor exposição ao longo da vida, seja por dificuldades em adaptar-se a novos paradigmas digitais (Rödel *et al*., 2014).

Outro fator determinante foi o nível educacional: indivíduos com maior escolaridade e aqueles que trabalham diretamente com IA demonstraram atitudes mais positivas em relação à tecnologia. Esse resultado pode ser explicado pelo maior conhecimento sobre os benefícios da IA, bem como pela familiaridade com seus mecanismos de funcionamento, reduzindo, assim, o medo do desconhecido (Gille; Jobin; Ienca, 2020).

4.2 TEORIA DO VALE DA ESTRANHEZA

A reação humana às máquinas já vem sendo estudada há algum tempo, e, na década de setenta, o japonês Masahiro Mori (1970) desenvolveu a teoria "vale da estranheza". Tal teoria sugere que, conforme uma máquina se torna mais semelhante ao ser humano, sua aceitação social pode cair abruptamente devido a uma sensação de desconforto ou repulsa (Mori; Macdorman; Kageki, 2012). Nesse contexto, o conceito do "vale da estranheza" (*uncanny valley*) torna-se relevante para analisar reações dos internautas às imagens geradas por IAG. Assim como ocorre com vozes sintéticas, representações visuais produzidas por IAG que se aproximam da aparência humana podem ativar reações ambíguas – oscilando entre aceitação e desconfiança.

A pesquisa de Ciechanowski *et al*. (2018) sugere que a presença do "vale da estranheza" pode levar a uma rejeição emocional de agentes artificiais. Se esse efeito também ocorrer com representações elaboradas por sistemas generativos, pode-se inferir que conteúdos visuais que parecem "quase humanos", mas não totalmente realistas, podem ser percebidos como menos íntegros ou benevolentes, despertando desconfiança.

Outros pesquisadores que apresentam achados relevantes para essa discussão são Lima e Feijó (2019), os quais afirmam que a complexidade da percepção em interações humano-robô dialoga diretamente com a teoria do vale da estranheza, ao evidenciar que a naturalização da IA em ambientes sociais exige mais do que aparência realista: requer comportamento emocionalmente legível. Segundo os autores, o êxito de robôs em contextos sociais depende da capacidade de reconhecer, interpretar e simular emoções humanas, utilizando estratégias de *machine learning* que promovem o ajustamento do comportamento em tempo real. No entanto, quando essa simulação não atinge padrões satisfatórios de coerência semântica e emocional, os usuários tendem a perceber tais agentes como perturbadores ou ambíguos – precisamente o efeito descrito por Mori *et al.* (2012) na formulação do *uncanny valley*.

Nesse sentido, a percepção de estranhamento diante de conteúdos artificiais – incluindo imagens hiper-realistas produzidas por IAG – pode ser intensificada pela ausência de pistas emocionais congruentes com as expectativas humanas. Lima e Feijó (2019) destacam que, para que a IA inspire confiança e empatia, é necessário que sua expressão emocional esteja alinhada a códigos sociais culturalmente assimiláveis, por conseguinte, a ausência desse alinhamento promove um hiato perceptivo que afeta negativamente o julgamento de autenticidade e confiabilidade, mesmo que, visualmente, o conteúdo seja realista. Andrea L. Guzman (2018) aprofunda essa reflexão ao propor que a resposta humana diante de agentes artificiais antropomórficos está relacionada à percepção de fronteiras ontológicas entre humanos e máquinas. Segundo a autora, quando essas fronteiras são transgredidas – por exemplo, quando uma IA assume forma ou comportamento que sugere intencionalidade sem efetivamente possuí-la – ocorre uma espécie de dissonância ontológica: o observador reconhece traços familiares, mas não encontra neles coerência existencial. Tal ruptura entre forma e essência provoca rejeição ou inquietação, fenômeno também descrito por Mori (1970), mas aqui interpretado a partir de uma lente fenomenológica e semântica. Isso reforça a hipótese de que o estranhamento frente a imagens criadas por IAG, sobretudo quando excessivamente realistas, não decorre unicamente da aparência, mas da percepção implícita de que há um "algo" que deveria ter alma, mas não tem.

A pesquisa de Guzman (2018) revela ainda que os limites entre humano e máquina são simbólicos e morais, ou seja, ultrapassam a

tecnicidade ou os aspectos visuais. As pessoas tendem a atribuir humanidade com base em categorias relacionais, como a capacidade de experimentar emoções, exercer julgamento ético ou agir com empatia. No entanto, ao se depararem com agentes artificiais que simulam essas capacidades sem ancoragem subjetiva, emerge a percepção de que há uma quebra no contrato ontológico da interação social. Essa percepção pode afetar diretamente a confiança atribuída aos conteúdos produzidos por IAG, uma vez que, apesar da aparência verossímil, o sujeito que observa permanece ciente, em algum nível, de que não há ali um outro genuinamente presente. Por conseguinte, as imagens podem ser rejeitadas, mesmo que tecnicamente perfeitas, por não cumprirem as expectativas ontológicas que regem as relações interpessoais reais.

No campo da comunicação, a teoria do "vale da estranheza" revela-se significativamente valiosa porque personagens digitais são utilizados como representantes de marcas e produtos. A publicidade tradicional já demonstrou que personagens de marca, como mascotes e avatares, podem aumentar a identificação do consumidor com a marca e melhorar a retenção da mensagem (Aggarwal; Mcgill, 2007). No entanto, quando esses personagens se tornam excessivamente realistas, mas não alcançam a perfeição dos traços humanos, podem entrar no "vale da estranheza" e gerar desconfiança ou rejeição por parte do público (Tinwell, 2014).

Por outro lado, pesquisas um pouco mais recentes desafiam essa teoria do "vale da estranheza". O estudo de Hanson *et al*. (2005) confronta essa visão tradicional ao demonstrar que a aceitação de robôs realistas pode ser maior do que se pensava. A criação de um androide baseado no escritor Philip K. Dick (PKD) trouxe evidências experimentais de que robôs humanizados podem ser bem recebidos, dependendo da qualidade estética e da integração social da máquina. Assim, a noção do "vale da estranheza" pode ser substituída por uma nova abordagem, denominada "caminho do envolvimento" (*Path of Engagement*), que enfatiza o papel do *design* e da interação na aceitação dos robôs.

Hanson *et al*. (2005) questionaram a validade do conceito do "vale da estranheza" ao conduzirem dois experimentos para avaliar a recepção de robôs humanoides pelo público. No primeiro experimento, foram exibidos vídeos de um androide e de outro robô expressivo para um grupo de participantes, que avaliaram suas reações. Os resultados

indicaram que 73% dos entrevistados consideraram os robôs atraentes, 0% os classificaram como perturbadores e 85% afirmaram que os robôs pareciam vivos, contrariando a ideia de que representações robóticas altamente realistas seriam rejeitadas automaticamente pelos humanos. No segundo experimento, os pesquisadores apresentaram uma sequência de imagens que variavam de representações caricaturais a realistas, pedindo aos participantes que avaliassem a aceitabilidade de cada uma. Os resultados mostraram que não houve uma queda significativa na aceitação conforme as imagens se tornavam mais realistas, o que sugere que a rejeição a robôs hiper-realistas pode não ser uma regra universal (Hanson *et al*., 2005).

Esses achados desafiam a hipótese de que robôs realistas são necessariamente inquietantes e indicam que a recepção de tais tecnologias pode depender de outros fatores, como o contexto da interação e a qualidade da animação. Hanson *et al*. (2005) propõem que a estética e a naturalidade dos movimentos podem influenciar positivamente a aceitação de robôs humanoides, sugerindo uma reformulação da teoria do "vale da estranheza" para considerar uma "trajetória de engajamento" em vez de uma queda abrupta na aceitação conforme o realismo aumenta.

Trazendo a pesquisa de Hanson *et al*. (2005) para os escopos específicos desta dissertação, pode-se compreender que as reações dos usuários a representações visuais criadas por IAG, especialmente no que tange à percepção de confiabilidade, revela a possibilidade de que elas não sejam necessariamente percebidas como menos confiáveis do que imagens reais, desde que apresentem coerência visual e correspondam às expectativas do público. Esse resultado é particularmente relevante, pois sugere que a aceitação de conteúdos visuais criados por IAG pode depender menos da origem artificial e mais de fatores como familiaridade estética, naturalidade das expressões e contexto da apresentação. Dessa forma, a investigação das associações implícitas e explícitas dos usuários em relação à credibilidade dessas imagens pode contribuir para uma compreensão mais aprofundada dos mecanismos cognitivos que influenciam a confiança na era digital.

A partir da revisão dos principais estudos sobre atitudes explícitas e implícitas e da compreensão dos fatores emocionais e cognitivos que moldam a percepção dos usuários frente a conteúdos gerados por IAG, torna-se possível avançar para a análise dos dados

empíricos desta pesquisa. A próxima seção detalha o delineamento experimental, justificando a escolha dos instrumentos (IAT e EAAT), a construção dos estímulos visuais, os critérios de seleção dos participantes e os cuidados éticos envolvidos na coleta de dados, permitindo a investigação sistemática das associações entre confiança e imagens produzidas por IAG.


5

5 METODOLOGIA

A pesquisa seguiu uma abordagem metodológica quantitativa, por meio, principalmente, da aplicação do IAT, uma ferramenta amplamente validada para medir associações implícitas entre conceitos. Esse método permite avaliar como os participantes associam conteúdos gerados por IAG e conteúdos humanos a atributos como competência, integridade e benevolência, conforme o modelo de Mayer, Davis e Schoorman (1995).

Trata-se de uma abordagem nomotética, pois visa a identificar padrões gerais de percepção e confiança em *outputs* desenvolvidos por IAG, a partir da aplicação de instrumentos padronizados como o IAT e o EAAT. A utilização de técnicas quantitativas e a análise de reações implícitas e explícitas têm por objetivo construir inferências que possam ser generalizadas para populações mais amplas, em consonância com o paradigma explicativo das ciências sociais empíricas.

Do ponto de vista praxiológico, este estudo adota uma angulação crítico-analítica, voltada à compreensão reflexiva das dinâmicas de confiança e julgamento associadas a conteúdos produzidos por IAG. A investigação buscou explicitar os mecanismos cognitivos e sociotécnicos que moldam a percepção dos usuários, contribuindo para o debate epistemológico e ético sobre o uso da IAG em ambientes digitais.

Na presente pesquisa, foi aplicado o Teste de Associação Implícita (IAT) junto a participantes usuários da internet em geral, com idades entre 18 e 75 anos. O teste consistiu em tarefas de categorização rápida, nas quais os respondentes deveriam associar imagens de profissionais do sistema de justiça – reais ou geradas por IAG – a atributos relacionados à confiança. A confiabilidade, nesse contexto, é entendida como composta pelas dimensões de competência, integridade e benevolência (Mayer; Davis; Schoorman, 1995). O objetivo foi verificar se as imagens produzidas por IA generativa evocam padrões diferenciados de associação de confiança em comparação às imagens reais, permitindo identificar vieses implícitos que operam de forma automática e não consciente no julgamento dos participantes.

O Teste de Associação Implícita – em inglês, *Implicit Association Test*, o que originou a forma abreviada IAT, amplamente referenciada – , proposto por Greenwald, McGhee e Schwartz (1998), emergiu como uma ferramenta inovadora para estudar associações cognitivas inconscientes, uma vez que ele mede a força de vínculos associativos entre conceitos e representações cognitivas, sendo amplamente utilizado em diversos campos, como psicologia social, clínica e organizacional. Essa metodologia se destaca por acessar atitudes, crenças e estereótipos que, muitas vezes, escapam à introspecção consciente. O IAT é um paradigma experimental que mede a força das associações entre conceitos, utilizando tempos de reação (TR) como indicador principal. A premissa básica é que indivíduos processam mais rapidamente informações congruentes, ou seja, aquelas que estão associadas em suas redes cognitivas (Greenwald; Mcghee; Schwartz, 1998).

O IAT baseia-se em uma tarefa de categorização dupla, segundo o qual o tempo de reação é usado como indicador da força das associações mentais entre conceitos. Greenwald, McGhee e Schwartz (1998) demonstraram que combinações consistentes com associações automáticas geram respostas mais rápidas, ao passo que combinações inconsistentes exigem maior esforço cognitivo e tempos de reação mais longos.

Essa abordagem oferece vantagens únicas: flexibilidade, já que pode ser adaptado para medir estereótipos de gênero, preconceitos raciais, atitudes políticas, entre outros (Nosek *et al.*, 2011); aplicação ampla, sendo utilizado em áreas como psicologia clínica, desenvolvimento infantil e neurociência (Richeson *et al.*, 2003); previsibilidade, havendo estudos que mostram que o IAT pode prever comportamentos discriminatórios e preferências intergrupais que medidas explícitas não conseguem captar (Greenwald *et al.*, 2003).

Acerca do teste, os autores esclarecem que "O procedimento IAT busca mensurar atitudes implícitas medindo sua avaliação automática subjacente. O IAT é, portanto, similar em intenção aos procedimentos de preparação cognitiva para medir afeto ou atitude automáticos" (Greenwald *et al.*, 2003, p. 1464)[27]. Em nota, acrescentam que "O IAT também pode ser usado para medir estereótipos implícitos e

[27] Tradução livre do original: "The IAT procedure seeks to measure implicit attitudes by measuring their underlying automatic evaluation. The IAT is therefore similar in intent to cognitive priming procedures for measuring automatic affect or atitude".

autoconceito implícito por meio da seleção apropriada do conceito-alvo e discriminações de atributos" (Greenwald *et al.*, 2003, p. 1466)[28] .

Nesse modelo, a memória é construída como uma rede metafórica, na qual conceitos são interconectados entre si e, quanto mais congruentes, mais próximos eles se encontram na rede. Quando se pensa na marca *Apple*, por exemplo, facilmente são lembrados produtos e pessoas, tais como: *iPhone*, *iPad*, Steve Jobs etc. A ativação desses conceitos, que estão inter-relacionados mentalmente, ocorre de maneira automática e subconsciente (Psychology, 2020). De acordo com Zaltman (2003), a mente humana opera por meio de metáforas e associações implícitas que influenciam profundamente a forma como se percebem marcas e produtos. Em que pese Zaltman (2003) ter aplicado seus estudos ao "consumo", ele mesmo reconhece que esse mecanismo vale para qualquer forma de julgamento social e simbólico – inclusive instituições, profissões e figuras de autoridade. Transportando essa lógica à IAG, a forma como os usuários associam *outputs* gerados por máquinas pode depender das metáforas que possuem sobre tecnologia e confiança.

O IAT é uma ferramenta valiosa para complementar as metodologias tradicionais de medição da confiança na tecnologia, pois permite detectar vieses implícitos que os usuários podem não estar conscientes ou dispostos a relatar. Isso decorre do fato de que os participantes não têm tempo para refletir sobre suas respostas, por conseguinte o teste capta associações inconscientes que podem não ser reveladas em questionários tradicionais. A confiança eletrônica frequentemente depende de fatores não totalmente racionais ou conscientes, como a percepção intuitiva de segurança em um *site*, a familiaridade com a marca ou o efeito de avaliações de terceiros (Mcknight; Chervany, 2001). Assim, o tempo de resposta é analisado estatisticamente, permitindo comparações sistemáticas entre diferentes tecnologias e populações de internautas.

Neste trabalho, utilizou-se a plataforma de Teste Implícito de Associação da empresa *Millisecond Software*, amplamente reconhecida por desenvolver e validar versões robustas do IAT (*Implicit Association Test*) por meio do seu *software Inquisit*, que permite a implementação precisa de testes baseados em tempo de reação. A

[28] Tradução livre do original: "The IAT can be used also to measure implicit stereotypes and implicit self-concept (see Greenwald & Banaji, 1995) by appropriate selection of target concept and attribute discriminations".

Millisecond é utilizada em pesquisas acadêmicas e institucionais de alta credibilidade e oferece suporte à replicação dos paradigmas propostos por Greenwald, McGhee e Schwartz (1998), garantindo a confiabilidade psicométrica dos resultados obtidos.

O IAT foi realizado *on-line* pelos participantes, utilizando-se de um *software* especializado acima descrito, para registrar o tempo de reação dos participantes em milissegundos. Em linhas gerais, convém esclarecer que os participantes foram apresentados a imagens, divididos em duas categorias: conteúdo criado por IAG e conteúdo real.

Esses conteúdos foram intercalados com palavras com categorizações positivas ou negativas. Essas categorias estão relacionadas a atributos específicos, quais sejam: "competente", "qualificado", "capaz", "confiável", "seguro", "honesto", "simpático", "incompetente", "desqualificado", "incapaz", "injusto", "duvidoso", "inseguro", "desonesto" e "antipático", com objetivo de medir a rapidez com que os participantes associam esses atributos às imagens ou textos apresentados. A escolha dos atributos utilizados nos estímulos do IAT foi baseada no modelo de confiança proposto por Mayer, Davis e Schoorman (1995), que define a confiança como a disposição de um indivíduo em ser vulnerável às ações de outro, com base na expectativa de que esse outro agirá de forma competente, íntegra e benevolente. A partir dessas três dimensões centrais – competência, integridade e benevolência –, foram selecionadas palavras que representam semanticamente esses atributos, tanto em seu aspecto positivo quanto negativo. Essa categorização visa a captar as reações implícitas dos participantes diante de imagens que evocam julgamentos sobre esses elementos fundamentais da confiança.

O processo de seleção e criação das imagens utilizadas como estímulos no IAT foi conduzido de forma sistemática, de modo a garantir comparabilidade visual entre os conjuntos de imagens reais e artificiais. Inicialmente, foram selecionadas imagens reais de profissionais e ambientes do sistema de justiça. As imagens criadas por IAG e utilizadas no teste foram obtidas de duas formas: algumas selecionadas em banco de imagens de acesso livre (FREEPIK, 2025), e outras geradas artificialmente por meio de plataforma de IA (SHAKKER AI, 2025). A escolha desse contexto jurídico decorre do objetivo central da pesquisa, que busca investigar a percepção de confiança em representações associadas ao sistema de justiça brasileiro.

Para assegurar equivalência perceptiva entre os estímulos, foram controladas variáveis visuais como enquadramento, composição, iluminação, ângulo da câmera, posição corporal, vestimenta formal, inclusão de togas, brasões, martelos, ambientes de tribunal e presença de policiais fardados. Esse procedimento visou evitar que diferenças irrelevantes ao objetivo da pesquisa (por exemplo, qualidade estética ou estilo visual) influenciassem as respostas dos participantes, de modo que as associações implícitas fossem atribuídas preferencialmente à distinção entre origem real ou artificial, e não a características visuais secundárias. Assim, cada imagem artificial foi construída para apresentar estrutura, tema e elementos iconográficos semelhantes às imagens reais correspondentes, permitindo comparar grupos de estímulos com alto grau de paralelismo visual.

Esse processo de pareamento entre estímulos reais e artificiais buscou aumentar o rigor experimental e reduzir vieses decorrentes de pistas visuais não controladas, assegurando que as diferenças observadas nas medidas implícitas e explícitas refletissem predominantemente a percepção da artificialidade e não variações estéticas independentes.

Apesar de sua aplicação amplamente validada pelo meio acadêmico, o IAT não está isento de críticas e limitações. Uma das principais questões está relacionada à interpretação dos efeitos, haja vista que ele mede associações relativas entre categorias, e não atitudes absolutas. Ou seja, os resultados indicam se o participante associa com mais rapidez determinado conceito (como “confiança” ou “competência”) a imagens de IA ou humanas, mas não quantificam o nível de confiança individual em cada tipo de imagem de forma isolada. Essa característica exige uma interpretação cuidadosa, centrada na comparação entre categorias, e não na avaliação direta de uma atitude específica. Cita-se, a título de exemplo, um escore positivo em um IAT de raça, que pode indicar associações mais fortes entre “branco e positivo” do que entre “negro e positivo”, mas isso não implica necessariamente preconceito explícito ou hostilidade em relação a pessoas negras (Nosek; Banaji; Greenwald, 2007).

Fiedler, Messner e Bluemke (2006) alertaram que a probabilidade de um falso positivo é elevada, ou seja, um escore elevado em um teste individual pode não refletir uma atitude real, mas, sim, variações contextuais ou estratégicas, por conseguinte, pode-se

dizer que o IAT é mais adequado para análises em nível grupal do que para diagnósticos individuais.

Para complementar a análise dos resultados implícitos, realizou-se, também, o *Explicit Attribute Assignment Test* (EAAT), da *Millisecond Software*, aqui referido apenas como Teste de Associação Explícita ou EAAT. No caso desta pesquisa, as respostas são dadas por meio de uma escala de avaliação – como "muito confiável", "confiável", "neutro", "pouco confiável" e "nada confiável" – o que impede a neutralidade total e força um julgamento consciente (Millisecond Software, 2023). Assim, enquanto o IAT mede associações inconscientes entre estímulos e atributos, o EAAT capta respostas explícitas dos participantes, permitindo uma análise comparativa entre percepções automáticas e conscientes sobre confiança em *outputs* gerados por IAG.

Além de sua aplicabilidade prática, o EAAT apresenta fundamentos teóricos consistentes no campo da psicologia social e cognitiva, especialmente no que tange ao estudo das atitudes proposicionais. De acordo com De Houwer, Gawronski e Barnes-Holmes (2013), as atitudes explícitas derivam de processos proposicionais que envolvem validação consciente de informações, por conseguinte, julgamentos que os indivíduos conseguem acessar, verbalizar e justificar. Ademais, diferentemente das atitudes implícitas, que operam de forma automática, as atitudes explícitas são moldadas por normas sociais, valores conscientes e contextos discursivos. Nesse sentido, o EAAT permite acessar esses julgamentos racionais e socialmente mediados, oferecendo uma importante métrica para análise de conteúdos cuja confiabilidade pode estar sujeita à reflexão crítica dos participantes – como ocorre frequentemente em *outputs* digitais desenvolvidos por IAG.

O uso do EAAT destaca-se por sua capacidade de capturar respostas conscientes padronizadas por meio de escolhas dicotômicas, normalmente estruturadas em torno de escalas binárias (sim/não, confiável/não confiável), o que facilita a quantificação e a análise estatística dos dados. Essa técnica, também conhecida como *forced-choice format*, é utilizada em estudos sobre percepção social e julgamento moral, justamente por reduzir a ambiguidade na resposta e forçar o participante a tomar uma posição definida (Paulhus, 1991). Em ambientes digitais, em que a ambivalência e a desinformação são frequentes, a clareza exigida pelo EAAT pode revelar como fatores tais quais reputação da fonte, aparência visual ou textual e

transparência da origem influenciam diretamente os julgamentos conscientes de credibilidade. No caso específico deste estudo, a aplicação do EAAT contribuiu para verificar se, uma vez informados de que determinada imagem foi produzida por IAG, os participantes mantinham ou rejeitavam sua confiança no conteúdo – um dado crucial para compreender os efeitos cognitivos da rotulagem de conteúdo artificial.

Além disso, a triangulação entre medidas implícitas (via IAT) e explícitas (via EAAT) segue recomendações metodológicas clássicas para estudos sobre atitudes, conforme discutido por Nosek e Smyth (2007). Os autores afirmam que a comparação entre esses dois níveis de resposta permite identificar possíveis dissociações entre cognição automática e avaliação deliberada, fenômeno comum em situações que envolvem julgamentos éticos, preconceitos, ou, como neste caso, formação de confiança. Por meio do EAAT, foi possível observar como os participantes racionalizam seus julgamentos sobre confiança ao interagir com imagens rotuladas como artificiais, mesmo quando essas imagens ativam positivamente suas associações implícitas. A presença ou ausência dessa coerência entre atitudes implícitas e explícitas contribui para a compreensão dos mecanismos de decisão e percepção de risco no uso de *outputs* digitais mediados por IAG.

5.1 PROCEDIMENTOS METODOLÓGICOS DA PESQUISA EXPERIMENTAL

A metodologia adotada neste estudo é embasada em um público-alvo previamente estabelecido, qual seja: usuários da *internet* em geral, seja em redes sociais, seja em *sites* diversos, com idades entre 18 e 75 anos. Não se exigiu um perfil específico em termos de ocupação ou interesses, mas optou-se por dar preferência aos estudantes da área de Comunicação Social, Ciência da Computação e Direito, bem como profissionais da área jurídica. Ademais, entendeu-se imprescindível que os participantes fossem usuários regulares de *internet*, utilizando-a tanto para lazer quanto para consumo de produtos e serviços. O recrutamento de participantes e os testes performados foram realizados de modo *on-line*, via *instagram* e grupos de *Whatsapp* do pesquisador, bem como presencial para alunos de graduação e pós-graduação. A combinação desses canais de recrutamento resultou na obtenção da amostra final de participantes que concluíram integralmente a pesquisa demográfica e os dois testes (IAT e o EAAT).

Embora o tamanho amostral assegure poder estatístico adequado para os contrastes planejados, trata-se de uma amostra de conveniência, obtida por convites em redes sociais e grupos específicos. Assim, as conclusões apresentadas nesta dissertação referem-se apenas ao grupo estudado e não podem ser generalizadas para a população em sentido amplo, devendo ser interpretadas com a devida cautela.

Para determinar o tamanho adequado da amostra para o IAT utilizado nesta pesquisa, considerou-se a necessidade de alcançar poder estatístico suficiente para detectar efeitos moderados, característicos em estudos sobre atitudes implícitas (Greenwald *et al.*, 1998). Seguindo as recomendações de Fulcher, Dean e Trufil (2016), adotaram-se como parâmetros de cálculo: tamanho de efeito (d = 0,5), nível de significância ($\alpha = 0,05$), poder estatístico ($1-\beta = 0,80$) e correlação entre medidas repetidas (r = 0,5). Esses valores considerariam e validariam uma amostra mínima de 34 (trinta e quatro) participantes. Entretanto, a literatura sugere a ampliação dessa amostra, de modo a compensar possíveis exclusões decorrentes de respostas extremas ou inválidas, sendo recomendável a inclusão de 50 (cinquenta) a 100 (cem) participantes em estudos experimentais dessa natureza (Fulcher; Dean; Trufil, 2016).

A pesquisa utilizou amostra de conveniência e foram definidos critérios de inclusão, a saber: (a) idade mínima de 18 anos; (b) proficiência em língua portuguesa; (c) acesso estável a dispositivo compatível (computador ou *smartphone*); (d) aceite do Termo de Consentimento Livre e Esclarecido (TCLE); e (e) cumprimento integral do protocolo experimental, incluindo os blocos do Teste de Associação Implícita (IAT) e todos os itens do Teste de Atribuição Explícita de Atributos (EAAT).

Os critérios de exclusão contemplaram: (a) duplicidade de respostas identificada por metadados técnicos; (b) taxa de erro elevada no IAT (superior a 20%); (c) presença de latências inválidas ou extremas (por exemplo, mais de 10% dos ensaios com tempo inferior a 300 ms, ou latências superiores a 10.000 ms); e (d) abandono do experimento antes da etapa de *debriefing*. Os critérios de exclusão referentes à qualidade dos dados (como taxa de erro superior a 20% no IAT e presença de latências inválidas ou extremas – mais de 10% dos ensaios abaixo de 300 ms ou respostas acima de 10.000 ms) foram aplicados automaticamente pela plataforma *Inquisit*, conforme parâmetros previamente definidos pelo *software*. Assim, a depuração da amostra

não foi realizada manualmente pelo pesquisador, mas seguiu os procedimentos técnicos padronizados pela ferramenta, garantindo uniformidade na exclusão de registros inconsistentes.

Durante a fase de coleta de dados, uma das principais dificuldades enfrentadas foi a obtenção de um número significativo de voluntários dispostos a participar dos testes experimentais. Apesar dos testes estarem hospedados em plataformas seguras e especializadas, como o *Inquisit Web* da *Millisecond Software*, houve resistência perceptível por parte de potenciais respondentes ao receberem os convites para participação por meio de *links* eletrônicos. Esse fenômeno pode ser atribuído a uma crescente desconfiança digital no ambiente *online*, especialmente diante do aumento de fraudes virtuais, golpes de *phishing* e infecções por *malware*, que levam os usuários a evitarem clicar em *links* desconhecidos ou não verificados, mesmo quando compartilhados por instituições, pesquisadores ou contatos de confiança.

Essa barreira limitou a velocidade da coleta de dados e refletiu a composição e a diversidade da amostra, uma vez que muitas pessoas demonstraram receio em acessar a pesquisa mesmo após esclarecimentos sobre a natureza acadêmica do estudo e a proteção de dados pessoais envolvidos. Em algumas situações, foi necessário fornecer explicações adicionais sobre a legitimidade da ferramenta utilizada, bem como o respaldo institucional do projeto. Esse cenário evidencia um desafio metodológico contemporâneo para pesquisadores que empregam testes *on-line* baseados em reações rápidas e coleta de dados psicométricos: a erosão da confiança digital, que afeta diretamente a adesão voluntária a experimentos científicos e pode se tornar uma variável crítica nos estudos que utilizam tecnologia na interface entre cognição e comportamento.

Além da resistência inicial em acessar os *links* da pesquisa por desconfiança digital, alguns participantes que aceitaram contribuir relataram dificuldades técnicas durante a execução do experimento, especialmente no uso do aplicativo *web* do *Inquisit*. Foram mencionados casos em que o sistema travava, no entanto, não foi identificado exatamente em qual circunstância ocorria a falha, se era em razão do sistema operacional do computador do participante, insuficiência de memória RAM, instabilidade da conexão de *internet*, uso de dispositivos móveis ou outro fator técnico específico. Apesar disso, é importante registrar essa limitação como um fator dificultador

na obtenção dos dados. Esses problemas, embora pontuais, não afetaram de forma significativa a pesquisa, entretanto diminuíram a amostra, uma vez que a experiência do usuário é fator determinante para a continuidade em testes digitais que exigem atenção e precisão.

O procedimento do IAT foi introduzido aos participantes da seguinte maneira: eles receberam uma explicação geral sobre o teste, sem detalhes que pudessem influenciar suas respostas. Após isso, foram expostos aos estímulos: imagens e palavras (atributos) foram exibidas em sequência, com instruções para classificar rapidamente os pares de conteúdos e atributos apresentados. As respostas foram registradas por meio de teclas ou cliques, dependendo do dispositivo utilizado.

Em seguida, houve a interpretação dos resultados: o tempo de reação foi considerado para identificar diferenças significativas entre as associações feitas com representações visuais produzidas por IAG e humanos.

O uso do IAT é particularmente relevante porque permite identificar associações implícitas que os participantes podem não conseguir expressar conscientemente. Estudos anteriores, como os de Wąsikowska (2021), indicam que métodos indiretos, como o IAT, revelam aspectos subliminares das emoções e percepções, muitas vezes divergentes das respostas conscientes obtidas por entrevistas ou questionários.

Embora o IAT seja uma ferramenta robusta, é importante considerar que ele mede apenas associações implícitas, não capturando integralmente as percepções conscientes ou as reações emocionais espontâneas dos participantes. A aplicação do IAT neste estudo oferece uma oportunidade única de explorar associações implícitas entre *outputs* gerados por IAG e atributos críticos da confiança, frequentemente confundidos com a noção de credibilidade, mas aqui delimitados nos termos do modelo de Roger Mayer *et al.* (1995), como competência, integridade e benevolência, contribuindo para um entendimento mais profundo sobre os efeitos desses conteúdos no comportamento dos usuários.

Já em relação ao EAAT, a estrutura foi composta por estímulos textuais, segmentados em duas categorias principais: (a) conteúdos gerados por IA e (b) conteúdos gerados por humanos. Cada estímulo foi acompanhado por uma afirmação, à qual o participante deveria

responder em uma escala de avaliação explícita, variando de “Muito confiável” a “Nada confiável”. O questionário utilizado no EAAT buscou avaliar a percepção consciente dos usuários sobre confiança de representações visuais produzidas por IAG, bem como sua susceptibilidade à manipulação.

As frases foram elaboradas a partir de três dimensões centrais da pesquisa: confiança e manipulação percebida. Para avaliar a confiança, foram propostas perguntas como: “O quanto você confia em imagens geradas por IA?” e “O quanto você confia em imagens reais?”, utilizando uma escala de cinco pontos, variando de “Muito confiável” a “Nada confiável”.

Por fim, para investigar a percepção de manipulação, os participantes responderam a sentenças como “O quanto você acredita que imagens geradas por IA podem ser manipuladas?” e “O quanto você acredita que imagens reais podem ser manipuladas?”. Dessa forma, o EAAT complementou a análise do IAT, permitindo uma comparação entre as percepções automáticas dos participantes e seus julgamentos explícitos sobre confiabilidade dos *outputs*.

A aplicação dos testes seguiu um protocolo estruturado. Inicialmente, os participantes visualizavam um estímulo (imagem ou texto). Em seguida, deveriam selecionar uma das opções disponíveis (“positivo” ou “negativo; “real” ou “IA”; etc.), avançando para o próximo estímulo. A testagem seguiu até que todas as afirmações fossem respondidas para os diferentes tipos de conteúdo.

A análise dos resultados foi realizada a partir da frequência das respostas para cada categoria de conteúdo (IA ou humano). Além disso, os resultados do EAAT foram comparados aos do IAT, permitindo a identificação de discrepâncias entre percepções implícitas e explícitas. Adicionalmente, buscou-se verificar se a transparência na informação sobre a origem do conteúdo influenciava as respostas dos participantes.

O EAAT foi operacionalizado por meio da plataforma da *Millisecond Software*, garantindo a geração automatizada de dados e permitindo a análise quantitativa das respostas. Os resultados extraídos do EAAT complementaram a compreensão das associações inconscientes medidas pelo IAT, fornecendo uma visão mais ampla sobre a forma como os usuários interpretam e reagem a conteúdos gerados por IAG.

Os dados coletados foram analisados por meio de procedimentos de estatística descritiva e inferencial. Foi realizada análise exploratória de correlação entre as medidas implícitas (IAT) e explícitas (EAAT), com o objetivo de investigar a relação entre julgamentos automáticos e conscientes. Em todas as análises, adotou-se nível de significância de 5% ($p < 0,05$).

A pesquisa foi conduzida em conformidade com os princípios éticos aplicáveis às Ciências Humanas e Sociais, observando integralmente as diretrizes da Resolução CNS nº 510/2016, a qual estabelece os marcos normativos para pesquisas envolvendo seres humanos nesse campo. Os procedimentos adotados garantiram a proteção dos direitos, do bem-estar e da privacidade dos participantes, assegurando anonimato absoluto, ausência de coleta de dados sensíveis ou identificadores pessoais e liberdade para desistência a qualquer momento.

À luz do art. 1º, incisos I e VII, e do art. 2º, parágrafo único, da Resolução 510/2016, o estudo enquadra-se entre as modalidades de pesquisa isentas de apreciação por Comitê de Ética em Pesquisa (CEP), por envolver exclusivamente aplicação de instrumentos anônimos, sem riscos previsíveis e sem qualquer intervenção sobre os participantes. Ainda assim, todo o processo investigativo observou rigor ético compatível com as boas práticas científicas e com a proteção dos participantes.

5.2 ILUSTRAÇÃO DE COMO FOI O IAT E O EAAT

Para fins ilustrativos, apresentam-se a seguir algumas telas exemplares dos experimentos conduzidos (IAT e EAAT), com o intuito de demonstrar a estrutura básica e a lógica de aplicação dos testes. Optou-se pela apresentação de 7 (sete) imagens representativas, de modo a preservar o fluxo da leitura. As demais telas que compõem integralmente os blocos dos experimentos encontram-se disponíveis no Apêndice A, garantindo transparência metodológica e plena reprodutibilidade da pesquisa. Ambas as ferramentas foram desenvolvidas com base nos princípios estabelecidos pelo *Project Implicit*, da Universidade de *Harvard*, cuja plataforma disponibiliza testes validados para uso em ambientes acadêmicos e educacionais (*Project Implicit*, 2025).

O *Project Implicit* foi fundado em 1998 pelos pesquisadores Tony Greenwald (*University of Washington*), Mahzarin Banaji (*Harvard University*) e Brian Nosek (*University of Virginia*). A iniciativa tem como missão divulgar o conhecimento científico sobre os vieses implícitos e desenvolver ferramentas que permitam medi-los de forma rigorosa e replicável. Para isso, criou-se um "laboratório virtual" que aplica testes comportamentais baseados em tempo de reação e escolhas proposicionais conscientes, possibilitando identificar dissociações entre atitudes automáticas e deliberadas (*Project Implicit*, 2025).

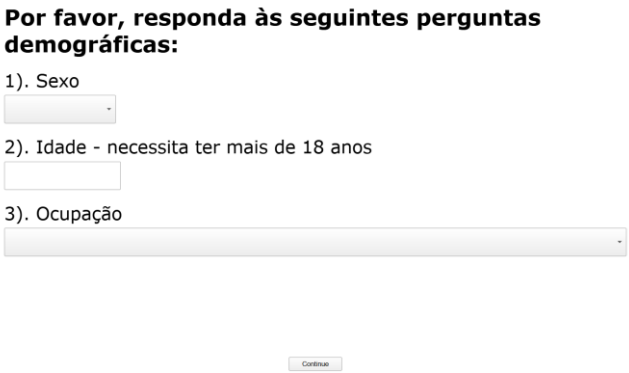
A construção dos *slides* utilizados nesta pesquisa foi guiada diretamente pela estrutura proposta nos testes disponíveis no *site* oficial da plataforma (https://implicit.harvard.edu). Foram mantidos integralmente os critérios já validados pela Universidade de *Harvard* como o número de imagens e de atributos avaliados, bem como a ordem dos blocos de apresentação e as instruções fornecidas ao início de cada teste, de modo a preservar a confiabilidade psicométrica do modelo original.

Na sequência, são apresentadas miniaturas dos *slides* (Figuras 1 a 7) que compõem a aplicação de ambos os testes. Essa etapa visa a dar transparência ao processo experimental e destacar a coerência metodológica entre a proposta teórica e a execução prática da pesquisa.

Figura 1 – Slide 1

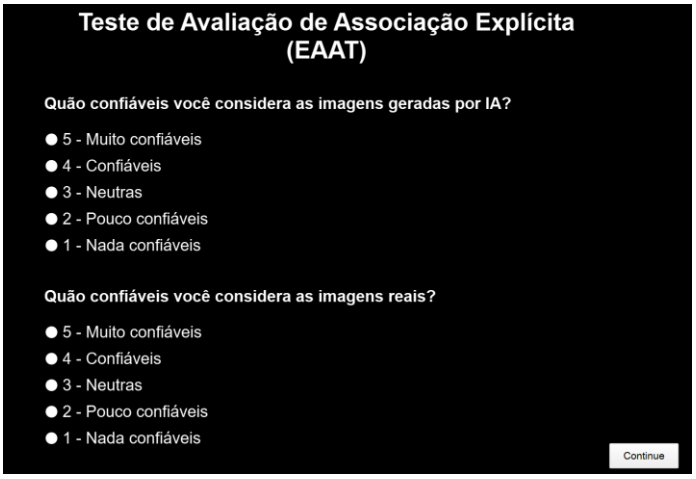
Fonte: Elaborado pelo pesquisador (2025).

Figura 2 – Slide 2



Fonte: Elaborado pelo pesquisador (2025).

Figura 3 – Slide 4

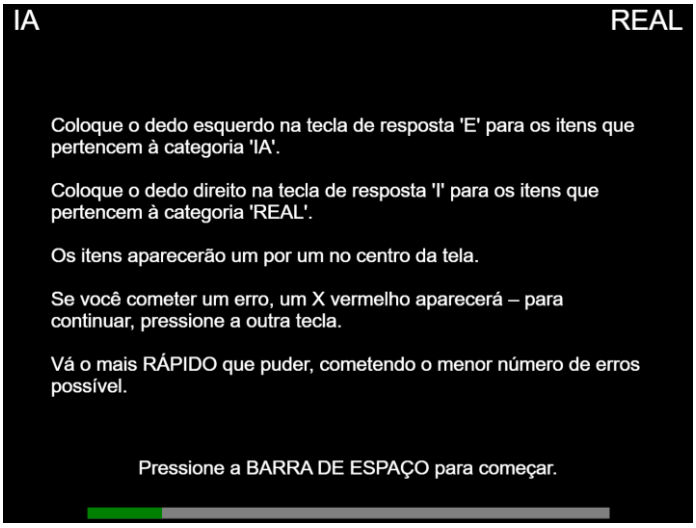


Fonte: Elaborado pelo pesquisador (2025).

Figura 4 – Slide 6

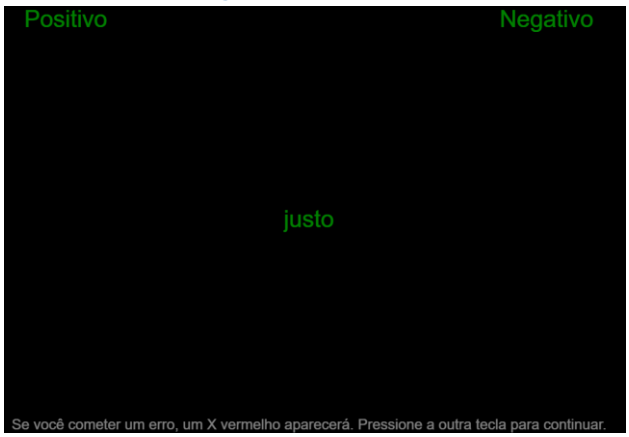
Fonte: Elaborado pelo pesquisador (2025).

Figura 5 – Slide 7



Fonte: Elaborado pelo pesquisador (2025).

Figura 6 – Slide 28



Fonte: Elaborado pelo pesquisador (2025).

Mais 15 (quinze) *slides* de atributos positivos e negativos são apresentados sequencialmente no teste.

Figura 7 – Slide 44

Fonte: Elaborado pelo pesquisador (2025).

Importa ressaltar que, embora os participantes sejam informados antecipadamente sobre as categorias (imagens IA *versus* reais) e os atributos positivos e negativos a serem avaliados, esse procedimento segue o protocolo padrão do IAT e não compromete a mensuração implícita das associações cognitivas. Estudos demonstram que o IAT permanece sensível à força das associações automáticas mesmo quando os rótulos das categorias são conhecidos pelo participante, pois o núcleo da tarefa é o tempo de reação em associações congruentes *versus* incongruentes, e não o simples reconhecimento das categorias (Greenwald *et al.*, 1998). Além disso, utilizou-se o contrabalanceamento da ordem dos blocos e das teclas de resposta para reduzir efeitos de aprendizagem ou estratégias conscientes, conforme recomendado por Nosek, Banaji e Greenwald (2007). Assim, mesmo que a tarefa pareça "fácil" em termos de categorização, ela exige rapidez e automatismo que tornam improvável a aplicação de estratégias deliberadas capazes de anular as diferenças de tempo de resposta inerentes às associações implícitas. Como demonstrado, o teste se mostra uma importante ferramenta para produção de dados deste trabalho e a análise dos dados gerados representa um contributo ao meio acadêmico.

5.3 PROCEDIMENTOS E CONSIDERAÇÕES ÉTICAS

No ambiente digital, a ética é um pilar indispensável para estabelecer e manter a confiança no sistema como um todo, especialmente quando se trata de inteligência artificial. A ética na IA pode ser entendida como um campo da ética aplicada que busca examinar os efeitos morais do desenvolvimento e da implementação de tecnologias baseadas em IA (Lamb, 2024). A literatura filosófica, como destacam Bostrom e Yudkowsky (2014), argumenta que tecnologias disruptivas, como a IA, amplificam dilemas éticos clássicos.

Imperioso trazer à baila dois desses dilemas clássicos para enriquecer este trabalho, quais sejam: "Leviatã", de Thomas Hobbes (1979)[29], e princípios relativos à "regra de ouro" e à ética, de Immanuel Kant (2003)[30]. As discussões sobre o contrato social de Hobbes, inseridas nas discussões atuais acerca de novas tecnologias, podem trazer uma

[29] "Leviatã", a mais importante criação de Thomas Hobbes e considerada por muitos sua obra-prima, foi originalmente publicada em 1651.
[30] Importa esclarecer que Kant discute a "regra de ouro" relacionando-a à ética em sua obra "Fundamentação da Metafísica dos Costumes", originalmente publicado em 1785, uma das mais influentes em toda a história da filosofia moral.

reflexão acerca de como tecnologias de IA demandam uma "autoridade soberana" para regular seu uso e evitar o caos. A ausência de regulamentação eficaz poderia levar a um "estado de natureza", segundo o qual os agentes tecnológicos agem de forma egoísta, potencialmente causando danos irreparáveis à sociedade. A filosofia kantiana, por sua vez, propõe o imperativo categórico, que orienta as ações humanas com base na máxima universalizável: "aja apenas de acordo com a máxima que você pode ao mesmo tempo querer que se torne uma lei universal". No contexto da IA e *internet* em geral, surge a questão: como é possível programar sistemas para agir de maneira moral com base em normas universalmente aceitáveis?

A ética na IA e a IA ética são apresentadas como campos interligados, porém distintos (Lamb, 2024). Enquanto a primeira reflete sobre os princípios morais aplicáveis ao desenvolvimento tecnológico – incluindo seus riscos, limites e responsabilidades. Ou seja, interroga-se o "dever-ser" da tecnologia. Já a IA ética visa à construção de sistemas que integrem valores éticos em sua lógica de funcionamento, seja por meio de regras explícitas, aprendizado supervisionado ou mecanismos de governança algorítmica. Ocorre que a celeuma não termina de forma simplista, haja vista que não existe consenso social sobre o que venha a ser "valores éticos".

Para ilustrar essas situações que envolvem a ética no uso de novas tecnologias, convém citar o caso amplamente divulgado do *Google Photos*, em 2015, no qual indivíduos negros foram rotulados como "gorilas", exemplificando os perigos do uso inadequado de IA em sistemas de reconhecimento facial (Garcia, 2016). O erro foi identificado pelo usuário Jacky Alcine, que compartilhou a falha no *Twitter*, atraindo ampla atenção pública e gerando repercussão global (BBC News, 2015). Esses erros, além de serem moralmente inaceitáveis, geram impactos legais e financeiros significativos para as empresas envolvidas. Do ponto de vista ético, o incidente expõe como a falta de representatividade nos dados de treinamento pode levar a resultados discriminatórios.

Como Garcia (2016) argumenta, não há neutralidade em se tratando de IA, uma vez que reflete os preconceitos implícitos nas sociedades que geram os dados. No caso do *Google Photos*, a rotulagem ofensiva reforçou estereótipos raciais e causou danos emocionais às pessoas afetadas. Conforme destaca Dal Verne (2023, p. 87), "[...] a possibilidade de os profissionais de *marketing* e as empresas

detectarem e explorarem os dados emocionais dos consumidores em tempo real levanta importantes questões éticas e legais"[31], uma vez que a assimetria informacional pode comprometer a autonomia dos usuários, por conseguinte, o autor defende a necessidade de regulamentação para evitar que a IA seja usada de maneira injusta ou discriminatória.

Um ponto abordado por Garcia (2020, p. 16) é a questão que envolve a qualidade dos dados utilizados pelas Ias. Veja-se:

> A inteligência da máquina depende da qualidade dos dados e dos exemplos a que ela é submetida, e vai reproduzir o conhecimento que está impregnado nesses dados. [...] Se a máquina receber dados e informações carregados de vieses e preconceitos [...] ela irá não só aprender com eles como perpetuá-los, durante o seu processo de aprendizado, quando exposta a novos dados.

Pesquisadores como Marcus e Davis (2019) argumentam que a explicabilidade dos sistemas é um componente essencial para conquistar a confiança pública, uma vez que é necessário o desenvolvimento de sistemas éticos que incorporem princípios de transparência e consentimento informado. A natureza de "caixa-preta" de muitos sistemas de IA, conforme mencionado por Lamb (2024), cria barreiras para a conformidade legal, pois dificulta a auditabilidade e a explicação das decisões tomadas pelos algoritmos.

Trazendo à discussão os escopos mais específicos do presente estudo, observa-se que Lamb (2024) destaca a necessidade de que sistemas de IA reflitam valores éticos consensuais. No contexto da *internet*, isso significa que as imagens desenvolvidas que utilizam a IA devem respeitar a privacidade, bem como evitar manipulação. A percepção de que os *outputs* gerados por IAG pode ser enviesado ou utilizado de forma enganosa desencadeia consequências diretamente na confiança e nas decisões dos usuários. Já no contexto específico do *neuromarketing*, é possível que os profissionais da área consigam identificar e explorar as vulnerabilidades emocionais dos usuários por meio da IAG, surgindo – portanto – preocupações sobre práticas injustas e violação do princípio da autonomia das pessoas, conforme advertem Morozov (2022) e Zuboff (2020). Os referidos autores afirmam

[31] Tradução livre do inglês: "the possibility for marketers and businesses to detect and exploit consumers' emotional data in real time raises important ethical and legal issues".

que a utilização de tecnologias digitais para mapear e influenciar comportamentos individuais compromete a liberdade de escolha, expondo os usuários a processos de manipulação emocional invisível que restringem sua autodeterminação em ambientes digitais.

A IAG utilizada pelo *Instagram* – por exemplo – analisa o comportamento do usuário por meio de algoritmos preditivos, rastreando tempo de visualização, interações, padrões de engajamento e preferências de conteúdo sem que os usuários tenham plena consciência de como esses dados são processados e utilizados. Assim como no *neuromarketing*, no qual a análise do comportamento inconsciente é usada para otimizar estratégias de persuasão, os algoritmos de IA ajustam continuamente o *feed* do usuário nas redes sociais, por exemplo, para maximizar seu tempo de engajamento e sua receptividade a anúncios personalizados. Exatamente nesse ponto que Wiederhold (2020, p. 2) reflete acerca da preocupação ética envolvida nessa dinâmica. *In verbis*:

> Uma das principais preocupações éticas no *neuromarketing* é a potencial violação da privacidade do consumidor. Ao contrário da pesquisa de *marketing* tradicional, o *neuromarketing* não depende somente de dados autorrelatados, mas, em vez disso, usa imagens cerebrais e medições fisiológicas para prever o comportamento do consumidor. Isso levanta questões sobre os limites da autonomia do consumidor e até que ponto as empresas devem ter permissão para acessar processos subconscientes para influenciar decisões de compra.

Ainda nesse contexto, sistemas de IAG empregados em decisões críticas, como aprovação de crédito, diagnósticos médicos, decisões judiciais, exemplificam a dificuldade de se lidar com a questão ética. Os vieses presentes nos dados de treinamento muitas vezes reforçam desigualdades preexistentes, como destacou Noble (2018) em seu estudo sobre algoritmos discriminatórios. Henrique Alves Pinto (2020), na pesquisa "A Utilização da Inteligência Artificial no Processo de Tomada de Decisões: Por uma Necessária *Accountability*", argumenta que, embora a IA possa trazer benefícios como eficiência e precisão, sua implementação em processos decisórios deve ser acompanhada por mecanismos de *accountability* para evitar prejuízos aos direitos fundamentais e assegurar a confiança pública.

De acordo com Ana Cristina Bicharra Garcia (2020, p. 22), "[...] para um uso consciente e com menos vieses, faz-se indispensável uma

abordagem multidisciplinar", ou seja, isso incluiria especialistas em ética, cientistas sociais, e especialistas que melhor entendem as nuances de cada área de aplicação da IA. Modelos de aprendizado de máquina frequentemente apresentam múltiplos caminhos para alcançar soluções semelhantes, mas a ausência de uma compreensão do "porquê" desses caminhos pode levar a decisões opacas ou discriminatórias. Por conseguinte, o desafio de compreender o comportamento humano por meio da IA exige integrar dados e análises comportamentais, bem como valores sociais.

Os resultados que foram obtidos nos testes, e serão analisados na seção a seguir, remetem diretamente à necessidade de se refletir sobre os pressupostos éticos do uso da IA em ambientes sociais e institucionais. Se, por um lado, a IAG amplia as possibilidades de representação e comunicação, por outro, seus efeitos sobre a confiança pública podem comprometer direitos fundamentais, sobretudo em áreas sensíveis como justiça, saúde e finanças. É nesse ponto que a discussão sobre procedimentos e considerações éticas se torna indispensável, fornecendo o arcabouço normativo e filosófico capaz de orientar a utilização responsável da IA e de mitigar os riscos inerentes à sua aplicação.


6

6 ANÁLISE DE DADOS

Os dados quantitativos coletados nesta pesquisa foram organizados em três partes principais: (i) a caracterização demográfica dos participantes; (ii) as respostas à Escala de Avaliação de Associação Explícita (EAAT); e (iii) os resultados do IAT.

Ao longo do processo de coleta, 501 (quinhentos e um) participantes iniciaram os testes, no entanto, 493 (quatrocentos e noventa e três) concluíram o questionário demográfico inicial, enquanto 464 (quatrocentos e sessenta e quatro) completaram integralmente o EAAT. Contudo, apenas 344[32] (trezentos e quarenta e quatro) indivíduos finalizaram as três etapas – o questionário demográfico, o EAAT e o IAT. Esse conjunto constitui a amostra final considerada válida para as análises de correlação e comparação entre atitudes explícitas e implícitas. Assim, embora o leitor encontre diferentes valores de "n" nas seções descritivas de cada teste, as conclusões inferenciais e estatísticas apresentadas nesta dissertação referem-se exclusivamente ao grupo completo de 344 participantes, garantindo a consistência dos resultados e a validade das inferências comparativas.

A imagem que segue – Figura 8 – deixa essa informação mais didática.

Figura 8 – Participantes da pesquisa empírica

Fonte: Elaborado pelo pesquisador (2025).

A primeira etapa corresponde à análise demográfica contemplou variáveis como idade, sexo/gênero, nível de escolaridade, familiaridade com tecnologias digitais e experiências prévias com inteligência artificial, de modo a traçar um panorama diversificado do

[32] Montante já excluindo os 36 (trinta e seis) participantes que não atenderam aos critérios de qualidade recomendados por Greenwald *et al.* (2003).

perfil dos participantes. A caracterização demográfica é relevante na medida em que aspectos individuais podem influenciar tanto as associações implícitas captadas pelo IAT quanto os julgamentos conscientes de confiança aferidos pela EAAT, oferecendo subsídios para uma interpretação contextualizada dos resultados experimentais.

Em relação à idade, a amostra apresentou variação entre 18 e 75 anos, com média de 31,3 anos (DP = 12,3) e mediana de 28 anos. O primeiro quartil concentrou-se em 20 anos, enquanto o terceiro quartil foi de 40 anos, revelando uma amostra composta majoritariamente por jovens adultos e adultos em fase de consolidação profissional.

No tocante ao sexo/gênero, observou-se relativa paridade entre os respondentes: 51,1% identificaram-se como masculinos e 47,5% como femininos, enquanto 1,4% optou por não responder. Essa distribuição indica a presença de diferentes identificações de gênero na amostra, permitindo a descrição de respostas provenientes de perfis diversos.

No que diz respeito à escolaridade, houve predominância de indivíduos com ensino superior incompleto (37,7%), seguidos por participantes com pós-graduação *lato sensu* (21,8%), ensino superior completo (15,8%) e ensino médio completo (12,2%). Percentuais menores foram registrados para pós-graduação *stricto sensu* – indicado por participantes que declararam mestrado (6,0%) e doutorado (1,6%) –, ensino médio incompleto (1,2%) e ensino fundamental completo (1,0%), além de 1,2% em outras formas de escolaridade e 1,6% que não responderam.

Em relação à familiaridade com IA, constatou-se um uso elevado: 42,1% declararam utilizar ferramentas dessa natureza várias vezes ao dia, enquanto 9,0% relataram uso diário e 26,1% algumas vezes por semana. Frequências menores foram atribuídas a algumas vezes por mês (6,4%), menos de uma vez por mês (6,2%) e uma vez por semana (4,0%), ao passo que 4,6% nunca haviam utilizado tais tecnologias. Esses dados indicam que parcela expressiva da amostra possui contato recorrente com sistemas de IA.

Com o escopo de assegurar maior didática na exposição dos resultados, os dados coletados foram sistematizados na Tabela 1, a qual apresenta de forma resumida o perfil sociodemográfico da amostra pesquisada. A disposição tabular das informações possibilita uma visualização organizada das principais distribuições de frequências,

percentuais e médias, favorecendo a compreensão dos padrões observados.

Tabela 1 – Perfil sociodemográfico da amostra (n=344)

Categoria	f (n)	%
Masculino	176	51,2
Feminino	163	47,4
Não respondeu	5	1,5
Idade (Média/DP)	31,3 (12,3)	–
Sup. incompleto	135	39,2
Pós-grad. lato sensu	78	22,7
Sup. completo	57	16,6
Ensino médio completo	43	12,5
Mestrado	21	6,1
Ens. médio incompleto	6	1,7
Fundamental completo	4	1,2

Fonte: Elaborado pelo pesquisador (2025).

Na sequência dos testes, houve a aplicação da Escala de Avaliação de Associação Explícita (EAAT)[33], o qual – por se tratar de uma escala de autorrelato em formato *Likert* (Likert, 1932; Devellis, 2017) – não está sujeito a critérios de exclusão baseados em tempo de resposta ou acurácia. Nesse tipo de medida, as exclusões concentram-se em situações de questionários incompletos ou padrões de resposta inválidos (como a escolha repetida da mesma opção em todos os itens).

O EAAT foi aplicado com o objetivo de mensurar as atitudes conscientes dos participantes em relação às imagens apresentadas. O instrumento foi estruturado em dois fatores principais: confiança e manipulação. A dimensão "confiança" foi avaliada por meio de quatro itens que investigaram em que medida as imagens transmitiam credibilidade, integridade, benevolência e competência institucional. Já a dimensão "manipulação" foi medida por outros quatro itens, que

[33] Participaram dessa etapa 464 (quatrocentos e sessenta e quatro) respondentes, os quais todos finalizaram integralmente esse instrumento (EAAT), resultando em uma taxa de evasão nula. Esse índice contrasta positivamente com a etapa demográfica, na qual, embora tenha havido elevada completude (98,4%), registrou-se a ausência parcial de respostas em sete questionários. No presente estudo, entretanto, todos os 464 (quatrocentos e sessenta e quatro) respondentes completaram integralmente a escala e não foram observados padrões que indicassem respostas artificiais ou inválidas. Dessa forma, nenhum caso foi excluído do EAAT, sendo todos os participantes considerados válidos para a análise estatística e, portanto, seguiram para a etapa subsequente do estudo, que era o IAT. O resultado indica que, uma vez ultrapassada a fase inicial de caracterização sociodemográfica, os participantes demonstraram maior engajamento e comprometimento com a tarefa avaliativa, reduzindo a ocorrência de desistências e assegurando a totalidade dos dados necessários à análise das atitudes explícitas em relação à confiança nas imagens apresentadas.

buscavam captar a percepção de artificialidade ou distorção nas imagens.

Todos os itens foram respondidos em uma escala do tipo *Likert* de 5 (cinco) pontos, com âncoras variando de 1 = nada confiável/manipulado a 5 = muito confiável/manipulado. Não houve a necessidade de se colocar itens invertidos recodificados, assim os escores finais de cada dimensão foram calculados pela média aritmética das respostas aos respectivos itens.

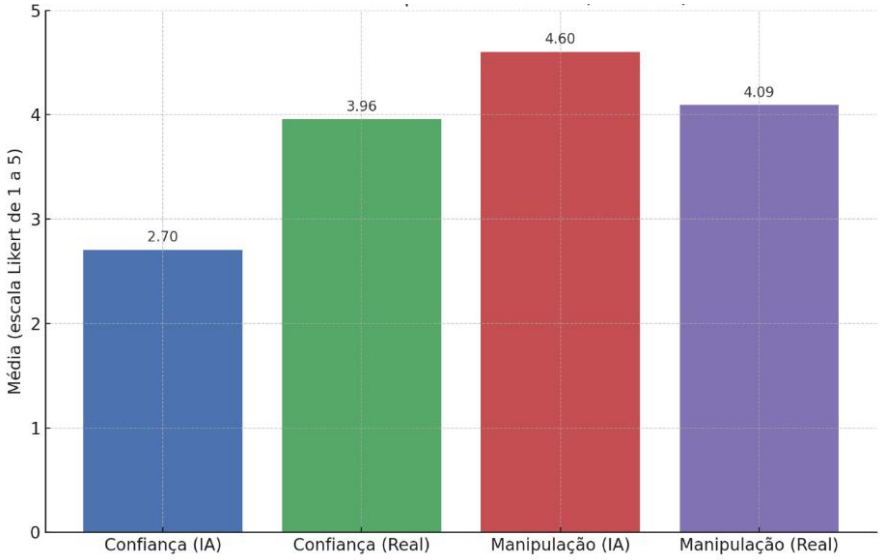
Os resultados indicam diferenças expressivas entre a confiança atribuída a imagens reais e a imagens produzidas por IA. A variável "confiança" em imagens reais apresentou média de 3,96 (DP = 0,87), mediana de 4,0 e quartis em 4,0 e 5,0, sugerindo valores mais elevados de atribuição de credibilidade a representações autênticas. Em contraste, a confiança em imagens de IA obteve média de 2,70 (DP = 0,95), mediana de 3,0 e quartis em 2,0 e 3,0, indicando valores inferiores de confiança atribuída. As médias e desvios-padrão foram calculados a partir das respostas obtidas na Escala *Likert* de 5 pontos aplicada no EAAT, variando de 1 ("nada confiável") a 5 ("muito confiável"). As medidas de tendência central (média e mediana) e dispersão (desvio-padrão e quartis) foram estimadas utilizando estatística descritiva, com base nas respostas válidas dos 344 participantes que concluíram integralmente todas as etapas da pesquisa.

Esse conjunto de medidas permite observar a tendência geral das percepções de confiança, bem como a consistência das respostas entre os participantes, oferecendo um panorama descritivo das diferenças observadas entre as imagens reais e aquelas geradas por inteligência artificial, considerando exclusivamente a amostra consolidada de 344 participantes.

No que se refere à percepção de manipulação, verificou-se padrão distinto. As imagens de IA foram avaliadas com média de 4,60 (DP = 0,59), mediana de 5,0 e forte concentração no limite superior da escala (Q1 = 4,0; Q3 = 5,0), sugerindo reconhecimento predominante de artificialidade. Já as imagens reais receberam média de 4,09 (DP = 0,84), mediana de 4,0 e quartis em 4,0 e 5,0, o que indica a presença de percepções de possível manipulação, ainda que em níveis médios inferiores aos atribuídos às imagens artificiais.

Sequencialmente, apresentam-se esses percentuais no gráfico, a fim de que seja permitida a melhor visualização dos dados gerados nesse tocante.

Gráfico 1 – Médias das respostas do EAAT (n=344)



Legenda da Escala Likert (1 a 5) utilizada no EAAT

Valor	Confiança	Manipulação
1	Não confio nem um pouco	Nada manipulada
2	Confio pouco	Pouco manipulada
3	Confio moderadamente	Moderadamente manipulada
4	Confio bastante	Bastante manipulada
5	Confio totalmente	Totalmente manipulada

Fonte: Elaborado pelo pesquisador (2025).

A última etapa dos testes se refere ao IAT[34], no qual 380 (trezentos e oitenta) participantes concluíram integralmente. No entanto, 36 (trinta e seis) serão excluídos em razão do não atendimento aos critérios de qualidade recomendados por Greenwald *et al.* (2003), operacionalizados pela variável *excludeCriteriaMet*. Esses critérios incluem a presença de mais de 10% de respostas com latências inferiores a 300 ms, bem como taxas de acerto inferiores a 70%, parâmetros que comprometem a validade psicométrica dos resultados. Dessa forma, a amostra final válida para análise foi composta por 344 (trezentos e quarenta e quatro) participantes, assegurando maior

[34] Verificou-se que 407 (quatrocentos e sete) participantes iniciaram a tarefa, dos quais 380 (trezentos e oitenta) a concluíram integralmente, resultando em uma taxa de evasão interna de 27 (vinte e sete) indivíduos (6,4%). Tal resultado sugere que o IAT, por demandar maior tempo de execução e exigir respostas rápidas a estímulos sucessivos, pode ter imposto maior carga cognitiva e gerado maior propensão ao abandono em relação às etapas anteriores do estudo.

consistência metodológica e alinhamento às boas práticas de uso do IAT em pesquisas empíricas.

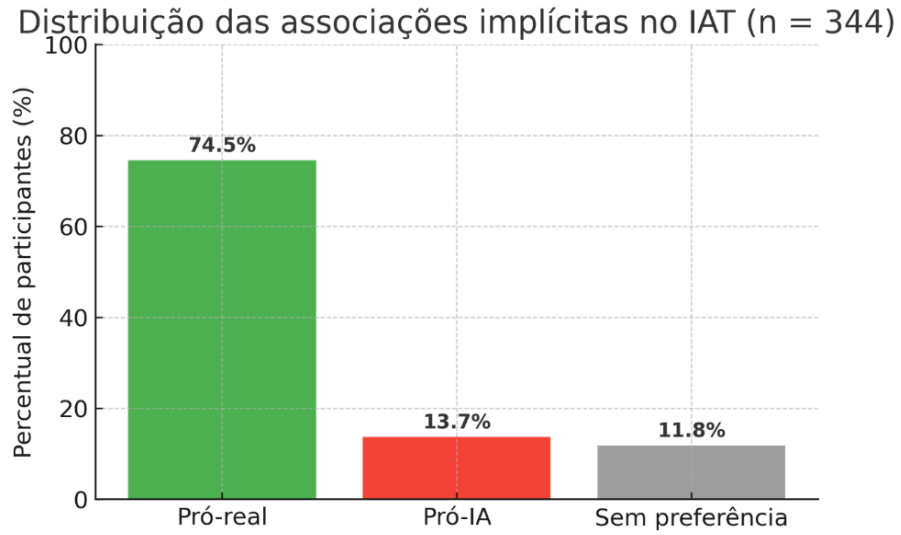
A análise dos escores do IAT indicou padrões compatíveis com associações implícitas entre as categorias avaliadas. O *D-score* agregado (d), calculado segundo o algoritmo proposto por Greenwald, Nosek e Banaji (2003) – que consiste na padronização das diferenças de tempos médios de resposta entre blocos compatíveis e incompatíveis, ajustados pela variabilidade intraindivíduos – apresentou média de 0,43 (DP = 0,39), mediana de 0,46, com variação entre –0,64 e 1,33. Valores positivos correspondem a maior facilidade relativa de associação entre as imagens reais e atributos de confiança, enquanto valores negativos correspondem a um padrão inverso, observado com menor frequência na amostra.

No que se refere ao desempenho, a proporção média de acertos foi de 83,2% (DP = 12,0%), com mediana de 86,9%, valor compatível com níveis elevados de acurácia na tarefa. A análise da proporção de respostas rápidas (RT < 300 ms) mostrou média de apenas 0,27% (DP = 1,15%), valor bastante inferior ao limite de 10% recomendado como parâmetro de qualidade metodológica, o que sugere que os participantes compreenderam as instruções e não responderam de forma precipitada ou aleatória.

A dimensão analisada na presente pesquisa foi mensurada por meio do *D-score*, indicando que 74,5% dos participantes apresentaram escores compatíveis com preferência implícita por imagens reais, distribuídos entre preferência leve (14,7%), moderada (29,1%) e forte (30,7%). Em contrapartida, 13,7% apresentaram escores compatíveis com preferência implícita por imagens de IA, sendo 7,6% em grau leve, 3,7% moderado e 2,4% forte. Por fim, 11,8% situaram-se na faixa de neutralidade (–0,15 ≤ D ≤ +0,15), sem indicação de preferência implícita predominante. Esses resultados são indicativos de maior frequência de associações implícitas favoráveis às imagens reais, em consonância com os padrões observados nas medidas explícitas do EAAT, ainda que uma parcela minoritária da amostra apresente associações favoráveis às imagens produzidas por inteligência artificial.

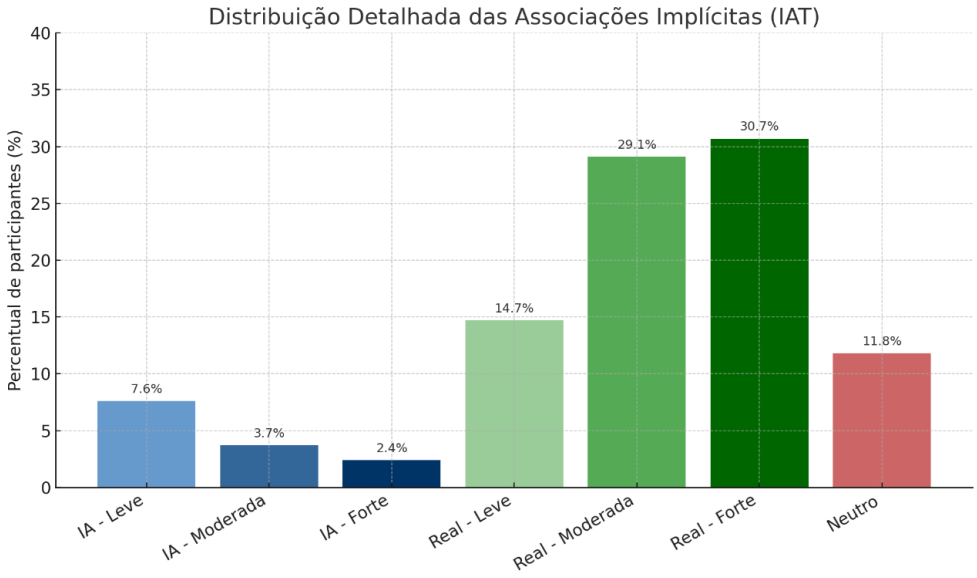
Os Gráfico 2 e Gráfico 3, sequencialmente apresentados, elucidam tais resultados:

Gráfico 2 – Distribuição das Associações Implícitas no IAT (n=344)



Fonte: Elaborado pelo pesquisador (2025).

Gráfico 3 – Distribuição Detalhada das Associações Implícitas (IAT)



Fonte: Elaborado pelo pesquisador (2025).

A Tabela 2, na sequência, sintetiza a categorização acima indicada – leve, moderada e forte; pró-real, pró-IA e sem preferência.

Tabela 2 – Classificação das Associações Implícitas segundo faixas do D-score (n=344)

Categoria	Leve	Moderada	Forte	Total (%)
Pró-real	14,7%	29,1%	30,7%	74,5%
Pró-IA	7,6%	3,7%	2,4%	13,7%
Sem preferência	–	–	–	11,8%
Total	–	–	–	100%

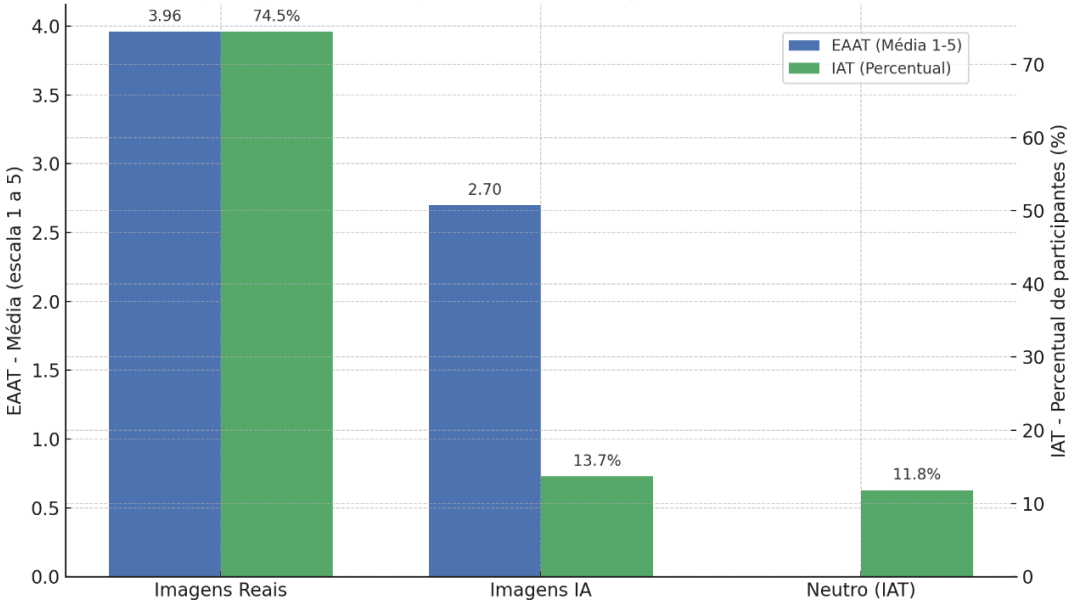
Fonte: Elaborado pelo pesquisador (2025).

Com o objetivo de manter a comparabilidade a comparabilidade entre medidas explícitas e implícitas, a análise foi conduzida exclusivamente com os participantes que concluíram integralmente todas as etapas da pesquisa, isto é, o questionário demográfico, o EAAT e o IAT. Dessa forma, embora o EAAT tenha contado inicialmente com 464 (quatrocentos e sessenta e quatro) respondentes, apenas os 344 (trezentos e quarenta e quatro) participantes que finalizaram todos os instrumentos e atenderam aos critérios de qualidade do IAT foram considerados na análise comparativa. Essa decisão metodológica permite que os contrastes realizados sejam baseados nos julgamentos de um mesmo grupo de indivíduos, possibilitando a descrição de padrões de coerência ou dissociação entre os níveis consciente e automático de avaliação.

A comparação entre os resultados do IAT e do EAAT indica a presença de dissociação entre atitudes implícitas e explícitas. No EAAT, os participantes atribuíram maior confiabilidade às imagens reais (M = 3,96) em comparação às imagens de IA (M = 2,70), ao mesmo tempo em que identificaram maior manipulação nas imagens artificiais (M = 4,60) do que nas reais (M = 4,09). No IAT, por sua vez, os escores implícitos apontam que 74,5% da amostra apresentaram escores compatíveis com preferência por imagens reais, enquanto 13,7% apresentaram escores compatíveis com preferência por imagens de IA e 11,8% situaram-se na faixa de neutralidade. Esses resultados são indicativos de que, embora as avaliações conscientes apontem menor atribuição de confiança às imagens produzidas por IAG e valores mais elevados de confiança às imagens reais, as associações automáticas captadas pelo IAT apresentam padrão semelhante, com maior frequência de associações favoráveis às representações autênticas. Esse conjunto de achados é compatível com a hipótese de dissociação entre julgamentos implícitos e explícitos, sugerindo que a confiança atribuída às imagens pode ser processada de maneira distinta nos níveis consciente e inconsciente.

O Gráfico 4, em continuidade, evidencia a comparação EAAT (explícito) *versus* IAT (implícito):

Gráfico 4 – Comparação EAAT (explícito) versus IAT (implícito) – incluindo Neutro



Fonte: Elaborado pelo pesquisador (2025).

O gráfico comparativo indica a presença de dissociação entre atitudes explícitas e implícitas frente às imagens analisadas. No nível consciente, mensurado pelo EAAT, os participantes atribuíram maior confiança às imagens reais (M = 3,96) em relação às de IA (M = 2,70), resultado coerente com a literatura sobre credibilidade em ambientes digitais. Já no plano implícito, captado pelo IAT, 74,5% dos participantes apresentaram preferência por imagens reais, enquanto apenas 13,7% favoreceram imagens de IA, havendo, ainda, um contingente neutro de 11,8% que se situou na faixa de neutralidade. Esse resultado corrobora o modelo de confiança de Mayer, Davis e Schoorman (1995), ao indicar que dimensões como competência, integridade e benevolência são mais prontamente vinculadas a representações autênticas. Ao mesmo tempo, confirma a tese de Greenwald *et al.* (1998) de que vieses implícitos podem operar de maneira distinta das declarações conscientes, indicando a existência de uma camada automática de julgamento. Ademais, em consonância com Zaltman (2003), os achados sugerem que as imagens reais ativam associações simbólicas mais profundas relacionadas à legitimidade e autoridade institucional, ao passo que as imagens artificiais, ainda que reconhecidas cognitivamente, despertam maior estranhamento no plano implícito. Assim, a análise integrada dos dois testes aponta para a presença de um descompasso entre a confiança declarada e a confiança automática, o que pode ter implicações para o uso de imagens de IA em contextos jurídicos e institucionais.

Assim, os dados obtidos indicam que, apesar da crescente familiaridade com imagens criadas por IA no cotidiano digital, subsiste uma diferença significativa de confiança e percepção crítica quando comparadas a imagens reais. Esses achados sugerem possíveis repercussões para o uso de IAG em contextos institucionais, como campanhas públicas, comunicação jurídica e produção de provas visuais, apontando para a necessidade de consideração de critérios relacionados à transparência, veracidade e rotulagem do conteúdo visual.

Com o objetivo de organizar a apresentação dos resultados, os dados obtidos foram organizados em tabelas – Tabela 3 e Tabela 4 – que sintetizam as estatísticas descritivas das medidas explícitas (EAAT) e implícitas (IAT). As tabelas apresentam de forma sistematizada as principais distribuições de frequências, percentuais e médias associadas às variáveis analisadas.

Tabela 3 – Estatísticas descritivas do EAAT (n=344)

Variável	M	DP	Md	Q1	Q3
Confiança (Reais)	3,96	0,87	4,0	4,0	5,0
Confiança (IA)	2,70	0,95	3,0	2,0	3,0
Manipulação (Reais	4,09	0,84	4,0	4,0	5,0
Manipulação (IA)	4,60	0,59	5,0	4,0	5,0

Fonte: Elaborado pelo pesquisador (2025).

Tabela 4 – Distribuição das Associações Implícitas (IAT) (n=344)

Categoria de associação implícita	% participantes
Preferência leve – Real	14,7%
Preferência moderada – Real	29,1%
Preferência forte – Real	30,7%
Preferência leve – IA	7,6%
Preferência moderada – IA	3,7%
Preferência forte – IA	2,4%
Neutro	11,8%

Fonte: Elaborado pelo pesquisador (2025).

Além disso, uma tabela consolidada – Tabela 5 – foi elaborada a fim de comparar, lado a lado, as medidas explícitas e implícitas de confiança e manipulação, evidenciando a dissociação entre julgamentos conscientes e automáticos.

Tabela 5 – Comparação consolidada entre EAAT (n=344) e (IAT) (n=344)

Medida	Resultado
Confiança (Reais) – EAAT	M = 3,96
Confiança (IA) – EAAT	M = 2,70
Manipulação (Reais) – EAAT	M = 4,09
Manipulação (IA) – EAAT	M = 4,60
Preferência Real – IAT	74,5%
Preferência IA – IAT	13,7%
Neutro – IAT	11,8%

Fonte: Elaborado pelo pesquisador (2025).

A análise dos resultados permite retomar as hipóteses previamente formuladas. A H1 foi corroborada, pois as imagens reais apresentaram maior média de confiança no EAAT (M = 3,96; DP = 0,87) em relação às imagens geradas por IA (M = 2,70; DP = 0,95). A H2 também encontrou suporte empírico, uma vez que o IAT, considerando os 344 (trezentos e quarenta e quatro) participantes válidos, revelou predominância de associações favoráveis às imagens reais (74,5%), em contraste com 13,7% que favoreceram imagens de IA e 11,8% que permaneceram neutros, confirmando a tendência pró-real. A H3, que postulava a existência de dissociação entre atitudes explícitas e implícitas, mostrou-se compatível com os resultados obtidos, na medida em que, embora as respostas conscientes apresentem valores mais elevados de aceitação da IA, os julgamentos automáticos, captados pelo *D-score* (M = 0,43; DP = 0,39), mantêm forte viés pró-real. Por fim, a H4 igualmente foi sustentada pelos dados, visto que os participantes atribuíram maior percepção de manipulação às imagens artificiais (M = 4,60; DP = 0,59) do que às reais (M = 4,09; DP = 0,84). Em conjunto, os achados são indicativos de consonância entre os resultados empíricos e as hipóteses formuladas, sem prejuízo de interpretações mais amplas a serem desenvolvidas na seção de Discussão.


7

7 CONSIDERAÇÕES FINAIS

As implicações deste estudo concentram-se na compreensão de como a confiança é construída e modulada diante de imagens produzidas por IAG, especialmente em contextos jurídicos e institucionais. Os resultados evidenciaram que, embora exista certa aceitação consciente das imagens artificiais, as associações implícitas mantêm forte viés pró-real, favorecendo representações visuais reais. Essa dissociação entre atitudes explícitas e implícitas indica que a credibilidade de instituições de justiça, quando mediada por recursos visuais, pode ser impactada de forma distinta nos níveis consciente e inconsciente de processamento. Nesse sentido, compreender os mecanismos que sustentam a confiança em ambientes digitais valoriza a pertinência social da pesquisa, além de oferecer subsídios concretos para o uso ético e estratégico de imagens em campanhas institucionais, no Tribunal do Júri e em outras práticas comunicacionais do sistema de justiça.

A análise dos dados coletados permitiu confirmar a relevância de investigar a confiança em conteúdos visuais gerados por IAG. O exame empírico demonstrou que as imagens reais de profissionais do sistema de justiça são percebidas como mais confiáveis do que as artificiais, tanto no plano consciente, captado pelo EAAT, quanto no plano inconsciente, evidenciado pelo IAT. Essa constatação reforça a ideia de que a autenticidade visual permanece como referência central para a construção da credibilidade no ambiente digital.

Outra revelação fundamental foi a evidência de uma dissociação entre atitudes explícitas e implícitas. Enquanto as avaliações conscientes revelaram certa aceitação das imagens artificiais, especialmente quando comparadas ao cotidiano de exposição tecnológica dos usuários, as associações automáticas mantiveram uma preferência marcante pelas representações autênticas. Essa divergência confirma que a confiança não se reduz a julgamentos racionais, mas envolve também predisposições emocionais e simbólicas, alinhando-se à literatura que aponta para a complexidade dos processos de percepção no ambiente digital. Essa dissociação entre atitudes explícitas e implícitas confirma a perspectiva simbólica de Zaltman (2003), ao evidenciar que percepções de confiança são

moldadas por associações afetivas e culturais que ultrapassam o raciocínio lógico.

Noutro giro, o presente trabalho verificou que as imagens de IAG foram mais fortemente associadas à ideia de manipulação, evidenciando que os usuários reconhecem limitações éticas e riscos de engano nesses conteúdos. Essa constatação reforça a necessidade de maior transparência e rotulagem das representações artificiais, sobretudo em contextos institucionais sensíveis, como o sistema de justiça, nos quais a confiança é elemento estruturante da legitimidade social. A associação entre imagens artificiais e percepções de manipulação confirma o que a literatura tem descrito como crise de transparência algorítmica (Gillespie, 2018). A opacidade dos processos de criação e difusão de conteúdos digitais fragiliza a confiança pública e amplia a necessidade de mecanismos claros de identificação e rastreabilidade das representações produzidas por IA. Essa constatação está em consonância com as proposições éticas de Simões e Caldeira (2025), que defendem o equilíbrio entre inovação técnica e integridade comunicacional.

A análise dos dados obtidos e compilados permitiu, ainda, retomar as hipóteses previamente formuladas. A H1 foi confirmada, uma vez que as imagens reais apresentaram médias significativamente mais altas de confiança no EAAT (M = 3,96; DP = 0,87), em comparação às imagens produzidas por IAG (M = 2,70; DP = 0,95). Esse achado demonstra que, no nível consciente, os participantes atribuíram maior credibilidade às representações autênticas, reforçando a centralidade da autenticidade para a construção da confiança.

No que tange à H2, houve, também, confirmação, haja vista que os escores do IAT revelaram predominância de associações implícitas favoráveis às imagens reais em detrimento das artificiais, com 74,5% dos participantes apresentando preferência pró-real, contra apenas 13,7% pró-IA e 11,8% sem preferência. Esse padrão indica que a confiança automática, menos sujeita a racionalizações conscientes, permanece fortemente inclinada em direção às representações reais.

A H3, por sua vez, que postulava a existência de dissociação entre atitudes explícitas e implícitas, também encontrou suporte nos resultados obtidos. Enquanto as respostas conscientes (EAAT) sugerem uma relativa aceitação da IA, as associações automáticas (IAT) reforçam de modo ainda mais acentuado a preferência por imagens reais. Essa

discrepância confirma que os vieses implícitos podem operar de forma distinta dos julgamentos conscientes, revelando diferentes camadas no processo de atribuição de confiança.

Por fim, a H4 igualmente foi corroborada, visto que os participantes atribuíram maior percepção de manipulação às imagens artificiais (M = 4,60; DP = 0,59) do que às reais (M = 4,09; DP = 0,84). Esse resultado explicita que a artificialidade desperta, em nível consciente, uma maior sensação de distorção ou falta de autenticidade, o que reforça a cautela necessária no uso de representações produzidas por inteligência artificial em contextos institucionais.

Depois de confrontar os resultados obtidos com as hipóteses que foram formuladas na proposição deste trabalho, passa-se a analisar as importantes contribuições teóricas para a compreensão da confiança em ambientes digitais mediados por IA.

Conforme exposto na seção 2.1, a literatura sobre algoritmos aponta que tais sistemas não são meros instrumentos neutros, mas, sim, estruturas mediadoras do conhecimento, das escolhas e da percepção social (Cotter, 2019). Ao produzirem *outputs* – como imagens, textos ou recomendações – os algoritmos materializam processos de seleção que refletem tanto critérios técnicos quanto valores implícitos em seus dados de treinamento. Os resultados reunidos neste estudo reforçaram essa visão: no EAAT, os participantes atribuíram menor confiança às imagens geradas por IA (M = 2,70) em comparação às imagens reais (M = 3,96), sinalizando que, no plano consciente, os *outputs* algorítmicos ainda são percebidos como menos íntegros e menos transparentes. Essa menor credibilidade parece ecoar os alertas teóricos de que algoritmos operam como "caixas-pretas" (Pasquale, 2015), cujas lógicas opacas dificultam a construção de confiança social em seus resultados.

Entretanto, os resultados do IAT revelaram uma camada distinta: os participantes tenderam a associar com maior naturalidade as imagens reais a atributos de confiança, enquanto as artificiais não despertaram o mesmo padrão de associação automática. Isso evidencia que, mesmo quando não há julgamento consciente, os algoritmos de produção de imagens ainda enfrentam dificuldades para evocar credibilidade equivalente à das representações autênticas. Essa constatação dialoga com a crítica de Gillespie (2018) sobre os chamados algoritmos de destruição em massa, que, ao reproduzirem vieses

invisíveis, podem consolidar desconfiança e exclusão social. Ao mesmo tempo, confirma que a confiança atribuída a *outputs* algorítmicos é modulada não apenas pela percepção consciente de manipulação (captada no EAAT), mas também por associações emocionais e simbólicas mais profundas (captadas no IAT), reforçando que a lógica algorítmica não é recebida como neutra, mas como impregnada de significados que afetam diretamente sua aceitação social.

Na concepção de Damasio (1994), as emoções exercem um papel determinante no processo de decisão, funcionando como "marcas somáticas" que orientam a escolha entre diferentes alternativas. A partir dessa perspectiva, compreende-se que o raciocínio humano não se afasta das reações emocionais, mas é constantemente informado por elas. Os resultados deste estudo caminharam nesse sentido: os participantes demonstraram confiar mais nas imagens reais do que nas artificiais, o que sugere que a credibilidade consciente tende a se apoiar em estímulos capazes de transmitir autenticidade e familiaridade. Além disso, os *outputs* produzidos por IAG foram percebidos como mais manipulados, reforçando a ideia de que sinais emocionais negativos influenciam diretamente o julgamento e funcionam como alertas contra potenciais riscos de falsificação ou engano.

Contudo, os resultados do IAT revelaram que essa influência emocional não se limita ao nível consciente. Ao demonstrar que 74,5% dos participantes apresentaram associações implícitas de confiança em favor de imagens reais, contra apenas 13,7% para imagens artificiais, os dados obtidos e analisados indicaram que os mecanismos emocionais automáticos reforçam a tendência de vincular credibilidade à autenticidade visual. Essa constatação dialoga com as contribuições de Zaltman (2003), para quem grande parte dos significados que orientam decisões humanas é formada por associações simbólicas profundas, muitas vezes inconscientes. Em um cenário digital marcado pela presença de conteúdos produzidos por IAG, isso significa que, ainda que os usuários possam reconhecer a utilidade tecnológica das imagens artificiais, suas reações emocionais – tanto explícitas quanto implícitas – continuam a privilegiar representações percebidas como reais, confirmando a centralidade das emoções na mediação entre tecnologia e confiança social.

A literatura sobre credibilidade no ambiente digital enfatiza que a confiança *on-line* depende de múltiplos fatores, como *design*, reputação da fonte, clareza das informações e percepção de

autenticidade (Fogg, 2003; Flanagin; Metzger, 2013). Nesse sentido, os dados do EAAT confirmaram que a credibilidade das imagens está diretamente associada à percepção de autenticidade: imagens reais receberam maior confiança (M = 3,96) do que as artificiais (M = 2,70), além de menor percepção de manipulação (M = 4,09 contra 4,60). Esses resultados reforçam a tese de que, no ambiente digital, a credibilidade decorre da aparência formal do conteúdo, bem como de uma avaliação crítica dos usuários sobre a origem e a confiabilidade do processo de produção da informação. Ao revelar que a manipulação percebida foi mais acentuada em *outputs* artificiais, o trabalho ora apresentado demonstrou que a credibilidade *on-line* continua ancorada em parâmetros de veracidade e transparência, mesmo diante da sofisticação tecnológica da IAG.

Por outro lado, os resultados do IAT revelaram a presença de uma ambivalência cognitiva na forma como os participantes reagem às reproduções visuais desenvolvidas por IAG. Apesar da crescente familiaridade com esse tipo de conteúdo, as associações automáticas continuam a privilegiar as representações reais, indicando que, no plano implícito, persiste uma resistência em atribuir plena confiança às las. Essa discrepância mostra que, mesmo quando há disposição consciente para aceitar a IA como legítima, permanecem barreiras inconscientes que reforçam a autenticidade como critério central de credibilidade. É possível que num futuro próximo isso possa mudar, haja vista que a familiaridade com a tecnologia tende a 112orna-la mais "tolerável" ao ser humano. Essa afirmação dialoga com a teoria da aculturação tecnológica (Mcluhan, 1964; Postman, 1985), pela qual a repetição e a integração cotidiana de novas mídias transformam o que inicialmente causa estranhamento em algo socialmente naturalizado. Tal fenômeno está em consonância com o conceito de ambivalência cognitiva, segundo o qual é possível sustentar simultaneamente avaliações racionais positivas e reações implícitas negativas diante de uma mesma tecnologia (Rydell; Mackie, 2008). No caso deste trabalho, isso significa que os usuários podem reconhecer a utilidade ou inevitabilidade da IA, mas, inconscientemente, mantêm predisposições pró-real, o que explica a vantagem das imagens reais no nível das associações implícitas.

O conceito de ambivalência cognitiva descreve a coexistência de atitudes contraditórias em relação a um mesmo objeto, combinando avaliações conscientes positivas com reações inconscientes negativas (Rydell; Mackie, 2008). No contexto da pesquisa relatada por meio desta

dissertação, essa ambivalência manifestou-se empiricamente: no EAAT, ainda que as imagens artificiais tenham sido avaliadas com menor confiança média (M = 2,70) em comparação às reais (M = 3,96), sua presença no ambiente digital foi reconhecida pelos participantes como parte legítima do ecossistema comunicacional. Esse dado sugere que, em nível consciente, os usuários já aceitam a inevitabilidade da IAG como elemento integrante da comunicação digital, ainda que mantenham reservas quanto à sua autenticidade e integridade.

Contudo, os resultados do IAT evidenciaram a face implícita dessa avaliação: ainda que as imagens geradas por IAG sejam aceitas no discurso consciente, as associações automáticas demonstraram uma tendência consistente de privilegiar as representações reais. Isso indica que, enquanto a IA pode ser tolerada ou até valorizada por sua utilidade, no nível inconsciente prevalece um viés pró-real, que vincula maior confiança ao que é percebido como genuíno. Esse fenômeno confirma o diagnóstico de que tecnologias disruptivas, como a IA, desencadeiam tensões entre a familiaridade cultural e a resistência simbólica, produzindo uma ambivalência cognitiva estrutural capaz de impactar diretamente a credibilidade de mensagens visuais em contextos institucionais e jurídicos.

De certo, o próprio desenvolvimento do IAT por Greenwald *et al.* (1998) já mostrava que julgamentos automáticos podem contrariar o que as pessoas declaram em seus relatos. De forma semelhante, Rydell e Mackie (2008) conceituaram esse fenômeno como ambivalência cognitiva, ressaltando que indivíduos podem conviver com avaliações positivas conscientes e resistências inconscientes. Em um campo mais próximo ao presente estudo, Fietta *et al.* (2021) verificaram que, apesar de a maioria dos participantes expressar simpatia declarada pela IA, apenas uma fração reduzida demonstrava associações automáticas favoráveis, enquanto predominavam vieses implícitos contrários à tecnologia.

Ainda, nesse esforço que se está fazendo em confrontar os resultados empíricos dos testes com as teorias apresentadas na parte teórica desta dissertação, imperioso relembrar a chamada Teoria do Vale da Estranheza, formulada por Mori (1970), a qual sustenta que, quanto mais um artefato artificial se aproxima da aparência humana, maior tende a ser a aceitação – até certo ponto. Quando a semelhança atinge um limiar em que o objeto é quase humano, mas não plenamente convincente, gera-se uma sensação de estranhamento,

repulsa ou desconfiança. Esse fenômeno, amplamente explorado em pesquisas sobre robótica e interfaces digitais, pode ser estendido às imagens criadas por IAG. O presente estudo corroborou essa leitura: os participantes atribuíram níveis elevados de confiança às imagens reais, mas foram mais críticos em relação às artificiais, avaliando-as como mais manipuladas e menos autênticas. A discrepância sugere que as imagens produzidas por IA se situam justamente nesse "vale" – suficientemente realistas para evocar comparação com representações autênticas, mas incapazes de transmitir plena naturalidade ou legitimidade.

Em que pese os participantes tenham reconhecido conscientemente a presença da IA como parte do ecossistema digital, suas reações automáticas privilegiaram sistematicamente as imagens reais, razão pela qual esse viés implícito pró-real pode ser lido como expressão do desconforto inconsciente que o *uncanny valley* descreve: o objeto artificial é processado como familiar, mas, simultaneamente, desperta uma sensação de artificialidade que mina a confiança. Nesse diapasão, os achados deste trabalho atualizam a teoria de Mori (1970) para o contexto jurídico-comunicacional, sugerindo que, quando aplicadas em ambientes sensíveis como o sistema de justiça, imagens artificiais podem desencadear reações emocionais negativas que comprometem a legitimidade institucional.

Do ponto de vista teórico, esta dissertação contribui para o aprofundamento dos estudos sobre confiança no ambiente digital ao demonstrar empiricamente a dissociação entre atitudes explícitas e implícitas frente a conteúdos visuais gerados por IAG. Ao retomar o modelo de Mayer, Davis e Schoorman (1995), foi possível confirmar que atributos clássicos da confiança – competência, integridade e benevolência – continuam sendo associados prioritariamente a representações autênticas. Além disso, os resultados dialogam com a literatura de Greenwald *et al.* (1998) e Rydell e Mackie (2008), reforçando a ideia de que julgamentos inconscientes podem divergir significativamente das avaliações conscientes, configurando um quadro de ambivalência cognitiva.

No campo prático, os resultados oferecem subsídios importantes para a discussão sobre o uso ético e responsável da IA em contextos institucionais. No sistema de justiça, por exemplo, a forte preferência implícita por imagens reais sugere que o emprego de representações artificiais sem a devida transparência pode comprometer a

legitimidade e a credibilidade das instituições. Do mesmo modo, em campanhas públicas e na comunicação jurídica, o reconhecimento de que a IA ainda se apresenta com percepções de manipulação reforça a necessidade de critérios rigorosos de rotulagem, transparência e *accountability*. Para seara do direito digital, os dados empíricos indicam que a regulação deve considerar os aspectos objetivos da produção de conteúdo, além dos efeitos subjetivos da confiança, que operam em níveis conscientes e inconscientes. Assim, esta dissertação mostra-se um contributo tanto para o avanço acadêmico da área quanto para a formulação de práticas e políticas públicas capazes de mitigar riscos e ampliar a legitimidade no uso de tecnologias emergentes.

Como possíveis implicações práticas, aplicadas a dois contextos institucionais concretos do Ministério Público, exemplificam-se: (i) a produção de campanhas institucionais do Ministério Público; e (ii) a utilização de simulações visuais no plenário do Tribunal do Júri. Em ambos os casos, os dados obtidos e analisados indicaram que o uso de pessoas reais reforça significativamente a percepção de credibilidade e confiança – elementos essenciais à legitimação do discurso institucional. Em campanhas de conscientização pública, a substituição de rostos humanos por avatares criados por IA pode ser contraproducente, desencadeando desconfiança cognitiva mesmo que a imagem seja esteticamente convincente. Da mesma forma, no plenário do júri, a adoção de reconstituições animadas por IA sem transparência, laudo técnico e controle do contraditório pode ser percebida como manipulação, comprometendo a eficácia argumentativa do operador do direito (tanto o Promotor de Justiça quanto ao Advogado). Dessa forma, os resultados empíricos sustentam a recomendação de que o uso da IAG em comunicações institucionais e atividades processuais deve ser criterioso, transparente e subsidiário, com prioridade para representações visuais que evoquem maior confiança junto ao público e ao corpo de jurados.

Noutro giro, é possível refletir sobre o papel da educação e da comunicação no contexto de uma sociedade amplamente mediada pelas tecnologias digitais. McLuhan e Strate (2008) sugerem que é necessário um novo tipo de alfabetização, uma “alfabetização midiática”, que vá além da simples decodificação de conteúdos e permita aos indivíduos compreenderem os meios de comunicação como ambientes que moldam suas percepções e comportamentos. Os referidos autores afirmam que o desafio contemporâneo não se resume

a entender o conteúdo, mas, sim, estudar a mídia por ela mesma, reconhecendo os impactos profundos que os meios têm sobre as estruturas sociais e culturais (Strate, 2008).

Postman (2000) também alerta para a necessidade de educar as novas gerações para entenderem criticamente os meios de comunicação, haja vista que a educação tradicional, centrada no texto escrito, não é mais suficiente para preparar os indivíduos para um mundo onde as imagens, os sons e as interações digitais dominam o cenário comunicacional. A abordagem ecológica da mídia, portanto, exige uma mudança de paradigma na educação, que precisa reconhecer a centralidade das novas tecnologias e ensinar os alunos a navegar por esses ambientes com criticidade e responsabilidade.

Por outro lado, o texto de Yigael (2011) aprofunda a discussão ao introduzir uma crítica ontológica e epistemológica fundamental à IA. O autor chama atenção para a diferença estrutural entre organismos vivos e sistemas artificiais no que tange à capacidade de produzir significados, agir com intencionalidade e atribuir valor às experiências. Essas dimensões, segundo Yigael (2011), são constitutivas da cognição humana, pois envolvem o processamento de informações e vivências subjetivas, bem como engajamento existencial com o mundo. Ao passo que os organismos vivos orientam suas ações a partir de finalidades intrínsecas e contextos interpretativos, a IA apenas executa operações funcionais baseadas em padrões estatísticos, sem qualquer horizonte de sentido próprio. Essa ausência de interioridade torna-se particularmente relevante quando se observa a crescente tentativa de atribuir à IA atributos como “benevolência” ou “integridade”, os quais dependem, em última instância, da experiência ética situada e da consciência moral – condições que a IA não possui nem simula autenticamente.

Essa reflexão converge com as proposições de Simões e Caldeira (2025), que defendem a necessidade de harmonizar a inovação técnica com a integridade ética na criação digital, haja vista que o avanço das IAGs exige eficiência algorítmica, além de um real compromisso com valores humanos fundamentais, como autenticidade, responsabilidade e respeito à dimensão simbólica das interações. Segundo os pesquisadores, “a criação sintética deve ser acompanhada por uma reflexão ética contínua, capaz de equilibrar o poder criativo da máquina e a integridade da experiência humana” (Simões; Caldeira, 2025, p. 43-59).

Tal perspectiva alinha-se à defesa, desenvolvida nesta pesquisa, de que a adoção de imagens artificiais em instituições do sistema de justiça requer prudência e avaliação criteriosa de seus impactos sobre a confiança pública, por conseguinte, mais do que uma questão estética ou tecnológica, trata-se de um imperativo ético que busca assegurar que a transformação digital não erosione os fundamentos simbólicos que sustentam a legitimidade e a credibilidade das instituições democráticas.

Nesse sentido, os dados obtidos por meio do IAT e do EAAT revelam os limites da própria IA em alcançar a complexidade da confiança humana. A confiança, tal como abordada neste estudo, não é um simples reconhecimento de padrão visual, mas uma resposta afetiva e semântica a sinais percebidos como legítimos, coerentes e moralmente consistentes. A despeito do realismo técnico das imagens produzidas por modelos generativos, a ausência de um lastro existencial e de uma intencionalidade atribuível a um "outro" efetivamente consciente pode ser o que subjaz às reações de estranhamento ou ceticismo captadas implicitamente. Desse modo, as evidências desta dissertação se somam à crítica contemporânea sobre os limites da IA enquanto entidade cognitiva, reforçando a necessidade de abordagens que considerem a dimensão ética, afetiva e semântica das interações mediadas por IA.

7.1 LIMITAÇÕES E SUGESTÕES PARA PESQUISAS FUTURAS

Embora os resultados tenham se mostrado consistentes e em linha com as teorias e artigos científicos apresentados ao longo da pesquisa, é imperioso reconhecer algumas limitações metodológicas deste estudo. Em primeiro lugar, a amostra foi composta por participantes recrutados majoritariamente em ambientes digitais, por meio de convites em redes sociais e aplicativos de mensagens, o que a caracteriza como uma amostra não probabilística. Esse aspecto limita a generalização dos achados para a população em geral, ainda que o número final de respondentes tenha sido bem razoável para os escopos da pesquisa realizada. Além disso, a aplicação do EAAT envolve autorrelato, o que pode ser influenciado por vieses conscientes, como desejabilidade social ou pressões normativas, reduzindo a espontaneidade das respostas. No caso do IAT, ainda que o instrumento seja consolidado para captar associações automáticas, seu

desempenho também pode ser afetado por fatores externos, como fadiga, distração ou familiaridade prévia com testes cognitivos.

Outra limitação metodológica refere-se à ausência de um pré-teste das tarefas experimentais. Embora as instruções e os estímulos tenham sido elaborados com base em protocolos validados pelo *Project Implicit* (*Project Implicit*, 2025) e adaptados às recomendações de Greenwald *et al.* (1998), a realização prévia de um piloto poderia ter permitido a identificação de eventuais ambiguidades, dificuldades técnicas ou necessidade de ajustes na apresentação dos estímulos, especialmente na etapa inicial do IAT. A inexistência desse pré-teste não compromete os resultados obtidos, mas indica um ponto de aprimoramento importante para futuras pesquisas que desejem aumentar a robustez e o controle experimental das medidas implícitas e explícitas.

A pesquisa apresenta caráter transversal, uma vez que articula fundamentos da Comunicação Digital, Psicologia Cognitiva, Ciência de Dados e Direito, evidenciando que a percepção de confiança em ambientes mediados por tecnologia não pode ser compreendida de forma isolada por um único campo disciplinar. Essa transversalidade reforça a pertinência de abordagens interdisciplinares para analisar fenômenos sociotécnicos complexos, como o uso de imagens geradas por IAG em instituições públicas.

Do ponto de vista temporal, os achados refletem um momento específico da evolução da IAG, marcado pela rápida disseminação de ferramentas como *Midjourney*, *DALL·E* e modelos de linguagem avançados. As percepções identificadas — tanto implícitas quanto explícitas — são influenciadas por esse contexto e podem se modificar conforme a tecnologia se torna mais difundida, naturalizada ou regulamentada. Assim, a temporalidade constitui elemento central para a interpretação dos resultados e aponta para a necessidade de estudos longitudinais que examinem a evolução dessas percepções ao longo dos próximos anos.

Outra limitação do estudo refere-se à estratégia de aprofundamento estatístico adotada. Embora tenham sido aplicados testes inferenciais adequados ao delineamento da pesquisa, como o teste t pareado para comparação entre imagens reais e artificiais e o teste t para uma amostra para verificação do D-score em relação a zero, não foram exploradas análises estatísticas adicionais que poderiam

ampliar a compreensão da variabilidade individual dos resultados, tais como modelos mais complexos ou análises estratificadas por subgrupos. A inclusão desses procedimentos poderia oferecer maior refinamento interpretativo, especialmente no que se refere à magnitude e à distribuição dos efeitos observados, sem, contudo, comprometer a validade dos achados apresentados.

Diante dessas limitações, sugerem-se direções para pesquisas futuras. A uma, recomenda-se a replicação do estudo com amostras específicas do sistema de justiça – magistrados, jurados, promotores e advogados – a fim de verificar se os padrões identificados na população geral se mantêm em grupos diretamente envolvidos com a legitimidade institucional. Outra possibilidade é a realização de experimentos em contextos controlados, nos quais a transparência das imagens (rotulagem de "real" ou "gerada por IA") seja manipulada, permitindo avaliar seus efeitos diretos sobre a confiança. Ademais, seria relevante ampliar a investigação para outros tipos de estímulos visuais, como vídeos ou *deepfakes*, de modo a compreender se a dissociação entre atitudes explícitas e implícitas se reproduz em mídias mais dinâmicas. Por fim, estudos comparativos em diferentes culturas poderiam contribuir para identificar se os resultados observados refletem predisposições universais ou se estão modulados por fatores socioculturais específicos.

Além das agendas de continuidade já mencionadas, os achados deste estudo também abrem espaço para investigações em outras áreas que vivenciam processos de mediação algorítmica e dependem da construção de confiança. Setores como saúde digital, educação mediada por tecnologias, segurança pública, jornalismo, publicidade, comércio eletrônico e serviços financeiros já utilizam imagens sintéticas, avatares e interfaces baseadas em IA em escala crescente. A aplicação combinada de medidas implícitas e explícitas nesses contextos pode revelar como diferentes públicos atribuem confiança a sistemas automatizados, como percebem autenticidade em representações digitais e de que maneira reagem à crescente substituição de elementos humanos por recursos artificiais. A replicação do presente modelo metodológico em tais domínios pode contribuir para compreender a aceitabilidade social da IA, bem como os limites éticos e comunicacionais de sua adoção.

REFERÊNCIAS

REFERÊNCIAS

REFERÊNCIAS

ADOLPHS, Ralph; ANDERSON, David J. **The neuroscience of emotion**: A new synthesis. Princeton: Princeton University Press, 2018.

AGGARWAL, Pankaj; MCGILL, Ann L. Is that car smiling at me? Schema congruity as a basis for evaluating anthropomorphized products. **Journal of Consumer Research**, v. 34, n. 4, p. 468-479, 2007.

AKERLOF, George. A. The market for "lemons": Quality uncertainty and the market mechanism. **The Quarterly Journal of Economics**, v. 84, n. 3, p. 488-500, 1970.

ALVES, Gabriela Mesquita Martins. **A influência da inteligência artificial no processo de decisão de compra online do consumidor**. 2023. Dissertação (Mestrado em Direção Comercial e *Marketing*) – Instituto Superior de Administração e Gestão, Porto, 2023.

BARONE, Michael J.; TAYLOR, Valerie A.; URBANY, Joel E. Advertising signaling effects for new brands: the moderating role of perceived brand differences. **Journal of Marketing Theory and Practice**, [S. l.], v. 13, n. 1, p. 1-13, 2005. Disponível em: https://doi.org/10.1080/10696679.2005.11658534. Acesso em: 16 out. 2025.

BAUER, Raymond A. Consumer behavior as risk taking. *In*: HANCOCK, Robert S. (Ed.). **Dynamic Marketing for a Changing World**, Proceedings of the 43rd. Conference of the American Marketing Association, p. 389-398, 1960.

BBC NEWS. Google apologises for photos app's racist blunder. **BBC News**, 1 jul. 2015. Disponível em: https://www.bbc.com/news/technology-33347866. Acesso em: 6 dez. 2024.

BILAL, Muhammad; ZHANG, Yunfeng; CAI, Shukai; AKRAM, Umair; HALIBAS, Alrence. Artificial intelligence is the magic wand making customer-centric a reality! An investigation into the relationship between consumer purchase intention and consumer engagement through affective attachment. **Journal of Retailing and Consumer Services**, v. 77, 2024. Disponível em: https://doi.org/10.1016/j.jretconser.2023.103674. Acesso em: 19 jun. 2024.

BOMMASANI, Rishi; NARAYANAN, Deepak; CHUANG, Kelly; WISTER, William; KRUEGER, Gretchen; BROWN, Peter. **On the opportunities and risks of foundation models**. Preprint (arXiv), 2021. Disponível em: https://arxiv.org/abs/2108.07258. Acesso em: 10 fev. 2025.

BOSTROM, Nick; YUDKOWSKY, Eliezer. The ethics of artificial intelligence. *In*: FRANKISH, Keith; RAMSEY, William M. (Eds.). **The Cambridge handbook of artificial intelligence**. Cambridge: Cambridge University Press, 2014. p. 316-334.

BUCHER, Taina. **If... then:** Algorithmic power and politics. Oxford: Oxford University Press, 2018.

CHANG, Yu-Wen; CHIEN, Shih-Yi; CHAN, Yao-Cheng; TSAO, Ching-Chih. Human-Robot Interaction in E-Commerce: The Role of Personality Traits and Chatbot Mechanisms - A Neuromarketing Research. **ACM/IEEE International Conference on Human-Robot Interaction**, p. 312-316, 2024. Disponível em: https://doi.org/10.1145/3610978.3640742. Acesso em: 19 jun. 2024.

CIECHANOWSKI, Leon; PRZEGALINSKA, Aleksandra; MAGNUSKI, Mikolaj; GLOOR, Peter. **In the shades of the uncanny valley:** An experimental study of human–chatbot interaction. Future Generation Computer Systems, 2018. Disponível em: https://doi.org/10.1016/j.future.2018.01.055. Acesso em: 15 mar. 2025.

CHIEN, Sung-En; CHU, Li; LEE, Hsing-Hao; YANG, Chin-Chun; LIN, Fo-Hui; YANG, Pei-Ling; WANG, Te-Mei; YEH, Su-Ling. Age difference in perceived ease of use, curiosity, and implicit negative attitude toward robots. **ACM Transactions on Human-Robot Interaction**, v. 8, n. 2, p. 9-14, 2019.

COLEMAN, James S. **Foundations of Social Theory**. Cambridge, MA: Belknap Press of Harvard University Press, 1990.

COLQUITT, Jason A.; SCOTT, Brent A.; LEPINE, Jeffery A. Trust, Trustworthiness, and Trust Propensity: A Meta-Analytic Test of Their Unique Relationships With Risk Taking and Job Performance. **Journal of Applied Psychology**, v. 92, n. 4, p. 909-927, 2007.

COMITÊ GESTOR DA INTERNET NO BRASIL. **Pesquisa sobre o uso das tecnologias de informação e comunicação no Brasil:** TIC Domicílios 2022. São Paulo: CGI.br, 2022. Disponível em: https://www.cgi.br/pesquisa/tic/. Acesso em: 8 mar. 2025.

COMITÊ GESTOR DA INTERNET NO BRASIL. **Pesquisa sobre o uso das tecnologias de informação e comunicação nos domicílios brasileiros:** TIC Domicílios 2023 [livro eletrônico] = Survey on the use of information and communication technologies in Brazilian households: ICT Households 2023 / [editor] Núcleo de Informação e Coordenação do Ponto BR. 1. ed. São Paulo: Comitê Gestor da Internet no Brasil, 2024. Disponível em: https://cetic.br/media/docs/publicacoes/2/20241104102822/tic_domicilios_2023_livro_eletronico.pdf. Acesso em: 24 mar.2025

COMITÊ GESTOR DA INTERNET NO BRASIL. TIC Domicílios 2023. **We Are Social**; Meltwater. Digital 2024: Brazil. Disponível em: https://datareportal.com/reports/digital-2024-brazil. Acesso em: 24 mar.2025

COTTER, Karen. Playing the visibility game. **New Media & Society**, v. 21, n. 4, p. 895-913, 2019.

COTTEN, Sheila R.; GUPTA, Subhendu S. Characteristics of online and offline health information seekers and factors that discriminate between them. **Social Science & Medicine**, v. 59, n. 9, p. 1795-1806, 2004. DOI: 10.1016/j.socscimed.2004.02.020.

CRAWFORD, Kate. **Atlas of AI:** Power, Politics, and the Planetary Costs of Artificial Intelligence. New Haven: Yale University Press, 2021.

DAMASIO, António R. **Descartes' error:** emotion, reason, and the human brain. New York: G.P. Putnam, 1994.

DAVIDSON, Richard J.; EKMAN, Paul; SARON, Clifford D.; SENULIS, Joseph A.; FRIESEN, Wallace V. Approach-withdrawal and cerebral asymmetry: emotional expression and brain physiology. I. **Journal of Personality and Social Psychology**, v. 58, n. 2, p. 330-341, 1990.

DAL VERNE, Tommaso De Mari Casareto. **Artificial Intelligence, Neuroscience and Emotional Data. What Role for Private Autonomy in the Digital Market?** Erasmus Law Review, 2023. Disponível em: https://arxiv.org/abs/2309.10776. Acesso em: 15 jan. 2025.

DE HOUWER, Jan; GAWRONSKI, Bertram; BARNES-HOLMES, Dermot. What do implicit measures measure? Theory and application in the study of attitudes. *In*: DE HOUWER, Jan; MOORS, Agnes. (org.). **Cognitive science perspectives on personality and emotion**. New York: Psychology Press, 2013. p. 155-194.

DEVELLIS, Robert F. **Scale development:** theory and applications. 4. ed. Los Angeles: Sage Publications, 2017.

DIRKS, Kurt T.; FERRIN, Donald L. Trust in leadership: Meta-analytic findings and implications for research and practice. **Journal of Applied Psychology**, v. 87, n. 4, p. 611-628, 2002.

DIETVORST, Berkeley J.; BHARTI, Soaham. **People reject algorithms in uncertain decision domains because they have diminishing sensitivity to forecasting error.** University of Chicago Booth School of Business, 2019. Disponível em: https://ssrn.com/abstract=3424158. Acesso em: 15 jan. 2025.

EBBEN, Maureen; BULL, Elizabeth. **Constructing Authenticity:** Social Media Influencers and the Shaping of Online Identity. Disponível em: www.intechopen.com. Acesso em: 19 jun. 2024.

EKMAN, Paul. An argument for basic emotions. **Cognition & Emotion**, v. 6, n. 3-4, p. 169-200, 1992.

ELLIS, Sean; BROWN, Morgan. **Hacking Growth**: How Today's Fastest-Growing Companies Drive Breakout Success. New York: Crown Business, 2017.

SWAIT, Joffre; ERDEM, Tülin. **Characterizing brand effects on choice and choice set formation under uncertainty.** Edmonton: University of Alberta; New York: New York University, Stern School of Business, 2006. Disponível em: https://ssrn.com/abstract=965472. Acesso em: 15 fev. 2025.

FAST, Ethan; HORVITZ, Eric. Long-term trends in the public perception of artificial intelligence. *In*: **Proceedings of the 31st AAAI Conference on Artificial Intelligence (AAAI-17)**. San Francisco, CA: AAAI Press, 2017, p. 963-969. Disponível em: https://www.aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14972. Acesso em: 15 fev. 2025.

FERREIRA, Sérgio Rodrigo da Silva. O que é (ou o que estamos chamando de) 'Colonialismo de Dados'? **PAULUS: Revista de Comunicação Da FAPCOM**, v. 5, n. 10, 2021. Disponível em: https://doi.org/10.31657/rcp.v5i10.458. Acesso em: 19 jun. 2024.

FIEDLER, Klaus; MESSNER, Claude; BLUEMKE, Matthias. Unresolved problems with the “I”, the “A”, and the “T”. **European Review of Social Psychology**, v. 17, p. 74-147, 2006.

FIETTA, Valentina; ZECCHINATO, Francesca; DI STASI, Brigida; POLATO, Mirko; MONARO, Merylin. **Dissociation Between Users’ Explicit and Implicit Attitudes Toward Artificial Intelligence:** An Experimental Study. IEEE Transactions on Human-Machine Systems, 2021.

FLANAGIN, Andrew J.; METZGER, Miriam J. Perceptions of Internet information credibility. **Journalism and Mass Communication Quarterly**, v. 77, n. 3, p. 515-540, 2000.

FLANAGIN, Andrew J.; METZGER, Miriam J. **Digital Media and Youth:** Unparalleled Opportunity and Unprecedented Responsibility. Cambridge, MA: The MIT Press, 2008. p. 5-28. DOI: 10.1162/dmal.9780262562324.005.

METZGER, Miriam J.; FLANAGIN, Andrew J. Credibility and trust of information in online environments: the use of cognitive heuristics. **Journal of Pragmatics**, v. 59, p. 210-220, 2013. Disponível em: https://doi.org/10.1016/j.pragma.2013.07.012. Acesso em: 20 jun. 2024.

TSENG, Shawn; FOGG, B. J. Credibility and computing technology. **Communications of the ACM**, v. 42, n. 5, p. 39-44, maio 1999. Disponível em: https://doi.org/10.1145/301353.301402. Acesso em: 20 jun. 2024.

FOGG, Brian Jeffrey. **Persuasive technology:** using computers to change what we think and do. Amsterdam: Morgan Kaufmann Publishers, 2003.

FOGG, Brian Jeffrey; SOOHOO, Cheryl; DANIELSON, David; MARABLE, Leslie; STANFORD, Joseph; TAUBER, Ellen. How do users evaluate the credibility of Web sites? A study with over 2,500 participants. *In*: Conference On Designing For User Experiences, 2003, San Francisco. **Proceedings...** New York: ACM, 2003. p. 1-15. DOI: 10.1145/997078.997097.

FORTUNATI, Leopoldina; O’SULLIVAN, John. Convergence crosscurrents: Analog in the digital and digital in the analog. **The Information Society**, v. 36, n. 3, p. 160-166, 2020. Disponível em: https://doi.org/10.1080/01972243.2020.1737608. Acesso em: 20 jun. 2024.

FREEPIK. **Imagem** [fotografia digital]. Freepik, 2025. Disponível em: https://www.freepik.com. Acesso em: 30 jan. 2025.

FLUSSER, Vilém. **Filosofia da Caixa Preta**: ensaios para uma filosofia da fotografia. Rio de Janeiro: Relume Dumará, 2002.

FLORIDI, Luciano. Digital time: latency, real-time, and the onlife experience of everyday time. **Philosophy & Technology**, v. 34, p. 407-412, 2021. Disponível em: https://ssrn.com/abstract=3935638. Acesso em: 16 out. 2025.

FRITH, Chris D. **Other minds:** The octopus, the sea, and the deep origins of consciousness. London: Yale University Press, 2020.

FULCHER, E.; DEAN, A.; TRUFIL, G. Neurosense and packaging: understanding consumer evaluations using implicit technology. *In:* BURGESS, Paul; VILAS-BOAS, Ana Alice (org.). **Integrating the packaging and product experience in food and beverages.** Cambridge: Woodhead Publishing, 2016. (Woodhead Publishing Series in Food Science, Technology and Nutrition). p. 121-138. Disponível em: https://doi.org/10.1016/B978-0-08-100356-5.00006-1. Acesso em: 8 mar. 2025.

GAMBETTA, Diego. **Can we trust trust?** Oxford: University Press, 2000.

GAWRONSKI, Bertram; BODENHAUSEN, Galen V. Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change. **Psychological Bulletin**, Washington, DC, v. 132, n. 5, p. 692-731, 2006. DOI: 10.1037/0033-2909.132.5.692.

GARCIA, Ana Cristina Bicharra. Ética e Inteligência Artificial. **Computação Brasil**, São Paulo, v. 16, n. 11, p. 14-22, nov. 2020.

GARCIA, Megan. Racist in the machine: the disturbing implications of algorithmic bias. **World Policy Journal**, v. 33, n. 4, p. 111-117, 2016.

GILES, Jim. Internet encyclopaedias go head to head. **Nature**, v. 438, n. 7070, p. 900-901, 2005. Disponível em: https://www.nature.com/articles/438900a. Acesso em: 8 mar. 2025.

GILLE, Felix; JOBIN, Anna; IENCA, Marcello. What we talk about when we talk about trust: Theory of trust for AI in healthcare. **Intelligence-Based Medicine**, [s.l.], v. 1-2, p. 100001, nov. 2020. DOI: 10.1016/j.ibmed.2020.100001.

GILLESPIE, Tarleton. The relevance of algorithms. *In:* GILLESPIE, Tarleton; BOCZKOWSKI, Pablo; FOOT, Kirsten (org.). **Media**

technologies: essays on communication, materiality, and society. Cambridge, MA: MIT Press, 2013. p. 167-194. Disponível em: https://www.tarleton-gillespie.org/wp-content/uploads/2016/07/Gillespie_2013_The-Relevance-of-Algorithms.pdf. Acesso em: 26 jan. 2025.

GILLESPIE, Tarleton. **Custodians of the Internet:** Platforms, Content Moderation, and the Hidden Decisions That Shape Social Media. Yale University Press, 2018.

GILLESPIE, Tarleton. **The Relevance of Algorithms**. Cambridge: MIT Press, 2018.

GREENWALD, Anthony G.; BANAJI, Mahzarim R. Implicit social cognition: Attitudes, self-esteem, and stereotypes. **Psychological Review**, v. 102, n. 1, p. 4-27, 1995.

GREENWALD, Anthony G.; MCGHEE, Debbie E.; SCHWARTZ, Jordan L. K. Measuring individual differences in implicit cognition: The Implicit Association Test. **Journal of Personality and Social Psychology**, v. 74, n. 6, p. 1464-1480, 1998.

GREENWALD, Anthony G.; NOSEK, Brian A.; BANAJI, Mahzarin R. Understanding and using the Implicit Association Test: I. An improved scoring algorithm. **Journal of Personality and Social Psychology**, Washington, DC, v. 85, n. 2, p. 197-216, 2003. DOI: 10.1037/0022-3514.85.2.197. Disponível em: https://doi.org/10.1037/0022-3514.85.2.197. Acesso em: 26 jan. 2025.

GOODFELLOW, Ian; BENGIO, Yoshua; COURVILLE, Aaron. **Deep Learning**. Cambridge: MIT Press, 2016.

GUZMAN, Andrea L. Ontological boundaries between humans and machines: A response to the call for a critical AI studies. **Communication Theory**, v. 28, n. 3, p. 346-365, 2018. DOI: 10.1093/ct/qty010.

HADDON, Leslie; SILVERSTONE, Roger. Information and Communication Technologies and Everyday Life: Individual and Social Dimensions. *In*: DUCATEL, Ken; WEBSTER, Juliet; HERRMAN, Werner (Eds.). **The Information Society in Europe**: Work and Life in an Age of Globalization. Lanham, MD: Rowman and Littlefield, 2000. p. 233-258.

HANSON, David; OLNEY, Andrew; PEREIRA, Ismar A.; ZIELKE, Marge. Upending the Uncanny Valley. *In*: **Proceedings of the Association for the Advancement of Artificial Intelligence Open Interaction**, 2005.

HARRIS, Paul R.; SILLENCE, Elizabeth; BRIGGS, Pam. Perceived threat and corroboration: key factors that improve a predictive model of trust in internet-based health information and advice. **Journal of Medical Internet Research**, v. 13, n. 3, p. e51, 2011. DOI: 10.2196/jmir.1821.

HOBBES, Thomas. **Leviatã ou matéria, forma e poder de uma república eclesiástica e civil**. São Paulo: Abril Cultural, 1979.

HSIAO, Kuo-Lun; CHEN, Chien-Hsiung. What drives in-app purchase intention for mobile games? An examination of perceived values and loyalty. **Electronic Commerce Research and Applications**, v. 16, p. 18-29, 2016.

IHDE, Don. **Technology and the lifeworld:** from garden to earth. Bloomington: Indiana University Press, 1990.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **O impacto transformador da inteligência artificial na geração de conteúdo e imagem:** uma jornada evolutiva. Rio de Janeiro: IBGE Digital, 2025. Disponível em: https://www.ibge.gov.br/ibge-digital/38980-o-impacto-transformador-da-inteligencia-artificial-na-geracao-de-conteudo-e-imagem-uma-jornada-evolutiva.html. Acesso em: 26 jan. 2025.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA. **84,9% das indústrias de médio e grande porte utilizaram tecnologia digital avançada**. Agência de Notícias, 2025. Disponível em: https://agenciadenoticias.ibge.gov.br/agencia-noticias/2012-agencia-de-noticias/noticias/37973-84-9-das-industrias-de-medio-e-grande-porte-utilizaram-tecnologia-digital-avancada. Acesso em: 26 jan. 2025.

ISRAFILZADE, Khalil; SADILI, Nuraddin. Beyond interaction: Generative AI in conversational marketing - foundations, developments, and future directions. **Journal of Life Economics**, v. 11, n. 1, p. 13-29, 2024. Disponível em: https://doi.org/10.15637/jlecon.2294. Acesso em: 19 jun. 2024.

KAHNEMAN, Daniel; FREDERICK, Shane. A Model of Heuristic Judgment. *In*: HOLYOAK, Keith J.; MORRISON, Robert G. **The Cambridge Handbook of Thinking and Reasoning**. Cambridge University Press, 2005. p. 267-293

KANG, Hyunjin.; LOU, Chen. AI agency vs. human agency: Understanding human-AI interactions on TikTok and their implications for user engagement. **Journal of Computer-Mediated Communication**, v. 27, n. 5, 2022. Disponível em: https://doi.org/10.1093/jcmc/zmac014. Acesso em: 19 jun. 2024.

KANT, Immanuel. **Fundamentação da metafísica dos costumes**. São Paulo: Martins Fontes, 2003.

KLAYMAN, Joshua; HA, Young-Won. Confirmation, disconfirmation, and information in hypothesis testing. **Psychological Review**, v. 94, n. 2, p. 211-228, 1987.

KOTLER, Philip. KARTAJAYA, Hermawan; SETIAWAN, Iwan. **Marketing 5.0 [recurso eletrônico]:** tecnologia para a humanidade. Rio de janeiro: Sextante, 2021.

KRIEGESKORTE, Nikolaus; DOUGLAS, Pamela K. Cognitive computational neuroscience. **Nature Neuroscience**, v. 21, n. 9, p. 1148-1160, 2018. Disponível em: https://pmc.ncbi.nlm.nih.gov/articles/PMC6706072/. Acesso em: 05 mar. 2025.

LAMB, Luís C. Ética em IA e IA ética: prolegômenos e estudo de casos significativos. **Revista USP**, São Paulo, n. 141, p. 107-120, abr./maio/jun. 2024.

LANIER, Jaron. **Dez argumentos para você deletar agora suas redes sociais**. São Paulo: Intrínseca, 2018.

LEE, Nick; BRODERICK, Amanda J.; CHAMBERLAIN, Laura. What is 'neuromarketing'? A discussion and agenda for future research. **International Journal of Psychophysiology**, v. 63, n. 2, p. 199-204, 2007. DOI: 10.1016/j.ijpsycho.2006.03.007.

LEGG, Shane; HUTTER, Marcus. A collection of definitions of intelligence. **Frontiers in Artificial Intelligence and Applications**, v. 157, p. 17, 2007.

LEMMERS-JANSEN, Imke; VELTHORST, Eva; FETT, Anne-Kathrin. The social cognitive and neural mechanisms that underlie social functioning in individuals with schizophrenia - a review. **Transl Psychiatry**, v. 13, n. 1: 327, Oct 21, 2023. DOI: 10.1038/s41398-023-02593-1. PMID: 37865631; PMCID: PMC10590451.

LEVENSON, Robert. Human emotion: a functional view. *In*: EKMAN, Paul; DAVIDSON, Richard (eds.). **The nature of emotion:** fundamental questions. New York: Oxford University Press, 1994. p. 123-126.

LIKERT, Rensis. A technique for the measurement of attitudes. **Archives of Psychology**, n. 140, p. 1-55, 1932.

LIMA, Edirlei Soares de; FEIJÓ, Bruno. Artificial Intelligence in Human-Robot Interaction. *In*: AYANOĞLU, H.; DUARTE, E. (ed.). **Emotional Design in Human-Robot Interaction**. Cham: Springer, 2019. (Human–Computer Interaction Series). p. 187-199. DOI: 10.1007/978-3-319-96722-6_11.

MADEIRA, Afonso Celso Magalhães; NEVES, Barbara Coelho; BRANCO, Daniel de Jesus Barcoso Cautela. O Uso da Inteligência Artificial Aplicada ao Marketing Digital**. Journal of Digital Media & Interaction**, v. 3, n. 8, p. 95-111, 2020.

MARCUS, Gary; DAVIS, Ernest. **Rebooting AI:** building artificial intelligence we can trust. New York: Vintage, 2019.

MARTINS, Maura. Conversando com robôs: como a IA ajuda no relacionamento com clientes? **TecMundo**, 26 abr. 2024. Disponível em: https://www.tecmundo.com.br/mercado/282248-conversando-robos-ia-ajuda-relacionamento-com-clientes.htm. Acesso em: 16 out. 2025.

MARTÍN-BARBERO, Jesús. **Dos meios às mediações:** comunicação, cultura e hegemonia. Rio de Janeiro: UFRJ, 2003.

MAYER, Roger C.; DAVIS, James H.; SCHOORMAN, F. David. An integrative model of organizational trust. **The Academy of Management Review**, v. 20, n. 3, p. 709-734, jul. 1995. Disponível em: http://www.jstor.org/stable/258792. Acesso em: 24 jan. 2025.

MCKNIGHT, D. Harrison; CHERVANY, Norman L. What Trust Means in E-Commerce Customer Relationships: An Interdisciplinary Conceptual Typology. **International Journal of Electronic Commerce**, v. 6, n. 2, p. 35-69, 2001.

MCLUHAN, Marshall. **Os meios de comunicação como extensões do homem**. Tradução de Décio Pignatari. 1. ed. São Paulo: Cultrix, 1964.

MCLUHAN, Marshall; STRATE, Lance. **O que é mídia?** Epistemologia da comunicação. São Paulo: Paulus, 2008.

MCMILLAN, Sally J.; MACIAS, Wendy. Strengthening the safety net for online seniors. **Journal of Health Communication**, v. 13, n. 8, p. 778-792, 2008. DOI: 10.1080/10810730802487471.

MEDLOCK, Stephanie; ETTEMA, Rosalie Gerarda Adriana; VAN DEKKER, Anita; KAPTEIN, Maurits; KROON, Annemarie Adriana; VAN BUNGE, Bart; POELMAN, Johan; KONIGS, Maurits; SMITH, Steffie; EBERS, Sarah; BONTEN, Tobias Nicolaas; VAN DER BOOM, Hilde; TATES, Kiek; VAN WEERT, Julia Catharina Maria. Health information-seeking behavior of seniors who use the internet: a survey. **Journal of Medical Internet Research**, v. 17, n. 1, p. e10, 2015. DOI: 10.2196/jmir.3749.

MEDINA, Marco; FERTIG, Cristina. **Algoritmos e Programação:** Teoria e Prática. 2. ed. São Paulo: Novatec Editora, 2022.

MILLISECOND SOFTWARE. **Inquisit by Millisecond:** Reaction time software for cognitive and psychological experiments. Seattle, WA: Millisecond Software, 2023. Disponível em: https://www.millisecond.com. Acesso em: 24 jan. 2025.

MOLM, Linda D.; TAKAHASHI, Nobuyuki; PETERSON, Gretchen. Risk and Trust in Social Exchange: An Experimental Test of a Classical Proposition. **American Journal of Sociology**, v. 105, n. 5, p. 1396-1427, 2000.

MONTAG, Christian.; ALI, Raian; DAVIS, Kenneth L. Affective neuroscience theory and attitudes towards artificial intelligence. **AI and Society**, 2024b. Disponível em: https://doi.org/10.1007/s00146-023-01841-8. Acesso em: 19 jun. 2024.

MORGAN, Robert M.; HUNT, Shelby D. The commitment-trust theory of relationship marketing. **Journal of Marketing**, v. 58, n. 3, p. 20-38, 1994.

MORI, Masahiro. The uncanny valley. Translation by Karl F. MacDorman and Norri Kageki. **IEEE Robotics & Automation Magazine**, v. 19, n. 2, p. 98-100, June 2012. DOI: 10.1109/MRA.2012.2192811. Disponível em: https://ieeexplore.ieee.org/document/6213128. Acesso em: 16 out. 2025.

MORI, Masahiro; MACDORMAN, Karl F.; KAGEKI, Norri. The Uncanny Valley [From the field]. **IEEE Robotics & Automation Magazine**, v. 19, n. 2, p. 98-100, 2012.

MORIUCHI, Emi. Okay, Google!: an empirical study on voice assistants on consumer engagement and loyalty. **Psychology & Marketing**, v. 36,

n. 5, p. 489-501, 2019. DOI: 10.1002/mar.21192. Disponível em: https://doi.org/10.1002/mar.21192. Acesso em: 19 jun. 2024.

MOROZOV, Evgeny. **Big Tech:** a ascensão dos dados e a manipulação da liberdade. São Paulo: Ubu Editora, 2022.

MOSSERI, Adam. **News Feed Ranking in Three Minutes Flat:** How Does News Feed Work? [S. l.: s. n.], 2018. Disponível em: https://newsroom.fb.com/news/2018/05/inside-feed-news-feed-ranking/. Acesso em: 1 nov. 2024. https://doi.org/10.18254/s0000057-4-1.

MULHOLLAND, Caitlin; FRAJHOF, Isabella Z. Entre as leis da robótica e a ética: regulação para o adequado desenvolvimento da Inteligência Artificial. *In*: MULHOLLAND, Caitlin; FRAZÃO, Ana (Coords.). **Inteligência Artificial e Direito:** ética, regulação e responsabilidade. São Paulo: Thomson Reuters, 2019, v. 1, p. 1-21.

NICOLAU, Mark. Artificial Intelligence - Friend or Foe. **IPI Letters**, p. 34-41, 2024. Disponível em: https://doi.org/10.59973/ipil.54. Acesso em: 19 jun. 2024.

NOBLE, Safiya. **Algorithms of oppression:** How search engines reinforce racism. New York: NYU Press, 2018.

NOSEK, Brian A.; SMYTH, Frederick L. A multitrait-multimethod validation of the Implicit Association Test: Implicit and explicit attitudes are related but distinct constructs. **Experimental Psychology, Göttingen**, v. 54, n. 1, p. 14-29, 2007.

NOSEK, Brian A.; SMYTH, Frederick L. Implicit social cognitions predict sex differences in math engagement and achievement. **American Educational Research Journal**, v. 48, n. 5, p. 1125-1156, out. 2011. DOI: 10.3102/0002831211410683. Disponível em: https://doi.org/10.3102/0002831211410683. Acesso em: 19 jun. 2024.

OH, Changhoon; CHOI, Jinhan; LEE, Sungwoo; PARK, SoHyun; KIM, Daeryong; SONG, Jungwoo; KIM, Dongwhan; LEE, Joonhwan; SUH, Bongwon. Understanding User Perception of Automated News Generation System. *In*: **CHI Conference on Human Factors in Computing Systems (CHI 2020)**, 2020, Honolulu, HI, USA. Proceedings [...]. New York: ACM, 2020. p. 1-10. Disponível em: http://dx.doi.org/10.1145/3313831.3376811. Acesso em: 9 mar. 2025.

OPENAI. **Resposta gerada pelo modelo de linguagem ChatGPT**. Disponível em: https://www.openai.com/chatgpt. Acesso em: 2 jul. 2024.

PALETTA, Francisco Carlos; COSTA DO LAGO, Jader Jaime. Plataformização e o uso da informação para a criação de estímulos de consumo. **e-Ciências da Informação**, v. 12, n. 1, 2022. DOI: 10.15517/eci.v12i1.48095.

PANKSEPP, Jaak. **Affective neuroscience:** the foundations of human and animal emotions. Oxford: Oxford University Press, 1998.

PANKSEPP, Jaak. Cross-species affective neuroscience decoding of the primal affective experiences of humans and related animals. **PLoS ONE**, v. 6, n. 9, p. e21236, 2011. Disponível em: https://doi.org/10.1371/journal.pone.0021236. Acesso em: 19 jun. 2024.

PASK, Gordon. **Conversation Theory:** Applications in Education and Epistemology. Amsterdam: Elsevier, 1976.

PASQUALE, Frank. **The Black Box Society:** the secret algorithms that control money and information. Cambridge, Massachusetts; London, England: Harvard University Press, 2015. ISBN 978-0-674-36827-9. Disponível em: https://doi.org/10.4159/harvard.9780674736061. Acesso em: 16 out. 2025.

PAULHUS, Delroy L. Measurement and control of response bias. *In*: ROBINSON, John P.; SHAVER, Phillip R.; WRIGHTSMAN, Lawrence S. (eds.). **Measures of personality and social psychological attitudes**. San Diego: Academic Press, 1991. p. 17-59.

PAVLOU, Paul A. Consumer Acceptance of Electronic Commerce: Integrating Trust and Risk with the Technology Acceptance Model. **International Journal of Electronic Commerce**, v. 7, n. 3, p. 101-134, 2003.

PEW RESEARCH CENTER. **Internet/Broadband Fact Sheet**. Washington, D.C.: Pew Research Center, 2021. Disponível em: https://www.pewresearch.org/internet/fact-sheet/internet-broadband/. Acesso em: 8 mar. 2025.

PINTO, Henrique Alves. A utilização da inteligência artificial no processo de tomada de decisões: por uma necessária accountability. **Revista de Informação Legislativa**, Brasília, v. 57, n. 225, p. 43-60, 2020. Disponível em:

https://www2.senado.leg.br/bdsf/bitstream/handle/id/596785/001178601_RIL_v.57_n.225_p.043-060.pdf. Acesso em: 6 dez. 2024.

POSTMAN, Neil. The humanism of media ecology. **Proceedings of the Media Ecology Association**, v. 1, p. 10-16, 2000.

POTDAR, Vidyasagar; JOSHI, Suman; HARISH, Ranjit; BASKERVILLE, Richard; WONG THONGTHAM, Phayung. A process model for identifying online customer engagement patterns on Facebook brand. **Information Technology & People**, v. 31, n. 2, p. 595-614, 2018.

PROJECT IMPLICIT. **About Us**. Disponível em: https://implicit.harvard.edu/implicit. Acesso em: 21 abr. 2025.

PSYCHOLOGY. **Associative Networks - IResearchNet**. 2020. Disponível em: http://psychology.iresearchnet.com/social-psychology/social-cognition/associative-networks/. Acesso em: 24 jan. 2025.

RICHESON, Jennifer A.; BAIRD, Abigail A.; GORDON, Heather L.; HEATHERTON, Todd F.; WYLAND, Carrie L.; TRAWALTER, Sophie; SHELTON, J. Nicole. An fMRI investigation of the impact of interracial contact on executive function. **Nature Neuroscience**, v. 6, n. 12, p. 1323-1328, 2003.

RÖDEL, Christian; STADLER, Sebastian; MESCHTSCHERJAKOV, Alexander; TSCHELIGI, Manfred. Towards autonomous cars: The effect of autonomy levels on acceptance and user experience. *In*: **Proceedings of the 6th International Conference on Automotive User Interfaces and Interactive Vehicular Applications (AutomotiveUI '14)**. Seattle, WA: ACM, 2014. p. 1-8. DOI: 10.1145/2667317.2667330.

ROTTER, Julian B. Generalized expectancies for interpersonal trust. **American Psychologist**, v. 26, p. 443-452, 1971.

ROUVROY, Antoinette; BERNS, Thomas. Algorithmic Governmentality and Prospects of Emancipation: Disparateness as a Precondition for Individuation through Relationships? *In*: HAMMERSLEV, Ole; MADSEN, J.M. **Rethinking Law, Democracy and the Digital**. Copenhagen: Ex Tuto Publishing, 2015. p. 3-17.

ROUVROY, Antoinette; BERNS, Thomas. Governamentalidade algorítmica e perspectivas de emancipação: o díspar como condição de individuação pela relação. **Revista Eco-Pós**, v. 18, n. 2, 2015.

RUMELHART, David E.; HINTON, Geoffrey E.; WILLIAMS, Ronald J. Learning representations by back-propagating errors. **Nature**, London, v. 323, n. 6088, p. 533-536, 1986. DOI: 10.1038/323533a0. Disponível em: https://www.nature.com/articles/323533a0. Acesso em: 25 jun. 2025.

RUSSELL, Stuart J.; NORVIG, Peter. **Artificial intelligence:** a modern approach. Kuala Lumpur: Pearson Education Limited, 2016.

RYDELL, Robert J.; MACKIE, Diane M. Consequences of discrepant explicit and implicit attitudes: Cognitive dissonance and increased information processing. **Journal of Experimental Social Psychology**, San Diego, v. 44, n. 6, p. 1526-1532, nov. 2008. DOI: 10.1016/j.jesp.2008.07.006.

SANTAELLA, Lúcia; KAUFMAN, Dora. A inteligência artificial generativa como quarta ferida narcísica do humano. **Matrizes**, São Paulo, v. 18, n. 1, p. 37-53, jan./abr. 2024. DOI: https://doi.org/10.11606/issn.1982-8160.v18i1p37-53. Disponível em: https://www.revistas.usp.br/matrizes/article/view/221116. Acesso em: 16 out. 2025.

SHAKKER AI. **Imagem** [gerada por inteligência artificial]. Shakker AI, 2025. Disponível em: https://www.shakker.ai/pt/home. Acesso em: 15 fev. 2025.

SHETH, Amit; THIRUNARAYAN, Krishnaprasad. **The duality of data and knowledge across the three waves of AI**. arXiv preprint arXiv:2103.13520, [S.l.], mar. 2021. Disponível em: https://arxiv.org/abs/2103.13520. Acesso em: 24 jun. 2025.

SHIN, Taylor; RAZEGHI, Yasaman; LOGAN IV, Robert L.; WALLACE, Eric; SINGH, Sameer. AutoPrompt: Eliciting Knowledge from Language Models with Automatically Generated Prompts. *In*: **Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)**, p. 4222-4235, 2020. Disponível em: https://aclanthology.org/2020.emnlp-main.346.pdf. Acesso em: 15 mar. 2025.

SIMÕES, José Manuel; CALDEIRA, Wilson. Harmonizing creation and integrity: navigating ethical complexities in generative artificial intelligence. *In:* CAMPONEZ, Carlos; CHRISTOFOLETTI, Rogério; SUÁREZ VILLEGAS, Juan Carlos (orgs.). **Comunicação, ética e IA:** diálogos sobre

desafios e perspetivas na era digital. Braga: CECS – Centro de Estudos de Comunicação e Sociedade, Universidade do Minho, 2025. p. 43-59.

SILVA, Francilene de Oliveira; PALETTA, Rita de Cássia Romeiro. Ética no jornalismo em tempos de máquinas comunicadoras e inteligência artificial generativa. *In:* CAMPONEZ, Carlos; CHRISTOFOLETTI, Rogério; SUÁREZ VILLEGAS, Juan Carlos (orgs.). **Comunicação, ética e IA:** diálogos sobre desafios e perspetivas na era digital. Braga: CECS – Centro de Estudos de Comunicação e Sociedade, Universidade do Minho, 2025. p. 81-99.

SMITH, Aaron. For Media Or Other Inquiries. **Pew Research Center**, v. 16, 2018. Disponível em: www.pewresearch.org. Acesso em: 19 jun. 2024.

STRATE, Lance. Studying media as media: McLuhan and the media ecology approach. **MediaTropes eJournal**, v. 1, p. 127-142, 2008.

WĄSIKOWSKA, Barbara. Consumer Neuroscience - The Application of Selected Neurobiological Methods in Consumer Research. **European Research Studies**, v. 24, n. 2B, p. 1153-1162, 2021.

WIEDERHOLD, Stijn G. H. **Neuromarketing & ethics:** how far should we go with predicting consumer behavior? 2020. Bachelor's Thesis – University of Twente, Netherlands, 2020.

WIRTH, Norbert. Hello marketing, what can artificial intelligence help you with? **International Journal of Market Research**, v. 60, n. 5, p. 435-438, 2018.

WISE, Roy A. Forebrain substrates of reward and motivation. **Journal of Comparative Neurology**, v. 493, n. 1, p. 115-121, 2005. Disponível em: https://onlinelibrary.wiley.com/doi/10.1002/cne.20689. Acesso em: 15 jan. 2025.

WOLTON, Dominique. **Informar não é comunicar**. Tradução de Juremir Machado da Silva. Porto Alegre: Sulina, 2010.

TADDEO, Mariarosaria; FLORIDI, Luciano. The case for e-trust. **Ethics and Information Technology**, v. 13, p. 1-3, 2011.

TARABORELLI, Dario. How the Web is changing the way we trust. *In*: WAELBERS, Katinka; BRIGGLE, Adam; BREY, Philip (Eds.). **Current Issues in Computing and Philosophy.** Amsterdam: IOS Press, 2008. p. 194-204.

TINWELL, Angela. **The Uncanny Valley in Games and Animation**. Boca Raton: CRC Press, 2014.

TURING, Alan M. Computing machinery and intelligence (1950). *In*: COPPEL, B. Jack (Ed.). **The essential Turing:** the ideas that gave birth to the computer age. Oxford: Clarendon Press, 2012. p. 433-464.

TURKLE, Sherry. **Alone together:** why we expect more from technology and less from each other. New York: Basic Books, 2011.

YIGAEL, Yoav. Fundamental Issues in Artificial Intelligence. **World Futures**, v. 67, n. 8, p.564–568, 7 nov. 2011.

ZAJONC, Robert B. Feeling and thinking: Preferences need no inferences. **American Psychologist**, Washington, DC, v. 35, n. 2, p. 151-175, fev. 1980. DOI: 10.1037/0003-066X.35.2.151.

ZALTMAN, Gerald. **How customers think:** essential insights into the mind of the market. Boston: Harvard Business School Press, 2003.

ZUBOFF, Shoshana. **A era do capitalismo de vigilância:** a luta por um futuro humano na nova fronteira do poder. Rio de Janeiro: Intrínseca, 2020.

APÊNDICES

APÊNDICES

APÊNDICES

APÊNDICE A – Contextualização e Ilustrações sobre o IAT e o EAAT

A fim de tornar mais visível a aplicação metodológica deste estudo, apresenta-se a seguir a ilustração dos testes IAT e EAAT, construídos especialmente para a pesquisa desenvolvida. Ambas as ferramentas foram desenvolvidas com base nos princípios estabelecidos pelo *Project Implicit*, da Universidade de Harvard, cuja plataforma disponibiliza testes validados para uso em ambientes acadêmicos e educacionais (*Project Implicit*, 2025).

O *Project Implicit* foi fundado em 1998 pelos pesquisadores Tony Greenwald (*University of Washington*), Mahzarin Banaji (*Harvard University*) e Brian Nosek (*University of Virginia*). A iniciativa tem como missão divulgar o conhecimento científico sobre os viéses implícitos e desenvolver ferramentas que permitam medi-los de forma rigorosa e replicável. Para isso, criou-se um "laboratório virtual" que aplica testes comportamentais baseados em tempo de reação e escolhas proposicionais conscientes, possibilitando identificar dissociações entre atitudes automáticas e deliberadas (*Project Implicit*, 2025).

A construção dos *slides* utilizados na pesquisa desenvolvida foi guiada diretamente pela estrutura proposta nos testes disponíveis no *site* oficial da plataforma (https://implicit.harvard.edu). Foram mantidos critérios já validados pela Universidade de Harvard como o número de imagens e de atributos avaliados, bem como a ordem dos blocos de apresentação e as instruções fornecidas ao início de cada teste, de modo a preservar a confiabilidade psicométrica do modelo original.

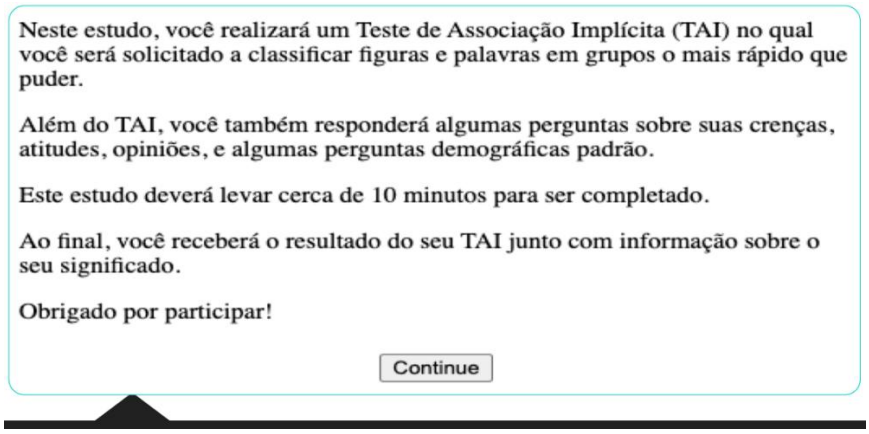
Na sequência, são apresentadas miniaturas dos *slides* que compuseram a aplicação de ambos os testes, visando a dar transparência ao processo experimental e destacar a coerência metodológica entre a proposta teórica e a execução prática da pesquisa.

Estrutura do Teste

O IAT, na forma padrão, é montado em **blocos de ensaios** onde os participantes são instruídos a classificar estímulos rapidamente, pressionando teclas específicas. Geralmente, utilizam-se duas categorias centrais (imagens reais x imagens IA) e dois atributos ("Competente" vs. "Sem Decoro"). A estrutura típica segue um **formato de 7 blocos** (Greenwald et al., 2003; Lane, Banaji, Nosek, & Greenwald, 2007):

1. **Bloco 1 de Aprendizagem Simples**: **Treinamento Inicial** – Classificação de imagens como **REAL** ou **IA**.
2. **Bloco 2 de Aprendizagem** : **Treinamento de Palavras** – Associação de palavras.
3. **Bloco 3** : **Associação de Imagens e Atributos**.
4. Bloco 4 **Teste Crítico** – Medida de velocidade e precisão das associações; teste dessa mesma combinação; mede-se o tempo de reação
5. **Bloco 5 de Aprendizagem Simples 2**: mudança de posições das categorias; inversão das teclas pressionadas para evitar aprendizado mecânico
6. **Bloco 6 Associação Invertida** – Teste da dificuldade; associação contrabalanceada
7. **Bloco 7 Teste Final**: Mesma estrutura do bloco 4, para testar consistência das respostas, teste dessa segunda combinação; mede-se novamente o tempo de reação.

O diferencial está na comparação entre o tempo de reação e taxa de acertos nos blocos de teste (4 e 7). Se o participante responde mais rapidamente quando Imagem Real está pareada com Agradável do que quando Imagem IA está pareada com Falsa (e inversamente para Real), infere-se que existem associações implícitas mais fortes entre Real e Agradável, por exemplo.



Na próxima tarefa, será apresentado um conjunto de palavras ou imagens para você classificar em grupos. Esta tarefa requer que você classifique os itens o mais rápido que puder, fazendo o menor número de erros possíveis. Se você for muito lento ou fizer muitos erros os resultados não poderão ser interpretados adequadamente. Esta parte do estudo levará cerca de 5 minutos. Segue uma lista de rótulos de categorias e os itens que pertencem a cada uma dessas categorias.

Categoria	Itens
Positivo	**Competente, qualificado, capaz, justo, confiável, seguro, honesto, simpático**
Negativo	**Incompetente, desqualificado, incapaz, injusto, duvidoso, inseguro, desonesto, antipático**
REAL	**Imagem Real (veja na próxima pagina)**
IA	**Imagem IA (veja na próxima pagina)**

Tenha em mente

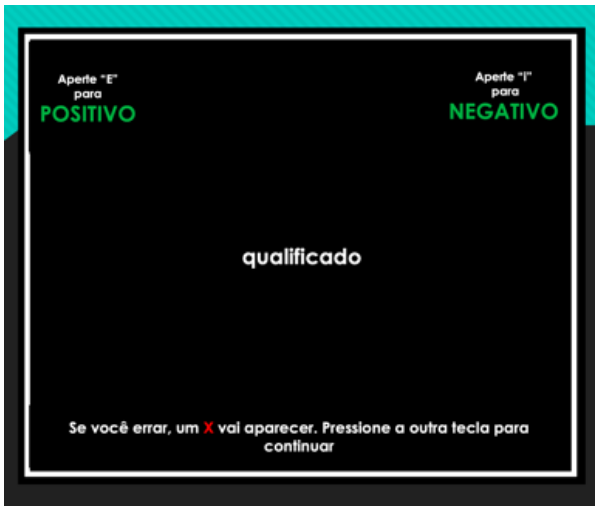
- Mantenha os dedos indicadores nas teclas 'e' e 'i', respectivamente, para que a sua resposta seja rápida.
- Dois rótulos no topo indicarão que palavras ou imagens correspondem a cada uma das teclas.
- Cada palavra ou imagem possui uma classificação correta. A maioria é fácil.
- O teste não produzirá resultados se você responder lentamente -- Por favor tente responder o mais rápido possível.
- É de se esperar que você cometa alguns erros, já que terá de responder rapidamente. Não tem problema.
- Para melhores resultados, evite distrações e permaneça concentrado.

Estou pronto para começar

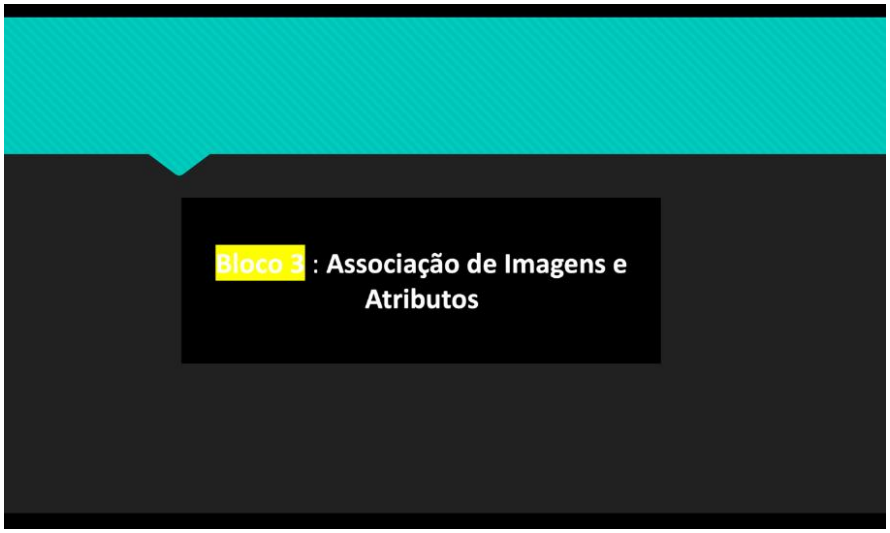

REAL
Imagens Reais

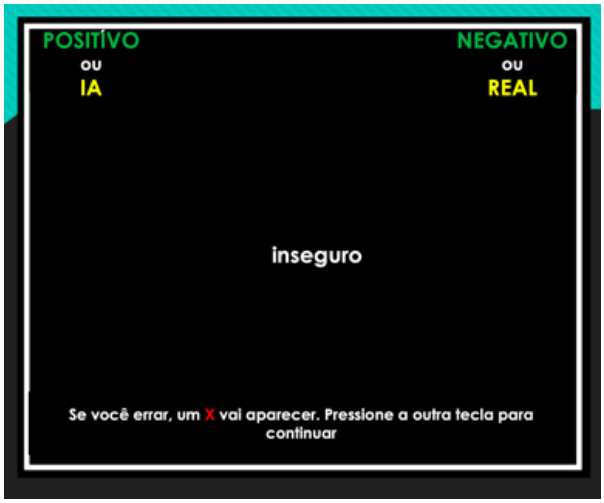

Bloco 1 de Aprendizagem Simples:
Treinamento Inicial – Classificação de
imagens como IA ou REAL.

Seguem, no teste, mais 18 (dezoito) *slides* de imagens geradas por IAG e reais.

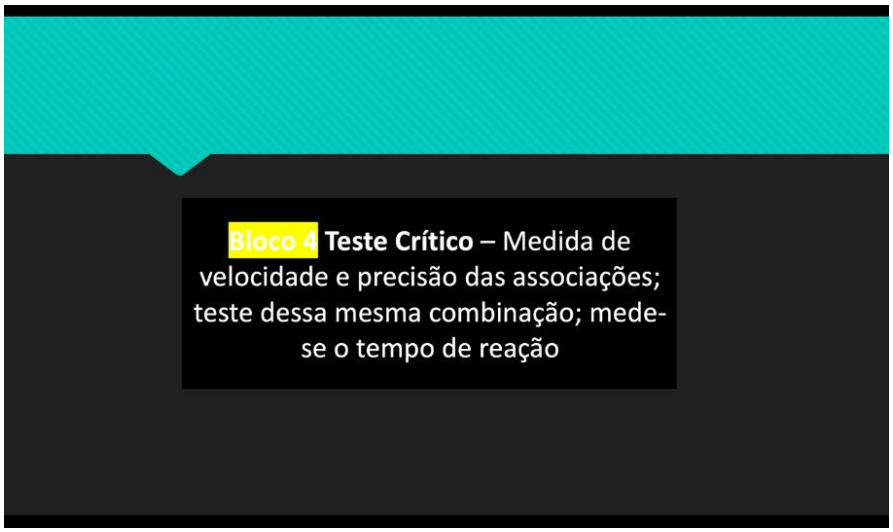


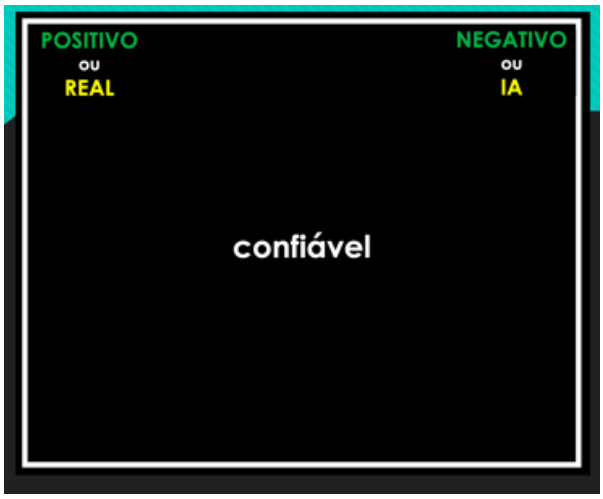
Mais 14 (quatorze) *slides* de atributos positivos e negativos são apresentados sequencialmente no teste.



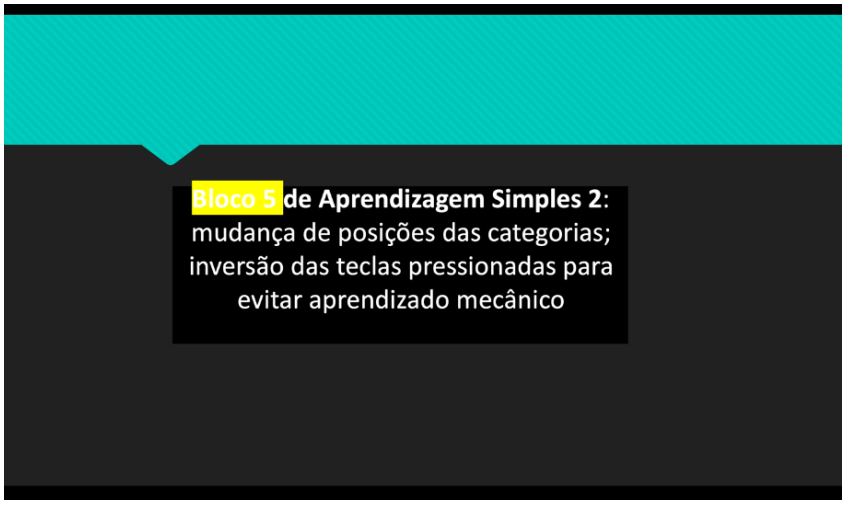


Seguem, no teste, mais 18 (dezoito) *slides* mesclando atributos positivos ou negativos e imagens geradas por IAG ou reais.

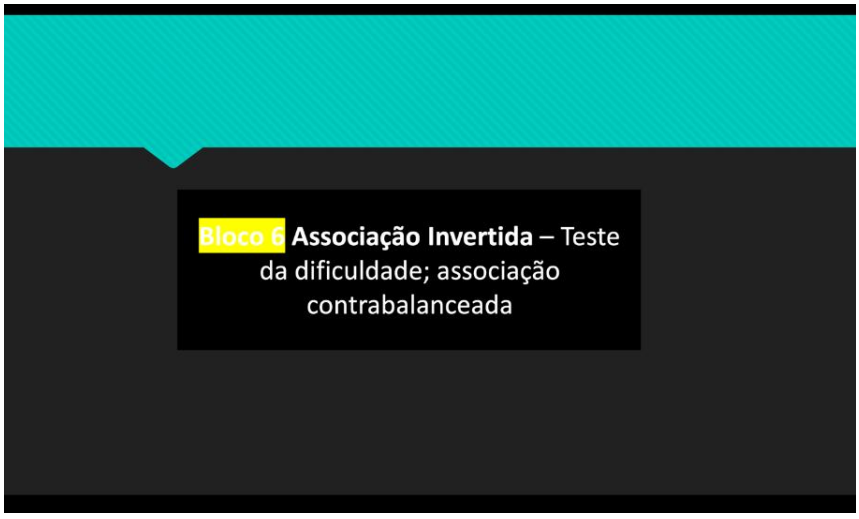


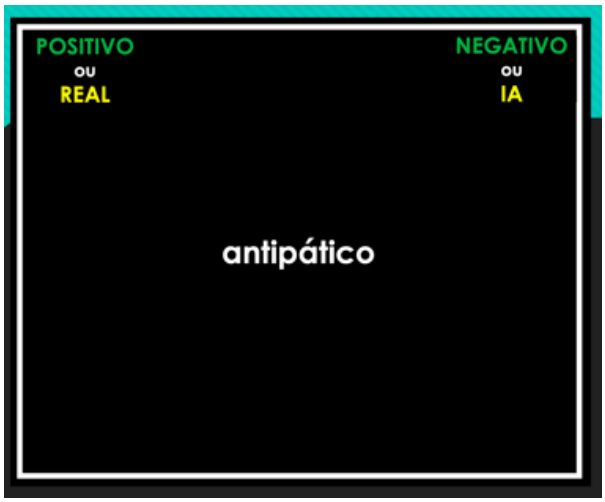


São apresentados, no teste, mais 38 (trinta e oito) *slides* mesclando atributos positivos ou negativos e imagens geradas por IAG ou reais.

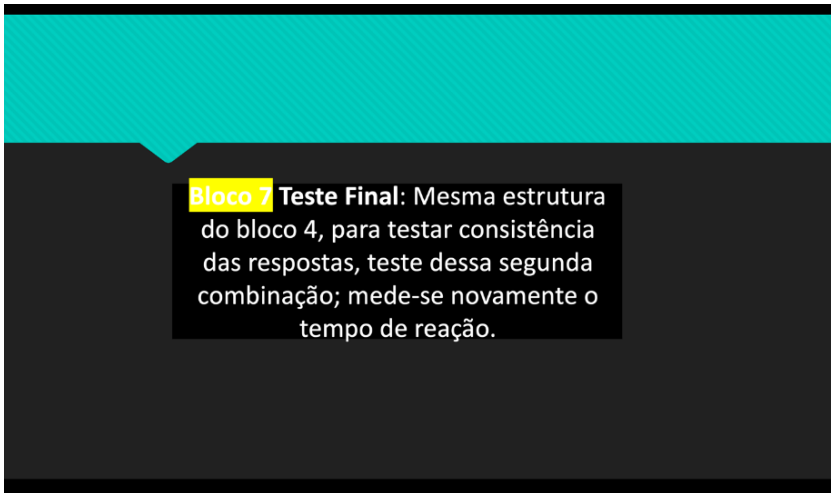


Sequencialmente, apresentam-se mais 26 (vinte e seis) *slides* no teste com imagens geradas por IAG ou reais.

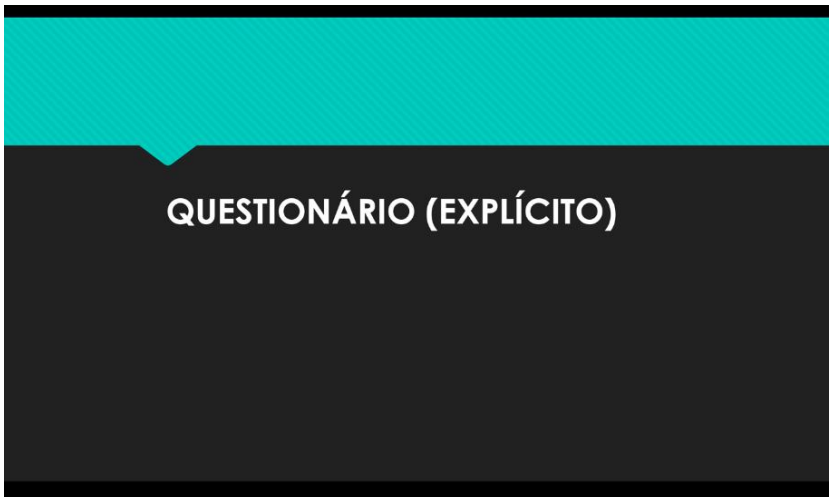


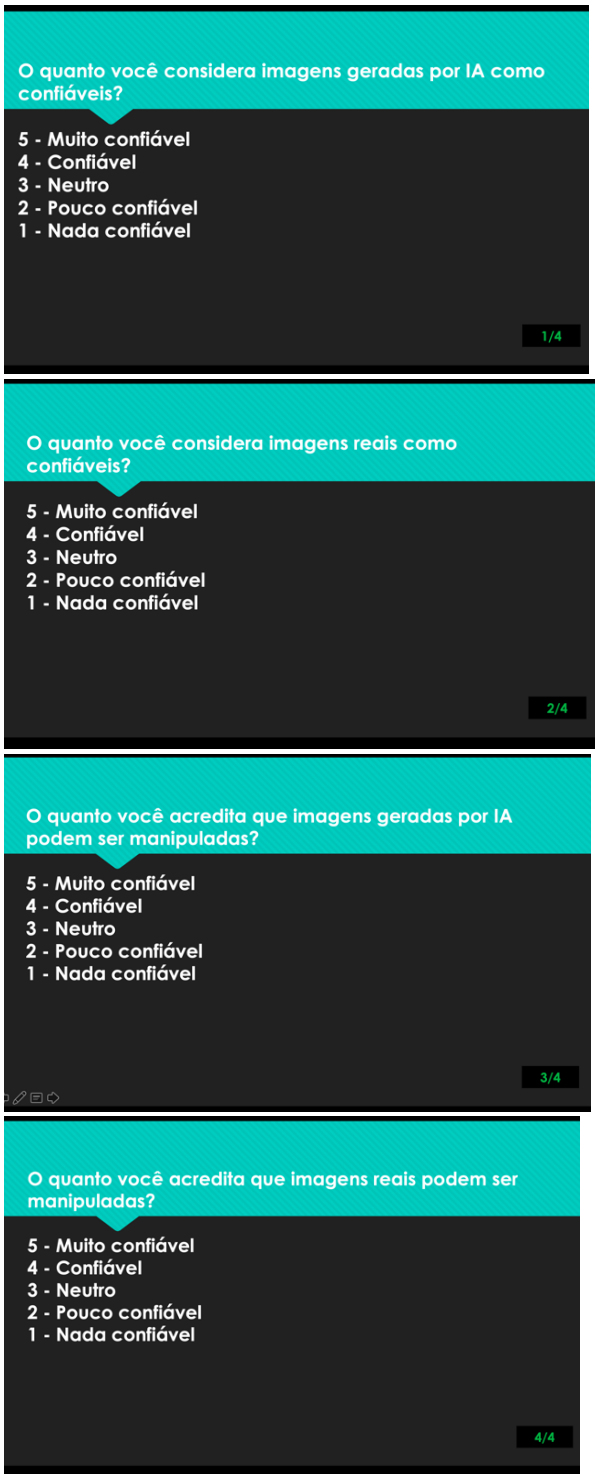


Apresentam-se, no teste, mais 18 (dezoito) *slides* mesclando atributos positivos ou negativos e imagens geradas por IAG ou reais.



Seguem, sequencialmente, no teste, mais 38 (trinta e oito) *slides* mesclando atributos positivos ou negativos e imagens geradas por IAG ou reais.





Como demonstrado, o teste se mostrou uma importante ferramenta para geração de dados deste trabalho e a análise dos dados gerados representa um contributo ao meio acadêmico.


idp
A ESCOLHA QUE
TRANSFORMA
O SEU CONHECIMENTO