

# Qualidade de Software para Engenheiros de IA: Um Estudo Inicial da Realidade Brasileira

Lidiane C. Silva, Lucinara K. S. Fernandes, Eder F. Soares,  
Paulo Henrique M. Maia, Ismayle S. Santos, Francisco C. M. B. Oliveira  
Universidade Estadual do Ceará  
Fortaleza, Ceará, Brasil

{lidiane.castro,lucinara.fernandes,eder.furtado}@aluno.uece.br,{pauloh.maia,ismayle.santos,fran.oliveira}@uece.br

## Abstract

This article presents an analysis of the Brazilian reality regarding software quality for Artificial Intelligence (AI). The study seeks to investigate the software quality assurance strategies adopted during the lifecycle of AI/ML components, covering the development, integration and maintenance phases. We conducted a survey with 40 participants, and the results indicate that the Brazilian industry faces several quality issues in the development of AI systems, including accuracy, performance, interpretability and scalability. In addition, challenges related to maintainability, documentation, reuse and code quality are identified.

**CCS Concepts:** • Software and its engineering → Software testing and debugging.

**Keywords:** Qualidade de software, Inteligência Artificial, Machine Learning

## 1 Introdução

A Inteligência Artificial (IA) têm ganhado cada vez mais destaque e adoção em diversos setores, impulsionando transformações significativas nos softwares utilizados pelas empresas. Segundo o estudo de Song et al. [7], a IA e a *Machine Learning* (ML) têm se tornado uma parte indispensável das soluções de software empresarial, fornecendo recursos avançados para automação de tarefas, análise de dados complexos e tomada de decisões inteligentes.

De acordo com estudos recentes, a adoção da IA/ML é um dos principais motores da transformação digital nas organizações. Segundo Varshney e Mojsilović [8], a IA desempenha um papel fundamental na criação de sistemas capazes de serem compreendidos e confiáveis, o que é essencial para a transformação digital bem-sucedida. Além disso, Bughin et al. [2] destacam que a IA é considerada a próxima fronteira digital, proporcionando novas oportunidades e desafios para as empresas que desejam se manter competitivas.

Diante desse cenário de grande adesão e utilização de IA/ML em softwares, surge a necessidade de garantir a qualidade desses sistemas. A qualidade de software desempenha um papel crucial na satisfação do cliente, na confiabilidade do sistema e na minimização de riscos [3]. No entanto, o

desenvolvimento, a integração e a manutenção de componentes de IA/ML apresentam desafios específicos, que exigem estratégias adequadas de garantia de qualidade.

Para tanto, este artigo tem como desafio adaptar as questões da entrevista realizada para um *survey*, adequando para o contexto brasileiro, visando entender como a qualidade de componentes de IA e ML é trabalhada no Brasil e reproduzir o estudo realizado no artigo de Valentina Golendukhina et al. [5] onde este foi direcionado à realidade Austríaca. O objetivo deste artigo é então investigar as estratégias de garantia de qualidade de software adotadas no Brasil durante o ciclo de vida de componentes de IA/ML, abrangendo as fases de desenvolvimento, integração e manutenção.

No contexto deste estudo, são formuladas duas questões de pesquisa (QP1 e QP2) que norteiam as investigações. A QP1 busca responder à pergunta: "Quais problemas de qualidade foram enfrentados pela indústria em sistemas desenvolvidos com componentes de Inteligência Artificial?". Essa questão permitirá identificar os desafios específicos enfrentados pelas empresas no que diz respeito à qualidade dos sistemas de IA/ML. Já a QP2 aborda a seguinte questão: "Quais práticas de qualidade são aplicadas na indústria para manter a qualidade do código de Inteligência Artificial e Machine Learning?". Essa questão ajudará a compreender as estratégias e práticas adotadas pelas organizações para assegurar a qualidade do código em sistemas de IA/ML.

A fim de obter informações sobre a realidade da qualidade de software nas empresas brasileiras que desenvolvem soluções contendo componentes de IA e ML, foi realizado um estudo por meio de um *survey*. O questionário foi divulgado em redes sociais amplamente utilizadas, como *LinkedIn*, *Facebook* e *Instagram*, além de convites enviados por e-mail. Como resultado, foram obtidas 40 respostas de profissionais atuantes no mercado brasileiro, com experiência em projetos que envolvem IA e ML.

O restante deste artigo está organizado da seguinte maneira: na seção 2 e 3, será apresentado um panorama das estratégias de garantia de qualidade de software adotadas durante o ciclo de vida de componentes de IA/ML. Em seguida, na seção 4, serão discutidos os problemas de qualidade enfrentados pela indústria nesse contexto, com base nas respostas obtidas por meio do *survey*. Posteriormente, na seção 5, serão analisadas as práticas de qualidade aplicadas pela indústria para

manter a qualidade do código de IA/ML. Por fim, nas seções 6 e 7, serão identificadas as ameaças à validade desse trabalho, apresentando as conclusões do estudo e possibilidade de trabalhos futuros.

## 2 Trabalhos relacionados

Nesta seção, são explorados estudos e contribuições significativas que abrangem direta ou indiretamente os aspectos abordados neste artigo. Os trabalhos relacionados cobrem uma variedade de áreas, desde engenharia de software e sistemas baseados em IA até qualidade de software. Ao examinar esses estudos, pretende-se identificar lacunas existentes, destacar as contribuições originais deste trabalho e estabelecer um contexto sólido para a pesquisa apresentada.

Como base deste estudo, tem-se o artigo de Golendukhina et. al. [5] que aborda a questão da qualidade de software para engenheiros de IA dentro do contexto austríaco, buscando trazer clareza para esse tema complexo. Os autores destacam que, devido às características únicas da IA, como a natureza probabilística dos modelos e a dependência dos dados, as métricas tradicionais de qualidade de software podem não ser adequadas. O artigo explora abordagens alternativas para avaliar a qualidade de software em IA, incluindo aspectos como robustez, transparência, explicabilidade e ética. Além disso, discute a importância de adotar boas práticas de engenharia de software desde o início do desenvolvimento de sistemas de IA, como testes rigorosos, versionamento de modelos e documentação clara. O estudo empírico realizado envolveu entrevistas semiestruturadas com representantes de dez Pequenas e Médias Empresas (PMEs) austríacas que desenvolvem sistemas habilitados para IA. Foram coletadas informações demográficas e estratégias de *Software Quality Assurance* (SQA), abordando questões relacionadas ao desenvolvimento de IA/ML, desafios encontrados, questões de qualidade e práticas implementadas.

Felderer et. al. [4], por sua vez, destacam a importância de se estabelecer processos adequados de garantia de qualidade desde a fase de desenvolvimento até a implantação desses sistemas. No estudo, são discutidos os desafios específicos relacionados à IA, como a falta de transparência dos modelos, a dificuldade em lidar com a evolução contínua dos algoritmos e a necessidade de adaptar as práticas de teste tradicionais. O artigo deles também aborda técnicas e ferramentas para a garantia de qualidade em sistemas baseados em IA, incluindo abordagens de teste, validação, verificação e monitoramento.

Amershi [1] apresenta um estudo de caso que explora a aplicação de princípios de engenharia de software no contexto de *Machine Learning*. O autor destaca os desafios únicos enfrentados pelos engenheiros de software ao trabalhar com sistemas de aprendizado de máquina, incluindo a necessidade de lidar com dados complexos, modelos em constante

evolução e requisitos de desempenho em tempo real. O estudo de caso examina como uma equipe de engenharia aplica práticas de engenharia de software, como controle de versão, testes automatizados, revisões de código e depuração, para melhorar a qualidade e a manutenibilidade de um sistema de *Machine Learning*. O artigo enfatiza a importância de considerar os aspectos de engenharia de software desde as fases iniciais do desenvolvimento de sistemas de aprendizado de máquina, a fim de garantir a robustez e a escalabilidade desses sistemas.

Neste estudo, pretende-se aplicar de forma sistemática a questão da qualidade de software para engenheiros de IA, como proposto por Valentina Golendukhina et. al. [5], porém trazendo para o contexto brasileiro, adquirindo informações a partir de questionário com perguntas objetivas e discursivas direcionadas a profissionais brasileiros da área. Além disso, aspectos considerados relevantes nos estudos de Federer et. al. [4] e Amershi [1] contribuíram para elaboração do *survey*.

## 3 Metodologia

A presente pesquisa empregou uma metodologia baseada na realização de um *survey*. O planejamento do *survey* seguiu o processo proposto por Kasunic [6] para o design efetivo de *surveys* na área de engenharia de software.

O questionário foi elaborado com cautela, visando explorar aspectos relacionados à qualidade de software em sistemas de IA. Isso envolveu a definição de perguntas relevantes e a estruturação lógica do questionário para obter informações valiosas dos participantes. Utilizando a interface do *Google Forms*, o questionário foi construído com diferentes tipos de perguntas, como perguntas de múltipla escolha, perguntas abertas e escalas de classificação. Foram incluídas perguntas relacionadas a diversos aspectos da qualidade de software, como confiabilidade, desempenho, usabilidade, segurança e manutenibilidade.

O público-alvo do estudo foi definido como profissionais da área de IA, desenvolvedores, engenheiros de software ou qualquer pessoa com experiência e conhecimento relevantes sobre o tema. Para a avaliação do questionário foi realizada uma execução piloto com 2 pessoas convidadas. Os resultados do piloto permitiram a melhoria do planejamento do *survey* (e.g. melhor forma de análise de alguns dados) e também do questionário (e.g. com o refinamento de algumas questões).

Para a distribuição do questionário para a população foi utilizada a *Self selection*, que é uma técnica de amostragem não probabilística no qual os o *survey* foi distribuído online e os participantes que atendiam aos critérios poderiam escolher se participavam ou não da pesquisa. Neste sentido, o link para o questionário foi compartilhado por meio de diversas redes sociais, como *Facebook*, *Instagram* e *LinkedIn*, além de convites por e-mail e em grupos profissionais. Os critérios para a adequação do participante, divulgados junto com o questionário, foram: experiência profissional, formação

acadêmica ou envolvimento em projetos relacionados à IA. Além disso, foi utilizada a estratégia de *Convenience Sample*, na qual os autores entraram em contato com profissionais da área, buscando atingir um número maior de participantes.

Foi estipulado um período de 10 dias para que os participantes respondessem ao questionário. As respostas foram registradas automaticamente, garantindo a integridade e a confidencialidade dos dados.

Após a conclusão da coleta, os dados foram exportados do *Google Forms* para uma ferramenta de análise estatística, como o *Microsoft Excel*. Foram empregadas técnicas de análise descritiva e, quando apropriado, análise estatística, a fim de identificar padrões, tendências e conclusões relevantes. A figura 1 representa de forma gráfica o fluxo adotado na execução do estudo.

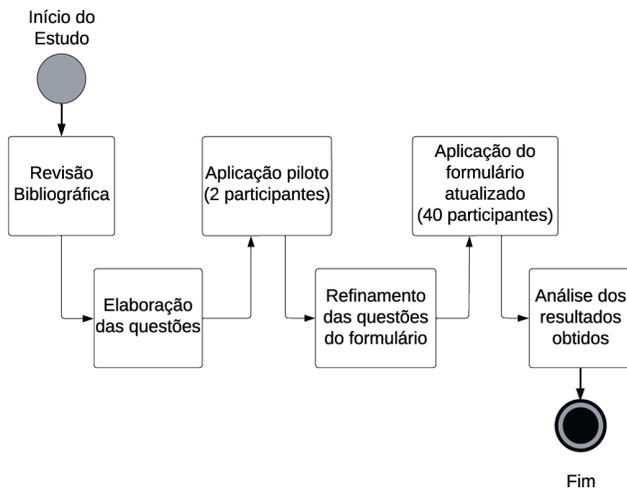


Figure 1. Metodologia do estudo

## 4 Resultados

Nesta seção, são apresentados os resultados da análise dos dados coletados por meio de um questionário distribuído para profissionais da área de IA. Os resultados fornecem informações sobre diferentes aspectos relacionados aos desafios enfrentados pelas empresas no desenvolvimento de sistemas baseados em IA, bem como as práticas adotadas para solucioná-los. O questionário foi distribuído utilizando a plataforma *Google Forms*, contando com a participação de 40 participantes, que contribuíram com informações relacionadas ao perfil de atuação profissional e aspectos relacionados à qualidade de código de IA nas aplicações que participam do desenvolvimento. O formulário aplicado, assim como os dados obtidos estão disponíveis em um repositório no *GitHub*<sup>1</sup>. A seguir, serão apresentados os principais pontos encontrados e serão discutidas suas implicações para a área de IA.

<sup>1</sup><https://bit.ly/4851v7n>

### 4.1 Análise de perfil

Na primeira parte do questionário, foram propostas perguntas para compreender o perfil dos profissionais participantes e das empresas ao qual atuam. Os participantes representaram diversas organizações, sendo a maioria, 77.5%, empresas de médio porte (até 499 funcionários) e grandes empresas (acima de 500 funcionários). Para capturar os perfis profissionais dos participantes, foi incluída uma pergunta de múltipla escolha, revelando que a maioria se identificou como "Cientista de Dados", 60%, que seriam profissionais responsáveis por coletar, organizar e analisar grandes volumes de dados para extrair percepções e criar modelos preditivos. O segundo perfil mais frequentemente citado foi o de "Engenheiro de *Machine Learning*", 42.5%, que são profissionais envolvidos na criação, treinamento e implementação de modelos de IA capazes de realizar tarefas específicas, como reconhecimento de padrões, processamento de linguagem natural ou visão computacional.

Ainda sobre o perfil, também foram investigados os tipos de software com IA/ML desenvolvidos pelas empresas, onde foi constatado que 42.5% dos participantes afirmaram desenvolver software para uso interno, enquanto 45% mencionaram que suas empresas desenvolvem software empresarial para comercialização. Outro aspecto relevante foi a identificação do papel da IA no software desenvolvido pelas empresas, no qual foi observado que 50% dos participantes consideraram a IA como um componente crítico do software, enquanto 40% a utilizam como uma funcionalidade secundária. Em relação ao processo de desenvolvimento de componentes de IA/ML adotado pelas empresas, verifica-se que a maioria dos participantes utiliza metodologias ágeis, 75%, sendo a frequência de dias por sprint a métrica mais citada dentre as mais utilizadas por estas para acompanhar o processo de desenvolvimento desses componentes, com 70% das respostas.

### 4.2 Problemas de qualidade em sistemas desenvolvidos com componentes de IA/ML

Na segunda parte do questionário, o foco das perguntas está direcionado à caracterização dos principais problemas enfrentados pelas empresas que desenvolvem sistemas habilitados para IA (QP1).

De acordo com as respostas dos participantes, é possível observar que a indústria tem enfrentado diversos problemas de qualidade no desenvolvimento de sistemas com componentes de Inteligência Artificial (IA). Em relação à variedade de problemas enfrentados no desenvolvimento de códigos de IA/ML, os principais incluem:

- Problemas de precisão: Identificados por 75% dos participantes, estes se referem à complicações relacionadas à capacidade do software de IA em fornecer resultados corretos e confiáveis;

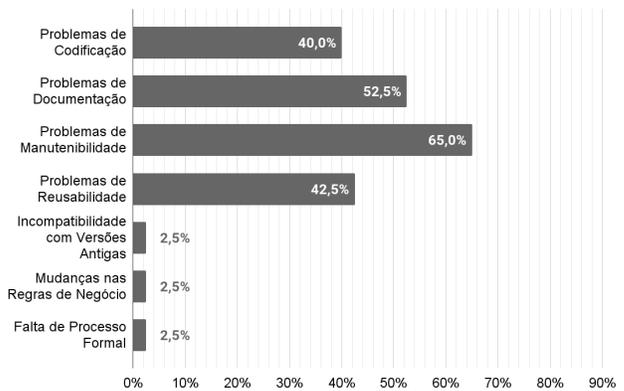
- Desempenho abaixo do esperado: Apontado por 70% dos respondentes, este pode resultar em respostas lentas ou imprecisas do sistema;
- Problemas com interpretabilidade: Apontados por 57.5% dos participantes, estes indicam dificuldades na compreensão, pelos desenvolvedores, de como ou porque foram tomadas certas decisões;
- Problemas de escalabilidade: Com aproximadamente 52.5% dos participantes indicando que os sistemas de IA podem não ser considerados escaláveis o suficiente, é possível que haja dificuldade em lidar com grandes quantidades de dados ou em funcionar de maneira eficiente em ambientes de alta demanda.

Além disso, existem problemas de qualidade mais amplos associados ao desenvolvimento de sistemas de IA/ML, por isso no questionário foi interrogado sobre os relacionados aos códigos de IA/ML. Todos esses problemas relatados pelos participantes da pesquisa se encontram no gráfico da Figura 2, porém os que se destacaram foram:

- Manutenibilidade: Cerca de 65% dos participantes relataram problemas de manutenibilidades, o que resulta em um contexto onde sistemas de IA enfrentam desafios em termos de facilidade de manutenção e atualização. Isso pode dificultar a correção de erros ou a introdução de melhorias.
- Documentação: A documentação adequada é essencial para garantir a compreensão e o uso correto dos sistemas de IA. No entanto, de acordo com os dados obtidos, 52.5% dos participantes afirmam que os sistemas ao qual atuam possuem problemas de documentação;
- Reuso: Na pesquisa, cerca de 42.5% dos profissionais afirmam enfrentar problemas relacionados a reuso;
- Qualidade de Código: A qualidade do código é um fator importante para garantir a robustez e a confiabilidade de um sistema de IA. Porém, 40% dos participantes declararam enfrentar problemas na qualidade de código.

Além disso, os participantes foram questionados no *survey* sobre as práticas de qualidade aplicadas para identificar problemas no código de IA/ML durante o processo de desenvolvimento. De acordo com as respostas obtidas, constatou-se que 82.5% das empresas realizam revisões de código, 52.5% realizam testes de unidade, 30% realizam testes de integração, 25% fazem testes de sistema, e 82.5% utilizam análises estatísticas (como acurácia, matriz de confusão, entre outros) como parte do processo de avaliação da qualidade.

Outro questionamento abordado durante a pesquisa dizia respeito à forma como os testes de componentes de IA/ML são conduzidos nos sistemas das empresas participantes. A análise dos dados revelou que 47.5% das empresas realizam testes manuais, 27.5% optam por testes automatizados e 52.5% utilizam uma abordagem de testes híbridos. Essas proporções refletem as práticas adotadas pelas empresas participantes



**Figure 2.** Respostas à pergunta: Quais problemas de qualidade você costuma enfrentar em códigos de IA/ML?

e fornecem uma visão das estratégias de teste empregadas em seus sistemas de IA/ML. Cabe ressaltar que esses dados específicos são baseados nas respostas obtidas na pesquisa e podem variar entre as diferentes empresas e contextos dentro do cenário nacional.

Na fase final da análise para responder à questão de pesquisa QP1, foi realizado um questionamento aos participantes sobre qual membro da equipe realiza os testes dos componentes de IA/ML nos sistemas de suas respectivas empresas. Os resultados revelaram que 70% dos participantes indicaram que os próprios desenvolvedores são responsáveis pela realização dos testes, enquanto apenas 40% mencionaram que os analistas de testes desempenham essa função. Essas informações retratam as práticas adotadas pelas empresas participantes e fornecem uma visão sobre a distribuição de responsabilidades nos processos de teste de IA/ML.

Esta etapa da análise revelou que a indústria enfrenta diversos problemas de qualidade no desenvolvimento de sistemas de IA, incluindo precisão, desempenho, interpretabilidade e escalabilidade. Também foram identificados desafios relacionados à manutenibilidade, documentação, reuso e qualidade do código. As práticas de qualidade mais comuns para identificação dos problemas de qualidade incluem revisões de código e análises estatísticas, enquanto os testes manuais são mais prevalentes do que os automatizados. A maioria das empresas atribui aos desenvolvedores a responsabilidade pelos testes de componentes de IA/ML. Essas conclusões destacam a necessidade de aprimoramentos e estratégias abrangentes para garantir a qualidade dos sistemas de IA no contexto empresarial brasileiro.

### 4.3 Práticas de qualidade aplicadas para resolver problemas em sistemas desenvolvidos com componentes de IA/ML

Na QP2, tem-se o intuito de entender quais práticas costumam ser aplicadas para resolver os problemas enfrentados

na QP1. Por exemplo, as empresas podem aplicar práticas conhecidas de controle de qualidade ou diferentes tipos de soluções alternativas.

No survey havia uma pergunta inicial para compreender a periodicidade com que a varredura de vulnerabilidades dos sistemas é realizada. Com base nas respostas dos participantes, foi constatado que 57.7% não possuem uma periodicidade definida para essa atividade, enquanto 12.5% realizam a varredura semanalmente e 12,5% a realizam diariamente. Esses dados fornecem uma visão sobre as práticas adotadas pelas empresas participantes em relação à segurança e identificação de vulnerabilidades em seus sistemas.

Foi elaborada também uma pergunta adicional para compreender como as buscas por vulnerabilidades nos sistemas são conduzidas, incluindo a detecção de vazamentos de dados e outras potenciais ameaças à segurança. Com base nas respostas dos participantes, foi observado que 47.5% das empresas não possuem um processo estabelecido para busca de vulnerabilidades e detecção de vazamento de dados. Por outro lado, 35% utilizam ferramentas automatizadas para realizar varreduras de segurança periódicas em busca de vulnerabilidades e vazamentos de dados. Além disso, 25% monitoram constantemente os logs do sistema e utilizam sistemas de detecção de intrusões para identificar possíveis ameaças e vazamentos de dados, enquanto 20% realizam auditorias regulares do sistema com o objetivo de identificar possíveis vulnerabilidades e vazamentos de dados. Tais informações fornecem uma visão sobre as abordagens adotadas pelas empresas participantes em relação à segurança cibernética e demonstram a diversidade de práticas para garantir a proteção dos sistemas contra vulnerabilidades e vazamentos de dados, apesar do número significativo de empresas que não possuem processo estabelecido, como já citado.

Um dos pontos também abordados foi o da adoção de padrões de SQA, *Software Quality Assurance*, pelas empresas ao qual os profissionais participantes atuam. Ao serem questionados sobre a utilização de tais padrões e identificar quais seriam eles, constatou-se que 50% dos participantes afirmaram que as empresas em que estão vinculados não adotam nenhum padrão de SQA. Por outro lado, 28.9% responderam que as empresas adotam padrões internos, como *templates* ou planilhas próprias padronizadas, para garantir a qualidade do software. Além disso, 18.4% mencionaram a adoção de padrões externos, como ISO (*International Organization for Standardization*) ou CMMI (*Capability Maturity Model Integration*). Esses resultados destacam a diversidade de abordagens adotadas pelas empresas em relação à implementação de padrões de SQA, com uma parcela significativa não adotando padrões específicos, enquanto outras recorrem a padrões internos ou externos reconhecidos internacionalmente.

Por fim, ainda relacionado as práticas de qualidade utilizadas pelas empresas ao qual os profissionais participantes atuam, foi abordada a questão sobre como estes selecionam

os componentes de IA/ML para integrar em seus sistemas. Mais especificamente, questionou-se se essa seleção é baseada em uma avaliação da qualidade de código. Os resultados revelaram que 60% dos participantes realizam uma avaliação da qualidade de código antes de proceder com a integração do componente. Por outro lado, 27.5% dos participantes consideram outras características além da qualidade de código na seleção dos componentes. Além disso, 10% dos participantes afirmaram que não realizam avaliação da qualidade de código antes da integração. Esses dados destacam a relevância dada à qualidade de código na seleção de componentes de IA/ML, somada a significativa consideração em aplicar também outras características durante esse processo de seleção, embora uma parcela desconsidere tais práticas.

## 5 Discussões

Foram elaboradas duas questões abertas para obter a resposta aberta dos participantes do *survey*, analisamos as questões objetivas sobre propostas de melhorias no fluxo de qualidade de código de IA na sua empresa e Quais práticas de qualidade de código de IA você conhece e que sua empresa ainda não usa. Cada questão englobava uma série de sugestões dadas pelos participantes.

### Questão 01: O que você melhoraria no fluxo de qualidade de código de IA da sua empresa?

Dentre as respostas fornecidas, algumas melhorias sugeridas são:

- Automatização de Testes e Dashboards;
- Treinamento e Conhecimento Geral;
- Revisão de Código;
- Testes Realizados por Testadores Independentes;
- Integração da Pipeline de Testes;
- Adoção de Práticas de MLOps;
- Documentação;
- MLOps/DataOps e IAOps;
- Processo de Ingestão de Dados;
- Testes Automatizados.

É importante ressaltar que as necessidades e desafios podem variar entre as empresas, e a melhoria do fluxo de qualidade de código de IA depende do contexto e dos requisitos específicos de cada organização.

### Questão 02: Quais práticas de qualidade de código de IA você conhece e que sua empresa ainda não usa?

Dentre as respostas fornecidas, algumas práticas de qualidade de código de IA que são mencionadas, mas ainda não são usadas pelas empresas incluem:

- Análise Estática de Código;
- Testes Automatizados;
- Pipeline Bem Definido;
- Verificação de Qualidade dos Dados de Treinamento;
- Testes de Penetração;
- Revisão de Código;
- Métricas de Código;

- Testes Unitários e Testes de Sistema;
- Design Patterns AI;
- TDD (Test-Driven Development) e OOP (Programação Orientada a Objetos);
- MLOps e DataOps.

Embora as empresas mencionadas ainda não estejam usando essas práticas, é importante reconhecer que as necessidades e os recursos de cada empresa podem variar, e a implementação de práticas de qualidade de código específicas para IA pode depender do contexto e dos requisitos do projeto.

## 6 Ameaças à validade

Na presente seção, abordaremos as potenciais ameaças à validade do estudo, que incluem as dimensões de validade interna, externa, de construção e de conclusão.

**Validade de Construção.** Para mitigar potenciais ameaças na criação do questionário, foi conduzida uma execução piloto com a participação de 2 profissionais especializados em IA/ML. Com base nos dados e *feedbacks* obtidos, o questionário passou por uma revisão, abrangendo tanto o formato quanto a formulação das perguntas. Essa etapa teve como objetivo melhorar a clareza e a adequação das questões, visando a validade interna do questionário.

**Validade interna.** A utilização de um questionário por meio de *survey* apresenta riscos, como a impossibilidade de esclarecer as perguntas aos participantes, bem como a não modificação ou esclarecimento posterior das respostas. Para mitigar esse problema, foi realizado um piloto com 2 participantes, no qual as perguntas foram previamente validadas, levando em consideração as dificuldades e ajustes identificados, a fim de evitar dúvidas por parte dos participantes ao responderem.

**Validade externa.** O questionário foi distribuído para um público diversificado por meio de múltiplos canais de contato, resultando em 40 respostas. Embora reconheçamos que nossos resultados podem não ser generalizáveis, eles oferecem uma visão das práticas adotadas pelas equipes de IA/ML das pessoas entrevistadas. Apesar das limitações, conduzimos o procedimento de pesquisa de maneira adequada e disponibilizamos os dados para permitir futuras replicabilidades deste estudo.[5]

**Validade de conclusão.** A análise e interpretação dos resultados foram realizadas considerando as informações coletadas por meio do questionário. No entanto, os resultados obtidos nesta pesquisa estão devidamente documentados e podem ser replicados em trabalhos futuros.

## 7 Conclusão

O estudo apresentado neste artigo teve como objetivo fornecer os primeiros insights a respeito do cenário brasileiro em relação à qualidade de software para engenheiros de IA. Para isso, foram realizadas duas perguntas principais: quais são

os problemas enfrentados pelas empresas brasileiras no desenvolvimento de sistemas de IA/ML e quais são as práticas de qualidade mais comuns adotadas por essas empresas?

A análise dos dados coletados revelou que a indústria brasileira enfrenta diversos problemas de qualidade no desenvolvimento de sistemas de IA, incluindo precisão, desempenho, interpretabilidade e escalabilidade. Além disso, foram identificados desafios relacionados à manutenibilidade, documentação, reuso e qualidade do código.

Em relação às práticas de qualidade, as mais comuns incluem revisões de código e análises estatísticas, enquanto os testes manuais são mais prevalentes do que os automatizados. A maioria das empresas atribui aos desenvolvedores a responsabilidade pelos testes de componentes de IA/ML.

Essas conclusões destacam a necessidade de aprimoramentos e estratégias abrangentes para garantir a qualidade dos sistemas de IA no contexto empresarial brasileiro. O estudo fornece uma visão das práticas adotadas pelas equipes de IA/ML que atuam no Brasil e disponibiliza os dados para permitir futuras replicações deste estudo.

Em relação às perspectivas de desenvolvimentos futuros, é notável identificar um parâmetro notório, que consiste na comparação dos resultados obtidos neste estudo com aqueles encontrados em pesquisas conduzidas na Áustria. Além disso, uma consideração adicional pode ser a expansão da amostra de respondentes, a fim de aprimorar e validar ainda mais as conclusões alcançadas nesta pesquisa.

## References

- [1] Saleema Amershi, Andrew Begel, Christian Bird, Robert DeLine, Harald Gall, Ece Kamar, Nachiappan Nagappan, Besmira Nushi, and Thomas Zimmermann. 2019. Software engineering for machine learning: A case study. In *2019 IEEE/ACM 41st International Conference on Software Engineering: Software Engineering in Practice (ICSE-SEIP)*. IEEE, 291–300.
- [2] Jacques Bughin, Eric Hazan, Sree Ramaswamy, Michael Chui, Tera Allas, Peter Dahlstrom, Nicolaus Henke, and Monica Trench. 2017. Artificial intelligence: The next digital frontier? (2017).
- [3] Paulo Jorge Borges da Costa. 2022. *Garantia de qualidade de software (SQA)*. Ph. D. Dissertation.
- [4] Michael Felderer and Rudolf Ramler. 2021. Quality assurance for AI-based systems: overview and challenges. *arXiv preprint arXiv:2102.05351* (2021).
- [5] Valentina Golendukhina, Valentina Lenarduzzi, and Michael Felderer. 2022. What is software quality for AI engineers? towards a thinning of the fog. In *Proceedings of the 1st International Conference on AI Engineering: Software Engineering for AI*. 1–9.
- [6] Mark Kasunic. 2005. Designing an effective survey.
- [7] Wang L. Zhang Z. Zhang B. Song, X. 2020. Research on the Application of Machine Learning and Artificial Intelligence in Modern Enterprise Software. In *Proceedings of the 6th International Conference on Education, Language, Art and Inter-cultural Communication (ELAIC 2020)*. Atlantis Press, 18–23.
- [8] Mojsilović A. Varshney, K. R. 2018. Towards trustworthy AI systems: Designing AI systems that people can understand and trust. In *IEEE Computer Society*. IEEE, 26–35.