

A Process for Eliciting Trustworthy Requirements in Machine Learning Systems

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

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

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

{carlailane, ingrid}@ufc.br

Abstract. *This work proposes a process for Requirements Elicitation (RE) for Trustworthy Artificial Intelligence (AI), focused on Machine Learning (ML) systems. The process promotes the participation of various stakeholders involved in the development of these systems. It was developed based on the Brainwriting technique and insights obtained from a study that investigated this technique in the elicitation of requirements for Trustworthy AI. The study involved six female participants discussing Trustworthy AI based on the gender bias case of Amazon's algorithm, resulting in the creation of requirements for Trustworthy AI. This research contributes to filling gaps in RE for Trustworthy AI and promotes reflections on the challenges and aspects of this topic.*

Keywords. *Artificial intelligence; Machine learning; Trustworthy artificial intelligence; Requirement elicitation process; Brainwriting.*

Resumo. *Este trabalho propõe um processo para Elicitação de Requisitos (ER) para Inteligência Artificial (IA) Confiável, focado em sistemas de Aprendizado de Máquina (AM). O processo promove a participação dos diversos stakeholders envolvidos no desenvolvimento desses sistemas. Ele foi desenvolvido com base na técnica de Brainwriting e insights obtidos a partir de um estudo que investigou essa técnica na elicitação de requisitos para IA Confiável. O estudo envolveu seis participantes mulheres, discutindo IA confiável com base no caso de viés de gênero do algoritmo da Amazon, resultando na criação de requisitos para IA confiável. Esta pesquisa contribui para preencher lacunas na ER para IA confiável e promover reflexões sobre os desafios e aspectos desse tema.*

Palavras-Chave. *Inteligência artificial; Aprendizado de máquina; Inteligência artificial confiável; Processo de elicitação de requisitos; Brainwriting.*

1. Introdução

Parte das tarefas que antes eram executadas totalmente por humanos passaram a ser apoiadas 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 tratar o assunto, com abordagens que 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 relevantes 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 nesse processo que permite o envolvimento de grupos é a *Brainwriting* [Calazans et al. 2018]. Baseada em *Brainstorming*, essa técnica busca levantar soluções por meio de descrição a partir do envolvimento de grupos [Pazmino 2015]. Essa é uma abordagem eficiente para a geração de ideias, com o potencial de ajudar os participantes a superar a redução de produtividade observada no processo de *Brainstorming* [Michinov 2012].

Este trabalho é uma extensão de Silva et al. (2023), publicado no VIII Workshop sobre Aspectos Sociais, Humanos e Econômicos de Software e tem como objetivo propor um processo para a elicitação de requisitos para IA confiável. Esse processo foi desenvolvido a partir de *insights* obtidos da investigação do uso da técnica *Brainwriting*, com foco em sistemas de AM aplicados ao recrutamento e seleção de candidatos, através da colaboração multidisciplinar dos envolvidos. Para isso, foi conduzido um estudo de caso exploratório com um grupo de seis participantes mulheres, de diferentes perfis de atuação, utilizando a técnica e questões abertas para discutir IA confiável. O estudo tomou como base o caso apresentado por Dastin (2018), que aborda o algoritmo de recrutamento da *Amazon*, considerado sexista devido a problemas de viés de gênero. A pesquisa

fundamentou-se nos princípios de IA confiável apresentados por Kaur et al. (2022), que incluem Justiça, Explicabilidade, Responsabilidade, Privacidade e Aceitação.

As principais contribuições deste trabalho são: (1) O desenvolvimento de um processo para a elicitação de requisitos para IA confiável, que promove a participação e colaboração de diversos *stakeholders*; e, (2) A demonstração da viabilidade da técnica de *Brainwriting* na elicitação de requisitos colaborativos e na obtenção de uma diversidade de opiniões, essencial para a área de Sistemas de Informação.

Este artigo está organizado da seguinte forma: a fundamentação teórica (Seção 2), os trabalhos relacionados (Seção 3), o estudo de caso exploratório sobre o uso da técnica *Brainwriting* na elicitação de requisitos para IA confiável (Seção 4), os resultados do estudo de caso exploratório (Seção 5), as discussões e o relato das participantes do estudo de caso exploratório (Seção 6), a apresentação do processo baseado em *Brainwriting* para a elicitação de requisitos para IA confiável (Seção 7), as limitações e ameaças à validade (Seção 8) e a conclusão e os trabalhos futuros (Seção 9).

2. Fundamentação teórica

Nesta seção, são apresentados os principais conceitos para a definição deste estudo: (2.1) Engenharia de Requisitos para AM; (2.2) *Brainwriting*; e, (2.3) IA confiável.

2.1. Engenharia de Requisitos para Aprendizado de Máquina

A Engenharia de Requisitos (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]. De acordo com Villamizar et al. (2021), a ER possui grande relevância no processo de desenvolvimento de aplicações que envolvem AM, tendo em vista que um modelo de AM é uma especificação feita a partir de dados de treinamento, que estabelece o comportamento do sistema com base no conhecimento adquirido.

A ausência da ER ou sua realização inadequada pode acarretar inconsistências e, consequentemente, 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, com 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 interfiram, deixando os participantes inseguros ao expressar suas ideias diante de pessoas com outro *status* social ou nível de conhecimento [VanGundy 1984].

Na literatura, são encontradas algumas variantes da técnica de *Brainwriting*. Com base em Boy (1996) e VanGundy (1984), na Tabela 1 são descritos o *Brainwriting* tradicional e algumas variantes.

Tabela 1. *Brainwriting* tradicional e algumas variantes

Nome da Técnica	Descrição da Técnica
<i>Brainwriting Tradicional</i>	Os participantes escrevem individualmente pontos de vista, perspectivas ou opiniões sobre um assunto em folhas de papel. As ideias são compartilhadas e enriquecidas através da leitura e adição de pontos de vista dos outros participantes, estimulando a reflexão crítica e a interação.
<i>Nominal Group Technique (NGT)</i>	Os participantes geram ideias silenciosamente por escrito, sem discussão. Ideias são listadas e esclarecidas verbalmente. Em seguida, os participantes votam para classificar as ideias por importância, resultando em uma lista priorizada.
<i>Collective Notebook (CNB)</i>	Os participantes recebem cadernos e escrevem ideias diariamente sobre um problema específico. Após um mês, os cadernos são coletados, e as ideias são organizadas em categorias e discutidas pelo grupo.
<i>Brainwriting Pool</i>	Um pequeno grupo gera ideias individualmente e as passa para outros membros, que as usam como estímulo para gerar novas ideias. Esse processo de troca e expansão das ideias é repetido para enriquecer a geração criativa.
<i>Pin Cards</i>	Os participantes escrevem ideias em fichas e as passam para o membro ao lado. O processo de escrita e passagem continua, permitindo a inspiração a partir das ideias dos outros. As fichas são coletadas, organizadas e discutidas pelo grupo.

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]. De acordo com Li et al. (2023), a IA confiável é um campo multidisciplinar que vai além da Ciência da Computação e do AM, impactando profundamente a sociedade e exigindo soluções rápidas para diversos tópicos de pesquisa.

De acordo com Kaur et al. (2022), existe uma grande quantidade de princípios propostos para tornar a IA confiável, sendo difícil encontrar um conjunto comum. Desta forma, neste estudo serão utilizados como base os cinco princípios/requisitos essenciais para projetar sistemas de IA confiáveis, revisados e discutidos por Kaur et al. (2022), que são:

- **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 consequências prejudiciais ao usuário; e
- **Aceitação:** aumentar a aceitação do sistema pelos usuários.

3. Fundamentação teórica

Nesta seção, são abordados os conceitos fundamentais relacionados à Engenharia de Requisitos para AM, *Brainwriting*, e IA confiável. A compreensão desses conceitos é crucial para o desenvolvimento e aplicação eficaz de métodos para a elicitación de requisitos em sistemas de IA confiáveis.

3.1. Engenharia de Requisitos para Aprendizado de Máquina

A Engenharia de Requisitos (ER) é uma fase essencial no desenvolvimento de sistemas de software, crucial para a definição de produtos que atendam às necessidades dos *stakeholders* [Ishikawa and Matsuno 2020]. No contexto de AM, a ER assume uma importância adicional, pois um modelo de AM é essencialmente uma especificação baseada em dados de treinamento, que define o comportamento do sistema com base no conhecimento adquirido [Villamizar et al. 2021].

A realização inadequada da ER pode levar a inconsistências e custos adicionais para o projeto [Gruber et al. 2017]. Desafios específicos surgem na definição e gestão de requisitos para sistemas de IA/AM, como observado por Horkoff (2019), devido à natureza complexa e dinâmica desses sistemas.

3.2. *Brainwriting*

O *Brainwriting* é uma técnica eficaz para a geração de ideias, aplicada silenciosamente por escrito, o que estimula a criatividade e a concentração dos participantes [VanGundy 1984]. Em contraste com o *Brainstorming*, o *Brainwriting* permite que ideias sejam desenvolvidas sem a influência de status social ou pressão imediata, promovendo um ambiente mais equitativo e produtivo [Heslin 2009].

Embora existam diversas variantes do *Brainwriting*, como o *Nominal Group Technique* e o *Collective Notebook*, todas visam enriquecer o processo criativo e podem ser aplicadas na elicitación de requisitos, facilitando a contribuição de diferentes *stakeholders* e a coleta de perspectivas diversas.

4. Trabalhos relacionados

Kaur et al. (2022) realizaram uma análise abrangente dos requisitos essenciais para garantir a confiabilidade dos sistemas de IA. Seu estudo não apenas destaca os requisitos cruciais, mas também discute estratégias de mitigação de riscos, validação e verificação, oferecendo uma visão completa dos avanços recentes em IA confiável.

Por sua vez, Ahmad et al. (2021) conduziram uma revisão sistemática da literatura sobre métodos de Engenharia de Requisitos (ER) para IA, identificando uma ampla gama de abordagens e técnicas utilizadas na área. Embora seu foco não esteja diretamente em requisitos de IA confiável, sua análise destaca lacunas importantes na compreensão e abordagem dos requisitos em sistemas de IA.

Pei et al. (2022) ofereceram uma visão detalhada do processo de ER para AM, enfatizando a importância da colaboração entre os diferentes papéis envolvidos. Embora não tenham explorado especificamente os requisitos de IA confiável, seu estudo fornece *insights* sobre o desenvolvimento colaborativo de sistemas de AM, que podem ser aplicados à construção de sistemas de IA confiáveis.

Carvalho et al. (2022) propuseram um modelo conceitual baseado na Engenharia Semiótica para refletir sobre a IA Explicável (MoReXAI), enfatizando a importância da explicabilidade na compreensão e aceitação dos sistemas de IA. Embora seu foco principal seja na explicabilidade, a reflexão sobre temas éticos é fundamental para a confiabilidade dos sistemas de IA.

Por fim, Cerqueira et al. (2022) desenvolveram o Guia Ético RE4AI para elicitación de requisitos éticos de IA, oferecendo uma abordagem prática para lidar com questões éticas na concepção de sistemas de IA. Embora seu foco principal esteja na ética, a consideração dessas questões é crucial para garantir a confiabilidade dos sistemas de IA.

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, não abordando 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. Cerqueira et al. (2022) apresentam um guia para elicitación de requisitos éticos de IA, já Carvalho et al. (2022) propõem um modelo para refletir sobre IA explicável. Embora o tópico de confiabilidade esteja inserido nesses trabalhos, eles não se dedicam a falar de confiabilidade como tema principal.

Este trabalho vai além das investigações anteriores ao explorar a técnica de *Brainwriting* como uma abordagem eficaz para a elicitación de requisitos para IA confiável. Além da investigação, é proposto um modelo de processo que incorpora o *Brainwriting* como uma ferramenta fundamental. Este processo visa envolver ativamente os diversos *stakeholders* interessados, considerando os sistemas de AM e as questões de confiabilidade associadas. Essa abordagem holística não apenas identifica requisitos para IA confiável, mas também estabelece um método estruturado que promove uma participação mais ampla e uma abordagem mais abrangente para lidar com os desafios emergentes na área.

5. Materiais e métodos

Esta seção descreve a abordagem adotada para a condução da pesquisa, cujo objetivo é desenvolver um processo para a elicitación de requisitos de sistemas de IA confiáveis. Para isso, foi conduzido um estudo de caso inicial, investigando o uso da técnica *Brainwriting* na elicitación de requisitos para sistemas de AM. Com base nos *insights* obtidos, foi cri-

ado o processo. Esta seção está dividida em duas partes principais: a primeira aborda o estudo de caso exploratório, enquanto a segunda parte descreve o desenvolvimento e a estruturação do processo.

5.1. Estudo de caso exploratório

O estudo inicial foi um estudo de caso exploratório, com foco na aplicação da técnica de *Brainwriting* para a eliciação de requisitos para sistemas de IA confiável. O contexto escolhido foi o de sistemas de AM aplicados ao recrutamento e seleção de candidatos, dada a relevância e atualidade do tema. O estudo de caso foi conduzido em três etapas principais:

1. **Planejamento e Preparação:** nesta fase, foram definidos o contexto do estudo, a escolha dos participantes e a preparação dos materiais necessários. Participaram seis mulheres com perfis multidisciplinares e um mediador. Utilizou-se um caso real do algoritmo de recrutamento da Amazon, que apresentou viés sexista, como base para as discussões.
2. **Execução do Brainwriting:** durante a execução, os participantes participaram de sessões de *Brainwriting*, uma técnica que promove a geração de ideias de forma colaborativa e anônima. As sessões foram estruturadas para permitir que os participantes expressassem suas opiniões sobre os requisitos para IA confiável, com base nos princípios estabelecidos por Kaur et al. (2022).
3. **Análise e Extração de Requisitos:** após as sessões de *Brainwriting*, as ideias geradas foram analisadas e categorizadas de acordo com os princípios de IA confiável. Essa análise possibilitou a identificação de requisitos relevantes para o desenvolvimento de sistemas de IA confiáveis.

5.1.1. Participantes e Contexto

Os participantes foram selecionados com base em suas experiências e conhecimentos multidisciplinares, garantindo uma diversidade de perspectivas. O estudo utilizou um caso real de um sistema de recrutamento da Amazon para destacar a importância de considerar aspectos éticos e de confiança no desenvolvimento de sistemas de IA.

5.1.2. Métodos de Coleta e Análise de Dados

Os dados foram coletados durante as sessões de *Brainwriting* e nas discussões subsequentes. A análise envolveu a categorização das ideias geradas e a identificação de padrões. Os princípios de Kaur et al. (2022) foram usados como estrutura para a análise e extração dos requisitos.

5.2. Desenvolvimento do Processo

Com base nos *insights* obtidos no estudo de caso, foi desenvolvido o Processo para Elicitação de Requisitos para IA Confiável (PERIAC), composto por quatro fases principais:

1. **Planejamento e Preparação:** definição do escopo do projeto, seleção dos *stakeholders* e preparação dos materiais necessários.
2. **Execução:** aplicação da técnica de *Brainwriting* para a geração colaborativa de ideias e requisitos, utilizando um conjunto de perguntas baseadas no estudo de Carvalho et al. (2022).
3. **Análise e Extração de Requisitos:** categorização e análise das ideias geradas, com base nos princípios de Justiça, Explicabilidade, Responsabilidade, Privacidade e Aceitação.
4. **Validação dos Requisitos:** revisão e validação dos requisitos identificados, com a participação dos *stakeholders* para garantir a relevância e viabilidade dos mesmos.

A concepção do processo proposto resulta da junção de etapas da metodologia tradicional de *Brainwriting*, incorporando *insights* de um estudo de caso exploratório apresentado na Seção 4 deste trabalho. Além disso, é enriquecido pela seleção criteriosa de questões formuladas no Modelo para Reflexão sobre IA Explicável (MoReXAI), proposto no estudo Carvalho et al. (2022). Essas questões foram cuidadosamente escolhidas para refletir os objetivos deste estudo, que visa investigar e levantar requisitos para IA Confiável. Adicionalmente, um conjunto de cartas foi desenvolvido, cada uma contendo a definição de um princípio, exemplos e um conjunto de perguntas relevantes. Essa combinação de abordagens confere ao PERIAC uma estrutura flexível e abrangente para a elicitación de requisitos de IA confiável.

Durante o desenvolvimento do processo, tornou-se evidente a necessidade premente de estabelecer um conjunto de princípios para a IA Confiável, que servissem como norteadores. Diante da vasta diversidade de princípios propostos para garantir a confiabilidade da IA, foi desafiador encontrar um conjunto coeso de diretrizes para lidar com esse tema complexo e em constante evolução. Assim, este estudo optou por adotar como referência os cinco requisitos essenciais para IA Confiável, conforme revisados e discutidos por Kaur et al. (2022).

A escolha desse estudo foi motivada pela abordagem metódica na revisão e discussão de um conjunto de requisitos fundamentais para a IA confiável, culminando na identificação de cinco requisitos comuns, já apresentados na Seção 2.4, sendo eles: Justiça, Explicabilidade, Responsabilidade, Privacidade e Aceitação.

Diante da necessidade de estabelecer um conjunto de perguntas que guiasse o processo, fornecendo direcionamento para a reflexão e o levantamento de requisitos, foram buscadas questões que pudessem ser integradas ao processo. Identificou-se que no estudo Carvalho et al. (2022) é proposto um conjunto de 27 perguntas, relacionadas a tópicos considerados importantes pelo autor para IA explicável (Privacidade e Controle Humano; Responsabilidade e Prestação de Contas; Confiabilidade e Segurança; Transparência e Explicabilidade; Justiça, Equidade e Não Discriminação), das quais 15 são centradas em dados e 12 no modelo de AM. O objetivo dessas perguntas é auxiliar na reflexão e na elicitación de requisitos para IA Explicável. Essas perguntas resultaram de mapeamentos de conjuntos de perguntas abordadas em outros trabalhos da literatura.

Dessa forma, optou-se por selecionar algumas perguntas relevantes para o processo, auxiliando na reflexão e na elicitación de requisitos para IA confiável. Inicialmente,

todas as perguntas relacionadas ao tópico de Confiabilidade e Segurança foram escolhidas por serem o foco central deste trabalho. Em seguida, o conjunto de perguntas descartadas foi analisado para verificar se alguma pergunta importante poderia ser reconsiderada. Após essa análise, todas as perguntas selecionadas foram reavaliadas com o intuito de identificar sua relevância para o contexto do projeto. Na Tabela 2 são apresentadas as perguntas selecionadas.

Com o objetivo de enriquecer e orientar o processo de *Brainwriting*, proporcionando aos participantes uma compreensão mais abrangente das questões a serem consideradas e auxiliando o facilitador na organização e extração dos requisitos, um conjunto de cartas foi desenvolvido. Cada carta oferece uma definição do princípio em questão, um exemplo fictício e uma série de perguntas destinadas a orientar os participantes na reflexão e no levantamento de requisitos.

O fluxo de aplicação do PERIAC e conjunto de cartas desenvolvidas são apresentados na Seção 9 deste trabalho.

6. Estudo de caso exploratório

Este estudo foi conduzido como um estudo de caso exploratório com elementos experimentais, visando investigar o uso da técnica de *Brainwriting* na elicitação de requisitos para IA confiável. Nesta seção são descritas as etapas e atividades seguidas no estudo. O estudo é constituído por três etapas principais: (6.1) Planejamento e preparação; (6.2) Aplicação; e (6.3) Análise e extração de requisitos. O fluxo das etapas é ilustrado na Figura 1.

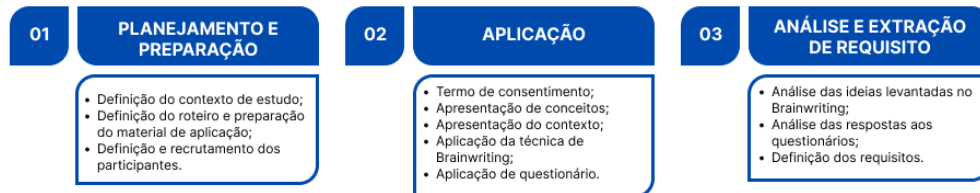


Figura 1. Etapas do estudo.

6.1. 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 nesta ferramenta, 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

Tabela 2. Perguntas selecionada do estudo Carvalho et al.(2022)

ID	Perguntas
Perguntas centradas em dados	
P1	O que representam as instâncias que compõem o conjunto de dados? Existem vários tipos de instâncias? As relações entre instâncias individuais são explicitadas (por exemplo, classificação de filmes dos usuários, links de redes sociais)?
P2	O conjunto de dados representa todas as instâncias, ou é uma amostra de um conjunto maior? Se é uma amostra, qual foi a estratégia de amostragem?
P3	Há alguma informação faltando em instâncias individuais? Que estratégia foi usada para balancear o conjunto de dados? Como foi validada essa estratégia? Existem erros, fontes de ruído ou redundâncias no conjunto de dados?
P4	Os conjuntos de dados permanecem constantes ou podem ser modificados, ou excluídos ao longo do tempo?
P5	O conjunto de dados está relacionado a pessoas? Contém dados que podem ser considerados confidenciais ou sensíveis? Contém dados que, se visualizados diretamente, podem ser ofensivos, insultuosos, ameaçadores ou podem causar ansiedade?
P6	O conjunto de dados identifica alguma subpopulação (por idade, gênero)? Se sim, como essas subpopulações são identificadas e como elas estão distribuídas no conjunto de dados?
P7	É possível identificar indivíduos, direta ou indiretamente, a partir do conjunto de dados?
P8	Como foi feita a coleta de dados? Quais mecanismos ou procedimentos foram usados? Como esses mecanismos ou procedimentos foram validados? Quem estava envolvido no processo de coleta de dados?
P9	Durante qual período de tempo os dados foram coletados? Esse período de tempo corresponde ao período de criação dos dados associados às instâncias?
P10	Foram realizados processos de revisão ética (por exemplo, por um conselho de revisão institucional)?
P11	Há algo sobre a composição do conjunto de dados ou a maneira como ele foi coleta e pré-processado/ limpo/rotulado que possa afetar usos futuros? Há algo que um futuro usuário deve saber para evitar usos que resultem em tratamentos injustos de indivíduos ou grupos, ou outros danos indesejáveis? Existe algo que um futuro usuário poderia fazer para mitigar esses danos indesejáveis?
P12	Existem tarefas para as quais o conjunto de dados não deve ser usado?
Perguntas centradas no modelo	
P13	Qual o tipo de modelo de AM está sendo desenvolvido? Poderia explicar como ele funciona?
P14	Qual a principal intenção de uso do modelo? Quem são os usuários pretendidos, que serão atendidos pelo modelo? Algum outro cenário de uso fora desse escopo?
P15	O modelo pode afetar de alguma forma grupos demográficos ou fenotípicos? Qual a influência que esses fatores podem ter no desempenho do modelo? Como é feita a avaliação da influência desses fatores no desempenho do modelo?
P16	Os instrumentos ou o ambiente de coleta dos conjuntos de dados pode influenciar de alguma forma o resultado do modelo? Qual a influência que esses fatores podem ter no desempenho do modelo? Como é feita a avaliação da influência desses fatores no desempenho do modelo?
P17	Quais métricas de desempenho do modelo estão sendo relatadas e por que foram selecionadas em vez de outras métricas de desempenho? Como elas são calculadas?
P18	Quais conjuntos de dados foram usados para avaliar o modelo? Por que esses conjuntos de dados foram utilizados? Como os dados foram pré-processados para avaliação?
P19	São feitas análises quantitativas com relação a subgrupos populacionais desagregados do conjunto de dados? Como o modelo se saiu em relação a cada fator? Como o modelo se saiu em relação à intersecção dos fatores avaliados?

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. Visando assegurar padrões éticos, foi utilizado um termo de consentimento informado para todos os participantes, que foram claramente informados sobre os objetivos, métodos, benefícios, riscos e seu direito de se retirar do estudo a qualquer momento. O material de apoio pode ser visualizado no relatório técnico¹.

A ferramenta de suporte adotada para o estudo foi o Google Jamboard², 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 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?*”.

Posteriormente, 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 nos chats das plataformas de redes sociais WhatsApp³ e LinkedIn⁴, porém, devido a limitações de tempo, apenas seis pessoas aceitaram participar do estudo. Na Tabela 3 é apresentado o perfil das participantes. O responsável por aplicar e conduzir o processo foi o primeiro autor do artigo, que possui experiência com Pesquisa, Requisitos e *Design*.

Para finalizar, o estudo passou por uma revisão realizada por especialistas nas áreas de Engenharia de Software e Interação Humano-Computador, que avaliaram e aprovaram os procedimentos e materiais utilizados.

6.2. 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 Meet⁵. Inicialmente, o condutor do estudo distribuiu para todas as participantes um documento de

¹<https://zenodo.org/record/7811847>

²<https://jamboard.google.com/>

³<https://web.whatsapp.com/>

⁴<https://www.linkedin.com>

⁵<https://meet.google.com/>

Tabela 3. Perfil das participantes.

ID	Perfil
P1	Design
P2	Desenvolvimento
P3	Qualidade/Pesquisa
P4	Ensino/Pesquisa
P5	Pesquisa
P6	Recursos Humanos/Recrutamento

apoio com o termo de consentimento para participação na pesquisa, a definição dos principais conceitos utilizados e as questões para investigar a IA confiável. O termo de consentimento teve como objetivo assegurar padrões éticos, informar os objetivos do estudo, garantir o anonimato das participantes e esclarecer o direito de se retirar do estudo a qualquer momento. 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 o contexto de discussão baseado em Dastin (2018).

Após a compreensão de todos os envolvidos sobre o processo adotado no estudo e o contexto de discussão, iniciou-se a etapa de *Brainwriting*, sendo apresentada a seguinte questão: “*Como desenvolver sistemas de recrutamento e seleção de candidatos utilizando IA, livres de discriminação e confiáveis para os usuários?*”. As participantes foram incentivadas 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 escritas na ferramenta Google Jamboard, em quadros separados, preservando a visualização individual de cada participante, durante um período de 10 minutos. Em seguida, as ideias geradas 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 precisou deixar o estudo devido a outros compromissos pessoais que surgiram. Desta forma, os seus dados foram desconsiderados para o estudo. A Figura 2 mostra o registro do momento da aplicação do estudo. Esta etapa teve duração de 2 horas e 30 minutos.

Após a etapa de *Brainwriting*, as participantes foram liberadas, com o pedido de refletirem 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 técnico⁶.

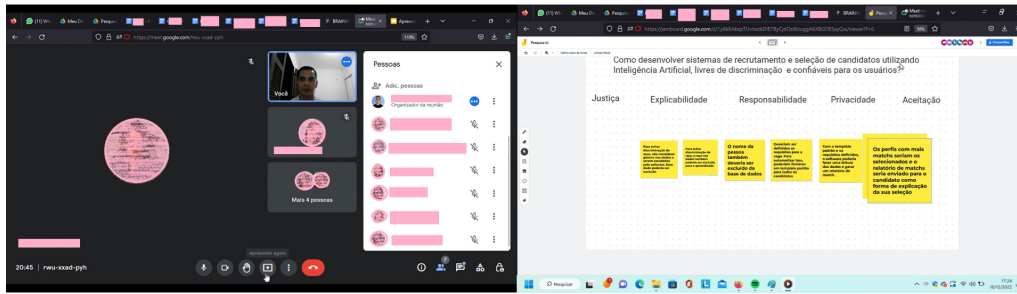


Figura 2. Registro da aplicação do estudo.

6.3. Análise e extração de requisitos

Inicialmente, as ideias originadas da aplicação da técnica de *Brainwriting* foram registradas em uma planilha eletrônica, analisadas, categorizadas com identificadores e alinhadas aos princípios para abordar a confiança em sistemas de IA conforme delineado por Kaur et al. (2022).

Posteriormente, foram agrupadas por similaridade e transformadas em requisitos. O relatório técnico contendo as 22 ideias está disponível no relatório técnico⁷. Assim, essas 22 ideias foram condensadas em 14 requisitos, detalhados na próxima seção.

Em seguida, as questões sobre IA confiável, distribuídas para que as participantes respondessem, passaram por uma análise qualitativa devido à sua natureza descritiva. Os métodos de análise qualitativa lidam com variáveis que não podem ser quantificadas, apenas observadas [Lazar et al. 2017].

Esta análise levou em conta a percepção das participantes sobre IA confiável. O processo de análise envolveu a transcrição das respostas para uma planilha eletrônica, identificação de relações entre as respostas das participantes e identificação de padrões e temas importantes, levando em consideração as limitações de cada questão. Todas as atividades desta etapa foram conduzidas pelo primeiro autor do artigo. Os resultados são detalhados na próxima seção.

7. Resultados do estudo de caso exploratório

Esta Seção apresenta os resultados obtidos através do estudo de caso exploratório. A Seção está organizada da seguinte forma: (7.1) Resultados sobre a aplicação da técnica de *Brainwriting* e requisitos extraídos; e, (7.2) Resultados sobre IA confiável a partir da visão das participantes.

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

No processo de mapeamento entre os requisitos e os princípios para IA Confiável, as ideias geradas pela técnica de *Brainwriting* foram registradas, analisadas e categorizadas.

⁶<https://zenodo.org/record/7811847>

⁷<https://zenodo.org/record/7811877>

A análise envolveu a transcrição das respostas, a identificação de relações entre elas e a detecção de padrões e temas importantes. Esse processo proporcionou uma compreensão estruturada dos requisitos necessários para garantir a confiança em sistemas de IA.

A Tabela 4 apresenta 14 Requisitos Não Funcionais extraídos das 22 ideias levantadas durante a aplicação da técnica de *Brainwriting* e princípios para IA confiável abordados por Kaur et al. (2022) relacionados.

7.2. Resultados sobre IA confiável a partir da visão das participantes

Esta Subseção apresenta uma análise detalhada das percepções das participantes sobre IA confiável, com base em suas respostas a uma série de perguntas estruturadas. Cada Subsubseção a seguir explora um aspecto específico sobre IA confiável, destacando as diversas opiniões e *insights* das participantes.

7.2.1. Q1 - Por que abordar confiança em IA?

A primeira pergunta busca entender as razões para abordar a confiança em IA.

A participante P1 cita que “As IAs estão cada vez mais presentes no nosso cotidiano, algumas empresas já adotam IAs para atendimento aos clientes, outras em processos criativos. Ao mesmo tempo que vemos essa alta, também observamos notícias sobre o comportamento de algumas dessas IAs, e isso pode gerar desconfiança por ser uma abordagem relativamente nova. É importante abordar a confiança para que os usuários se sintam confortáveis e possam ter uma melhor experiência justamente pela confiabilidade repassada.”. De acordo com a visão da participante P2, “Porque a IA deve simular os comportamentos humanos. Dessa forma, uma vez que a ética é tão importante no mundo real, ela também deve ser levada em consideração na atuação de sistemas cuja IA é implementada.”. A participante P3 cita que “Para que possa ter resultados mais efetivos.”. De acordo com a participante P4, “Para garantir que os sistemas utilizados em nossa sociedade estejam livres de preconceitos, discriminação e exclusão de grupos.”. Já a participante P5 diz, “Para possibilitar o desenvolvimento de um sistema seguro, justo, confiável e que gere resultados satisfatórios para os usuários.”.

Como pode-se observar, a motivação para abordar a confiança em IA apresenta-se de maneira complementar entre as participantes. A confiança em IA pode tratar de questões éticas, permitir o desenvolvimento de sistemas livres de preconceitos, discriminação e exclusão de grupos, possibilitar resultados mais efetivos e proporcionar uma melhor experiência para os usuários pela confiabilidade transmitida.

7.2.2. Q2 - O que abordar dentro do contexto de confiança em IA?

A segunda pergunta explora os elementos que devem ser abordados para promover a confiança em IA.

De acordo com as participantes, deve-se abordar:

Tabela 4. Requisitos extraídos das ideias levantadas a partir da aplicação da técnica de *Brainwriting*

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

(I) Transparência de Dados e Forma de Tratamento - P1 afirma: “Acredito que a transparência sobre como os dados vão ser tratados é essencial...” e P4 complementa: “Transparência de dados...”;

(II) Imparcialidade - P1 menciona que “...é sempre bom mostrar que essa IA segue princípios que não desrespeitam ou excluem grupos.” P2 acrescenta: “É preciso abordar se problemas de xenofobia, homofobia, racismo e discriminações de modo geral, não estão sendo praticados com o uso da IA...” e P3 reforça com “Imparcialidade”;

(III) Princípios Éticos - P4 sugere “... princípios éticos.”;

(IV) Leis Locais - P2 destaca que “...leis do local em que a IA está sendo aplicada também podem ser levadas em consideração nessa análise de confiança, pois dependendo do contexto elas também podem acabar sendo infringidas.”;

(V) Princípios de Confiança - P5 diz: “Considerar os princípios de confiança em IA (Justiça, Explicabilidade, Responsabilidade, Privacidade, Aceitação).”;

Desta forma, observa-se que os itens levantados pelas participantes possuem relação com os princípios propostos por Kaur et al. (2022) para tratar a confiança em sistemas de IA.

7.2.3. Q3 - Quando abordar confiança em IA?

Para P1, a confiança em IA deve ser abordada “Sempre que ela for ser utilizada.”. Já as demais participantes sugerem abordagens em momentos ou situações específicas. P2 afirma que a confiança deve ser abordada “Quando a IA a ser utilizada provê resultados difíceis de serem revertidos...”. P3 diz “Sempre que for lidar com dados sensíveis.”. P5 sugere abordar a confiança “... ao se trabalhar em sistemas que façam o uso de informações críticas e confidenciais.”. E P4 complementa “Durante todo o processo inicial, para que de fato a IA possa diagnosticar e corrigir possíveis falhas futuras.”.

As respostas destacam que a confiança em IA deve ser abordada tanto de forma contínua quanto em momentos críticos específicos, garantindo que a IA opere de maneira confiável ao longo de seu uso e desenvolvimento.

7.2.4. Q4 - Como abordar confiança em IA?

A quarta pergunta trata das maneiras de abordar a confiança em IA.

Para P1, a confiança deve ser tratada através de “princípios regulamentadores.”. P4 sugere “Envolvendo discussões amplas para garantir que a equipe minimize os riscos e perceba a importância dos aspectos éticos.”. P2 propõe “Estudando os resultados gerados em sistemas que utilizam a IA para tomar alguma decisão para que seja possível verificar se os resultados são justos ou se, de alguma forma, foram discriminatórios ou infringiram alguma lei.”. P5 recomenda “Realizar um estudo de como aplicar os princípios de confiança (Justiça, Explicabilidade, Responsabilidade, Privacidade, Aceitação) no sistema a ser desenvolvido.” P5 ressalta que esses estudos devem “ser feitos 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.”

As sugestões indicam que a confiança em IA pode ser abordada de várias maneiras, incluindo a implementação de princípios regulamentadores, discussões éticas, estudos de resultados e aplicação de princípios de confiança, com o objetivo de garantir segurança e justiça para os usuários.

7.2.5. Q5 - Onde abordar confiança em IA?

A quinta pergunta discute os locais e etapas onde a confiança em IA deve ser abordada.

P3 sugere abordar ”Em todas as etapas do desenvolvimento.”. Essa ideia é reforçada por P5 ”No processo de desenvolvimento e manutenção do sistema. É importante aplicar os princípios de confiança, desde a elicitação dos requisitos do sistema até a sua fase de manutenção e implantação. Para garantir que o sistema continue atendendo aos requisitos exigidos.”. P4 complementa ”Com a equipe, dentro da empresa, antes que o produto final seja apresentado para a sociedade.”. P1 sugere que o assunto seja abordado ”Em debates, pesquisas e empresas desenvolvedoras.”. P2 afirma que a confiança deve ser abordada ”Em sistemas que utilizam a IA para gerar resultados que afetam as pessoas.”.

As participantes concordam que a confiança em IA deve ser abordada em todas as etapas do desenvolvimento e manutenção do sistema, bem como em debates e pesquisas, assegurando a inclusão dos princípios de confiança desde o início até a implementação e além.

7.2.6. Q6 - Por quem deve ser abordada confiança em IA?

A sexta pergunta investiga quem deve ser responsável por abordar a confiança em IA.

As participantes indicam que o assunto deve ser abordado por diversos grupos, P1 menciona ”Pela empresa que desenvolveu...”, P3 diz ”Pela equipe de requisitos”, P4 acrescenta ”A equipe de desenvolvimento...” e P5 inclui ”... stakeholders envolvidos no projeto do sistema a ser desenvolvido.”. Além disso, a abordagem do assunto também deve envolver os usuários, P2 afirma que deve ser ”Pelos pessoas que a utilizam como forma de facilitar o seu trabalho...”, P4 menciona ”... possíveis usuários.” e P5 reforça ”Pelos usuários.”. Pesquisadores da área também são citados por P1: ”... por pesquisadores da área...”

De acordo com as participantes, a confiança em IA deve ser abordada por todos os envolvidos, incluindo desenvolvedores, equipes de requisitos, usuários e pesquisadores, garantindo uma abordagem abrangente e colaborativa.

7.2.7. Q7 - Para quem abordar confiança em IA?

A sétima pergunta aborda o público-alvo para quem a confiança em IA deve ser direcionada.

As participantes enfatizam que deve ser para os usuários da aplicação e pessoas afetadas pelos resultados gerados, P1 menciona "Para pessoas que vão utilizá-la...". P3 diz "Em prol dos usuários da aplicação.", P4 afirma "Para os usuários que utilizaram a 'ferramenta'." e P2 destaca "Para as pessoas que são afetadas com o resultado da IA.". P1 ressalta que a abordagem deve ser feita "... com uma linguagem simples." para garantir a compreensão de todos.

A confiança em IA deve ser abordada de maneira acessível e clara, visando principalmente os usuários e aqueles afetados pelos resultados da IA, garantindo que todos compreendam os princípios e práticas envolvidos.

8. Discussões e relatos das participantes do estudo de caso exploratório

A forma como o estudo foi conduzido contribuiu para a disseminação do assunto. P2 expressou: *Achei bastante interessante a forma como foi explicada a importância da confiança em IA...*. Além disso, P2 relatou que *O exemplo real do sistema da Amazon foi de bastante ajuda tanto para entender conceitos, como para compreender a importância da abordagem do tema*. Trazer exemplos reais pode auxiliar na compreensão do tema e de seus conceitos.

Os princípios para tratar a confiança em sistemas de IA, apresentados por Kaur et al. (2022) e utilizados no estudo, tiveram relevância neste estudo e podem ser relevantes em outros projetos, como citou P5: *"...Para desenvolver um sistema mais confiável para os usuários, é importante considerar os princípios abordados na pesquisa..."*.

A técnica de *Brainwriting* mostrou-se eficiente para o estudo, como comentado por P2: *"O Brainwriting foi muito interessante para que nós, participantes, pudéssemos absorver ainda mais o conteúdo abordado na pesquisa, além de promover discussões que facilitaram o entendimento dos pontos..."*. A técnica promoveu interação, compartilhamento de conhecimentos e levantamento de ideias, consequentemente, o 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. P4 deixou o seguinte comentário: *... Outro ponto foi o tempo, levando a um desgaste das participantes...*. P4 também mencionou: *...durante as 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, como justiça, privacidade, aceitação, etc...*. Em relação à multidisciplinaridade das participantes, é importante para enriquecer a discussão e reflexão, considerando diferentes perspectivas. No entanto, é necessário investigar melhores formas de envolver esses grupos diversos.

9. Processo para elicitación de requisitos para IA confiável

Esta Seção apresenta o Processo para Elicitación de Requisitos para IA Confiável (PERIAC) e o conjunto de cartas elaboradas para auxiliar. O PERIAC é concebido para fornecer um guia aos projetistas de sistemas de IA durante a etapa de elicitação de requisitos. Seu objetivo primordial é auxiliar na reflexão e facilitar a identificação e especificação de requisitos para IA Confiável. Além disso, o PERIAC foi projetado para promover a participação e colaboração dos diversos *stakeholders* envolvidos no desenvolvimento de sistemas de IA.

9.1. Fluxo de Aplicação do PERIAC

O processo proposto para a elicitação de requisitos para IA confiável foi criado com base na técnica *Brainwriting* e nos conhecimentos obtidos no estudo apresentado na Seção 4 deste trabalho. As fases principais do processo são: (I) Planejamento e Preparação (Seção 9.1.1); (II) Execução (Seção 9.1.2); (III) Análise e Extração de Requisitos (Seção 9.1.3); e (IV) Validação dos Requisitos (Seção 9.1.4). O fluxo de aplicação é ilustrado na Figura 3.

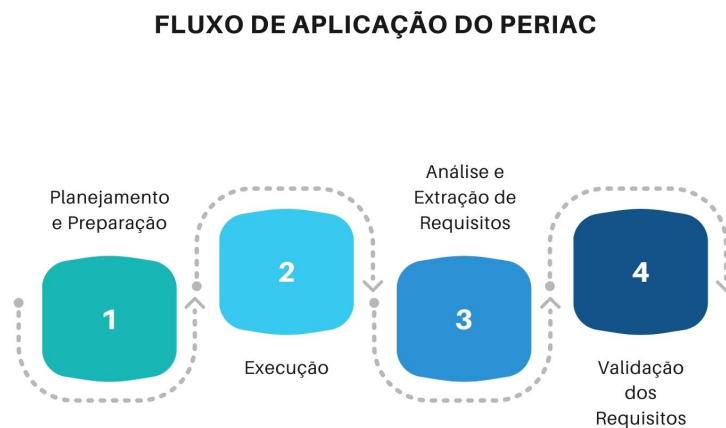


Figura 3. Fluxo de aplicação do PERIAC.

9.1.1. Planejamento e Preparação

Nesta fase, o objetivo é estabelecer as bases para a condução do processo. As atividades desta fase são:

1. **Levantamento de Informações do Projeto:** coleta de informações básicas sobre o projeto, incluindo nome, descrição contextual e objetivos.
2. **Definição de Objetivos do PERIAC:** estabelecimento claro dos objetivos da aplicação do Processo de Elicitación de Requisitos para IA Confiável (PERIAC) no contexto do projeto.

3. **Revisão da Literatura:** investigação da literatura existente para identificar projetos similares e *insights* relevantes que possam orientar o desenvolvimento do processo.
4. **Planejamento da Condução e Elaboração de Roteiro:** definição dos métodos de condução das sessões (presenciais ou online), elaboração de roteiros detalhados para guiar as discussões e atividades e seleção de perguntas (ver Tabela 6.1).
5. **Preparação do Material de Apoio:** preparação de todo o material necessário para facilitar as sessões, incluindo termos de consentimento, definições de conceitos e ferramentas de suporte. No caso de sessões conduzidas presencialmente, é necessário preparar um material físico composto por cartas e canetas para distribuir aos participantes. Esses materiais permitem que registrem suas ideias e contribuições de forma prática e tangível durante a sessão. Para sessões realizadas online, é essencial selecionar as ferramentas de suporte adequadas, como plataformas de videoconferência e ferramentas de colaboração online.
6. **Seleção e Recrutamento dos Participantes:** identificação e recrutamento dos participantes ideais para compor uma equipe multidisciplinar, considerando suas habilidades e experiências relevantes.

9.1.2. Execução

Esta fase é dividida em dois momentos distintos para promover discussões e reflexões sobre os dados e modelos a serem desenvolvidos, seguido pelo levantamento de requisitos. O primeiro momento visa promover uma discussão e reflexão entre os participantes, levantando entendimento sobre os dados que serão utilizados e o modelo a ser desenvolvido. O segundo momento tem o objetivo de conduzir uma sessão de *Brainwriting* com os participantes para o levantamento de requisitos. As atividades são as seguintes:

Momento 1: Entendimento do Projeto

1. **Contextualização do Projeto:** apresentação do projeto, explicando seu contexto, objetivos e os princípios de IA confiável.
2. **Roda de Conversa:** discussão com base no roteiro, utilizando as questões selecionadas para nivelar o conhecimento de todos os *stakeholders* em relação aos dados e ao modelo a ser desenvolvido.

Momento 2: Brainwriting

1. **Apresentação do Projeto:** reforço do contexto, objetivos e princípios de IA confiável, seguido pela explicação da dinâmica da sessão.
2. **Leitura e Reflexão sobre o Assunto:** fornecimento de cartas contendo princípios relevantes para reflexão individual dos participantes. Cada cartão é composto pela descrição do princípio, exemplo e conjunto de questões.
3. **Contribuição de Pontos de Vista:** registro individual das respostas dos participantes para expressar suas visões sem influências externas.
4. **Interação e Discussão:** compartilhamento das respostas e discussão em grupo para enriquecer a compreensão coletiva.

5. **Aprofundamento e Enriquecimento das Respostas:** tempo adicional para aprimorar as respostas com base nas discussões anteriores, reiniciando o processo com novos princípios.
6. **Conclusão e Síntese:** consolidação dos *insights* obtidos para identificar conclusões importantes.

9.1.3. Análise e Extração de Requisitos

Esta fase visa organizar as respostas dos participantes, classificá-las e extrair os requisitos para IA Confiável. As atividades desta fase são:

1. **Organização das Respostas:** organização das respostas em uma planilha por princípio e pergunta para facilitar a visualização e análise.
2. **Extração de Potenciais Requisitos:** análise cuidadosa das respostas para convertê-las em potenciais requisitos.

9.1.4. Validação dos Requisitos

Esta fase visa validar os potenciais requisitos com os participantes para garantir sua adequação às necessidades do projeto. As atividades são as seguintes:

1. **Análise dos Potenciais Requisitos:** apresentação dos potenciais requisitos aos participantes para avaliação e discussão de melhorias.
2. **Revisão dos Requisitos:** revisão dos requisitos selecionados, solicitando *feedback* em relação à relevância e possíveis melhorias.
3. **Refinamento dos Requisitos:** ajuste e refinamento dos requisitos com base nos *feedback* recebidos para garantir sua precisão e adequação.

9.2. Conjunto de Cartas

Para enriquecer e orientar o PERIAC, um conjunto de cartas foi desenvolvido. Cada carta define um princípio, apresenta um exemplo fictício e inclui perguntas que ajudam os participantes na reflexão e no levantamento de requisitos. Essas cartas facilitam a compreensão e a organização dos requisitos essenciais para sistemas de IA confiáveis.

A carta de Justiça, apresentada na Figura 4, define que a Justiça visa assegurar que o sistema opere de forma imparcial, sem viés ou discriminação, promovendo equidade e igualdade de oportunidades para todos os usuários. Como exemplo, no recrutamento com IA, vieses nos dados históricos podem resultar em exclusão injusta de grupos minoritários, exigindo revisão do algoritmo e políticas de recrutamento para promover equidade. As questões a serem respondidas são: O que é justiça em relação ao projeto que está sendo desenvolvido? Qual o objetivo da justiça no projeto que está sendo desenvolvido? Para quem é essa justiça no projeto que está sendo desenvolvido? Como ela deve ser envolvida no projeto que está sendo desenvolvido? Quando ela deve ser envolvida no projeto que está sendo desenvolvido? Onde ela deve ser envolvida no projeto que está sendo desenvolvido?

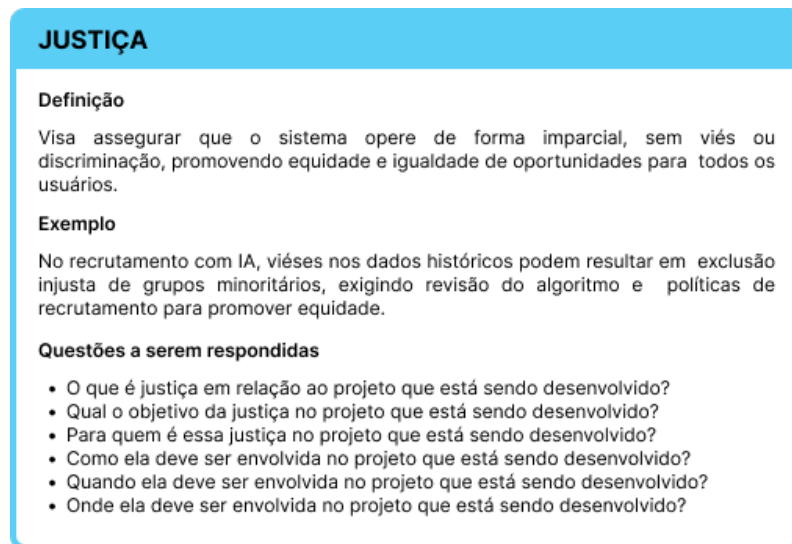


Figura 4. Carta do princípio de Justiça.

A carta de Explicabilidade, apresentada na Figura 5, define que os critérios e processos que fundamentam as decisões dos sistemas de IA devem ser compreensíveis para os diferentes envolvidos, incluindo desenvolvedores, usuários e *stakeholders*. Isso ajuda a garantir que as decisões sejam compreendidas e justificadas, mitigando preocupações sobre opacidade e falta de controle. Por exemplo, um sistema de IA é utilizado para conceder ou negar empréstimos a indivíduos, mas os critérios exatos usados pelo algoritmo para tomar essas decisões não são claramente compreendidos pelos usuários, levantando preocupações sobre transparência e justiça. As questões a serem respondidas são: O que é explicabilidade em relação ao projeto que está sendo desenvolvido? Qual o objetivo da explicabilidade no projeto que está sendo desenvolvido? Para quem é essa explicabilidade no projeto que está sendo desenvolvido? Como ela deve ser envolvida no projeto que está sendo desenvolvido? Quando ela deve ser envolvida no projeto que está sendo desenvolvido? Onde ela deve ser envolvida no projeto que está sendo desenvolvido?

A carta de Responsabilidade, apresentada na Figura 6, define a importância de estabelecer mecanismos para identificar e responsabilizar os agentes envolvidos nas decisões e ações dos sistemas de IA. Isso inclui tanto os desenvolvedores quanto os operadores desses sistemas. A prestação de contas é essencial para mitigar riscos e garantir que eventuais danos ou consequências adversas sejam tratados de forma adequada. Por exemplo, um sistema de IA autônomo, como um veículo autônomo, causa um acidente. Identificar quem é responsável pelo acidente - seja o fabricante do veículo, o desenvolvedor do algoritmo, ou o operador humano no momento do acidente - e determinar como responsabilizá-los por eventuais danos ou consequências adversas pode ser complexo e desafiador. As questões a serem respondidas são: O que é responsabilidade em relação ao projeto que está sendo desenvolvido? Qual o objetivo da responsabilidade no projeto que está sendo desenvolvido? Para quem é essa responsabilidade no projeto que está sendo desenvolvido? Como ela deve ser envolvida no projeto que está sendo desenvolvido? Quando ela deve ser envolvida no projeto que está sendo desenvolvido? Onde ela deve

EXPLICABILIDADE

Definição

Os critérios e processos que fundamentam as decisões dos sistemas de IA devem ser compreensíveis para os diferentes envolvidos, incluindo desenvolvedores, usuários e stakeholders. Isso ajuda a garantir que as decisões sejam compreendidas e justificadas, mitigando preocupações sobre opacidade e falta de controle.

Exemplo

Um sistema de IA é utilizado para conceder ou negar empréstimos a indivíduos, mas os critérios exatos usados pelo algoritmo para tomar essas decisões não são claramente compreendidos pelos usuários, levantando preocupações sobre transparência e justiça.

Questões a serem respondidas

- O que é explicabilidade em relação ao projeto que está sendo desenvolvido?
- Qual o objetivo da explicabilidade no projeto que está sendo desenvolvido?
- Para quem é essa explicabilidade no projeto que está sendo desenvolvido?
- Como ela deve ser envolvida no projeto que está sendo desenvolvido?
- Quando ela deve ser envolvida no projeto que está sendo desenvolvido?
- Onde ela deve ser envolvida no projeto que está sendo desenvolvido?

Figura 5. Carta do princípio de Explicabilidade.

ser envolvida no projeto que está sendo desenvolvido?

A carta de Privacidade, apresentada na Figura 7, destaca a importância da proteção da privacidade e confidencialidade dos dados dos usuários na era da IA. Os sistemas devem coletar, armazenar e utilizar os dados de forma ética e responsável, respeitando os direitos individuais e evitando qualquer forma de exploração ou violação. Isso requer o uso de práticas de segurança robustas e a conformidade com regulamentações de privacidade. Por exemplo, um assistente virtual que coleta e armazena dados pessoais dos usuários sem seu conhecimento ou consentimento está violando sua privacidade. Isso pode resultar em exploração indevida desses dados ou em sua divulgação a terceiros sem autorização, comprometendo a confidencialidade e a segurança das informações pessoais dos usuários. As questões a serem respondidas são: O que é privacidade em relação ao projeto que está sendo desenvolvido? Qual o objetivo da privacidade no projeto que está sendo desenvolvido? Para quem é essa privacidade no projeto que está sendo desenvolvido? Como ela deve ser envolvida no projeto que está sendo desenvolvido? Quando ela deve ser envolvida no projeto que está sendo desenvolvido? Onde ela deve ser envolvida no projeto que está sendo desenvolvido?.

A carta de Aceitação, apresentada na Figura 8, destaca a importância de considerar a aceitação dos usuários para a eficácia dos sistemas de IA. É essencial que esses sistemas sejam aceitos e adotados pelos usuários, o que requer a consideração de suas necessidades, expectativas e valores durante o projeto e implementação. Por exemplo, um sistema de recomendação em uma plataforma de streaming de vídeo que não leva em consideração as preferências individuais dos usuários pode resultar em recomendações imprecisas ou inadequadas. Isso pode levar à frustração dos usuários e à recusa em utilizar o serviço, prejudicando a experiência de uso e a aceitação da plataforma. As questões a serem respondidas são: O que é aceitação em relação ao projeto que está sendo de-

RESPONSABILIDADE

Definição

É importante estabelecer mecanismos para identificar e responsabilizar os agentes envolvidos nas decisões e ações dos sistemas de IA. Isso inclui tanto os desenvolvedores quanto os operadores desses sistemas. A prestação de contas é essencial para mitigar riscos e garantir que eventuais danos ou consequências adversas sejam tratados de forma adequada.

Exemplo

Um sistema de IA autônomo, como um veículo autônomo, causa um acidente. Identificar quem é responsável pelo acidente, se é o fabricante do veículo, o desenvolvedor do algoritmo, ou o operador humano no momento do acidente e determinar como responsabilizá-los por eventuais danos ou consequências adversas pode ser complexo e desafiador.

Questões a serem respondidas

- O que é responsabilidade em relação ao projeto que está sendo desenvolvido?
- Qual o objetivo da responsabilidade no projeto que está sendo desenvolvido?
- Para quem é essa responsabilidade no projeto que está sendo desenvolvido?
- Como ela deve ser envolvida no projeto que está sendo desenvolvido?
- Quando ela deve ser envolvida no projeto que está sendo desenvolvido?
- Onde ela deve ser envolvida no projeto que está sendo desenvolvido?

Figura 6. Carta do princípio de Responsabilidade.

PRIVACIDADE

Definição

A proteção da privacidade e confidencialidade dos dados dos usuários é uma preocupação central na era da IA. Os sistemas devem coletar, armazenar e utilizar os dados de forma ética e responsável, respeitando os direitos individuais e evitando qualquer forma de exploração ou violação. Isso envolve o uso de práticas de segurança robustas e a conformidade com regulamentações de privacidade.

Exemplo

Um assistente virtual coleta e armazena dados pessoais dos usuários sem seu conhecimento ou consentimento, violando sua privacidade. Isso pode resultar em exploração indevida desses dados ou em sua divulgação a terceiros sem autorização, comprometendo a confidencialidade e a segurança das informações pessoais dos usuários.

Questões a serem respondidas

- O que é privacidade em relação ao projeto que está sendo desenvolvido?
- Qual o objetivo da privacidade no projeto que está sendo desenvolvido?
- Para quem é essa privacidade no projeto que está sendo desenvolvido?
- Como ela deve ser envolvida no projeto que está sendo desenvolvido?
- Quando ela deve ser envolvida no projeto que está sendo desenvolvido?
- Onde ela deve ser envolvida no projeto que está sendo desenvolvido?

Figura 7. Carta do princípio de Privacidade.

ACEITAÇÃO

Definição

Para que os sistemas de IA sejam eficazes, é essencial que sejam aceitos e adotados pelos usuários. Isso requer considerar suas necessidades, expectativas e valores ao projetar e implementar esses sistemas. Uma experiência de uso positiva e inclusiva é fundamental para promover a aceitação e confiança na IA.

Exemplo

Um sistema de recomendação em uma plataforma de streaming de vídeo não leva em consideração as preferências individuais dos usuários, resultando em recomendações imprecisas ou inadequadas. Isso pode levar à frustração dos usuários e à recusa em utilizar o serviço, prejudicando a experiência de uso e a aceitação da plataforma.

Questões a serem respondidas

- O que é aceitação em relação ao projeto que está sendo desenvolvido?
- Qual o objetivo da aceitação no projeto que está sendo desenvolvido??
- Para quem é essa aceitação no projeto que está sendo desenvolvido?
- Como ela deve ser considerada no projeto que está sendo desenvolvido?
- Quando ela deve ser considerada no projeto que está sendo desenvolvido?
- Onde ela deve ser considerada no projeto que está sendo desenvolvido?

Figura 8. Carta do princípio de Aceitação.

sendo desenvolvido? Qual o objetivo da aceitação no projeto que está sendo desenvolvido? Para quem é essa aceitação no projeto que está sendo desenvolvido? Como ela deve ser considerada no projeto que está sendo desenvolvido? Quando ela deve ser considerada no projeto que está sendo desenvolvido? Onde ela deve ser considerada no projeto que está sendo desenvolvido?.

10. Limitações e ameaças à validade

Em relação às ameaças internas, a possível exaustão e desmotivação das participantes devido ao tempo necessário para a aplicação do estudo podem ter afetado 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. Apesar dessa divisão, ainda existe a possibilidade de que alguns participantes possam ter sido afetados pela carga de trabalho, o que poderia influenciar suas respostas.

Quanto às ameaças externas, a seleção cuidadosa das 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. Os resultados podem não ser extrapoláveis para diferentes contextos ou grupos de pessoas, o que pode afetar a aplicabilidade das descobertas em cenários diversos.

Além dessas questões, algumas outras ameaças à validade também devem ser consideradas, como a influência da visão do pesquisador. A seleção das participantes, a condução do estudo e a análise e interpretação dos dados podem ter sido afetadas pelo viés do pesquisador, introduzindo subjetividades que podem comprometer a objetividade dos resultados. A consciência desse potencial viés é crucial para uma avaliação crítica

dos achados.

Limitações relacionadas ao *Brainwriting* devem ser consideradas, pois, embora seja uma técnica valiosa para eliciar ideias, ela não pode capturar todos os aspectos relevantes. A técnica tem a vantagem de promover a geração de ideias de forma rápida e em um ambiente colaborativo, mas pode falhar em abordar questões mais profundas ou complexas que exigem maior reflexão e análise detalhada. Além disso, o *Brainwriting* pode não ser igualmente eficaz para todos os participantes, dependendo de suas habilidades individuais de expressão escrita e criatividade.

No âmbito deste estudo, é essencial reconhecer uma ameaça significativa à validade, relacionada à falta de validação externa do processo proposto. O processo em questão, delineado neste estudo, foi desenvolvido e validado exclusivamente pelos autores da pesquisa, sem ter sido submetido a uma validação externa. A falta de validação externa do processo de software proposto representa uma ameaça à validade do estudo em duas frentes. Em relação à validade de construção, embora os autores tenham delineado o processo com base em sua experiência, a falta de validação pela comunidade científica suscita incertezas sobre a validade das medições e conceitos empregados. Quanto à validade externa, a falta de validação restringe a generalização dos resultados para contextos diferentes. A aplicabilidade e eficácia do processo em diversas organizações, cenários industriais ou contextos culturais permanecem desconhecidas devido à ausência de validação externa, limitando a compreensão das possíveis variações de resultados em diferentes ambientes.

11. Conclusão e trabalhos futuros

Este trabalho apresentou o Processo para Elicitação de Requisitos para IA Confiável (PERIAC), desenvolvido para auxiliar na criação de sistemas de IA confiáveis e éticos. O PERIAC promove a participação e colaboração dos diversos *stakeholders* envolvidos no desenvolvimento de sistemas de IA e é composto por quatro fases principais: Planejamento e Preparação, Execução, Análise e Extração de Requisitos, e Validação dos Requisitos. O processo adota os princípios de Justiça, Explicabilidade, Responsabilidade, Privacidade e Aceitação, conforme propostos por Kaur et al. (2022), e inclui um conjunto de perguntas para auxiliar durante a fase de execução, retiradas do estudo de Carvalho et al. (2022). Além disso, o PERIAC apresenta um conjunto de cartas dedicadas a cada um dos princípios base, contendo a descrição do princípio, um exemplo e perguntas.

O PERIAC foi fundamentado em *insights* obtidos a partir de um estudo de caso exploratório que investigou o uso da técnica de *Brainwriting* na elicitação de requisitos para IA confiável. Este estudo focou em sistemas de AM aplicados ao recrutamento e seleção de candidatos, através da colaboração multidisciplinar dos envolvidos. O estudo utilizou os princípios para IA confiável de Kaur et al. (2022) como base e foi conduzido em três etapas, com a participação de seis mulheres com perfis multidisciplinares e um mediador. O contexto do estudo envolveu um caso real 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 identificar requisitos para uma IA confiável utilizando os princípios apresentados por Kaur et al. (2022) e a técnica de *Brainwriting*. O

estudo também facilitou o envolvimento multidisciplinar na discussão sobre IA confiável, promovendo a disseminação de conhecimento e reflexão sobre os aspectos e desafios relacionados ao tema. Os requisitos extraídos oferecem diretrizes cruciais para o desenvolvimento de sistemas de IA confiáveis e éticos no contexto do recrutamento e seleção de candidatos. Ao priorizar a proteção de dados, o treinamento sem viés, a diversidade nos dados e na equipe de desenvolvimento, bem como a transparência e a responsabilidade nas decisões do sistema, esses requisitos promovem um ambiente de recrutamento mais justo, transparente e responsável.

Os resultados obtidos a partir das percepções das participantes oferecem *insights* sobre os aspectos cruciais relacionados à confiança em IA. As diversas opiniões expressas destacam a importância de abordar a confiança em IA em várias frentes, incluindo a transparência de dados, a imparcialidade, os princípios éticos, o cumprimento das leis locais, e a aplicação dos princípios de confiança propostos por Kaur et al. (2022). Além disso, as participantes enfatizaram a necessidade de abordar a confiança em todas as etapas do desenvolvimento e manutenção dos sistemas de IA, envolvendo todos os *stakeholders relevantes*.

Para trabalhos futuros, pretende-se: (i) validar o processo proposto por meio de um novo estudo; e (ii) aplicar o processo proposto em um projeto real.

Referências

- [Ahmad et al. 2021] 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 2021] Bellaby, R. W. (2021). Can ai weapons make ethical decisions? *Criminal Justice Ethics*, 40(2):86–107.
- [Boy 1996] Boy, G. (1996). The group elicitation method: An introduction. In *Advances in Knowledge Acquisition: 9th European Knowledge Acquisition Workshop, EKAW'96 Nottingham, United Kingdom, May 14–17, 1996 Proceedings 9*, pages 290–305. Springer.
- [Calazans et al. 2018] Calazans, A. T. S., Paldês, R. Á., Masson, E. T. S., and Guimarães, F. d. A. (2018). As técnicas de elicitação e de documentação de requisitos nos métodos ágeis. In *CIbSE*, pages 487–500.
- [Carvalho et al. 2022] Carvalho, N. d. O., 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.
- [Cerqueira et al. 2022] 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.

- [Dastin 2018] 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.
- [Gruber et al. 2017] 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 2009] 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 2019] 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 and Matsuno 2020] 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 et al. 2022] Kaur, D., Uslu, S., Rittichier, K. J., and Durrresi, A. (2022). Trustworthy artificial intelligence: A review. *ACM Comput. Surv.*, 55(2).
- [Lazar et al. 2017] Lazar, J., Feng, J. H., and Hochheiser, H. (2017). *Research methods in human-computer interaction*. Morgan Kaufmann.
- [Li et al. 2023] Li, B., Qi, P., Liu, B., Di, S., Liu, J., Pei, J., Yi, J., and Zhou, B. (2023). Trustworthy ai: From principles to practices. *ACM Comput. Surv.*, 55(9).
- [Michinov 2012] 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 2015] Pazmino, A. V. (2015). *Como se cria: 40 métodos para design de produtos*. Editora Blucher.
- [Pei et al. 2022] 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.
- [Silva et al. 2023] Silva, F. L. Q. d., Sampaio, A. L., Bezerra, C. I. M., and Monteiro, I. T. (2023). Brainwriting na elicitação de requisitos para ia confiável. In *Anais do VIII Workshop sobre Aspectos Sociais, Humanos e Econômicos de Software*, pages 111–120. SBC.
- [Thiebes et al. 2021] Thiebes, S., Lins, S., and Sunyaev, A. (2021). Trustworthy artificial intelligence. *Electronic Markets*, 31:447–464.
- [VanGundy 1984] VanGundy, A. B. (1984). Brain writing for new product ideas: an alternative to brainstorming. *Journal of Consumer Marketing*.
- [Villamizar et al. 2021] Villamizar, H., Escovedo, T., and Kalinowski, M. (2021). Requirements engineering for machine learning: A systematic mapping study. In *2021 47th*

Euromicro Conference on Software Engineering and Advanced Applications (SEAA), pages 29–36.

- [Yoshioka et al. 2021] 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.