

INSTITUTO TECNOLÓGICO DE AERONÁUTICA



João Victor dos Santos

**APRIMORAMENTO DA FERRAMENTA DE
DIAGNÓSTICO PARA GERENCIAMENTO DE RISCO EM
AERÓDROMOS: MODELAGEM E IMPLEMENTAÇÃO**

Trabalho de Graduação
2025

Curso de Engenharia Civil-Aeronáutica

João Victor dos Santos

**APRIMORAMENTO DA FERRAMENTA DE
DIAGNÓSTICO PARA GERENCIAMENTO DE RISCO EM
AERÓDROMOS: MODELAGEM E IMPLEMENTAÇÃO**

Orientador

Prof. Dr. Evandro José da Silva (ITA)

ENGENHARIA CIVIL-AERONÁUTICA

**SÃO JOSÉ DOS CAMPOS
INSTITUTO TECNOLÓGICO DE AERONÁUTICA**

Dados Internacionais de Catalogação-na-Publicação (CIP)
Divisão de Informação e Documentação

dos Santos, João Victor

Aprimoramento da Ferramenta de Diagnóstico para Gerenciamento de Risco em Aeródromos:
Modelagem e Implementação / João Victor dos Santos.

São José dos Campos, 2025.

55f.

Trabalho de Graduação – Curso de Engenharia Civil-Aeronáutica– Instituto Tecnológico de Aeronáutica,
2025. Orientador: Prof. Dr. Evandro José da Silva.

1. Aeroportos. 2. Segurança operacional. 3. Gerenciamento de riscos. 4. Análise de falhas.
5. Desenvolvimento de software. 6. Inteligência artificial. 7. Computação. 8. Transportes. I. Instituto
Tecnológico de Aeronáutica. II. Título.

REFERÊNCIA BIBLIOGRÁFICA

DOS SANTOS, João Victor. **Aprimoramento da Ferramenta de Diagnóstico para Gerenciamento de Risco em Aeródromos: Modelagem e Implementação**. 2025. 55f. Trabalho de Conclusão de Curso (Graduação) – Instituto Tecnológico de Aeronáutica, São José dos Campos.

CESSÃO DE DIREITOS

NOME DO AUTOR: João Victor dos Santos

TÍTULO DO TRABALHO: Aprimoramento da Ferramenta de Diagnóstico para Gerenciamento de Risco em Aeródromos: Modelagem e Implementação.

TIPO DO TRABALHO/ANO: Trabalho de Conclusão de Curso (Graduação) / 2025

É concedida ao Instituto Tecnológico de Aeronáutica permissão para reproduzir cópias deste trabalho de graduação e para emprestar ou vender cópias somente para propósitos acadêmicos e científicos. O autor reserva outros direitos de publicação e nenhuma parte deste trabalho de graduação pode ser reproduzida sem a autorização do autor.

João Victor dos Santos
Rua H8B, 216
12228-461 – São José dos Campos, SP

APRIMORAMENTO DA FERRAMENTA DE DIAGNÓSTICO PARA GERENCIAMENTO DE RISCO EM AERÓDROMOS: MODELAGEM E IMPLEMENTAÇÃO

Essa publicação foi aceita como Relatório Final de Trabalho de Graduação

João Victor dos Santos

Autor

Evandro José da Silva (ITA)

Orientador

São José dos Campos, 18 de novembro de 2025.

*À minha esposa, Bianca Gabriela da Silva
Ibraim, por ter me acompanhado em cada
passo do sonho de me formar no ITA.*

Agradecimentos

À minha mãe, Elissalma, pelo apoio e pela força diante de todas as adversidades e distâncias enfrentadas, por sempre acreditar no quanto a educação pode transformar vidas e não medir esforços para que eu pudesse chegar aqui. Ao meu irmão Daryston, por ter cuidado de mim quando mais jovem e por me proteger dos perigos da vida. Ao meu irmão Vailson, que me fez abrir os olhos para esse sonho e me ofereceu a oportunidade de trilhá-lo, auxiliando-me com seus ensinamentos intelectuais para que eu alcançasse a aprovação no ITA e mostrou-me o caminho para me tornar um homem responsável.

À minha esposa, Bianca, por ter me acompanhado e encorajado a alçar novos ares para atingir este sonho; por sua compreensão, cuidado, companhia, palavras de carinho e incentivo, pelo companheirismo e por ser meu apoio diante das dificuldades acadêmicas e extra acadêmicas.

Aos amigos que fiz no H8 - Robalinho, Guilherme Ernando, Raphael Rosendo, Hermiro, Matheus Vilela - e a tantos outros que, embora não citados, me acompanharam nesta jornada e contribuíram para tornar este ciclo da graduação um período mais leve e alegre.

Ao meu orientador, professor Evandro, pela sua visão dinâmica, pelo cuidado sempre fraterno com nossa turma da Civil e pela paciência em compartilhar seu vasto conhecimento, sempre com um olhar inteligente e multidisciplinar.

Aos meus professores de graduação, especialmente a uma professora ímpar, Mayara Condé, que me ajudou a trilhar, escolher e desejar o ramo de transporte e tráfego aéreo como a área em que gostaria de atuar. Seu ensino em sala de aula é espetacular e despertou em mim a carreira almejada; é uma pessoa por quem nutro grande admiração e respeito.

Por fim, agradeço a Nosso Senhor Jesus Cristo que, sendo o sentido e sustento da minha vida, não coloca limites em Sua generosidade e misericórdia para comigo. Consagro este trabalho a Nossa Senhora, Mãe e Rainha da humanidade, a fim de que ela o ofereça, com suas santas mãos, para a maior glória de Deus.

*“Bem-aventurado és tu, quem quer que sejas,
que temes o Senhor, que andas nos seus caminhos.”*

— SALMOS 127:1

Resumo

O presente trabalho consiste no estudo de relatórios de incidentes e acidentes aeronáuticos para o desenvolvimento de uma ferramenta de análise e predição das causas prováveis e fatores contribuintes para a ocorrência do respectivo evento, bem como indicar possíveis falhas infraestruturais em casos fictícios. Utilizou-se inteligência artificial através de Modelos de Linguagem em Grande Escala (LLM) treinado e comparado com a sua base original para análise de resultados. Como modelo teste, obteve-se uma ferramenta que, à partir de narrativas, consegue identificar causas de incidentes e problemas infraestruturais, mas que necessita de supervisão, pois ainda incorre em inferências que podem ser não factuais. Nessa perspectiva, observa-se que o estudo representa um marco inicial em uma ferramenta teste e o modelo apresenta ganho na agilidade do processamento e entendimento situacional, mas necessita da supervisão humana para garantir seu resultado.

Abstract

This study consists of analyzing aeronautical incident and accident reports to develop a tool for analyzing and predicting the probable causes and contributing factors leading to a given event, as well as indicating possible infrastructural failures in fictitious cases. Artificial intelligence was employed through Large Language Model (LLM), which were trained and compared with their original base models for result analysis. As a test model, a tool was obtained that, based on narratives, can identify causes of incidents and infrastructural problems, but which still requires supervision, since it may incur inferences that are not factual. From this perspective, the study represents an initial milestone in a prototype tool: the model shows gains in processing speed and situational understanding, but still requires human oversight to ensure the reliability of its outputs.

Lista de Figuras

FIGURA 1.1 – Distribuição dos aeródromos públicos no Brasil por classe e por perfil operacional Fonte: Preparado com dados da AGÊNCIA NACIONAL DE AVIAÇÃO CIVIL (2025).	15
FIGURA 1.2 – Ferramenta Única da ANAC: interface do checklist com escala Likert . .	16
FIGURA 1.3 – Exemplo de análise bowtie para a categoria RI	16
FIGURA 2.1 – Exemplo de FTA. Fonte: (National Aeronautics and Space Administration, 2002).	22
FIGURA 2.2 – Exemplo de <i>bowtie</i> . Fonte: (AUST; PONS, 2020)	23
FIGURA 3.1 – Descrição geral dos passos feitos para obtenção da ferramenta.	26
FIGURA 3.2 – Guia da ferramenta CAROL do NTSB para descrição de filtros e extração de dados de incidentes/acidentes aéreos.	33
FIGURA 3.3 – Exemplo de formatação padrão do bloco de informações acerca do aeroporto contida nos relatório.	35
FIGURA 4.1 – CSV em Excel dos principais dados dos incidentes/acidentes aéreos para utilização na ferramenta.	40
FIGURA 4.2 – Caso 1 de problema infraestrutural ligado à <i>wildlife</i>	41
FIGURA 4.3 – Caso 2 de problema infraestrutural ligado à faixa de pista insuficiente para operação.	41
FIGURA 4.4 – Caso 3 de problema infraestrutural ligado à sinalização horizontal e placas informacionais.	42
FIGURA 4.5 – Caso 4 de problema infraestrutural ligado à sinalização luminosa e horizontal.	42
FIGURA 4.6 – Caso 5 de problema infraestrutural ligado à drenagem, resultando em acúmulos de água.	43

FIGURA 4.7 – Caso 6 de problema infraestrutural ligado à ausência de torre de controle.	43
FIGURA 4.8 – Caso de 7 problema infraestrutural ligado à problemas de superfície da pista de pouso e decolagem.	44
FIGURA 4.9 – Resultado das métricas BLEU-1, ROUGUE-L e SemSim.	47
FIGURA 4.10 – Matriz de comparação de classificadores binários gerada para o primeiro cenário.	48
FIGURA 4.11 – Matriz de comparação de classificadores binários gerada para o segundo cenário.	50
FIGURA 4.12 – Matriz de comparação de classificadores binários gerada para o terceiro cenário.	50

Lista de Tabelas

TABELA 1.1 – Classificação dos aeródromos quanto ao número de passageiros processados por ano (AGÊNCIA NACIONAL DE AVIAÇÃO CIVIL, 2024)	14
TABELA 3.1 – Hiperparâmetros que foram ajustados para o treinamento.	38
TABELA 4.1 – Valores dos hiperparâmetros após treinamento e validação.	45
TABELA 4.2 – Tabela de amostra do resultado das causas prováveis. As colunas de ProbableCause, PC_base e PC_ft se referem, respectivamente, às causas prováveis reais, às causas prováveis geradas pelo modelo base e às causas prováveis geradas pelo modelo treinado.	46

Sumário

1	INTRODUÇÃO	14
1.1	Contexto e Definição do Problema	14
1.2	Objetivos do Trabalho	17
1.3	Estrutura do Trabalho	17
2	A LITERATURA SOBRE IDENTIFICAÇÃO DE PERIGOS	19
2.1	Processos de Identificação de Perigos	20
2.1.1	<i>Faut-tree Analysis</i> (FTA) na Aviação	21
2.1.2	Método <i>Bowtie</i> na Identificação de Perigos	22
2.2	Inteligência Artificial e o Cenário da Aviação	24
3	METODOLOGIA	26
3.1	Base de Dados de Ocorrências Aeronáuticas	26
3.1.1	Base de Dados: <i>Airbus</i>	27
3.1.2	Base de Dados: <i>Boeing</i>	27
3.1.3	Base de Dados: <i>Flight Safety Fondation</i> (FSF)	28
3.1.4	Base de Dados: <i>International Civil Aviation Organization</i> (ICAO)	29
3.1.5	Base de Dados: <i>Aviation Safety Reporting System</i> (ASRS)	29
3.1.6	Base de Dados: <i>International Air Transport Association</i> (IATA)	30
3.1.7	Base de Dados: Painel SIPAER	30
3.1.8	Base de Dados: Centro de Investigação e Prevenção de Acidentes Aero- náuticos (CENIPA)	31
3.1.9	Base de Dados: <i>National Transportation Safety Board</i> (NTSB)	31
3.2	Sistematização dos Dados	33

3.3	Relatórios de Ocorrências Aeronáuricas	34
3.4	Incremento de Informações dos Relatórios na Base de Dados	35
3.5	Criação de Cenários Hipotéticos	36
3.6	<i>Large Language Models</i> (LLM) para Análise dos Dados e Geração de Causas e Perigos	36
3.7	Testes do Modelo	39
4	RESULTADOS E DISCUSSÕES	40
4.1	Banco de Dados	40
4.2	Aeródromos Hipotéticos e Seus Cenários	41
4.3	LLM Treinado	44
4.4	Teste com Causas Possíveis	45
4.5	Teste com Perigos Relacionados à Infraestrutura	47
5	CONCLUSÕES E TRABALHOS FUTUROS	52
	REFERÊNCIAS	54

1 Introdução

1.1 Contexto e Definição do Problema

O cenário de segurança operacional (*safety*) nos aeródromos brasileiros passa por constantes atualizações, de modo a manter o transporte aéreo em níveis de risco tão baixos quanto razoavelmente praticável e alinhados às metas e aos padrões internacionais estabelecidos pela ICAO (*International Civil Aviation Organization*). Entretanto, cada Estado possui condições operacionais, regulatórias e socioeconômicas particulares e, por isso, necessita desenvolver métodos que permitam atender a tais objetivos de forma compatível com sua realidade. No caso brasileiro, esses requisitos vêm sendo sistematicamente revisados pela ANAC, especialmente por meio do RBAC n.º 153, que define critérios de classificação dos aeródromos e requisitos mínimos de segurança operacional associados a cada classe.

A Tabela 1.1 apresenta a classificação regulamentar dos aeródromos brasileiros segundo o número de passageiros processados por ano, conforme previsto no RBAC n.º 153.

TABELA 1.1 – Classificação dos aeródromos quanto ao número de passageiros processados por ano (AGÊNCIA NACIONAL DE AVIAÇÃO CIVIL, 2024)

Classe	Descrição
I	Aeródromo em que o número de passageiros processados por ano é inferior a 200.000 .
II	Aeródromo em que o número de passageiros processados por ano é igual ou superior a 200.000 e inferior a 1.000.000 .
III	Aeródromo em que o número de passageiros processados por ano é igual ou superior a 1.000.000 e inferior a 5.000.000 .
IV	Aeródromo em que o número de passageiros processados por ano é igual ou superior a 5.000.000 .

Segundo a ANAC, o Brasil possui atualmente **490 aeródromos públicos abertos ao tráfego** (AGÊNCIA NACIONAL DE AVIAÇÃO CIVIL, 2025). Essa ampla rede aeroportuária impõe desafios significativos para a garantia da segurança operacional, considerando as dificuldades enfrentadas pelos operadores aeroportuários, pelas companhias aéreas e pelo órgão regulador

na gestão integrada do risco operacional.

Diferentemente dos grandes aeroportos — nos quais a infraestrutura, os sistemas e os procedimentos passam por processos formais de certificação e contam com operadores altamente especializados — grande parte dos aeródromos brasileiros é composta por instalações de menor porte, muitas vezes administradas por prefeituras ou governos estaduais. Nesses casos, limitações técnicas, operacionais e de recursos humanos impactam diretamente a capacidade de manter padrões elevados de segurança.

A Figura 1.1 apresenta a distribuição espacial dos aeródromos públicos por classe e por perfil operacional. Observa-se que a maior parte dos aeródromos pertence à Classe I e, dentro desta, predominam perfis não certificados, o que reforça a heterogeneidade das condições operacionais no País.

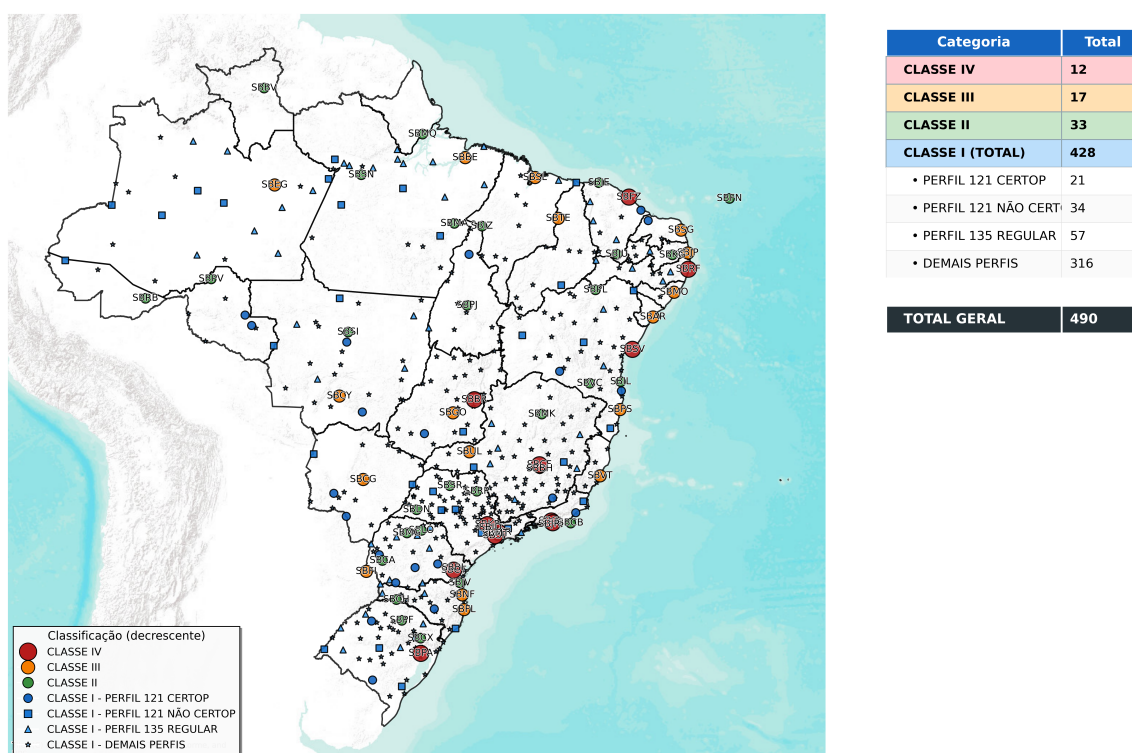


FIGURA 1.1 – Distribuição dos aeródromos públicos no Brasil por classe e por perfil operacional Fonte: Preparado com dados da AGÊNCIA NACIONAL DE AVIAÇÃO CIVIL (2025).

Nesse contexto, a atualização de ferramentas, processos e procedimentos em diferentes ambientes operacionais torna-se essencial para acompanhar o desenvolvimento de novas tecnologias voltadas à melhoria do desempenho, da capacidade e da agilidade dos processos de monitoramento e gestão da segurança operacional. Todavia, esse processo é complexo, uma vez que os aeródromos brasileiros apresentam grande variabilidade em termos de desafios de infraestrutura, operação, contexto socioeconômico e maturidade organizacional.

A Agência Nacional de Aviação Civil (ANAC) tem empregado a chamada Ferramenta Única (Agência Nacional de Aviação Civil (ANAC), 2022) como apoio ao diagnóstico de riscos de segurança

operacional e à identificação de potenciais mitigações. Esta ferramenta foi desenvolvida em alinhamento com *stakeholders* da indústria e inspirada no método *bowtie*, permitindo avaliar a existência de barreiras de infraestrutura, procedimentos e treinamento a partir de um *checklist* em que o analista utiliza uma escala Likert de 1 a 5 (ver Figura 1.2). Com estas entradas, para as categorias RI, RE, FOD, BIRD e WILD, obtém-se uma análise no formato *bowtie* (Figura 1.3).

Uma das deficiências apontadas nesta ferramenta em seu formato atual é que a planilha possui as mesmas perguntas e pesos independentemente do perfil operacional e das condições locais. Além disso, as perguntas são baseadas em consenso de especialistas e não em dados sistemáticos sobre ocorrências de segurança.

SUPERINTENDÊNCIA DE INFRAESTRUTURA AEROPORTUÁRIA - SIA Ferramenta de Diagnóstico para Gerenciamento de Risco de Aeródromos				RI	RE	FOD	BIRD	WILD	
				X	X	X	X	X	
				REATIVAS		ORGANIZACIONAIS			
				X			X		
Elaborado por: Organização Avaliadora 1: _____ Nome do avaliador: _____ Cargo/Função: _____ Organização Avaliadora 2: _____ Nome do avaliador: _____ Cargo/Função: _____ Tipo de Fiscalização: _____ Visita Técnica (VT)				Nome do Aeródromo: _____ Indicador ICAO: _____ Data: _____					
Organismo de avaliação em Processo Conjunto de Governamento de Risco e Garantia da Segurança Operacional, emitido para: info@anac.gov.br									
TIPIC	TEMJ	BARREIRA (ARJM)	COMENTÁRIOS	AValiação	RI	RE	FOD	BIR	WIL
1	FREV	AVF01	Existência de TWR operante	Moderada	3	3	3	3	3
2	FREV	AVF02	Existência de AFIS operante	Muito Fraca	1	1	1	1	1
3a	FREV	AVF03	Proteção adequada da área operacional (cercamento)	Forte	5	5	5	5	5
3b	FREV	AVF04	Proteção adequada da área operacional (cercamento)	Forte	5	5	5	5	5
4	FREV	AVF04	Configuração adequada do sistema de pistas, pátio e TFS	Moderada	3	3	3	3	3
5a	FREV	AVF04	Sinalização horizontal adequada	Moderada	3	3	3	3	3
5b	FREV	AVF04	Sinalização horizontal adequada	Moderada	3	3	3	3	3
6a	FREV	AVF04	Sinalização luminosa adequada [Não Aplicável se o aeródromo estiver cadastrado sob a condição operacional tipo Diurno]	Fraca	1	1	1	1	1
6b	FREV	AVF04	Sinalização luminosa adequada [Não Aplicável se o aeródromo estiver cadastrado sob a condição operacional tipo Diurno]	Moderada	3	3	3	3	3
7	FREV	AVF04	Sinalização vertical adequada	Fraca	1	1	1	1	1

FIGURA 1.2 – Ferramenta Única da ANAC: interface do checklist com escala Likert

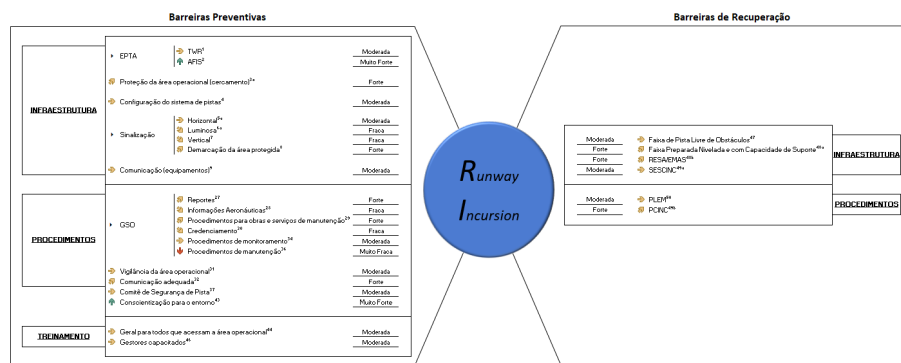


FIGURA 1.3 – Exemplo de análise bowtie para a categoria RI

Este Trabalho de Graduação busca contribuir com metodologias de análise da segurança operacional, considerando a possibilidade de utilizar grandes bases de dados sobre incidentes e acidentes aeroportuários. A próxima seção detalha o objetivo declarado no escopo deste estudo.

1.2 Objetivos do Trabalho

O presente Trabalho de Graduação tem por objetivo geral desenvolver e testar uma ferramenta baseada em Inteligência Artificial para a identificação de causas de incidentes e acidentes aeronáuticos associados a problemas de infraestrutura aeroportuária. O propósito é auxiliar na análise de segurança operacional em grandes volumes de relatórios, permitindo que analistas foquem nos eventos mais relevantes para o gerenciamento de riscos, a fim de sistematizar perigos e propor ações de mitigação.

Para atingir o objetivo geral, o trabalho estabeleceu os seguintes objetivos específicos:

1. Coletar e tratar dados reais de relatórios de ocorrências aeronáuticas para análise retroativa e criação de cenários fictícios para teste;
2. Treinar uma ferramenta de Inteligência Artificial através do uso de um *Large Language Model* (LLM);
3. Realizar a comparação do modelo treinado com o modelo base (pré-treinado) para avaliar o ganho de desempenho no domínio de segurança operacional de aeródromos.

1.3 Estrutura do Trabalho

A seguir, apresenta-se um resumo do trabalho, elaborado com o intuito de auxiliar na compreensão dos capítulos subsequentes:

- **Capítulo 1: Introdução**

Este capítulo apresenta o contexto, definição do problema, motivação e os objetivos do estudo, bem como uma visão geral da estrutura do trabalho.

- **Capítulo 2: A Literatura Sobre Identificação de Perigos**

Uma visão geral e uma análise crítica das pesquisas, teorias e metodologias existentes relevantes para o estudo.

- **Capítulo 3: Metodologia**

Uma descrição detalhada dos métodos, materiais e procedimentos utilizada para conduzir o presente trabalho.

- **Capítulo 4: Resultados e Discussões**

Apresentação dos resultados da pesquisa, seguida de discussão e interpretação dos mesmos.

- **Capítulo 5: Conclusão e Trabalhos Futuros**

Um resumos das principais descobertas, conclusões extraídas dos resultados e recomendações para estudos futuros.

2 A Literatura Sobre Identificação de Perigos

No âmbito de segurança operacional de aeródromos, para que se possam gerar planos de ação e melhorias estruturais capazes evitar incidentes e minimizar seus danos, é necessário, antes de tudo, analisar quais são os perigos aos quais os aeródromos estão submetidos. Nessa perspectiva, considerando que a identificação dos perigos constitui a etapa inicial e fundamental do processo de gerenciamento de riscos à segurança operacional, torna-se imprescindível definir o que pode ser configurado como um perigo e possuir uma compreensão clara daqueles aos quais a aviação está suscetível. Cabe salientar que a etapa de identificação é de suma importância, pois não é possível analisar consequências ou adotar medidas mitigadoras sem antes conhecer os elementos que podem originar um cenário de perigo.

De acordo com o Doc 9859 *Safety Management Manual* da *International Civil Aviation Organization* (ICAO) (International Civil Aviation Organization, 2018), um perigo, no setor da aviação, é definido como algo ou alguma situação que possa gerar um potencial dano - seja na forma de um incidente ou acidente aeronáutico -, seja no sistema ou em seu ambiente, podendo manifestar-se de formas variadas. Na operação aeroportuária, isso pode ocorrer por meio de condições ambientais (chuva intensa, *wind shear*, nevoeiro), estados técnicos e características físicas do aeródromo (aderência e contaminação de pista, drenagem insuficiente, sinalização luminosa/horizontal degradada, *Runway Safety End Area* inadequada), processos e serviços (procedimentos *Air Traffic Control* locais, *ground handling*, inspeções de pista) ou interações com terceiros (fauna, veículos, obras em área de movimento, *Foreign Object Debris*).

Somada à definição da ICAO, a ABNT NBR ISO 31000 (TÉCNICAS, 2018), que trata, em seu escopo, das diretrizes de gestão de risco, define risco como o efeito da incerteza nos objetivos. Transportando essa definição para o contexto em que o objetivo é a segurança operacional, um perigo instaura incerteza sobre esse objetivo (aumenta a chance de incidentes), enquanto o risco expressa como essa incerteza se materializa (combinação de probabilidade e consequência). Assim, perigo e risco não são sinônimos: o perigo é a fonte potencial; o risco é a exposição mensurável a esse potencial, dada uma operação específica, barreiras existentes e contexto local.

Dessa forma, pode-se entender que o objetivo da identificação é encontrar, reconhecer e

descrever sistematicamente os perigos que possam gerar algum risco no cenário da segurança aeroportuária. Essa etapa permite que sejam realizadas análises a partir de seus resultados, de modo a subsidiar planos de antecipação e mitigação dos perigos identificados.

Vale salientar que, de acordo com o Doc 9859 da ICAO, há distinção de eventos caracterizados como “perigo” e “consequência”. O referido manual apresenta como exemplo ilustrativo que um vento transversal à pista constitui um perigo, ao passo que uma excursão de pista configura uma possível consequência desse perigo, dependendo de fatores como o estado da pista, a técnica de pilotagem, os limites de operação e o suporte do aeródromo. Contudo, para fins de análise abrangente dos fatores de risco aos quais o aeródromo pode estar submetido, no presente trabalho os cenários descritos por ambas as definições serão considerados como risco ou perigo aplicáveis ao contexto do aeródromo.

2.1 Processos de Identificação de Perigos

Considerando os perigos e riscos aos quais o setor aeronáutico estão sujeitos e suas diversas fontes, torna-se imprescindível seguir metodologias claras, de modo a obter, com maior abrangência e precisão, a identificação dos perigos e, assim, atuar de modo eficiente na minimização dos danos.

O Manual de Gerenciamento de Segurança da ICAO, Doc 9859, promove uma divisão entre os métodos de identificação da seguinte forma:

- Reativos: São métodos que identificam perigos com base na análise de incidentes, acidente e eventos indesejados já ocorridos;
- Proativos: São métodos que objetivam antecipar perigos antes que eles causem danos por meio da análise do funcionamento normal da organização.

Eles permitem a identificação do perigo antes que os eventos indesejados possam ocorrer. Seus dados provêm, principalmente, de programas de análise de dados de voo e sistemas de reporte de segurança.

Métodos baseados na análise de dados históricos, projeções e simulações para antecipar tendências emergentes de risco também são citados no documento, podendo ser definidos como métodos preditivos, apesar de não haver menção explícita ao termo no manual. Somado ao manual da ICAO, a NBR IEC 31010 (TÉCNICAS, 2012) apresenta diversos métodos e técnicas para o processo de avaliação de risco. Nas próximas seções, serão brevemente detalhados dois métodos em específico: *Fault-tree Analysis* e Bowtie. Esses métodos foram escolhidos para descrição por dois motivos:

- Eles descrevem todo o ambiente que produz um perigo, gerando as causas de risco para o aeródromo. Dessa forma, são métodos de identificação que não focam apenas no perigo, mas em suas causas e consequências;
- São amplamente utilizados no setor aeronáutico.

2.1.1 *Faut-tree Analysis* (FTA) na Aviação

Estudos da *National Aeronautics and Space Administration* (NASA) (National Aeronautics and Space Administration, 2002) e da *National Examination Board in Occupational Safety and Health* (NEBOSH) (RRC Training,) demonstram que o FTA permanece um dos métodos mais robustos para decomposição de falhas técnicas e operacionais, permitindo a quantificação de probabilidades e a priorização de barreiras de mitigação. No contexto aeronáutico, o FTA é amplamente empregado em sistemas críticos (propulsão, comandos de voo, geração/ distribuição elétrica, hidráulica) para avaliar como combinações de falhas latentes e ativas podem culminar, por exemplo, em *loss of control in flight* (LOC-I).

O *Fault-Tree Analysis* (FTA) foi desenvolvido na *Bell Laboratories*, em 1962, para a força aérea dos Estados Unidos, com o objetivo de avaliar a confiabilidade do míssil Minuteman II, sendo rapidamente adotado por setores que exigiam alto nível de segurança, entre eles o aeroespacial. Trata-se de uma técnica *top-down*: parte-se de um “evento indesejado” (chamado de *top event*) — por exemplo, perda de controle em voo — e, por meio de portas lógicas (como *AND*, *OR*), decompõem-se as combinações de falhas que podem conduzir a esse evento. O método suporta dois níveis de análise:

- Qualitativo: identificação estruturada de causas que revelam caminhos críticos para o *top event* e orientam ações de mitigação. Ou seja, identifica potenciais causas pelas quais a falha ocorreu.
- Quantitativo: cálculo da probabilidade do *top event* a partir das probabilidades (ou taxas) dos eventos básicos e de suas relações lógicas, permitindo derivar medidas de importância (Birnbaum, Fussell-Vesely, risco atribuível) para priorizar barreiras com maior eficácia risco-custo.

No contexto de aeródromos, a FTA é particularmente útil para transformar um problema amplo (por exemplo, “excursão de pista”) em hipóteses verificáveis conectadas a perigos controláveis pelo gestor aeroportuário.

Analisando o modo qualitativo (é o curso pelo qual será útil para o trabalho) o processo se decompõe do seguinte modo:

- Definição do *top event* (Por exemplo, *runway excursion*);

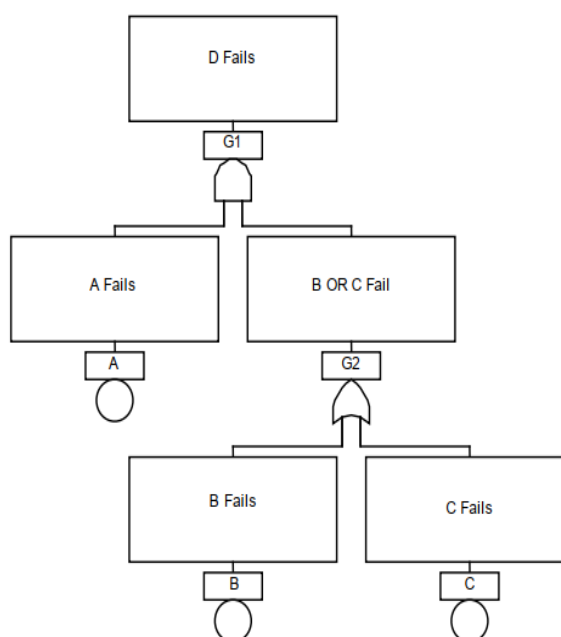


FIGURA 2.1 – Exemplo de FTA. Fonte: (National Aeronautics and Space Administration, 2002).

- Construção do diagrama utilizando portas lógicas para relacionar as falhas numa sequência estruturada;
- Qualificação das falhas básicas.

Dessa forma, observa-se que há uma regra geral de análise: observa-se uma narrativa e busca-se o que pode gerar e causar aquela narrativa.

2.1.2 Método *Bowtie* na Identificação de Perigos

O método *bowtie* emergiu na década de 1990 como uma resposta ao desafio de comunicar riscos complexos a públicos não especialistas, unindo a lógica causal do *Fault-Tree Analysis* (FTA) – ameaças – à lógica consequencial do *Event-Tree Analysis* (ETA) – consequências – tendo no centro o evento de topo (*top event*), entendido como a materialização imediata do perigo. O resultado é um mapa visual de risco que mostra, de um lado, como o perigo pode ser ativado (cadeias causais) e, de outro, o que ocorre posteriormente (cascatas de consequências), tornando explícitas as barreiras envolvidas: aquelas que previnem a ocorrência do evento de topo (barreiras preventivas) e aquelas que limitam sua severidade caso ele ocorra (barreiras de recuperação/mitigação).

Assim, em sua essência, o método consiste na formulação de um conjunto estruturado de perguntas em sequência lógica, de modo que se obtenha, ao final, a ilustração do perigo, do evento indesejado, dos eventos de segurança, dos potenciais resultado e dos controles de risco

implementados para sua minimização. Trata-se de um método esquemático simples que descreve e analisa os caminhos de um perigo desde suas causas até suas consequências (AUST; PONS, 2020).

O controle de riscos se dá pelo estabelecimento de barreiras que previnam a ocorrência dos eventos ou perigos indesejáveis, bem como por medidas regulatórias que atuem no sentido de manter um estado desejado, ou seja, mitigar a gravidade das consequências que, ocorrido o evento, possam surgir.

No contexto de aeródromos, o *bowtie* é particularmente útil por aproximar o risco do ponto de intervenção gerencial. Por exemplo: para o *top event* “excursão de pista na aterrissagem”, as ameaças à esquerda podem incluir pista (*runway*) contaminada por água ou borracha, drenagem insuficiente, balizamento ou sinalização horizontal degradados, aproximação instável, vento de cauda, pavimento com macrotextura ou microtextura inadequada, ou ainda FOD ou fauna. As barreiras preventivas correspondem a projetos e manutenções (pavimento, *grooving*, drenagem), processos (inspeções e medições de condição de pista, reporte padronizado, limpeza), tecnologia (balizamento e *Precision Approach Path Indicator* em conformidade), pessoas (treinamento de *airside ops*, *standard phraseology* e coordenação) e gestão (NOTAMs, controle de obras em área de movimento). À direita, consequências como danos à aeronave, interrupção de pista e ferimentos são mitigadas por barreiras de recuperação: RESA/EMAS, planos de emergência, posicionamento e tempo de resposta *Aircraft Rescue and Fire Fighting* (ARFF), rotas de acesso, equipamentos de remoção e protocolos de evacuação.

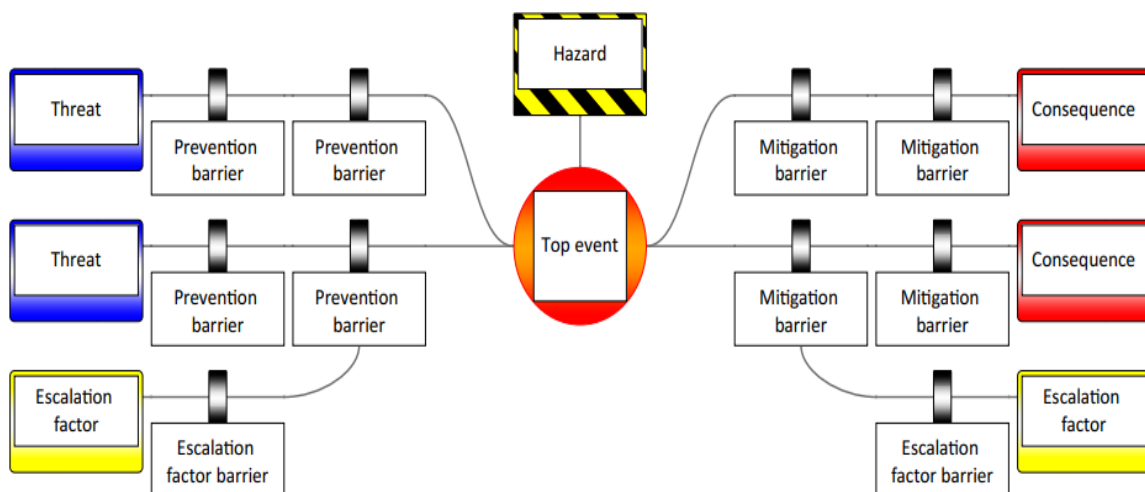


FIGURA 2.2 – Exemplo de *bowtie*. Fonte: (AUST; PONS, 2020)

O método pode ser empregado de forma qualitativa (inventariando causas, barreiras, fraquezas) ou semi-quantitativa, acoplando noções de frequência e severidade em cada ramo - ou mesmo importâncias derivadas de FTA.

Dessa forma, o método *bowtie* gera informações importantes acerca de identificação de perigos: por estar acoplado ao FTA, incorpora as causas possíveis de um *top event*, e, adicional-

mente, oferece a rastreabilidade das consequências, as quais também podem ser consideradas como perigos potenciais no contexto do aeródromo.

2.2 Inteligência Artificial e o Cenário da Aviação

Considerando os fatores históricos, a identificação de perigos na aviação depende da análise combinada de dados de voo, registros operacionais, inspeções e relatórios investigativos. Dessa forma, com o crescimento exponencial do volume e da diversidade de dados, técnicas e métodos que utilizam inteligência artificial tem surgido para automatizar seu processamento e melhorar a velocidade e a proatividade na análise dos perigos decorrentes de relatórios, elevando a eficiência dos Sistemas de Gerenciamento da Segurança Operacional (SGSO).

Um eixo particularmente promissor envolve Modelos de Linguagem em Grande Escala (*Large Language Models* — LLMs) aplicados a tarefas como resumo, classificação multi-rótulo de perigos, extração de fatores contribuintes e Recuperação Aumentada por Conhecimento (*Retrieval-Augmented Generation* - RAG) em acervos de relatórios, transformando a dinâmica de trabalho no setor aeronáutico.

Estudos recentes de Inteligência Artificial exploraram LLMs como ferramentas de apoio à decisão para controladores e gestores de tráfego aéreo, desenvolvendo o *chatATC*, uma interface conversacional que integra dados históricos de GDPs (*Ground Delay Programs*) e suas características para embasar decisões sobre novos GDPs. O achado central é que, quando corretamente instruídos por *prompts* e conectados a bases de dados confiáveis, os LLMs reduzem o atrito informacional na tomada de decisão, oferecendo contexto sintetizado sem substituir o julgamento humano (ABDULHAK *et al.*, 2024). O trabalho desenvolvido nesse artigo fez uso de duas abordagens de treinamento para o LLM: *in-prompt* e *fine-tuned*. Diante dos resultados, ambos os modelos de treinamento mostraram-se assertivos para as requisições dos usuários, indicando desempenho satisfatório. Apesar do rendimento equivalente, o modelo de treinamento *in-prompt* possui a desvantagem da limitação imposta ao comprimento dos *prompts* de entrada, o que reduz drasticamente a quantidade de dados que podem ser importados. Em contrapartida, o modelo treinado por *fine-tuned* não possui essa limitação, sendo apenas restringido no quesito de reduzir a carga computacional. Além disso, o *fine-tuned* permite a personalização da saída para adaptar à alguma interface gráfica. Nessa perspectiva, observa-se o benefício da customização do modelo ao formato desejado pelo desenvolvedor, contraposto à necessidade de mais teste para avaliar seu desempenho e ao maior poder computacional requerido para seu desenvolvimento.

Somado ao citado anteriormente, o estudo de Archana Tikayat Ray (RAY *et al.*, 2023) desenvolveu, também utilizando métodos baseados em LLMs, uma ferramenta para geração de resumos mais sucintos de incidentes aeronáuticos utilizando o ChatGPT com base nos relató-

rios disponibilizados pelo *Aviation Safety Reporting System* (ASRS) em sua base de dados. O estudo foi conduzido utilizando a engenharia de *prompt* de modo a estruturar comandos que retornassem respostas satisfatórias. Além disso, foi possível realizar uma comparação com as sinopses reais a fim de verificar a congruência da ferramenta e sua assertividade. Por fim, o estudo também obteve uma análise de possíveis fatores humanos que estariam relacionados aos incidentes, bem como a determinação da entidade à qual o acidente poderia ser atribuído.

Além do que já foi citado, o estudo de Simon Diemert e Jens H. Weber (DIEMERT; WEBER, 2023) apresenta a atuação de um modelo baseado em LLM's para contribuir com as investigações relativas à geração de ideias no processo de análise de perigos. O projeto desenvolveu uma ferramenta que, a partir da descrição preliminar de um sistema, é capaz de identificar condições de perigo. Assim, os autores propuseram um sistema de equipe humano-IA para auxiliar analistas, denominando essa abordagem de *Co-hazard Analysis* (CoHA), especialmente útil em seções de *brainstorming*.

Em síntese, a literatura recente e os resultados práticos convergem para o seguinte: LLMs não substituem a análise de segurança, mas potencializam sua escala, velocidade e coerência, sobretudo quando conectados a bases confiáveis e métodos estruturados de engenharia de risco.

3 Metodologia

Para facilitar a visualização do processo metodológico para a obtenção da ferramenta, apresenta-se a seguir um mapa mental descritivo dos passos realizados:

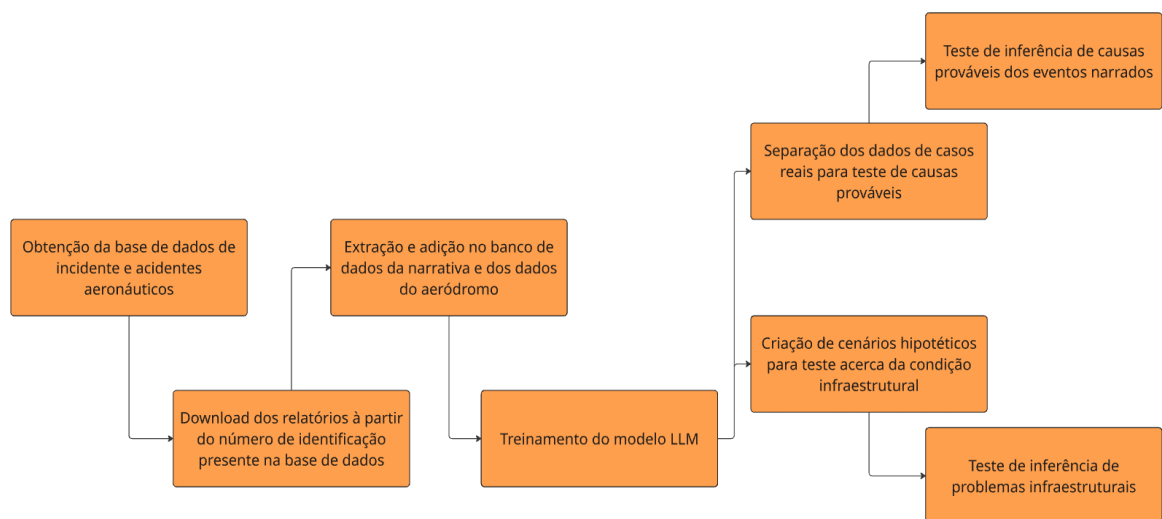


FIGURA 3.1 – Descrição geral dos passos feitos para obtenção da ferramenta.

Segue, nas próximas seções, a descrição de cada passo do processo.

3.1 Base de Dados de Ocorrências Aeronáuticas

Para identificar perigos operacionalmente controláveis ou mitigáveis pelos gestores de aeródromos, foi conduzido um mapeamento de fontes públicas e institucionais de segurança da aviação com o objetivo de obter relatórios de incidentes/acidentes aéreos com foco em ocorrências, fatores contribuintes e recomendações. O planejamento dessa etapa seguiu quatro critérios de seleção, pensados para maximizar valor analítico e rastreabilidade: (I) escopo internacional ou nacional relevante, de modo a permitir comparabilidade entre contextos e períodos; (II) extraibilidade técnica, privilegiando repositórios com exportação estruturada (CSV/JSON) e com campos textuais ricos — como *Analysis*, narrativas, *Probable Cause* e dados do aeródromo —, imprescindíveis para mineração semântica e validação qualitativa; (III) taxonomias consolidadas, contemplando tipo de ocorrência, fase do voo, tipo de aeronave, aeródromo, lesões,

dano material e condições meteorológicas, o que reduz ambiguidade classificatória e favorece integração entre bases; e (IV) rastreabilidade metodológica, assegurada por metadados claros (datas, versões, origem), condição necessária para auditoria e reprodutibilidade dos resultados.

Aplicando esses critérios, procedeu-se a uma triagem com testes de extração e de qualidade de campos (completude, padronização, consistência temporal). Nessa análise comparativa, a base da *National Transportation Safety Board* (NTSB) destacou-se por combinar amplitude histórica, granularidade de variáveis e riqueza narrativa, além de oferecer uma interface de consulta e extração enxuta e intuitiva, que facilita a construção de filtros reprodutíveis. Em termos práticos, a NTSB possibilita: (I) vincular variáveis estruturadas (datas, local, tipo de aeronave, fase de voo, lesões, danos, meteorologia) a descrições textuais de alta qualidade (seção *Analysis* e *Probable Cause*); (II) recuperar detalhes de aeródromo e pista quando presentes nos relatórios; e (III) exportar os resultados em formatos adequados a *pipelines* analíticos. Essa combinação atende diretamente ao objetivo do trabalho: transformar evidência documental em identificação de perigos sob governança do aeródromo.

Com base nisso, a NTSB foi eleita como fonte primária para a análise principal. Na sequência deste capítulo, apresenta-se um resumo crítico das bases avaliadas e, por fim, descrevem-se os procedimentos operacionais adotados para obtenção, enriquecimento e integração dos dados da NTSB, que compõem o núcleo empírico das análises subsequentes.

3.1.1 Base de Dados: *Airbus*

O *Airbus Accident Statistics* (Airbus, 2024) é um portal público voltado a sínteses estatísticas de acidentes envolvendo jatos comerciais no período 1958–2024. O conteúdo é organizado em seções como *Fatal Accidents*, *Hull Losses*, *Accident Categories*, *Accident Rates by Category & Generation* e *Accidents by Flight Phase*, permitindo visualizar tendências de longo prazo, taxas por categoria e diferenças por geração de aeronaves. A experiência é centrada em gráficos e painéis interativos, com possibilidade de download das figuras (PNG) mas não há mecanismos de consulta linha-a-linha. O objetivo do site é, portanto, contextualizar o risco em nível setorial, oferecendo um contexto macro para comparações e *benchmarking*.

Como limitação, o portal não disponibiliza microdados exportáveis (CSV/JSON) nem relatórios técnicos por evento. Assim, não suporta mineração analítica no nível de aeródromo. Trata-se, portanto, de um portal focado em sumários estatísticos — útil para *benchmarking* e *framing*.

3.1.2 Base de Dados: *Boeing*

A *Boeing* mantém um *hub* público de segurança cujo destaque é a *Aerospace Safety Timeline* (The Boeing Company,), um recurso em forma de linha temporal interativa, de caráter “narrativo”,

pensado para aprendizado e pesquisa, que vai dos primórdios da aviação até programas atuais. O recurso organiza eventos marcantes, decisões regulatórias, recomendações e melhorias técnicas em cartões com cores e legendas temáticas, oferecendo uma leitura histórica e contextual do “porquê” e “quando” determinadas mudanças ocorreram. A ênfase está na fruição visual e no aprendizado — mais em contar a história do sistema de segurança aeronáutica do que em permitir consultas analíticas profundas.

Do ponto de vista de usabilidade, a exploração é feita dentro da própria *timeline*: o usuário clica em um marco para abrir o cartão associado e, em muitos casos, é redirecionado à fonte primária (por exemplo, relatórios do NTSB ou materiais da *Flight Safety Foundation*) para leitura completa. Não há uma página única que liste todos os incidentes ou um índice tabular tradicional; além disso, o site não oferece exportação de microdados (CSV/JSON) nem filtros avançados multivariados típicos de repositórios orientados a modelagem.

Dessa forma, o *hub* é focado em conectar eventos, regulamentos e aprendizados ao longo do tempo, contando com recursos visuais. Para análises quantitativas em nível de aeródromo, modelagem ou mineração de dados por ocorrência, entretanto, é necessário recorrer a outra base que disponibiliza microdados estruturados e narrativas completas.

3.1.3 Base de Dados: *Flight Safety Foundation* (FSF)

A *Flight Safety Foundation* (FSF) (*Flight Safety Foundation*,) disponibiliza *dashboards* interativos construídos sobre a base global do *Aviation Safety Network* (ASN), com foco em apresentar indicadores agregados (contagem anual de acidentes/fatalidades, distribuição por fase de voo, natureza de operação e região). A proposta do portal é sintetizar tendências e facilitar uma leitura rápida do panorama de segurança, sem expor o nível transacional dos registros.

Contudo, a versão gratuita do painel é bastante restrita: em geral, cobre apenas o último ano e exibe métricas gerais, com poucos filtros e sem acesso a microdados (evento a evento). O acesso avançado — tanto aos *dashboards* completos quanto aos *datasets* em profundidade (incluindo opções de *download* e consulta via API) — é pago e direcionado a membros ou assinantes. Na prática, isso limita o valor do FSF/ASN para pesquisas que exigem mineração analítica por aeródromo, cruzamento de variáveis ou extração de narrativas.

Vale salientar que, como a versão gratuita é limitada, pouco foi agregado de informação acerca do conteúdo da base disponível pelo portal, visto que apenas a versão paga tem acesso aberto e amplo dos dados.

3.1.4 Base de Dados: *International Civil Aviation Organization (ICAO)*

A ICAO (International Civil Aviation Organization,) mantém um portal que reúne, sobretudo, relatórios de situação da segurança (globais e regionais) e documentos estratégicos (como o *Safety Report*, o *Global Aviation Safety Plan*) e materiais de padronização taxonômica (por exemplo, ADREP) e de monitoramento (por exemplo, iSTARS/daTIS). Esses conteúdos são valiosos para contextualizar tendências (categorias de risco prioritárias, taxas por região/segmento), definir metas e indicadores de desempenho de segurança e harmonizar nomenclaturas e classificações utilizadas na indústria. Em outras palavras, o portal funciona como referência normativa e estratégica, consolidando *guidance* e terminologia que embasam políticas de segurança e a comparabilidade entre Estados/operadores. Apesar de todo aparato, o site não oferece um banco público de microdados evento-a-evento com exportação direta (CSV/JSON) que permita mineração analítica de incidentes/acidentes em nível de aeródromo (por exemplo, cruzar condições de pista, fatores contribuintes, narrativas).

Desse modo, se mostrou um portal importante no quesito de documentação acerca de recomendações para melhoria contínua de segurança, como também nomenclaturas importantes acerca de incidentes, mas não foi utilizado como base para extração de relatórios.

3.1.5 Base de Dados: *Aviation Safety Reporting System (ASRS)*

O ASRS (National Aeronautics and Space Administration,) é um repositório voluntário, confidencial e não punitivo de relatos de segurança operado pela NASA, cuja base é voltada à coleta de relatos de segurança enviados por profissionais de linha de frente (pilotos, controladores, comissários, mecânicos). Cada registro combina uma narrativa sanitizada (com remoção de identificadores) com campos codificados por analistas (data, local, fase do voo, anomalias, fatores humanos), o que torna a base particularmente rica para estudos de fatores humanos, contexto operacional e condições latentes que nem sempre emergem em relatórios investigativos formais.

Para fins de consulta e extração, o ASRS disponibiliza um “*Query Wizard*” com amplo conjunto de filtros (por campos estruturados e por busca textual em *Narrative/Synopsis*), operadores lógicos e opções de exportação (visualização *web*, *Word*, planilha/CSV). Essa flexibilidade, porém, exige familiarização prévia: o leque de campos, operadores e particularidades da sintaxe (curingas, combinações *AND/OR/NOT*, proximidade) pode dificultar a intuição de uso e demanda um percurso iterativo de testes até se chegar a consultas reproduzíveis.

Somado a isso, os relatos são de conteúdo predominantemente informal (a própria página ressalta que o ASRS não “verifica/valida” os relatos), mostrando, assim, que não se enquadra na base de dados requerida para o trabalho.

3.1.6 Base de Dados: *International Air Transport Association* (IATA)

A IATA concentra, em uma única página, recursos numéricos e gráficos voltados à contextualização da segurança operacional e ao apoio à prevenção (International Air Transport Association,). O núcleo é um *dashboard* interativo que organiza indicadores por categorias de risco prioritárias (por exemplo, *Loss of Control In-Flight*, *Controlled Flight Into Terrain*, *Runway Excursions*), permitindo *benchmarking* entre regiões/anos e a visualização de tendências (taxas, distribuição por tipo de evento, recortes geográficos). A ênfase está em sínteses de alto nível, com visualizações que facilitam leituras comparativas e comunicação executiva.

Do ponto de vista de pesquisa aplicada, entretanto, trata-se de um painel de tendências e resumos, não de um repositório de microdados por ocorrência. A interface não expõe, para *download*, registros linha-a-linha com variáveis operacionais (aeródromo, condições de pista, narrativa do evento, fatores contribuintes), o que inviabiliza mineração analítica no nível de detalhe requerido por este trabalho.

Além do painel, a IATA disponibiliza documentos de referência—como guias de recomendações de prevenção e relatórios de situação de segurança—que consolidam boas práticas, taxonomias e prioridades de risco. Dessa forma, apesar do acervo de conteúdo, a página não se enquadra como banco de dados para exportação e estudo dos casos.

3.1.7 Base de Dados: Painel SIPAER

O Painel do Sistema de Investigação e Prevenção de Acidentes Aeronáuticos (SIPAER) (Brasil. Centro de Investigação e Prevenção de Acidentes Aeronáuticos, a) é a interface pública de *business intelligence* de visualização interativa do CENIPA para ocorrências aeronáuticas no Brasil. Assim como o painel da IATA, ele privilegia *dashboards*, gráficos e recursos visuais interativos para leitura rápida e comparação de indicadores das tendências nacionais em vez de oferecer um repositório transacional linha a linha. A navegação é organizada por módulos temáticos — Ocorrências, Acidentes e Recomendações — com filtros por período, tipo/classificação, fase do voo, fatores contribuintes, natureza da operação e recortes regionais. Esse desenho favorece comparações sintéticas, *benchmarking* entre recortes e comunicação executiva do panorama de segurança. Soma-se também ao painel uma lista de recomendações de segurança para ações mitigadoras vinculadas a eventos e tendências nacionais.

Do ponto de vista analítico, porém, o painel impõe limitações importantes. A exportação em CSV disponível na interface geralmente traz um conjunto enxuto de campos (identificação básica, datas, tipificação agregada e poucas variáveis operacionais), sem narrativas detalhadas por evento, sem padronização completa de atributos de aeródromo e sem metadados suficientes para reconstruir o caso no nível requerido por modelos e mineração textual.

Dessa forma, ele não atende aos requisitos deste estudo para extração de dados estruturada e reprodutível de dados por ocorrência com narrativas.

3.1.8 Base de Dados: Centro de Investigação e Prevenção de Acidentes Aeronáuticos (CENIPA)

O CENIPA (Brasil. Centro de Investigação e Prevenção de Acidentes Aeronáuticos, b) é a autoridade brasileira responsável pela investigação de ocorrências aeronáuticas e pela produção de Relatórios Finais (RF) que documentam fatos, análises, fatores contribuintes e recomendações de segurança. Como fonte primária e oficial, seus relatórios são particularmente valiosos para compreender fatores operacionais e infraestruturais em aeródromos brasileiros (pista, sinalização, serviços, gestão de fauna/FOD, coordenação local), além de oferecerem rastreabilidade documental para decisões de mitigação. Em alguns períodos há, ainda, dados estruturados disponibilizados em portais governamentais, o que amplia a utilidade para análises quantitativas. Contudo, a abrangência, a padronização e a extraibilidade desses dados variam ao longo do tempo, o que impacta a reprodutibilidade.

No portal do CENIPA, os RFs são apresentados em uma listagem com campos descritivos sintéticos e links diretos para os PDFs de cada caso. Os relatórios mais recentes costumam ter boa formatação e organização interna, facilitando a leitura técnica (escopo, fatos, análise, fatores e recomendações). Entretanto, do ponto de vista de automação, há duas limitações críticas: (I) inconsistência histórica de formato — ao ampliar a janela temporal, observa-se heterogeneidade de estrutura, terminologia e *layout* entre relatórios antigos e recentes; (II) ausência de mecanismos nativos de exportação — não há na página de RF nem de referência de dados únicos para download em CSV/JSON nem API pública que permita coletar, em lote, metadados e trechos relevantes (por exemplo, fatores e menções de pista).

Metodologicamente, ainda que o CENIPA permaneça referência formal para consulta qualitativa e validação de achados (especialmente quando o foco é o contexto brasileiro) e ser uma boa fonte primária e oficial para conteúdos de incidentes aeronáuticos, tais restrições de extração estruturada, filtragem avançada e padronização longitudinal reduziram sua adequação como base empírica principal deste estudo, cujo objetivo requer microdados reprodutíveis e narrativas acessíveis em escala.

3.1.9 Base de Dados: *National Transportation Safety Board* (NTSB)

A NTSB mantém um banco público e extensivo de acidentes/incidentes aeronáuticos, com campos estruturados (data, local/aeródromo, tipo de aeronave, fase de voo, danos, lesões, meteorologia) e campos textuais detalhados de alta qualidade (como *Analysis* e *Probable Cause*). As investigações têm metodologia transparente, padronização consistente e ampla cobertura

histórica, o que habilita quatro capacidades centrais para esse trabalho: (I) extração replicável em CSV do próprio site, abarcando vários dados e a rotulação do relatório para otimização de extração; (II) vinculação entre variáveis estruturadas e narrativas; (III) classificação/rotulagem de perigos (excursão ou incursão de pista, *Foreign Object Debris* (FOD), fauna); e (IV) filtros riscos por condição operacional e tipo de evento. Em comparação às demais fontes avaliadas, a NTSB oferece a melhor combinação de granularidade, qualidade narrativa e acessibilidade pública, atendendo ao objetivo do estudo de identificar perigos controláveis por aeródromos e quantificá-los com base em evidências.

A plataforma *Case Analysis and Reporting Online* (CAROL) (National Transportation Safety Board,) foi utilizada como fonte de pesquisa, filtragem e obtenção dos dados para o estudo, pois além de conter o aparato necessário (CSV exportável com informações relevantes, arquivos dos relatórios na íntegra), possui uma interface intuitiva para filtragem da pesquisa objetivada.

Para obtenção inicial dos dados base, utilizou-se a *custom search* do CAROL para se ter o controle acerca das características principais que eram necessárias nos dados para posterior obtenção dos relatórios. O portal possibilita o agrupamento de regras de filtro para melhor separação e análise. Dessa forma, foram divididos três grupos de regras para filtragem e obtenção dos dados.

No primeiro grupo foi concentrado os filtros fixos de contexto e escopo com o objetivo de definir o universo geral investigativo. Foi definida uma janela temporal de vinte anos tanto para a data do evento quanto para a data de publicação do relatório, assegurando uma base numericamente robusta e, ao mesmo tempo, representativa das práticas e da infraestrutura atuais. Somado a isso, limitou-se o país aos Estados Unidos da América pelo fato de que há o acesso direto do NTSB ao acervo investigativo doméstico e, conseqüentemente, a maior granularidade dos campos estruturados e textuais. Ainda, para assegurar qualidade informacional, restringiu-se o *status* a relatórios completos, evitando registros preliminares ou inconclusos. Por fim, foi delimitado o tipo de aeronave a “*Airplane*”, mantendo coerência com o objetivo do estudo, focado em operações de asa fixa e sua interação com a infraestrutura do aeródromo.

O segundo grupo tratou da fase do voo, com o intuito de aproximar a amostra dos eventos em que fatores sob governança do aeródromo — pista, áreas de movimento, sinalização, serviços e procedimentos locais — exercem influência direta. Para isso, dada a opção de retirar características pelo CAROL, aplicou-se exclusões às fases de voo *Enroute*, *Uncontrolled Descent*, *Emergency Descent* e *Initial Climb*. Essas fases descrevem trechos do envelope operacional predominantemente fora do âmbito infraestrutural do aeródromo. Dessa forma ao excluí-las, foi privilegiado ocorrências relacionadas a táxi, decolagem, aproximação, pouso e movimentação em superfície, que são mais pertinentes ao escopo deste trabalho.

Por fim, o terceiro grupo estabeleceu as categorias de evento de interesse e foi aplicado de maneira isolada, uma a uma, em consultas repetidas. Essa escolha metodológica decorre

de uma limitação prática do CSV exportado pelo CAROL: o arquivo não delimita a que categoria de evento está referenciado aquele incidente, ou seja, não possui uma aba para indicar o *event cause*. Dessa forma, caso fosse colocado mais de uma categoria de evento, os arquivos poderiam pertencer a um ou ambos tipos, sem conseguir ser identificado sem sua leitura. Assim, conduziu-se buscas separadas para cada uma das seguintes categorias: *Runway Excursion*, *Runway Incursion*, *Airport Occurrence*, *Collision on Takeoff or Landing*, *Wildlife (non-bird)*, *Birdstrike*, *Ground Collision*, *Abnormal Runway Contact*, *Wrong Fuel*, *Fuel Contamination*, *Air Traffic*, *Ground Handling*, *Loss of Control (Ground)*, *Wrong Surface or Wrong Airport*, *Undershoot/Overshoot*, *Windshear or Thunderstorm*, *Evacuation* e *Hazardous Material Leak/Spill*.

Cada execução manteve integralmente os filtros dos Grupos 1 e 2 e variou apenas a categoria do Grupo 3, gerando um CSV exportado específico por tipo de evento. Vale salientar que, da mesma forma que o grupo 2, as categorias especificadas no grupo 3 tem por objetivo restringir a casos que possam ser colocados dentro da infraestrutura do aeroporto e que sofrem sua influência direta.

A decisão por estruturar a busca dessa forma teve dois objetivos principais: garantir reprodutibilidade e manter a rastreabilidade entre os resultados exportados e as categorias investigativas de interesse do trabalho. Assim, obteve-se 18 bases de incidentes de acordo com a classificação do *event cause* para serem analisadas.

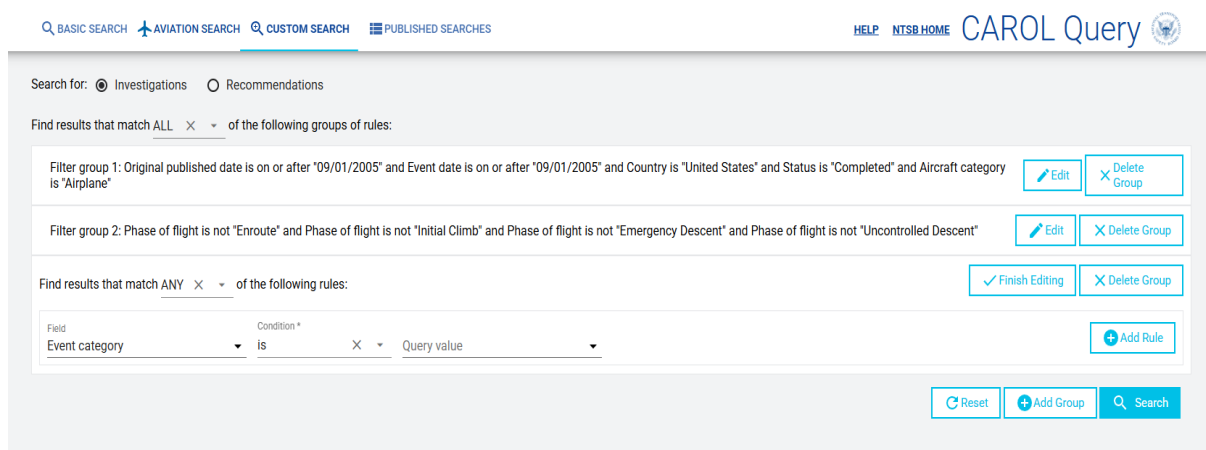


FIGURA 3.2 – Guia da ferramenta CAROL do NTSB para descrição de filtros e extração de dados de incidentes/acidentes aéreos.

3.2 Sistematização dos Dados

Obtido os dados como dito anteriormente, cada um dos CSV's foram nomeados de forma a ser identificado com o tipo de evento ao qual ele se refere (foi adicionado o fator "_Event-Cause" em que, no lugar do termo "*EventCause*", foi colocado o nome do evento para identi-

ficação em cada arquivo). Somado a isso, uma coluna foi adicionada nos arquivos do NTSB para identificar cada um dos incidentes como classificados com aquele evento. Assim, ao total, foram obtidos 13830 incidentes extraídos das classificações de tipos de evento.

Após isso, concentrou-se os arquivos para montar um CSV único contendo todos os incidentes aeronáuticos para facilitar os estudos. Nessa etapa, percebeu-se que alguns dados foram exportados com uma má configuração, sendo excluídos da base, resultando um total de 13789 casos resultantes.

Por fim, para finalizar a varredura, foi feita a integração dos dados para obter a duplicação de ocorrências que se enquadrariam em mais de um tipo de evento para serem deduplicadas e suas categorias serem agregadas na coluna de rótulo, preservando simultaneamente todos os demais campos estruturados e textuais do arquivo. Para isso, desenvolveu-se um código em *python* para otimizar a leitura dos dados e identificação dos incidentes (eles possuem número de identificação único) duplicados, exclusão da base de dados as repetições e incremento das categorias de evento. Ao final, obteve-se um total de 9015 incidentes aeronáuticos únicos da base do NTSB. Cada um dos incidentes contam com as seguintes colunas de informação : *NtsbNo*, *EventType*, *Mkey*, *EventDate*, *City*, *State*, *Country*, *ReportNo*, *N#*, *SerialNumber*, *HasSafetyRec*, *Mode*, *ReportType*, *OriginalPublishedDate*, *DocketOriginalPublishedDate*, *HighestInjuryLevel*, *FatalInjuryCount*, *SeriousInjuryCount*, *MinorInjuryCount*, *OnboardInjuryCount*, *OnGroundInjuryCount*, *ProbableCause*, *Findings*, *EventID*, *Latitude*, *Longitude*, *Make*, *Model*, *AirCraftCategory*, *AirportID*, *AirportName*, *AmateurBuilt*, *NumberOfEngines*, *EngineType*, *Scheduled*, *PurposeOfFlight*, *FAR*, *AirCRAFTDamage*, *WeatherCondition*, *Operator*, *BroadPhaseofFlight*, *ReportStatus*, *RepGenFlag*, *MostRecentReportType*, *DocketUrl*, *ReportUrl*, *Event Cause* (adicionado para identificação da classificação do incidente pelo NTSB).

3.3 Relatórios de Ocorrências Aeronáuticas

Após a consolidação do CSV com os metadados gerais dos incidentes — incluindo o identificador único de cada evento —, optou-se por realizar a extração integral dos relatórios em PDF diretamente do repositório do NTSB. Essa etapa foi planejada porque, embora o CSV já trouxesse campos estruturados relevantes, a análise de perigos no nível do aeródromo demanda o conteúdo textual completo para mais detalhes da narrativa do ocorrido, sobretudo a seção *Analysis* constante de cada relatório.

Dessa forma, projetou-se uma rotina de coleta automatizada que, a partir do identificador presente no CSV, reconstruía a URL do relatório: na base do NTSB, o endereço segue um padrão estável, variando essencialmente pelo número/ID do caso. Essa padronização viabilizou um ciclo programático simples: ler o ID, montar a URL, requisitar o PDF e salvá-lo localmente, registrando metadados de coleta.

Concluído o processo, foram obtidos 9.015 relatórios na íntegra a partir do repositório do NTSB, cada um associado ao seu registro no CSV por meio do identificador único e do caminho local do arquivo. Com isso, a base de trabalho passou a dispor, além dos campos estruturados originais, do conteúdo textual completo necessário às etapas seguintes de análise.

3.4 Incremento de Informações dos Relatórios na Base de Dados

Para enriquecer o conjunto analítico com dados, evidências textuais e atributos operacionais relevantes ao nível de aeródromo par melhor análise do cenário global envolvendo os acontecimentos, procedeu-se a um incremento sistemático da base.

Em primeiro lugar, para que se tivesse o relato geral do que ocorreu em cada um dos incidentes para ser uma fonte de informação futura, extraiu-se integralmente os textos da seção *Analysis* constantes nos relatórios completos do NTSB e vinculou-se cada narrativa ao respectivo registro do CSV-base. A vinculação foi feita por identificadores únicos do caso, preservando o encadeamento com a fonte oficial. O texto foi mantido integralmente para não se ter nenhuma perda. O resultado foi a criação de um novo campo no dataset contendo a narrativa completa por ocorrência. Essa escolha metodológica permite que, nas etapas seguintes, técnicas de análise textual (como a confirmação de perigos tendo como base o relato oficial) possam ser aplicadas diretamente sobre o conteúdo investigativo (extração de perigos, fatores contribuintes, menções a condições de pista), sem perda de contexto.

Em segundo lugar, agregou-se informações acerca do aeródromo sempre que disponíveis no relatório (os dados do NTSB apenas relata o nome dos aeroportos) e foram incorporadas ao CSV em campos explícitos. Quando presentes nos relatórios, foram mapeados itens de identificação do aeródromo, suas características e dados de pista.

Airport Information

Airport:	Lubbock Executive Airpark F82	Runway Surface Type:	Concrete
Airport Elevation:	3178 ft msl	Runway Surface Condition:	Dry
Runway Used:	17	IFR Approach:	None
Runway Length/Width:	5144 ft / 70 ft	VFR Approach/Landing:	None

FIGURA 3.3 – Exemplo de formatação padrão do bloco de informações acerca do aeroporto contida nos relatórios.

Para casos sem informação disponível, foi mantido valores nulos controlados sem imputação nos campos criados. Essa etapa gerou os seguintes campos de informação: `airport_code`, `airport_name`, `airport_state`, `runway_id`, `runway_length_ft`, `runway_surface`, `runway_condition_` `notes`.

Em ambas as frentes, o princípio foi não sobrescrever dados previamente existentes e não reduzir a narrativa a resumos; ao contrário, preservar o máximo de granularidade com proveniência clara. Por fim, executou-se verificações de consistência para assegurar que o enriquecimento fosse reprodutível, íntegro e pronto para alimentar as etapas analíticas subsequentes.

3.5 Criação de Cenários Hipotéticos

A fim de testar a inferência automática de identificação de falhas infraestruturais em aeródromos — tratadas aqui como perigos sob responsabilidade do gestor aeroportuário — foram elaborados cenários sintéticos com apoio do ChatGPT. Esses cenários funcionam como um conjunto de casos-teste controlados, nos quais cada texto descreve uma operação aeroportuária com contexto operacional completo (dimensões do aeródromo, número de pistas, tipo de operação IFR/VFR e período dia/noite, comprimento \times largura da pista, meteorologia predominante, tipos de aeronaves acomodadas) e introduz, de forma explícita ou sutil, um ou mais problemas de infraestrutura (p.ex., drenagem insuficiente, rachaduras e FOD recorrente, sinalização luminosa/horizontal degradada, perímetro sem cercamento/ingresso de fauna, RESA insuficiente/ausência de EMAS, sinalização horizontal confusa/placas desatualizadas, balizamento inoperante em baixa visibilidade).

Os cenários foram gerados com *prompts* estruturados, pedindo ao ChatGPT que mantivesse um formato padronizado (parágrafos concisos, presença das variáveis operacionais acima, linguagem técnica clara) e variasse somente os elementos relevantes ao perigo testado. Em seguida, cada rascunho foi analisado manualmente para que estivessem de acordo com o requerido e pudesse ser utilizado nos testes.

3.6 *Large Language Models* (LLM) para Análise dos Dados e Geração de Causas e Perigos

Possuindo um *dataset* enriquecido com as informações descritas anteriormente, o passo final é utilizar-se de uma ferramenta de análise e interpretação textual treinada para que possa ser identificada as causas prováveis de incidentes aeronáuticos à partir de informações fornecidas para tal ferramenta, bem como seu teste no ambiente de cenário de falha infraestrutural do aeródromo.

A partir desse objetivo, optou-se pela utilização de ferramentas de inteligência artificial na direção dos *Large Language Models* (LLMs). Tal aparato tecnológico foi selecionado pelo fato de que esses modelos têm capacidade de processar, compreender e gerar textos em linguagem natural, tornando possível a interpretação dos dados fornecidos, adequação aos cenários, e re-

torno de respostas voltadas ao objetivo de obter causas de acidentes aéreos.

Decidido a opção pelos modelos de linguagem de grande porte, necessitou-se fazer a busca e escolha de LLMs diante de três variáveis escolhidas: custo, disponibilidade e compatibilidade com o *hardware* utilizável. Em relação ao custo, optou-se por priorizar ferramenta gratuita por dois motivos: não se ter custo para treinar (visto que o processo para se obter o resultado gera erros, e a cada erro aumenta o valor gasto) bem como ser algo reprodutivo e adaptável sem necessidade de investimento monetário. Do ponto de vista da disponibilidade, precisa-se ter acesso de modo fácil e transparente aos modelos (independente da versão) para uso gratuito, seja em *hardware* próprio ou nuvem. Por último, a compatibilidade com o *hardware* é de extrema importância por dois motivos: o primeiro, e mais importante, é ter a condição necessária para conseguir treinar o modelo e, o segundo, conseguir otimizar o tempo de treinamento do LLM.

Para conseguir atender aos três requisitos, optou-se por obter o modelo através da plataforma do *Hugging Face* que consiste de uma plataforma dedicada especialmente ao compartilhamento de modelos de linguagem, *datasets* e inferências. Vale salientar que são modelos de linguagem abertos que visam a democratização da tecnologia. Assim, possuindo um amplo repositório de LLMs, o seguinte modelo foi escolhido: Meta Llama 3.1. Além dos requisitos, o modelo conta com a característica de ser focado na geração e inferência de texto, ponto principal do trabalho.

Dessa forma, procedeu-se com o emprego de modelos de LLM para ler e interpretar os dados coletados dos incidentes (principalmente as narrativas investigativas) e retornar possíveis causas (com validação em trechos do *Analysis*) e identificação de perigos infraestruturais através de treinamento com *Supervised Fine-Tuning* (SFT). O SFT consiste de uma técnica na qual se utiliza de um modelo pré-treinado e se ajusta os pesos com pares de entrada/saída rotulados (nesse caso, a *Analysis* gerar a *Probable Cause*) para que ele siga instruções e produza o formato desejado no seu domínio. Assim, o SFT ensina o modelo a responder utilizando exemplos corretos e direcionados tornando a ferramenta especializada em seu domínio.

Como etapa inicial para o treinamento dos modelos, fez-se a preparação dos dados na seguinte divisão: 70% para treinamento, 15% para validação e 15% para teste. Isso foi feito com base no procedimento de *machine learning* para que o fosse possível ajustar hiperparâmetros e melhor evitar os casos de *undeffiting* ou *overfitting*. Os dados de treino e validação foram salvos na extensão JSONL e os de teste em CSV para que, de forma facilitada, fossem consultados para análise e verificação.

Na próxima etapa, fez-se o treinamento propriamente dito com a utilização da técnica de *Low-Rank Adaptation* (LoRA), uma forma de ajuste fino de parâmetros com eficiência. Ela consiste de uma técnica que, ao invés de treinar todos os parâmetros, ela divide o modelo em matrizes e treina apenas aqueles necessários para o aprendizado de novas informações. Dessa forma, eles diminuem tanto o custo computacional quanto a possibilidade de um sobreajuste. Assim, objetivando os pilares de capacidade do modelo, estabilidade do treinamento e genera-

lização sob restrições de custo, os seguintes hiperparâmetros foram ajustados:

Hiperparâmetro	Descrição	Valores
LR (learning_rate)	Taxa de aprendizado do otimizador. É o principal controlador de estabilidade e velocidade de convergência.	$[1e^{-5} - 3e^{-4}]$
r	Define o tamanho do subespaço aprendido pelas matrizes LoRA (decomposição de baixo posto). Controla capacidade de adaptação.	{8,16,32}
lora_alpha	Fator de reescala aplicado à atualização LoRA. Atua como um ganho que ajusta a força da atualização aprendida. Controla o ganho efetivo das atualizações.	{16,32,64}
lora_dropout	<i>Dropout</i> aplicado nas projeções LoRA durante o treino. Regulariza a cabeça LoRA durante o treino.	{0.0,0.05,0.1}
max_seq_length	Tamanho máximo de sequência e/ou empacotamento de múltiplos exemplos no mesmo <i>context window</i> . <i>Context</i> curto corta trechos de <i>Analysis</i> ; longo aumenta custo de memória/tempo.	{1024,1536,2048}
warmup_ratio	Aquece a taxa de aprendizado do zero até o valor-alvo. Evita picos de gradiente nas primeiras iterações.	{0.03,0.05,0.1}

TABELA 3.1 – Hiperparâmetros que foram ajustados para o treinamento.

A escolha dos hiperparâmetros a serem explorados foi guiada por três objetivos: (I) capacidade de adaptação do modelo ao domínio (r, lora_alpha), (II) estabilidade e velocidade de convergência (learning_rate, warmup_ratio) e (III) cobertura informacional do contexto sem inviabilizar custo computacional (max_seq_length) com regularização para evitar sobreajuste (lora_dropout). Assim, priorizou-se variar os parâmetros que, segundo a literatura e prática de SFT/LoRA, mais afetam o equilíbrio viés-variância, a dinâmica de otimização e a fidelidade ao texto completo da *Analysis*.

Feito as variações dos hiperparâmetros com os dados de validação, utilizou-se a métrica de perda de validação (eval_loss) como indicador primário de ajuste ao corpus e estabilidade de treinamento. A eval_loss é calculada sobre um conjunto de validação não visto no treino e reflete a probabilidade média que o modelo atribui aos *tokens* corretos segundo o objetivo de aprendizado (SFT). Em termos práticos, perdas menores indicam melhor modelagem da distribuição dos textos de *Analysis* e maior regularidade da saída.

3.7 Testes do Modelo

Para a primeira análise do modelo treinado, utilizou-se os dados de teste para avaliar a assertividade em relação a identificar causas prováveis de ocorrência de incidente. A identificação de causas prováveis é importante pelo fato de que elas agregam fatores que podem ser considerados perigos para o aeródromo, além de que foram através de tais dados que o modelo foi treinado. Observou-se também o resultado do modelo base para se ter uma comparação de desempenho, utilizando os mesmos dados de *input* e instrução para a ferramenta.

O segundo experimento avaliou o desempenho do classificador baseado em LLM em cenários hipotéticos construídos com apoio do ChatGPT, comparando suas respostas com a percepção humana de referência. O objetivo foi medir se o modelo consegue detectar perigos relacionados à infraestrutura quando o contexto é controlado (texto padronizado, variáveis operacionais explícitas, ausência de ruído documental típico de relatórios), isolando assim o sinal semântico associado aos perigos sob governança do aeródromo. Criou-se, por fim, uma matriz de comparação de classificadores binários para análise dos resultados do LLM treinado e do seu modelo base.

4 Resultados e Discussões

4.1 Banco de Dados

A pesquisa e extração de dados feita no portal NTSB, somado à adição da coluna de classificação de cada incidente/acidente de acordo com o seu *Event Cause*, bem como a concentração e exclusão de dados repetidos, fez com que fosse gerado um CSV com 9015 casos aeronáuticos.

As colunas de informações adicionais do aeroporto e da narrativa na íntegra dos incidentes foram obtidas após a extração otimizada dos relatórios no portal do NTSB. O acervo dos relatórios gerou um repositório de 6,73 GB de conteúdo informacional dos eventos. Apenas dois casos não possuíam a narrativa, gerando ao fim do alinhamento um total de 9013 casos para análise. Abaixo segue um exemplar dos dados contidos no CSV utilizado para o estudo.

ProbableCause	Findings	BroadPhaseofFlight	EventCategory	airport	airport_elevation	runway_used	runway_length_width	runway_surface_type	runway_surface_condition	ifr_approach	ifr_approach_Landing	Analysis
The pilot's failure to Aircraft - Air Landing	RE, Loss of Control (Lake County 4734 ft msl		RE, Loss of Control (Lake County 4734 ft msl		17	5318 ft / 100 ft	Asphalt	Dry	None	None	Full stop	The pilot of the tailwheel
The pilot's failure to Personnel Is Landing	RE, Abnormal Runw: NUT TREE V 116 ft msl		RE, Abnormal Runw: NUT TREE V 116 ft msl		02	4700 ft / 75 ft	Asphalt	Dry	None	None	Full stop;Traffic pattern	The pilot of the tailwheel
The pilot's decision Aircraft - Air Landing	RE Merrill Field 143 ft msl		RE Merrill Field 143 ft msl		34	2640 ft / 75 ft	Asphalt	Dry	None	None	Full stop;Straight-in	The operator reported tha
The pilot's failure to Personnel Is Landing	RE, Abnormal Runw: Anoka Cour 912 ft msl		RE, Abnormal Runw: Anoka Cour 912 ft msl		27	5000 ft / 100 ft	Concrete	Airport ANE	Dry	None	Full stop;Traffic pattern	The pilot reported the acc
The pilot's failure to Aircraft - Air Takeoff	RE, Loss of Control (Shannons F 80 ft msl		RE, Loss of Control (Shannons F 80 ft msl		NE/SW	1400 ft / 100 ft	Gravel	Dry	None	None	None	The pilot of the tailwheel-
The pilot's failure to Aircraft - Air Landing	RE Toledo Sub 671 ft msl		RE Toledo Sub 671 ft msl		27	4807 ft / 50 ft	Asphalt	Dry	None	None	Full stop	The pilot reported that he
The pilot's failure to Personnel Is Landing	RE, Loss of Control (LAKE ELMO 932 ft msl		RE, Loss of Control (LAKE ELMO 932 ft msl		14/32	3504 ft / 75 ft	Asphalt	Dry	None	None	Full stop;Go around;Trafi	After the completion of th
The student pilot's fi Personnel Is Takeoff	RE, Collision on Taki WICHITA DV 1332 ft msl		RE, Collision on Taki WICHITA DV 1332 ft msl		01R/19L	7302 ft / 150 ft	Concrete	NTL ICT	Dry	None	None	The student pilot was plai
The pilot's failure to Personnel Is Landing	RE, Loss of Control (Avenger Fie 2380 ft msl		RE, Loss of Control (Avenger Fie 2380 ft msl		22	5658 ft / 75 ft	Asphalt	Dry	None	None	Straight-in	The pilot reported that aft
The pilot's failure to Aircraft - Air Landing	RE, Abnormal Runw: Eastern Ore 1494 ft msl		RE, Abnormal Runw: Eastern Ore 1494 ft msl		26	6301 ft / 150 ft	Asphalt	Pendleton PDT	Dry	None	Full stop;Traffic pattern	The pilot of the tailwheel
The pilots' failure to Environment Takeoff	RE Kayenta OV 5688 ft msl	05-23	RE Kayenta OV 5688 ft msl		05-23	7101 ft / 75 ft	Asphalt	Dry	None	None	None	The right-seated pilot stat
The pilot's failure to Aircraft - Air Takeoff	RE, GOOL SKY MEADO 2350 ft msl		RE, GOOL SKY MEADO 2350 ft msl		16	3100 ft / 100 ft	Grass/turf	Dry	None	None	None	The pilot of a tailwheel eq
The pilot's decision Aircraft - Air Landing	RE Hanvey Airfr 21 ft msl		RE Hanvey Airfr 21 ft msl		15	2672 ft / 32 ft	Asphalt	Dry	None	None	Full stop;Go around;Trafi	The pilot reported that he
The pilot's failure to Personnel Is Takeoff	RE, Loss of Control (San Luis Ob 212 ft msl		RE, Loss of Control (San Luis Ob 212 ft msl		29	6101 ft / 150 ft	Asphalt	Airport SBP	Dry	None	None	The pilot reported that, d
The pilot's failure to Personnel Is Landing	RE Henderson 2491 ft msl		RE Henderson 2491 ft msl		17R	6501 ft / 100 ft	Asphalt	Dry	None	None	Traffic pattern	The pilot reported that he
The failure of the left Aircraft - Air Landing	RE, Loss of Control (St. James M 1068 ft msl		RE, Loss of Control (St. James M 1068 ft msl		15	3999 ft / 75 ft	Asphalt	Dry	None	None	Full stop;Traffic pattern	The flight instructor and p
The pilot's failure to Personnel Is Landing	RE, Loss of Control (LAKE COUN 4734 ft msl		RE, Loss of Control (LAKE COUN 4734 ft msl		35	5318 ft / 100 ft	Asphalt	Dry	None	None	Straight-in	According to the pilot, up
The pilot's failure to Personnel Is Landing	RE FULLERTON 96 ft msl		RE FULLERTON 96 ft msl		24	3121 ft / 75 ft	Asphalt	Dry	None	None	Full stop	The pilot of the airplane r
The pilot's failure to Aircraft - Air Landing	RE, Loss of Control (SWANBOR 2394 ft msl		RE, Loss of Control (SWANBOR 2394 ft msl		09/27	3100 ft / 50 ft	Asphalt	Dry	None	None	Full stop;Traffic pattern	The pilot of the tailwheel
The pilot's failure to Aircraft - Air Takeoff	RE, Collision on Taki South Valle 4606 ft msl		RE, Collision on Taki South Valle 4606 ft msl		16	5862 ft / 100 ft	Asphalt	Dry	None	None	None	The flight instructor and s
The pilot's failure to Personnel Is Landing	RE, Loss of Control (A L Mangha 343 ft msl		RE, Loss of Control (A L Mangha 343 ft msl		36	5000 ft / 75 ft	Asphalt	Dry	None	None	Full stop	The pilot reported that the
The failure of the U.S Personnel Is Landing	RE, Abnormal Runw: RUSSELLV 409 ft msl		RE, Abnormal Runw: RUSSELLV 409 ft msl		07/25	5505 ft / 75 ft	Asphalt	Dry	None	None	Full stop;Traffic pattern	A flight of three U.S. Army
The pilots' overcorre Aircraft - Air Landing	RE, Loss of Control (ANDERSON 761 ft msl		RE, Loss of Control (ANDERSON 761 ft msl		23	6002 ft / 149 ft	Asphalt	Dry	None	None	Straight-in	The flight instructor and if
The pilot's loss of ai Aircraft - Air Landing	RE, Abnormal Runw: GRENADA M 207 ft msl		RE, Abnormal Runw: GRENADA M 207 ft msl		13	7000 ft / 150 ft	Asphalt	Dry	None	None	Full stop	The pilot reported that aft
The pilot's failure to Personnel Is Landing	RE, Loss of Control (Henderson 2490 ft msl		RE, Loss of Control (Henderson 2490 ft msl		35L	6501 ft / 100 ft	Asphalt	Dry	Unknown	None	Straight-in	The pilot reported that, d
The pilot's failure to Aircraft - Air Landing	RE, Collision on Taki Aero Estate 445 ft msl		RE, Collision on Taki Aero Estate 445 ft msl		27	3100 ft / 60 ft	Grass/turf	Dry	None	None	Full stop;Traffic pattern	The pilot reported that du

FIGURA 4.1 – CSV em Excel dos principais dados dos incidentes/acidentes aéreos para utilização na ferramenta.

A maioria das colunas do CSV original do NTSB foram ocultadas pelo fato de que não possuem informações relevantes para o estudo.

4.2 Aeródromos Hipotéticos e Seus Cenários

Para utilização no teste de inferência acerca de identificação de falhas infraestruturais em aeródromos que são consideradas um perigo para a atuação aeronáutica, os seguintes cenários foram criados com o auxílio do ChatGPT:

Tamanho: ~90 ha • **Pistas:** 1 • **Operação:** VFR diurno (sem noturno)

Pista: 1.200 × 23 m • **Meteo:** tropical úmido; pancadas vespertinas

Aeronaves: monomotores/bimotores leves, helicópteros

Problema: cercamento perimetral ineficaz; invasão de animais.

Narrativa ampliada: A patrulha reporta trilhas de capivaras cruzando perto da interseção com TWY A. Buracos no alambrado e porteira de serviço sem fechamento automático permitem incursão ocasional de cães.

Ocorrências cresceram 60% no semestre, com duas rejeições de decolagem de aeronaves leves ao avistar fauna em eixo. As armadilhas de captura e roçada do canteiro lateral reduziram parcialmente o risco, mas sem solução estrutural o problema persiste.

FIGURA 4.2 – Caso 1 de problema infraestrutural ligado à *wildlife*.

Tamanho: ~180 ha • **Pistas:** 1 • **Operação:** IFR/VFR diurno; noturno restrito

Pista: 1.900 × 30 m • **Meteo:** inverno seco; verão com trovoadas

Aeronaves: E175, ATR 72

Problema: RESA insuficiente; ausência de EMAS.

Narrativa ampliada: Excursões por pista molhada aproximam-se de talude curto após a cabeceira 22. A faixa de pista não atende plenamente as recomendações de extensão para a categoria de aeronaves que opera regularmente.

Um ATR 72 saiu para a grama sem danos, mas ficou a 8 m do declive. A ampliação de faixa depende de desapropriação de área privada e desvio de drenagem pluvial.

FIGURA 4.3 – Caso 2 de problema infraestrutural ligado à faixa de pista insuficiente para operação.

Tamanho: ~150 ha • **Pistas:** 1 • **Operação:** IFR (RNAV) e VFR

Pista: 1.650 × 30 m • **Meteo:** verão chuvoso; vento variável

Aeronaves: ATR 72, EMB-120, Caravan

Problema: sinalização horizontal confusa; placas de taxi com codificação desatualizada.

Narrativa ampliada: As marcas de "holding position" estão gastas nas interseções B1 e B2; placas ainda exibem nomenclaturas antigas, divergentes do AIP atual. Dois eventos de alinhamento indevido foram registrados em 90 dias, ambos com aeronaves regionais.

Tripulações relatam hesitação em pontos de espera, alongando o tempo de ocupação de interseção e elevando a chance de incursão sob visibilidade reduzida. O pacote de repintura e substituição de placas foi licitado, porém sem ordem de início.

FIGURA 4.4 – Caso 3 de problema infraestrutural ligado à sinalização horizontal e placas informacionais.

Tamanho: ~200 ha • **Pistas:** 1 • **Operação:** IFR/VFR dia/noite, RNAV (sem ILS)

Pista: 1.800 × 30 m • **Meteo:** chuva orográfica; teto baixo à tarde

Aeronaves: ATR 72, Dash 8, E170 (restrições)

Problema: sinalização luminosa e horizontal degradada.

Narrativa ampliada: Setores do balizamento de borda apagam intermitentemente após garoa; o circuito apresenta fuga à terra, com DRs atuando durante picos de umidade. Marcas de eixo e cabeceira estão desbotadas, confundindo percepção lateral sob chuva noturna.

As PAPI exibem brilho assimétrico (um box com intensidade inferior), induzindo ajustes tardios de trajetória. O cronograma de repintura está 10 meses atrasado e existem NOTAMs alternando trechos "u/s".

FIGURA 4.5 – Caso 4 de problema infraestrutural ligado à sinalização luminosa e horizontal.

Tamanho do aeródromo: ~240 ha • **Pistas:** 1 • **Operação:** IFR/VFR diurno-noturno, ILS CAT I

Pista (CxL): 2.300 × 45 m • **Meteo predominante:** ventos E-NE; chuvas fortes nov-mar

Aeronaves: ATR 72, E190, A320/B737 (peso limitado)

Problema: drenagem insuficiente (lâmina d'água).

Narrativa ampliada: Após pancadas de chuva com taxa >20 mm/h, formam-se poças persistentes entre os abalos B e C, sobretudo no terço médio da RWY 10/28. Relatos de "falsa sensação de frenagem" aparecem nos ASRs desde o último verão.

Patrulhas de pista demoram ~25–35 min para dispersar água com squeegee e caminhão vácuo, o que gera sequência de arremetidas em finais de tarde. O PIREP de um E190 indicou aquaplanagem característica (roda livre + vibração leve) a 90–100 kt. A vala perimetral está assoreada e os poços de visita próximos ao eixo ficam no limite de capacidade.

FIGURA 4.6 – Caso 5 de problema infraestrutural ligado à drenagem, resultando em acúmulos de água.

Tamanho: ~120 ha • **Pistas:** 1 • **Operação:** VFR diurno (IFR não autorizado)

Pista: 1.400 × 30 m • **Meteo:** nevoeiro matinal outono-inverno; vento de vale variável

Aeronaves: C172/182, King Air, PC-12, Caravan

Problema: ausência de torre de controle (somente AFIS parcial).

Narrativa ampliada: Nos fins de semana, o fluxo de instrução e táxi-aéreo dobra e a CTAF congestionada. Houve duas aproximações simultâneas em pernas do vento opostas por falha de coordenação em tráfego local/voos de subida.

Sem TWR, coordenações de backtrack em RWY única ficam lentas, alongando FIS-intervals e gerando espera no ar. O operador implantou marcações "self-announce" nos pontos de espera, mas incidentes de "base curta não reportada" persistem, sobretudo entre 10h–12h.

FIGURA 4.7 – Caso 6 de problema infraestrutural ligado à ausência de torre de controle.

Tamanho: ~310 ha • **Pistas:** 1 • **Operação:** IFR (RNAV/GNSS) e VFR dia/noite
Pista: 2.100 × 45 m • **Meteo:** vento NW-SE, tempo seco; rajadas no inverno
Aeronaves: E175/E195, A319, B737-700

Problema: rachaduras longitudinais e FOD recorrente.

Narrativa ampliada: Inspeções pós-banco de horas decolagens encontram lascas de asfalto (2–5 cm) entre A e D. Vibração perceptível em pousos “pesados” tem sido reportada a partir do ponto 1.200 m do cabeceira 10.

O levantamento PCI mostrou queda de 20 pontos no índice superficial em 18 meses, com selagem de trincas vencida. A varredora recolhe material granulado todos os dias úteis; houve dois pneus cortados em Q3. O PCN segue adequado, porém a macrotextura está irregular.

FIGURA 4.8 – Caso de 7 problema infraestrutural ligado à problemas de superfície da pista de pouso e decolagem.

Os cenários foram criados através de um *prompt* requisitando narrativas que descrevessem aeroportos com problemas de infraestrutura em setores diversos e que pudessem trazer características de tamanho do aeródromo, quantidade, comprimento e largura de pista, tipo de operação, predominância meteorológica e aeronaves acomodadas. Assim, os sete casos anteriores foram criados em uma descrição um tanto quanto ampla mas que continham algum problema específico.

Vale salientar que, como toda a codificação e treinamento do LLM foram feitos em inglês e com relatórios de mesma língua, traduziu-se também os cenários para estarem no mesmo idioma fonte do que foi utilizado.

4.3 LLM Treinado

O *script* de treinamento do modelo Meta foi desenvolvido em codificação *python* e utilizado no editor de código *Visual Studio Code* (VScode) para reprodução. Um ponto a ser salientado é de que, para a utilização do modelo Meta, necessitou-se fazer uma breve solicitação na plataforma do *Hugging Face* para o seu uso, mesmo ela sendo gratuita. O processo de solicitação e resposta foram rápidos, não sendo um empecilho para o seu uso ou replicação.

Na separação dos dados, obteve-se 6309 eventos para treinamento, 1352 para validação e 1352 para teste. Eles foram guardados de modo a fazer com que tanto o modelo base e o treinado pudessem utilizá-los. Para a análise, apenas os dados de *Analysis*, *ProbableCause* e os referentes ao aeroporto foram utilizados.

Para o treino, em cada um dos dados, também é adicionado um *prompt* de descrição para a função específica de identificador de causas prováveis de incidentes aeronáuticos.

Um total de 13 horas por teste de treinamento foi necessário para o modelo do Meta. Vale salientar que esse tempo é o período de um treinamento, podendo se perceber que é algo custoso tanto em relação ao *hardware* quanto à duração. Terminado o treinamento, os hiperparâmetros resultantes foram os seguintes:

Hiperparâmetro	Valor Adotado
LR (learning_rate)	$1.5e^{-4}$
r	16
lora_alpha	16
lora_dropout	0.05
max_seq_length	2048
warmup_ratio	0.05

TABELA 4.1 – Valores dos hiperparâmetros após treinamento e validação.

Eles foram obtidos fazendo a variação durante a validação do modelo no treinamento. A escolha foi feita baseada no menor valor do `eval_loss` à cada conjunto de passos. O treinamento resultou no valor de 1.141 como mínimo resultado obtido. Dessa forma, obteve-se o modelo treinado para a fase de testes.

4.4 Teste com Causas Possíveis

Assim como no treinamento, a parte de teste conta com a flexibilidade de instruções extras, ou seja, apenas é obrigatório ceder a informação da narrativa (o *Analysis* dos relatórios).

A partir dos dados de relatórios para teste, utilizou-se o modelo base e o treinado para obter inferências acerca de causas prováveis dos incidentes à partir das narrativas e dados coletados. Abaixo segue uma tabela de amostra com os dados resultantes da análise:

ProbableCause	PC_base	PC_ft
<i>The flightcrew's misjudged flare while landing.</i>	<i>Hard landing due to inadequate recovery from a previous hard landing event. Contributing factor: Failure by both pilots to report the initial hard landing event to maintenance personnel.</i>	<i>The pilot's failure to maintain aircraft control while landing. Contributing was the pilot's decision to land after experiencing a hard landing event.</i>
<i>The pilot's failure to maintain directional control during the takeoff roll which resulted in a runway excursion and nose-over event.</i>	<i>Pilot failure to maintain directional control due to inadequate rudder authority. Contributing factor may have been the pilot's perception that he lacked sufficient rudder authority, likely influenced by unfamiliarity with the aircraft's handling characteristics.</i>	<i>The pilot's failure to maintain directional control during the takeoff roll. Contributing was his lack of rudder authority.</i>
<i>The pilot's failure to extend the landing gear prior to touchdown. Contributing to the accident was the pilot's failure to utilize the pre-landing checklist.</i>	<i>Pilot's failure to ensure the landing gear was extended before touchdown resulted in a hard landing with substantial damage to the aircraft. Contributing factor may be inadequate adherence to standard operating procedures during critical phases of flight.</i>	<i>The pilot's failure to extend the landing gear before touchdown. Contributing to the accident was the pilot's inadequate use of checklists.</i>

TABELA 4.2 – Tabela de amostra do resultado das causas prováveis. As colunas de ProbableCause, PC_base e PC_ft se referem, respectivamente, às causas prováveis reais, às causas prováveis geradas pelo modelo base e às causas prováveis geradas pelo modelo treinado.

Para o estudo do rendimento da ferramenta, utilizou-se de três métricas associadas a geração textual:

- BLEU-1 : Precisão de unigramas (palavras individuais) do texto gerado em relação às referências, com penalidade de brevidade. Quanto maior, mais palavras iguais às da referência. Não “entende” sinônimos.
- ROUGUE-L : Mede precisão baseado no LCS (maior subsequência comum). Captura ordem global sem exigir contiguidade. Quanto maior o valor, maior a sobreposição estrutural (ainda lexical).
- SemSim : Mede proximidade de sentido mesmo com palavras diferentes. Quanto maior,

mais semelhante semanticamente.

Os resultados gerados pelas métricas seguem abaixo:

```
== MÉTRICAS (média) ==
Base | BLEU-1: 0.209 | ROUGE-L: 0.227 | SemSim: 0.642
FT   | BLEU-1: 0.155 | ROUGE-L: 0.222 | SemSim: 0.644
```

FIGURA 4.9 – Resultado das métricas BLEU-1, ROUGUE-L e SemSim.

A métrica de BLEU-1 mostrou menor sobreposição literal de palavras com a referência, ou seja, ele indica que há maior parafraseamento ou usando vocabulários diferentes. Uma análise visual nos resultados constatou que as saídas estão mais longas do que a referência de causas prováveis. Isso é também um fator que faz com que a métrica penalize, pois há uma relativa diferença de comprimento com a referência.

A métrica de ROUGUE-L obteve maior valor do que a BLEU-1 por justamente tratar de estrutura/ordem global, sendo coerente com mais paráfrase ou pequenas reordenações. Apesar disso, seu valor baixo acaba indicando o mesmo que foi constatado anteriormente de que o tamanho da resposta gerada está em comprimento relativamente diferente daquele de referência.

Por fim, a métrica SemSim mostrou o melhor resultado dentre as três. Apesar da menor coincidência lexical mostrada pelas outras métricas, a SemSim mostra Sinal que o FT aprendeu o conteúdo, mas não necessariamente o verbatim/formato esperado. Seu valor pouco acima dos 50% pode indicar que a ferramenta está pondo outros fatores como contribuintes além do que está nas causas prováveis reais. Isso foi constatado visualmente em alguns exemplos no fator de adicionar fatores contribuintes para a causa provável.

Nessa perspectiva, observou-se que a ferramenta de teste gerou bons resultados de identificação de modo semântico, acima de 0.64, mostrando que com ajustes acerca do campo textual resultante pode aumentar seus valores de métricas e se tornar mais assertiva. Um ponto que pode ser testado para melhora (não garantido tal resultado) seria uma limitação da estrutura textual da resposta, seja por número de caracteres, ou de sentenças.

4.5 Teste com Perigos Relacionados à Infraestrutura

Para o teste de inferência acerca de falhas de infraestrutura do aeroportos fictícios criados, optou-se por escolher 3 *prompts* de pedido de um usuário objetivando 3 cenários:

- Inferência acerca de um pedido amplo, que não há uma guia para o problema: *"Please analyze the following airport scenario and identify if there is a description of any infrastructural problem."*;

- Inferência acerca de um pedido específico e bem rotulado de infraestrutura: *"Please analyze the following airport scenario and identify whether there is an infrastructure issue related to airport perimeter fencing."*;
- Inferência acerca de um pedido específico mas que pode ser inferido com outros fatores: *"Please analyze the following airport scenario and identify whether it indicates the absence of an ATC tower (only AFIS/Information service available)."*

Os três cenários serviram para analisar o comportamento do LLM treinado (bem como do LLM base) quando é solicitado de diversas formas, visto que o usuário pode tanto não visar ser específico em sua requisição, como também não ter grande conhecimento acerca da engenharia de *prompt* para gerar um guia detalhado para o seu pedido na ferramenta. Os rótulos também foram utilizados em inglês para manter a linguagem do treinamento do LLM.

Ressalta-se que outros *prompts* de requisição de usuário foram feitos mas que não agregam discussões maiores do que os escolhidos. Apesar disso, os dados completos podem ser encontrados em drive.

Com os resultados, gerou-se uma matriz de confusão para, de modo mais objetivo e direto, obter a comparação acerca das afirmativas geradas pelos modelos. Abaixo segue o resultado, onde "M1" se refere ao modelo treinado e "M2" ao modelo base, de cada cenário, bem como sua análise para busca de informações acerca das respostas geradas pelo modelo e fazer observações em relação ao seu comportamento.

Pontua-se também que os valores marcados com o símbolo "*" referem-se à respostas em que o modelo não afirma diretamente a resposta, mas pelo contexto pode ser inferido a qual classificador quis se referir.

Matriz de Comparação de Classificadores Binários			
Please analyze the following airport scenario and identify if there is a description of any infrastructural problem.			
Caso	Correto (V/F)	M1	M2
1	V	V	V*
2	V	F	V
3	V	F	F
4	V	F	V
5	V	F	V
6	V	F	V
7	V	F	V

FIGURA 4.10 – Matriz de comparação de classificadores binários gerada para o primeiro cenário.

Cenário 1:

Para o modelo treinado, observou-se que apenas o caso infraestrutural envolvendo recursos físicos em dano (explicitamente citado), que é o caso da cerca com avarias e porteira de serviço sem fechamento automático, foi identificado como problema infraestrutural. Entretanto, apesar de responder, para os outros casos, que não se tratavam de narrativas de problemas infraestruturais, o modelo identificou as falhas específicas de cada caso, considerando a questão de infraestrutura como ora voltado para o ramo de manutenção ou reparo (no casos que envolvem materiais físicos desgastados ou falhando), ora voltado para o ramo operacional ou de procedimento (no caso da falta de torre de controle). À partir disso, concluiu-se que o modelo treinado, para uma inferência ampla, consegue identificar os problemas infraestruturais, mas entende de forma diferente o que seria o termo "infraestrutural".

Observando o treinamento, *prompt* de requisição e as respostas, duas possíveis causas foram analisadas nesse cenário:

- O modelo treinado está direcionado à ser o mais coerente com o que ele está exposto, visto que foi treinado para extrair dos textos as causas prováveis do incidente, ou seja, ele, em seu primeiro comportamento, irá analisar para onde o texto direciona de forma incisiva. Como há um único caso no qual está explícito uma falha crua e estrutural que está gerando um problema, apenas esse foi identificado como falha de infraestrutura;
- Ter utilizado o termo "*infrastructural*" ao invés de "*infrastructure*" (termo melhor padronizado para se referir aos casos de infraestrutura) no *prompt* de requisição. Os termos podem ser interpretados de modo diferente, podendo causar alguma confusão.

Já o modelo base considerou o todos os cenários como falha infraestrutural, com exceção do caso 3, o qual ele delimitou como falha de manutenção ou operacional. Um ponto a se observar do modelo base, é a questão de que no caso 4, que também envolve repintura, ele inicia falando que é um caso infraestrutural, mas ao fim retorna o termo de manutenção. Assim, casos parecidos podem gerar também confusão no que forçar em relação ao modelo base. Nessa perspectiva, observa-se que o modelo base conseguiu melhor gerenciar termos de significado amplo e gerais, abarcando maiores fatores para o termo de infraestrutura.

Matriz de Comparação de Classificadores Binários			
Please analyze the following airport scenario and identify whether there is an infrastructure issue related to airport perimeter fencing.			
Caso	Correto (V/F)	M1	M2
1	V	V	V
2	F	F	F
3	F	F	F
4	F	F	F
5	F	F	F
6	F	F	V
7	F	F	F

FIGURA 4.11 – Matriz de comparação de classificadores binários gerada para o segundo cenário.

Cenário 2:

Considerando o cenário no qual há uma referência direta a um fator infraestrutural no *prompt* de requisição, houve uma assertividade de 100% do modelo treinado. O modelo manteve o padrão de identificar os problemas de cada um dos cenários, mostrando que, mesmo em uma pergunta direcionada, ele permanece analisando todo o aeroporto.

Em relação ao modelo base, ele manteve a identificação dos perigos e boa relação de acertos. Entretanto, o caso referente ao qual houve erro, o modelo gerou uma resposta com bastante inferência sobre os dados, justificando a escolha por conta do tamanho do aeroporto.

Matriz de Comparação de Classificadores Binários			
Please analyze the following airport scenario and identify whether it indicates the absence of an ATC tower (only AFIS/Information service available).			
Caso	Correto (V/F)	M1	M2
1	F	V	V
2	F	F	V
3	F	V*	V*
4	F	F	F
5	F	F*	F
6	V	V	V
7	F	F	F*

FIGURA 4.12 – Matriz de comparação de classificadores binários gerada para o terceiro cenário.

Cenário 3:

O cenário 3 obteve um resultado bastante interessante. Quando se pergunta a falta de uma estrutura direcionada e não há nada específico na narrativa mas se pode inferir de acordo com as características do aeroporto e do cenários, os modelos inferem. Quando há na narrativa explicitamente a não existência da torre, ambos os modelos tem assertividade. Entretanto, quando não há, os modelos observam as características e operações do aeródromo para extrair dados que possam sustentar a existência ou a falta de uma torre de controle.

Um exemplo disso ocorreu foi no primeiro caso cuja narrativa citava rastros de capivara cruzando uma TWY. Isso fez com que os modelos inferissem que não há vigilância ou controle adequando naquele aeródromo. Outro fator também para o mesmo caso foi a citação da operação do aeródromo, que apenas opera VFR e diurno.

Assim, pode-se observar dois pontos que podem ter causado o erro dos modelos:

- Perguntas acerca de fatores infraestruturais que possam ser respondidos através de inferências acerca do modelo do aeroporto e não na narrativa, pode ser uma fonte de confusão para as respostas;
- Perguntas amplas que, de acordo com o contexto narrado, possa também ser inferido uma falta infraestrutural, também pode ser uma fonte de confusão.

Vale salientar que, um ponto interessante do modelo treinado é que sua assertiva é direta e permanente, mostrando um bom comportamento para treinamento.

Como sugestão para possíveis métodos para mitigação dos erros do modelo treinado (vale destacar que não é garantido de que isso irá melhorar sua assertividade), obteve-se os seguintes pontos:

- Utilização de termos mais adequados nos *prompts* de solicitação (vide o termo referente a infraestrutura);
- Direcionar o treinamento do modelo para casos de infraestrutura (vide que os casos de treinamento estavam muito ligados a fatores humanos);
- Contornar perguntas que possam ser inferidas pela característica do aeródromo através de omissão de dados quando for feita as requisições de esfera ampla.

5 Conclusões e Trabalhos Futuros

Analisando o mundo das ferramentas de inteligência artificial, observa-se a vastidão de possibilidades acerca do que se pode gerar com um treinamento. Nessa perspectiva, observa-se que as ferramentas analisadas nesse trabalho teste tiveram um resultado que mostram a necessidade e indispensável interpretação e ação do ser humano para identificação de perigos para aeródromos.

O resultado das métricas de desempenho textual, para o modelo treinado e o base, mostram que ainda necessita de melhora no treinamento e direcionamento para que resultem respostas mais próximas daquelas reais. Um ponto importante é que, mesmo com textos, expressões e tamanho de saída diferente daquelas esperadas, a semântica dos modelos mostrou bom potencial caso alguns ajustes sejam feitos.

Comparado com a interpretação humana, o modelo ainda necessita de melhora significativa, pois, por se tratar de uma ferramenta utilizada no âmbito de gerenciamento de segurança de aeroportos, é crítico sua assertividade. Desse modo, a ferramenta se mostra um modelo de apoio, que necessita da confirmação humana por estar inserida em um contexto tão sensível.

Como ferramenta de auxílio, ela se mostrou promissora no quesito de conseguir analisar fatores além da narrativa descritiva do problema infraestrutural (como, por exemplo, a dinâmica de funcionamento do aeródromo). E, como identificação de causas prováveis através de narrativas, mostrou-se uma ferramenta de automação útil, mas que precisa de supervisão.

Fazendo um adendo ao caso treinado, ele obteve alguns resultados abaixo do modelo base. Isso pode ter sido causado pela configuração da saída das respostas do modelo treinado que podem ser melhor ajustadas.

Como sugestão para trabalhos futuros, deixo aqui alguns pontos:

- Poder explorar outras ferramentas de LLM e, bem como combinações de hiperparâmetros para análise;
- No treinamento, criar uma base de dados direcionada para problemas infraestruturais para que seu escopo seja mais específico;
- Utilização de narrativa teste mais longa, menos detalhada e com termos aeronáuticos mais

específico para analisar a ferramenta diante desse cenário;

- Agregar dados de desempenho de aeronave pode ser algo relevante, visto que sua análise pode resultar em eventos que deveriam ocorrer arremetidas mas geraram incidentes.

Apesar da ferramenta única da ANAC, utilizada para análise de segurança dos aeródromos, não focar nas causas do incidentes e sim nas barreiras preventivas e mitigatórias, o presente trabalho é de suma importância para seu desenvolvimento pois direciona o caminho para identificação e análise das barreiras, pois para saber prevenir e mitigar, necessita conhecer o evento ao qual será necessário uma barreira.

Assim, conclui-se que o trabalho é um bom teste de partida para análise de melhores ferramentas de apoio para identificação de perigos que possam ser de responsabilidade infraestrutural do aeródromo e demonstra a importância de testar o desempenho dessas ferramentas em casos técnicos e específicos.

Referências

ABDULHAK, S. *et al.* Chatatc: Large language model-driven conversational agents for supporting strategic air traffic flow management. *In: Proceedings of the International Conference on Research in Air Transportation (ICRAT). Proceedings [...]*. Singapore: [s.n.], 2024. Trabalho apresentado em conferência

AGÊNCIA NACIONAL DE AVIAÇÃO CIVIL. **Regulamento Brasileiro da Aviação Civil - RBAC nº 153: Aeródromos – Operação, Manutenção e Resposta à Emergência**. July 2024. Com alterações posteriores até a Emenda nº 08, com vigência a partir de 15.07.2024. Available at: <https://www.anac.gov.br/assuntos/legislacao/legislacao-1/rbha-e-rbac/rbac/rbac-153>. Accessed: 31 dec. 2024.

AGÊNCIA NACIONAL DE AVIAÇÃO CIVIL. **Lista dos aeródromos de uso público por classes do RBAC nº 153 e por perfil operacional da Classe I**. 2025. Disponível em: <https://www.gov.br/anac/pt-br/assuntos/regulados/aeroportos-e-aerodromos/lista-dos-aerodromos-de-uso-publico-por-classes-do-rbac-no-153-e-por-perfil-operacional-da-classe-i>. Acesso em: 15 nov. 2025.

Agência Nacional de Aviação Civil (ANAC). **Ferramenta de Diagnóstico de Segurança Operacional para Gerenciamento de Risco de Aeródromos**. 2022. Disponível em: <https://www.gov.br/anac/pt-br/centrais-de-conteudo/aeroportos-e-aerodromos/guias-e-outras-publicacoes/ferramenta-de-diagnostico-de-seguranca-operacional-para-gerenciamento-de-risco-de-aerodromos/view>. Acesso em: 16 nov. 2025. Arquivo: *Ferramenta de Diagnóstico de Segurança Operacional_ARDM.xlsx* (171 KB). Atualizado em 03 ago. 2022 às 12h22.

Airbus. **A statistical analysis of commercial aircraft accidents 1958–2024**. 2024. <https://accidentstats.airbus.com/>. Acesso: 16 Nov. 2025

AUST, J.; PONS, D. A systematic methodology for developing bowtie in risk assessment: Application to borescope inspection. **Aerospace**, v. 7, n. 7, p. 86, 2020

Brasil. Centro de Investigação e Prevenção de Acidentes Aeronáuticos. **Painel SIPAER – Ocorrências aeronáuticas**. . <https://painelsipaer.cenipa.fab.mil.br/extensions/Sipaer/Ocorrencias.html>. Acesso em: 16 nov. 2025

Brasil. Centro de Investigação e Prevenção de Acidentes Aeronáuticos. **Relatórios Finais**. . <https://sistema.cenipa.fab.mil.br/cenipa/paginas/relatorios/relatorios.php>. Acesso em: 16 nov. 2025

DIEMERT, S.; WEBER, J. H. **Can Large Language Models Assist in Hazard Analysis?** 2023. ArXiv preprint. Disponível em: <https://arxiv.org/abs/2303.15473>

Flight Safety Foundation. **Toolkits & Resources.** .
<https://flightsafety.org/toolkits-resources/>. Acesso em: 16 nov. 2025

International Air Transport Association. **Recommendations for accident prevention.** .
<https://www.iata.org/en/publications/safety-report/accident-prevention-strategies/>. IATA Annual Safety Report – Accident Prevention Strategies. Acesso em: 16 nov. 2025

International Civil Aviation Organization. **Safety Reports.** .
<https://www.icao.int/safety-reports>. Acesso em: 16 nov. 2025

International Civil Aviation Organization. **Safety Management Manual (SMM). Doc 9859.** 4. ed. Montréal: ICAO, 2018

National Aeronautics and Space Administration. **ASRS Database Online – Query Wizard.** .
https://akama.arc.nasa.gov/ASRSDBOnline/QueryWizard_Filter.aspx. Acesso em: 16 nov. 2025

National Aeronautics and Space Administration. **Fault Tree Handbook with Aerospace Applications.** Washington, DC, 2002

National Transportation Safety Board. **CAROL – Case Analysis and Reporting Online.** .
<https://carol.ntsb.gov/>. Acesso em: 16 nov. 2025

RAY, A. T. *et al.* Examining the potential of generative language models for aviation safety analysis: case study and insights using the aviation safety reporting system (asrs). **Aerospace**, v. 10, n. 9, p. 770, 2023

RRC Training. **Fault Tree Analysis (FTA) and Event Tree Analysis (ETA).** . Course material, NEBOSH National Diploma – Unit A: Managing Health and Safety, Element A3 – Identifying Hazards, Assessing and Evaluating Risks. Material institucional, sem data

TÉCNICAS, A. B. de N. **ABNT NBR ISO/IEC 31010:2012 – Gestão de riscos – Técnicas para o processo de avaliação de riscos.** Rio de Janeiro: [s.n.], 2012. Norma técnica

TÉCNICAS, A. B. de N. **ABNT NBR ISO 31000:2018 – Gestão de riscos – Diretrizes.** Rio de Janeiro: [s.n.], 2018. Norma técnica

The Boeing Company. **Aerospace Safety Timeline – Safety Experience.** .
<https://www.boeing.com/safety/safetyexperience?boeingPath=%2F>. Acesso em: 16 nov. 2025

FOLHA DE REGISTRO DO DOCUMENTO

1. CLASSIFICAÇÃO/TIPO TC	2. DATA 17 de novembro de 2025	3. DOCUMENTO Nº DCTA/ITA/TC-081/2025	4. Nº DE PÁGINAS 55
5. TÍTULO E SUBTÍTULO: Aprimoramento da Ferramenta de Diagnóstico para Gerenciamento de Risco em Aeródromos: Modelagem e Implementação			
6. AUTOR(ES): João Victor dos Santos			
7. INSTITUIÇÃO(ÕES)/ÓRGÃO(S) INTERNO(S)/DIVISÃO(ÕES): Instituto Tecnológico de Aeronáutica - ITA			
8. PALAVRAS-CHAVE SUGERIDAS PELO AUTOR: Perigos; Riscos; Segurança Operacional; Ferramenta de Identificação; Modelos de Linguagem em Grande Escala.			
9. PALAVRAS-CHAVE RESULTANTES DE INDEXAÇÃO: Aeroportos; Segurança operacional; Gerenciamento de riscos; Análise de falhas; Desenvolvimento de software; Inteligência artificial; Computação; Transportes.			
10. APRESENTAÇÃO: (X) Nacional () Internacional ITA, São José dos Campos. Curso de Graduação em Engenharia Civil-Aeronáutica. Orientador: Prof. Dr. Evandro José da Silva. Publicado em 2025.			
11. RESUMO: O presente trabalho consiste no estudo de relatórios de incidentes e acidentes aeronáuticos para o desenvolvimento de uma ferramenta de análise e predição das causas prováveis e fatores contribuintes para a ocorrência do respectivo evento, bem como indicar possíveis falhas infraestruturais em casos fictícios. Utilizou-se inteligência artificial através de Modelos de Linguagem em Grande Escala (LLM) treinado e comparado com a sua base original para análise de resultados. Como modelo teste, obteve-se uma ferramenta que, à partir de narrativas, consegue identificar causas de incidentes e problemas infraestruturais, mas que necessita de supervisão, pois ainda incorre em inferências que podem ser não factuais. Nessa perspectiva, observa-se que o estudo representa um marco inicial em uma ferramenta teste e o modelo apresenta ganho na agilidade do processamento e entendimento situacional, mas necessita da supervisão humana para garantir seu resultado.			
12. GRAU DE SIGILO: (X) OSTENSIVO () RESERVADO () SECRETO			