Universidade Federal de Pernambuco

Centro de Informática

Graduação em Engenharia da Computação

Identificação e Caracterização de Requisitos de Confiabilidade em Sistemas de Machine Learning

Ana Letícia Albuquerque Santos

Trabalho de Graduação

Recife, Pernambuco

2025

Universidade Federal de Pernambuco
Centro de Informática

Ana Letícia Albuquerque Santos

Identificação e Caracterização de Requisitos de Confiabilidade em Sistemas de Machine Learning

Trabalho apresentado ao Curso de Graduação em Engenharia da Computação do Centro de Informática da Universidade Federal de Pernambuco como requisito parcial para obtenção do grau de Bacharel em Engenharia da Computação.

Área de concentração: Engenharia de Software e Linguagens de Programação
Orientador: Carla Taciana Lima Lourenco Silva

Recife, Pernambuco
2025

Ficha de identificação da obra elaborada pelo autor,
através do programa de geração automática do SIB/UFPE

Santos, Ana Letícia Albuquerque.
Identificação e caracterização de requisitos de confiabilidade em sistemas de Machine Learning / Ana Letícia Albuquerque Santos. - Recife, 2025.
82 p. : il., tab.

Orientador(a): Carla Taciana Lima Lourenço Silva Schuenemann
Trabalho de Conclusão de Curso (Graduação) - Universidade Federal de Pernambuco, Centro de Informática, Engenharia da Computação - Bacharelado, 2025.
Inclui referências, apêndices.

1. Engenharia de Requisitos. 2. Sistemas Confiáveis. 3. Sistemas baseados em Machine-Learning. I. Schuenemann, Carla Taciana Lima Lourenço Silva. (Orientação). II. Título.

000 CDD (22.ed.)

Ana Letícia Albuquerque Santos

Identificação e Caracterização de Requisitos de Confiabilidade em Sistemas de Machine Learning

Trabalho de Conclusão de Curso apresentado como pré-requisito para conclusão do Curso de Bacharelado em Engenharia da Computação do Centro de Informática da Universidade Federal de Pernambuco. Defendida e aprovada em 5 de agosto de 2025 pela seguinte banca examinadora:

Carla Taciana Lima Lourenço Silva Schuenemann
Orientador/CIn-UFPE

Cleber Zanchettin
Examinador/CIn-UFPE

Dedico este trabalho à Deus e minha família.

Agradecimentos

À minha mãe, Renata, que me mostrou desde pequeninha que eu tinha que batalhar muito para conquistar o que eu queria, enquanto lutava suas próprias batalhas para garantir que eu tivesse tudo o que eu precisava. Mãe, você é a razão de tudo o que eu sou e de tudo o que eu ainda serei. Esse trabalho e todas as minhas conquistas são frutos de todo o seu esforço, de todas as noites mal dormidas, de todos os dias corridos. Obrigada por todo o incentivo, por me dar colo sempre que eu precisei e por nunca ter desistido de mim mesmo quando eu não era fácil. Obrigada, mainha. Você é minha estrela guia para todos os oceanos que eu irei navegar nessa vida.

Ao meu namorado, Zilde. Que sorte a minha ter te encontrado tão cedo nessa vida. Você me inspira todos os dias a ser melhor, a tentar mais, a querer mais. Obrigada por ser minha fonte de força, confiança, esperança. Acho que eu devo agradecer também ao CIn por ter cruzado nossos caminhos no passado, mas agradeço muito mais a nós dois por lutarmos por um futuro. Juntos.

Aos meus irmãos, Ana Laura e Arthur. É um enorme privilégio ter crescido com vocês e saber que estamos ligados pra vida. Gosto de pensar que eu influenciei de alguma forma os seus caminhos, mas, apesar de tudo o que eu falo, não se deixem limitar às minhas escolhas. Vocês são muito melhores do que eu jamais poderia ser.

Aos meus avós, Antônio e Maria Elza. Palavras não são suficientes para dizer o quanto eu sou grata pela presença de vocês. Vó, cada gesto seu transborda o cuidado que a senhora sempre teve por mim. Obrigada por todos os conselhos e por toda comida feita e guardada especialmente para mim. Vô, o senhor sempre foi o meu maior incentivador. Lembro de, quando passei na prova do Detran, ter ficado mais feliz por saber que eu ia lhe deixar orgulhoso. Eu espero um dia fazer jus a tudo que o senhor já fez por mim.

Aos meus tios e tias, Rosane, Rejane, Rômulo, Hélder e Cássia, que sempre foram uma enorme rede de apoio na minha vida. Eu jamais chegaria onde eu estou hoje sem o privilégio que eu tive de poder contar com a ajuda de vocês, seja por uma roupa lavada, uma carona ou um almoço pronto. E também à minha madrinha, Flaviana, que me acolheu como filha e sempre esteve lá quando eu precisei.

Às minhas magnólias, Lethycia e Camila, que viveram comigo os meus melhores anos e, mesmo distantes fisicamente, sempre se fizeram presentes. Lethycia, eu acho que nem teria feito esse curso se não fosse por você. Camila, você não tem ideia do que fez quando acolheu aquela menininha introvertida nas assustadoras primeiras semanas do colégio. Obrigada por compartilharmos nossas histórias e por sermos grandes incentivadoras umas da outras.

Às minhas amigas, Ana Carolina, Marcela e Maria Fernanda, que me fazem explicar com o que eu trabalho todas as vezes que a gente se encontra. Obrigada pelo apoio incondicional nos últimos 13 anos. A presença de vocês sempre deu muito mais cor pra minha vida.

Aos meus amigos, Caio, Lopes, Marcos, Plácido, Renato e Yano. Obrigada por todos os cinemas e todas as sessões de jogos. Cada momento e cada risada vendo Yano jogar Resistência são um combustível indispensável para os meus dias.

Aos meus companheiros de curso, Alice, Dayane, Morone, Lucca, Elias, Victória e Williams. Que loucura pensar que entramos juntos nessa experiência que foi o Centro de Informática recém saídos do ensino médio e hoje podemos ver nossas carreiras se formando. Esse TCC só saiu por causa de todos os projetos, todas as revisões pré prova e todo o apoio que eu recebi de vocês. Um adendo especial à Williams, que sempre foi um ombro amigo e esteve comigo muito mais além dos limites da universidade. E também às minhas Winx, porque a gente sabe que ser mulher nessa área não é nada fácil, mas vocês fizeram tudo ser muito mais leve.

Aos meus eternos monitores, Mendes e Ambrósio, que olharam para aquela menina "baixinha de óculos" perdida no primeiro período da faculdade e viram potencial nela. A ajuda de vocês foi essencial para formar a profissional que eu sou hoje e eu serei para sempre grata.

E também à minha orientadora, professora Carla Silva, que confiou em mim para dar vida a este trabalho e foi indispensável para que ele pudesse acontecer. Muito obrigada!

"Everything you lose is a step you take"

— Taylor Swift

Resumo

Os avanços da tecnologia e as pesquisas no campo da Inteligência Artificial tornaram os sistemas que utilizam modelos de Aprendizagem de Máquina (ML) cada vez mais presentes no cotidiano das pessoas, inclusive em domínios críticos como saúde e mobilidade. O fato de que esses sistemas podem trazer riscos a indivíduos e para a sociedade torna-os fortes candidatos a se tornarem sistemas que precisarão demonstrar que são confiáveis por meio de artefatos chamados de casos de garantia. Este trabalho busca elaborar um conjunto estruturado de requisitos para serem utilizados na construção desses artefatos. Para isso, foi realizada uma revisão da literatura e os requisitos foram avaliados por profissionais da área por meio de um questionário. Resultou em um conjunto de requisitos organizados nas categorias de escopo, segurança, dados, desempenho e validação. Posteriormente, a análise das respostas dos especialistas permitiu revisar e refinar a proposta inicial com base em divergências observadas e sugestões fornecidas pelos participantes. Os achados oferecem uma base sólida que pode orientar as equipes que desenvolvem ML e querem garantir mais qualidade, segurança e confiança em seus sistemas.

Palavras-chave: Engenharia de Requisitos, Sistemas Confiáveis, Sistemas baseados em Machine-Learning

Abstract

Advances in technology and research in the field of Artificial Intelligence have led to the widespread presence of systems based on Machine Learning (ML) models in people's daily lives, including in critical domains such as healthcare and mobility. The fact that these systems can pose risks to individuals and society makes them strong candidates to become systems that will need to demonstrate trustworthiness through artifacts called assurance cases. This work aims to develop a structured set of requirements to be used in the construction of these artifacts. To achieve this, a literature review was conducted and the requirements were evaluated by professionals from the field through a survey. Resulted in a set of requirements grouped into the categories of scope, security, data, performance and validation. Subsequently, the analysis of the experts' feedback enabled the refinement and revision of the initial proposal based on divergences and suggestions provided by the participants. These findings offer a solid foundation that can guide teams developing ML to improve the quality, safety, and reliability in their systems.

Keywords: Requirements Engineering, Reliable Systems, Machine Learning-Based Systems.

Lista de Tabelas

Lista de Figuras

Sumário

Capítulo 1

Introdução

Este capítulo apresenta o contexto inicial deste trabalho, abordando as motivações que justificam a pesquisa, os objetivos que orientam seu desenvolvimento, as principais contribuições esperadas e a organização do documento. Com isso, busca-se fornecer uma visão geral que permita compreender a importância do tema e o caminho seguido para alcançar os resultados propostos.

1.1 Motivação

Nos últimos anos, tem-se observado um crescimento acelerado do uso de sistemas baseados em *Machine Learning* (ML) em diversas aplicações, muitas delas de caráter crítico, como saúde, transporte e finanças, o que eleva as preocupações quanto à sua confiabilidade. Diferentemente de sistemas tradicionais, modelos de ML apresentam comportamentos que dependem de dados de treinamento, os quais podem não abranger todos os cenários do mundo real [1, 2].

Estudos recentes mostram que ainda é desafiador estabelecer práticas consolidadas para especificar e validar a confiabilidade em sistemas de ML, especialmente em contextos críticos [3, 4]. Com isso, torna-se difícil assegurar que os modelos atendam a níveis aceitáveis de desempenho e segurança.

No campo de sistemas automotivos, padrões como a ISO 26262 e a ISO 21448 (SOTIF) têm desempenhado um papel importante no tratamento dos riscos associados ao uso de ML, mas ainda existem lacunas na definição de processos e artefatos específicos para componentes baseados em aprendizado de máquina [5].

Nesse cenário, mapear e caracterizar requisitos de confiabilidade para sistemas de ML é

essencial para mitigar riscos e possibilitar o uso seguro desses modelos em aplicações sensíveis, contribuindo para processos de desenvolvimento mais seguros e confiáveis [6]. Esses requisitos oferecerão uma base estruturada que possa servir de apoio à elaboração de casos de garantia, que são artefatos utilizados para demonstrar, com base em evidências, a confiabilidade de sistemas [7].

1.2 Objetivos

Este trabalho tem como objetivo central propor um conjunto de requisitos que favoreçam a construção de sistemas baseados em ML mais confiáveis, principalmente em contextos críticos.

Entre os objetivos específicos, incluem-se:

- Investigar requisitos voltados à confiabilidade de sistemas de ML por meio de uma revisão da literatura direcionada para publicações de autores reconhecidos na área, complementada pela análise de artigos que os referenciem ou sejam por eles referenciados;
- Organizar os requisitos encontrados em grupos que auxiliem na sua aplicação em diferentes cenários e facilitem o entendimento por profissionais e pesquisadores;
- Submeter os requisitos levantados para avaliação por especialistas da área, buscando verificar sua aplicabilidade prática e adequação às demandas de sistemas que utilizam ML em domínios críticos.

1.3 Contribuições

As principais contribuições deste trabalho incluem avanços que podem apoiar tanto a prática profissional quanto a pesquisa acadêmica na área de confiabilidade em sistemas de ML:

- Proposta de um conjunto de requisitos específicos para promover a construção de sistemas de ML mais confiáveis;
- Disponibilização de resultados analisados e verificados por especialistas do setor, aumentando a relevância prática;
- Fornecimento de uma base para futuras pesquisas relacionadas a requisitos e confiabilidade em sistemas de ML, com potencial aplicação na construção de casos de garantia.

1.4 Estrutura do Documento

Este trabalho está dividido nas seguintes seções:

- O Capítulo 2 apresenta o **referencial teórico**, que aborda os principais conceitos relacionados à Engenharia de Requisitos, Casos de Garantia, além de normas, certificações e publicações relevantes para a construção deste estudo.
- O Capítulo 3 descreve a **metodologia** adotada, detalhando as etapas de extração dos requisitos e o processo de validação.
- O Capítulo 4 apresenta os **resultados** obtidos nas etapas descritas na metodologia, bem como as análises realizadas até o conjunto final de requisitos propostos.
- O Capítulo 5 traz as **conclusões** do trabalho, discutindo as principais contribuições, limitações e sugestões de trabalhos futuros.

Capítulo 2

Referencial Teórico

Este capítulo tem como objetivo apresentar os fundamentos conceituais que sustentam a pesquisa, abordando os principais temas relacionados à confiabilidade em sistemas baseados em aprendizado de máquina.

Para isso, são discutidos aspectos da Engenharia de Requisitos, certificações e padrões aplicáveis à segurança de sistemas críticos, metodologias para construção de casos de garantia e revisão da literatura que analisam práticas e desafios na construção de sistemas confiáveis com componentes de *Machine Learning*. Por fim, são apresentados trabalhos que compartilham objetivos semelhantes ao desta pesquisa, contribuindo para contextualizar a proposta.

2.1 Engenharia de Requisitos

A Engenharia de Requisitos (ER) é o ramo da Engenharia de Software responsável por identificar, analisar, documentar e gerenciar as necessidades e restrições de um sistema de software [8]. É uma etapa essencial durante todo o ciclo de vida do produto que visa garantir que o sistema final atenda às expectativas e demandas dos diversos *stakeholders*, sejam eles clientes, usuários ou desenvolvedores.

Existem diversas técnicas para modelar e compreender tanto requisitos funcionais quanto não funcionais (como desempenho, segurança, usabilidade). *Nuseibeh & Easterbrook (2000)* [8] separaram as seis classes de abordagens a seguir:

- **Tradicionais**: Incluem diferentes formas genéricas de coletar dados, como questionários, entrevistas e análise de documentação preexistente.

- **Coletivas**: Exigem maior participação dos *stakeholders*, realizando dinâmicas em grupo para promover discussão e alinhamento entre os envolvidos.
- **Prototipagem**: Consiste em criar versões preliminares do sistema para obter feedback que ajude a refinar os requisitos.
- **Orientadas a modelos**: Utilizam modelos específicos do tipo de informação que deve ser coletada para a definição, documentação e análise de requisitos.
- **Cognitivas**: Visam compreender o pensamento e comportamento dos usuários para garantir que o sistema atenda às suas necessidades.
- **Contextuais**: Levam em consideração o contexto dos usuários, realizando análises para identificar padrões nos comportamentos e interações sociais.

Depois do passo de elicitação de requisitos, o processo da ER também prevê as seguintes fases [9]:

- **Análise**: Refinamento e detalhamento das informações coletadas para criar um conjunto coerente de requisitos.
- **Especificação**: Documentação de funcionalidades, restrições e comportamento desejado do sistema a partir dos requisitos analisados.
- **Validação e Verificação (V & V)**: Validação da documentação, além de garantir consistência, completude e corretude dos requisitos.

Nuseibeh & Easterbrook [8] também destacaram inconsistências nas especificações, gerenciamento da evolução dos requisitos e tratamento de requisitos não funcionais como alguns dos principais desafios da ER. Hoje, com o crescimento significativo da inteligência artificial, surgem novos obstáculos, já que começa a demanda de alta qualidade para os modelos baseados em aprendizado de máquina [3].

Para um software convencional, os requisitos funcionais especificam comportamentos estáveis e claros, como, por exemplo, funcionalidades, casos de uso e regras de negócio bem estabelecidos. São passíveis de verificação por testes funcionais e análise de desempenho, já que são esperados resultados determinísticos (uma mesma entrada retorna uma mesma saída) [10]. Mas para soluções de ML, os requisitos são baseados em dados, então o modelo pode variar conforme a qualidade e a distribuição dos dados de treinamento [10].

Fase	ER Tradicional	Extensões para ML
Elicitação	Entrevistas, questionários e *workshops* Prototipagem rápida Análise de documentação	+ Mapeamento de fontes de dados + Envolvimento de cientistas de dados e especialistas em ética
Análise	Priorização e refinamento de requisitos Verificação de consistência	+ Definição e interpretação de métricas de desempenho + Avaliação de qualidade, distribuição e proveniência dos dados
Especificação	Casos de uso Requisitos funcionais Requisitos não funcionais	+ Metas quantitativas para métricas de ML + Critérios de robustez, imparcialidade e transparência + Restrições regulatórias de dados
V & V	Validação da documentação Testes funcionais e de integração Revisão estática	+ Monitoramento contínuo em produção + Detecção de deriva de dados + Políticas de *retraining* e auditoria de viés

Tabela 2.1: Fases da ER tradicional e respectivas extensões para sistemas de *Machine Learning*

Fonte: Adaptado de *Vogelsang & Borg* [9]

Então, cada fase do processo tradicional da ER deve passar a incorporar passos voltados às especificidades dos modelos, como sintetizado na Tabela 2.1. Na elicitação, busca-se mapear fontes de dados relevantes e envolver desde o início cientistas de dados e especialistas em privacidade e ética. Na análise, faz-se necessário a definição e interpretação de métricas de desempenho, assim como a avaliação da qualidade e diversidade dos dados de treinamento. Na especificação, os requisitos ganham limites aceitáveis para essas métricas e detalham critérios de robustez, imparcialidade e transparência. Por fim, em V&V, as práticas tradicionais devem ser complementadas por monitoramento contínuo, detecção de deriva de dados e políticas de

retreinamento [9].

2.2 Certificações e Padrões de Segurança

Nesta seção, serão apresentados os principais referenciais, normas e metodologias empregados para garantir a segurança de sistemas críticos que incorporam aprendizado de máquina. Inicialmente, é introduzido o AMLAS, metodologia específica para estruturar a garantia de segurança de componentes de ML em sistemas autônomos. Na sequência, discutido os padrões automotivos ISO 26262 e SOTIF, destacando suas abordagens complementares. E por fim, o DO-178C, o padrão de certificação de software embarcado em aeronaves, apontando seus desafios e possíveis extensões para aplicações de ML.

2.2.1 AMLAS

O AMLAS (*Garantia de Aprendizagem de Máquina para Sistemas Autônomos*) é um método estruturado para garantir a segurança e confiabilidade de componentes baseados em ML utilizados em sistemas autônomos críticos [4]. Ele fornece um conjunto de passos para gerar um plano de segurança bem fundamentado que demonstra que o modelo opera de forma segura dentro do contexto operacional, considerando as limitações e incertezas dos modelos treinados com dados.

O processo AMLAS requer como entrada os requisitos de segurança em nível de sistema, definidos por especialistas ou normas, e organiza-se em seis estágios iterativos [4]:

- **Definição do Escopo**: Caracterizar o contexto operacional e limites do sistema.
- **Especificação de Requisitos**: Com base no escopo e riscos associados, definir os requisitos de segurança.
- **Gestão de Dados**: Coletar e preparar os dados necessários, garantindo que atendam aos requisitos de segurança.
- **Treinamento do Modelo**: Desenvolver o modelo usando os dados coletados, incluindo treinamento e validação interna.
- **Verificação do Modelo**: Utilizando dados distintos dos dados de treinamento, verificar o modelo, assegurando corretude e bom desempenho sob condições adversas.

- **Operação do Modelo**: Implantar o modelo no sistema autônomo e realizar monitoramento contínuo, identificando possíveis falhas ou desvios que possam afetar a segurança.

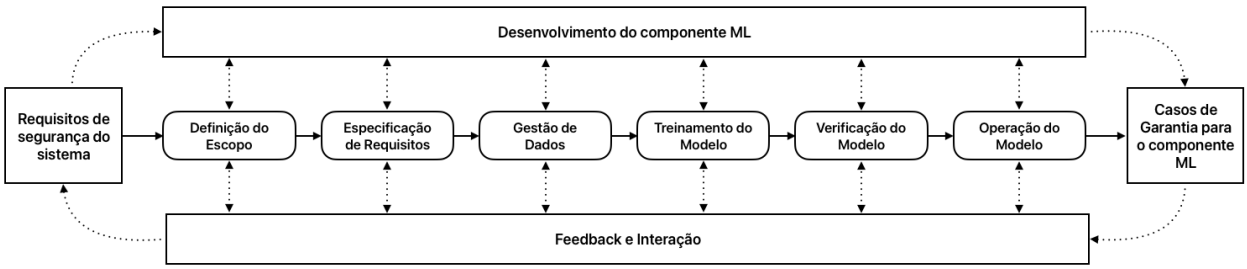

Figura 2.1: Estágios do Processo AMLAS

Por ser iterativo, o AMLAS permite revisitar estágios sempre que novas evidências exijam algum tipo de refinamento. Como pode ser visto na Figura 2.1, o fluxo de **Feedback e Interação** assegura que o componente de ML satisfaça continuamente os requisitos de segurança alocados ao longo de seu desenvolvimento e implantação [4].

2.2.2 ISO 26262

A ISO 26262 é o padrão internacional de referência para avaliação de segurança em sistemas automotivos, voltado para circuitos elétricos, eletroeletrônicos e software embarcado em veículos rodoviários. Tem como objetivo garantir que os sistemas operem de forma segura mesmo quando há falhas ou defeitos no hardware e software [11].

Para estruturar a avaliação de risco, é definido o ASIL (Nível de Integridade de Segurança Automotiva), que classifica os sistemas do nível A de menor risco para o nível D de maior risco conforme a severidade das consequências de uma falha e capacidade de controle por parte do motorista. Cada ASIL impõe um conjunto de passos obrigatórios ao longo de todo o ciclo de vida do produto, que vai desde definição de requisitos até verificação e validação [11]. A figura Figura 2.2 mostra este ciclo de vida simplificado [12].

No entanto, a ISO não cobre as particularidades de modelos de ML nem trata perigos resultantes do funcionamento esperado do sistema, mas que podem ser inseguros devido a deficiências funcionais ou ambientais, como, por exemplo, falhas no reconhecimento do ambiente ou limitações de percepção, algo comum em sistemas com ML. Por isso, embora essencial, sua

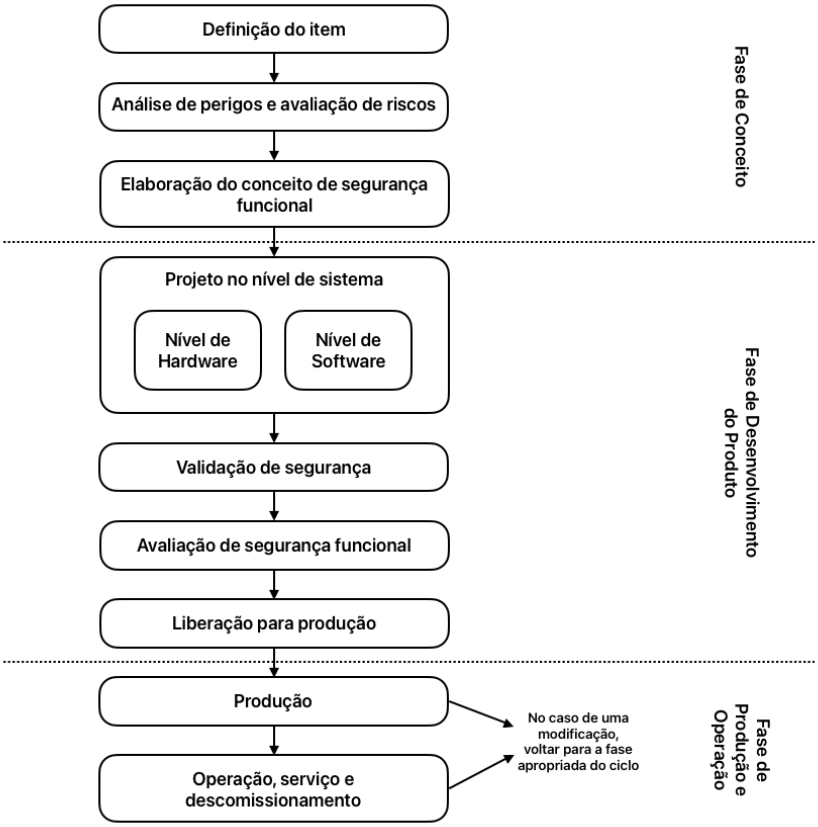

Figura 2.2: Ciclo de vida simplificado da ISO 26262

aplicação isolada é insuficiente para certificar veículos autônomos baseados em aprendizado de máquina [13].

2.2.3 SOTIF

A ISO 21448, mais conhecida como SOTIF (Segurança da Funcionalidade Intencional), abrange as lacunas deixadas pela ISO 26262. Foi projetada para sistemas de assistência ao condutor (ADAS) e tem como objetivo tratar de condições em que o sistema funciona conforme projetado, mas o comportamento esperado ainda pode resultar em riscos, como falsos positivos ou negativos, decorrentes de limitações do modelo ou dos dados [11].

O processo também é muito similar à ISO 26262. Primeiro, é feita uma análise de perigos resultantes de deficiências funcionais e depois a identificação de cenários críticos que podem desencandear esses perigos. Em seguida, o sistema passa pelo processo de V & V, onde são realizados testes em condições conhecidas e são explorados cenários novos ou não previstos. Por fim, são feitas as modificações funcionais ou de requisitos caso o risco residual ainda seja inaceitável [11].

Ambas as normas requerem processos iterativos e documentação rigorosa para a construção de um plano de segurança, especialmete quando integradas com metodologias como o AMLAS. Juntas, elas formam uma base importante para a certificação de sistemas automotivos, especialmente para os que incorporam aprendizado de máquina [11].

2.2.4 DO-178C

A DO-178C é o padrão internacionalmente aceito para a certificação de software embarcado em sistemas de aeronaves, de modo a garantir que cumpram requisitos rigorosos de segurança e confiabilidade ao longo de todo o seu ciclo de vida [11].

O padrão define os seguintes 5 níveis de DAL (Nível de Garantia de Design) baseados no impacto potencial de uma falha de software [14]:

- **Nível A (Catástrofe)**: Falha pode causar perda da aeronave ou fatalidades.
- **Nível B (Perigo Grave)**: Falha pode resultar em graves lesões aos ocupantes.
- **Nível C (Menor Perigo)**: Falha provoca lesões leves ou desconforto.
- **Nível D (Falha Limitada)**: Falha reduz funcionalidade, mas sem riscos significativos à segurança.
- **Nível E (Sem Impacto)**: Falha sem efeito prático na segurança do voo.

Embora forneça uma base sólida para softwares determinísticos, ele não aborda aspectos centrais de sistemas baseados em ML, mas já existem estudos que estão adaptando a abordagem do DO-178C. Como, por exemplo, a FAA (Administração Federal de Aviação dos EUA), que tem buscado implementar o padrão juntamente com a noção de *Overarching Properties* (propriedades abrangentes), que incluem intenção, correção e aceitabilidade [11].

Sridhar et al. (2025) [15] também estenderam a certificação através da criação de métodos semi-automatizados com foco em sistemas de criticidade baixa com modelos estáticos, com o intuito de facilitar a aplicação prática imediata. O novo processo inclui avaliações específicas para a natureza de ML, como validação contínua do desempenho, análise do conjunto de dados e robustez frente a variações reais.

2.3 Casos de garantia

Casos de garantia surgem da evolução dos casos de segurança para abranger qualquer propriedade crítica de um sistema. Enquanto os casos de segurança focavam em demonstrar que um produto era seguro, os casos de garantia generalizam essa abordagem, exigindo que o sistema seja estruturado pelos seguintes elementos [7]:

- **Reivindicações**: São afirmações sobre propriedades particulares ou caracterísicas do sistema.
- **Argumentos**: São conexões lógicas usadas para explicar o motivo das evidências sustentam as reivindicações.
- **Evidências**: Consistem em testes, análises formais, simulações ou revisões que comprovam a viabilidade das reivindicações.

Ao explicitar tanto o *"o quê"* (reivindicações) quanto o *"por quê"* (argumento) e o *"com que"* (evidências), é possível construir um raciocínio transparente, capaz de resistir a auditorias e revisões [7].

Para montar os casos, é formada uma estrutura em árvore, onde cada argumento desce um nível, usando sub-reivindicações para sustentar as reivindicações de nível superior até atingir as evidências nos nós folhas. A abordagem mais conhecida para fazer uma representação gráfica é o GSN (Notação de Estruturação de Metas). Nele, são definidos nós de **objetivo**, **estratégia**, **solução**, **pressuposição** e **justificativa**, além de nós de **contexto** para delimitar escopo e termos. O tipo do nó é indicado por sua forma e as setas conectam objetivos às estratégias até chegar nas soluções [7]. É possível ver um exemplo dessa árvore na Figura 2.3.

Quando aplicados a sistemas baseados em aprendizagem de máquina, esses conceitos não podem se limitar a documentos estáticos, pois é preciso lidar diretamente com a incerteza e a complexidade dos modelos probabilísticos. Fazem-se necessárias revisões periódicas de cada objetivo e a coleta contínua de novas evidências à medida que surgem situações de uso real [16].

Burton & Herd (2023) [16] sugerem inserir nos casos de garantia sub-objetivos que tratem explicitamente de deficiências de especificação, performance, dados e operação. Em um diagrama de GSN, a reivindicação principal é decomposta em quatro sub-objetivos correspondentes a essas categorias, cada um sustentado por soluções específicas. Pressuposições e justificativas ajudam a explicitar, por exemplo, quais distribuições de dados foram consideradas ou por que determinadas métricas de robustez foram julgadas suficientes.

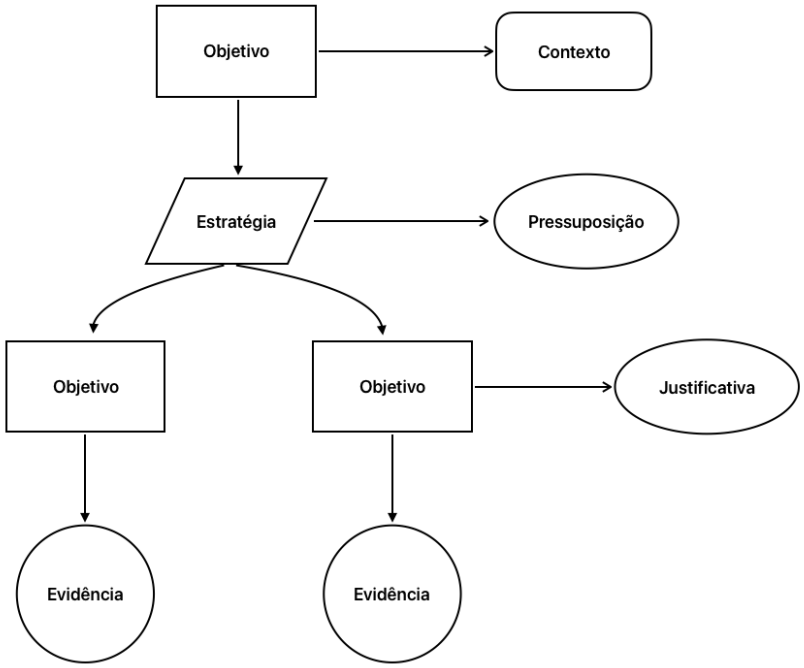

Figura 2.3: Exemplo de árvore do GSN

2.4 Revisões e Estudos Gerais sobre Confiabilidade

Esta seção apresenta os artigos utilizados na revisão da literatura que discutem, de forma direta ou indireta, desafios, métodos e práticas relacionados à confiabilidade em sistemas baseados em *Machine Learning*. Esses estudos fornecem visões gerais e diretrizes que abordam requisitos para que modelos de ML operem de forma segura, robusta e responsável. Ao comentar cada um deles, busca-se destacar seus principais objetivos, métodos, resultados e contribuições para a fundamentação deste trabalho. A Tabela 2.2 mostra uma comparação dos artigos em relação a essas categorias.

Borg et al. (2023) [13] demonstraram como construir um plano de segurança completo para um componente de ML em um sistema crítico de frenagem automática de emergência para pedestres (PAEB), o SMIRK. Foi utilizada a metodologia AMLAS em conjunto com o processo SOTIF para guiar as atividades de desenvolvimento, especificação de requisitos, coleta e validação de dados, treinamento e verificação do modelo e elaboração dos casos de garantia. O estudo contribui de forma significativa ao detalhar os requisitos de desempenho, de segurança, de dados e de robustez, além de explicitar práticas para rastreabilidade e validação contínua. O artigo também evidencia a importância de requisitos de dados como relevância, completude e acurácia, que são essenciais para a confiabilidade de modelos em domínios críticos.

Sridhar et al. (2025) [15] propuseram um processo de certificação semi-automatizado para sistemas inteligentes aplicados em contextos aeronáuticos de baixa criticidade. O método combina processos manuais e automáticos para avaliar a validade do modelo e a robustez do sistema, apresentando uma abordagem de classificação que adapta a profundidade das verificações à criticidade do sistema, ao nível de autonomia e à complexidade do modelo ML utilizado. Como estudo de caso, os autores aplicaram a metodologia em um sistema de detecção de objetos em tempo real projetado para aeronaves de reconhecimento. O trabalho propõe práticas para a avaliação contínua de desempenho, validação de dados e monitoramento pós-certificação.

Odu et al. (2025) [17] investigaram a criação manual e automática de planos de segurança para o componente de previsão de trajetória do sistema de direção autônoma *Baidu Apollo*. Inicialmente, foram construídos casos de garantia pelos autores utilizando a AMLAS combinada com análise de riscos e, em seguida, foi utilizado um *Large Language Model* (LLM) para gerar os mesmos casos. Após a comparação das abordagens, foi possível perceber que o LLM pode gerar casos de garantia próximos ao manual quando alimentado com informações contextuais e exemplos. O estudo destaca requisitos do sistema como robustez em diferentes cenários operacionais e gerenciamento de mudanças, além de ressaltar a importância de processos formais para garantir a rastreabilidade dos requisitos de segurança.

Gujral et al. (2024) [18] descreveram o desenvolvimento de uma ferramenta projetada para avaliar riscos de segurança em sistemas de aeronaves não tripuladas (UAS). O objetivo foi permitir que gerentes de frotas dessas aeronaves identifiquem, monitorem e mitiguem riscos antes da decolagem, usando serviços que avaliam riscos como perda de sinal de GPS, proximidade de obstáculos, entre outros. O método consistiu em integrar os serviços em uma interface com mapas interativos. Foram citados requisitos do sistema, como atualizações em tempo real dos dados e visualizações claras dos perigos nas rotas, demonstrando como fatores determinantes para a confiabilidade são essenciais para sistemas sensíveis a falhas.

Hong at al. (2025) [19] mostraram como análises de risco sistemáticas podem identificar perigos e antecipar perdas. Usando como exemplo uma aplicação ML para recomendação de trilhas, foi visto que muitos perigos podem ser previstos e mitigados se métodos estruturados forem aplicados desde as fases iniciais de desenvolvimento. O trabalho contribui para a confiabilidade em soluções que empregam aprendizado de máquina ao propor a integração de processos de análises de risco como parte do ciclo de vida das aplicações, facilitando a identificação e mitigação de cenários críticos.

Sivakumar (2024) [20] construiu manualmente um plano de segurança para um componente ML de um veículo autônomo, seguindo metodologias como AMLAS e HARA (Análise de Perigos e Avaliação de Riscos) e avaliou o uso do GPT-4 para gerar o plano automaticamente. Foi observado que o GPT-4 produziu casos de garantia com similaridade moderada, mas ainda requer validação humana para assegurar qualidade e correção. O trabalho enfatiza a necessidade de casos de garantia completos e acessíveis para fundamentar a certificação de sistemas críticos com aprendizado de máquina. Destacam-se requisitos como precisão, robustez à variação de cenário e rastreabilidade dos dados utilizados no treinamento dos modelos como indispensáveis para garantir confiabilidade em sistemas baseados em ML.

Zeller et al. (2024) [6] apresentaram um *pipeline* de operações para aprendizado de máquina (MLOps) aplicado a sistemas no setor ferroviário, com objetivo de possibilitar o desenvolvimento e a validação contínuos de funções críticas como detecção de obstáculos em trens autônomos. O método inclui ferramentas para análise de qualidade de dados, verificação automatizada dos modelos, capacidade de detecção de anomalias, além do gerenciamento de casos de garantia. A abordagem destaca que a escolha cuidadosa das métricas e a definição de limites para os indicadores de segurança são desafios determinantes para a qualidade dos casos de garantia.

Sivakumar et al. (2024) [21] investigaram como LLMs podem ser utilizados para gerar casos de garantia. Foram conduzidos experimentos em dois grupos: o primeiro avaliou o conhecimento do GPT-4 sobre regras estruturais e sintáticas do GSN (Notação de Estruturação de Objetivos), amplamente usada para representar o plano de segurança de forma gráfica ou em texto estruturado; o segundo, a capacidade do GPT-4 de gerar casos de garantia para sistemas reais, incluindo um componente de detecção de ruído em pneus baseado em aprendizado de máquina. O trabalho evidencia que LLMs podem ajudar na especificação de requisitos de confiabilidade, especialmente ao reduzir o esforço na documentação de segurança.

Zeller (2023) [22] propôs uma abordagem para avaliar a segurança de sistemas autônomos que incorporam modelos de ML. Foram utilizadas CFDTs (Árvores de Falhas e Deficiências por Componentes) para modelar as relações de causa e efeito, o que possibilitou a identificação de riscos e a definição de medidas de mitigação. Neste trabalho, a confiabilidade está relacionada à capacidade do sistema de evitar falhas e deficiências funcionais que possam levar a riscos inaceitáveis.

Socha et al. (2022) [23] apresentaram um sistema de frenagem automática de emergência para pedestres desenvolvido com ML e implementado em um simulador de grau industrial. O

estudo construiu um demonstrador aberto e completo, combinando o código-fonte, dados de treinamento e um plano de segurança. Os resultados incluem a criação de casos de garantia baseados na metodologia AMLAS e requisitos específicos de confiabilidade, como precisão mínima na detecção, robustez contra entradas inesperadas e mitigação de falsos positivos, ressaltando a necessidade de lidar com deficiências funcionais para garantir segurança em aplicações autônomas.

Tabela 2.2: Comparação dos artigos revisados quanto a objetivo, método, sistema e requisitos abordados

Artigo	**Objetivo Principal**	**Método / Abordagem**	**Sistema / Exemplo**	**Requisitos Abordados**
Borg et al. (2023) [13]	Construir caso de segurança completo para ML crítico	AMLAS + SOTIF	SMIRK (PAEB)	Contexto operacional, desempenho, segurança, representatividade dos dados
Sridhar et al. (2025) [15]	Propor certificação semi-automatizada para ML aeronáutico	DO-178C adaptado	Detector de objetos em aeronaves	Validação formal e contínua, consistência, gerenciamento de dados, adequação à criticidade
Odu et al. (2025) [17]	Comparar casos de segurança manuais e gerados por LLM	AMLAS + LLM	*Baidu Apollo* (previsão de trajetória)	Robustez em cenários diversos, rastreabilidade
Gujral et al. (2024) [18]	Avaliar riscos em planejamento pré-voo de drones	Ferramenta de avaliação de riscos	Sistemas UAS	Atualização em tempo real, clareza na apresentação dos riscos
Hong et al. (2025) [19]	Antecipar perigos via análise proativa de riscos	STPA + suporte de LLM	Recomendação de trilhas	Identificação de perigos, integração de análises de risco, *fairness* e privacidade
Sivakumar (2024) [20]	Criar manualmente e avaliar geração automática de *safety cases*	AMLAS + HARA + GPT-4	Veículo autônomo	Precisão, robustez a variações de cenário, rastreabilidade, validação de rótulos
Zeller et al. (2024) [6]	Desenvolver *pipeline* contínuo com garantia de segurança	MLOps + análise automatizada	*safe.trAIn* (domínio ferroviário)	Qualidade de dados e rótulos, métricas de robustez, monitoramento contínuo
Sivakumar et al. (2024) [21]	Avaliar uso de LLMs na criação de *safety cases*	GSN + GPT-4	Sistema de detecção de ruído em pneus	Especificação automática, rastreabilidade
Zeller (2023) [22]	Avaliar segurança com CFDTs para ML em sistemas autônomos	CFDT	*PANORover* (veículo autônomo)	Mitigação de deficiências, rastreabilidade, segurança funcional
Socha et al. (2022) [23]	Construir demonstrador aberto com *safety case* completo	AMLAS + SOTIF	SMIRK (PAEB)	Precisão, robustez a entradas inesperadas, mitigação de falsos positivos

2.5 Trabalhos Relacionados

Nos últimos anos, o avanço e a adoção crescente de sistemas autônomos têm ampliado a preocupação com a sua confiabilidade, segurança e responsabilidade. Diante da natureza probabilística desses sistemas e da dificuldade de interpretar e validar seu comportamento, surge a necessidade de estabelecer requisitos específicos que garantam seu funcionamento adequado, mesmo sob condições adversas.

Esta seção apresenta alguns trabalhos que têm objetivos semelhantes ao desta pesquisa, abordando o desafio de definir critérios confiáveis para sistemas inteligentes e de propor soluções que sustentem seu uso seguro e responsável.

Carvalho (2024) [24] propõe um conjunto de 32 requisitos não funcionais (RNFs) baseados no Projeto de Lei 2338/2023, que visa regulamentar o uso de inteligência artificial no Brasil. Os requisitos são categorizados segundo uma taxonomia unificada e agrupados em aspectos como segurança, ética, transparência, confiabilidade e usabilidade. A metodologia inclui análise documental da legislação e revisão de literatura, com mapeamento entre os artigos da lei e os RNFs definidos. Embora não tenha realizado avaliação com especialistas, o trabalho oferece uma estrutura detalhada e promissora para avaliação de conformidade legal em sistemas de IA. Este estudo foi identificado durante a aplicação do questionário, portanto, não foi utilizado como base de identificação dos requisitos de confiabilidade.

Sadowski et al. (2022) [25] realizaram uma revisão sistemática da literatura sobre métodos de garantia de segurança aplicados a sistemas baseados em ML. Os autores organizam os métodos em categorias como técnicas de verificação formal, testes, monitoramento em tempo de execução e construção de casos de garantia. O trabalho destaca a ausência de consenso sobre requisitos específicos para ML e a dificuldade de adaptar metodologias tradicionais de Engenharia de Software a esse novo contexto.

Balagurunathan et al. (2021) [26] discutem os principais requisitos técnicos, éticos e operacionais para garantir a confiabilidade de sistemas de inteligência artificial baseados em ML no domínio médico, com foco na oncologia. Os autores organizam os desafios enfrentados ao longo do ciclo de vida do desenvolvimento de modelos, propondo um processo dividido em fases (inicial, intermediária e final) que contempla desde a preparação de dados até a validação clínica e a integração em ambientes hospitalares. Este trabalho não foi utilizado para a identificação dos requisitos de confiabilidade, pois foi encontrado após todas as etapas terem sido concluídas.

Paterson et al. (2025) [27] reorganizam e detalham a aplicação prática do AMLAS, vi-

sando torná-lo mais acessível. Para isso, em cada uma das etapas definidas pelo AMLAS, os autores apresentam recomendações práticas, métricas específicas e critérios de aceitação, além de discutirem desafios enfrentados em aplicações reais. O trabalho contribui significativamente para a sistematização de requisitos de confiabilidade ao transformar as etapas do AMLAS em atividades concretas e auditáveis, facilitando sua adoção em domínios críticos como transporte e robótica.

Capítulo 3

Metodologia de Pesquisa

Este capítulo apresenta a metodologia seguida para atingir o objetivo deste trabalho, que é identificar e caracterizar requisitos de confiabilidade em sistemas de *Machine Learning*.

Nele, são descritos a abordagem da pesquisa, o processo de seleção e extração dos requisitos, a forma de classificação utilizada, o método de validação com especialistas e a análise dos dados coletados.

3.1 Abordagem da pesquisa

Esta pesquisa possui dois passos principais:

- **Revisão da literatura**: mapeamento e síntese dos principais requisitos relatados na literatura acadêmica sobre ML confiável.
- **Questionário**: pesquisa quantitativa acompanhado de análise qualitativa dos comentários abertos para avaliar e refinar os requisitos extraídos.

3.2 Revisão da Literatura

3.2.1 Processo de seleção dos artigos

O processo de seleção dos artigos foi conduzido em três etapas, de modo a garantir que as referências utilizadas reflitam as práticas mais relevantes e recentes. O período de seleção iniciou-se em 30 de Março de 2025 e terminou-se em 10 de Maio de 2025. Primeiramente,

identificaram-se autores com histórico consolidado em confiabilidade de *software*, casos de garantia e requisitos voltados para ML. Nesse estágio, destacaram-se autores como *Markus Borg* [13], *Mithila Sivakumar* [20] e *Marc Zeller* [22], que possuem trabalhos que abordam metodologias detalhadas e estudos de caso voltados ao tema.

Em seguida, a busca bibliográfica foi expandida por meio de análise de citações, utilizando ferramentas como *Google Scholar* [1] e *Connected Papers* [2] para localizar artigos que citem ou sejam citados pelos autores inicialmente selecionados. Essa estapa permitiu identificar trabalhos complementares como os de *Chandrasekar Sridhar* [15] e *Oluwafemi Odu* [17], garantindo cobertura ampla de abordagens, desafios e soluções relacionados à confiabilidade em ML.

Por fim, aplicou-se um filtro temporal, restringindo a seleção a publicações de 2022 em diante, de modo a focar em práticas e desafios contemporâneos e assegurar que a revisão estivesse alinhada com as tendências mais recentes do tema.

3.2.2 Extração dos requisitos

A fase de extração de requisitos foi iniciada com a leitura completa de cada artigo selecionado, com atenção às seções que descrevem requisitos explícitos, critérios de qualidade ou estruturas de caso de garantia.

A partir da leitura, todos os requisitos foram registrados em um ambiente do *Notion* [3], por meio de uma base de dados customizada que contemplava colunas para descrição do requisitos, referência bibliográfica e observações adicionais. Esse formato no *Notion* permitiu organização, busca e filtragens rápidas durante o processo.

A seguir, um exemplo de como foi realizada a extração de requisitos a partir de *Borg et al. (2023)* [13]. Este trabalho possui seções que indicam explicitamente os requisitos do sistema utilizado como objeto de estudo. Um dos citados foi:

> *"SMIRK shall commence automatic emergency braking if and only if collision with a pedestrian on collision course is imminent"* [13]

que, em tradução livre, significa "O SMIRK deve iniciar a frenagem de emergência automática se e somente se uma colisão com um pedestre em rota de colisão for iminente".

[1] https://scholar.google.com/
[2] https://www.connectedpapers.com/
[3] https://www.notion.com/

Esse enunciado estabelece de forma clara a condição sob a qual o sistema deve acionar a mitigação. Em um contexto genérico de sistemas de ML, foi reescrito como "O sistema deve ser capaz de detectar situações de risco com alta confiabilidade e realizar ações de mitigação que mantenham a segurança dos usuários e do ambiente".

De modo análogo, a especificação de dados no artigo contêm:

> *"The data samples shall include the complete range of environmental factors within the scope of the ODD."* [13]

> *"The data samples shall include images representing all types of pedestrians according to the demographics of the ODD."* [13]

que, em tradução livre, diz "As amostras de dados devem incluir toda a gama de fatores ambientais dentro do escopo do ODD" e "As amostras de dados devem incluir imagens que representem todos os tipos de pedestres conforme a demografia definida no ODD", onde ODD é a sigla para *Operational Design Domain* (ou Domínio de Design Operacional), que é um termo para definir o conjunto de condições ambientais, operacionais e contextuais em que o sistema pode funcionar com segurança.

Esses requisitos asseguram que os dados de treinamento e teste cubram todas as possíveis variações de ambiente e de usuários, evitando viés e melhorando a robustez do modelo. Deles, foi retirado que "Os conjuntos de dados usados para treinamento, validação e testes devem ser representativos do domínio operacional definido, contemplando diversidade de condições ambientais e perfis de usuários".

Dessa forma, é possível manter a essência técnica dos requisitos originais, mas que pode ser aplicável a qualquer sistema baseado em aprendizagem de máquina.

3.2.3 Classificação e caracterização

Com os requisitos extraídos, todos os itens foram organizados em cinco categorias principais, de modo a facilitar a compreensão de suas finalidades e aplicação prática em sistemas de ML. As categorias são:

- **Escopo**: Relacionados à delimitação do alcance funcional e das condições de operação do sistema.

- **Segurança**: Englobam a proteção do sistema contra falhas internas e ataques externos que possam comprometer tanto a integridade dos dados quanto o funcionamento correto do modelo.

- **Dados**: Focam na origem, qualidade, diversidade e governança do conjunto de dados utilizado em todas as fases do ciclo de vida do sistema.

- **Desempenho**: Garantem a qualidade preditiva do modelo, sua robustez e competitividade em cenários operacionais reais.

- **Validação**: Relacionados a estratégias e métricas para testar, monitorar e retreinar o modelo ao longo do tempo.

3.3 Pesquisa com especialistas

Para organizar os requisitos identificados e coletar informações sobre sua relevância, clareza e completude, foi aplicado um formulário online hospedado no *Google Forms*[4]. O instrumento foi estruturado em seções que refletiam o fluxo de coleta de dados adotado neste trabalho. O conteúdo completo do formulário encontra-se no Apêndice A.

A primeira seção apresentava um Termo de Consentimento Livre e Esclarecido, onde informava que a participação era voluntária, que os dados seriam coletados de forma anônima e que nenhuma informação pessoal identificável seria retida. Apenas os respondentes que confirmaram o termo puderam prosseguir.

Na sequência, a seção de Perfil Profissional era composta por perguntas sobre a função principal do participante, o tempo de experiência com *Machine Learning*, o número de projetos de ML em que atuou e quais deles envolveram elaboração de casos de garantia. Também foi questionado como cada profissional teve acesso ao formulário para ser possível avaliar o alcance e a diversidade do público alcançado.

As próximas seções foram dedicadas às categorias dos requisitos, sendo uma seção para cada categoria. Foram feitas perguntas diretamente relacionadas à lista de requisitos onde os participantes indicaram seu grau de concordância em uma escala de 1 (Discordo totalmente) a 5 (Concordo totalmente). A abordagem quantitativa forneceu medidas diretas de aceitação e apontou requisitos que poderiam demandar reformulação ou maior detalhamento.

[4] https://docs.google.com/forms/

Por fim, foi aberto um espaço para que os especialistas sugerissem novos requisitos ou fizessem observações sobre os apresentados. Esse campo livre capturou percepções que não cabiam nas perguntas fechadas, enriquecendo a análise posterior.

Para compor o painel de respondentes, o formulário foi divulgado em grupos de *WhatsApp*[5], listas de e-mail voltadas a estudantes, professores e profissionais de computação, além de ser compartilhado no feed do *LinkedIn* e enviado diretamente pela plataforma[6] a especialistas com histórico em ML e Ciência de Dados. O formulário permaneceu disponível por duas semanas, no período de 26 de junho de 2025 a 10 de julho de 2025. Dessa forma, foi possível garantir uma amostra diversa de visões e experiências, fundamental para validar e refinar o conjunto final de requisitos.

3.4 Análise dos dados

A análise seguiu três etapas complementares para combinar evidências quantitativas e qualitativas, finalizando com uma síntese final dos requisitos.

Inicialmente, todas as respostas às perguntas de escala foram consolidadas em uma planilha, permitindo o cálculo de métricas para cada requisito. Com o *Google Forms*, foram gerados gráficos de barras que ilustram a distribuição de concordância por requisito e por categoria, identificando itens com alto grau de aceitação e aqueles que apresentaram maior dispersão de opiniões.

Depois, foi conduzida uma análise dos comentários livres coletados na seção final do formulário. Cada comentário foi lido e agrupados em categorias temáticas. A partir desse mapeamento, foi possível destacar as recomendações recorrentes e os pontos de melhorias para a lista de requisitos.

Combinando os achados quantitativos e qualitativos, os requisitos foram revisados e, quando necessário, ajustados. O resultado desse processo foi a consolidação de um conjunto final de requisitos validado por métricas objetivas e enriquecido por opiniões de praticantes e pesquisadores.

[5] https://www.whatsapp.com/
[6] https://br.linkedin.com/

Capítulo 4

Resultados

Este capítulo apresenta os achados obtidos a partir da etapa de extração inicial de requisitos e de sua validação junto a especialistas. Primeiro, é apresentado o conjunto de requisitos provindos apenas da revisão da literatura organizado por categoria. Em seguida, as análises quantitativas e qualitativas da pesquisa.

4.1 Requisitos extraídos da revisão da literatura

Esta seção apresenta o conjunto de requisitos de confiabilidade elaborados, organizados nas categorias de escopo, segurança, dados, desempenho e validação. A Tabela 4.1 mostra a rastreabilidade entre cada requisito e os artigos que fundamentaram sua formulação.

4.1.1 Requisitos de Escopo

REQ-ESC-01 - Definição do Contexto Operacional

O sistema deve possuir uma definição formal e documentada do seu *Operational Design Domain* (ODD), incluindo:

- Condições ambientais, geográficas, temporais e operacionais válidas para sua execução segura.
- Exemplos de cenários operacionais típicos e críticos (ex.: rodovias urbanas, ambientes internos com baixa luminosidade).
- Atribuição de pesos de criticidade para cada condição do ODD, com o objetivo de orientar o esforço de teste e validação.

REQ-ESC-02 - Especificação das Interfaces e Limites do Sistema ML

O sistema deve definir claramente os limites funcionais do componente de *Machine Learning*, incluindo:

- Entradas esperadas: formatos, faixas válidas, taxas de chegada e restrições temporais (ex.: tempo máximo de resposta).
- Saídas geradas: formatos, escalas e intervalos esperados.
- Contratos formais de interação com outros componentes do sistema global.

REQ-ESC-03 - Classificação do Sistema ML

O sistema deve ser classificado com base nos seguintes critérios:

- Criticidade do contexto de uso, considerando o impacto potencial de falhas (ex.: segurança de pessoas, impacto financeiro, reputacional).
- Nível de autonomia do sistema em cada fase (treinamento, inferência, decisão final), especificando o grau de supervisão humana esperado.
- Complexidade do modelo, categorizando-o como simples (ex.: regressão logística) ou complexo (ex.: redes neurais profundas).

4.1.2 Requisitos de Segurança

REQ-SEG-01 - Detecção e Resposta a Cenários Críticos

O sistema deve ser capaz de detectar situações de risco com alta confiabilidade e realizar ações de mitigação que mantenham a segurança dos usuários e do ambiente.

REQ-SEG-02 - Implementação de *Fail-Safe*

Em caso de falhas internas, entradas inválidas ou condições fora do ODD, o sistema deve adotar automaticamente um comportamento *fail-safe*[1], entrando em um estado seguro previamente definido.

[1] *Fail-safe* é a capacidade de um sistema adotar um estado seguro em caso de falhas.

REQ-SEG-03 - Garantia de *Fairness*

O sistema deve evitar vieses discriminatórios, assegurando a *fairness*[2] nas decisões automatizadas.

REQ-SEG-04 - Garantia de Privacidade

O sistema deve cumprir com legislações de proteção de dados (ex.: LGPD [3], GDPR [4]), incluindo:

- Mecanismos de consentimento para coleta e uso de dados pessoais.
- Políticas de anonimização ou pseudonimização.

REQ-SEG-05 - Monitoramento de Ameaças

O sistema deve possuir mecanismos de monitoramento contínuo para:

- Detecção de padrões de entrada suspeitos.
- Identificação de tentativas de ataques adversariais.
- Geração de alertas de segurança.

REQ-SEG-06 – Robustez a Perturbações e Condições Fora do Domínio

O sistema deve manter comportamento seguro e consistente mesmo quando exposto a situações não ideais, como:

- Ruídos ou perturbações nos dados de entrada (ex.: *motion blur* [5], oclusões, ruído de sensores, variações extremas de iluminação).
- Entradas inesperadas ou fora do domínio conhecido de operação (fora do ODD).

Mesmo diante de incertezas, o sistema não deve tomar ações perigosas ou instáveis.

[2] *Fairness* é a propriedade de um sistema de aprendizado de máquina de produzir decisões justas, sem discriminar grupos com base em atributos sensíveis como raça, gênero ou idade.

[3] Lei Geral de Proteção de Dados (LGPD), legislação brasileira que regula o uso de dados pessoais.

[4] General Data Protection Regulation (GDPR), regulamentação europeia sobre proteção de dados pessoais.

[5] *Motion blur* é o desfoque causado por movimento rápido da câmera ou do objeto durante a captura da imagem.

4.1.3 Requisitos de Dados

REQ-DAD-01 - Representatividade dos Dados

Os conjuntos de dados usados para treinamento, validação e testes devem ser representativos do domínio operacional definido, contemplando diversidade de situações e cenários.

REQ-DAD-02 - Validação de Rótulos (*Labels*)

Deve-se estabelecer e documentar processos formais para garantir a acurácia dos rótulos, incluindo:

- Revisão manual por especialistas.
- Validação cruzada automatizada.
- Definição de critérios mínimos de qualidade de rotulagem.

REQ-DAD-03 - Qualidade e Rastreabilidade dos Dados

Os dados utilizados devem ser completos, corretos e rastreáveis, com documentação de origem, versão e histórico de transformações.

REQ-DAD-04 - Gerenciamento de Dados e Controle de Viés

Devem existir políticas de governança de dados para:

- Monitoramento contínuo de viés (*data drift* [6] e *concept drift* [7]).
- Atualização periódica dos datasets de treinamento.
- Garantia de integridade e representatividade ao longo do ciclo de vida.

[6] *Data drift* é a mudança na distribuição dos dados de entrada ao longo do tempo
[7] *Concept drift* refere-se à mudança na relação entre entradas e saídas esperadas.

4.1.4 Requisitos de Desempenho

REQ-DES-01 - Métricas Quantitativas de Performance

O desempenho do modelo de ML deve ser avaliado com base em métricas previamente definidas (ex.: precisão [8], *recall*[9], *F1-score* [10], mAP[11]).

Os valores mínimos aceitáveis para cada métrica devem ser especificados antes da implantação.

REQ-DES-02 - Consistência em Condições Reais

O sistema deve manter um desempenho consistente frente às variações normais de entrada esperadas no ambiente operacional.

REQ-DES-03 - Comparação com Sistemas de Referência

Antes da implantação, o sistema deve demonstrar desempenho igual ou superior a soluções existentes ou legadas, considerando tanto aspectos de segurança quanto de eficiência.

4.1.5 Requisitos de Validação

REQ-VAL-01 - Validação Formal de Conformidade

Deve ser realizada validação formal demonstrando que os requisitos de *Machine Learning* atendem aos requisitos de sistema de mais alto nível, especialmente em termos de segurança.

REQ-VAL-02 - Cobertura de Cenários Operacionais e de Falha

Os testes de validação devem abranger tanto os cenários operacionais normais quanto as condições de falha conhecidas e possíveis.

REQ-VAL-03 - Validação Contínua ao Longo do Ciclo de Vida

Toda nova versão do modelo, alteração de dados ou modificação de código relevante deve passar por um processo de revalidação antes da liberação para produção.

[8] *Precisão* mede a proporção de previsões positivas feitas pelo modelo que estão corretas.
[9] *Recall* mede a proporção de casos relevantes que foram corretamente identificados pelo modelo.
[10] *F1-score* é a média harmônica entre precisão e *recall*, usada para balancear os dois aspectos.
[11] *mAP* (mean Average Precision) é a média da precisão obtida em diferentes limiares de decisão.

Tabela 4.1: Rastreabilidade entre os requisitos e os artigos da literatura

Requisito	**Fontes de referência**
REQ-ESC-01	*Borg et al. (2023)*[13], *Zeller et al. (2024)*[6]
REQ-ESC-02	*Borg et al. (2023)*[13]
REQ-ESC-03	*Sridhar et al. (2025)*[15]
REQ-SEG-01	*Borg et al. (2023)*[13]
REQ-SEG-02	*Sridhar et al. (2025)*[15], *Zeller (2023)*[22]
REQ-SEG-03	*Hong et al. (2025)*[19]
REQ-SEG-04	*Hong et al. (2025)*[19]
REQ-SEG-05	*Hong et al. (2025)*[19], *Sivakumar (2024)*[20]
REQ-SEG-06	*Sridhar et al. (2025)*[15], [17], *Sivakumar (2024)*[20], *Socha et al. (2022)*[23]
REQ-DAD-01	*Borg et al. (2023)*[13], *Zeller et al. (2024)*[6]
REQ-DAD-02	*Sridhar et al. (2025)*[15], *Sivakumar (2024)*[20], *Zeller et al. (2024)*[6]
REQ-DAD-03	*Sridhar et al. (2025)*[15], *Odu et al. (2025)*[17], *Sivakumar (2024)*[20], *Zeller et al. (2024)*[6], *Sivakumar et al. (2024)*[21], *Zeller (2023)*[22]
REQ-DAD-04	*Sridhar et al. (2025)*[15], *Sivakumar (2024)*[20]
REQ-DES-01	*Borg et al. (2023)*[13], *Sridhar et al. (2025)*[15], *Sivakumar (2024)*[20], *Socha et al. (2022)*[23]
REQ-DES-02	*Borg et al. (2023)*[13], *Sridhar et al. (2025)*[15]
REQ-DES-03	*Sivakumar (2024)*[20]
REQ-VAL-01	*Sridhar et al. (2025)*[15]
REQ-VAL-02	*Zeller et al. (2024)*[6]
REQ-VAL-03	*Sridhar et al. (2025)*[15]

4.2 Análise Quantitativa

O questionário de validação permaneceu disponível por um período de duas semanas e contou com a participação de 60 respondentes.. Nesta seção, é apresentada uma análise das

respostas relacionadas ao perfil profissional dos participantes, bem como da concordância em relação aos requisitos propostos.

4.2.1 Perfil Profissional

Q1. Como você encontrou esta pesquisa?

A pesquisa foi divulgada por meio de diferentes canais de comunicação com o intuito de alcançar uma amostra diversificada de profissionais. O canal com maior retorno de respostas foi o convite via *LinkedIn*, responsável por 56,7% das participações. Em seguida, destacaram-se os grupos de *WhatsApp*, com 23,3%, e listas de *e-mail*, com 13,3%. Os demais 6,7% dos participantes relataram ter conhecido a pesquisa por meio de convites diretos de colegas, redes sociais ou outras formas não especificadas.

A Figura 4.1 apresenta a distribuição percentual completa das respostas.

Q2. Há quanto tempo você trabalha com sistemas baseados em *Machine Learning*?

A maioria dos participantes (38,3%) declarou possuir entre 3 e 5 anos de experiência, evidenciando uma predominância de profissionais com familiaridade prática na área. Em seguida, 23,3% afirmaram atuar entre 1 e 2 anos, enquanto 21,7% possuem entre 6 e 10 anos de experiência.

Participantes com menos de 1 ano de atuação (8,3%) e aqueles com mais de 10 anos (8,3%) representaram as menores parcelas da amostra.

A Figura 4.2 apresenta a distribuição percentual completa das respostas.

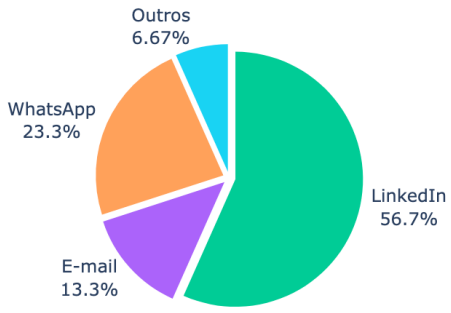

Figura 4.1: Canais de divulgação da pesquisa (Q1).

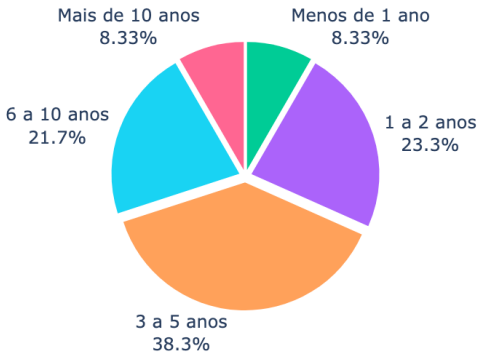

Figura 4.2: Tempo de experiência com ML (Q2).

Q3. Qual a sua principal função atual?

A maior parte dos respondentes atua como cientista de dados, representando 33,3% da amostra. As demais funções foram distribuídas de forma mais equilibrada: 15% atuam com pesquisa e academia, 15% com engenharia ou desenvolvimento de software e 15% foram agrupados na categoria "outros", já que possuiam um perfil mais diferenciado.

Funções relacionadas à gestão e liderança correspondem a 11,7% das respostas, enquanto engenheiros de ML e engenheiros de dados representaram 5% cada.

A Figura 4.3 apresenta a distribuição das funções atuais dos participantes.

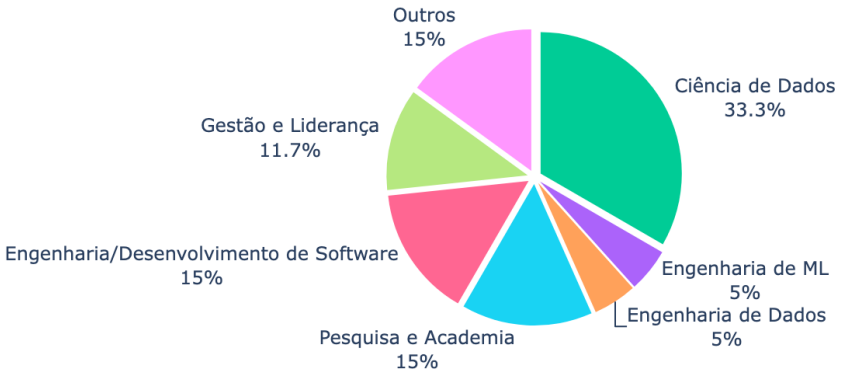

Figura 4.3: Função principal do participante (Q3).

Q4. Em quantos projetos relacionados a sistemas baseados em *Machine Learning* você já participou?

A maior parte dos participantes (43,3%) relatou ter participado de menos de cinco projetos relacionados a sistemas de *Machine Learning*. Outros 31,7% afirmaram ter atuado em entre cinco e dez projetos, enquanto 25% possuem experiência em mais de dez iniciativas. A Figura 4.4 apresenta a distribuição detalhada das respostas.

Q5. Nos projetos relacionados a sistemas baseados em *Machine Learning* nos quais você participou eram produzidos casos de garantia?

A produção de casos de garantia ainda não é uma prática amplamente disseminada. Apenas 20% dos participantes afirmaram que seus projetos envolviam esse tipo de artefato. A maioria (53,3%) declarou que os projetos não produziam casos de garantia, e 26,7% disseram não saber responder à pergunta. A Figura 4.5 ilustra essa distribuição de respostas.

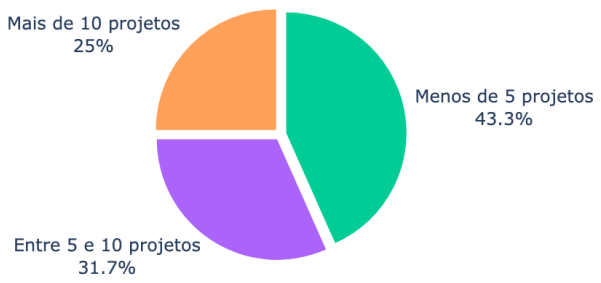

Figura 4.4: Quantidade de projetos com ML (Q4).

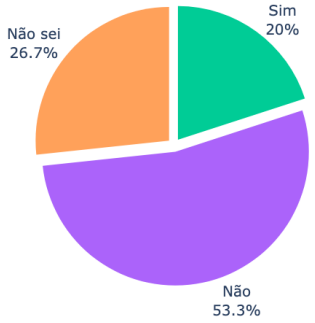

Figura 4.5: Participação em projetos com ML que produziram casos de garantia (Q5).

4.2.2 Requisitos de Escopo

Q6. Na sua opinião, é importante que sistemas de *Machine Learning* tenham uma definição formal e documentada do seu ODD, incluindo cenários típicos e críticos?

A maioria expressiva dos participantes (86,7%) demonstrou concordar com a importância de que sistemas de *Machine Learning* possuam uma definição do seu *Operational Design Domain* (ODD). Especificamente, 40 respondentes atribuíram a nota máxima (5) e outros 12 escolheram a nota 4 na escala de concordância. Apenas um participante discordou totalmente da afirmação (nota 1) e outros 7 foram neutros, escolhendo a nota 3.

A Figura 4.6 apresenta a distribuição geral das respostas, evidenciando um forte alinhamento dos especialistas em relação à necessidade de clareza sobre os limites operacionais dos sistemas de ML.

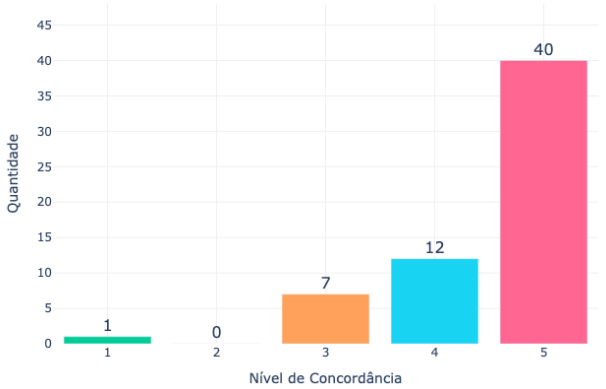

Figura 4.6: Definição formal do ODD (Q6).

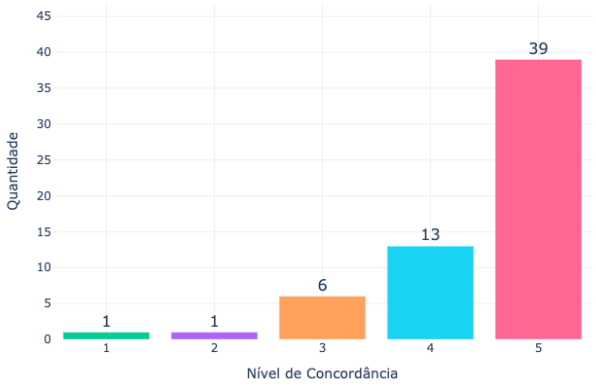

Figura 4.7: Atribuição de níveis de criticidade às condições do ODD (Q7).

Q7. Você considera útil atribuir diferentes níveis de criticidade às condições do ODD como forma de destacar quais situações exigem mais atenção do sistema?

Para essa pergunta, a maioria dos participantes (65%) atribuiu a nota máxima. Outros 13 respondentes (21,7%) marcaram o nível 4 de concordância, e apenas 2 pessoas (3,3%) atribuíram notas inferiores a 3.

A distribuição geral das respostas, apresentada na Figura 4.7, reforça o reconhecimento da relevância de mecanismos que permitam priorizar e classificar situações dentro do escopo operacional dos sistemas.

Q8. Você concorda que a definição explícita das entradas e saídas de um componente de ML ajuda a garantir maior controle sobre o comportamento do sistema?

A maioria dos participantes demonstrou forte concordância com a afirmação. Aproximadamente 91,7% dos respondentes atribuíram as notas 4 ou 5, com destaque para os 38 que marcaram o nível máximo de concordância. Apenas uma pessoa discordou totalmente, e 4 selecionaram a nota 3. A distribuição é apresentada na Figura 4.8.

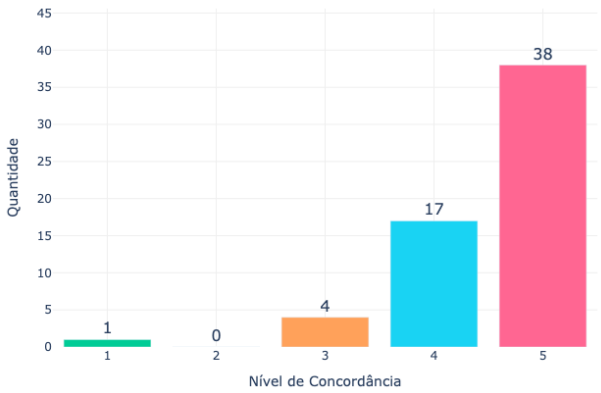

Figura 4.8: Definição explícita de entradas e saídas (Q8).

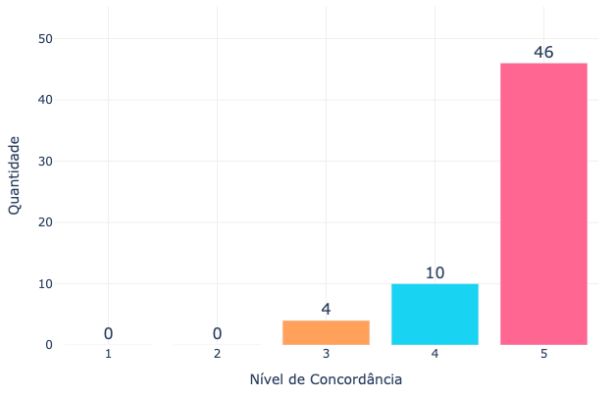

Figura 4.9: Classificação com base na criticidade de uso (Q9).

Q9. Você considera relevante classificar um sistema de ML com base na criticidade de uso?

A concordância com essa pergunta foi quase unânime, como pode ser visto na Figura 4.9. A nota máxima foi atribuída por 46 respondentes (76,7%), enquanto outros 10 marcaram nota 4. Nenhum participante discordou da proposição, e apenas 4 pessoas indicaram neutralidade.

Q10. Você considera relevante classificar um sistema de ML com base no nível de autonomia com que atua?

Aproximadamente 81,7% dos participantes atribuíram as notas 4 ou 5, sendo que 37 marcaram a nota máxima.

Esse resultado, apresentado na Figura 4.10, reforça o entendimento de que o grau de autonomia influencia diretamente o nível de risco e, portanto, deve ser um fator considerado ao definir os requisitos.

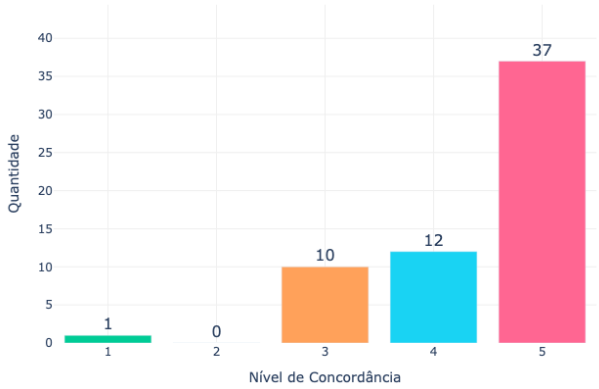

Figura 4.10: Classificação com base na autonomia (Q10).

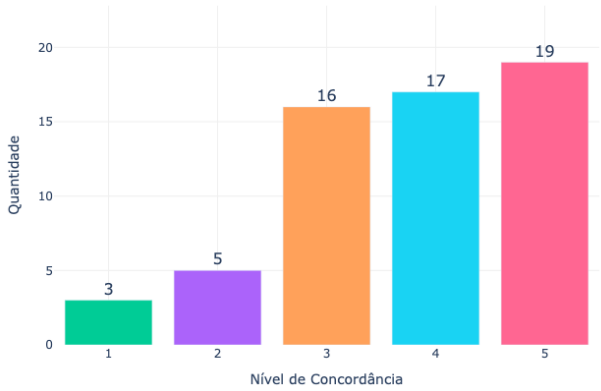

Figura 4.11: Classificação com base na complexidade do modelo (Q11).

Q11. Você considera relevante classificar um sistema de ML com base na complexidade do modelo utilizado?

As respostas para essa pergunta apresentaram maior dispersão em relação às anteriores, como pode ser observado na Figura 4.11. Embora a maioria ainda tenha atribuído notas altas (63,3% nos níveis 4 e 5), houve uma parcela considerável de participantes com opiniões mais neutras ou céticas: 16 marcaram nota 3 (26,7%), 5 marcaram 2 e 3 marcaram 1.

4.2.3 Requisitos de Segurança

Q12. Você concorda que a capacidade que um sistema baseado em ML tem para detectar situações de risco e executar ações de mitigação é uma exigência essencial para promover a segurança operacional?

A maioria dos participantes (78,3%) demonstrou forte concordância com a afirmação, sendo que 29 atribuíram a nota máxima (5) e 18 marcaram o nível 4. Outros 11 participantes selecionaram o nível 3, enquanto apenas dois marcaram níveis de discordância (1 e 2).

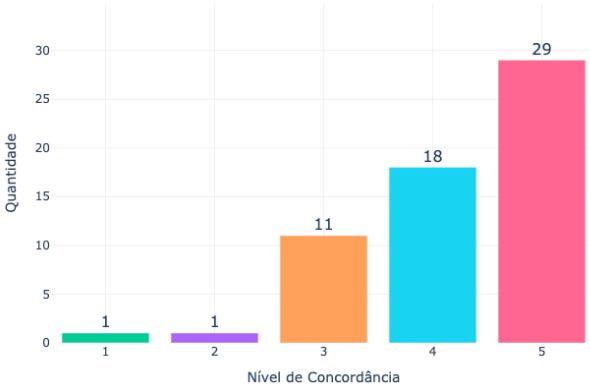

Figura 4.12: Classificação com base na autonomia (Q12).

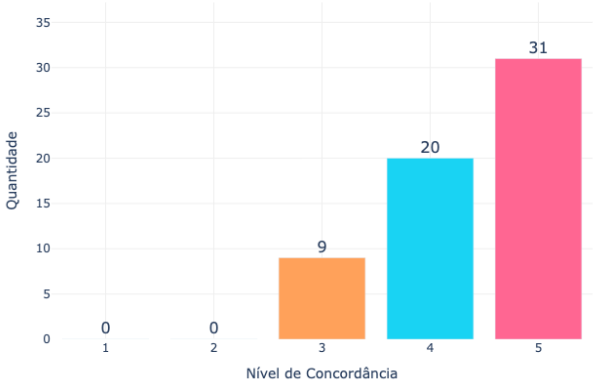

Figura 4.13: Classificação com base na complexidade do modelo (Q13).

Q13. Você concorda com a exigência de que um sistema baseado em ML entre automaticamente em um estado seguro diante de falhas internas, entradas inválidas ou condições fora do seu ODD?

A Figura 4.13 mostra uma tendência de concordância: 31 respondentes (51,7%) atribuíram a nota máxima e outros 20 (33,3%) marcaram nota 4. Não houve respostas nos níveis 1 ou 2, e apenas 9 participantes (15%) demonstraram neutralidade (nota 3).

Q14. Você concorda que o desenvolvimento de sistemas baseados em ML deve incluir medidas para evitar vieses discriminatórios nas decisões automatizadas?

A concordância com a necessidade de combater vieses discriminatórios em sistemas de *Machine Learning* foi quase unânime. Mais de 85% dos participantes (51 respondentes) atribuíram a nota máxima (5) e 7 escolheram a nota 4. Apenas 2 marcaram níveis inferiores: um participante marcou nota 1 e outro nota 3.A Figura 4.14 ilustra essa distribuição.

Q15. Você concorda que sistemas baseados em ML devem estar em conformidade com legislações de proteção de dados

Esta foi a questão com mais escolhas da nota máxima, como é possível observar na Figura 4.15. 55 participantes (91,7%) atribuíram a nota 5 e outros 3 selecionaram nota 4. Apenas dois participantes marcaram níveis inferiores (2 e 3), e não houve nenhuma resposta no nível de discordância mais acentuado.

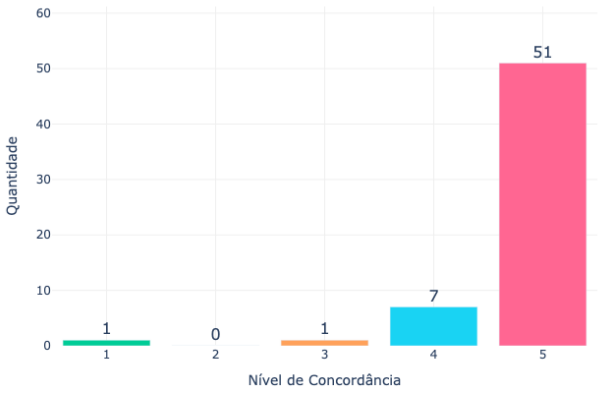

Figura 4.14: Evitar vieses discriminatórios (Q14).

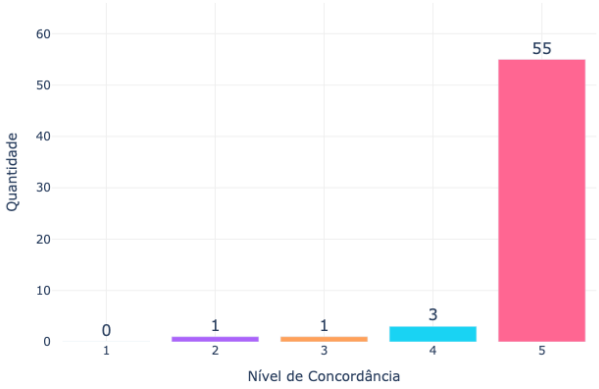

Figura 4.15: Conformidade com legislações de proteção de dados (Q15).

Q16. Você considera relevante que sistemas baseados em ML tenham mecanismos de segurança para lidar com ataques adversariais?

Como apresentado na Figura 4.16, a maioria dos participantes (80%) atribuiu a nota máxima de concordância (5) à relevância de proteger sistemas de *Machine Learning* contra ataques adversariais. Outros 9 marcaram nota 4, enquanto apenas 3 participantes atribuíram nota 3. Nenhuma resposta foi registrada nos níveis 1 ou 2.

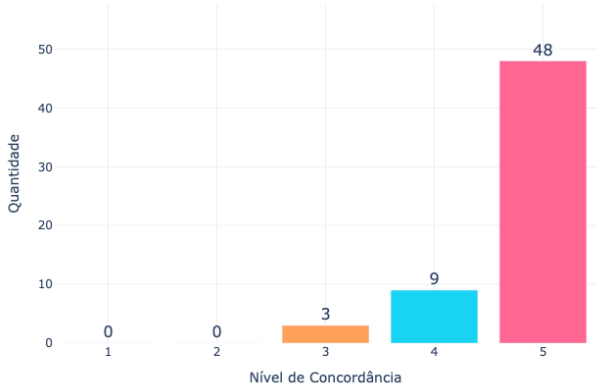

Figura 4.16: Segurança contra ataques adversariais (Q16).

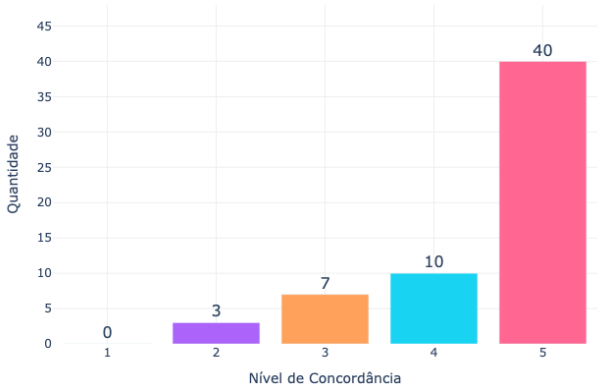

Figura 4.17: Robustez diante de situações não ideais (Q17).

Q17. Você considera importante que sistemas baseados em ML consigam manter um comportamento seguro e previsível mesmo diante de situações não ideais?

A Figura 4.17 demonstra que 40 participantes (66,7%) selecionaram o nível máximo de concordância, enquanto outros 10 marcaram nota 4. As respostas restantes se distribuíram entre os níveis 2 e 3, com apenas 3 participantes discordando parcialmente (nota 2) e nenhum

totalmente (nota 1).

4.2.4 Requisitos de Dados

Q18. Você concorda que é importante que os conjuntos de dados utilizados para treinamento, validação e testes sejam representativos do domínio real de operação do sistema?

Nesta pergunta, nota 5 foi selecionada por 51 participantes (85%), enquanto outros 6 marcaram nota 4. As notas inferiores representaram casos isolados: apenas 2 marcaram nota 3 e 1 respondeu com nota 2. A Figura 4.18 evidencia esse consenso.

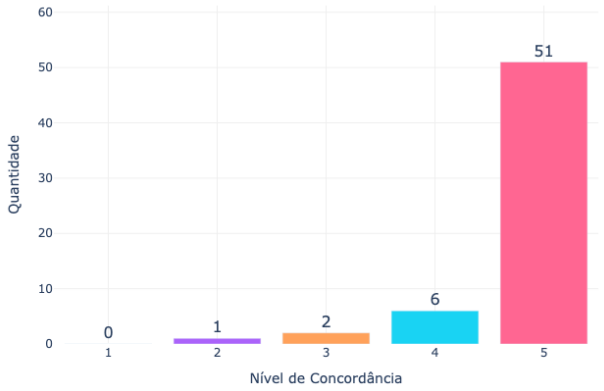

Figura 4.18: Conjunto de dados representativos do ODD (Q18).

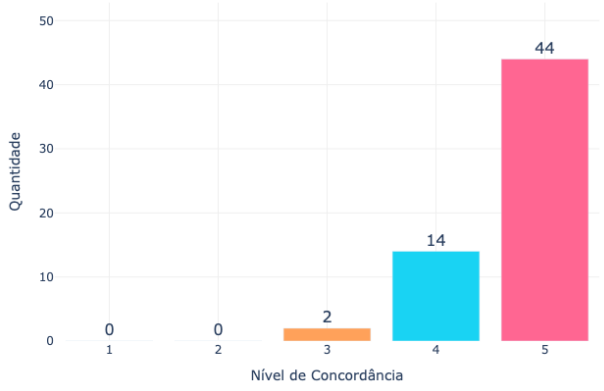

Figura 4.19: Qualidade e acurácia dos rótulos dos dados (Q19).

Q19. Você considera importante garantir a qualidade e a acurácia dos rótulos dos dados por meio de processos formais?

A maioria dos especialistas também demonstrou forte concordância com a necessidade de assegurar a qualidade dos rótulos utilizados nos dados de treinamento. 44 respondentes (73,3%) atribuíram nota máxima (5), e 14 outros marcaram nota 4. Apenas dois participantes indicaram neutralidade (nota 3), e não houve discordância. A Figura 4.19 apresenta a distribuição das respostas.

Q20. Você concorda que a rastreabilidade dos dados é importante para a confiabilidade de sistemas baseados em ML?

A Figura 4.20 ilustra concordância quase unânime quanto à importância da rastreabilidade dos dados. A nota máxima (5) foi atribuída por 43 respondentes (71,7%) e outros 13 marcaram

nota 4. Apenas 4 pessoas atribuíram nota 3, e não houve registros nos níveis 1 ou 2.

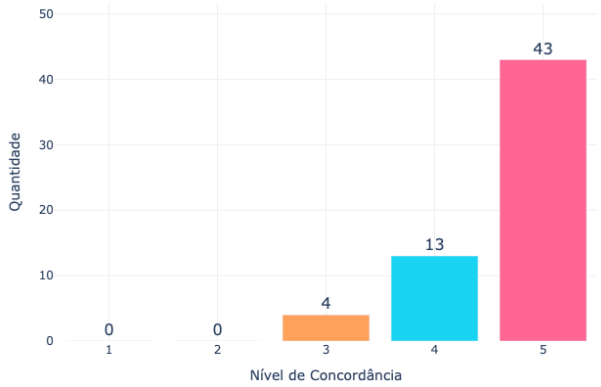

Figura 4.20: Rastreabilidade dos dados (Q20).

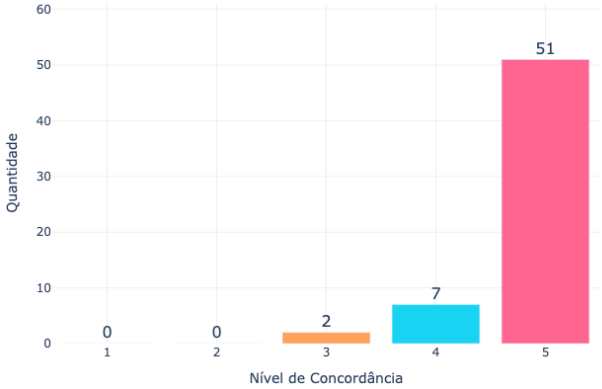

Figura 4.21: Políticas de governança dos dados (Q21).

Q21. Você concorda que é importante que existam políticas de governança de dados que incluam o monitoramento contínuo de viés e a atualização periódica dos datasets?

Uma das afirmações mais bem avaliadas do questionário está apresentada na Figura 4.21: 51 participantes (85%) atribuíram nota 5, e outros 7 marcaram nota 4. Apenas 2 pessoas indicaram neutralidade com nota 3, e nenhuma marcou níveis de discordância.

4.2.5 Requisitos de Desempenho

Q22. Você concorda que o desempenho de sistemas baseados em ML deve ser avaliado com base em métricas quantitativas previamente definidas, com valores mínimos estabelecidos antes da implantação?

Como é possível ver na Figura 4.22, 37 participantes (61,7%) atribuíram a nota máxima (5) e 15 marcaram nota 4. As notas intermediárias (2 e 3) somaram 8 respostas, sendo 3 com nota 2 e 5 com nota 3. Nenhuma resposta indicou discordância total (nota 1).

Q23. Você concorda que é relevante que, antes da implantação, um sistema baseado em ML seja comparado com soluções existentes ou legadas, de forma a garantir desempenho igual ou superior, tanto em termos de segurança quanto de eficiência?

48 participantes (80%) atribuíram a nota máxima de concordância, e outros 6 marcaram nota 4. Apenas seis participantes deram notas inferiores, com destaque para 4 que marcaram

nota 3. Dois participantes discordaram (nota 1) e nenhum marcou nota 2. A Figura 4.23 mostra esta distribuição.

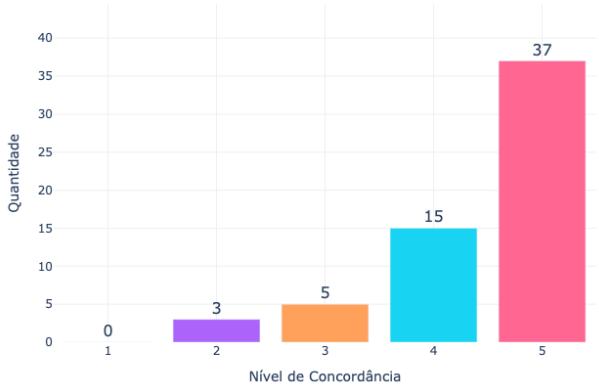

Figura 4.22: Desempenho avaliado com métricas quantitativas (Q22).

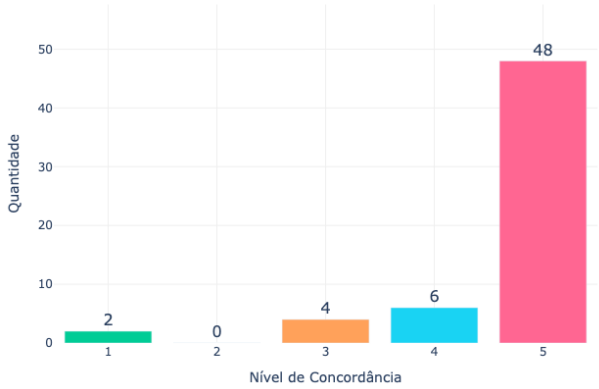

Figura 4.23: Comparação com soluções existentes ou legadas (Q23).

4.2.6 Requisitos de Validação

Q24. Você concorda que sistemas baseados em ML devam ser formalmente verificados e validados para demonstrar conformidade com os requisitos de segurança?

A questão sobre verificação e validação formais com foco em requisitos de segurança obteve alta concordância: 43 participantes (71,7%) atribuíram nota máxima (5), e 9 marcaram nota 4. Outros 8 distribuíram-se entre os níveis 3 e 2, com nenhum caso de discordância total. A Figura 4.24 mostra a distribuição das respostas.

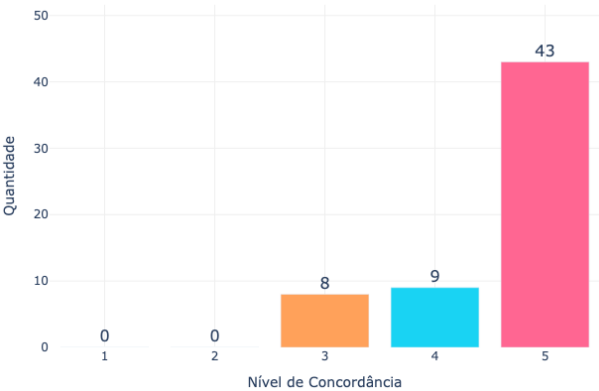

Figura 4.24: Validação de requisitos de segurança (Q24).

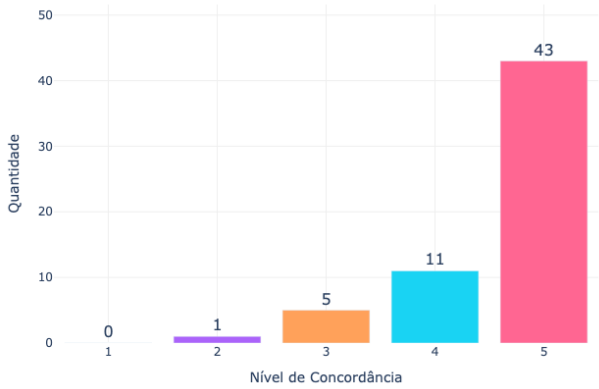

Figura 4.25: Validação de requisitos de desempenho (Q25).

Q25. Você concorda que sistemas baseados em ML devam ser formalmente verificados e validados para demonstrar conformidade com os requisitos de desempenho?

Para a validação formal voltada ao desempenho, os resultados foram muito semelhantes aos da questão anterior, como visto na Figura 4.25. A nota máxima (5) foi atribuída por 43 participantes (71,7%) e 11 marcaram nota 4. Apenas 6 pessoas marcaram nota 3, 1 marcou 2 e nenhuma marcou 1.

Q26. Você concorda que sistemas baseados em ML devam ser formalmente verificados e validados para demonstrar conformidade com os requisitos de dados?

A Figura 4.26 demonstra alto grau de aceitação. A nota máxima (5) foi marcada por 46 respondentes (76,7%) e 10 indicaram nota 4. As notas inferiores foram pontuais, com 3 respostas no nível 3 e 1 no nível 2.

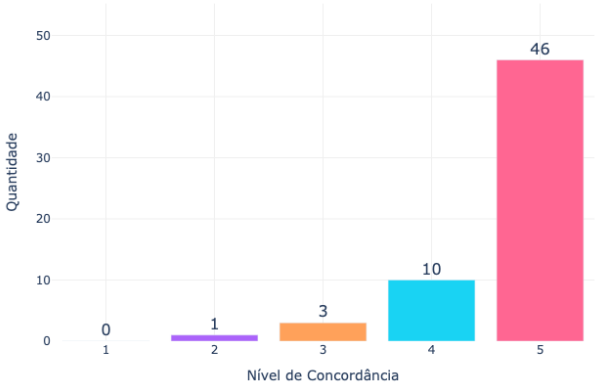

Figura 4.26: Validação de requisitos de dados (Q26).

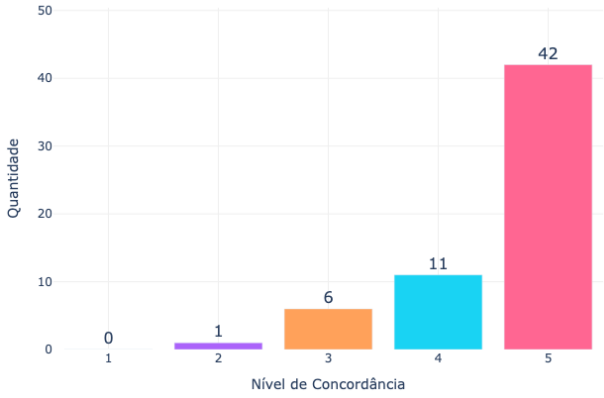

Figura 4.27: Validação de requisitos de escopo (Q27).

Q27. Você concorda que sistemas baseados em ML devam ser formalmente verificados e validados para demonstrar conformidade com os requisitos de escopo?

A última questão manteve o padrão de concordância elevado observado nas anteriores: 42 participantes (70%) atribuíram nota 5 e 11 marcaram nota 4. Apenas 7 pessoas deram notas entre 2 e 3, e nenhuma marcou nota 1. A Figura 4.27 mostra essa distribuição.

4.2.7 Síntese da Análise

Tabela 4.2: Média e desvio padrão das respostas por pergunta (Q6 a Q27), organizados por categoria

Categoria	Pergunta	Média	Desvio Padrão
Escopo	Q6	4,50	0,83
	Q7	4,47	0,87
	Q8	4,52	0,77
	Q9	4,70	0,59
	Q10	4,40	0,89
	Q11	3,73	1,15
Segurança	Q12	4,22	0,92
	Q13	4,37	0,74
	Q14	4,78	0,64
	Q15	4,87	0,50
	Q16	4,75	0,54
	Q17	4,45	0,89
Dados	Q18	4,78	0,58
	Q19	4,70	0,53
	Q20	4,65	0,61
	Q21	4,82	0,47
Desempenho	Q22	4,43	0,85
	Q23	4,63	0,88
Validação	Q24	4,58	0,72
	Q25	4,60	0,72
	Q26	4,68	0,65
	Q27	4,57	0,74

Com a análise quantitativa, foram avaliadas 22 perguntas representando os 19 requisitos propostos e, para cada pergunta, foram calculadas a média e o desvio padrão das respostas, como mostrado na Tabela 4.2. De forma geral, os dados revelam uma tendência de alta concordância

com os requisitos, onde a maioria das perguntas obteve média superior a 4,5. Requisitos relacionados à segurança operacional, governança de dados e conformidade legal foram especialmente bem avaliados.

Algumas perguntas apresentaram médias ligeiramente inferiores ou desvios padrão mais elevados, sugerindo visões variadas entre os participantes. Para investigar possíveis origens dessas variações, a próxima seção analisa os resultados segmentados por diferentes perfis profissionais.

4.3 Análise por Perfil

A Tabela 4.3 apresenta a taxa de concordância (notas 4 ou 5) por perfil de respondente para cada pergunta do questionário. A análise foi segmentada por experiência profissional (até 5 anos ou mais de 5 anos), área de atuação (profissionais de Ciência de Dados, Engenharia de Dados ou Engenharia de ML em comparação com outras áreas), e experiência com projetos que utilizaram casos de garantia.

Participantes com até 5 anos de experiência e aqueles com mais de 5 anos demonstraram níveis de concordância semelhantes, mas diferenças pontuais se destacam. Aqueles com mais de 5 anos de experiência, que representam 30% do total de respondentes, apresentaram concordância ligeiramente superior em requisitos ligados a validação e desempenho, como a comparação com soluções legadas (Q23) e a verificação formal de conformidade com requisitos de segurança, desempenho, dados e escopo (Q24–Q27), além da capacidade de manter comportamento seguro mesmo em situações não ideais (Q17). Esses dados sugerem que profissionais mais experientes valorizam fortemente práticas formais de avaliação e robustez em cenários complexos.

Por outro lado, participantes com até 5 anos de experiência demonstraram maior concordância em requisitos como classificação por complexidade do modelo (Q11), a capacidade de detectar riscos (Q12), a transição para estado seguro (Q13) e a proteção contra ataques adversariais (Q16). Isso pode indicar que esses profissionais reconhecem a importância de mecanismos fundamentais de controle e segurança.

Analisando por área de atuação, profissionais de dados e ML (cerca de 43,3% dos respondentes) apresentaram níveis de concordância significamente mais baixos em muitas das questões, em comparação com os demais participantes. As maiores diferenças foram observadas em requisitos de escopo e segurança, como a definição formal do ODD (Q6), a classificação por criticidade do ODD (Q7), o nível de autonomia do sistema (Q10) e a complexidade do modelo

(Q11), além de aspectos relacionados à segurança operacional (Q12, Q13, Q17). Essa tendência se estendeu também a um requisito de desempenho, sobre o uso de métricas quantitativas previamente definidas (Q22). Essa diferença pode sugerir que profissionais da área tendem a ser mais críticos ou céticos em relação à aplicabilidade desses requisitos.

Tabela 4.3: Taxa de concordância (notas 4 ou 5) por perfil de respondente (%)

Questão	Até 5 anos de atuação	Mais de 5 anos de atuação	Área de Dados/ML	Outras áreas	Com casos de garantia	Sem casos ou não sabe
Q06	88.1%	83.3%	80.8%	91.2%	100%	83.3%
Q07	85.7%	88.9%	80.8%	91.2%	100%	83.3%
Q08	90.5%	94.4%	92.3%	91.2%	91.7%	91.7%
Q09	95.2%	88.9%	88.5%	97.1%	100%	91.7%
Q10	81.0%	83.3%	69.2%	91.2%	83.3%	81.2%
Q11	61.9%	55.6%	50.0%	67.6%	66.7%	58.3%
Q12	81.0%	72.2%	61.5%	91.2%	91.7%	75.0%
Q13	88.1%	77.8%	76.9%	91.2%	100%	81.2%
Q14	95.2%	100%	96.2%	97.1%	100%	95.8%
Q15	95.2%	100%	92.3%	100.0%	91.7%	97.9%
Q16	97.6%	88.9%	92.3%	97.1%	91.7%	95.8%
Q17	81.0%	88.9%	69.2%	94.1%	100%	79.2%
Q18	95.2%	94.4%	92.3%	97.1%	100%	93.8%
Q19	95.2%	100%	96.2%	97.1%	91.7%	97.9%
Q20	95.2%	88.9%	92.3%	94.1%	91.7%	93.8%
Q21	95.2%	100%	96.2%	97.1%	100%	95.8%
Q22	88.1%	83.3%	73.1%	97.1%	100%	83.3%
Q23	85.7%	100%	88.5%	91.2%	100%	87.5%
Q24	83.3%	94.4%	84.6%	88.2%	91.7%	85.4%
Q25	85.7%	100%	88.5%	91.2%	100%	87.5%
Q26	90.5%	100%	92.3%	94.1%	100%	91.7%
Q27	88.1%	88.9%	88.5%	88.2%	100%	85.4%

Quanto à participação em projetos com casos de garantia, que representam 20% dos res-

pondentes, foi registrado 100% de concordância em 13 das 22 questões. Esses participantes apresentaram taxas superiores especialmente em requisitos como definição formal e classificação do ODD (Q6 e Q7), transição automática para estado seguro (Q13), robustez diante de situações não ideais (Q17), uso de métricas quantitativas (Q22), e verificação formal de conformidade com requisitos de desempenho (Q25) e escopo (Q27). Essa uniformidade nas respostas sugere que profissionais familiarizados com casos de garantias têm maior alinhamento com práticas formais e preventivas, valorizando fortemente requisitos voltados à confiabilidade.

Essas diferenças entre os perfis ajudam a explicar os resultados observados na seção anterior, em que questões como Q10, Q11, Q12, Q13 e Q22 apresentaram médias mais baixas ou maior dispersão nas respostas. Foi possível observar que essas perguntas também estão entre aquelas em que houve maior contraste entre os grupos analisados, indicando que a diversidade de perfis profissionais pode gerar interpretações distintas sobre a importância ou aplicabilidade de certos requisitos.

4.4 Análise Qualitativa

A análise qualitativa foi conduzida com base nas respostas da última seção do questionário, que teve como objetivo capturar percepções complementares, críticas e sugestões dos especialistas. Dessa forma, foi possível identificar lacunas, ambiguidades e pontos de melhoria.

4.4.1 Comentários Finais

Q28. Você gostaria de sugerir algum requisito adicional que, na sua opinião, é importante para garantir a confiabilidade de sistemas baseados em ML? Ou gostaria de fazer algum comentário sobre os requisitos apresentados?

Dos 60 respondentes, 20 deixaram comentários. A Figura 4.28 apresenta a distribuição dos perfis desses participantes. Observa-se que há um equilíbrio entre respondentes com até 5 anos de experiência (11 pessoas) e com mais de 5 anos (9 pessoas). Além disso, 9 participantes atuam diretamente em áreas como Ciência de Dados, Engenharia de Dados ou Engenharia de ML, enquanto os demais têm formações ou funções diversas. A maioria dos respondentes participou de pelo menos 5 projetos com ML, sendo que 9 relataram envolvimento em mais de 10 projetos. Apesar disso, apenas 4 indicaram que tais projetos contavam com casos de garantia.

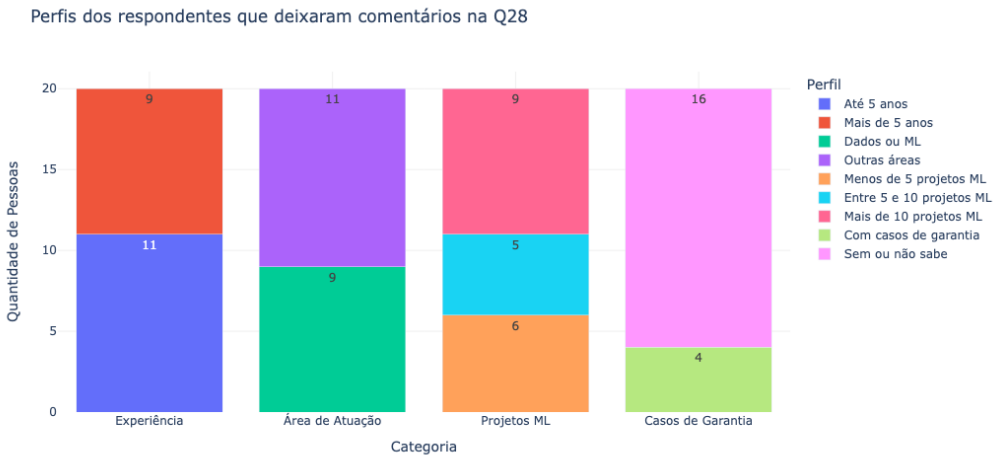

Figura 4.28: Distribuição do perfil dos participantes que deixaram comentários (Q28).

As contribuições foram analisadas e agrupadas nos seguintes eixos:

- **Validação e Verificação**

  Diversos participantes comentaram sobre os desafios da V&V formais em sistemas baseados em ML. Um deles apontou: *"Respondi com opção neutra [...] pois não tenho certeza se realmente é viável essa verificação formal."*

  Outros reforçaram a importância de múltiplas etapas de validação, incluindo dados, modelo e ambiente de produção.

- **Governança e Rastreabilidade**

  Alguns comentários abordaram a necessidade de versionamento de código e rastreabilidade de execuções. Outros reforçaram práticas como monitoramento de parâmetros, alarmes automáticos e mecanismos de pré-processamento robustos.

- **Contextualização dos requisitos**

  Houve respostas relacionadas ao uso dos requisitos conforme a fase do projeto ou o nível de maturidade da organização, como por exemplo: *"Projetos de ML começam com provas de conceito. Enquanto não há cliente, não vale a pena investir em todos os requisitos."*

  Além disso, foi destacada que nem todos os sistemas tem o mesmo grau de criticidade: *"Existem sistemas e modelos que sequer sabem o que significa risco [...] outro sistema pode ser responsável por isso."*

- **Uso de Métricas**

  Foi sugerido que as métricas de avaliação não se restringisem ao desempenho do modelo,

mas que métricas de impacto no negócio também fossem incluídas: *"Pensar em métricas do negócio que o modelo irá impactar [...] definir baseline e valor estimado para elas também."*

Outros participantes mencionaram métricas específicas para avaliar distribuições de dados, como testes de *chi-square* [12], KS1/KS2 [13] e acompanhamento pós-implantação.

Essa análise foi de grande importância para a validação, pois trouxe não apenas críticas construtivas, mas também novas dimensões que podem ser incorporadas ao conjunto final de requisitos.

4.5 Revisão dos Requisitos

Com base nas análises da pesquisa, alguns requisitos apresentaram sinais de que podem se beneficiar de revisão. Foram utilizados os comentários que indicaram críticas ou sugestões de melhoria como critério para realizar essa seleção.

4.5.1 Requisitos Revisados

A seguir estão listados os requisitos selecionados para revisão e suas respectivas justificativas. O conjunto final de requisitos é apresentado integralmente no Apêndice B.

REQ-ESC-03

Comentários indicaram que nem todos os sistemas exigem classificação por complexidade ou autonomia, dependendo do contexto de uso. Isso se refletiu nas perguntas relacionadas à este requisito (Q10 e Q11), que tiveram, respectivamente, médias 4,40 e 3,73, além de desvios padrão 0,89 e 1,15. Esses resultados indicam menor concordância e consenso entre os participantes.

REQ-SEG-01

Um comentário destacou que nem todo sistema ML está apto ou precisa atuar diretamente em situações de risco. Além disso, a pergunta relacionada (Q12) apresentou média 4,22 e desvio padrão 0,92.

[12] Testes de χ^2 (chi-square) avaliam se há independência entre variáveis categóricas ou se uma distribuição observada difere significativamente da esperada.

[13] KS1 e KS2 referem-se a testes utilizados para comparar distribuições acumuladas, sendo úteis para detectar mudanças entre conjuntos de dados.

REQ-DES-01

Com média 4,43 e desvio padrão 0,85, este requisito também recebeu sugestões qualitativas de novos tipos de métricas a serem incluídas.

REQ-VAL-01, REQ-VAL-02 e REQ-VAL-03

Apesar das médias altas, vários comentários indicaram dúvidas quanto à viabilidade da validação formal, sugerindo que os requisitos podem ser reformulados ou contextualizados por tipo de aplicação.

4.6 Considerações Finais

Diante dos resultados obtidos, torna-se evidente a necessidade de evoluir para práticas que ofereçam maior rastreabilidade, transparência e justificativa formal da confiabilidade em sistemas de Machine Learning. A identificação de requisitos estruturados e validados por profissionais da área demonstra que há insumos suficientes para avançar na elaboração de uma *framework*[14] de geração de casos de garantia específicos para esse tipo de sistema. Essa transição é fundamental para atender às exigências crescentes de segurança e auditabilidade, especialmente em domínios críticos onde falhas podem ter consequências severas.

[14]O termo *framework* refere-se a uma estrutura ou arcabouço reutilizável de componentes de software que oferece suporte à construção e organização de aplicações.

Capítulo 5

Conclusão e trabalhos futuros

Este trabalho descreve o processo de extração e revisão até alcançar um conjunto de requisitos de confiabilidade em sistemas baseados em aprendizado de máquina. O objetivo proposto foi alcançado por meio de uma revisão da literatura e da pesquisa com especialistas.

Os resultados indicam que há amplo consenso no mercado sobre a importância de práticas voltadas à governança de dados, rastreabilidade, imparcialidade, validação contínua e conformidade legal. Ao mesmo tempo, revelam pontos de maiores discussões, como a viabilidade da validação formal em diferentes contextos.

O estudo resultou em um conjunto de requisitos, sensível às diferentes realidades enfrentadas no desenvolvimento de soluções com ML.

5.1 Contribuições

Este trabalho oferece contribuições significativas para a área de engenharia de requisitos aplicada a sistemas de ML. Os requisitos podem auxiliar equipes técnicas na formalização de critérios de qualidade, segurança e conformidade desde as fases iniciais de desenvolvimento, promovendo maior previsibilidade e confiabilidade no ciclo de vida dos modelos. Além disso, as contribuições deste estudo também pavimentam o caminho para a construção de casos de garantia adaptados a sistemas de ML, oferecendo uma base inicial para justificar formalmente sua confiabilidade.

5.2 Limitações

Apesar dos resultados expressivos, este estudo possui algumas limitações, como:

- A amostra de profissionais utilizada na validação pode não refletir todos os setores e domínios críticos onde ML é aplicado.
- A coleta de dados via redes sociais (e.g. *LinkedIn*, *WhatsApp*) tende a atrair perfis semelhantes, o que pode ter limitado a variedade da amostra e enviesado os resultados.
- A análise qualitativa se baseou em comentários opcionais. Dessa forma, não foi possível aprofundar os resultados em pontos específicos que poderiam ter sido mais explorados com outros métodos, como entrevistas ou dinâmicas em grupo.
- Este trabalho não avaliou, na prática, a aplicação dos requisitos em um sistema real de ML, o que limita a comprovação de sua viabilidade em ambientes produtivos.
- O processo de seleção dos artigos não seguiu uma revisão sistemática da literatura, devido a restrições de tempo, o que pode ter resultado na exclusão de estudos relevantes. Além disso, por se tratar de uma área de pesquisa em rápida evolução, novos artigos significativos podem ter sido publicados após a conclusão desta análise.

5.3 Trabalhos futuros

Este trabalho abre espaço para diversas direções de estudo e aplicação, como:

- Utilizar os requisitos propostos em sistemas em desenvolvimento ou já implantados.
- Testar os requisitos em contextos como veículos autônomos, sistemas médicos ou financeiros para permitir a adaptação às particularidades de cada ambiente.
- Investigar o uso dos requisitos para a construção de argumentos verificáveis sobre a confiabilidade do sistema.

Apêndice A

Formulário aplicado na pesquisa

Este apêndice apresenta o conteúdo completo do formulário aplicado para coleta de percepções sobre requisitos de confiabilidade em sistemas baseados em *Machine Learning* (ML). O questionário foi estruturado em seis seções temáticas e aplicado via *Google Forms*.

As respostas foram registradas majoritariamente por meio de escalas de concordância, relevância ou importância de 1 a 5, onde 1 significava total discordância e 5, total concordância.

A.1 Termo de Consentimento Livre e Esclarecido (TCLE) em concordância com a pesquisa

Público-alvo

Profissionais com experiência em desenvolvimento, validação, operação ou gestão de sistemas que utilizam Machine Learning.

Coleta e Uso de Dados

Todas as informações fornecidas neste questionário serão **coletadas de forma anônima, mantidas em sigilo e armazenadas em ambiente seguro**. Os dados serão utilizados **exclusivamente para fins científicos**.

Voluntariedade e Direitos

A participação é **completamente voluntária**. Você pode **interromper o preenchimento a qualquer momento**, sem precisar justificar. Caso decida não enviar suas respostas, basta

sair do formulário ou utilizar a opção "Limpar formulário" antes de finalizar. **Após o envio**, como as respostas são anônimas, **não será possível excluir os dados**, pois não haverá como identificar suas informações individuais.

Privacidade e Proteção de Dados

Nenhuma informação que permita a sua identificação será solicitada ou registrada. Todas as respostas serão utilizadas **exclusivamente para fins de pesquisa acadêmica**, de acordo com os princípios éticos estabelecidos para pesquisas com seres humanos.

Acesso aos Resultados

Se desejar, você poderá solicitar uma cópia do relatório final com os resultados desta pesquisa, entrando em contato com o pesquisador responsável pelo e-mail informado abaixo.

Contato

Para dúvidas ou mais informações sobre esta pesquisa: **alas3@cin.ufpe.br**

Declaração de Consentimento

Ao prosseguir e enviar suas respostas, você declara que

- Leu e compreendeu as informações apresentadas acima

- Concorda, de forma livre e espontânea, em participar desta pesquisa.

- Autoriza a coleta e o uso das informações fornecidas para os fins descritos.

Declaro estar de acordo com os termos de participação nesta pesquisa.
( ) Sim ( ) Não

A.2 Perfil Profissional

1. Como você encontrou esta pesquisa?
( ) Post no LinkedIn
( ) Convite via LinkedIn
( ) Post em lista de e-mail
( ) Convite via e-mail

( ) Grupos de WhatsApp

( ) Outro: _____

2. Há quanto tempo você trabalha com Sistemas baseados em Machine Learning?

( ) Menos de 1 ano

( ) 1 a 2 anos

( ) 3 a 5 anos

( ) 6 a 10 anos

( ) Mais de 10 anos

3. Qual a sua principal função atual?

[Resposta aberta]

4. Em quantos projetos relacionados a Sistemas baseados em Machine Learning você já participou?

( ) Menos de 5 ( ) Entre 5 e 10 ( ) Mais de 10

5. Nos projetos relacionados a Sistemas baseados em Machine Learning nos quais você participou eram produzidos Assurance Cases (casos de garantia)?

( ) Sim ( ) Não ( ) Não sei

A.3 Requisitos de Escopo

6. Na sua opinião, é importante que sistemas de Machine Learning tenham uma definição formal e documentada do seu ODD (Operational Design Domain — conjunto de condições ambientais, operacionais e contextuais em que o sistema pode funcionar com segurança), incluindo cenários típicos e críticos?

(1) Não importante — (5) Muito importante

7. Você considera útil atribuir diferentes níveis de criticidade às condições do ODD como forma de destacar quais situações exigem mais atenção do sistema?

(1) Sem utilidade — (5) Muito útil

8. Você concorda que a definição explícita das entradas e saídas de um componente de ML ajuda a garantir maior controle sobre o comportamento do sistema?

(1) Discordo totalmente — (5) Concordo totalmente

9. Você considera relevante classificar um sistema de ML com base na criticidade de uso (por exemplo, se envolve riscos à segurança, impacto financeiro ou reputacional)?
(1) Irrelevante — (5) Muito relevante

10. Você considera relevante classificar um sistema de ML com base no nível de autonomia com que atua (por exemplo, se há supervisão humana ou decisões automáticas)?
(1) Irrelevante — (5) Muito relevante

11. Você considera relevante classificar um sistema de ML com base na complexidade do modelo utilizado (por exemplo, modelos simples como regressão logística versus modelos complexos como redes neurais profundas)?
(1) Irrelevante — (5) Muito relevante

A.4 Requisitos de Segurança

12. Você concorda que a capacidade que um sistema baseado em ML tem para detectar situações de risco e executar ações de mitigação é uma exigência essencial para promover a segurança operacional?
(1) Discordo totalmente — (5) Concordo totalmente

13. Você concorda com a exigência de que um sistema baseado em ML entre automaticamente em um estado seguro diante de falhas internas, entradas inválidas ou condições fora do seu ODD (Operational Design Domain - conjunto de condições nas quais o sistema foi projetado para operar com segurança)?
(1) Discordo totalmente — (5) Concordo totalmente

14. Você concorda que o desenvolvimento de sistemas baseados em ML deve incluir medidas para evitar vieses discriminatórios nas decisões automatizadas?
(1) Discordo totalmente — (5) Concordo totalmente

15. Você concorda que sistemas baseados em ML devem estar em conformidade com legislações de proteção de dados (como LGPD e GDPR), incluindo, por exemplo, mecanismos de consentimento e anonimização?

(1) Discordo totalmente — (5) Concordo totalmente

16. Você considera relevante que sistemas baseados em ML tenham mecanismos de segurança para lidar com ataques adversariais, como o monitoramento de entradas suspeitas e a implementação de defesas contra manipulações maliciosas?
(1) Irrelevante — (5) Muito relevante

17. Você considera importante que sistemas baseados em ML consigam manter um comportamento seguro e previsível mesmo diante de situações não ideais, como ruído nas entradas, condições inesperadas ou dados fora do domínio de treinamento?
(1) Pouco importante — (5) Muito importante

A.5 Requisitos de Dados

18. Você concorda que é importante que os conjuntos de dados (datasets) utilizados para treinamento, validação e testes sejam representativos do domínio real de operação do sistema?
(1) Discordo totalmente — (5) Concordo totalmente

19. Você considera importante garantir a qualidade e a acurácia dos rótulos dos dados por meio de processos formais, como revisão por especialistas ou validação automatizada?
(1) Discordo totalmente — (5) Concordo totalmente

20. Você concorda que a rastreabilidade dos dados (incluindo informações como origem, versão e histórico de transformações) é importante para a confiabilidade de sistemas baseados em ML?
(1) Discordo totalmente — (5) Concordo totalmente

21. Você concorda que é importante que existam políticas de governança de dados que incluam o monitoramento contínuo de viés (como *data drift* e *concept drift*) e a atualização periódica dos *datasets*?
(1) Discordo totalmente — (5) Concordo totalmente

A.6 Requisitos de Desempenho

22. Você concorda que o desempenho de sistemas baseados em ML deve ser avaliado com base em métricas quantitativas (ex.: precisão, *recall*, F1-score) previamente definidas, com valores mínimos estabelecidos antes da implantação?
(1) Discordo totalmente — (5) Concordo totalmente

23. Você concorda que é relevante que, antes da implantação, um sistema baseado em ML seja comparado com soluções existentes ou legadas, de forma a garantir desempenho igual ou superior, tanto em termos de segurança quanto de eficiência?
(1) Discordo totalmente — (5) Concordo totalmente

A.7 Requisitos de Validação

24. Você concorda que sistemas baseados em ML devam ser formalmente verificados e validados para demonstrar conformidade com os requisitos de segurança?
(1) Discordo totalmente — (5) Concordo totalmente

25. Você concorda que sistemas baseados em ML devam ser formalmente verificados e validados para demonstrar conformidade com os requisitos de desempenho?
(1) Discordo totalmente — (5) Concordo totalmente

26. Você concorda que sistemas baseados em ML devam ser formalmente verificados e validados para demonstrar conformidade com os requisitos de dados?
(1) Discordo totalmente — (5) Concordo totalmente

27. Você concorda que sistemas baseados em ML devam ser formalmente verificados e validados para demonstrar conformidade com os requisitos de escopo?
(1) Discordo totalmente — (5) Concordo totalmente

A.8 Comentário Aberto

28. Você gostaria de sugerir algum requisito adicional que, na sua opinião, é importante para garantir a confiabilidade de sistemas baseados em ML? Ou gostaria de fazer algum comentário

sobre os requisitos apresentados?

[Resposta aberta]

Apêndice B

Conjunto final de Requisitos

Este apêndice apresenta a consolidação dos achados do trabalho em um conjunto refinado de requisitos de confiabilidade para sistemas baseados em *Machine Learning*.

B.0.1 Requisitos de Escopo

REQ-ESC-01 - Definição do Contexto Operacional

O sistema deve possuir uma definição formal e documentada do seu *Operational Design Domain* (ODD), incluindo:

- Condições ambientais, geográficas, temporais e operacionais válidas para sua execução segura.
- Exemplos de cenários operacionais típicos e críticos (ex.: rodovias urbanas, ambientes internos com baixa luminosidade).
- Atribuição de pesos de criticidade para cada condição do ODD, com o objetivo de orientar o esforço de teste e validação.

REQ-ESC-02 - Especificação das Interfaces e Limites do Sistema ML

O sistema deve definir claramente os limites funcionais do componente de *Machine Learning*, incluindo:

- Entradas esperadas: formatos, faixas válidas, taxas de chegada e restrições temporais (ex.: tempo máximo de resposta).
- Saídas geradas: formatos, escalas e intervalos esperados.
- Contratos formais de interação com outros componentes do sistema global.

REQ-ESC-03 - Classificação Contextual do Sistema ML

Sempre que aplicável, o sistema deve ser classificado com base em características relevantes para sua confiabilidade, considerando:

- Criticidade do contexto de uso, considerando o impacto potencial de falhas (ex.: segurança de pessoas, impacto financeiro, reputacional).
- Nível de autonomia do sistema em cada fase (treinamento, inferência, decisão final), especificando o grau de supervisão humana esperado.
- Complexidade do modelo, categorizando-o como simples (ex.: regressão logística) ou complexo (ex.: redes neurais profundas).

A aplicação desse requisito deve considerar o contexto específico da solução e o papel desempenhado pelo componente de ML, reconhecendo que nem todos os sistemas exigem detalhamento de todos esses critérios.

B.0.2 Requisitos de Segurança

REQ-SEG-01 - Detecção e Resposta a Cenários Críticos

Nos casos em que o sistema de ML for responsável por decisões em contextos sensíveis, ele deve ser capaz de detectar situações de risco com alta confiabilidade e realizar ações de mitigação que mantenham a segurança dos usuários e do ambiente.

REQ-SEG-02 - Implementação de *Fail-Safe*

Em caso de falhas internas, entradas inválidas ou condições fora do ODD, o sistema deve adotar automaticamente um comportamento *fail-safe*, entrando em um estado seguro previamente definido.

REQ-SEG-03 - Garantia de *Fairness*

O sistema deve evitar vieses discriminatórios, assegurando a *fairness* nas decisões automatizadas.

REQ-SEG-04 - Garantia de Privacidade

O sistema deve cumprir com legislações de proteção de dados (ex.: LGPD, GDPR), incluindo:

- Mecanismos de consentimento para coleta e uso de dados pessoais.
- Políticas de anonimização ou pseudonimização.

REQ-SEG-05 - Monitoramento de Ameaças

O sistema deve possuir mecanismos de monitoramento contínuo para:

- Detecção de padrões de entrada suspeitos.
- Identificação de tentativas de ataques adversariais.
- Geração de alertas de segurança.

REQ-SEG-06 – Robustez a Perturbações e Condições Fora do Domínio

O sistema deve manter comportamento seguro e consistente mesmo quando exposto a situações não ideais, como:

- Ruídos ou perturbações nos dados de entrada (ex.: *motion blur*, oclusões, ruído de sensores, variações extremas de iluminação).

- Entradas inesperadas ou fora do domínio conhecido de operação (fora do ODD).

Mesmo diante de incertezas, o sistema não deve tomar ações perigosas ou instáveis.

B.0.3 Requisitos de Dados

REQ-DAD-01 - Representatividade dos Dados

Os conjuntos de dados usados para treinamento, validação e testes devem ser representativos do domínio operacional definido, contemplando diversidade de situações e cenários.

REQ-DAD-02 - Validação de Rótulos (*Labels*)

Deve-se estabelecer e documentar processos formais para garantir a acurácia dos rótulos, incluindo:

- Revisão manual por especialistas.
- Validação cruzada automatizada.
- Definição de critérios mínimos de qualidade de rotulagem.

REQ-DAD-03 - Qualidade e Rastreabilidade dos Dados

Os dados utilizados devem ser completos, corretos e rastreáveis, com documentação de origem, versão e histórico de transformações.

REQ-DAD-04 - Gerenciamento de Dados e Controle de Viés

Devem existir políticas de governança de dados para:

- Monitoramento contínuo de viés (*data drift* e *concept drift*).
- Atualização periódica dos datasets de treinamento.
- Garantia de integridade e representatividade ao longo do ciclo de vida.

B.0.4 Requisitos de Desempenho

REQ-DES-01 - Métricas Quantitativas de Performance

O desempenho do modelo de ML deve ser avaliado com base em métricas previamente definidas (ex.: precisão, *recall*, *F1-score*, mAP), com valores mínimos aceitáveis antes da implantação.

Sempre que aplicável, também devem ser consideradas métricas complementares voltadas ao acompanhamento pós-implantação (ex.: testes de distribuição como *chi-square*, KS1/KS2) que auxiliem no monitoramento contínuo da performance do sistema.

REQ-DES-02 - Consistência em Condições Reais

O sistema deve manter um desempenho consistente frente às variações normais de entrada esperadas no ambiente operacional.

REQ-DES-03 - Comparação com Sistemas de Referência

Antes da implantação, o sistema deve demonstrar desempenho igual ou superior a soluções existentes ou legadas, considerando tanto aspectos de segurança quanto de eficiência.

B.0.5 Requisitos de Validação

REQ-VAL-01 - Validação Formal de Conformidade

Sempre que viável, deve ser realizada validação formal demonstrando que os requisitos de *Machine Learning* atendem aos requisitos de sistema de mais alto nível, especialmente em termos de segurança. A necessidade de formalização pode variar de acordo com o grau de criticidade e o contexto de uso.

REQ-VAL-02 - Cobertura de Cenários Operacionais e de Falha

Os testes de validação devem abranger tanto os cenários operacionais normais quanto as condições de falha conhecidas e possíveis. A abrangência da validação deve considerar os riscos

associados ao uso pretendido do sistema.

REQ-VAL-03 - Validação Contínua ao Longo do Ciclo de Vida

Toda nova versão do modelo, alteração de dados ou modificação de código relevante deve passar por um processo de revalidação antes da liberação para produção. O grau de formalidade pode variar conforme o ambiente de uso (ex.: protótipos, produção restrita, ambientes críticos).

Referências Bibliográficas

[1] M. Rana and M. Bhushan, "Machine learning and deep learning approach for medical image analysis: diagnosis to detection," *Multimedia Tools and Applications*, vol. 82, pp. 26731–26769, 2023.

[2] M. Obthong, N. Tantisantiwong, W. Jeamwatthanachai, and G. Wills, "A survey on machine learning for stock price prediction: algorithms and techniques," in *Proceedings of the 13th International Joint Conference on Knowledge Discovery, Knowledge Engineering and Knowledge Management*, pp. 168–175, 2021.

[3] H. Villamizar, T. Escovedo, and M. Kalinowski, "Requirements engineering for machine learning: A systematic mapping study," in *47th Euromicro Conference on Software Engineering and Advanced Applications (SEAA)*, pp. 29–36, IEEE, 2021.

[4] R. Hawkins, C. Paterson, C. Picardi, Y. Jia, R. Calinescu, and I. Habli, "Guidance on the assurance of machine learning in autonomous systems (amlas)," tech. rep., Assuring Autonomy International Programme (AAIP), University of York, February 2021.

[5] K. Radlak, M. Szczepankiewicz, T. Jones, and P. Serwa, "Organization of machine learning based product development as per iso 26262 and iso/pas 21448," in *2020 IEEE 25th Pacific Rim International Symposium on Dependable Computing (PRDC)*, pp. 110–119, 2020.

[6] M. Zeller, T. Waschulzik, C. Carlan, M. Serahlazau, C. Bahlmann, Z. Wu, S. Spieckermann, D. Krompass, S. Geerkens, C. Sieberichs, K. Kirchheim, B. K. Özen, and L. Diez Robles, "Continuous development and safety assurance pipeline for ml-based systems in the railway domain," in *Computer Safety, Reliability, and Security. SAFECOMP 2024 Workshops*, (Cham), pp. 446–459, Springer Nature Switzerland, 2024.

[7] J. Rushby, X. Xu, M. Rangarajan, and T. L. Weaver, "Understanding and evaluating assurance cases," Contractor Report NASA/CR-2015-218802, NASA Langley Research Center, Hampton, VA, Sept. 2015. Contract No. NNL13AC555T.

[8] B. Nuseibeh and S. Easterbrook, "Requirements engineering: a roadmap," in *Proceedings of the Conference on The Future of Software Engineering*, ICSE '00, (New York, NY, USA), p. 35–46, Association for Computing Machinery, 2000.

[9] A. Vogelsang and M. Borg, "Requirements engineering for machine learning: Perspectives from data scientists," in *2019 IEEE 27th International Requirements Engineering Conference Workshops (REW)*, pp. 245–251, 2019.

[10] Z. Pei, L. Liu, C. Wang, and J. Wang, "Requirements engineering for machine learning: A review and reflection," in *2022 IEEE 30th International Requirements Engineering Conference Workshops (REW)*, pp. 166–175, 2022.

[11] F. Tambon, G. Laberge, L. An, A. Nikanjam, P. S. Nouwou Mindom, Y. Pequignot, F. Khomh, G. Antoniol, E. Merlo, and F. Laviolette, "How to certify machine learning based safety-critical systems? a systematic literature review," *Automated Software Engineering*, vol. 29, no. 38, 2022.

[12] L. Caminski, "Iso 26262: segurança funcional no desenvolvimento de sistemas automotivos," trabalho de conclusão de curso (especialização em sistemas embarcados para a indústria automotiva), Universidade Tecnológica Federal do Paraná, Curitiba, 2018.

[13] M. Borg, J. Henriksson, K. Socha, O. Lennartsson, E. S. Lönegren, T. Bui, P. Tomaszewski, S. R. Sathyamoorthy, S. Brink, and M. H. Moghadam, "Ergo, smirk is safe: a safety case for a machine learning component in a pedestrian automatic emergency brake system," *Software Quality Journal*, 2023.

[14] I. Dodd and I. Habli, "Safety certification of airborne software: An empirical study," *Reliability Engineering System Safety*, vol. 98, no. 1, pp. 7–23, 2012.

[15] C. Sridhar, V. Gupta, P. Jain, and K. Vaidhyanathan, "Approach towards semi-automated certification for low criticality ml-enabled airborne applications," tech. rep., Software Engineering Research Center, IIIT Hyderabad, India, 2025.

[16] S. Burton and B. Herd, "Addressing uncertainty in the safety assurance of machine-learning," *Frontiers in Computer Science*, vol. 5, p. 1132580, 2023.

[17] O. Odu, A. B. Belle, and S. Wang, "Llm-based safety case generation for baidu apollo: Are we there yet?," tech. rep., Lassonde School of Engineering, York University, 2025.

[18] V. Gujral, P. U. Lee, G. Costedoat, J. Feldman, L. Spirkovska, and C. Walter, "Development of a safety hazards risk assessment tool for uncrewed aircraft system traffic management during preflight planning." NASA Ames Research Center Technical Publication, 2024.

[19] Y. Hong, C. S. Timperley, and C. Kästner, "From hazard identification to controller design: Proactive and llm-supported safety engineering for ml-powered systems," tech. rep., Carnegie Mellon University, 2025.

[20] M. Sivakumar, "Design and automatic generation of safety cases of ml-enabled autonomous driving systems," master's thesis, York University, Toronto, Ontario, April 2024.

[21] M. Sivakumar, A. B. Belle, J. Shan, and K. K. Shahandashti, "Exploring the capabilities of large language models for the generation of safety cases: the case of GPT-4," in *Proceedings of the 2024 IEEE 32nd International Requirements Engineering Conference Workshops (REW)*, pp. 35–45, IEEE, 2024.

[22] M. Zeller, "Safety assurance of autonomous systems using machine learning: An industrial case study and lessons learnt," in *Proceedings of the INCOSE International Symposium*, Wiley, 2023.

[23] K. Socha, M. Borg, and J. Henriksson, "Smirk: A machine learning-based pedestrian automatic emergency braking system with a complete safety case," *Software Impacts*, vol. 13, p. 100352, 2022.

[24] G. B. T. d. Carvalho, "Proposta de requisitos não funcionais para aderência de sistemas computacionais ao projeto de lei de regulamentação da inteligência artificial." Trabalho de Conclusão de Curso (Graduação em Ciência da Computação) — Universidade Federal da Paraíba, João Pessoa, 2024. Disponível em: https://repositorio.ufpb.br/jspui/handle/123456789/34679.

[25] C. Sadowski, M. Mühlbauer, D. Schmidt, and B. Dieber, "A survey on methods for the safety assurance of machine learning based systems," *Journal of Systems Architecture*, vol. 127, p. 102426, 2022.

[26] Y. Balagurunathan, H. Sun, R. A. Bruno, C. Mitchell, J. G. Goldgof, D. B. Goldgof, L. O. Hall, and R. Gatenby, "Requirements and reliability of ai in the medical context," *Journal of Imaging*, vol. 7, no. 11, p. 236, 2021.

[27] C. Paterson, R. Hawkins, C. Picardi, Y. Jia, R. Calinescu, and I. Habli, "Safety assurance of machine learning for autonomous systems," *Reliability Engineering System Safety*, vol. 264, p. 111311, 2025.