

***Brainwriting* na eliciação de requisitos para IA confiável**

**Francisco Luciano Quirino da Silva¹, Andréia Libório Sampaio¹,
Carla Ilane Moreira Bezerra¹, Ingrid Teixeira Monteiro¹**

¹ Programa de Pós-Graduação em Computação (PCOMP) – Universidade
Federal do Ceará (UFC) – Quixadá – CE – Brasil

{lucianoquirino14, andreia.ufc}@gmail.com,

{carlailane, ingrid}@ufc.br

Abstract. *Failures identified in Machine Learning (ML) systems have affected users' trust, making it essential to address the issue. This study investigates the use of the Brainwriting technique in eliciting requirements for trustworthy Artificial Intelligence (AI), focusing on ML, through multidisciplinary collaboration among the involved roles. The study was conducted with a group of women with different professional profiles, facilitated by a moderator. At the end of the study, a set of open-ended questions was applied to understand the participants' perspectives regarding trustworthy AI. Through the study, it was possible to create requirements for trustworthy AI, with emphasis on ML, as well as facilitate multidisciplinary participation using Brainwriting.*

Resumo. *Falhas apontadas em sistemas de Aprendizado de Máquina (AM) têm afetado a confiança dos usuários, tornando-se essencial abordar o tema. Este estudo investiga o uso da técnica Brainwriting na eliciação de requisitos para Inteligência Artificial (IA) confiável, com foco em AM, por meio da colaboração multidisciplinar dos papéis envolvidos. O estudo foi conduzido com um grupo de mulheres com perfis de atuação diferentes, facilitado por um mediador. Ao final do estudo, um conjunto de perguntas abertas foi aplicado para compreender a perspectiva das participantes em relação à IA confiável. Com o estudo foi possível criar requisitos para a IA confiável, com ênfase em AM, além de proporcionar a participação multidisciplinar, utilizando Brainwriting.*

1. Introdução

Parte das tarefas que antes eram executadas totalmente por humanos passou a ser apoiada por tecnologias como o Aprendizado de Máquina (AM) e seus algoritmos de tomada de decisão, tornando-se cada vez mais presentes na vida cotidiana. AM é uma abordagem da Inteligência Artificial (IA) focada em técnicas e sistemas relacionados ao aprendizado [Bellaby 2021]. De acordo com Bellaby (2021), a implementação de sistemas com algoritmos automatizados traz algumas preocupações em relação à sua incorporação em infraestruturas críticas da sociedade, considerando estudos de casos que abordam os efeitos adversos das soluções de IA. Com a crescente necessidade de aprimoramento das técnicas computacionais ligadas ao aprendizado dos sistemas, foram desenvolvidos e propostos métodos e estruturas recentemente, visando garantir que os sistemas se comportem sem causar danos aos usuários e à sociedade [Kaur et al. 2022]. Esses métodos frequentemente concentram-se em torná-los mais confiáveis e seguros.

Muitos trabalhos estão sendo desenvolvidos pelo governo e por diferentes comunidades científicas no campo da IA confiável, propondo formas diversas de abordar o assunto, com abordagens envolvem propriedades como justiça, transparência, responsabilidade, controlabilidade, diretrizes éticas, leis, programas e estruturas para medir e aumentar a confiança do usuário, além de promover a prestação de contas e o uso responsável da IA, demonstrando a importância e a preocupação em estabelecer uma estrutura confiável no desenvolvimento de sistemas de IA [Kaur et al. 2022]. As diversas abordagens propostas concentram-se em diferentes fases do ciclo de vida da IA, como *design*, dados, modelagem, implementação e supervisão. Apesar da existência de vários trabalhos na literatura que abordam a confiança em IA, há uma escassez de pesquisas sobre requisitos para lidar com a confiança em IA/AM. Essa lacuna pode ser atribuída à falta de pesquisas sobre Engenharia de Requisitos (ER) para AM [Yoshioka et al. 2021]. Abordar requisitos confiáveis é uma tarefa de grande importância para o desenvolvimento de sistemas de IA seguros e confiáveis [Kaur et al. 2022].

Para manter e desenvolver requisitos viáveis e valiosos para sistemas de AM, é necessário existir uma colaboração entre cientistas de dados, engenheiros de software e especialistas de domínio [Yoshioka et al. 2021]. Buscar técnicas que permitam o envolvimento colaborativo na elicitação de requisitos de sistemas de AM é crucial para o desenvolvimento e sucesso dessas aplicações. Uma técnica utilizada na elicitação de requisitos que permite o envolvimento de grupos é a *Brainwriting* [Calazans et al. 2018]. A técnica *Brainwriting* é baseada em *Brainstorming* e busca levantar soluções por meio de descrição a partir do envolvimento de grupos [Pazmino 2015]. *Brainwriting* é uma abordagem eficiente para a geração de ideias, que tem o potencial de ajudar os participantes a superar a redução de produtividade observada no processo de *Brainstorming* [Michinov 2012].

O objetivo deste trabalho é investigar o uso da técnica *Brainwriting* na elicitação de requisitos para Inteligência Artificial (IA) confiável, com foco em AM, por meio da colaboração multidisciplinar dos papéis envolvidos. Foi conduzido um estudo com um grupo de seis participantes mulheres, de diferentes perfis de atuação, utilizando o *Brainwriting* e questões abertas para discutir IA confiável. O estudo adotou como contexto o caso apresentado por Dastin (2018), que aborda o algoritmo de recrutamento da *Amazon* considerado sexista devido a problemas de viés de gênero feminino em suas saídas. O estudo baseou-se nos princípios de IA confiável apresentados por Kaur et al. (2022), incluindo Justiça, Explicabilidade, Responsabilidade, Privacidade e Aceitação. Essa pesquisa contribui para preencher a lacuna sobre a ER para AM e IA confiável, além de promover a reflexão sobre os aspectos e desafios desse tema na sociedade.

Este artigo está organizado da seguinte forma: a fundamentação teórica (Seção 2), os trabalhos relacionados (Seção 3), a elicitação de requisitos para IA confiável (Seção 4), os resultados (Seção 5), as discussões e o relato das participantes (Seção 6), as limitações e ameaças à validade (Seção 7) e a conclusão e os trabalhos futuros (Seção 8).

2. Fundamentação teórica

Nesta Seção são apresentados os principais conceitos para a definição deste estudo.

2.1. Engenharia de Requisitos para Aprendizado de Máquina

A ER é uma fase essencial do processo de desenvolvimento de sistemas de software, desempenhando um papel crítico na definição de produtos adequados e de alto valor para

seus interessados [Ishikawa and Matsuno 2020]. A ausência da ER ou sua realização inadequada pode acarretar inconsistências e, conseqüentemente, custos adicionais para o projeto [Gruber et al. 2017]. De acordo com Gruber et al. (2017), um desafio enfrentado pela ER tem sido como desenvolver e gerenciar requisitos para sistemas de IA/AM. Segundo Horkoff (2019), quando se trata de soluções de software envolvendo AM, devido à forma de projetar, executar e manter essas soluções, parte do conhecimento sobre requisitos não pode ser aplicado.

2.2. *Brainwriting*

Brainwriting é uma abordagem alternativa à técnica de *Brainstorming*, aplicada a um grupo de pessoas, tendo foco na geração silenciosa e escrita de ideias [VanGundy 1984]. Nesta técnica a formulação de ideias é feita de forma escrita, evitando voz alta, visando que os participantes possam ser mais criativos, contribuindo para uma melhor concentração, levando a um maior fluxo de ideias [Heslin 2009]. Outro ponto importante é que a escrita evita que as diferenças existentes entre os perfis envolvidos na discussão não venham interferir, deixando os participantes inseguros ao expressar suas ideias diante de pessoas com outro *status* social ou nível de conhecimento [VanGundy 1984].

2.3. IA confiável

A IA confiável (*Trustworthy Artificial Intelligence - TAI*) é um campo de estudo interdisciplinar e dinâmico, que compreende que a utilização de todo o potencial da IA pela sociedade global só se dará a partir do estabelecimento de confiança, que é um fenômeno complexo e que apresenta diferenças em relação às perspectivas de aplicação, natureza e crenças [Thiebes et al. 2021]. Kaur et al. (2022) abordam os seguintes princípios para tratar a confiança em sistema de IA:

- **Justiça:** garantir que o sistema seja justo, livre de preconceito e discriminação;
- **Explicabilidade:** garantir que os diferentes *stakeholders* entendam os motivos que levaram a uma determinada decisão;
- **Responsabilidade:** garantir a rastreabilidade dos responsáveis por danos causados pelos algoritmos;
- **Privacidade:** garantir o uso dos dados de forma responsável, sem causar conseqüências prejudiciais ao usuário; e,
- **Aceitação:** aumentar a aceitação do sistema pelos usuários.

3. Trabalhos relacionados

Kaur et al. (2022) abordam uma análise da literatura sobre diferentes requisitos para tornar sistemas de IA confiáveis. Os autores discutem diversas abordagens para mitigar riscos, aumentar a confiança e promover a aceitação desses sistemas pela sociedade. Eles também exploram estratégias de validação e verificação. O estudo oferece uma visão dos avanços recentes em IA confiável, contribuindo para o entendimento e orientação de futuras pesquisas.

Ahmad et al. (2021) apresentam uma Revisão Sistemática da Literatura (RSL) sobre métodos de ER para IA. Os autores investigaram abordagens, ferramentas e técnicas de escrita e modelagem de requisitos para sistemas de IA/AM. Através da análise de 27 estudos, foram identificadas várias técnicas e notações utilizadas na área, assim como,

desafios e limitações existentes. Os resultados mostraram que ainda existe uma carência de pesquisas de ER para sistemas de IA/AM.

No estudo de Cerqueira et al. (2022) os autores apresentam um Guia para elicitación de requisitos éticos de IA, intitulado de Guia Ético RE4AI. Os autores adotaram a metodologia *Design Science Research*, com o intuito de investigar o problema, desenvolver um protótipo e avaliá-lo. O Guia visa ajudar equipes de desenvolvimento de software ágil a elicitación requisitos éticos para IA, consistindo em sistema baseado na web que conta com baralho de 26 cartas e considera 11 princípios éticos. O estudo tem como contribuição o preenchimento da lacuna entre princípios e práticas de alto nível e abstratos, auxiliando na elicitación de requisitos e implementação ética em IA, utilizando histórias de usuários.

Pei et al. (2022) abordam uma revisão sobre ER para AM. O estudo visa fornecer uma visão sobre processo de ER para AM, considerando a colaboração dos diferentes papéis. Os autores revisaram a literatura sobre ER para AM e investigaram em seguida a relação de cada função durante o processo de compreensão colaborativa de requisitos e desenvolvimento de sistemas. No estudo, os autores apresentam de maneira resumida um processo de análise colaborativa de requisitos. Os autores apresentam direções para pesquisas futuras.

Os estudos apresentados por Ahmad et al. (2021) e Pei et al. (2022) investigaram ER para IA e/ou AM, porém o objetivo dos autores foi identificar abordagens para tratar requisitos, nenhum deles abordou diretamente requisitos de IA confiável. Por outro lado, o estudo de Kaur et al. (2022) analisou requisitos para tornar os sistemas de IA confiáveis, mas não propôs um modelo de processo para lidar com essa questão. O trabalho de Cerqueira et al. (2022) apresentou um guia para tratar requisitos em IA, no contexto ético, sem tratar diretamente da IA confiável. Este trabalho investiga a técnica de *Brainwriting* como uma técnica para a elicitación de requisitos para IA confiável, envolvendo a participação dos papéis interessados, considerando sistemas de AM.

4. Elicitación de requisitos para IA confiável

Nesta seção, são apresentadas as etapas e atividades da pesquisa.

4.1. Planejamento e preparação

Inicialmente, com o objetivo de selecionar um contexto para o estudo, foram investigados problemas reais nos quais algoritmos de AM geraram resultados enviesados que afetaram a confiança dos usuários finais. Durante o processo de investigação, foi identificado o problema gerado por uma ferramenta experimental de IA para recrutamento da *Amazon*, considerada sexista por apresentar viés de gênero feminino em suas saídas. Dastin (2018) relata a descoberta de um grande problema existente em uma ferramenta experimental de recrutamento da *Amazon*, que utilizava IA para classificar candidatos com pontuações que variavam de 1 a 5 estrelas. No entanto, foi observado que o sistema não classificava candidatos de maneira neutra em termos de gênero para vagas de desenvolvedor de software e outros cargos técnicos. Isso foi consequência da forma de treinamento do modelo, que foi treinado com base em padrões de currículos enviados à empresa ao longo de um período de 10 anos, nos quais a maioria era composta por homens. Dessa maneira, decidiu-se adotar esse exemplo como contexto para aplicação do estudo.

Em seguida, foi definido o roteiro e preparado o material de apoio para a aplicação do estudo, com o termo de consentimento para participação na pesquisa, a definição dos principais conceitos utilizados, a ferramenta para a aplicação da técnica de *Brainwriting* e as questões para investigar a IA confiável a partir da visão das participantes da discussão. No material de apoio, foi criado um espaço para os relatos das participantes, destinado a opiniões sobre o assunto abordado, a forma de condução da pesquisa ou sugestões. A ferramenta de suporte adotada para o estudo foi o Google Jamboard¹, por viabilizar a colaboração entre equipes e ser de fácil manuseio. As questões para investigar a IA confiável a partir da visão das participantes foram adaptadas das questões utilizadas no estudo de Oliveira Carvalho et al. (2022) e são as seguintes: Q1 - “*Por que abordar confiança em IA?*”; Q2 - “*O que abordar dentro do contexto de confiança em IA?*”; Q3 - “*Quando abordar confiança em IA?*”; Q4 - “*Como abordar confiança em IA?*”; Q5 - “*Onde abordar confiança em IA?*”; Q6 - “*Por quem deve ser abordada confiança em IA?*”; e, Q7 - “*Para quem abordar confiança em IA?*”.

Para finalizar, foi definido o público-alvo e recrutados os participantes para o estudo. Levando em consideração equipes de projeto multidisciplinares e o exemplo adotado, que aborda o viés de gênero feminino no recrutamento com IA, o público-alvo foi estabelecido como mulheres atuantes nas áreas de Recursos Humanos, Desenvolvimento de Software e/ou Pesquisa em Computação. Considerando o público-alvo, buscou-se o contato por meio de mensagens enviadas nas plataformas de redes sociais WhatsApp² e LinkedIn³, porém, devido a limitações de tempo, apenas seis pessoas aceitaram participar do estudo. As participantes possuem os seguintes perfis: P1 - *Design*; P2 - *Desenvolvimento*; P3 - *Qualidade/Pesquisa*; P4 - *Ensino/Pesquisa*; P5 - *Pesquisa*; e P6 - *Recursos Humanos/Recrutamento*. O responsável por aplicar e conduzir o processo foi o primeiro autor do artigo, que possui experiência com Pesquisa, Requisitos e *Design*.

4.2. Aplicação

Nesta etapa, foi realizada a aplicação do estudo com as participantes selecionadas na etapa anterior. O estudo foi conduzido *online* através da plataforma Google Meet⁴. Inicialmente, o condutor do estudo distribuiu para todas as participantes um documento de apoio. Após o consentimento das participantes, o condutor prosseguiu com a explicação do estudo, processo adotado, termos importantes e princípios para tratar a confiança em sistemas de IA (Justiça, Explicabilidade, Responsabilidade, Privacidade e Aceitação) apresentados por Kaur et al. (2022), e contexto de discussão baseado em Dastin (2018).

A partir da compreensão de todos os envolvidos sobre o processo adotado no estudo e contexto de discussão, iniciou-se a etapa de *Brainwriting*, sendo apresentada a seguinte questão “*Como desenvolver sistemas de recrutamento e seleção de candidatos utilizando Inteligência Artificial, livres de discriminação e confiáveis para os usuários?*”. Os participantes foram incentivados a escrever ideias de soluções com base nos princípios de Justiça, Explicabilidade, Responsabilidade, Privacidade e Aceitação, para tratar a confiança em sistemas de IA. As ideias foram descritas na ferramenta Google Jamboard, em quadros separados, durante um período de 10 minutos. Em seguida, as ideias geradas

¹<https://jamboard.google.com/>

²<https://web.whatsapp.com/>

³<https://www.linkedin.com>

⁴<https://meet.google.com/>

foram compartilhadas e discutidas com todas as participantes, permitindo o refinamento das ideias e a disseminação de conhecimento entre as partes envolvidas. Após essa etapa de compartilhamento de ideias, foram disponibilizados 5 minutos para que as participantes pudessem retornar aos seus quadros e refletir sobre suas ideias, possíveis melhorias ou novos registros. Para finalizar, todas as ideias foram compartilhadas novamente, visando classificá-las de acordo com os princípios para tratar a confiança em sistemas de IA apresentados por Kaur et al. (2022). Durante o processo, a participante P6 deixou o estudo. Esta etapa teve duração de 2 horas e 30 minutos.

Após a etapa de *Brainwriting*, as participantes foram liberadas, com a responsabilidade de refletir sobre o assunto e, posteriormente, responder às questões sobre IA confiável. Além disso, foram encorajadas a deixar suas opiniões sobre o assunto abordado, a forma de condução da pesquisa ou sugestões. O prazo para responder às questões foi de 5 dias. As participantes tiveram a liberdade de decidir se retornariam ou não ao estudo e responderiam às questões. Todas as participantes optaram por retornar e responderam às questões. Os formulários com as respostas das participantes estão disponíveis no relatório técnico⁵.

4.3. Análise e extração de requisitos

Inicialmente, as ideias levantadas a partir da aplicação da técnica de *Brainwriting* foram transcritas para uma planilha eletrônica, analisadas, etiquetadas com identificadores e princípios para tratar a confiança em sistemas de IA apresentados por Kaur et al. (2022), agrupadas por semelhança e transformadas em requisitos. As 22 ideias estão disponíveis em relatório técnico⁶. Desta forma, as 22 ideias levantadas foram transformadas em 14 requisitos, apresentados na próxima Seção. Em seguida, as questões sobre IA confiável, liberadas para que as participantes respondessem, foram analisadas qualitativamente por se tratarem de dados descritivos. Métodos de análise qualitativa tratam de variáveis que não permitem serem medidas, apenas observadas [Lazar et al. 2017]. A análise considerou a percepção dos participantes em relação à IA confiável. O processo de análise se deu através da transcrição das respostas para uma planilha eletrônica, identificação de relação entre as respostas das participantes, padrões e temáticas importantes considerando as limitações de cada questão. Todas as atividades desta etapa foram realizadas pelo primeiro autor do artigo. Os resultados são apresentados na próxima Seção.

5. Resultados

Esta Seção apresenta os resultados obtidos através do estudo.

5.1. Resultados sobre a aplicação da técnica de *Brainwriting* e requisitos extraídos

A Tabela 1 apresenta 14 requisitos não funcionais extraídos das 22 ideias levantadas durante a aplicação da técnica de *Brainwriting*. Os 14 requisitos estão relacionados aos princípios para IA confiável abordados por Kaur et al. (2022) da seguinte forma:

- **Justiça:** R2, R3, R4, R5, R6, R7 e R8;
- **Explicabilidade:** R9, R10 e R11;
- **Responsabilidade:** R4, R12 e R13;
- **Privacidade:** R1 e R2; e,
- **Aceitação:** R7, R8, R9, R11, R12, R13 e R14.

⁵<https://doi.org/10.5281/zenodo.7811847>

⁶<https://doi.org/10.5281/zenodo.7811877>

Tabela 1. Requisitos extraídos das ideias levantadas a partir da aplicação da técnica de *Brainwriting*

ID	Requisitos Extraídos
R1	O sistema deve proteger dados relacionados a nome de pessoa.
R2	O sistema deve ser treinado para desconsiderar dados relacionados a gênero, raça, etnia, idade e religião, visando evitar discriminação.
R3	O sistema deve ser treinado com <i>datasets</i> diversificados e de maneira equilibrada.
R4	A equipe de desenvolvimento e manutenção do sistema deve ser diversa.
R5	O sistema deve levar em consideração apenas habilidades e competências dos candidatos.
R6	Deve ser estabelecido limitações de uso do sistema, visando evitar que eles sejam utilizados de forma inadequada beneficiando ou excluindo determinados grupos.
R7	O sistema deve fornecer <i>template</i> padrão para todos os candidatos e ter os requisitos para as vagas bem definidas.
R8	O sistema deve passar por simulações de cenários do mundo real antes de ser disponibilizado para uso final.
R9	O sistema deve gerar relatórios de <i>match</i> entre requisitos para a vaga e candidatos selecionados, justificando a seleção.
R10	O sistema deve enviar relatórios de <i>matches</i> para os candidatos explicando a forma de seleção.
R11	O sistema deve fornecer aos candidatos <i>feedback</i> confiável e justo, apresentando os motivos pelos quais não foi selecionado e indicando pontos que podem ser melhorados.
R12	Deve ser realizado testes comparativos entre resultados gerados pelo sistema e resultados de processos realizados por profissionais de RH.
R13	O sistema deve ser monitorado com frequência, receber contínuas avaliações e manutenções.
R14	Deve ser garantido que os resultados gerados pelo sistema não sejam fatores decisivos para a contratação.

5.2. Resultados sobre confiança em IA a partir da visão dos participantes

Q1 - Por que abordar confiança em IA? A motivação para abordar a confiança em IA, a partir da perspectiva das participantes, complementa-se entre a visão das mesmas. A IA confiável aborda questões éticas, possibilitando o desenvolvimento de sistemas livres de preconceito, discriminação e exclusão de grupos, resultando em uma experiência mais eficaz e satisfatória para os usuários, graças à confiança transmitida.

Q2 - O que abordar dentro do contexto de confiança em IA? Conforme a visão das participantes dentro do contexto de confiança em IA, deve-se abordar: (i) Transparência de Dados e Forma de Tratamento; (ii) Imparcialidade; (iii) Princípios Éticos; (iv) Leis Locais; e, (v) Princípios de Confiança. Esses itens estão relacionados aos princípios propostos por Kaur et al. (2022) para tratar a confiança em sistemas de IA.

Q3 - Quando abordar confiança em IA? No sentido de quando abordar confiança em IA, para as participantes as respostas apresentam-se de maneiras diversas. Para P1 a confiança em IA deve ser abordada “*Sempre que ela for ser utilizada*”. Já as demais participantes apresentaram respostas sobre a abordagem em momentos ou situações específicas, como P2 “*Quando a IA a ser utilizada provê resultados difíceis de serem revertidos...*”, P3 “*Sempre que for lidar com dados sensíveis...*”, P5 “*... ao se trabalhar em sistemas que façam o uso de informações críticas e confidenciais...*” e P4 “*Durante todo o processo inicial...*”.

Q4 - Como abordar confiança em IA? Para as participantes o assunto pode ser tratado através: (i) Princípios regulamentadores; (ii) Discussões amplas; (iii) Estudos aplicados a resultados gerados em sistemas que utilizam a IA; e, (iv) Estudos sobre como

aplicar princípios de confiança no sistema a ser desenvolvido. A participante P5 ressalta que estudos sobre como aplicar princípios de confiança no sistema a ser desenvolvido devem “... ser feito na fase inicial do projeto. Podendo ser feito um brainstorming com a equipe que será responsável pelo projeto”. Para P3 o assunto pode ser abordado “Em forma de que leve mais segurança aos usuários”.

Q5 - Onde abordar confiança em IA? Segundo a visão das participantes, o assunto confiança em IA deve ser abordado em todo o ciclo de desenvolvimento do projeto, assim como em debates e pesquisas.

Q6 - Por quem deve ser abordada confiança em IA? Segundo a visão das participantes, o assunto sobre confiança em IA deve ser abordado por diversos grupos, como: (i) Empresa que desenvolve a tecnologia; (ii) Equipe de requisitos; (iii) Equipe de desenvolvimento; (iv) Stakeholders; e, (v) Usuários. Deste modo, podemos concluir que conforme a visão das participantes, o assunto deve ser abordado não apenas por grupos específicos, mas por todos os interessados.

Q7 - Para quem abordar confiança em IA? Em relação a para quem abordar confiança em IA, deve ser para os usuários da aplicação e pessoas afetadas com os resultados gerados, P1 “Para pessoas que vão utilizá-la...”, P3 “Em prol dos usuários da aplicação..”, P4 “Para os usuários que utilizaram a ferramenta.” e P2 “Para as pessoas que são afetadas com o resultado da IA”. A P1 ressalta “... Por isso é bom abordar com uma linguagem simples”.

6. Discussões e relatos das participantes

A forma como o estudo foi conduzido contribuiu para a disseminação do assunto, P2 “Achei bastante interessante a forma como foi explicada a importância da confiança em IA...”, ainda P2 relata que “...O exemplo real do sistema da amazon foi de bastante ajuda tanto para entender conceitos, como para entender a importância da abordagem do tema”. Trazer exemplos reais pode auxiliar na compreensão do tema e seus conceitos.

Os princípios para tratar a confiança em sistemas de IA apresentados por Kaur et al. (2022) e utilizado no estudo, tiveram relevância neste estudo e podem ser relevantes em outros projetos, como cita P5 “... Para que seja desenvolvido um sistema mais confiável para os usuários, é importante considerar os princípios abordados na pesquisa...”.

A técnica de *Brainwriting* apresentou-se eficiente para o estudo como comentado por P2 “O *Brainwriting* foi muito interessante para que de fato nós, participantes, pudéssemos absorver ainda mais o conteúdo que foi abordado na pesquisa, além de conseguir promover discussões que facilitaram ainda mais o entendimento dos pontos...”. A técnica promoveu interação, compartilhamento de conhecimentos e levantamento de ideias, consequentemente levantamento de possíveis requisitos, podendo ser aprimorada e aplicada em futuras pesquisas e ao longo do ciclo de desenvolvimento do projeto.

É importante envolver diversos interessados na discussão para a disseminação e evolução do tema. No entanto, é necessário buscar novas formas de promover essa interação, garantindo uma experiência participativa positiva e evitando exaustão e desconforto. A P4 deixou o seguinte comentário “... Outro ponto foi o tempo, levando a um desgaste dos participantes...”. P4 ainda cita “... durante a realização das discussões percebeu-se que por se tratar de um grupo diverso, com formações e áreas distintas, foi

mais difícil encontrar um ponto de intersecção na hora de realizar a “categorização” e adequar as falas de acordo com os itens disponibilizados, justiça, privacidade, aceitação, etc...”. Em relação à multidisciplinaridade dos participantes, é importante para enriquecer a discussão e reflexão, considerando diferentes perspectivas. No entanto, é necessário investigar melhores formas de envolver esses grupos diversos.

7. Limitações e ameaças à validade

Em relação às ameaças internas, a possível exaustão e desmotivação dos participantes devido ao tempo necessário para a aplicação do estudo podem afetar suas contribuições. Para mitigar esse problema, o estudo foi dividido em duas partes: uma síncrona, utilizando a técnica de *Brainwriting*, e outra assíncrona, com respostas às questões sobre IA confiável. Quanto às ameaças externas, a seleção cuidadosa dos participantes do estudo é relevante. Definimos o público-alvo com base no contexto apresentado, garantindo que os participantes fossem representativos em relação ao exemplo abordado. No entanto, é importante reconhecer que a aplicação do estudo com um público-alvo específico, baseado no contexto do problema, limita a generalização dos resultados para outras situações.

8. Conclusão e trabalhos futuros

Este trabalho investigou o uso de *Brainwriting* para a eliciação de requisitos de uma IA confiável, focando em AM e envolvendo colaboração multidisciplinar dos papéis interessados. Foram utilizados princípios para IA confiável propostos por Kaur et al. (2022), como base. O estudo foi aplicado em três etapas e contou com a participação de seis mulheres com perfis multidisciplinares e um mediador. O contexto do estudo envolveu o caso de um algoritmo utilizado em uma ferramenta de recrutamento da *Amazon* que apresentou problemas sexistas em seus resultados, conforme relatado por Dastin (2018).

A partir do estudo foi possível criar requisitos para uma IA confiável utilizando como base os princípios apresentados no trabalho de Kaur et al. (2022) e a técnica de *Brainwriting*. O estudo também permitiu o envolvimento multidisciplinar na discussão sobre IA confiável. Desta forma, *Brainwriting* se mostrou útil como técnica de suporte para eliciação de requisitos para IA confiável, assim como, envolvimento multidisciplinar, disseminação de conhecimento e reflexão dos aspectos e desafios sobre o assunto.

Trabalhos futuros: (i) definir um processo baseado em *Brainwriting* para eliciar requisitos para IA confiável; (ii) replicar o estudo em um projeto real em estágio inicial; e (iii) investigar a viabilidade e a contribuição da técnica de *Brainwriting* para a discussão multidisciplinar de IA confiável nas diversas fases do ciclo de vida de sistemas de AM.

Referências

- Ahmad, K., Bano, M., Abdelrazek, M., Arora, C., and Grundy, J. (2021). What’s up with requirements engineering for artificial intelligence systems? In *2021 IEEE 29th International Requirements Engineering Conference (RE)*, pages 1–12.
- Bellaby, R. W. (2021). Can ai weapons make ethical decisions? *Criminal Justice Ethics*, 40(2):86–107.
- Calazans, A. T. S., Paldês, R. Á., Masson, E. T. S., and Guimarães, F. d. A. (2018). As técnicas de eliciação e de documentação de requisitos nos métodos ágeis. In *CIBSE*, pages 487–500.

- Dastin, J. (2018). Amazon scraps secret ai recruiting tool that showed bias against women. In *Ethics of Data and Analytics*, pages 296–299. Auerbach Publications.
- de Cerqueira, J. A. S., De Azevedo, A. P., Leão, H. A. T., and Canedo, E. D. (2022). Guide for artificial intelligence ethical requirements elicitation-re4ai ethical guide. In *HICSS*, pages 1–10.
- de Oliveira Carvalho, N., Libório Sampaio, A., and de Vasconcelos, D. R. (2022). Morexai - a model to reason about the explanation design in ai systems. In Degen, H. and Ntoa, S., editors, *Artificial Intelligence in HCI*, pages 130–148, Cham. Springer International Publishing.
- Gruber, K., Huemer, J., Zimmermann, A., and Maschotta, R. (2017). Integrated description of functional and non-functional requirements for automotive systems design using sysml. In *2017 7th IEEE International Conference on System Engineering and Technology (ICSET)*, pages 27–31.
- Heslin, P. A. (2009). Better than brainstorming? potential contextual boundary conditions to brainwriting for idea generation in organizations. *Journal of Occupational and Organizational Psychology*, 82(1):129–145.
- Horkoff, J. (2019). Non-functional requirements for machine learning: Challenges and new directions. In *2019 IEEE 27th International Requirements Engineering Conference (RE)*, pages 386–391.
- Ishikawa, F. and Matsuno, Y. (2020). Evidence-driven requirements engineering for uncertainty of machine learning-based systems. In *2020 IEEE 28th International Requirements Engineering Conference (RE)*, pages 346–351. IEEE.
- Kaur, D., Uslu, S., Rittichier, K. J., and Duresi, A. (2022). Trustworthy artificial intelligence: A review. *ACM Comput. Surv.*, 55(2).
- Lazar, J., Feng, J. H., and Hochheiser, H. (2017). *Research methods in human-computer interaction*. Morgan Kaufmann.
- Michinov, N. (2012). Is electronic brainstorming or brainwriting the best way to improve creative performance in groups? an overlooked comparison of two idea-generation techniques. *Journal of Applied Social Psychology*, 42:E222–E243.
- Pazmino, A. V. (2015). *Como se cria: 40 métodos para design de produtos*. Editora Blucher.
- Pei, Z., Liu, L., Wang, C., and Wang, J. (2022). Requirements engineering for machine learning: A review and reflection. In *2022 IEEE 30th International Requirements Engineering Conference Workshops (REW)*, pages 166–175.
- Thiebes, S., Lins, S., and Sunyaev, A. (2021). Trustworthy artificial intelligence. *Electronic Markets*, 31:447–464.
- VanGundy, A. B. (1984). Brain writing for new product ideas: an alternative to brainstorming. *Journal of Consumer Marketing*.
- Yoshioka, N., Husen, J. H., Tun, H. T., Chen, Z., Washizaki, H., and Fukazawa, Y. (2021). Landscape of requirements engineering for machine learning-based ai systems. In *2021 28th Asia-Pacific Software Engineering Conference Workshops (APSEC Workshops)*, pages 5–8.