Ramicés dos Santos Silva

COMPLIANCE APOIADO POR IA: MELHORIAS COM USO DE IA GENERATIVA PARA O PROCESSO DE CONFORMIDADE EM SEGURANÇA E PROTEÇÃO DE DADOS.

Itajaí (SC), dezembro de 2025

UNIVERSIDADE DO VALE DO ITAJAÍ
CURSO DE MESTRADO ACADÊMICO EM COMPUTAÇÃO APLICADA

COMPLIANCE APOIADO POR IA: MELHORIAS COM USO DE IA GENERATIVA PARA O PROCESSO DE CONFORMIDADE EM SEGURANÇA E PROTEÇÃO DE DADOS.

por

Ramicés dos Santos Silva

Dissertação apresentada como requisito parcial à obtenção do grau de Mestre em Computação Aplicada.
Orientadora: Anita Maria da Rocha Fernandes, Dra.

Itajaí (SC), dezembro de 2025

FOLHA DE APROVAÇÃO

Compliance Apoiado por IA: Melhorias com uso de IA Generativa para o Processo de Conformidade em Segurança e Proteção de Dados

Ramices dos Santos Silva.

Esta Dissertação foi julgada adequada para obtenção do Título de Mestre em Computação Aplicada, Área de Concentração Computação Aplicada e aprovada em sua forma final pelo Programa de Mestrado Acadêmico em Computação Aplicada da Universidade do Vale do Itajaí.

Jordan Passinato Sausen
Assinado de forma digital por Jordan Passinato Sausen
Dados: 2026.03.14 15:23:13 -03'00'

Prof. Jordan Passinato Sausen, Dr.
Coordenador do Curso de Mestrado em Computação Aplicada

Apresentado perante a Banca Examinadora composta pelos Professores:

Anita Maria da Rocha Fernandes
Assinado de forma digital por Anita Maria da Rocha Fernandes
Dados: 2026.03.03 10:52:50 -03'00'

Prof. Anita Maria da Rocha Fernandes, Dra. (UNIVALI)
Orientador

UNIVALI
Assinado de forma digital por Rudimar Luis Scaranto Dazzi:61166820904
Dados: 2026.03.03 10:50:34 -03'00'

Prof. Rudimar Luis Scaranto Dazzi, Dr. (UNIVALI)
Membro

gov.br
Documento assinado digitalmente
RAIMUNDO CELESTE GHIZONI TEIVE
Data: 03/03/2026 10:44:09-0300
Verifique em https://validar.iti.gov.br

Prof. Dr. Raimundo C. Ghizoni Teive (UNIVALI)
Membro

e0a9352e-da57-4899-aaa9-d96f729e59d8
Assinado de forma digital por e0a9352e-da57-4899-aaa9-d96f729e59d8
Dados: 2026.03.03 10:37:48 -03'00'

Prof. Michelle Wangham, Dra (UNIVALI)
Membro

gov.br
Documento assinado digitalmente
ROGERIO CID BASTOS
Data: 03/03/2026 10:19:23-0300
CPF: ***.425.409-**
Verifique as assinaturas em https://v.ufsc.br

Prof. Rogério Cid Bastos, Dr (UFSC)
Membro

Dedico este trabalho ao meu filho Eduardo, pela paciência e compreensão nos inúmeros momentos em que precisei estar trabalhando, à minha família, pelo apoio incondicional em todas as horas e especialmente à Professora Dra. Anita Maria da Rocha Fernandes, por sua amizade, orientação segura, pelo incentivo incondicional e por acreditar neste trabalho desde o início.

"Nossa tecnologia, nossas máquinas, são parte da nossa humanidade. Nós as criamos para estender a nós mesmos, e isso é o que há de único nos seres humanos." Ray Kurzweil

COMPLIANCE APOIADO POR IA: MELHORIAS COM USO DE IA GENERATIVA PARA O PROCESSO DE CONFORMIDADE EM SEGURANÇA E PROTEÇÃO DE DADOS

Ramicés dos Santos Silva

Dezembro / 2025

Orientadora: Anita Maria da Rocha Fernandes, Dra.

Área de Concentração: Computação Aplicada

Linha de Pesquisa: Inteligência Artificial

Palavras-chave: Compliance em Segurança da Informação, Inteligência Artificial Generativa, Automação de Relatórios.

Número de páginas: 226

RESUMO

O trabalho aborda a crescente demanda por soluções eficientes para o processo de compliance em segurança da informação e proteção de dados em empresas de médio porte, que enfrentam escassez de profissionais especializados para lidar com requisitos regulatórios cada vez mais complexos. Nesta dissertação, é proposta e implementada uma solução baseada em Inteligência Artificial Generativa para automatizar a geração de relatórios de compliance, personalizados para diferentes públicos na organização, como gestores, equipes técnicas e responsáveis pela conformidade. O objetivo principal é aumentar a eficiência, escalabilidade e confiabilidade desses processos, reduzindo o tempo e o esforço necessários para produzir relatórios e melhorando a clareza das informações apresentadas. A pesquisa adota uma revisão sistemática da literatura sobre aplicações de IA generativa em segurança da informação e compliance, seguida do desenvolvimento de uma arquitetura modular composta por integração de dados provenientes de um SIEM, processamento desses dados e geração automática de relatórios com uso de modelos de linguagem de grande escala. A solução foi avaliada com base em um framework multidimensional que considerou qualidade linguística, alinhamento ao propósito (persona-*fit*) e desempenho dos modelos de IA, a partir de dados reais de um ambiente monitorado. Os resultados indicam que a automação da criação de relatórios libera especialistas para atividades mais estratégicas, aumenta a consistência das informações e facilita a adaptação da organização a mudanças regulatórias, contribuindo para uma cultura de segurança mais acessível e sustentável.

AI-DRIVEN COMPLIANCE: ENHANCEMENTS USING GENERATIVE AI FOR SECURITY AND DATA PROTECTION COMPLIANCE PROCESSES

Ramicés dos Santos Silva

December / 2025

Advisor: Anita Maria da Rocha Fernandes, Ph.D.

Area of Concentration: Applied Computer Science

Research Line: Artificial Intelligence

Keywords: Information Security Compliance, Generative Artificial Intelligence, Automation of Reports.

Number of pages: 226

ABSTRACT

This dissertation examines the growing demand for efficient solutions to support information security and data protection compliance process in medium-sized companies, which often face a shortage of specialized professionals capable of handling increasingly complex regulatory requirements. It proposes and implements a solution based on Generative Artificial Intelligence to automate the generation of compliance reports tailored to different stakeholders within the organization, such as managers, technical teams, and compliance officers. The primary objective is to increase the efficiency, scalability, and reliability of compliance processes by reducing the time and effort required to produce reports and by improving the clarity of the information presented. The research adopts a systematic literature review on generative AI applications in information security and compliance, followed by the development of a modular architecture integrating data from a Security Information and Event Management (SIEM) system, processing these data, and generating automated reports using large language models. The proposed solution was evaluated using a multidimensional framework that considered linguistic quality, alignment with audience needs (persona-fit), and performance of the AI model, based on real data from a monitored environment. The results show that automating report generation frees specialists to focus on more strategic activities, increases the consistency of information, and facilitates organizational adaptation to regulatory changes, contributing to a more accessible and sustainable security culture.

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

AI	*Artificial Intelligence*
API	*Application Programming Interface*
CNN	*Convolutional Neural Network*
COBIT	*Control Objectives for Information and Related Technologies*
CSV	*Comma-Separated Values*
GAN	*Generative Adversarial Networks*
GPT	*Generative Pre-trained Transformer*
GDPR	*General Data Protection Regulation*
IA	Inteligência Artificial
ISO	*International Organization for Standardization*
LGPD	Lei Geral de Proteção de Dados
LLM	*Large Language Model*
LMH	*Language modeling head*
LSTM	*Long Short-Term Memory*
MCA	Mestrado em Computação Aplicada
MSE	*Mean Square Error*
NIST	*National Institute of Standards and Technology*
PLN	Processamento de Linguagem Natural
PRISMA	*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*
RSL	Revisão Sistemática da Literatura
SGSI	Sistema de Gestão de Segurança da Informação
SIEM	*Security Information and Event Management*
TI	Tecnologia da Informação
UNIVALI	Universidade do Vale do Itajaí
VAE	*Variational Autoencoder*

SUMÁRIO

1 INTRODUÇÃO

A sociedade está cada vez mais dependente dos sistemas de informação. Com isto, os desafios também crescem na mesma proporção, pois ataques e crimes cibernéticos estão cada vez mais sofisticados e registram aumentos constantes em abrangência (European Council, 2023). É evidente que as defesas cibernéticas atualmente implantadas não acompanham as técnicas de ataque usadas pelos invasores, já que a maioria das defesas ainda dependem exclusivamente da intervenção humana.

Como a atuação humana ainda é fortemente necessária em todas as camadas de proteção, entre os desafios mais críticos de hoje enfrentados pelas organizações modernas e hiper conectadas, está a falta de profissionais de cibersegurança qualificados necessários para apoiar os programas de segurança e proteção de dados (Singh et al., 2023). Harvard Business Review (2019) estimou que a lacuna da força de trabalho de cibersegurança representa 3,4 milhões até 2023. Para as equipes de segurança, as restrições de tempo é o fator que mais influencia negativamente na satisfação no trabalho, o que caracteriza a sobrecarga em face da escassez de mão de obra (ISC2, 2024).

Então, com o aumento de foco na cibersegurança, aumentou a demanda por profissionais de cibersegurança nas organizações públicas e privadas e, simultaneamente, alargou a lacuna entre a falta de profissionais qualificados e empregos disponíveis. Ainda em relação a dificuldade de mão de obra reportado pelo mercado, gerenciar as expectativas das partes interessadas e o senso de responsabilidade para com os clientes e equipe é frequentemente o aspecto mais estressante do trabalho (IBM, 2023). A dificuldade de mão de obra é então agravada pois os profissionais também precisam fazer a interface de comunicação com outras equipes, que muitas vezes têm menos experiência com os conceitos, mas que também fazem parte da estrutura organizacional.

Os grupos que normalmente fazem parte da estrutura organizacional são: a) Departamento Jurídico, *Compliance, C-Level* e Auditorias internas e externas (Sacomano Neto, Escrivão Filho, 2000). Para estas estruturas, o tempo, já escasso, ainda é consumido na atividade de confecção de relatórios de compliance, de forma a caracterizar a conformidade por meio de evidências documentais, onerando os diversos perfis profissionais. Tal situação faz com que o cuidado com a linguagem e necessidade de explicações tragam mais complexidade ao trabalho dos profissionais de segurança.

É por isso que as organizações, precisam de analistas sofisticados e treinados para classificar, analisar e executar processos relacionados à segurança e proteção de dados, isso exige conhecimento de linguagens de consulta específicas e configurações relevantes para cada produto (Grinnel, 2023). Entretanto, esta tarefa difícil que compõe o conjunto de conhecimentos necessários ainda é acrescida de talento para traduzir tudo isso em uma linguagem corporativa para outros times. Sendo assim, se observa a movimentação das gigantes das tecnologias mundiais, que têm buscado contribuir com esta dificuldade de mercado.

Dentre as soluções que despontam estão as da Microsoft e da Google. A Microsoft, por exemplo, está apostando no terreno com sua ferramenta generativa AI Security, onde busca aumentar a capacidade de seus produtos de segurança de dados para análises e respostas integradas e profundas. Ao integrar a IA Generativa GPT-4 ao seu produto, o Security Copilot, a Microsoft espera trabalhar com empresas para identificar mais facilmente atividades maliciosas. Também neste mesmo caminho, a Google anunciou a Cloud Security AI Workbench baseada em um modelo de linguagem de inteligência artificial denominada Sec-PaLM (TechCrunch+, 2023).

Estudos propõem o desenvolvimento de agentes autônomos inteligentes de defesa cibernética (AICA) utilizando técnicas de Aprendizado Computacional Distribuído (ACyD) para enfrentar as crescentes cibernéticas. Como apresentado por Oreyomi e Jahankhani (2022) em *Challenges and Opportunities of Autonomous Cyber Defence (ACyD) Against Cyber Attacks*, onde algoritmos de aprendizado de máquina, como Naïve Bayes, Redes Neurais Convolucionais (CNN – Convolutional Neural Network), LSTM (*Long-Short Term Memory*) e Máquina de Suporte Vetorial com Rede Neural, podem ser utilizados para detectar e neutralizar ataques cibernéticos gerados automaticamente, demonstrando o potencial da IA (Inteligência Artificial) e do Aprendizado de Máquina (AM) para estratégias de segurança e proteção de dados.

A contribuição percebida da IA neste contexto é trazida por, George, George e Martin (2023), em seu trabalho, *ChatGPT and the Future of Work*: *A Comprehensive Analysis of AI´s Impacto n Jobs and Employment*, afirma que medida que a IA avança, tecnologias como o ChatGPT exercem forte influência em empregos e setores diversos. De um lado, a IA propicia aumentos na produtividade e eficiência, automatizando tarefas repetitivas e liberando trabalhadores para atividades mais complexas e criativas, auxilia na tomada de decisões ao analisar grandes volumes de dados com rapidez e precisão, gerando *insights* valiosos para o sucesso empresarial.

Diante deste cenário observa-se que é de extrema importância o desenvolvimento de soluções que possam contribuir de forma positiva para os pontos citados. Com isso, este trabalho propõe o uso de Inteligência Artificial Generativa como ferramenta de apoio a confecção de relatórios de compliance onde a origem dos dados sejam eventos de segurança vindos de plataforma de mercado e que já atuam de forma bastante automatizada, mas que apresentam como resultado uma linguagem que depende de interpretação especializada, mão de obra escassa.

1.1 PROBLEMA DE PESQUISA

As empresas de médio porte enfrentam um cenário de crescente complexidade regulatória em segurança da informação e proteção de dados, impulsionado por normas e leis como a LGPD, o GDPR e padrões internacionais de segurança.

Ao mesmo tempo, há uma escassez significativa de profissionais especializados em cibersegurança e compliance, o que dificulta a condução de avaliações contínuas de riscos e a elaboração de relatórios técnicos e executivos consistentes.

Na prática, muitos processos de compliance permanecem fortemente manuais e dependentes de poucos especialistas, resultando em documentos extensos, pouco padronizados e, por vezes, difíceis de serem interpretados por diferentes públicos, como gestores, equipes técnicas e responsáveis pela conformidade.

Esse contexto limita a capacidade das organizações de monitorar sua postura de segurança de forma sistemática, responder rapidamente a vulnerabilidades e demonstrar conformidade perante órgãos reguladores e auditorias externas.

Para que evidenciem os resultados deste trabalho, foram elaboradas hipóteses de pesquisa que ao final da análise demonstram as contribuições propostas. A seguir serão apresentadas as hipóteses e evidências para as questões de pesquisa formuladas.

Questão de Pesquisa 1:

Q1: Como processo de automação por meio de IA pode contribuir para empresas de médio porte em sua missão de desenvolver melhor programas de segurança e proteção de dados?

Hipóteses:

- **H1a:** A implementação da solução de geração automatizada de relatórios de compliance apoiada por IA Generativa reduz o tempo necessário para produzir esses relatórios, em comparação com a elaboração manual por profissionais de cibersegurança.

- **H1b:** Os relatórios de compliance gerados pela solução baseada em IA Generativa apresentam clareza e compreensibilidade, na percepção de especialistas.

As evidências esperadas para aceitar tais hipóteses e responder a esta questão de pesquisa, referem-se a dados quantificáveis que demonstrem uma redução no tempo de conclusão das tarefas de compliance e no consumo de recursos após a implementação da automação orientada por IA; e feedback qualitativo de profissionais de compliance indicando melhor compreensão e capacidade de tomada de decisão devido aos relatórios gerados por IA.

Questão de Pesquisa 2:

Q2: A abordagem proposta contribui para melhorar a escalabilidade de atuação dos profissionais de segurança da informação, ainda escassos no mercado?

Hipóteses:

H2a: A adoção de uma abordagem baseada em IA Generativa reduz o esforço necessário dos profissionais de cibersegurança em tarefas rotineiras de elaboração de relatórios de compliance, aumentando o tempo disponível para atividades mais estratégicas.

H2b: A utilização da solução proposta permite ampliar o volume de relatórios de compliance gerados em um determinado intervalo de tempo, em comparação com a elaboração manual por especialistas, contribuindo para a escalabilidade das operações de compliance.

Como evidências esperadas para aceitar tais hipóteses e responder a esta questão de pesquisa, espera-se dados que mostrem uma diminuição no tempo gasto por profissionais de segurança da informação em tarefas rotineiras de compliance após a introdução de ferramentas com tecnologia de

IA; e um aumento no número de casos de compliance tratados por equipes de segurança da informação após a adoção de soluções orientadas por IA.

Questão de Pesquisa 3:

Q3: O uso de IA generativa gera resultados confiáveis para as equipes envolvidas no compliance de segurança e proteção de dados?

Hipóteses:

H3a: Os relatórios de compliance gerados pela solução baseada em IA Generativa apresentam nível adequado de precisão e fidelidade aos dados de entrada do SIEM, segundo a avaliação de especialistas, atendendo aos padrões esperados para uso em auditorias e processos de compliance.

H3b: Os relatórios gerados pela solução apresentam consistência interna e baixa taxa de erros factuais e gramaticais, de acordo com as métricas do framework de avaliação adotado (dimensões de desempenho do modelo de IA e qualidade linguística).

As evidências esperadas para aceitar estas hipóteses e responder a esta questão de pesquisa, referem-se à validação de relatórios de compliance gerados por IA contra padrões e regulamentações de compliance estabelecidos; e uma diminuição no número de erros e omissões relacionados à compliance após a implementação de ferramentas orientadas por IA.

A partir das hipóteses definidas em cada uma das questões de pesquisa e as evidências esperadas criou-se a hipótese de pesquisa que caracteriza de forma geral este trabalho. A seguir é apresentada esta **Hipótese Geral** e suas evidências.

HG: A implementação de uma solução de geração de relatórios de compliance apoiada por IA Generativa em empresas de médio porte contribui para a otimização das atividades de compliance em cibersegurança e proteção de dados, ao reduzir o tempo e o esforço para elaboração de relatórios, aumentar a escalabilidade do processo e produzir relatórios com níveis adequados de clareza, precisão e consistência, alinhados às necessidades de diferentes personas organizacionais.

Esta hipótese abrange principais características que se espera obter com este estudo:

- Otimização das atividades de compliance: refere-se à automação de tarefas rotineiras e a geração de relatórios personalizados com IA Generativa, a fim de liberar os profissionais especializados para se concentrarem em atividades mais complexas e estratégicas, otimizando o tempo e os recursos empregados em compliance.

- Maior eficiência: a automatização de processos e a geração de relatórios precisos e consistentes contribuem para a redução de erros e omissões, agilizando a tomada de decisões e o gerenciamento de riscos, com impacto positivo na eficiência geral das atividades de compliance.

- Escalabilidade: a capacidade da solução de lidar com um grande volume de dados e gerar relatórios personalizados para diferentes áreas da organização permite que as empresas de médio porte escalem suas operações de compliance de forma eficiente, acompanhando seu crescimento e demandas.

- Confiabilidade: a utilização de técnicas de IA Generativa para gerar relatórios confiáveis e precisos, baseados em dados e padrões de compliance estabelecidos.

Diante desse cenário, o problema de pesquisa que orienta esta dissertação é:

Como utilizar técnicas de Inteligência Artificial Generativa para automatizar e melhorar a geração de relatórios de compliance em segurança da informação e proteção de dados, de modo a torná-los mais eficientes, consistentes e adequados a diferentes públicos dentro da organização?

Para tal implementou-se a solução descrita a seguir.

1.1.1 SOLUÇÃO PROPOSTA

Para enfrentar esse problema, este trabalho propõe o desenvolvimento e a avaliação de uma solução de automação de relatórios de compliance apoiada por Inteligência Artificial Generativa.

A solução é baseada em uma arquitetura modular que integra dados provenientes de um Sistema de Gestão de Eventos e Informações de Segurança (SIEM), processa esses dados para extrair informações relevantes e utiliza modelos de linguagem de grande escala para gerar relatórios personalizados para diferentes personas (como CEO, Compliance Officer e equipe de TIC).

A proposta busca reduzir o esforço manual na elaboração de relatórios, aumentar a padronização e a clareza das análises apresentadas e permitir que especialistas concentrem seu tempo em atividades de maior valor estratégico, como a priorização de ações de mitigação e a definição de políticas de segurança.

A solução proposta é uma arquitetura modular com o objetivo de desacoplar a implementação em componentes de aquisição e integração de dados, processamento e geração dos relatórios. Os componentes principais serão: integração de dados, processamento de dados e geração de relatórios. Estes componentes serão apresentados e detalhados no capítulo relativo ao desenvolvimento.

A proposta utiliza APIs (*Application Program Inteface*) para interagir com um SIEM (Sistemas de Gestão de Vulnerabilidades) de mercado e recebe documentos com o conteúdo de *frameworks* de boas práticas em compliance de segurança e proteção de dados, facilitando a integração com os sistemas existentes da organização. O cuidado com as informações de contexto da IA Generativa permite que o modelo trate a novas informações e requisitos ao longo do tempo a partir do fornecimento de novos dados que caracterizam o ambiente avaliado. A solução poderá, em trabalhos futuros, ser integrada a outras ferramentas de segurança da informação, como plataformas de gerenciamento de riscos e incidentes, para fornecer uma visão holística da postura de segurança da organização.

Ao automatizar a geração de relatórios personalizados, a solução demonstrou oferecer diversos benefícios para empresas de médio porte:

- Maior eficiência e produtividade: Reduz o tempo e o esforço despendidos na criação manual de relatórios, liberando profissionais de segurança para se concentrarem em tarefas mais estratégicas.

- Visibilidade aprimorada da postura de segurança: Fornece aos stakeholders uma visão clara e precisa da postura de segurança da organização, facilitando a tomada de decisões estratégicas.

- Ao aliar a expertise em segurança da informação com o poder da IA Generativa, a solução oferece uma ferramenta para empresas de médio porte que permite fortalecerem seus programas de compliance.

1.1.2 DELIMITAÇÃO DE ESCOPO

Este trabalho teve como foco principal o desenvolvimento e avaliação de uma solução que melhore o processo de compliance de segurança. Especificamente visa contribuir para o processo de geração de relatórios de compliance em segurança da informação e proteção de dados, utilizando técnicas de inteligência artificial generativa (IA Generativa).

A solução proposta vem automatizar a criação de relatórios personalizados, adaptáveis às necessidades de diferentes personas, como alta gerência, profissionais de segurança da informação e gestores de conformidade a partir de informações vindas de um sistema de gestão de eventos e informações de segurança SIEM *(Security Information and Event Management)*, assim como *frameworks* de boas práticas de *compliance* em segurança.

A arquitetura da solução proposta será restrita a três módulos principais:

- Integração de Dados: Este módulo será responsável por coletar dados de diversos sistemas, como SIEMs e frameworks de boas práticas de compliance, como ISO 27001 e LGPD (Lei Geral de Proteção de Dados Pessoais)

- Processamento de Dados: Este módulo processará os dados coletados, extraindo informações relevantes como vulnerabilidades identificadas, eventos de segurança, status de conformidade com frameworks e regulamentações, entre outros.

- Geração de Relatórios: Este módulo utilizará técnicas de IA Generativa para gerar relatórios personalizados, adaptando a linguagem e o conteúdo dos relatórios às necessidades de cada persona.

É importante destacar que esta dissertação se concentrou em um protótipo de validação para a melhoria proposta que possa gerar relatórios verificáveis para as estruturas organizacionais de médio porte. Como o objetivo foi desenvolver uma solução integrável, as interações de entrada e saída da solução serão basicamente APIs que recebem e geram os dados propostos, sem o intuito de desenvolvimento de uma aplicação de interface com usuário final, e ao final são executadas em um pipeline que orquestra os módulos criados utilizando linguagem python. Todo o ambiente será detalhado mais adiante no capítulo de Desenvolvimento.

1.1.3 JUSTIFICATIVA

Este trabalho justifica-se pelo cenário de crescente complexidade na cibersegurança e pela escassez de profissionais especializados em segurança da informação e proteção de dados, impondo desafios significativos para empresas de médio porte na implementação e manutenção de sólidos programas de compliance. A solução proposta, baseada no uso de Inteligência Artificial Generativa, visa automatizar a geração de relatórios de compliance em segurança, personalizados para diferentes tipos de personas dentro da organização, como alta gerência, profissionais de segurança da informação e usuários finais.

Com a solução buscou-se otimizar os processos de compliance. Sua relevância reside na redução do tempo e dos recursos necessários para a geração de relatórios de conformidade, além de melhorar a clareza e a compreensibilidade dessas informações. Isso permite uma melhor tomada de decisão e gerenciamento de riscos na organização. Ademais, a automação de tarefas rotineiras libera os profissionais de segurança para se concentrarem em atividades mais estratégicas, aumentando a eficiência e a produtividade.

O trabalho é provido de complexidade considerável já que existe a necessidade de integrar dados de diversas fontes, como Sistemas de Gestão de Vulnerabilidades e frameworks de boas práticas de compliance, tais como a ISO 27001 e a Lei Geral de Proteção de Dados Pessoais. Ademais, avaliação de comportamento dos modelos de IA Generativa requer a cuidadosa curadoria de um conjunto de dados que engloba relatórios de compliance reais, informações sobre vulnerabilidades e eventos de segurança, bem como a linguagem específica de cada perfil de usuário na organização, além do desempenho de modelos de LLM distintos já disponíveis no mercado.

Além da relevância prática para organizações de médio porte, este trabalho também apresenta uma contribuição científica ao explorar um nicho ainda pouco abordado na literatura: o uso de IA Generativa para a geração automática de relatórios de compliance baseados em dados de SIEM e orientados a diferentes personas organizacionais.

A revisão sistemática da literatura apresentada no Capítulo 3 evidenciou que a maior parte dos estudos foca em aplicações de IA para detecção de ameaças, resposta a incidentes ou análise de riscos, com pouca ênfase na automação da comunicação de compliance em formatos adaptados a públicos específicos, como CEOs, gestores de TIC e compliance officers.

Os trabalhos identificados na revisão sistemática mostram avanços em monitoramento, dashboards e apoio à decisão, mas não abordam de forma integrada: (i) o uso de dados reais de SIEM, (ii) a geração automatizada de relatórios textuais voltados a múltiplas personas e (iii) a avaliação sistemática desses relatórios em dimensões como qualidade linguística, alinhamento ao propósito e desempenho do modelo de IA.

Essa lacuna fundamenta a proposta desta dissertação e justifica a necessidade de investigar em que medida uma solução baseada em IA Generativa pode otimizar o processo de compliance em cibersegurança em empresas de médio porte.

A solução proposta se mostrou adequada a empresas de médio porte que enfrentam dificuldades na gestão de compliance em segurança da informação e proteção de dados. A automatização da geração de relatórios personalizados permite que essas empresas ampliem suas operações de compliance de forma eficiente, acompanhando seu crescimento e demandas. O foco em empresas de médio porte foi decidido em decorrência de três fatores principais.

Primeiro, esse segmento enfrenta, de forma mais acentuada, a escassez de especialistas em cibersegurança e compliance, não dispondo de equipes extensas como grandes corporações.

Segundo, essas organizações estão igualmente sujeitas a exigências regulatórias relevantes (como LGPD e normas de segurança), mas muitas vezes não possuem estrutura para manter processos de compliance altamente manualizados.

Terceiro, soluções comerciais avançadas, frequentemente direcionadas ao grande porte, podem ser financeiramente inviáveis para esse público, o que torna particularmente relevante avaliar alternativas baseadas em IA Generativa que ampliem eficiência e escalabilidade com menor custo incremental de compliance.

1.2 OBJETIVOS

1.2.1 OBJETIVO GERAL

O objetivo geral desta dissertação é otimizar o processo de compliance em cibersegurança e proteção de dados em empresas de médio porte por meio da concepção, implementação e avaliação de uma solução de geração automatizada de relatórios de compliance apoiada por Inteligência Artificial Generativa.

1.2.2 OBJETIVOS ESPECÍFICOS

a) Realizar uma revisão sistemática da literatura sobre aplicações de IA generativa em cibersegurança, proteção de dados e automação de relatórios de compliance.

b) Identificar requisitos funcionais e não funcionais para a geração automatizada de relatórios de compliance a partir de dados de SIEM.

c) Propor e implementar uma arquitetura modular que integre coleta de dados, processamento e geração de relatórios com modelos de linguagem de grande escala.

d) Definir um framework de avaliação multidimensional para analisar a qualidade linguística, o alinhamento ao propósito (persona-fit) e o desempenho dos modelos de IA na geração dos relatórios.

e) Avaliar a solução proposta com base em dados reais de um ambiente monitorado, discutindo benefícios, limitações e oportunidades de evolução.

1.3 METODOLOGIA

A pesquisa adota uma abordagem aplicada e exploratória, combinando revisão sistemática da literatura, desenvolvimento de artefato tecnológico e avaliação empírica por meio de uma abordagem hipotético-dedutiva. Foram formuladas hipóteses relacionadas ao impacto da IA Generativa sobre as atividades de compliance em cibersegurança e proteção de dados, derivadas implicações observáveis dessas hipóteses e, em seguida, confrontadas com evidências empíricas obtidas a partir da solução implementada.

Do ponto de vista da natureza dos dados, o estudo combina métodos qualitativos e quantitativos. A componente quantitativa está presente nas métricas de avaliação dos relatórios (pontuações nas dimensões de qualidade linguística, alinhamento ao propósito e desempenho do modelo de IA, além de indicadores como tempo estimado de geração de relatórios). A componente qualitativa envolve a análise de conteúdo dos relatórios gerados, a interpretação dos julgamentos dos especialistas participantes da avaliação e a discussão das implicações práticas e científicas dos resultados obtidos.

Na primeira etapa, foi conduzida uma revisão sistemática da literatura para identificar aplicações de IA generativa em cibersegurança e compliance, bem como métodos e métricas de avaliação de relatórios gerados por modelos de linguagem.

Em seguida, foi desenvolvida uma arquitetura de referência para geração automatizada de relatórios a partir de dados de SIEM, incluindo módulos de integração, processamento e geração de texto com modelos de IA generativa.

Por fim, a solução foi avaliada por meio de um framework de avaliação que considera dimensões de qualidade linguística, alinhamento ao propósito e desempenho do modelo de IA utilizando dados reais e a percepção de especialistas para compor as métricas.

1.4 ESTRUTURA DA DISSERTAÇÃO

Esta dissertação está estruturada em cinco capítulos. No Capítulo 1, são apresentados o contexto, o problema de pesquisa, a solução proposta, os objetivos e a metodologia adotada. O Capítulo 2 descreve a fundamentação teórica em compliance em segurança da informação e proteção de dados, Inteligência Artificial Generativa e automação aplicada ao compliance. O Capítulo 3 apresenta os trabalhos correlatos e a revisão sistemática da literatura, destacando lacunas e oportunidades de pesquisa. O Capítulo 4 detalha o desenvolvimento da solução proposta, incluindo arquitetura, implementação, testes e metodologia de avaliação. Por fim, o Capítulo 5 traz as conclusões, contribuições, limitações e sugestões para trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

Neste Capítulo são fornecidos os conceitos e referências que sustentam a solução proposta de geração automatizada de relatórios de compliance apoiada por IA Generativa, apresentada nos capítulos seguintes.

A crescente dependência das organizações em relação aos sistemas de informação tem impulsionado a necessidade de estratégias robustas para mitigar riscos cibernéticos e garantir a conformidade regulatória. Aqui serão abordados temas fundamentais, incluindo a definição e importância do compliance, os principais *frameworks* e normas aplicáveis, além de discutir como a IA Generativa pode ser utilizada para otimizar processos específicos no domínio da segurança da informação.

Antes de detalhar as contribuições da IA Generativa, é essencial compreender o cenário atual do compliance em segurança e proteção de dados, seus principais desafios e melhores práticas. A seguir, apresenta-se uma análise sobre compliance em segurança e proteção de dados, destacando sua relevância no ambiente corporativo moderno.

2.1 COMPLIANCE EM SEGURANÇA E PROTEÇÃO DE DADOS

A crescente dependência das organizações em relação aos sistemas de informação, os quais são essenciais para suas operações, trouxe consigo uma série de riscos inerentes que têm exigido cada vez mais atenção à segurança da informação e à proteção de dados.

Nesse contexto, a conformidade regulatória, também conhecida como *compliance*, surge como um elemento crucial para garantir a segurança e a integridade dos ativos de informação. (Matas e Keegan, 2020)

O *compliance* é definido como o conjunto de ferramentas e ações de promoção da conduta ética e da conformidade com leis, normas e regulamentos aplicáveis a uma organização (Medeiros e Codignoto, 2022).

Para atingir a conformidade, as empresas devem implementar uma série de medidas, como a definição de políticas, procedimentos e controles internos alinhados aos requisitos legais e regulatórios relevantes para suas atividades.

Dentre as principais normas e frameworks utilizados para a gestão da segurança da informação e da proteção de dados, destacam-se a ISO 27001, o NIST *Cybersecurity Framework*, o COBIT, o GDPR e a LGPD.

2.1.1 PRINCIPAIS NORMAS E FRAMEWORKS DE COMPLIANCE

A ISO 27001 é uma das normas internacionais mais renomadas para a implementação de sistemas de gestão da segurança da informação, pois fornece uma estrutura abrangente com requisitos detalhados para o estabelecimento, implementação, operação, monitoramento, análise crítica, manutenção e melhoria contínua de um SGSI (Sistema de Gestão de Segurança da Informação) (Calder, 2023; Disterer, 2013).

Já o NIST *Cybersecurity Framework* é um conjunto de diretrizes, normas e melhores práticas amplamente utilizadas para a gestão da cibersegurança nas organizações. Desenvolvido pelo Instituto Nacional de Padrões e Tecnologia dos Estados Unidos, o *framework* fornece uma abordagem estruturada para identificar, detectar, proteger, responder e recuperar-se de ameaças e incidentes de segurança. O modelo é flexível e adaptável, que pode ser aplicado a organizações de diversos setores, ajudando-as a avaliar e melhorar sua postura de cibersegurança. Ao adotar o NIST *Cybersecurity Framework*, as empresas podem implementar ações eficazes de identificação, proteção, detecção, resposta e recuperação, mitigando riscos e garantindo a continuidade de seus negócios (PASCOE, 2023)

Outro conjunto de boas práticas bastante difundido é o COBIT que se destaca como um framework amplamente adotado pelas organizações para a governança e gestão de TI (Tecnologia da Informação), uma vez que visa alinhar as atividades de TI aos objetivos estratégicos do negócio, garantindo a conformidade com leis, regulamentos e políticas internas aplicáveis (Blokdyk, 2020; Disterer, 2013)

No que tange à proteção de dados pessoais, dois regulamentos têm se destacado, o GDPR e a LGPD. O GDPR, sigla para *General Data Protection Regulation*, é um conjunto de normas da União Europeia que estabelece regras para a coleta, tratamento, armazenamento e compartilhamento de dados pessoais de cidadãos europeus (Oreyomi e Jahankhani, 2022; Disterer, 2013).

Já a LGPD, Lei Geral de Proteção de Dados do Brasil, alinha-se ao GDPR e visa regular a proteção de dados pessoais no país, impondo obrigações e responsabilidades às organizações que lidam com tais informações.

De fato, a implementação de um programa efetivo de *compliance* em segurança da informação e proteção de dados é fundamental para as empresas e instituições, pois garante não apenas a conformidade legal, mas também a mitigação de riscos, a proteção dos ativos de informação e a preservação da confiança dos clientes, parceiros e demais partes interessadas.

2.1.2 IMPORTÂNCIA DA CONFORMIDADE REGULATÓRIA

A conformidade regulatória em segurança da informação e proteção de dados é de grande relevância para as organizações. Além de evitar sanções e multas impostas por órgãos reguladores, que podem gerar impactos financeiros e reputacionais significativos, a adesão às normas e regulamentos aplicáveis também contribui para a construção de uma cultura de segurança da informação e proteção de dados dentro das empresas. Isso porque a estrutura de governança estabelecida pelas organizações deve buscar alinhar os interesses e responsabilidades de proprietários, partes interessadas e gestores, garantindo o respeito mútuo e a efetividade do programa de compliance (Saqib et al., 2018).

Nesse sentido, a implementação de um programa robusto de compliance requer a incorporação, internalização e disseminação da cultura da integridade dentro da organização. A integridade é a fundação que assegura que as empresas operem de forma ética e cumpram suas obrigações legais e regulatórias" (Medeiros e Codignoto, 2022).

De fato, a adoção de práticas de compliance em segurança da informação e proteção de dados não se trata apenas de uma obrigação legal, mas também de uma estratégia fundamental para a sustentabilidade e a competitividade das organizações no longo prazo. Assim, é essencial que as empresas se empenhem na implementação de programas robustos de compliance, alinhados às melhores práticas e normas internacionais, a fim de garantir a conformidade regulatória, mitigar riscos, proteger seus ativos de informação e preservar a confiança de seus *stakeholders*.

2.1.3 DESAFIOS E MELHORES PRÁTICAS EM COMPLIANCE

Apesar dos benefícios evidentes da conformidade regulatória em segurança da informação e proteção de dados, a implementação de um programa de compliance efetivo ainda enfrenta diversos desafios.

Um dos principais desafios consiste na necessidade de lidar com a chamada "falha humana". A maior parte das violações de segurança da informação está relacionada a fatores comportamentais, como a falta de conscientização e de treinamento adequado dos funcionários. MATAS & KEEGAN, 2020)

Assim, é essencial que as organizações invistam continuamente em estratégias de conscientização e capacitação de seus colaboradores, a fim de consolidar uma cultura organizacional comprometida com a segurança da informação e a proteção de dados. Além disso, a complexidade e a constante evolução dos requisitos legais e regulatórios também representam um desafio significativo para as empresas.

Para enfrentar esses desafios, as organizações devem adotar melhores práticas de compliance, como (Al Kalbani et al., 2016):

- Implementar um programa de compliance integrado e alinhado aos objetivos estratégicos da empresa;
- Designar um profissional ou equipe responsável pela coordenação das atividades de compliance;
- Realizar avaliações periódicas de riscos e conformidade; e
- Desenvolver e atualizar constantemente políticas, procedimentos e controles de segurança da informação e proteção de dados.

Os desafios e exigências descritos evidenciam a necessidade de mecanismos mais eficientes e escaláveis de geração de relatórios, alinhados à proposta.

2.2 INTELIGÊNCIA ARTIFICIAL GENERATIVA E SUA APLICABILIDADE

A Inteligência Artificial Generativa é um subcampo da inteligência artificial que se concentra na criação de novos conteúdos, como texto, imagens e vídeos, a partir de dados existentes. Utilizando modelos avançados de aprendizado de máquina, a IA Generativa é capaz de identificar padrões complexos e gerar resultados que imitam características humanas. Esses modelos são treinados em grandes volumes de dados, permitindo que aprendam a distribuição subjacente dos dados e gerem novas instâncias que apresentam características estatísticas similares aos dados de treinamento. (Goodfellow et al., 2014).

A trajetória da IA Generativa está intrinsecamente ligada ao desenvolvimento da inteligência artificial como um todo. Desde os primeiros estudos na década de 1950, com as ideias pioneiras de Alan Turing sobre máquinas que poderiam "pensar", a IA evoluiu significativamente.

Um marco importante foi o desenvolvimento das redes neurais artificiais na década de 1980, que permitiram a criação de modelos capazes de aprender representações hierárquicas dos dados. Mais recentemente, a introdução de arquiteturas transformadoras, como o GPT (*Generative Pre-trained Transformer)* da OpenAI, revolucionou a capacidade dos modelos de processar e gerar linguagem natural, possibilitando avanços significativos em tarefas como tradução automática e geração de texto (Vaswani et al., 2017).

Grande parte dos modelos generativos utiliza aprendizado não supervisionado ou auto-supervisionado, nos quais os algoritmos aprendem representações úteis dos dados sem necessidade de rótulos explícitos (Leão, 2023). Além disto, modelos generativos aprendem distribuições de probabilidade sobre os dados. O objetivo é modelar P(x), onde x representa dados observáveis (ex.: uma frase ou imagem), permitindo gerar novas amostras a partir dessa distribuição (Leão, 2023).

2.2.1 PRINCIPAIS MODELOS DE IA GENERATIVA

A Inteligência Artificial Generativa tem evoluído significativamente, resultando no desenvolvimento de modelos capazes de criar conteúdo original, incluindo texto, imagens, vídeos e código-fonte. Esses avanços foram impulsionados pelo aprimoramento das arquiteturas de redes neurais e pelo crescimento da capacidade computacional, permitindo a criação de modelos mais sofisticados e eficientes (Goodfellow et al., 2014).

Entre os principais modelos de IA Generativa tem-se os Autoencoders; GANs (*Generative Adversarial Networks)*; *Transformers* e LLMs (*Large Language Models*); e *Diffusion Models*. Os *Transformers* e LLMs terão um maior detalhamento por servirem de base para o entendimento dos modelos de IA generativa que foram utilizados como ferramenta principal deste trabalho.

2.2.1.1 Autoencoders

Um *autoencoder* é um tipo de rede neural não supervisionada projetada para aprender representações compactas (codificações) de dados, normalmente com o objetivo de compressão, redução de dimensionalidade ou geração de dados (Baker, 2024).

Na IA Generativa, o *autoencoder* é utilizado para aprender uma representação latente dos dados que pode ser usada para reconstruir ou gerar novas amostras semelhantes às originais (Clark, 2024).

Um *autoencoder* possui duas partes principais (Baker, 2024):

- **Encoder**: Transforma a entrada x em uma representação latente z.

$$z = f_\theta(x)$$

- **Decoder**: Reconstrói a entrada a partir da codificação z.

$$\hat{x} = g_\phi(z)$$

O objetivo é minimizar o erro entre a entrada original x e a reconstrução $\hat{x}$ geralmente por meio da função de perda de reconstrução, como o erro quadrático médio (MSE – *Mean Square Error*) (Baker, 2024).

Embora os *autoencoders* tradicionais não sejam intrinsecamente modelos generativos (pois não amostram diretamente uma distribuição), variantes como o *Variational Autoencoder* (VAE) são projetadas especificamente para esse fim (Baker, 2024).

A seguir tem-se o passo a passo de como funciona o VAE:

1. **Encoder** aprende uma distribuição probabilística q(z|x), normalmente gaussiana.

2. **Latent space** é amostrado: $z \sim N(\mu(x), \sigma^2(x))$

3. **Decoder** reconstrói x a partir dez, gerando dados novos ao amostrar novos vetores latentes.

Isso permite gerar novos dados a partir de pontos arbitrários no espaço latente, o que caracteriza a capacidade generativa do modelo.

Como exemplo de aplicações com *autoencoders* tem-se:

- Geração de imagens: aprendem a representar rostos, objetos ou cenas e gerar novas variações.
- Interpolação semântica: criação de transições suaves entre duas imagens, atuando no espaço latente.
- Detecção de anomalias: ao comparar reconstruções, é possível detectar padrões fora do esperado.
- Geração de áudio e sinais biomédicos.

2.2.1.2 GANs - Generative Adversarial Networks

GAN (*Generative Adversarial Network*) é uma arquitetura de aprendizado profundo criada por Goodfellow et al. (2014) com o objetivo de gerar novos dados realistas a partir de uma distribuição de probabilidade aprendida. É um dos modelos mais populares da IA generativa, especialmente para geração de imagens, vídeos, música e textos visuais.

Uma GAN é composta por duas redes neurais artificiais que competem entre si (Goodfellow et al, 2014):

- Gerador (G): Cria dados falsos (ex: imagens) a partir de vetores aleatórios z.
- Discriminador (D): Distingue entre dados reais e falsos (gerados pelo G).

As etapas do processo que uma GAN executa são:

1. O gerador (G) recebe como entrada um vetor de ruído z~N(0,1) e tenta produzir uma amostra G(z) semelhante aos dados reais.

2. O discriminador (D) recebe amostras reais e geradas, e aprende a classificá-las como real (1) ou falsa (0).

3. G tenta enganar D, enquanto D tenta não ser enganado.

4. Esse processo é repetido até que G aprenda a produzir amostras indistinguíveis dos dados reais.

A ideia do processo de execução das GANs é (Langr e Bok, 2019):

1. A rede neural geradora analisa o conjunto de treinamento e identifica os atributos dos dados

2. A rede neural discriminadora também analisa os dados do treinamento inicial e distingue os atributos de forma independente

3. A geradora modifica alguns atributos de dados adicionando ruído (ou alterações aleatórias) a determinados atributos

4. A geradora passa os dados modificados para a discriminadora

5. A discriminadora calcula a probabilidade de que a saída gerada pertença ao conjunto de dados original

6. A discriminadora fornece algumas orientações à geradora para reduzir a randomização do vetor de ruído no próximo ciclo

Algumas aplicações comuns das GANs são: geração de rostos realistas; criação de estilo artístico (ex: transformar foto em pintura); aprimoramento de imagem (super-resolução); criação de mundos virtuais e animações; *Deepfakes*; design de moda e objetos com prototipagem visual.

2.2.1.3 Diffusion Models

Diffusion models (Modelos de Difusão) são uma classe de modelos generativos que aprendem a criar dados revertendo um processo gradual de adição de ruído (Sanseviero et al, 2024). Eles têm se destacado pela alta qualidade na geração de imagens, sendo a base de sistemas como DALL·E 2,

Stable Diffusion e Imagem. São utilizados principalmente para a geração de imagens e tarefas de visão computacional. Eles aprendem a adicionar ruído a uma imagem, e aprendem posteriormente a reverter esse processo para reconstruir uma imagem limpa (Sanseviero et al, 2024, Clark, 2024).

A intuição por trás dos modelos de difusão é inspirada na física, tratando os pixels como as moléculas de uma gota de tinta se espalhando em um copo de água ao longo do tempo. Assim como o movimento aleatório das moléculas de tinta acabará por levar à sua dispersão uniforme no vidro, a introdução aleatória de ruído em uma imagem acabará resultando no que parece ser estática de TV. Ao modelar esse processo de difusão e, em seguida, aprender de alguma forma a revertê-lo, um modelo de IA pode gerar novas imagens simplesmente "reduzindo o ruído" de amostras de ruído aleatório (Vemula, 2024).

No treinamento, os modelos de difusão difundem gradualmente um ponto de dados com ruído aleatório, passo a passo, até que seja destruído, então aprendem a reverter esse processo de difusão e reconstruir a distribuição de dados original (Clark, 2024).

Um modelo de difusão treinado pode gerar novos pontos de dados que se assemelham aos dados de treinamento, bastando eliminar o ruído de uma amostra inicial aleatória de ruído puro. Conceitualmente, isso é semelhante a um autocodificador de redução de ruído, no qual as imagens ruidosas atuam como variáveis latentes (Vemula, 2024).

Transformar diretamente o ruído aleatório em uma imagem coerente é extremamente difícil e complexo, mas transformar uma imagem ruidosa em uma imagem *um pouco menos ruidosa* é relativamente fácil e simples. Portanto os modelos de difusão formulam o processo de difusão reversa como uma transformação incremental, passo a passo, de uma distribuição simples (como o ruído gaussiano) para uma distribuição mais complexa (como uma imagem coerente) (Vemula, 2024; Sanseviero et al, 2024).

O processo de treinamento e, em seguida, a implementação de uma difusão pode ser dividida em três estágios principais (Sanseviero et al, 2024):

- **O processo de difusão para a frente,** em que uma imagem do conjunto de dados de treinamento é transformada em ruído puro, geralmente uma distribuição gaussiana.

- **O processo de difusão reversa,** em que o modelo aprende o inverso de cada etapa anterior no processo de difusão original para a frente.

- **Geração de imagens,** em que o modelo treinado coleta amostras de uma distribuição aleatória de ruído e as transforma em uma produção de alta qualidade usando o processo de difusão reversa, no qual ele aprendeu a eliminar o ruído de uma amostra aleatória de ruído gaussiano.

Alguns exemplos de modelos difusos são:

- *Stable Diffusion*, que gera imagens a partir de texto com controle visual;

- DALL·E 2, da OpenAI, que cria imagens criativas a partir de descrições em linguagem natural;

- Imagen da Google, que gera imagens foto realistas com texto; e

- AudioDiffusion, que gera música ou falas.

2.2.1.4 Transformers e os Large Language Models (LLMs)

Transformer é uma arquitetura de rede neural profunda e paralelizável, introduzida por Vaswani et al. (2017) no artigo "*Attention is all you need*", que revolucionou o campo de processamento de linguagem natural (PLN) e IA generativa. Esse trabalho introduziu o uso de uma estrutura que revolucionou a forma como o aprendizado profundo estava sendo conduzido até então: mecanismos de atenção que compreendem contexto.

Na época, outros tipos de redes neurais profundas eram utilizados para resolver tarefas de IA, como as Redes Neurais Recorrentes, e as LSTMs (*Long Short Term Memory*) (Sanseviero et al, 2024). Cada modelo baseado nestas redes precisa de um treinamento supervisionado para solucionar uma tarefa específica, como classificação de texto ou análise de sentimento, por exemplo, e funcionam em uma arquitetura *sequence-to-sequence*, ou seja, trabalham sequencialmente, analisando uma palavra após a outra (BAKER, 2024). Já no Transformers, os mecanismos de atenção trouxeram o grande benefício da paralelização. Com a paralelização, a sentença inteira é vista ao mesmo tempo. Isso traz uma série de vantagens, tais como uma maior velocidade de treinamento; eficiência no reconhecimento de dependência entre palavras; melhoria no reconhecimento de padrões; e capacidade de analisar problemas não sequenciais (Sanseviero et al, 2024).

A arquitetura base de um *Transformer* é composta por dois blocos principais*: Encoder* e *Decoder.* O *Encoder* recebe a entrada (ex: sentença em inglês) e codifica essa entrada em vetores de contexto ricos em significado. Já o *Decoder*, usa os vetores codificados para gerar uma sequência (ex: tradução para francês, texto, imagem etc.) (Sanseviero et al, 2024; Backer, 2024).

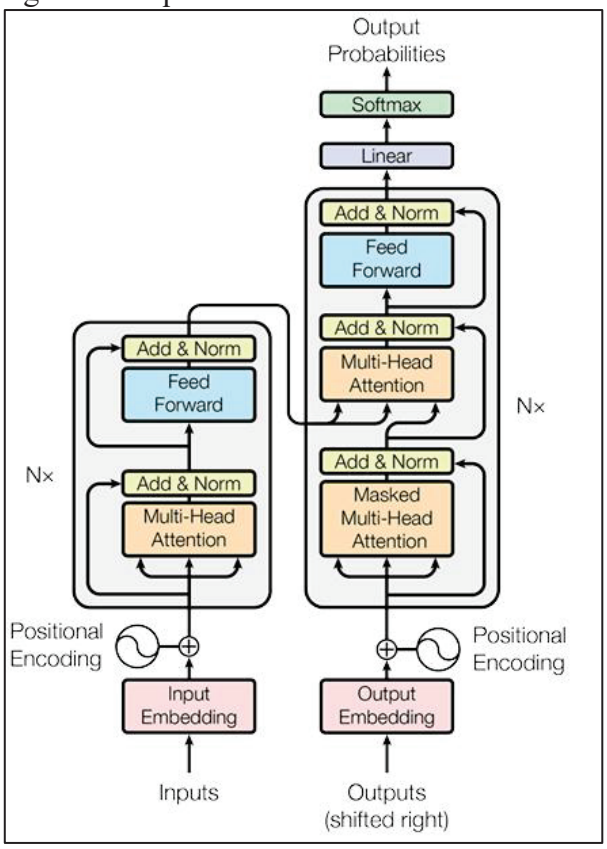
Cada um desses blocos é repetido algumas vezes, formando um conjunto de *encoders* e outro conjunto de *decoders*, que processam as informações uma após a outra. Essa repetição dos mecanismos faz com que o resultado seja de maior qualidade, pois, a cada repetição, os resultados ficam mais assertivos. Além desses dois componentes, há também uma camada denominada LMH - *language modeling head*, responsável pelo processamento final do texto (Sanseviero et al, 2024).

Além destes blocos, existem os componentes do *Transformer* (Vaswani et al, 2017):

- ***Embedding***, cuja função é representar palavras como vetores numéricos;
- **Self- Attention**, que mede a importância de cada palavra em relação às outras;
- ***Feedforwards Layers***, que capturam padrões mais profundos com redes mais densas;
- ***Layer normalization***, que estabiliza e acelera o treinamento; e
- ***Positional Encoding***, que adiciona a noção de ordem (posição das palavras na sequência).

A Figura 1 apresenta o fluxo de execução de um *Transformer*.

Figura 1. Arquitetura do Modelo Transformer



Fonte: (Vaswani et al., 2017)

De acordo com Vaswani et al (2017), a função do *encoder* é "entender" o texto de entrada. Para isso, ele possui duas camadas principais:

- **Self Attention**: que é o mecanismo de atenção responsável por identificar quais palavras de uma sentença estão mais relacionadas entre si. Ele permite que o modelo capture as sutilezas de contexto e significado de um texto.

- **Feed forward**: uma rede neural tradicional que processa as informações extraídas da camada *self-attention*, refinando a compreensão do contexto.

Além dessas, ainda existem duas camadas auxiliares:

- **Normalização**: responsável pelo cálculo que mantém as saídas de uma camada em formato adequado para a próxima; e

- **Conexões residuais**: um mecanismo que faz o "resgate" do *input*, garantindo que o contexto inicial não seja perdido ao longo do processamento.

A estrutura do *decoder*, de acordo com Vaswani et al (2017), se assemelha com a do *encoder*:

- **Self-Attention**: o mecanismo de atenção visto anteriormente. Porém, esse tem uma atenção parcial.
- **Encoder-Decoder Attention**: outro mecanismo de atenção.
- **Feedforward**: é a mesma estrutura presente no *encoder*. Um mecanismo simples de rede neural que tem a função de refinar as saídas dos mecanismos de atenção.
- **Normalização e conexão residual**: camadas intermediárias que funcionam como no *encoder*, mantendo o padrão e o contexto inicial.

A LMH (*Language modeling head*) é a última camada de processamento do Transformer e é composta por duas partes principais:

- **Linear**: conecta a saída dos *decoders* ao vocabulário do modelo.
- **Softmax**: calcula a probabilidade de cada palavra ser a próxima na sequência.

A camada linear recebe as representações matemáticas geradas pelos *decoders*, chamadas de vetores. Esses vetores são versões mais complexas e refinadas dos *embeddings* criados pelos *encoders*.

Nessa etapa, cada um dos vetores é mapeado pela camada linear, que tem acesso ao vocabulário completo do modelo. Para cada palavra desse vocabulário, o vetor recebe uma pontuação que reflete a probabilidade de essa palavra ser a próxima na sequência gerada.

Em seguida, a camada softmax processa as pontuações atribuídas pela camada linear e as converte em probabilidades. A palavra com probabilidade mais alta será escolhida como a próxima palavra gerada pelo modelo.

Os LLMs são Transformers treinados em larga escala, com bilhões ou trilhões de parâmetros, e grandes corpus textuais, para aprender estatísticas da linguagem e realizar geração de texto coerente

(Sanseviero et al, 2024). O funcionamento dos LLMs pode ser compreendido como a combinação de estrutura matemática (baseada nos Transformers) com grande volume de dados e técnicas de treinamento estatístico.

Sanseviero et al (2014) apresentam as seguintes etapas de funcionamento para os modelos LLMs

Pré-processamento

- Todo o texto é convertido em tokens (palavras, pedaços de palavras ou caracteres).
- Exemplo: "Cibersegurança é útil" → ["Cyber", "Segurança", "é", "útil"]

Embedding

- Cada token vira um vetor numérico.
- Esses vetores contêm informações semânticas aprendidas durante o treinamento.

Cálculo de Atenção

- O modelo analisa como cada palavra da frase se relaciona com as outras.
- Exemplo: Em "o usuário que estava logado invadiu o sistema", o modelo percebe que "invadiu" se refere a "usuário", mesmo que estejam distantes.

Camadas do Transformer

- Os vetores passam por múltiplas camadas neurais, refinando a compreensão contextual.
- Cada camada “enriquece” o significado do texto com base no contexto.

Geração de Saída

- A partir de um prompt de entrada (ex: "Complete: Compliance é ..."), o modelo prevê token por token qual será a próxima palavra mais provável.
- A saída é gerada autoregressivamente, ou seja, um token por vez, usando o anterior como base.

2.2.2 APLICAÇÕES DA IA GENERATIVA EM SEGURANÇA E PROTEÇÃO DE DADOS

A IAG tem sido aplicada em diversas áreas da cibersegurança, oferecendo novas ferramentas para a proteção de dados e a detecção de ameaças:

- Detecção de Intrusões: A IAG pode ser usada para analisar o tráfego de rede e identificar padrões ou comportamentos incomuns que possam indicar violações de segurança. Modelos generativos podem ser treinados para reconhecer atividades maliciosas com base em dados históricos de ataques (Chandola, Banerjee e Kumar, 2009).

- Simulação de Cenários de Ataque: A IAG permite a criação de cenários de ataque realistas para treinamento de equipes de segurança. Esses cenários podem incluir ataques de dia zero e ameaças persistentes avançadas, ajudando as organizações a se prepararem para possíveis ameaças (Sommer e Paxson, 2010).

- Análise de Comportamento: A IAG pode ser usada para monitorar atividades e identificar comportamentos anômalos que possam representar ameaças internas ou externas. Modelos generativos podem detectar desvios de padrões normais de comportamento, alertando para possíveis riscos (Garcia-Teodoro et al., 2009).

Essas aplicações destacam o potencial da IAG em fortalecer as defesas cibernéticas, embora também apresentem desafios relacionados à privacidade e à ética. A capacidade de gerar dados realistas pode ser usada tanto para proteger quanto para enganar, levantando questões sobre o uso responsável da tecnologia (Calo, 2017).

As características dos modelos de IA Generativa apresentados nesta seção fundamentam o uso de LLMs na automação de relatórios de compliance explorada neste trabalho.

2.3 AUTOMAÇÃO E INTELIGÊNCIA ARTIFICIAL NO COMPLIANCE

A crescente complexidade do cenário regulatório global tem impulsionado a busca por soluções inovadoras que permitam às organizações manterem a conformidade de forma eficiente e

eficaz. (Jain et al., 2024) Nesse contexto, a automação e a inteligência artificial emergem como ferramentas poderosas para otimizar os processos de compliance, mitigar riscos e reduzir custos (Garcia-Segura, 2024).

A automação, em sinergia com a inteligência artificial, redefine a abordagem das empresas em relação ao compliance, capacitando-as a analisar extensos conjuntos de dados, identificar padrões complexos e detectar anomalias sutis que escapariam à análise humana (Al-Shabandar et al., 2019). A capacidade da IA de processar grandes volumes de dados de forma rápida e precisa proporciona que as organizações monitorem continuamente suas operações, identifiquem potenciais violações de compliance e tomem medidas corretivas em tempo real (Ikudabo; Kumar, 2024).

2.3.1 AUTOMAÇÃO NO COMPLIANCE

A automação no compliance transcende a simples execução de tarefas repetitivas, abrangendo a implementação de sistemas inteligentes capazes de aprender, adaptar-se e tomar decisões autônomas (Hashmi; Governatori; Wynn, 2015). Essa transformação impulsionada pela IA permite que as empresas otimizem seus processos de compliance, direcionando recursos para áreas de maior risco e agregando valor estratégico aos negócios.

A automação de tarefas rotineiras, como a coleta e análise de dados, a geração de relatórios e o monitoramento de transações, libera os profissionais de compliance para se concentrarem em atividades mais complexas e estratégicas, como a análise de riscos, a investigação de fraudes e a implementação de programas de treinamento. A automação, impulsionada pela IA, não apenas agiliza os processos de compliance, mas também aumenta a precisão e a confiabilidade dos resultados, minimizando o risco de erros humanos e inconsistências.

Ao integrar sistemas de IA com *frameworks* de compliance, as organizações conseguem automatizar o rastreamento e o reporte de atividades relacionadas à conformidade, assegurando o cumprimento de normas e regulamentações (Adelakun, 2022).

2.3.2 INTELIGÊNCIA ARTIFICIAL NO COMPLIANCE

A IA pode ser integrada aos frameworks de compliance para automatizar o rastreamento e o reporte de atividades relacionadas à conformidade, garantindo o cumprimento de normas e regulamentações (Kingston, 2017b).

Ao analisar dados de diversas fontes, a IA ajuda a identificar áreas de risco e a implementar medidas preventivas para mitigar potenciais problemas de compliance. Além disso, a IA pode auxiliar na detecção de fraudes e crimes corporativos, aumentando a eficiência e a precisão dos processos de investigação. A capacidade da IA de analisar grandes volumes de dados e identificar padrões complexos permite que as organizações monitorem continuamente suas operações, detectem potenciais violações de compliance e tomem medidas corretivas em tempo real (Garcia-Segura, 2024).

A aplicação de técnicas de IA generativa na elaboração de relatórios de compliance pode promover a criação de documentos informativos fidedignos e precisos, fundamentados em dados e padrões de conformidade predefinidos, o que, por sua vez, facilita a tomada de decisões estratégicas e o cumprimento de normas regulatórias.

A aplicação da inteligência artificial no compliance se manifesta em diversas frentes, desde a interpretação automatizada de checklists de conformidade até a análise preditiva de riscos e a detecção de anomalias em transações financeiras (Kingston, 2017b) . Ou como se propõe neste trabalho, como ferramenta para automatizar o processo de geração dos relatórios de compliance com base em alertas de detecção de vulnerabilidades e fontes de logs de eventos de segurança.

A capacidade da IA de aprender de forma iterativa com os dados e adaptar-se dinamicamente a cenários em constante evolução capacita as organizações a otimizarem continuamente seus programas de compliance, conferindo-lhes maior eficácia e eficiência. Além disso, a IA pode ser utilizada para personalizar programas de treinamento em compliance, adaptando o conteúdo e a metodologia às necessidades específicas de cada indivíduo, aumentando o engajamento e a retenção do conhecimento. Essa otimização se traduz em redução de custos operacionais, mitigação de riscos e melhoria da reputação corporativa.

Nesse cenário, a implementação bem-sucedida da automação e da IA no compliance requer uma abordagem estratégica que considere as particularidades de cada organização, seus objetivos de negócios e seu apetite por risco. É fundamental que as empresas invistam em infraestrutura tecnológica adequada, capacitem seus profissionais e estabeleçam políticas claras de governança de dados para garantir a utilização ética e responsável da IA (Oliveira; Melo, 2023; Prasad; Kalavakolanu, 2023).

3 TRABALHOS CORRELATOS

Este capítulo tem o objetivo de demonstrar o estudo sobre trabalhos correlatos que possuem um grau de similaridade com o tema da dissertação. O capítulo identifica as contribuições de cada trabalho e as compara com a solução proposta, demonstrando o diferencial da dissertação. Para tal, foi realizada uma revisão sistemática da literatura sobre o uso de Inteligência Artificial Generativa na conformidade de cibersegurança, com foco em empresas de médio porte.

3.1 METODOLOGIA

Com intuito de atingir os objetivos da pesquisa e responder às questões formuladas, adotou-se uma abordagem metodológica baseada em revisão sistemática da literatura (RSL). Esta revisão tem como objetivo identificar como a IA Generativa e outras técnicas de automação têm sido aplicadas em cibersegurança e compliance, bem como mapear lacunas que justificam a solução proposta nesta dissertação. A análise de conteúdo e síntese dos resultados foi conduzida conforme descrito a seguir.

A metodologia adotada seguiu a estrutura estabelecida pelo PRISMA (*Preferred Reporting Items for Systematic Reviews and Meta-Analyses*), garantindo transparência, replicabilidade e rigor na seleção e análise dos estudos para tal, foram realizadas as seguintes etapas:

- Pesquisa Bibliográfica e Revisão Sistemática da Literatura
- Análise de Conteúdo e Classificação Temática
- Síntese e Discussão dos Resultados

Cada uma dessas etapas foi estruturada para permitir a identificação de padrões e tendências no uso da Inteligência Artificial Generativa (IA Generativa) na conformidade em segurança da informação, alinhando-se aos objetivos do estudo. A revisão sistemática da literatura conduzida neste capítulo buscou responder à seguinte pergunta de pesquisa:

“Como técnicas de IA Generativa e outras abordagens de automação têm sido utilizadas em cibersegurança e compliance, em especial na geração e apoio à elaboração de relatórios, e quais

lacunas permanecem quanto à geração automatizada de relatórios de compliance orientados a diferentes personas organizacionais?”

A partir dessa pergunta central, a RSL teve como objetivos específicos: (i) identificar aplicações existentes de IA Generativa em cibersegurança e conformidade regulatória, (ii) mapear como essas soluções tratam a geração e comunicação de resultados de segurança para diferentes públicos e (iii) destacar lacunas e oportunidades que fundamentam a proposta desta dissertação.

3.1.1 PESQUISA BIBLIOGRÁFICA E REVISÃO SISTEMÁTICA DA LITERATURA

Esta etapa consistiu na RSL, e buscou mapear o estado da arte sobre o uso da IA Generativa na automação de compliance em cibersegurança.

Foram consultadas quatro bases de dados científicas amplamente reconhecidas na área:

- Google Scholar (https://scholar.google.com/)
- IEEE Xplore (https://ieeexplore.ieee.org/)
- CAPES Periódicos (https://periodicos.capes.gov.br/)
- Scopus (https://www.scopus.com/search)

Nas bases escolhidas, os termos utilizados para a busca foram definidos para abranger os principais conceitos relacionados à pesquisa, incluindo:

- "*Generative AI*"
- "*Compliance*"
- "*Cyber Security*"
- "*Automatic Report Generation*"

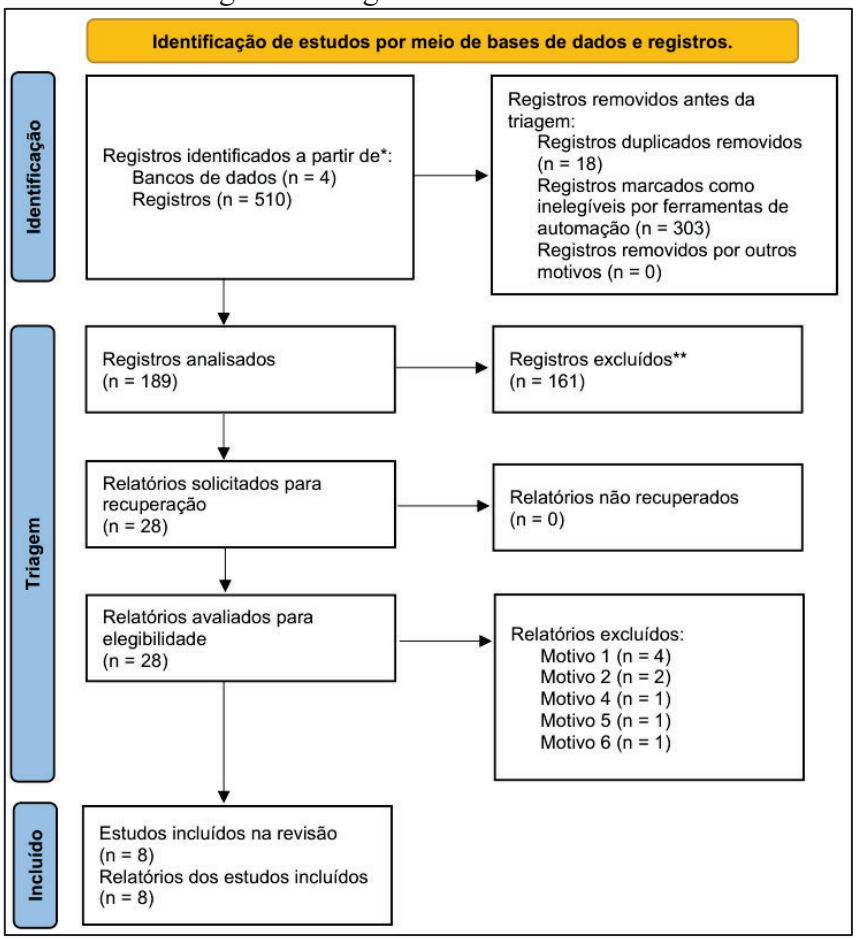
Os critérios para inclusão e exclusão de estudos foram estabelecidos conforme o método PRISMA.

Como o objetivo é analisar o papel da Inteligência Artificial Generativa (IA Generativa) na gestão da conformidade de segurança em organizações, iniciou-se pela fase de identificação. Nesta etapa, foi realizada uma busca abrangente nas bases de dados acadêmicas mais relevantes, incluindo Google Scholar, IEEE Xplore, CAPES e Scopus, usando uma string de busca booleana estruturada que combinava termos relacionados a tecnologias de IA generativa, automação de conformidade e plataformas de cibersegurança. A *string* de busca utilizada foi: *("generative AI" OR "generative artificial intelligence" OR "LLM" OR "large language model" OR "GPT" OR "ChatGPT" OR "AI generation") AND ("compliance report" OR "automated reporting" OR "report generation" OR "compliance automation" OR "automated compliance") AND ("XDR" OR "extended detection and response" OR "SIEM" OR "security platform" OR "threat detection" OR "security analytics" OR "SOC" OR "security operations")*. As buscas foram realizadas entre setembro e outubro de 2025, limitadas a trabalhos publicados entre 2022 e 2025, a fim de capturar as contribuições mais recentes para este campo emergente.

Após a identificação dos artigos relevantes, foi conduzida a fase de triagem. Durante esta fase, os títulos e resumos dos artigos identificados foram sistematicamente examinados para excluir aqueles que não estavam alinhados com os objetivos desta revisão. O processo de seleção focou especificamente em estudos empíricos e revisões sistemáticas que abordavam aplicações práticas da IA Generativa em contextos de cibersegurança. Artigos focados exclusivamente em regulação de IA, pré-impressões sem revisão por pares e estudos puramente teóricos sem aplicação prática foram sistematicamente excluídos da análise.

Considerando as etapas da metodologia PRISMA, as três fases principais da RSL foram: identificação, triagem e inclusão de estudos relevantes, as quais são ilustradas na Figura 2.

Figura 2. Diagrama do Fluxo Prisma



Fonte: Adaptado de (PRISMA, 2020)

Após a identificação dos artigos relevantes, foi conduzida a fase de triagem. Durante esta fase, os títulos e resumos dos artigos identificados foram sistematicamente examinados para excluir aqueles que não estavam alinhados com os objetivos desta revisão. O processo de seleção focou especificamente em estudos empíricos e revisões sistemáticas que abordavam aplicações práticas da IA Generativa em contextos de cibersegurança. Artigos focados exclusivamente em regulação de IA, pré-impressões sem revisão por pares e estudos puramente teóricos sem aplicação prática foram sistematicamente excluídas da análise. A Tabela 1 ilustra os critérios de seleção adotados na pesquisa.

Tabela 1. Critérios de seleção dos artigos

Tipo de dado	Critérios
Inclusão	Artigos publicados entre 2022 e 2025 Estudos empíricos ou revisões sistemáticas Foco em aplicações práticas de IA Generativa IA Generativa em Cibersegurança
Exclusão	Artigos focados apenas em regulação de IA Artigos publicados antes de 2022 Pré-impressões sem revisão por pares Estudos puramente teóricos sem aplicação prática

Fonte: Próprio Autor

Na sequência, foi realizada uma análise de texto dos estudos pré-selecionados para avaliar sua relevância em relação à questão de pesquisa deste artigo. Critérios como a profundidade com que os estudos abordavam a integração da IA Generativa com plataformas XDR e SIEM, a implementação prática de relatórios de conformidade automatizados, a metodologia utilizada e os resultados empíricos relatados foram considerados. Esta etapa culminou na seleção final de estudos a serem incluídos na revisão, garantindo o foco em aplicações práticas em vez de estruturas teóricas. A Tabela 2 ilustra, em números, o processo de seleção dos artigos em suas etapas.

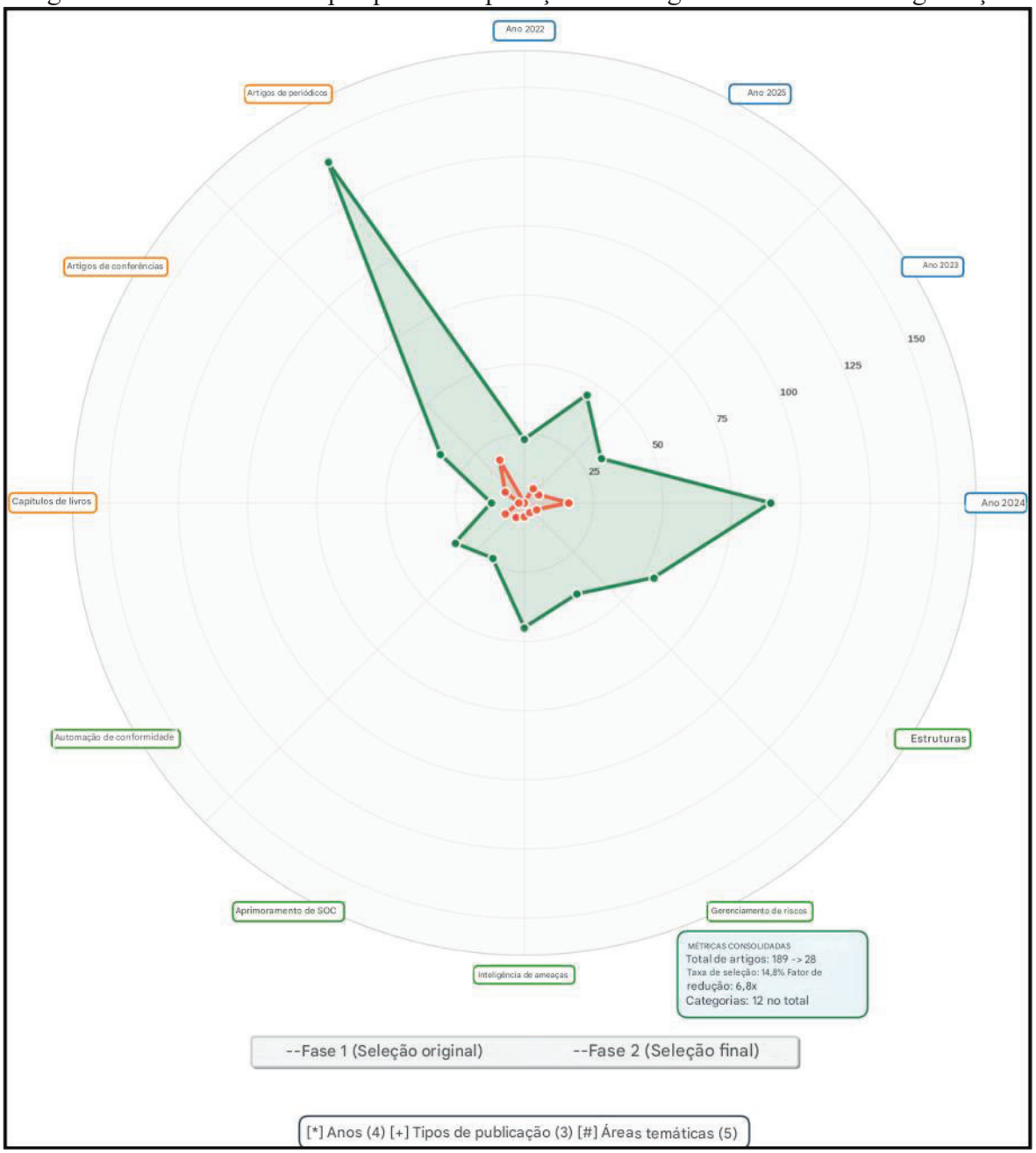
Tabela 2. Processo de seleção dos artigos

Base de dados	1ª seleção	2ª seleção	Seleção final
Google Scholar	89	4	2
IEEE Xplore	131	8	3
CAPES	221	12	2
Scopus	69	4	1
Duplicatas removidas	18	0	0
Registros inelegíveis	303	161	20
Total	189	28	8

Fonte: Próprio Autor

A seguir, a Figura 3 mostra que a distribuição da literatura selecionada revela padrões distintos nos domínios de aplicação e na distribuição temporal da pesquisa neste campo. O gráfico a seguir ilustra a distribuição das áreas de foco da pesquisa em diferentes aplicações de IA generativa na conformidade de cibersegurança e as tendências de publicação ao longo do período analisado, fornecendo insights sobre o estado atual e as direções emergentes desta área de pesquisa.

Figura 3. Áreas de foco da pesquisa em aplicações de IA generativa em cibersegurança

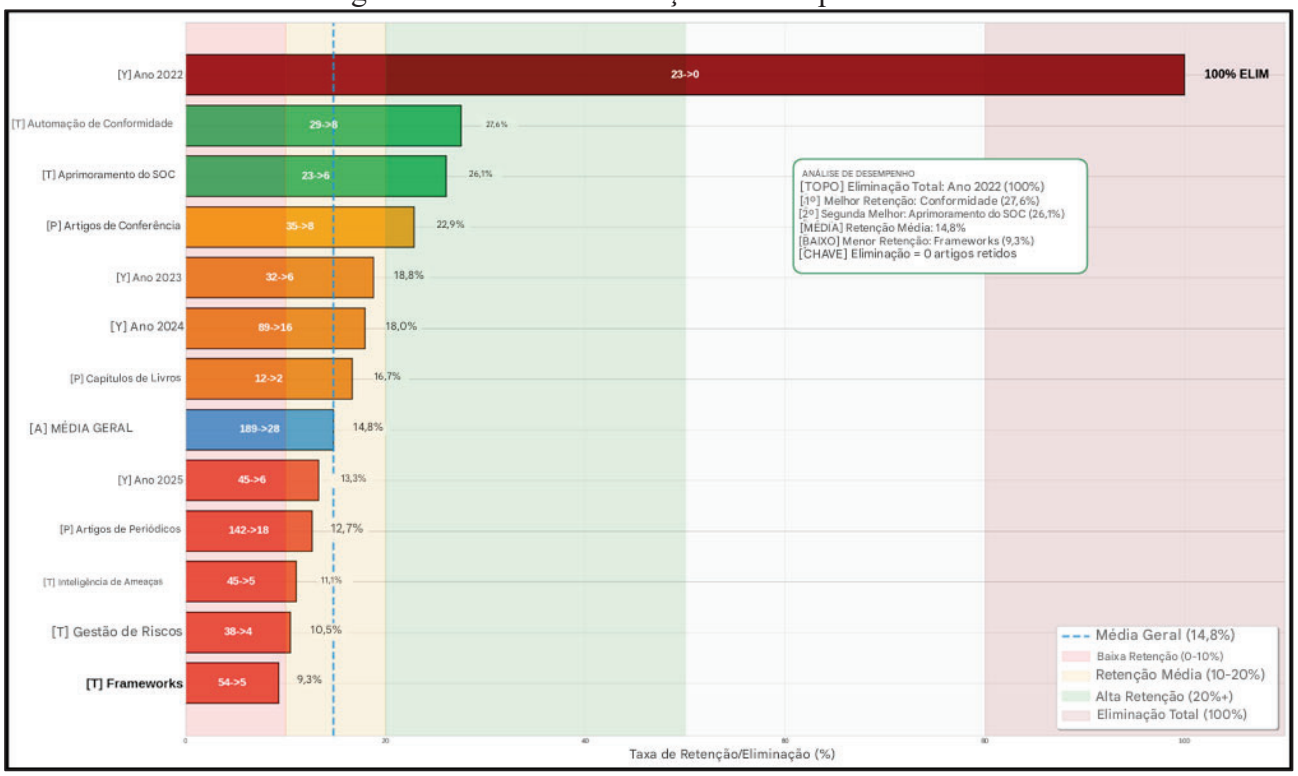


Fonte: Próprio Autor

Finalmente, na fase de inclusão, os estudos selecionados passaram por uma análise abrangente para extrair e sintetizar sistematicamente descobertas-chave relevantes para nossos objetivos de pesquisa. A Figura 4 ilustra o processo completo de seleção de artigos, demonstrando a Fase 2 com seu índice de retenção associado que reflete os rigorosos critérios de filtragem aplicados para garantir a relevância e qualidade da literatura. Esta síntese abrangente nos permitiu abordar sistematicamente nossas questões de pesquisa, fornecendo insights claros sobre o estado atual da aplicabilidade, benefícios demonstrados e limitações identificadas das tecnologias de IA generativa

especificamente em contextos de automação de conformidade em cibersegurança, com ênfase particular na evidência de aplicabilidade prática em ambientes empresariais e organizações de médio porte.

Figura 4. Análise de Retenção: Fase 1 para Fase 2

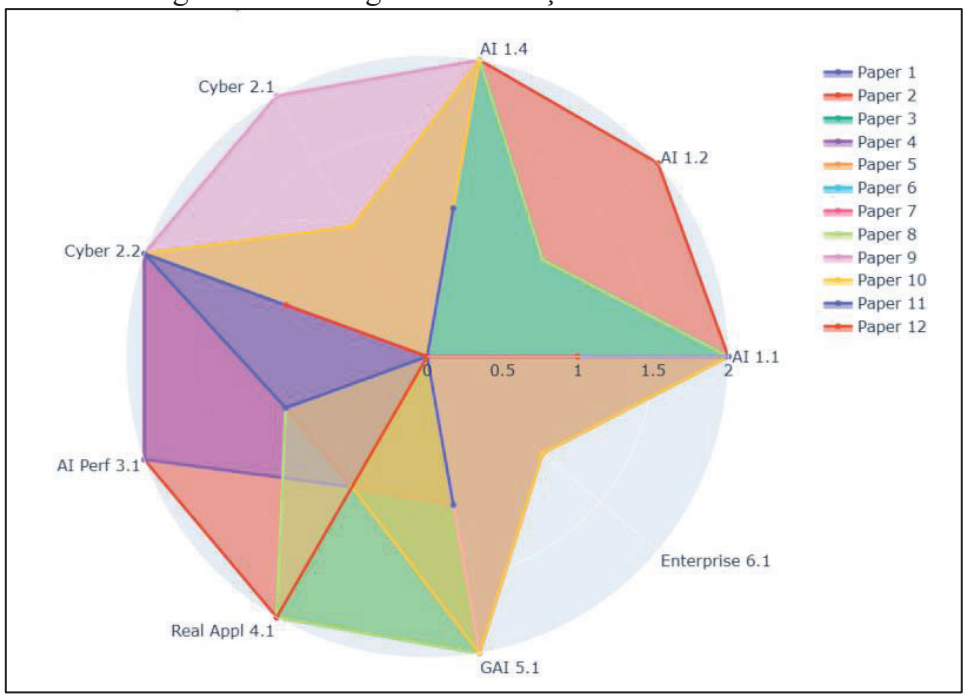


Fonte: Próprio Autor

Após o processo de seleção sistemática descrito na Fase 2, a Fase 3 envolveu uma avaliação de qualidade abrangente e uma análise de conteúdo detalhada dos estudos selecionados usando uma estrutura de avaliação estruturada. Esta fase empregou nove questões de pesquisa específicas organizadas em seis categorias temáticas para garantir uma avaliação rigorosa da relevância e qualidade de cada estudo. Os critérios de avaliação englobaram questões específicas de IA (Questões AI 1.1, 1.2 e 1.4) focando em recenticidade, métricas de avaliação e especificidade para aplicações de IA Generativa; questões focadas em cibersegurança (Questões Cybersecurity 2.1 e 2.2) examinando o foco em conformidade e a relevância para o estado da arte atual; considerações de desempenho de IA (*Questão AI Perform 3.1*) abordando questões de viés e justiça; avaliação de aplicabilidade no mundo real (Questão Real *Application Perform 4.1*) avaliando a implementação prática em empresas; verificação de especificidade da GAI (Questão GAI 5.1) garantindo o foco em IA generativa em vez de IA/ML geral; e relevância para o porte da empresa (Questão Enterprise Size 6.1) determinando a aplicabilidade a organizações de médio porte. Cada estudo selecionado foi

sistematicamente avaliado contra esses critérios usando uma estrutura de pontuação padronizada, com resultados visualizados por meio de uma análise abrangente de gráfico de radar, conforme ilustrado na Figura 5. Esta abordagem de avaliação multidimensional permitiu a comparação objetiva de estudos em diferentes dimensões de qualidade e relevância, garantindo que apenas pesquisas de alta qualidade e relevantes contribuíssem para nossa síntese. A visualização do gráfico de radar facilitou a identificação de estudos com forte desempenho em múltiplos critérios de avaliação, destacando ao mesmo tempo lacunas ou limitações potenciais em áreas específicas, apoiando, em última análise, decisões de seleção baseadas em evidências para a fase de análise final.

Figura 5. Abordagem de avaliação multidimensional



Fonte: Próprio Autor

A seleção final de 8 estudos abrange de 2023 a 2025, garantindo a cobertura dos desenvolvimentos mais recentes em aplicações de IA Generativa para conformidade em cibersegurança. A Figura 6 fornece uma lista dos estudos selecionados (em verde).

Figura 6. Publicações selecionadas (2022-2025)

ID	Article Title	Year of Publication
Paper 1	A Survey on Threat Hunting in Enterprise Networks	2023
Paper 2	AGIR: Automating Cyber Threat Intelligence Reporting with Natural Language Generation	2023
Paper 3	Applying Large Language Model Analysis and Backend Web Services in Regulatory Technologies for Continuous Compliance Checks	2025
Paper 4	CyKG-RAG: Towards knowledge-graph enhanced retrieval augmented generation for cybersecurity	2024
Paper 5	Empowering Security Operation Center With Artificial Intelligence and Machine Learning—A Systematic Literature Review	2025
Paper 6	Enhancing Security Operations Center: Wazuh Security Event Response with Retrieval-Augmented-Generation-Driven Copilot	2025
Paper 7	Human-in-the-Loop Intelligence: Advancing {AI}-Centric Cybersecurity for the Future	2023
Paper 8	Organizational Adaptation to Generative AI in Cybersecurity: A Systematic Review	2025
Paper 9	Position Paper: Leveraging Large Language Models for Cybersecurity Compliance	2024
Paper 10	Revolutionizing Third Party Risk Management Using Generative AI and RAG	2024
Paper 11	Empowering Security Operation Center With Artificial Intelligence and Machine Learning—A Systematic Literature Review	2025
Paper 12	ThreatInsight: Innovating Early Threat Detection Through Threat-Intelligence-Driven Analysis and Attribution	2024

Fonte: Próprio Autor

3.1.2 PESQUISA BIBLIOGRÁFICA E REVISÃO SISTEMÁTICA DA LITERATURA

A busca sistemática da literatura foi fundamental para identificar o estado da arte em relação à aplicação da IA Generativa como ferramenta que possa contribuir positivamente na gestão da conformidade de segurança e proteção de dados pessoais. Os estudos selecionados fornecerão uma base sólida para o desenvolvimento da pesquisa e para a construção de um conhecimento aprofundado sobre essa temática demonstrando que não existem trabalhos no contexto específico de aplicação proposta neste trabalho, porém que os benefícios esperados já têm sido encontrados em outros contextos.

Após os critérios de seleção atualizados e o refinamento no foco de pesquisa, oito estudos recentes foram identificados e analisados, representando o estado da arte atual neste campo em rápida evolução. A análise revela avanços significativos em tecnologias GAI aplicadas a operações de cibersegurança, automação de conformidade e inteligência de ameaças.

3.1.3 Aplicações de IA Generativa em Operações de Cibersegurança

Inteligência de Ameaças e Relatórios Automatizados

A aplicação da IA Generativa na geração de inteligência de ameaças representa uma das áreas mais maduras identificadas em nossa análise.(Perrina et al., 2023) apresentam o AGIR, um sistema híbrido sofisticado de Geração de Linguagem Natural que combina módulos determinísticos baseados em modelo de linguagem grandes tipo *transformers*, como o ChatGPT. O sistema demonstra métricas de desempenho notáveis, alcançando "um alto valor de recall (0,99) sem introduzir alucinação e reduzindo o tempo de escrita de relatórios em mais de 40%.

O sistema AGIR aborda um desafio operacional crítico nos Centros de Operações de Segurança (SOCs) automatizando a tarefa intensiva em trabalho de gerar relatórios de inteligência abrangentes a partir de representações formais de grafos de entidades. Esta abordagem representa um avanço significativo em relação aos sistemas de relatórios automatizados anteriores.

Construindo sobre esta base, (Ismail et al., 2025) estendem as aplicações GAI para operações de segurança em tempo real por meio de seu *Security Event Response Copilot* (SERC). O sistema "integra dois componentes principais: (1) extração de dados de eventos de segurança usando métodos de Geração Aumentada por Recuperação (RAG), e (2) orientação de resposta a incidentes baseada em LLM". Isto representa uma evolução significativa da geração de relatórios estáticos para a automação de resposta a incidentes dinâmica e ciente do contexto. A implementação do SERC demonstra aplicabilidade empresarial prática ao aproveitar plataformas de código aberto, pois os autores observam que "muitas empresas estão adotando cada vez mais plataformas de Centro de Operações de Segurança (SOC) de código aberto como alternativas econômicas às soluções proprietárias".

Sistemas de Recuperação Aprimorados por Conhecimento

A integração de conhecimento estruturado de cibersegurança com IA Generativa representa outro avanço significativo identificado em nossa análise. (Kurniawan, 2024) abordam limitações fundamentais dos Modelos de Linguagem Grandes em contextos de cibersegurança, observando que "esses modelos enfrentam sérias limitações. Eles frequentemente lutam com tarefas de raciocínio complexo e em manter a consistência em múltiplas consultas".

Seu framework CyKG-RAG demonstra como os grafos de conhecimento podem melhorar a confiabilidade dos LLMs, fornecendo relações semânticas estruturadas. Esta abordagem aborda diretamente o problema de alucinação que tem dificultado a adoção da GAI em aplicações críticas de cibersegurança. A inovação técnica reside em combinar a representação simbólica do conhecimento de bancos de dados estabelecidos de cibersegurança (CVE, CAPEC, MITRE ATT&CK). Esta abordagem híbrida permite o processamento de inteligência de ameaças estruturada e dados de log não estruturados por meio de representações de grafos unificadas.

3.1.4 Automação de Compliance e Gestão de Riscos

Avaliação de Conformidade Baseada em Evidências

Abordagens tradicionais de conformidade em cibersegurança confiaram fortemente em auditorias baseadas em documentação, que (Salman; Creese; Goldsmith, 2024) identificam como fundamentalmente falhas. Eles argumentam que "as técnicas atuais de conformidade e auditoria de cibersegurança dependem principalmente da disponibilidade de documentos de política e documentação procedural, em vez de evidências de implementação real". Esta observação destaca uma lacuna crítica entre as políticas documentadas e as posturas de segurança reais.

Os autores propõem aproveitar os Modelos de Linguagem Grandes para transformar a avaliação de conformidade de metodologias baseadas em documento para metodologias baseadas em evidências. Eles postulam que "os LLMs podem abordar os desafios de conformidade devido à sua notável capacidade de analisar dados não estruturados". O framework proposto aborda uma lacuna de mercado significativa, pois os autores observam que "a indústria atualmente carece de soluções que possam fornecer avaliação de conformidade contínua, cobrindo também um amplo espectro de aspectos organizacionais". O monitoramento contínuo de conformidade representa uma melhoria substancial em relação às abordagens baseadas em auditoria periódica, potencialmente permitindo a avaliação da postura de conformidade em tempo real para empresas de médio porte.

Automação do Gerenciamento de Riscos de Terceiros

Estendendo as aplicações GAI para o gerenciamento de riscos de terceiros, (Arora, 2024) apresenta um framework abrangente integrando IA Generativa e Geração Aumentada por Recuperação para automação de TPRM. Eles argumentam que "essa integração de IA Generativa e Geração Aumentada por Recuperação (RAG) abrirá novas fronteiras no Gerenciamento de Riscos de Terceiros (TPRM), tornando a criação de avaliações de risco e verificações de conformidade mais precisa, eficiente e escalável".

O sistema proposto demonstra capacidades sofisticadas de processamento de documentos, permitindo que "terceiros façam upload de suas políticas em uma pasta zip que contém documentos em diferentes formatos, como PDFs". Esta capacidade garante que as avaliações de conformidade permaneçam atualizadas com os requisitos regulatórios em evolução, um desafio crítico para os frameworks de conformidade estáticos tradicionais. O sistema incorpora mecanismos de supervisão

humana, observando que este processo automatizado introduz ainda a capacidade para os auditores conduzirem um processo de verificação manual no qual podem verificar, por meio de verificação cruzada, as descobertas geradas por IA usando hiperlinks fornecidos". Esta abordagem equilibra a eficiência da automação com a supervisão necessária para decisões críticas de conformidade.

3.1.5 Desafios e Implicações

Apesar dos avanços significativos, vários desafios técnicos persistem entre os estudos analisados. O problema da alucinação permanece uma preocupação crítica, com (Kurniawan, 2024) reconhecendo que os LLMs "muitas vezes falham ao lidar com conjuntos de dados factuais, às vezes introduzindo informações irrelevantes ou incorretas, um problema comumente referido como alucinação". Embora frameworks como o CyKG-RAG abordem algumas questões de confiabilidade por meio do fundamento no conhecimento, soluções abrangentes para viés, justiça e explicabilidade permanecem subdesenvolvidas, com a maioria dos estudos focando em métricas de desempenho técnico enquanto dão atenção insuficiente a considerações de segurança de IA críticas para a implantação empresarial.

A análise revela lacunas significativas na orientação de implementação específica para empresas, particularmente para organizações de médio porte, onde a consideração detalhada das restrições das PMEs, como recursos de TI limitados, considerações orçamentárias e requisitos de expertise técnica, permanece amplamente ausente. Além disso, uma limitação notável em vários estudos é a falta de validação empírica abrangente, com apenas (Perrina et al., 2023) fornecendo métricas de desempenho quantitativas e (Ismail et al., 2025) demonstrando integração prática de sistema, enquanto outros como (Salman; Creese; Goldsmith, 2024) e (Arora, 2024) permanecem mais conceituais sem validação de desempenho empírico.

No entanto, os estudos demonstram potencial significativo para melhorias de eficiência operacional em empresas de médio porte, com a redução de 40% no tempo de escrita de relatórios alcançada pelo AGIR e a abordagem de código aberto econômica do SERC sugerindo que as tecnologias GAI podem proporcionar ganhos substanciais de produtividade sem aumentos proporcionais nos custos de licenciamento. Como observam (Ismail et al., 2025), as soluções GAI podem "auxiliar os analistas a responder e mitigar violações de segurança de forma mais eficaz" enquanto "permitem que as organizações estabeleçam posturas de segurança eficazes sem incorrer

em custos substanciais", o que poderia beneficiar particularmente as empresas de médio porte que normalmente operam com pessoal limitado, em especial, na área de cibersegurança.

Não obstante, os desafios de implementação requerem consideração cuidadosa, pois a complexidade técnica de sistemas como o CyKG-RAG, que requer "tanto a manutenção do grafo de conhecimento de cibersegurança quanto a infraestrutura LLM", pode exceder os recursos técnicos disponíveis para muitas PMEs. Já (Karpatou, 2025) fornece um contexto histórico importante, observando que as ameaças de cibersegurança evoluíram para aproveitar "*Phishing* e Engenharia Social alimentadas por IA, Malware Autônomo e Auto Evolutivo, Deepfakes e Fraude de Identidade", ressaltando a urgência para que empresas de médio porte adotem mecanismos de defesa avançados alimentados por GAI, mas os caminhos práticos de implementação permanecem pouco claros, requerendo uma avaliação cuidadosa sobre se a sofisticação técnica justifica o esforço de implementação em comparação com as ferramentas e processos de cibersegurança existentes.

3.1.6 Evolução das Abordagens Anteriores

A geração atual de aplicações GAI representa uma evolução significativa em relação às soluções automatizadas de cibersegurança anteriores. Abordagens anteriores para automação de conformidade normalmente dependiam de sistemas rígidos baseados em regras ou algoritmos simples de correspondência de palavras-chave que lutavam com a interpretação de nuances necessárias para requisitos regulatórios complexos. Os sistemas SIEM tradicionais, embora eficazes para agregação de logs e correlação básica, careciam das capacidades sofisticadas de processamento de linguagem natural demonstradas pelas soluções GAI atuais.

A integração de modelos baseados em *transformers* com grafos de conhecimento específicos do domínio de conhecimento, representa um avanço fundamental na capacidade de processar e analisar dados não estruturados de cibersegurança. Soluções anteriores de gerenciamento de riscos de terceiros sofriam de estruturas de avaliação estáticas que não podiam se adaptar aos requisitos regulatórios em evolução. As abordagens aprimoradas por RAG descritas por (Arora, 2024) demonstram como o fundamento regulatório em tempo real pode abordar essas limitações, embora a validação empírica dessas afirmações permaneça incompleta.

3.1.7 Resumo comparativo das soluções correlatas

Para sintetizar as principais contribuições identificadas na revisão sistemática, a Tabela 3 apresenta um quadro comparativo dos estudos discutidos, destacando seu contexto de aplicação, abordagens tecnológicas e limitações em relação ao escopo desta dissertação.

Esse resumo evidencia como diferentes linhas de pesquisa têm explorado o uso de IA Generativa e técnicas correlatas em operações de cibersegurança, avaliação de conformidade e gestão de riscos de terceiros, ao mesmo tempo em que permite visualizar de forma estruturada as lacunas que motivam a solução proposta neste trabalho.

Tabela 3. Comparativo RSL x Proposta

AUTOR(ES) / ANO	CONTEXTO / DOMÍNIO	TIPO DE ABORDAGEM / TECNOLOGIA	FOCO PRINCIPAL DA SOLUÇÃO	PRINCIPAIS CONTRIBUIÇÕES	LIMITAÇÕES / LACUNAS EM RELAÇÃO A ESTE TRABALHO
PERRINA ET AL., 2023	Centros de Operações de Segurança (SOC) e inteligência de ameaças	AGIR – sistema híbrido de Geração de Linguagem Natural com LLMs tipo transformer (ex.: ChatGPT)	Geração automatizada de relatórios de inteligência de ameaças	Reduz em mais de 40% o tempo de escrita de relatórios, com alto recall (0,99) e baixa alucinação; automatiza relatórios abrangentes em SOCs	Não trata relatórios de compliance orientados a personas; não integra diretamente SIEM com múltiplos perfis (CEO, TIC, Compliance Officer) nem aplica framework de avaliação multidimensional A/B/C
ISMAIL ET AL., 2025	Operações de segurança em tempo real e resposta a incidentes	Security Event Response Copilot (SERC) com LLM + RAG	Automação de resposta a incidentes baseada em IA Generativa	Integra extração de eventos via RAG e orientação de resposta com LLM; demonstra aplicação prática em SOCs de código aberto, reduzindo custo	Foco em resposta a incidentes e não em relatórios de compliance; não gera relatórios textuais persona-fit nem avalia qualidade linguística / alinhamento ao propósito como neste trabalho
KURNIAWAN, 2024	Inteligência de ameaças e confiabilidade de LLMs em cibersegurança	Framework CyKG-RAG (LLM + grafos de conhecimento)	Mitigação de alucinações e aumento de consistência em respostas de LLM	Integra grafos de conhecimento com LLM para reduzir alucinações; mostra como conhecimento estruturado melhora a confiabilidade em tarefas de segurança	Não aborda geração de relatórios de compliance completos; foco em respostas mais confiáveis, sem considerar múltiplas personas nem avaliação empírica de relatórios como no seu estudo
SALMAN; CREESE; GOLDSMITH, 2024	Conformidade em cibersegurança baseada em evidências	Uso de LLMs para avaliação de conformidade baseada em evidências	Transformar auditorias de conformidade de abordagem baseada em documentos para abordagem baseada em evidências	Destacam falhas de auditorias centradas em documentos; propõem LLMs para analisar dados não estruturados e apoiar monitoramento contínuo de conformidade	Abordagem principalmente conceitual, sem validação empírica abrangente; não gera relatórios persona-fit a partir de SIEM nem aplica framework de avaliação como o deste trabalho
ARORA, 2024	Gerenciamento de riscos de terceiros (TPRM)	IA Generativa + RAG para automação de TPRM	Automatização de avaliações de risco e verificações de conformidade de terceiros	Processa documentos diversos de terceiros; mantém avaliações atualizadas com requisitos regulatórios; combina automação com possibilidade de verificação manual via hiperlinks	Foco em TPRM, não em relatórios de compliance internos baseados em SIEM; ainda conceitual, sem validação empírica robusta; não considera recorte específico de PMEs nem relatórios por persona
KARPATOU, 2025	Evolução histórica das ameaças de cibersegurança	Análise conceitual / revisão sobre ameaças avançadas (não é uma solução de IA específica)	Contextualizar ameaças modernas (phishing com IA, malware auto-evolutivo, deepfakes)	Mostra como IA é usada em ataques (engenharia social, malware, fraude de identidade), reforçando a urgência de mecanismos de defesa avançados	Não propõe framework de compliance nem solução técnica de geração de relatórios; serve de contexto, mas não aborda SIEM, relatórios persona-fit ou avaliação empírica de GAI como o seu trabalho
PROPOSTA DA DISSERTAÇÃO	Compliance em cibersegurança em empresas de médio porte	Arquitetura modular com IA Generativa (LLMs) + dados de SIEM	Geração automatizada de relatórios de compliance orientados a personas (CEO, TIC, Compliance Officer), com avaliação multidimensional	Integra dados reais de SIEM com IA Generativa; produz relatórios persona-fit; propõe framework de avaliação com dimensões A/B/C e aplica avaliação empírica com métricas quantitativas e análise qualitativa dos relatórios	Preenche lacuna de validação empírica em contexto de PME, foca em relatórios de compliance (não apenas inteligência de ameaças), e trata explicitamente da adaptação por persona e da avaliação sistemática da qualidade dos relatórios

Fonte: Elaborado pelo Autor

A partir da comparação estruturada, com base no quadro, evidenciou-se a posição deste trabalho em relação as demais abordagens, tornou-se claro o seu diferencial: integrar dados reais de SIEM com

IA Generativa para produzir relatórios de compliance persona-fit e avaliá-los de forma empírica em múltiplas dimensões, preenchendo lacunas apontadas na literatura quanto à validação prática, ao foco em PMEs e à comunicação adaptada para públicos distintos.

3.1.8 Lacunas de Pesquisa e Direções Futuras

A análise revela uma necessidade crítica de validação empírica abrangente dos frameworks GAI propostos. Embora as contribuições teóricas sejam significativas, a implantação prática desses sistemas em ambientes empresariais reais requer testes rigorosos e benchmarking de desempenho em relação às bases estabelecidas. Pesquisas futuras devem priorizar experimentos controlados comparando processos de cibersegurança aprimorados por GAI versus tradicionais, incorporando métricas quantitativas para precisão, eficiência e relação custo-benefício. Estudos complementares examinando o desempenho do sistema em ambientes regulatórios dinâmicos forneceriam insights valiosos sobre adaptabilidade e requisitos de manutenção.

Além disso, a falta de pesquisa específica para PMEs representa uma lacuna significativa na literatura atual. Estudos futuros devem investigar considerações de implementação únicas para empresas de médio porte, incluindo restrições de recursos, requisitos de escalabilidade e desafios de integração com sistemas existentes. Pesquisas examinando a análise de custo-benefício para a adoção de soluções GAI por PMEs forneceriam orientação essencial para a tomada de decisão organizacional. Adicionalmente, a investigação de modelos de implantação simplificados poderia tornar as capacidades avançadas de GAI mais acessíveis a organizações com recursos limitados.

Já a atenção limitada às considerações de segurança de IA, entre os estudos analisados, destaca-se uma necessidade crítica de pesquisa. Trabalhos futuros devem abordar estratégias de detecção e mitigação de viés, requisitos de explicabilidade para aplicações de conformidade e frameworks de governança para implantação de GAI empresarial. A pesquisa sobre a aceitação regulatória de evidências de conformidade geradas por IA representa outra área importante, particularmente à medida que as empresas buscam demonstrar conformidade por meio de sistemas automatizados. Compreender as implicações legais e regulatórias da avaliação de conformidade alimentada por GAI será crucial para a adoção generalizada.

Os achados indicam que a IA Gerativa pode aliviar substancialmente a dificuldade associada à geração manual de relatórios de conformidade. Ao automatizar este processo, as organizações podem liberar recursos humanos valiosos, melhorando assim a eficiência operacional geral.

A capacidade da IA Gerativa de sintetizar informações a partir de alertas de detecção de vulnerabilidades e logs de eventos de segurança permite a geração rápida de relatórios de conformidade abrangentes. Isso não apenas *streamlines* o processo de conformidade, mas também minimiza o potencial para erros humanos, que são frequentemente prevalentes na geração manual de relatórios.

3.1.9 Desafios

Embora a aplicação de IA Gerativa em cibersegurança apresente oportunidades, ela não está isenta de desafios. Questões-chave incluem a necessidade de dados de alta qualidade, pois o treinamento eficaz de modelos de IA Gerativa requer acesso a grandes volumes de dados precisos e relevantes.

Além disso, preocupações sobre privacidade e segurança de dados são fundamentais, particularmente ao lidar com informações sensíveis. Estabelecer confiança na precisão e confiabilidade dos relatórios e simulações gerados por IA é crucial para a implementação bem-sucedida. As organizações devem navegar por esses desafios para aproveitar plenamente o potencial da IA Gerativa no fortalecimento da conformidade em cibersegurança. Em resumo, a automação da geração de relatórios de conformidade usando IA Gerativa pode reduzir significativamente o fardo da geração manual de relatórios, melhorando a eficiência geral do processo.

3.2 CONSIDERAÇÕES

Este estudo apresenta uma revisão abrangente das aplicações emergentes da IA Generativa no aprimoramento da conformidade em cibersegurança, particularmente para organizações de médio porte. Os resultados observados destacam o potencial significativo da IA Generativa em simplificar o processo de conformidade e fortalecer a postura geral de segurança das organizações.

Nesse contexto, o diferencial desta dissertação está em integrar dados reais de um SIEM com uma arquitetura modular baseada em IA Generativa para produzir relatórios de compliance adaptados a três personas específicas (CEO, TIC e Compliance Officer) e em propor um framework de avaliação

multidimensional que contempla qualidade linguística, alinhamento ao propósito e desempenho dos modelos de IA.

Essa abordagem permite não apenas demonstrar a viabilidade técnica da geração automatizada de relatórios de compliance orientados a personas, mas também quantificar, de forma estruturada, ganhos potenciais em termos de clareza, eficiência e confiabilidade, contribuindo para preencher uma lacuna identificada na literatura.

A revisão sistemática da literatura identifica avanços significativos nas aplicações de IA Generativa para conformidade em cibersegurança, com progresso demonstrado na automação de inteligência de ameaças, sistemas de recuperação aprimorados por conhecimento e *frameworks* de avaliação de conformidade. Os estudos revelam potencial substancial para ganhos de eficiência operacional e capacidades de segurança aprimoradas, particularmente relevantes para empresas de médio porte que buscam soluções de cibersegurança econômicas.

No entanto, desafios significativos permanecem, incluindo limitações técnicas relacionadas à segurança e confiabilidade da IA, lacunas de implementação específicas para empresas de médio porte e validação empírica insuficiente dos frameworks propostos. O campo requer foco contínuo de pesquisa em considerações práticas de implantação, validação de desempenho abrangente e orientação de implementação específica para empresas para que seja possível o aproveitamento do pleno potencial das tecnologias GAI em aplicações de conformidade em cibersegurança. A análise comparativa dos trabalhos correlatos fundamenta a contribuição científica deste estudo, que consiste em demonstrar, com base em dados empíricos e em um arcabouço de avaliação formal, como a IA Generativa pode ser utilizada para otimizar processos de compliance em cibersegurança em empresas de médio porte.

4 DESENVOLVIMENTO

Neste capítulo, apresenta-se a implementação da solução para automatizar a geração de relatórios de compliance em segurança da informação e proteção de dados. O trabalho buscou abordar os desafios enfrentados por empresas de médio porte no contexto da escassez de profissionais qualificados e da crescente necessidade de produzir relatórios técnicos adaptados às diferentes áreas corporativas. A solução utiliza Inteligência Artificial Generativa (IA Generativa) como núcleo para transformar dados brutos provenientes de um sistema SIEM (Sistema de Informação e Gestão de Eventos) e informações complementares de frameworks de compliance em relatórios personalizados e compreensíveis.

A seguir, detalha-se a arquitetura da solução, as interações entre os componentes principais e as tecnologias específicas preliminarmente testadas. Além disso, são discutidas as contribuições esperadas e as limitações do projeto.

4.1 ARQUITETURA DA SOLUÇÃO

A solução foi projetada com uma arquitetura modular, permitindo a separação clara dos processos de integração, processamento e geração de relatórios dentro de um fluxo de automação. Essa abordagem facilita a manutenção e a escalabilidade da solução, além de possibilitar a substituição ou atualização de módulos individuais sem comprometer o sistema como um todo.

Os três principais módulos da arquitetura são:

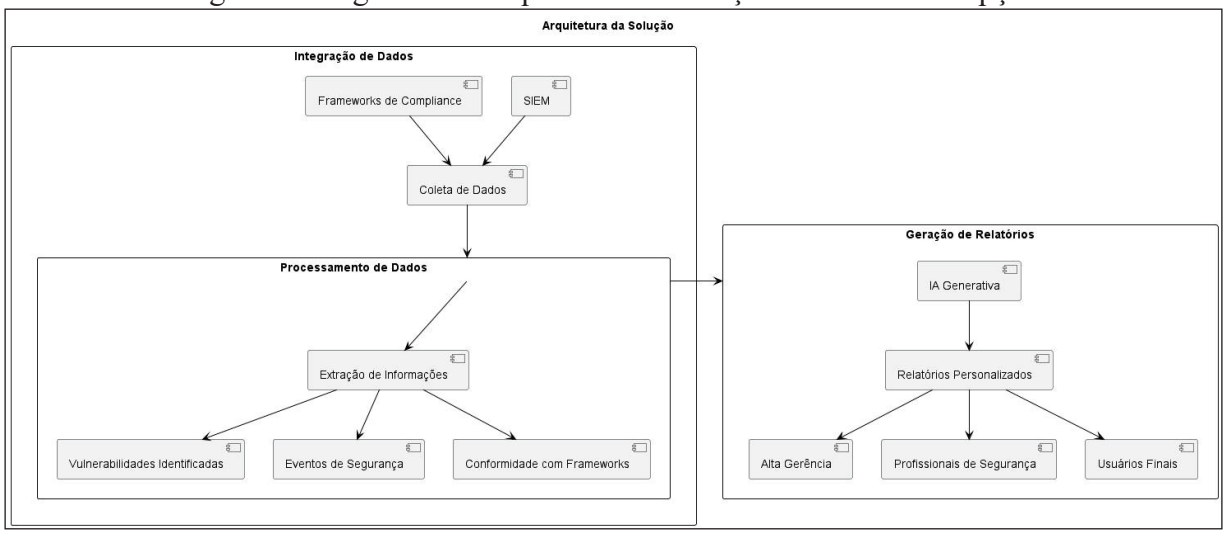
1. Integração de Dados: Coleta e integra dados de Sistemas de Gestão de Vulnerabilidades (SIEM) e frameworks de boas práticas de compliance em segurança, como ISO 27001 e Lei Geral de Proteção de Dados Pessoais (LGPD).

2. Processamento de Dados: Processa os dados coletados, extraindo informações relevantes como vulnerabilidades identificadas, eventos de segurança, status de conformidade com frameworks e regulamentações, entre outros.

3. Geração de Relatórios: Utiliza técnicas de IA Generativa para gerar relatórios personalizados e adaptados à linguagem e necessidades de cada persona, como alta gerência, profissionais de segurança da informação e usuários finais.

Os módulos da solução são ilustrados na Figura 7 a seguir.

Figura 7. Diagrama de Arquitetura da Solução: Fase de Concepção



Fonte: Elaboração Própria

O diagrama da Figura 7 ilustra os seguintes componentes:

- O componente “Integração de Dados” coleta informações de sistemas de gestão de vulnerabilidades (SIEM) e frameworks de compliance.
- O componente “Processamento de Dados” processa essas informações, extraindo detalhes relevantes.
- O componente “Geração de Relatórios” utiliza técnicas de IA generativa para criar relatórios personalizados para diferentes personas.

Além do modelo e fluxo de integração proposto nesta arquitetura, o processamento dos dados vindos dos frameworks e eventos de segurança são fatores importantes para o resultado que se espera, pois, e a partir desta curadoria e estruturação de informações que os relatórios são gerados e a melhoria proposta para o processo pôde ser demonstrada.

Uso da IA Generativa com Adequação de Contexto

Modelos de IA Generativa serão utilizados a partir de contexto fornecido com base no conjunto de dados que inclui:

- *Template* base de relatório de compliance: Para garantir a estrutura base, o modelo de relatório será estático e determinístico para reduzir o risco de variabilidade estrutural.

- Dados de vulnerabilidades e eventos de segurança: Dados processados a partir da consulta das apis do SIEM Wazuh e informações básicas de frameworks de boas práticas de segurança, fornecendo ao modelo o contexto necessário para melhor interpretar e analisar os dados de segurança.

- Linguagem específica de cada persona: Orientação do tom de linguagem e persona de destino para que o modelo gere relatórios adaptados.

A definição do contexto base para a tarefa a ser realizada pelo modelo será realizada através de um processo iterativo que envolve:

- Preparação de dados: Pré-processamento e formatação dos dados para serem utilizados pelo modelo de IA Generativa.

- Instrução do modelo: Fornecimento dos dados por meio de templates dinâmicos de prompt contextualizando a tarefa nos modelos de IA Generativa, ajustando seus parâmetros para gerar relatórios de alta qualidade.

- Avaliação e refinamento: Avaliação do desempenho de múltiplos modelos em um conjunto de dados de validação, observação visual e automação de verificação de termos para avaliação do método e seu refinamento para melhorar sua precisão e fluência.

4.2 PROTÓTIPO DE TESTES

No contexto deste projeto foi implementado um pipeline modular com objetivo de coletar, processar e transformar evidências de conformidade e vulnerabilidades em relatórios e blocos de transição de texto para maior fluência do texto gerado pelos agentes de IA.

Como protótipo de validação arquitetural, e até mesmo validação de possível utilização na solução final, iniciou-se o projeto com a construção do fluxo utilizando a plataforma de automação

N8N, observou-se ampla utilização no mercado e por disponibilizar versão *open-source* para uso local se mostrou uma forma rápida para o início dos testes de coleta e processamento dos dados.

O n8n constitui uma ferramenta web de automação de workflows caracterizada por sua flexibilidade operacional e baixo custo de implementação. A plataforma apresenta capacidade de integração com centenas de sistemas distintos por meio de nós especializados, além de suportar milhares de outros sistemas que utilizam interfaces REST API padronizadas. Entre suas funcionalidades principais, destaca-se a capacidade de manipulação e transformação de dados entre sistemas heterogêneos, eliminando a necessidade de uniformização de formatos durante processos de transferência de informações. Adicionalmente, a ferramenta oferece recursos analíticos que permitem a geração de insights a partir dos dados processados em seus workflows. (McFeetors; Pant, 2022).

O fluxo automatizado desenvolvido na plataforma N8N implementa uma versão inicial da arquitetura do projeto e estabelece um pipeline completo que inicia com a autenticação no sistema Wazuh (remoto) por meio de requisições HTTP com credenciais básicas, seguida por validação condicional do token de acesso. Uma vez autenticado, o fluxo foi implementado em três vertentes paralelas de coleta de dados: extração de vulnerabilidades diretamente do *Elasticsearch* via consultas estruturadas ao índice *wazuh-alerts*, recuperação de agentes associados a grupos específicos, e obtenção da listagem completa de agentes ativos no ambiente monitorado. Esta arquitetura paralela demonstra a capacidade da plataforma em orquestrar múltiplas fontes de dados simultaneamente, otimizando o tempo de execução e consolidando informações heterogêneas em um único fluxo de trabalho.

O processamento dos dados coletados foi implementado a partir de estruturas iterativas em nós de código *JavaScript* personalizados para normalização, agregação e transformação das informações brutas. Para o processamento de vulnerabilidades, o fluxo consolida os alertas por agente e prepara os dados para consumo pelos modelos de linguagem através do nó ConsolidateVulnerabilitiesByAgent. O processamento de *Security Configuration Assessment* (SCA) implementa loops aninhados que iteram sobre cada agente, recuperam suas políticas de conformidade associadas, executam consultas individuais para obter os resultados de cada verificação, e agregam esses dados com identificadores de agente e política antes de consolidá-los estatisticamente.

Simultaneamente, o fluxo implementa a coleta de informações de inventário de hardware e software através de múltiplos *endpoints* de API, aplicando loops de processamento que achatam

estruturas hierárquicas e mesclam informações de diferentes fontes em um conjunto de dados unificado através de um nó Merge, preparando assim uma base consolidada para análise.

A integração de Inteligência Artificial Generativa também foi testada a partir do uso múltiplos agentes especializados conectados a modelos de linguagem através da API *OpenRouter*[1]. O fluxo implementa quatro agentes distintos (*Agent, Agent1, Agent2 e Agent4*), cada um responsável por analisar um domínio específico dos dados consolidados: vulnerabilidades, conformidade SCA, inventário geral e síntese executiva. Estes agentes recebem dados estruturados através de nós preparatórios (*PrepareForLLM, PrepareVulnerabilityData, Statstics, PrepareIntroCAPInventory*) que formatam as informações segundo *templates* específicos, incluindo prompts contextualizados para orientar a análise do modelo de linguagem. Os insights gerados por cada agente são subsequentemente processados pelo nó *IAInsightsForReport,* que consolida as análises em um payload unificado destinado à geração do relatório. A conversão dos resultados em formato PDF ocorre através de nós *Puppeteer (PuppeterHtmlToPDF*), e uma requisição HTTP final para renderização do documento completo, demonstrando assim a viabilidade de automatizar integralmente o ciclo desde a coleta de dados de segurança até a produção de relatórios executivos enriquecidos com análise inteligente.

Os resultados obtidos nesta etapa demonstraram-se satisfatórios quanto à viabilidade técnica da abordagem proposta, validando a arquitetura geral do sistema e a integração entre coleta de dados, processamento e geração automatizada de relatórios através de agentes de IA. Contudo, as limitações inerentes à plataforma N8N tornaram-se evidentes à medida que os requisitos do projeto evoluíram, particularmente no que concerne à flexibilidade exigida para manipulação de estruturas de dados complexas, à necessidade de execução mais performática para processar volumes crescentes de informações, à implementação de testes automatizados para garantir a confiabilidade do sistema, ao emprego de módulos mais especializados e robustos que as capacidades nativas do JavaScript disponíveis na plataforma, e ao maior controle sobre o fluxo de execução e tratamento de exceções.

Diante destas constatações, optou-se por portar a solução inicial, já validada no ambiente N8N, para uma implementação utilizando linguagem tradicional de programação. A escolha pela

[1] Disponível em: https://openrouter.ai/

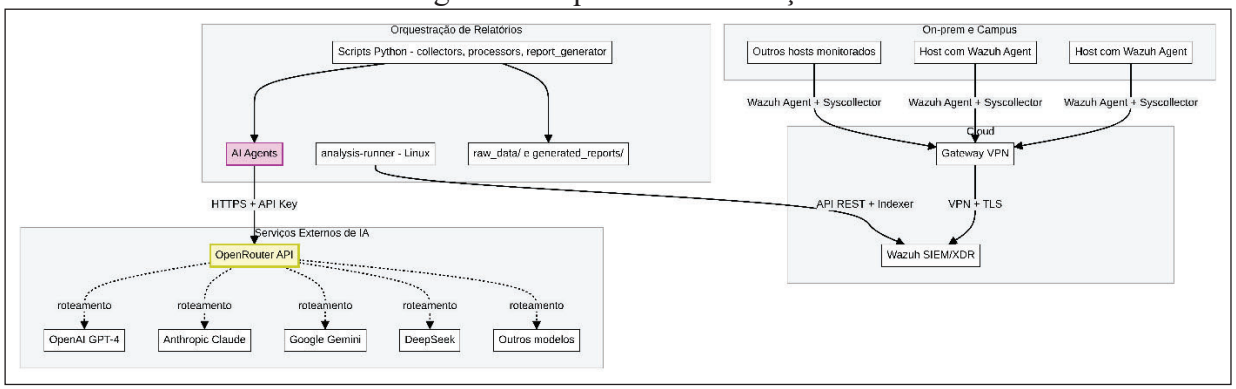
linguagem Python justificou-se pela familiaridade e experiência do autor com o ecossistema da linguagem, bem como pela ampla disponibilidade de bibliotecas especializadas em processamento de dados, integração com APIs, manipulação de documentos e frameworks para desenvolvimento de agentes de inteligência artificial, propiciando assim um ambiente de desenvolvimento mais adequado às demandas técnicas do projeto. De toda forma, o código gerado nesta etapa também é parte integrante do projeto e está disponibilizado no repositório[2].

4.3 IMPLEMENTAÇÃO

A partir dos testes e limitações observadas no protótipo partiu-se para implementação definitiva por meio dos agentes implementados em um pipeline utilizando linguagem de programação python. Esta Seção documenta a configuração do ambiente utilizado para a produção dos relatórios de conformidade. Primeiramente é feito o detalhamento da arquitetura global do projeto, a infraestrutura de monitoramento utilizada (Wazuh SIEM/XDR instalado em nuvem privada), o mecanismo de acesso dos coletores via VPN, e o pipeline de dados que vai do Wazuh até o relatório final gerado pelo módulo denominado *AIReportGenerator*. A seguir serão documentados os componentes de infraestrutura e integração essenciais (Wazuh SIEM/XDR em nuvem, pontes de rede/VPN, armazenamento de artefatos em *raw_data/ e generated_reports/*). É explicado o papel das camadas de coleta, processamento e geração de relatórios e o mecanismo e modelo de redução e controle de alucinações. A Figura 8 mostra o diagrama de arquitetura de alto nível do ambiente já a Figura 9 o fluxo de dados real do pipeline de coleta → processamento → consolidação → geração de relatório.

[2] Disponível em: https://github.com/ramisilva/wazuh-compliance-report/blob/main/scripts/Prototipo_fluxo_n8n_validacao.json

Figura 8. Arquitetura da Solução



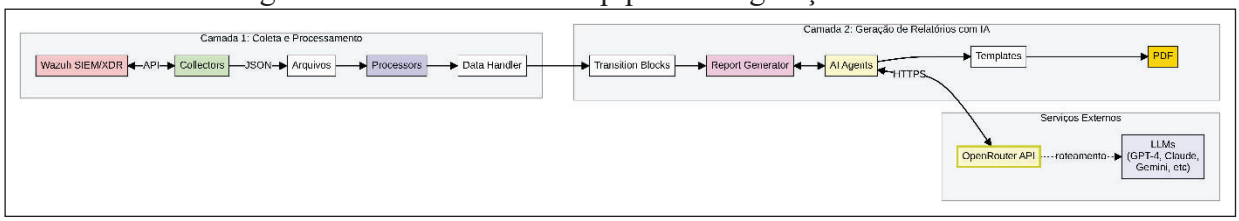
Fonte: Elaboração Própria

Como pode-se observar nas Figura 8 e 9, a implementação visa um pipeline modular para coletar, processar e transformar eventos de segurança e vulnerabilidades em relatórios e blocos de transição prontos para consumo humano ou por agentes de IA. Como camadas principais a solução apresenta:

- Coletores (raw_data/): ingestão de inventário, evidências SCA e saídas de *scanners* de vulnerabilidades.
- Processadores (modules/): normalização, agregação e enriquecimento (*CVE counts, SCA summaries*).
- Coleção de evidências (generated_reports/): CSV/JSON persistidos para auditoria e análise do resultado gerado.

- Módulo de transição (*scripts/build_transition.py*): empacotamento de blocos da transição com intuito de receber o contexto dos blocos de texto anterior e sucessor e proporcionar uma melhoria na fluidez do texto. (executive_summary, top-CVEs, evidences).
- Geração final de relatórios e validação pós-render (modules/report_generator.py e modules/report_utils/post_render_validator.py).

Figura 9. Fluxo de dados do pipeline de geração do relatório



Fonte: Elaboração Própria

A arquitetura da solução implementada estrutura-se em cinco módulos principais que operam em sequência para transformar dados brutos de segurança em relatórios analíticos enriquecidos. O fluxo inicia-se na plataforma Wazuh SIEM/XDR, que atua como fonte centralizada de dados de segurança, agregando logs, eventos e métricas de conformidade de toda a infraestrutura monitorada.

O ambiente utilizado para testes foi um conjunto de servidores de homologação da Universidade do Vale do Itajaí (6 servidores) em uma janela de coleta definida onde dados de inventário, *scan* de vulnerabilidades (Baseada na NVD (*Nist Vulnerability Database*)[3], verificação de conformidade com boas práticas de configuração e reforço de segurança (baseada no OpenScap[4]) foram persistidos pela plataforma Wazuh[5] para promover um conjunto maior de vulnerabilidades para fins de testes, foram incluídos ao ambiente de testes também um servidor linux propositalmente desatualizado e uma estação de trabalho do pesquisador para que, após a janela de coleta proporcionada pela Univali como testes, fosse possível continuar com dados novos coletados no cenário, bem como um conjunto maior de vulnerabilidades do que o encontrado no ambiente de homologação real.

O Wazuh é uma plataforma gratuita e de código aberto especializada em detecção de ameaças e monitoramento de segurança baseado em regras de segurança predefinidas. A plataforma possibilita o monitoramento de diversos tipos de *endpoints,* incluindo desktops, laptops, servidores, bem como dispositivos de rede tais como firewalls e roteadores, agregando e analisando dados em tempo real. Sua arquitetura fundamenta-se na coleta distribuída de dados através de agentes instalados nos sistemas monitorados, os quais reportam informações a um servidor central responsável pela

[3] Disponível em: https://nvd.nist.gov/
[4] Disponível em: https://www.open-scap.org/
[5] Disponível em: https://wazuh.com/

correlação e análise das informações de segurança. Esta abordagem permite às organizações uma visão consolidada e em tempo real de infraestrutura tecnológica. (Stanković; Gajin; Petrović, 2022)

A escolha pela plataforma se deu por ser de ampla utilização de mercado, com capacidade comparável as principais ferramentas que possuem custo de licenciamento e por ser de código aberto e gratuita permite a instalação em ambiente próprio. O comparativo de suas funcionalidades com sistemas correlatos de mercado é feito por (Šuškalo et al., 2023) onde define a solução como excelente custo benefício, com interface amigável e acessível a maioria das organizações. Como o foco deste trabalho é a aderência a empresas de médio porte se mostrou adequada como fonte de dados de segurança do projeto.

A plataforma expõe seus dados através de uma API REST autenticada via token JWT, permitindo a extração de dados, que no contexto deste trabalho são: inventário de ativos computacionais, resultados de auditorias de conformidade SCA (*Security Configuration Assessment*) e eventos de detecção de vulnerabilidades.

Os módulos Collectors, implementados na camada modules/collectors/ do sistema, estabelecem conexões bidirecionais autenticadas com a API do Wazuh, realizando requisições HTTP estruturadas para extrair dados específicos de cada categoria e persistindo as respostas em arquivos JSON na pasta raw_data. Dois são os módulos que fazem essa coleta:

- modules/inventory_collector.py (WazuhInventoryCollector): Responsável pela coleta dos dados de inventário, Syscollector do Wazuh.
- modules/data_collector.py (WazuhDataCollector): Responsável pela coleta dos eventos do Scan de vulnerabilidade e de conformidade SCA(OpenScap)

A etapa subsequente de processamento envolve a transformação e enriquecimento dos dados brutos através de dois componentes complementares: os processadores especializados e o módulo consolidador. Os processadores, implementados em modules/data_processor.py e módulos auxiliares como modules/vulnerability_processor.py, executam operações de agregação estatística, cálculo de métricas derivadas e enriquecimento determinístico dos dados coletados. Entre as transformações realizadas destacam-se: a contabilização de CVEs (*Common Vulnerabilities and Exposures*) categorizadas por nível de severidade, a consolidação de resultados de verificações SCA agrupadas por host, o cálculo de taxas de conformidade e scores de segurança, além do mapeamento enriquecido

entre endereços IP e nomes de host com suas respectivas distribuições de sistema operacional. Os dados resultantes deste módulo são persistidos em arquivos intermediários localizados no diretório raw_data:

- processed_inventory_data.json;
- processed_sca.json; e
- processed_vulnerabilities.json).

Já o módulo *Data Handler*, centralizado na função create_consolidated_data() do arquivo modules/report_utils/data_handler.py, assume então a responsabilidade de carregar os múltiplos arquivos JSON processados e unificá-los em uma estrutura de dados consolidada (consolidated_data) que organiza as informações em três domínios principais: inventário de ativos, vulnerabilidades detectadas e conformidade SCA. Adicionalmente, este módulo gera textos sumarizados (summary_text_for_ai) destinados ao consumo pelos agentes de inteligência artificial, além de construir tabelas determinísticas para renderização direta em relatórios, evitando assim possíveis alucinações dos modelos de linguagem.

A fase final de preparação para geração de relatórios é executada pelo *Transition Module*, implementado em scripts/build_transition.py, que encapsula a lógica de composição de blocos reutilizáveis de conteúdo. Este módulo produz componentes modulares como *executive_summary_block* (sumário executivo com contagens agregadas de vulnerabilidades e hosts), top_cves_table_block (tabela formatada em Markdown apresentando as vulnerabilidades mais críticas ordenadas por severidade) e evidence_links_block (referências estruturadas aos artefatos de evidência como arquivos JSON de dados brutos e gráficos estatísticos gerados). Estes blocos de transição são subsequentemente combinados com os dados consolidados e alimentados aos agentes de inteligência artificial especializados, cada um responsável pela análise de um domínio específico do relatório. Os agentes recebem como entrada tanto o objeto *consolidated_data* quanto uma base de conhecimento contextual (*knowledge_base*) carregada de arquivos de texto localizados em knowledge_base/, permitindo que as análises geradas pelos modelos de linguagem sejam fundamentadas não apenas nos dados coletados, mas também em políticas organizacionais, frameworks de conformidade e boas práticas de segurança previamente documentadas.

4.4 PRINCIPAIS CAMADAS

Cada uma das camadas da solução, descritas anteriormente, serão detalhadas a seguir para melhor compreensão da solução:

4.4.1 INTEGRAÇÃO DE DADOS

Este módulo é responsável por coletar e integrar os dados das fontes determinísticas do projeto, incluindo:

- Sistemas SIEM: Plataformas de gestão de eventos de segurança que são usadas para fornecer informações sobre vulnerabilidades, boas práticas de configuração de segurança bem como contexto de segurança de sistemas.

- *Frameworks de Compliance*: Regulamentações como ISO 27001 e LGPD servem como base para avaliar o status de conformidade da organização.

Como já detalhado anteriormente, a plataforma escolhida para uso foi o *Wazuh*. Como SIEM apresenta capacidade de monitoramento contínuo e disponibilidade de *APIs RESTful* que permitem a integração de dados. Essa integração garante que os eventos críticos de segurança sejam capturados em tempo real. Ainda neste Capítulo serão abordadas demais considerações acerca das escolhas tecnológicas de implementação da solução.

Também foi utilizada a plataforma SaaS *Openrouter* como broker de APIS para modelos de IA's generativas de mercado sua escolha se deu pela implementação de interface de API padronizada (Padrão de Fato percebido no mercado, OpenAI Chat API). A solução também encontra respaldo em outros trabalhos que fazem uso deste tipo de abordagem, onde é necessária a comparação de resultados a partir do uso de modelos diferentes disponíveis no mercado, um exemplo de uso semelhante do serviço é proposto por (Vijayakumar; Pyingkodi; Devi, 2025). A plataforma disponibiliza, atualmente, 573 modelos que podem ser consumidos via API. Este trabalho utilizou 6 APIs com base no critério: (listadas no *top used* da semana sendo 3 pagas e 3 com acesso gratuito.) Além disso, a plataforma também disponibiliza relatórios de bilhetagem que permitiram ampliar a análise também para uma perspectiva de custo que será discutida na Seção de resultados.

4.4.2 PROCESSAMENTO DE DADOS

Após a coleta, os dados são processados para extrair informações relevantes, tais como:

- Vulnerabilidades identificadas.
- Eventos de segurança classificados por severidade.
- Gerar resumos estatísticos como forma de reduzir a quantidade de dados encaminhada para o modelo de IA Generativa.

A camada de processamento de dados foi implementada como etapa intermediária entre a coleta bruta de informações e a análise por agentes de inteligência artificial, sendo responsável pela extração, transformação e enriquecimento das métricas de segurança coletadas. Esta camada foi implementada por meio de três processadores especializados que operam sobre os arquivos JSON brutos gerados pelos coletores. O processador de inventário (*inventory_processor.py*), implementado na classe *WazuhInventoryProcessor*, consome o arquivo *raw_inventory_data.json* e extrai métricas individuais de cada agente monitorado, incluindo sistema operacional completo (nome e versão concatenados), especificações de CPU (núcleos e frequência média), utilização de memória RAM (percentual calculado a partir das métricas *ram_usage e ram_total*), além de estatísticas agregadas do ambiente como total de hosts, distribuição de sistemas operacionais e médias de utilização de recursos computacionais. O resultado deste processamento é persistido em *processed_inventory_data.json*, estruturado de forma a facilitar tanto a renderização direta em *templates* quanto o consumo pelos agentes de IA em etapa posterior.

O processamento de vulnerabilidades, implementado em vulnerability_processor.py, consome o arquivo raw_vulnerabilities.json contendo os alertas do detector de vulnerabilidades do Wazuh e realiza múltiplas transformações analíticas. Primeiramente, extrai-se de cada alerta os campos relevantes de CVE (identificador, título, severidade, pontuação CVSS, vetor de ataque, referências e informações do pacote afetado), removendo-se dados redundantes ou de baixo valor informacional. Subsequentemente, o processador gera estatísticas agregadas incluindo contagem total de vulnerabilidades, distribuição por nível de severidade (Critical, High, Medium, Low), listagem das dez CVEs mais críticas ordenadas por pontuação CVSS, e mapeamento de vulnerabilidades por

agente com seus respectivos contadores. Um resumo textual estruturado em português (summary_text_for_ai) é gerado automaticamente para orientar os modelos de linguagem durante a análise, descrevendo quantitativamente o cenário de vulnerabilidades identificado. O processador de conformidade SCA (sca_processor.py), por sua vez, implementa a classe SCAProcessor que processa os resultados de *Security Configuration Assessment* extraindo, para cada política de segurança avaliada, métricas de conformidade incluindo score geral, total de verificações, contadores de aprovação/falha/não aplicável, identificação da política (ID e descrição) e listagem detalhada de cada verificação individual com título, resultado e descrição.

A consolidação final dos dados processados ocorre através do módulo *data_handler.py*, cuja função principal *create_consolidated_data()* atua como ponto de entrada único para construção da estrutura de dados que será consumida pelos agentes de inteligência artificial. Esta função carrega sequencialmente os arquivos *processed_inventory_data.json*, *processed_vulnerabilities.json* e *processed_sca.json*, organizando-os em um dicionário unificado (*consolidated_data*) com três chaves principais: *inventory, vulnerabilities e sca*. O enriquecimento adicional inclui a extração do arquivo bruto de inventário (*raw_inventory_data.json*) para compor uma lista *inventory['agents']* contendo, para cada host monitorado, os campos *agent_id, agent_name, agent_ip*, os (sistema operacional), sca_percent (percentual de conformidade agregado) e vulnerability_count (total de vulnerabilidades identificadas). Os resumos textuais (summary_text_for_ai) são incorporados tanto na seção de vulnerabilidades quanto na de conformidade SCA, fornecendo sínteses em linguagem natural que contextualizam os dados estruturados e orientam a análise dos modelos de linguagem. Este objeto consolidado é posteriormente distribuído integralmente aos três agentes especializados (*EnvironmentCharacterizationAgent, VulnerabilityAnalysisAgent e SCAAnalysisAgent*), assegurando consistência informacional e eliminando a necessidade de múltiplas leituras de arquivos durante a geração do relatório.

4.4.3 GERAÇÃO DE RELATÓRIOS

O terceiro módulo da solução é responsável por transformar os dados processados em relatórios personalizados, utilizando técnicas avançadas de IA Generativa. Este módulo recebe como entrada informações estruturadas provenientes do módulo anterior, que incluem dados específicos coletados pela plataforma SIEM, diretrizes sobre formato e linguagem, bem como referências a frameworks de boas práticas em cibersegurança.

A geração dos relatórios segue um processo sistemático, no qual cada persona-alvo (por exemplo, alta gestão, equipes técnicas ou *compliance officer*) é associada a prompts específicos desenvolvidos para orientar a construção textual gerada pela IA. Esses prompts são projetados com base em três dimensões fundamentais: o nível técnico do destinatário, as exigências regulatórias aplicáveis ao contexto organizacional e o estilo de comunicação preferido pela empresa.

Tal abordagem garante que os relatórios produzidos sejam não apenas precisos, mas também acessíveis e compreensíveis para diferentes públicos, promovendo uma melhor comunicação interna e alinhamento estratégico.

Os relatórios gerados neste módulo deverão adotar o formato *HTML e PDF,* uma escolha intencional que busca garantir compatibilidade com ferramentas amplamente utilizadas nas empresas e facilidade de acesso para as personas.

As APIs de IA Generativa empregadas neste módulo foram escolhidas de acordo com os critérios já descritos na subseção anterior com objetivo de avaliar sua eficiência em produzir:

- Relatórios de compliance reais, previamente gerados manualmente por especialistas, que servem como exemplos concretos da estrutura e conteúdo esperados;
- Linguagem específica adaptada a cada persona-alvo, visando refletir as nuances comunicacionais relevantes para cada grupo de destinatários;
- Por meio de dados brutos provenientes de eventos de segurança e vulnerabilidades, fornecendo à IA o conhecimento necessário para interpretar corretamente essas informações no contexto mais amplo da conformidade regulatória.

Por meio dessa combinação de dados, a IA Generativa tornou-se capaz de produzir relatórios que não apenas sintetizam informações críticas de segurança, mas também as apresentam de forma mais clara e contextualizada, atendendo às expectativas das personas-alvo. Esta abordagem representa uma contribuição significativa para a automação de processos de compliance, reduzindo o esforço humano necessário para a produção de relatórios e aumentando a eficiência operacional das organizações.

O módulo de geração de relatórios foi implementado na classe *AIReportGenerator* localizada em modules/report_generator.py, constitui-se como orquestrador central do processo de síntese

analítica, coordenando a execução sequencial de agentes especializados de inteligência artificial e consolidando suas produções em um documento HTML estruturado. A etapa inicial consiste no carregamento dos dados consolidados através do método *get_consolidated_data(),* que encapsula a função *create_consolidated_data()* do módulo *data_handler*, e na concatenação de todos os documentos de texto presentes no diretório *knowledge_base*/*.txt para enriquecimento contextual dos prompts. Os agentes de análise são executados sequencialmente, iniciando com o *EnvironmentCharacterizationAgent*, que gera a caracterização textual do ambiente monitorado; contudo, o gerador de relatórios sobrescreve deterministicamente os blocos de estatísticas de inventário (quantidade de hosts, distribuição de sistemas operacionais, métricas de memória RAM) para assegurar precisão factual. Subsequentemente, o VulnerabilityAnalysisAgent produz análise textual das vulnerabilidades identificadas, sendo que o gerador remove quaisquer listas ou tabelas de CVEs geradas pelo modelo de linguagem e reconstrói de forma determinística a tabela "Hosts mais afetados" a partir dos dados processados. O SCAAnalysisAgent complementa a camada analítica examinando o score de conformidade e os checks reprovados nas avaliações de configuração de segurança, esta técnica foi utilizada como forma de mitigar os riscos de geração de conteúdo sintético (alucinações) dos modelos de IA.

Após a geração das seções analíticas, o sistema executa três agentes de síntese que operam sobre o conteúdo previamente produzido e sobre os dados consolidados: o *ReportIntroductionAgent*, responsável pela introdução global do documento; o *ExecutiveSummaryAgent*, que produz o sumário executivo sintetizando as principais descobertas e métricas; e o *ConclusionAgent*, que elabora conclusões e recomendações estratégicas fundamentadas nas análises apresentadas. A montagem do relatório final ocorre através da renderização do template, utilizando o motor para templates, Jinja2[6] e a base para o relatório criada em *templates/report/base_report.html.j2*. O motor recebe como contexto todos os fragmentos HTML gerados pelos agentes e as métricas extraídas dos dados consolidados. O método _build_top_cves_block() constrói deterministicamente o bloco "Top CVEs (por impacto)", ordenando as vulnerabilidades por severidade e impacto potencial a partir do arquivo *processed_vulnerabilities.json*. O documento HTML resultante é persistido no diretório *generated_reports/Compliance+<PERSONA>+<MODEL>+<TIMESTAMP>/Relatorio_Final.ht*

[6] Disponível em: https://jinja.palletsprojects.com/en/stable/

ml, sendo opcionalmente convertido para formato PDF através da ferramenta wkhtmltopdf[7]. Adicionalmente, o sistema mantém um histórico de indicadores-chave de desempenho (KPIs) no arquivo *generated_reports/kpi_history.json*, possibilitando análises temporais da evolução das métricas de conformidade e segurança ao longo de múltiplas execuções do sistema, esta geração não é utilizada no contexto deste trabalho, mas pode ser de grande valia para pesquisas futuras como fonte histórica de dados de segurança do ambiente.

A etapa final do processo consiste na validação pós-renderização, implementada pela função *validate_and_clean_html()* localizada no módulo *post_render_validator.py*, que aplica sanitização e verificação de integridade referencial ao documento HTML gerado. Este validador remove elementos <li> que referenciam CVEs não presentes nos dados consolidados, elimina linhas de tabelas HTML que mencionam hosts inexistentes no inventário processado e normaliza todos os hyperlinks para o National Vulnerability Database (NVD) seguindo o padrão *https://nvd.nist.gov/vuln/detail/CVE-...*. O processo de validação gera um artefato de auditoria denominado *post_render_validation.json*, que documenta todas as remoções efetuadas, incluindo listas de CVEs e hosts excluídos do documento final, provendo rastreabilidade completa das transformações aplicadas e assegurando que o relatório entregue contenha exclusivamente informações verificáveis e consistentes com os dados coletados e processados nas etapas anteriores do pipeline.

A subseção a seguir é dedicada ao detalhamento da implementação da integração com as APIs dos modelos de IA generativa bem como a construção dos agentes utilizados no projeto.

4.4.4 Integração com OpenRouter e agentes de IA

A arquitetura central da solução fundamenta-se em um sistema multiagente especializado, implementado no módulo *agents/agents.py*, onde cada agente de inteligência artificial assume responsabilidade pela análise de um domínio específico do relatório de conformidade. Esta abordagem modular segue o padrão de design de separação de responsabilidades, permitindo que modelos de linguagem de grande escala (LLMs) concentrem sua capacidade analítica em contextos delimitados, reduzindo ambiguidades e aumentando a coerência técnica das produções textuais. A

[7] Disponível em: https://wkhtmltopdf.org/

Figura 10 ilustra o diagrama de arquitetura do módulo de agentes, evidenciando o fluxo de dados desde a consolidação até a geração de artefatos especializados por seção.

A classe base *BaseAgent*, definida em *agents/agents.py*, estabelece a infraestrutura comum para todos os agentes especializados, encapsulando a lógica de comunicação com provedores de modelos de linguagem, gerenciamento de *templates de prompts* e geração de artefatos de evidência. O método construtor *__init__* recebe como parâmetros o diretório de saída (output_dir), o identificador do modelo LLM a ser utilizado (model), e opcionalmente um *logger* customizado, inicializando estruturas internas para armazenamento. O método central *run(consolidated_data, knowledge_base)* orquestra o ciclo completo de execução do agente: carrega o template de prompt[8] correspondente à persona do agente a partir de *templates/prompts/<persona>/,* realiza a ingestão de variáveis com os dados consolidados e a base de conhecimento concatenada, invoca o modelo de linguagem através do cliente *OpenRouter*, processa a resposta estruturada e persiste tanto o HTML gerado quanto os metadados de execução em arquivos separados no diretório *generated_reports/<timestamp>/.* O método auxiliar *_make_openrouter_request* implementa a lógica de comunicação HTTP via API REST *do OpenRouter*, incluindo tratamento de erros, *retry* automático (máximo de três tentativas com backoff exponencial para contornar problemas com modelos que implementam limitação vazão de tokens no tempo) e timeout configurável de 60 segundos, retornando a resposta no formato padronizado *OpenAI-compatible*: *```python { "choices": [{"message": {"content": "..."}}], "usage": {"prompt_tokens": 1500, "completion_tokens": 800} }*

A partir desta resposta estruturada, o *BaseAgent* extrai o conteúdo textual gerado, e registra todas as métricas em logs de execução.

A hierarquia de agentes especializados deriva-se da classe base, implementando três agentes de análise primária e três agentes de síntese. O *EnvironmentCharacterizationAgent* produz a caracterização do ambiente tecnológico, processando dados de inventário (*consolidated_data['inventory']*) para gerar descrições de topologia, distribuição de sistemas operacionais, métricas de recursos computacionais (uso médio e máximo de RAM) e tabelas

[8] A documentação dos templates de prompt estão disponíveis no APÊNDICE 6 – DOCUMENTAÇÃO DOS TEMPLATES DE PROMPTS

estruturadas de hosts monitorados, incluindo identificador, nome, endereço IP e sistema operacional completo (nome e versão).

O *VulnerabilityAnalysisAgent* concentra-se na análise de vulnerabilidades detectadas, recebendo a lista completa de CVEs com suas métricas CVSS, severidades, pacotes afetados e referências técnicas (*consolidated_data['vulnerabilities']*), produzindo análise de distribuição por criticidade, identificação de vulnerabilidades de alta prioridade e geração de artefato CSV contendo a lista completa de achados com campos CVE, *Severity, Score, Package, Host e Description*. O *SCAAnalysisAgent* avalia os dados recebidos em relação a conformidade com políticas de configuração segura, analisando os resultados de *Security Configuration Assessment* (*consolidated_data['sca']*), calculando percentuais de conformidade por política CIS (Critical Security Controls[9]), identificando controles críticos reprovados e gerando tabelas de não-conformidades por host com recomendações técnicas de remediação. Os agentes de síntese (*ReportIntroductionAgent, ExecutiveSummaryAgent e ConclusionAgent)* operam em nível meta-analítico, recebendo os outputs dos agentes de análise primária e os blocos de transição pré-renderizados para produzir, respectivamente, a introdução contextual do relatório, o sumário executivo com indicadores-chave consolidados e a conclusão com recomendações estratégicas priorizadas.

A integração com provedores de modelos de linguagem realiza-se por meio da plataforma *OpenRouter*, que funciona como gateway unificado para múltiplos provedores LLM, permitindo alternância transparente entre diferentes modelos sem modificações no código dos agentes. As requisições seguem o padrão OpenAI Chat Completions API[10], com *endpoint*, cabeçalho de autenticação *Authorization: Bearer* <API_KEY> e *payload* JSON contendo o *array* de mensagens (*messages*), identificador do modelo (*model*), temperatura de amostragem (*temperature*, configurada como 0.3 para aumentar determinismo) e *token* máximo de resposta (*max_tokens*). A classe *BaseAgent* abstrai completamente esta camada de comunicação, permitindo que agentes especializados permaneçam agnósticos ao provedor específico. A Tabela 4. Custo dos modelos de IA como Serviço, apresenta os modelos de linguagem testados durante o desenvolvimento desta

[9] Informações em: https://www.cisecurity.org/controls/cis-controls-list
[10] Informações em: https://openrouter.ai/api/v1/chat/completions

dissertação, incluindo seus custos por milhão de tokens processados e características técnicas relevantes:

Tabela 4. Custo dos modelos de IA como Serviço

Provedor	***Custo (USD/1M tokens)***	***Características***
'z-ai/glm-4.5-air:free',	Grátis (experimental)	Qualidade alta, contexto 8k tokens, contexto 131K
'openai/gpt-oss-20b:free',	Grátis (experimental)	Excelente para análise técnica, contexto 200k
'x-ai/grok-4.1-fast',	Grátis (experimental)	Custo-benefício, bom raciocínio lógico, contexto 2M
'google/gemini-2.5-flash-lite',	Input: $0,5 / Output: $1,5	Rápido, contexto 1M tokens
'deepseek/deepseek-chat-v3-0324',	Input: $0,2 / Output: $0,88	Em testes, raciocínio avançado, contexto 164K
'google/gemini-3-pro-preview'	Input: $2,0 / Output: $12	Em testes, raciocínio avançado, contexto 1,05M

Fonte: Próprio Autor

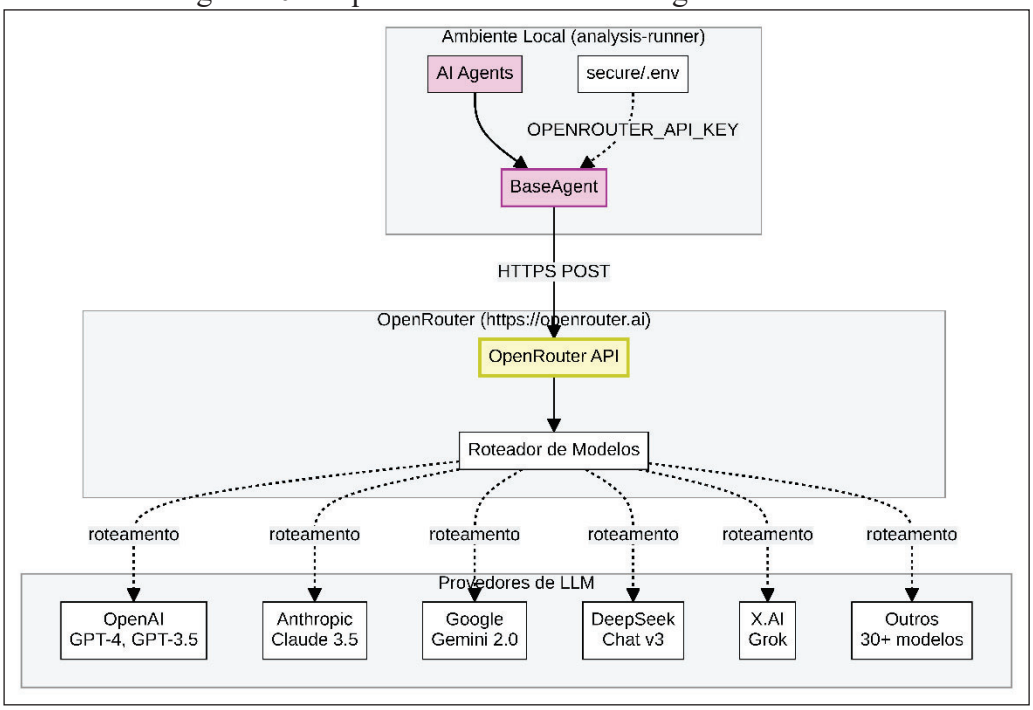
A configuração do módulo de agentes implementa práticas de segurança e auditabilidade essenciais para ambientes de produção. A chave de API do OpenRouter armazena-se no arquivo secure/.env (excluído do versionamento *Git via .gitignore*), sendo carregada automaticamente pelo módulo *python-dotenv* durante a inicialização dos agentes.

A questão da reprodutibilidade constitui desafio intrínseco a sistemas baseados em modelos de linguagem, dado o caráter não-determinístico das amostragens probabilísticas realizadas por LLMs, onde mesmo com temperatura configurada como zero há variabilidade estatística nas respostas geradas para prompts idênticos. Para mitigar este problema e assegurar reprodutibilidade metodológica adequada a um contexto de pesquisa acadêmica, a arquitetura implementa três mecanismos complementares.

Primeiro, todos os dados consolidados enviados como contexto aos agentes são persistidos em snapshots estruturados no diretório *evidence/data_context_for_<agent_name>.json*, permitindo reexecução exata do agente com contexto idêntico. Segundo os logs de requisições, arquivam tanto os prompts interpolados quanto as respostas completas dos modelos, possibilitando inspeção post-hoc de cada interação LLM e análise comparativa entre múltiplas execuções. Terceiro, quando o provedor oferece suporte (como em alguns modelos da OpenAI), o sistema utiliza *seeds*

determinísticas configuráveis no *payload* da requisição para reduzir variabilidade estatística, embora esta funcionalidade não seja universalmente disponível em todos os provedores integrados. A combinação destes três mecanismos assegura que, embora os textos gerados possam apresentar variações superficiais entre execuções, os achados técnicos, métricas e recomendações permaneçam consistentes. A arquitetura do módulo de agentes é ilustrada na Figura 10 a seguir:

Figura 10. Arquitetura do módulo de Agentes de IA



Fonte: Elaboração Própria

4.4.5 TECNOLOGIAS UTILIZADAS

A escolha das tecnologias foi guiada pelos critérios de funcionalidade, custo-benefício e viabilidade de implementação. Um detalhamento técnico das bibliotecas utilizadas no projeto encontra-se no Apêndice 1 e abaixo é feita uma descrição de alto nível para cada componente:

4.4.5.1 PLATAFORMA SIEM - WAZUH

O Wazuh foi selecionado como fonte primária de eventos de segurança devido às seguintes características:

- *Open-source* e gratuito, reduzindo custos operacionais.
- Alta capacidade de monitoramento e detecção de intrusões.
- API RESTful robusta para integração com outras ferramentas.
- Possibilidade de implementação em ambiente de testes

Comparativo com Outras Soluções SIEM

Foram pesquisadas 3 ferramentas similares de SIEM para fins de comparação e escolha conforme pode-se observar na Tabela 5. Comparativo das soluções SIEM a seguir.

Tabela 5. Comparativo das soluções SIEM

FERRAMENTA	OPEN SOURCE	FACILIDADE DE USO	INTEGRAÇÃO VIA API	CUSTO
Wazuh	Sim	Alta	Completa	Gratuito
Splunk	Não	Média	Parcial	Pago
ELK Stack	Sim	Baixa	Completa	Gratuito

Fonte: Elaborado pelo Autor

O Wazuh destacou-se por oferecer funcionalidades completas sem custos adicionais, tornando-o ideal implantação no contexto desta proposta e pela alta utilização observada no mercado atual em projetos de segurança.

4.4.5.2 FERRAMENTA DE AUTOMAÇÃO – N8N

A escolha do n8n como ferramenta de automação decorreu de suas capacidades avançadas de orquestração de workflows, permitindo conectar múltiplas APIs e realizar transformações complexas nos dados além de ser de código aberto.

Comparativo com Outras Ferramentas de Automação

Também foram pesquisadas 3 ferramentas similares de automação de fluxos e integração para fins de comparação e escolha conforme pode-se observar na Tabela 6 a seguir.

Tabela 6. Comparativo das soluções de automação

FERRAMENTA	OPEN SOURCE	FLEXIBILIDADE	FACILIDADE DE USO	ESCALABILIDADE
n8n	Sim	Alta	Média	Alta
Zapier	Não	Baixa	Alta	Baixa
Make.com	Não	Média	Alta	Média

Fonte: Elaborado pelo Autor

A escolha do n8n como ferramenta de automação decorreu de suas capacidades avançadas de orquestração de workflows, permitindo conectar múltiplas APIs e realizar transformações complexas nos dados. Além disso apresentou flexibilidade e escalabilidade, permitindo ajustes precisos ao fluxo de trabalho específico deste projeto além de ser de código aberto. Primeiro protótipo foi elabora por meio dessa plataforma, ne entanto, como já detalhado, optou-se pela implementação do pipeline proposto neste trabalho utilizando linguagem de programação *python*.

4.4.5.3 APIS DE IA GENERATIVA

As APIs de IA Generativa necessitariam de avaliação, e etapa de execução do projeto uma vez que dependiam da curadoria dos dados bem como definição dos formatos de saída. Testes preliminares, em etapa de projeto utilizou-se a diretamente o GROK[11] por facilidade e disponibilização de API sem custo, mas na etapa de desenvolvimento e execução do optou-se pelo Broker de Api's Openrouter já descrito anteriormente.

4.4.6 Testes e Execução do Fluxo Proposto

A seguir, apresenta-se a execução do fluxo de trabalho da solução implementada, desde a coleta inicial de dados até a geração final dos relatórios automatizados bem como a automação da execução da matriz de testes utilizando os 6 modelos de LLM para as 3 personas. Este processo foi executado para garantir que os dados brutos provenientes de plataformas SIEM fossem transformados em informações úteis e compreensíveis, adaptadas às necessidades específicas de diferentes personas dentro da organização, assim como a geração dos relatórios pudesse ser comparada em função do modelo de LLM utilizado, mantendo-se todas as demais condições idênticas.

[11] Disponível em: https://grok.com/

Este capítulo apresenta uma narrativa sequencial detalhada sobre o processo de execução do projeto, abrangendo desde a preparação do ambiente computacional até a geração de relatórios personalizados e a execução de testes matriciais. A execução segue uma arquitetura modular que separa claramente as etapas de coleta, processamento e geração de relatórios, permitindo execução independente e reutilização de dados intermediários.

4.4.6.1 Preparação do Ambiente de Execução

A preparação do ambiente constitui a etapa inicial e fundamental para garantir que todas as dependências necessárias estejam disponíveis e corretamente configuradas. Este processo envolve a configuração do interpretador Python, instalação de bibliotecas externas e preparação da estrutura de diretórios.

O projeto foi desenvolvido e testado utilizando Python 3.8 ou superior, sendo recomendada a versão 3.10 para compatibilidade com todas as funcionalidades implementadas. O ambiente de execução deve dispor de conectividade de rede para acesso à API do Wazuh e aos serviços de modelos de linguagem via OpenRouter.

A instalação das dependências Python é realizada por meio do gerenciador de pacotes pip, utilizando o arquivo ***requirements.txt*** que consolida todas as bibliotecas necessárias. O comando de instalação é executado no diretório raiz do projeto:

pip install -r requirements.txt

As principais dependências instaladas incluem:

- python-dotenv: Biblioteca para carregamento de variáveis de ambiente a partir de arquivos .env, facilitando a gestão de credenciais e configurações sensíveis sem hardcoding no código-fonte.

- requests: Cliente HTTP completo que gerencia as requisições à API REST do Wazuh, incluindo autenticação, tratamento de certificados SSL e gestão de timeouts.

- pandas: Framework de análise de dados que oferece estruturas de dados eficientes (DataFrame, Series) para manipulação, agregação e transformação de grandes volumes de dados de vulnerabilidades e inventário.

- matplotlib: Biblioteca de visualização que gera gráficos estatísticos incorporados aos relatórios, permitindo representação visual de distribuições de severidade e tendências temporais.

- seaborn: Extensão de alto nível para matplotlib que oferece paletas de cores profissionais e estilos de visualização otimizados para apresentação de dados de segurança.

Adicionalmente, para a geração de arquivos PDF a partir dos relatórios HTML, o sistema requer a instalação do utilitário *wkhtmltopdf,* que deve ser instalado diretamente a partir do gerenciador de pacotes do sistema operacional:

- *Ubuntu/Debian: sudo apt-get install wkhtmltopdf*
- *Red Hat/CentOS/Fedora: sudo yum install wkhtmltopdf*
- *macOSbrew install: wkhtmltopdf*

Na primeira execução, o sistema cria automaticamente os diretórios necessários para armazenamento de dados e relatórios:

- *raw_data/*: armazena dados brutos coletados diretamente da *API do Wazuh*

- *knowledge_base/*: Repositório de documentos de referência para contexto dos agentes de IA

- *generated_reports/*: Diretório de saída para relatórios gerados

- logs/: Registros de execução e diagnóstico (se habilitado)

Uma etapa também importante é a configuração das variáveis de ambiente que constitui aspecto crítico para a segurança e funcionamento correto do sistema. As credenciais e parâmetros operacionais são armazenados no arquivo .env, que deve ser criado a partir do template .env.example fornecido no repositório.

As variáveis relacionadas ao *Wazuh* configuram o acesso à plataforma SIEM/XDR e seus componentes:

WAZUH_API_URL: URL base da API REST do *Wazuh Manager*. Esta variável define o *endpoint* para todas as operações de coleta de dados, incluindo consultas a agentes, vulnerabilidades

e políticas de conformidade. Exemplo: https://192.168.194.10:55000. A configuração correta desta URL é essencial para o estabelecimento da conexão com o servidor Wazuh.

WAZUH_API_USER e WAZUH_API_PASSWORD: Credenciais de autenticação para acesso à API. Estas variáveis devem conter um usuário com permissões de leitura sobre os recursos de agentes, vulnerabilidades e SCA. Por questões de segurança, estas credenciais nunca devem ser versionadas em repositórios de código, permanecendo exclusivamente no arquivo .env local.

VERIFY_SSL: Flag booleana que controla a verificação de certificados SSL/TLS durante as requisições HTTPS. Em ambientes de produção, deve permanecer como *true* para garantir autenticidade e confidencialidade das comunicações. O valor false é aceitável apenas em ambientes de desenvolvimento ou laboratório com certificados auto assinados, após análise de risco apropriada.

WAZUH_INDEXER_URL, WAZUH_INDEXER_USER e WAZUH_INDEXER_PASSWORD: Variáveis opcionais que configuram acesso direto ao Wazuh Indexer (baseado em Elasticsearch) para consultas avançadas ou coleta de dados históricos não disponíveis via API REST. Estas credenciais são necessárias apenas se o módulo de coleta realizar consultas diretas ao índice de eventos.

Além disso, o sistema implementa mecanismos de controle de taxa (*rate limiting*) para evitar sobrecarga nos servidores Wazuh durante operações de coleta intensiva:

VULN_PAGE_SIZE: Define o número de registros de vulnerabilidades solicitados por requisição paginada. Valores elevados (1000) reduzem o número total de requisições, mas podem causar timeouts ou erros HTTP 429 (*Too Many Requests*) em servidores com recursos limitados. O ajuste fino deste parâmetro requer análise empírica: iniciar com 1000 e reduzir progressivamente (750, 500) caso ocorram falhas.

VULN_PAGE_DELAY_SECONDS: Intervalo de espera entre requisições consecutivas de páginas de vulnerabilidades. Este atraso artificial mitiga o risco de sobrecarga do servidor e evita bloqueios por *rate limiting*. Valores típicos variam entre 0.1 segundos (servidores robustos) e 1.0 segundo (ambientes com recursos limitados). O valor deve ser calibrado considerando o volume de dados e a capacidade do servidor Wazuh.

Já as variáveis de configuração para acesso aos modelos de linguagem de larga escala (LLMs) via OpenRouter definem o comportamento dos agentes de geração de conteúdo:

OPENROUTER_API_KEY: Chave de autenticação para acesso à plataforma OpenRouter. Esta chave deve ser obtida através do cadastro no serviço[12] e tratada como credencial sensível, pois autoriza o consumo de recursos computacionais que podem gerar custos monetários.

OPENROUTER_API_BASE_URL: URL base da API do OpenRouter. O valor padrão https://openrouter.ai/api/v1 deve ser mantido, exceto em casos de uso de proxies corporativos ou ambientes de teste isolados.

OPENROUTER_MODEL: Identificador do modelo de linguagem a ser utilizado para geração de conteúdo. A escolha do modelo impacta diretamente na qualidade, velocidade e custo da geração.

Na plataforma os modelos disponíveis, no momento da execução deste teste estavam na ordem de 573.

Na sequência, deve-se definiar os parâmetros para a Geração dos Relatórios, os parâmetros disponíveis são:

GENERATE_TRANSITIONS_BY_AI: Flag numérica (1 ou 0) que habilita ou desabilita a geração automática de transições narrativas entre seções do relatório. Quando ativada (valor 1), o sistema invoca o módulo de transições para gerar parágrafos de conexão que melhoram a coesão textual. A desativação (valor 0) resulta em relatórios mais diretos, com seções justapostas sem conectivos explícitos.

RENDER_PDF: Flag numérica que controla a geração automática de versão PDF do relatório HTML. O valor 1 ativa a conversão através do wkhtmltopdf, enquanto 0 gera apenas a versão HTML. A geração de PDF é recomendada para distribuição formal de relatórios, mas pode ser desabilitada para reduzir tempo de processamento em ambientes de desenvolvimento.

REPORT_OUTPUT_DIR: Caminho do diretório de destino para relatórios gerados. O valor padrão generated_reports/ pode ser personalizado para integração com sistemas de gestão documental ou repositórios corporativos.

---

[12] Disponível em: https://openrouter.ai

Depois da definição das variáveis necessárias para execução do projeto já se está apto para execução da primeira etapa, a coleta dos dados.

4.4.6.2 Etapa de Coleta de Dados

A etapa de coleta constitui o primeiro passo operacional do pipeline, sendo responsável pela extração de dados primários da plataforma Wazuh. Esta fase é executada através de dois módulos especializados que operam de forma independente.

Coleta de Inventário de Ativos

O módulo *inventory_collector* realiza a extração do inventário completo de agentes registrados no Wazuh Manager, incluindo metadados de sistema operacional, endereçamento de rede e estado de conectividade:

python3 -m modules.inventory_collector

Este comando executa as seguintes operações:

Estabelece conexão autenticada com a API REST do Wazuh utilizando as credenciais configuradas e consulta o endpoint /agents para obtenção da lista completa de agentes gerenciados.

Para cada agente, coleta informações detalhadas incluindo:

- Identificador único e nome do host
- Sistema operacional (distribuição, versão, arquitetura)
- Endereço IP e informações de rede
- Data de registro e último contato
- Estado de conexão (ativo, desconectado, pendente)
- Serializa os dados coletados em formato JSON estruturado
- Persiste o resultado no arquivo raw_data/wazuh_agents_inventory.json

Coleta de Dados de Vulnerabilidades e SCA

O módulo *data_collector* é responsável pela extração de dados de vulnerabilidades (CVEs) e resultados de Security Configuration Assessment:

python3 -m modules.data_collector

Esta operação executa um processo mais complexo devido ao volume potencialmente grande de vulnerabilidades:

Autentica na API do Wazuh e obtém lista de agentes ativos e, para cada agente, realiza requisições paginadas ao endpoint /vulnerability para coleta de CVEs detectados, implementa controle de taxa baseado nas variáveis *VULN_PAGE_SIZE e VULN_PAGE_DELAY_SECONDS,* consulta o *endpoint* /sca para obtenção de resultados de verificações de conformidade, agrega os dados coletados em estruturas JSON.

O módulo implementa mecanismos de checkpoint para permitir retomada da coleta em caso de interrupção, evitando reprocessamento de dados já coletados.

4.4.6.3 Etapa de Processamento de Dados

Após a coleta bem-sucedida, os dados brutos passam por uma etapa de processamento e enriquecimento que prepara as informações para consumo pelos agentes de geração de relatórios.

O processador de inventário realiza enriquecimento e agregação estatística dos dados de ativos a execução deste módulo se dá por meio do comando a seguir:

python3 -m modules.inventory_processor

As operações executadas incluem:

Leitura do arquivo de inventário bruto coletado na etapa anterior, aplicação de mapeamento de nomes de host para identificadores anônimos ou corporativos (se configurado em secure/host_mapping.json), classificação e agrupamento de agentes e geração de arquivo processado raw_data/processed_inventory.json com estrutura otimizada para consumo pelos agentes de relatório

4.4.6.4 Processamento de Vulnerabilidades e SCA

O processador de dados de segurança realiza análise, classificação e priorização de vulnerabilidades:

python3 -m modules.data_processor

Este módulo executa transformações complexas sobre os dados brutos:

Carrega dados de vulnerabilidades e SCA coletados, realiza enriquecimento de CVEs com informações de CVSS (score, vetor, severidade), agrupa vulnerabilidades sua classificação, identifica e prioriza Top 10 vulnerabilidades críticas com maior impacto, processa resultados de SCA, agrupa por política de conformidade aplicada, calcula taxa de compliance por política, identifica Top 10 verificações com não conformidade, gera estatísticas agregadas e persiste dados processados em arquivos estruturados para consumo rápido.

4.4.6.5 Geração de Relatórios por Persona

Com os dados processados disponíveis, o sistema pode gerar relatórios personalizados para diferentes perfis de audiência. Esta etapa utiliza o pipeline completo de geração, incluindo invocação dos agentes de IA para produção de conteúdo adaptado.

Execução para Persona CEO

A geração de relatório para o perfil CEO parametriza o agente com os templates de prompt que enfatizam impacto de negócio, riscos estratégicos e recomendações executivas, sua execução se dá por meio do comando:

python3 -m modules.run_full_pipeline --persona ceo --skip-collect --skip-process --render-pdf

Os parâmetros utilizados possuem as seguintes funções:

- --persona ceo: Define o perfil de audiência como Chief Executive Officer, ajustando linguagem, nível de detalhe técnico e foco em métricas de negócio
- --skip-collect: Omite a etapa de coleta, reutilizando dados já existentes em raw_data/
- --skip-process: Omite o processamento, utilizando dados já processados
- --render-pdf: Força a geração de versão PDF mesmo se RENDER_PDF estiver desabilitado no .env

O relatório gerado para CEO inclui:

- Sumário executivo com visão geral de riscos

- Indicadores-chave de segurança (KPIs)
- Impacto potencial de vulnerabilidades críticas no negócio
- Recomendações priorizadas por custo-benefício
- Roadmap de remediação com estimativas de esforço

4.4.6.6 Execução para Persona TIC (CISO/Diretor de TI)

O relatório para perfil técnico-gerencial equilibra profundidade técnica com visão de governança:

python3 -m modules.run_full_pipeline --persona tic --skip-collect --skip-process --render-pdf

Este relatório apresenta:

- Análise detalhada de vulnerabilidades por categoria
- Avaliação de postura de conformidade (SCA)
- Mapeamento de riscos por ativo crítico
- Recomendações técnicas de *hardening* e *patching*
- Indicadores de eficácia de controles de segurança

4.4.6.7 Execução para Persona Compliance Officer

O relatório voltado para conformidade regulatória enfatiza aderência a frameworks e normas:

python3 -m modules.run_full_pipeline --persona compliance_officer --skip-collect --skip-process --render-pdf

O conteúdo gerado foca em:

- Status de conformidade com políticas aplicáveis (PCI-DSS, CIS, etc.)
- Gaps de conformidade identificados

- Evidências de controles implementados
- Recomendações para remediação de não-conformidades

A arquitetura proposta faz com que, independentemente da persona, o fluxo de geração siga a dinâmica de múltiplos agentes onde, a partir do carregamento dos dados consolidados de inventário, vulnerabilidades e SCA a execução siga por:

- Invocação de agentes especializados: Cada seção do relatório é gerada por um agente dedicado:
- *ReportIntroductionAgent:* Contextualização e objetivos
- *EnvironmentCharacterizationAgent:* Descrição do ambiente tecnológico
- *VulnerabilityAnalysisAgent:* Análise de CVEs e riscos
- *SCAAnalysisAgent:* Avaliação de conformidade
- *ExecutiveSummaryAgent*: Síntese de descobertas
- *ConclusionAgent:* Conclusões e recomendações
- *Geração de transições:* Se habilitado, o *TransitionGenerator* cria parágrafos de conexão entre seções
- *Renderização HTML*: O conteúdo gerado é injetado em templates Jinja2 para produção do HTML final
- Validação pós-renderização: O *PostRenderValidator* verifica *placeholders* não resolvidos e inconsistências
- Conversão para PDF: Se solicitado, o HTML é convertido para PDF via *wkhtmltopdf*
- Persistência: O relatório é salvo em diretório *timestamped* sob *generated_reports*/
- Cada execução gera uma pasta nomeada no formato:

Compliance+<PERSONA>+<MODEL>+<TIMESTAMP>

- A pasta gerada contém:

  Relatorio_Final.html: Versão HTML do relatório

  Relatorio_Final.pdf: Versão PDF (se gerado)

  data_context_for_xxx.json: Metadados de contexto utilizado como parâmetro para cada agente)

  post_render_validation.json: indicação dos trechos e/ou dados que foram descartados pela barreira de alucinação

  pasta evidence: com os arquivos brutos de cada etapa, coleta e processamento para fins de auditabilidade e/ou consulta full para ações técnicas especializadas de correção do ambiente.

4.4.6.8 Execução da Matriz de Testes

A matriz de testes constitui uma funcionalidade que permite a geração em lote dos relatórios para avaliação comparativa de múltiplos modelos de linguagem para diferentes personas, gerando um conjunto abrangente de relatórios para análise de qualidade e consistência.

A matriz de testes foi desenvolvida para responder às seguintes questões:

Comparabilidade entre modelos: Diferentes LLMs produzem relatórios de qualidade equivalente?

Consistência por persona: Um mesmo modelo mantém coerência ao adaptar-se para diferentes audiências?

Relação custo-benefício: Modelos pagos justificam o investimento em comparação com alternativas gratuitas?

Para responder estas questões, o script *generate_matrix_reports.py* automatiza a execução de combinações persona × modelo, gerando uma matriz completa de relatórios. O script define

configurações padrão que cobrem os principais perfis de audiência e uma seleção representativa de modelos:

Personas padrão:

- ceo: Chief Executive Officer
- compliance_officer: Oficial de Conformidade
- tic: Diretor de TI/CISO

Modelos padrão:

- z-ai/glm-4.5-air:free: Modelo gratuito baseline
- openai/gpt-oss-20b:free: Alternativa gratuita com 20B parâmetros
- google/gemini-2.5-flash-lite: Modelo rápido otimizado
- x-ai/grok-4.1-fast: Modelo de alta performance
- deepseek/deepseek-chat-v3-0324: Especialista conversacional
- google/gemini-3-pro-preview: Modelo premium avançado

Esta configuração resulta em uma matriz de 3 × 6 = 18 execuções completas do pipeline de geração. A

Figura 11. Listagem dos relatórios gerados em lote ilustra os relatórios gerados, neste caso, em pasta única, para facilitar a análise visual e posteriormente a automação comparativa.

Figura 11. Listagem dos relatórios gerados em lote

Nome	Tamanho	Data da modificação
old reports/		23/11/2025, 17:41:30
CEO_deepseek_deepseek-chat-v3-0324_2025_11_24_00_57.html	20.7 kB	24/11/2025, 00:57:59
CEO_google_gemini-2.5-flash-lite_2025_11_24_00_52.html	26.6 kB	24/11/2025, 00:52:06
CEO_google_gemini-3-pro-preview_2025_11_24_01_01.html	30.9 kB	24/11/2025, 01:01:52
CEO_openai_gpt-oss-20b_free_2025_11_24_00_51.html	25.3 kB	24/11/2025, 00:51:27
CEO_x-ai_grok-4.1-fast_2025_11_24_00_55.html	20.2 kB	24/11/2025, 00:55:06
CEO_z-ai_glm-4.5-air_free_2025_11_24_00_50.html	24.8 kB	24/11/2025, 00:50:24
COMPLIANCE_OFFICER_deepseek_deepseek-chat-v3-0324_2025_11_24_01_15.html	28.0 kB	24/11/2025, 01:15:47
COMPLIANCE_OFFICER_google_gemini-2.5-flash-lite_2025_11_24_01_10.html	30.8 kB	24/11/2025, 01:10:13
COMPLIANCE_OFFICER_google_gemini-3-pro-preview_2025_11_24_01_19.html	33.4 kB	24/11/2025, 01:19:10
COMPLIANCE_OFFICER_openai_gpt-oss-20b_free_2025_11_24_01_09.html	27.9 kB	24/11/2025, 01:09:29
COMPLIANCE_OFFICER_x-ai_grok-4.1-fast_2025_11_24_01_13.html	28.5 kB	24/11/2025, 01:13:23
COMPLIANCE_OFFICER_z-ai_glm-4.5-air_free_2025_11_24_01_08.html	32.4 kB	24/11/2025, 01:08:29
TIC_deepseek_deepseek-chat-v3-0324_2025_11_24_01_36.html	24.0 kB	24/11/2025, 01:36:20
TIC_google_gemini-2.5-flash-lite_2025_11_24_01_29.html	30.1 kB	24/11/2025, 01:29:59
TIC_google_gemini-3-pro-preview_2025_11_24_01_39.html	33.9 kB	24/11/2025, 01:39:53
TIC_openai_gpt-oss-20b_free_2025_11_24_01_29.html	28.4 kB	24/11/2025, 01:29:22
TIC_x-ai_grok-4.1-fast_2025_11_24_01_33.html	24.3 kB	24/11/2025, 01:33:46
TIC_z-ai_glm-4.5-air_free_2025_11_24_01_28.html	28.9 kB	24/11/2025, 01:28:12

Fonte: Elaboração Própria

Pode observar pela

Figura 11, a partir da análise do *timestamp* de última modificação do arquivo, que os 18 relatórios foram gerados em 49 minutos, o que representa uma média de 2,7 min/relatório gerado. Observou-se que grande parte do tempo de geração do relatório está concentrada na resposta dos modelos de IA utilizados, alguns modelos demoram mais para responder e este tempo tem relação com o tipo de modelo (simples ou de raciocínio profundo) ou com custo por acesso, os modelos "premium" tendem a ter tempo de resposta mais rápido.

Uma observação importante, para fins de otimização e garantia de replicação idêntica do processo de geração, a coleta e o processamento dos dados foram feitos uma única vez, no início do processo em lote. Portanto, o tempo de coleta não está representado nesta análise. Para fins informativos observou-se que o processo todo, coleta e processamento, levou em média 50 s no ambiente de testes. Este dado pode variar muito em função da infraestrutura disponível e configuração da plataforma Wazuh. Para fins de análise os relatórios gerados neste testes e que mais performaram a partir da análise final podem ser conferidos no Apêndice 5 (foram incluídos 3 relatórios, 1 para cada persona)

A execução completa da matriz é iniciada com comando:

python3 scripts/generate_matrix_reports.py

Este script executa as seguintes operações:

- Validação de ambiente: Verifica presença de variáveis de ambiente necessárias (OPENROUTER_API_KEY)
- Criação de diretório de saída: Prepara *generated_reports/batch_outputs/* para consolidação dos resultados
- Geração de plano de execução: Computa todas as combinações persona × modelo (18 execuções)
- Execução iterativa: Para cada combinação: a. Configura variável de ambiente OPENROUTER_MODEL para o modelo atual; b. Invoca o *pipeline* completo: python3 -m modules.report_generator --persona <persona>; c. Aguarda conclusão da geração; d. Identifica o diretório de saída recém-criado em generated_reports/ e. Copia

o arquivo Relatorio_Final.html para o diretório de consolidação; f. Renomeia o arquivo seguindo padrão: <PERSONA>_<MODEL>_<TIMESTAMP>.html

- Consolidação: Todos os relatórios gerados são coletados em um único diretório para análise comparativa

Antes de executar a matriz completa (que pode consumir tempo e créditos de API significativos), é possível executar em modo de simulação:

python3 scripts/generate_matrix_reports.py --dry-run

Este modo executa todas as validações e exibe o plano de execução sem invocar os modelos de IA permitindo verificação de:

- Corretude das variáveis de ambiente
- Disponibilidade de dados processados

O script permite de geração em lote permite a personalização através de argumentos de linha de comando:

Seleção de personas específicas:

python3 scripts/generate_matrix_reports.py --personas ceo tic

Seleção de modelos específicos:

python3 scripts/generate_matrix_reports.py --models "google/gemini-3-pro-preview" "deepseek/deepseek-chat-v3-0324"

Alteração do diretório de saída:

python3 scripts/generate_matrix_reports.py --output-dir custom_comparison_reports

Ou a combinação de parâmetros:

```
python3 scripts/generate_matrix_reports.py \
  --personas ceo compliance_officer \
  --models "google/gemini-3-pro-preview" "x-ai/grok-4.1-fast" \
```

--output-dir premium_models_comparison

Após conclusão da execução em lote, os relatórios consolidados em *generated_reports/batch_outputs/* podem ser analisados através de:

Inspeção visual: Comparação manual da qualidade narrativa, coerência e profundidade técnica

- Scripts de análise quantitativa: Ferramentas auxiliares que extraem métricas:
- Comprimento de seções
- Densidade de termos técnicos
- Presença de *placeholders* não resolvidos
- Contagem de recomendações específicas vs. genéricas
- Avaliação de custo-efetividade: Correlação entre custo de geração e qualidade percebida
- Análise de consistência: Verificação de divergências entre múltiplas execuções do mesmo modelo

A execução da matriz de testes constitui elemento fundamental para validação científica da abordagem proposta, fornecendo evidências empíricas sobre a viabilidade e qualidade do uso de LLMs para geração automatizada de relatórios de segurança.

A cabo da descrição narrativa do processo de execução do projeto, cobrindo desde a preparação inicial do ambiente até a geração automatizada de matrizes de testes é possível concluir que a arquitetura modular implementada permite execução flexível e eficiente, com separação clara de responsabilidades e reaproveitamento de dados intermediários. Além disso, a capacidade de geração de relatórios personalizados para múltiplas personas e a execução de testes matriciais com diversos modelos de linguagem demonstram a robustez e escalabilidade da solução proposta.

Com os relatórios gerados buscou-se a avaliação dos resultados por meio de algumas perspectivas que são apresentadas a seguir.

5 AVALIAÇÃO E RESULTADOS

5.1 METODOLOGIA DE AVALIAÇÃO E ANÁLISE DOS RESULTADOS

5.1.1 Metodologia de Avaliação

A avaliação da solução proposta para a geração automatizada de relatórios de segurança, utilizando dados reais de SIEM (Wazuh) e modelos de Linguagem de Grande Escala (LLM), foi conduzida por meio de um framework de avaliação estruturado e um questionário administrado ao especialista no assunto e executivo de TIC do ambiente real utilizado nos testes. Este arcabouço metodológico visou garantir uma análise abrangente e multifacetada da qualidade e eficácia dos relatórios gerados.

5.1.1.1 Framework de Avaliação Proposto

O framework de avaliação, pode ser conferido em sua versão completa no Apêndice 2 – Metodologia de Avaliação), foi concebido para analisar os relatórios sob três dimensões ortogonais, consideradas cruciais para o contexto de relatórios de segurança gerados por IA: Qualidade Linguística (A), Alinhamento ao Propósito (Persona-Fit) (B) e Desempenho do Modelo de IA (C). Cada dimensão foi subdividida em métricas específicas, acompanhadas de descrições claras e indicadores de escala para facilitar a avaliação.

A. Qualidade Linguística: Esta dimensão focou na clareza, legibilidade, coerência, estrutura, adequação da terminologia e tom/formalidade dos relatórios. O objetivo foi verificar se o texto era facilmente compreensível pelo público-alvo, se apresentava um fluxo lógico entre as seções e se utilizava a linguagem apropriada para cada persona (CEO, TIC, *Compliance Officer*).

B. Alinhamento ao Propósito (Persona-Fit): Esta dimensão avaliou a relevância do conteúdo para as necessidades específicas de cada persona, a acionabilidade das recomendações, a comunicação de risco (mapeando achados técnicos para impactos de negócio/regulatórios) e a cobertura dos artefatos requeridos (tabelas de evidências, listas de CVEs, etc.). A métrica B5, especificamente, avaliou a concisão e utilidade do sumário executivo para a tomada de decisões.

C. Desempenho do Modelo de IA: Esta dimensão investigou a consistência dos dados factuais entre os relatórios gerados por diferentes modelos para a mesma persona, a fidelidade aos dados

originais (precisão dos valores numéricos), a profundidade da análise (discussão de tendências, mapeamento de frameworks, custo-benefício), a originalidade do texto (equilíbrio entre texto padrão e narrativa específica) e a taxa de erros factuais ou gramaticais.

A pontuação geral para cada relatório foi calculada como uma soma ponderada das médias das dimensões, com as seguintes ponderações: 30% para Qualidade Linguística (A), 40% para Alinhamento ao Propósito (B) e 30% para Desempenho do Modelo de IA (C). Esta ponderação foi escolhida para enfatizar a importância do alinhamento do conteúdo às necessidades da persona, que é um dos pilares da proposta de solução.

5.1.1.2 Instrumento de Avaliação Formal (Questionário)

O instrumento de avaliação formal consistiu em um questionário detalhado, apresentado também no (Apêndice 2 – Tabela 18. Questionário (Instrumento de Avaliação Formal), administrado a especialista no assunto (Responsável pelo ambiente real testado) para cada persona. Cada item do questionário foi respondido em uma escala *Likert* de 5 pontos (1 = Discordo Totalmente, 5 = Concordo Totalmente), exceto quando indicado de outra forma. Este questionário permitiu uma avaliação sistemática e padronizada dos relatórios, capturando a percepção qualitativa dos especialistas em relação a cada métrica do framework.

5.1.1.3 Coleta de Dados

A coleta de dados para a análise de resultados envolveu a geração de 18 relatórios distintos, compreendendo três personas (CEO, TIC, Compliance Officer) e seis modelos de LLM diferentes (*DeepSeek Coder, Google Gemini 2.5 Flash Lite, Google Gemini 3 Pro Preview, Mistral Large, OpenAI GPT-OSS 20B, XAI Grok 4.1 Fast e ZAI GLM 4.5*).

Os relatórios foram avaliados por um especialista e pela avaliação do Autor. Além da avaliação humana, foi realizada uma análise automatizada por uma IA Generativa, utilizando o modelo Gemini 3 Pro (o prompt utilizado para produzir a análise pode ser conferido no Apêndice 3), para complementar e validar as avaliações dos especialistas. As respostas brutas do questionário, que formaram a base para as pontuações apresentadas, estão no Apêndice 4 – Respostas Brutas do Questionário). A seguir apresentamos as análises em relação aos resultados alcançados.

5.1.2 Análise dos resultados

Pela abrangência deste trabalho a necessidade de avaliação comparativa de desempenho entre modelos se fez necessária, contudo, a eficiência e resultados desta pesquisa abrangem a análise do conjunto da arquitetura de geração automatizada de relatórios desenvolvida, considerando que a qualidade dos outputs não depende exclusivamente das capacidades intrínsecas dos LLMs, mas também da robustez do sistema que os orquestra. A arquitetura proposta foi projetada para endereçar já sabendo que haveriam desafios críticos sobre aplicação de IA generativa em contextos corporativos: (1) garantia de consistência entre múltiplas execuções, (2) mitigação de alucinações mediante ancoragem em dados verificáveis, (3) utilização de informações reais estruturadas em substituição a inferências não fundamentadas, (4) gerenciamento controlado de variabilidade entre modelos e personas, e (5) rastreabilidade auditável de todo o processo de geração. Os resultados dessa análise arquitetural, apresentados a seguir, demonstram que a combinação adequada de processamento estruturado de dados, *prompt engineering* contextualizado e templates especializados constituem fatores chave de sucesso para viabilizar a adoção de IA generativa em cibersegurança corporativa.

Como ponto inicial da análise, observou-se que a avaliação comparativa dos seis modelos de linguagem de grande escala (LLMs) na geração de relatórios de cibersegurança revelou variações de desempenho entre modelos e personas. A análise contemplou 18 relatórios (6 modelos × 3 personas), totalizando 252 avaliações individuais (com base no questionário de avaliação proposto) distribuídas em três dimensões principais: Qualidade Linguística (A), Alinhamento ao Propósito (B) e Desempenho do Modelo de IA (C).

O *framework* de avaliação aplicado utilizou uma escala *Likert* de 5 pontos para cada critério, permitindo a mensuração objetiva de características qualitativas dos relatórios. A pontuação geral foi calculada mediante ponderação diferenciada das dimensões (A: 30%, B: 40%, C: 30%), refletindo a importância relativa do alinhamento ao propósito da persona na utilidade prática dos relatórios gerados, já que o alinhamento do relatório com a persona constitui fator de eficiência proposta no contexto deste trabalho. A seguir já é apresentada a classificação geral quanto ao desempenho dos modelos utilizados na arquitetura proposta, assim, as demais análises pormenorizadas já podem contar com uma visão já apresentada.

Ranking Geral de Desempenho

A Tabela 7 apresenta a classificação consolidada dos modelos avaliados, considerando a média aritmética das pontuações obtidas nas três personas.

Tabela 7. Ranking Geral de Desempenho por Modelo

POSIÇÃO	MODELO	CEO	COMPLIANCE OFFICER	TIC	MÉDIA GERAL
1º	Google Gemini 2.5 Flash Lite	4.94	5.00	5.00	**4.98**
2º	X.AI Grok 4.1 Fast	4.94	4.94	4.94	**4.94**
3º	DeepSeek Chat v3	4.43	5.00	4.92	**4.78**
4º	Google Gemini 3 Pro Preview	4.14	4.44	4.29	**4.29**
5º	Z.AI GLM 4.5 Air Free	4.14	4.14	4.22	**4.17**
6º	OpenAI GPT-OSS 20B Free	3.43	3.51	3.43	**3.46**

Fonte: Próprio Autor

Os resultados demonstram uma amplitude significativa de desempenho, com o modelo líder (Gemini 2.5 Flash Lite, 4.98) superando o modelo de menor pontuação (GPT-OSS 20B Free, 3.46) em 1,52 pontos, equivalente a uma diferença de 43,9% no desempenho relativo. Esta variação substancial sugere que a escolha do modelo de IA tem impacto direto e mensurável na qualidade dos relatórios de segurança gerados.

Performance por Dimensão de Avaliação

A Tabela 8 detalha o desempenho médio dos modelos em cada dimensão avaliativa, permitindo identificar padrões de forças e fraquezas específicas.

Tabela 8. Performance Média por Dimensão de Avaliação

MODELO	A (LINGUÍSTICA)	B (PROPÓSITO)	C (IA)	CONSISTÊNCIA CROSS-PERSONA
Gemini 2.5 Flash Lite	5.00	5.00	4.93	★★★★★ Excelente
Grok 4.1 Fast	5.00	5.00	4.80	★★★★★ Excelente
DeepSeek Chat v3	4.83	4.73	4.80	★★★★ Boa
Gemini 3 Pro Preview	4.42	4.20	4.27	★★★ Regular
GLM 4.5 Air Free	4.25	4.07	4.20	★★★★ Boa
GPT-OSS 20B Free	3.75	3.27	3.40	★★ Fraca

Fonte: Próprio Autor

A partir das médias aritméticas das pontuações obtidas nas três personas observa-se que os modelos líderes (Gemini 2.5 e Grok 4.1) alcançaram pontuação máxima (5.00) nas dimensões A e B

em todas as personas, diferenciando-se apenas na dimensão C (Desempenho do Modelo de IA), onde questões de precisão factual e verificabilidade impactaram a avaliação final.

ANÁLISE POR PERSONA

Persona CEO (Chief Executive Officer)

Os relatórios gerados com foco na persona CEO demandam ênfase em impacto estratégico de negócio, linguagem executiva sem jargões técnicos excessivos e tradução de riscos técnicos em consequências financeiras e reputacionais mensuráveis. A Tabela 9 traz os resultados consolidados de cada dimensão avaliada, por modelo no contexto do CEO.

Tabela 9. Pontuações Detalhadas da Persona CEO

MODELO	A (LINGUÍSTICA)	B (PROPÓSITO)	C (IA)	PONTUAÇÃO GERAL
Gemini 2.5 Flash Lite	5.00	5.00	4.80	**4.94**
Grok 4.1 Fast	5.00	5.00	4.80	**4.94**
DeepSeek Chat v3	4.50	4.40	4.40	**4.43**
Gemini 3 Pro Preview	4.25	4.00	4.20	**4.14**
GLM 4.5 Air Free	4.25	4.00	4.20	**4.14**
GPT-OSS 20B Free	3.75	3.20	3.40	**3.43**

Fonte: Próprio Autor

Pode ser observado que os modelos**, Gemini 2.5 Flash Lite e Grok 4.1 Fast,** alcançaram pontuação idêntica (4.94), mas que se diferenciaram por abordagens complementares:

- **Gemini 2.5 Flash Lite**: Destacou-se pela inclusão de estimativas de investimento verificáveis (R$ 150-200 mil, fase 1, R$ 300-400, mil fase 2), permitindo ao CEO tomar decisões orçamentárias fundamentadas. Identificou com clareza o single point *of failure* (82 das 91 vulnerabilidades críticas concentradas em um host específico do ambiente de testes), traduzindo padrões técnicos em risco estratégico de negócio.

- **Grok 4.1 Fast**: apresentou quantificação financeira de riscos (R$ 500 mil a R$ 1 milhão por dia de *downtime* potencial, multas LGPD de até 2% do faturamento ou R$ 50 milhões). Introduziu abordagem inovadora de "investimento defensivo com ROI mensurável", tratando

segurança como ativo estratégico com projeção de redução de 70% do risco crítico em 90 dias.

Já o modelo **DeepSeek Chat v3** (avaliado em 4.43) demonstrou linguagem executiva e precisão factual sem alucinações, porém perdeu pontos pela ausência de métricas, elemento crítico para tomada de decisão executiva.

Ao final, o **GPT-OSS 20B Free** (nota 3.43) apresentou limitações severas em completude (B2: 3.0) e profundidade analítica (C3: 3.0), com recomendações genéricas ("priorizar correção", "melhorar score de conformidade") sem roadmap detalhado, prazos específicos ou análise com métricas corporativas.

Persona Compliance Officer

Para este perfil, os relatórios devem ser orientados a *frameworks* regulatórios (LGPD, ISO 27001, NIST, CIS), com mapeamento preciso de *gaps* de conformidade para artigos legais específicos, quantificação de exposição regulatória e ferramentas de rastreabilidade para gestão de múltiplas auditorias. A Tabela 10 apresenta os resultados consolidados, por modelo no contexto do Compliance Officer.

Tabela 10. Pontuações Detalhadas da Persona Compliance Officer

MODELO	A (LINGUÍSTICA)	B (PROPÓSITO)	C (IA)	PONTUAÇÃO GERAL
DeepSeek Chat v3	5.00	5.00	5.00	5.00
Gemini 2.5 Flash Lite	5.00	5.00	5.00	5.00
Grok 4.1 Fast	5.00	5.00	4.80	4.94
Gemini 3 Pro Preview	4.50	4.40	4.40	4.44
GLM 4.5 Air Free	4.25	4.00	4.20	4.14
GPT-OSS 20B Free	3.75	3.40	3.40	3.51

Fonte: Próprio Autor

Pode-se observar que os modelos**, DeepSeek Chat v3 e Gemini 2.5 Flash Lite** alcançaram pontuação máxima (5.00), representando excelência em relatórios de conformidade por abordagens complementares:

- **DeepSeek Chat v3**: Inovou com seção "Evidências para Auditoria", listando especificamente documentos a coletar por fase (GPOs aplicadas, logs de acesso, relatórios de conformidade de senha, screenshots de configuração

- **Gemini 2.5 Flash Lite**: Apresentou a "Matriz de Controles" com rastreabilidade cruzada entre múltiplos frameworks (LGPD ↔ ISO 27001 ↔ NIST ↔ CIS), permitindo gestão simultânea de auditorias paralelas. Quantificou precisamente exposição regulatória (72.59% de não conformidade).

Um ponto importante que foi observado no **Grok 4.1 Fast** (nota 4.94) foi a quantificação financeira de multas potenciais: "Exposição LGPD Art. 52: multa de até 2% do faturamento, estimada em R$ 50 milhões para empresa de médio porte". Incluiu "Checklist de Preparação para Auditoria" com 14 itens verificáveis, porém perdeu pontos em C1 (Precisão Factual: 4.0) por usar estimativas de multas baseadas em inferências sobre faturamento não confirmáveis nos dados brutos.

Persona TIC (Tecnologia da Informação e Comunicação)

Para este perfil o conteúdo deve ser mais técnico operacional com CVEs específicos, scores CVSS, comandos de remediação executáveis, análise de vetores de ataque e ferramentas de validação técnica (re-scan, logs, testes). A Tabela 11 apresenta os resultados consolidados, para os relatórios gerado para a persona TIC.

Tabela 11. Pontuações Detalhadas da Persona TIC

MODELO	A (LINGUÍSTICA)	B (PROPÓSITO)	C (IA)	PONTUAÇÃO GERAL
Gemini 2.5 Flash Lite	5.00	5.00	5.00	**5.00**
Grok 4.1 Fast	5.00	5.00	4.80	**4.94**
DeepSeek Chat v3	5.00	4.80	5.00	**4.92**
Gemini 3 Pro Preview	4.50	4.20	4.20	**4.29**
GLM 4.5 Air Free	4.25	4.20	4.20	**4.22**
Gemini 2.5 Flash Lite	5.00	5.00	5.00	**5.00**

Fonte: Próprio Autor

Com base no resultado, o modelo **Gemini 2.5 Flash Lite** alcançou a pontuação máxima (5.00), com destaque para inclusão de comandos operacionais específicos, importante no contexto técnico de possível remediação. Além disso, apresentou mapeamento completo de vetor de ataque: "Exploração remota via porta 80/443 → RCE (CVE-2024-56180) → escalação de privilégios → pivoting para rede interna 10.1.201.x". Incluiu KBs de patches específicos (KB5043936) e procedimentos de validação técnica (re-scan via nmap, confirmação de ausência de CVE).

Já o **Grok 4.1 Fast** (nota 4.94) inovou com "*Runbook* de Resposta a Incidentes" completo, incluindo IOCs operacionais e SIEM queries prontas para integração. Forneceu informações de ExploitDB IDs (#52341), Metasploit modules ('windows/http/deserialization_rce') e IOCs específicos (User-Agent: 'python-requests/2.28'). Ao final este modelo perdeu pontos em precisão factual (C1: 4.0) por usar inferências tecnicamente plausíveis, mas não verificáveis nos dados brutos originais.

Como destaque para o **DeepSeek Chat v3** (4.92) observou-se a análise técnica profunda: "host01 concentra 90% das vulnerabilidades críticas (82/91), configurando single point of failure. Também trouxe análise de dependências: host roda serviços web críticos, comprometimento permite pivoting para rede interna". Ao final ainda incluiu checklist operacional detalhado (backup pré-patch, aplicação em janela de manutenção, testes pós-patch, validação via Wazuh), mas sem comandos específicos prontos. A partir deste ponto serão apresentadas as análises consolidadas quanto as dimensões avaliadas por meio do modelo de avaliação proposta.

ANÁLISE DIMENSIONAL DETALHADA

Como primeira dimensão analisada a qualidade linguística traz como resultado as métricas observadas na Tabela 12.

Tabela 12. Performance Média em Qualidade Linguística por Critério

MODELO	A1 (CLAREZA)	A2 (ESTRUTURA)	A3 (VOCABULÁRIO)	A4 (AUSÊNCIA DE ERROS)	MÉDIA A
Gemini 2.5 Flash Lite	5.00	5.00	5.00	5.00	**5.00**
Grok 4.1 Fast	5.00	5.00	5.00	5.00	**5.00**
DeepSeek Chat v3	4.83	5.00	5.00	5.00	**4.83**
Gemini 3 Pro Preview	4.33	4.50	4.50	4.33	**4.42**
GLM 4.5 Air Free	4.17	4.33	4.17	4.33	**4.25**
GPT-OSS 20B Free	3.67	3.83	3.67	3.83	**3.75**

Fonte: Próprio Autor

Com principais achado tem-se que os modelos líderes (Gemini 2.5 e Grok 4.1) alcançaram pontuação máxima em todos os critérios linguísticos, demonstrando ausência total de erros gramaticais, estruturação lógica consistente e vocabulário perfeitamente calibrado por persona. O critério A3 (Adequação do Vocabulário) foi determinante na adaptação contextual: modelos top

utilizaram linguagem executiva para CEO ("alocação de recursos", "impacto reputacional"), terminologia jurídica para Compliance ("gaps críticos", "baseline de conformidade") e jargão técnico para TIC ("exploitabilidade", "vetor de ataque", "hardening").

Já a dimensão B, relativa ao alinhamento de propósito, traz como resultado métricas que permitem avaliar o alinhamento da solução na produção dos textos adaptados ao contexto das personas e os resultados podem ser observados na Tabela 13.

Tabela 13. Performance Média em Alinhamento ao Propósito por Critério

MODELO	B1 (RELEVÂNCIA)	B2 (COMPLETUDE)	B3 (ACIONABILIDADE)	B4 (OBJETIVIDADE)	B5 (PERSUASÃO)	MÉDIA B
Gemini 2.5 Flash Lite	5.00	5.00	5.00	5.00	5.00	**5.00**
Grok 4.1 Fast	5.00	5.00	5.00	5.00	5.00	**5.00**
DeepSeek Chat v3	5.00	4.67	4.67	5.00	4.33	**4.73**
Gemini 3 Pro Preview	4.33	4.00	4.00	4.33	4.33	**4.20**
GLM 4.5 Air Free	4.00	4.00	4.00	4.33	4.00	**4.07**
GPT-OSS 20B Free	3.33	3.00	3.33	3.33	3.33	**3.27**

Fonte: Próprio Autor

A análise do critério B3 (Acionabilidade) revela que: modelos que incluíram ferramentas operacionais específicas (comandos, checklists, matrizes, *runbooks*) obtiveram B3=5.0, enquanto modelos com recomendações genéricas ficaram limitados a B3≤3.0. Esta diferença de 67% (3.0→5.0) confirma operacionalização como multiplicador de utilidade prática, não característica secundária.

A métrica B2 (Completude) foi determinante para CEO: modelos sem estimativas de investimento (DeepSeek, GLM, GPT-OSS) perderam até 2,0 pontos neste critério (B2=3,0), pois CEOs não podem aprovar ações sem quantificação financeira. Para Compliance, completude exigiu mapeamento legal específico (artigos da LGPD, controles ISO 27001, CIS Controls) - modelos com referências genéricas a "normas" ficaram limitados a B2=3.0.

No contexto da dimensão C, que avalia o desempenho do modelo de IA, é um dos pontos-chave do trabalho e traz como resultado métricas que permitem avaliar de fato se o uso dos modelos foi satisfatório, ponto importante da proposta deste trabalho, os resultados dessa dimensão, agrupados

por: C1 (Precisão), C2 (Consistência), C3 (Profundidade), C4 (Criatividade) e C5 (Adaptação), podem ser observados na Tabela 14.

Tabela 14. Performance Média em Desempenho de IA por Critério

MODELO	C1	C2	C3	C4	C5	MÉDIA C
Gemini 2.5 Flash Lite	5.00	5.00	5.00	4.67	5.00	**4.93**
Grok 4.1 Fast	4.00	5.00	5.00	4.67	5.00	**4.80**
DeepSeek Chat v3	5.00	5.00	4.67	4.67	5.00	**4.80**
Gemini 3 Pro Preview	5.00	5.00	4.00	3.67	4.33	**4.27**
GLM 4.5 Air Free	5.00	5.00	4.00	3.00	4.00	**4.20**

Fonte: Próprio Autor

Ao analisar a métrica C1 (Precisão Factual) foi verificado que o resultado foi alto em todos os modelos (≥4.0), com apenas Grok 4.1 obtendo C1=4.0 (vs. 5.0 dos demais). Tal modelo teve desempenho inferior por usar inferências não verificáveis (ExploitDB IDs, IOCs específicos, estimativas de multas baseadas em faturamento inferido). Este resultado confirma maturidade geral dos LLMs avaliados em evitar alucinações críticas quando fornecidos com dados estruturados via prompt engineering adequado e a eficiência da arquitetura proposta com este objetivo.

Já em C3 (Profundidade de Análise) observou-se que: GPT-OSS ficou limitado a C3=3.0 por análises superficiais sem exploração de implicações (não identificou SPOF, não analisou vetores de ataque, não explorou dependências de serviços). Modelos top (C3=5.0) forneceram análises multicamadas: identificação de padrões (concentração 82/91 em host01) + diagnóstico de causa (SPOF) + análise de consequências (*pivoting* para rede interna) + recomendações específicas (isolamento + *hardening* + monitoramento).

Ao fina, a análise de C4 (Criatividade) permitiu medir a inovação em ferramentas: modelos que incluíram elementos únicos (Matriz de Controles, *Runbook* de Incidentes, Seção de Evidências para Auditoria) obtiveram C4≥4.67. Modelos com estrutura padrão sem diferenciação ficaram limitados a C4=3.0. Além das dimensões formais presentes na metodologia de avaliação proposta alguns achados em relação a solução serão debatidos a seguir.

PADRÕES IDENTIFICADOS

Ao analisar os resultados pode-se uma discussão também importante é a capacidade da solução proposta de adaptação contextual a partir da consistência Cross-Persona. Neste contexto, observou-se que o **Grok 4.1 Fast** demonstrou alta consistência sem variações entre personas (4.94 em todas), enquanto **DeepSeek Chat v3** apresentou variação de 0.57 pontos (4.43→5.00→4.92). Também pode-se destacar, a partir da análise dos relatórios, que o Grok mantém vocabulário específico calibrado por audiência sem comprometer estrutura como pode-se observar nos trechos extraídos dos relatórios:

- **CEO:** "paralisia operacional", "danos à marca", "alocação de recursos"
- **Compliance:** "gaps de conformidade", "exposição regulatória", "baseline de conformidade"
- **TIC:** "exploitabilidade", "vetor de ataque", "superfície de exposição"

Na sequência, o modelo, DeepSeek, demonstrou especialização evidente em Compliance (5.00 vs. 4.43 em CEO), sugerindo otimização deliberada para contextos regulatórios e, neste viés, destaca-se que**:** Para organizações que necessitam relatórios consistentes para múltiplas audiências, modelos generalistas com alta consistência cross-persona (Grok, Gemini 2.5) são preferíveis. Já para organizações com necessidade dominante em Compliance, especialistas como DeepSeek podem oferecer vantagem marginal.

Seguindo para a análise da dimensão B, em especial a B3 (Acionabilidade) pode-se observar que a presença de ferramentas operacionais específicas é ponto importante para eficiência do resultado, com isso destacaram-se aos modelos com operacionalização máxima (B3=5.0):

- **Gemini 2.5 TIC:** Comandos PowerShell executáveis, KBs de patches, comandos nmap de validação
- **Grok 4.1 TIC:** Runbooks, IOCs, SIEM queries prontas, Metasploit modules
- **DeepSeek Compliance:** Lista específica de evidências documentais por fase de auditoria
- **Gemini 2.5 Compliance:** Matriz de rastreabilidade multi-framework

Também é importante avaliar os modelos com resultado limitado (B3≤3.5):

- **GPT-OSS todas as personas:** Recomendações genéricas sem comandos, procedimentos ou ferramentas

- **GLM 4.5 Air TIC:** Ações descritas sem especificação de comandos ou validação técnica

Ao final da análise dessa dimensão específica, houve Gap de operacionalização entre Gemini 2.5 TIC (B3=5.0, comandos prontos) e GPT-OSS TIC (B3=3.0, recomendações genéricas) resultando em uma diferença de 1.57 pontos na pontuação geral (5.00 vs. 3.43), equivalente a 45% de gap de utilidade prática. Este achado demonstra que operacionalização é diferencial competitivo que determina viabilidade de uso corporativo.

Como próxima dimensão de análise considerou-se o Trade-off entre Impacto Persuasivo e Verificabilidade Factual e observou-se que, no modelo de avaliação proposto acarretou que, os modelos que incluíram quantificação financeira de riscos e investimentos alcançaram B5 (Persuasão) significativamente superior:

Com quantificação financeira (B5=5.0):

- **Gemini 2.5 CEO:** "Investimento fase 1: R$ 150-200 mil" (verificável via cotação de mercado)

- **Grok 4.1 CEO:** "Downtime: R$ 500k-1M/dia; Multa LGPD: R$ 50MM" (inferências baseadas em premissas)

Sem quantificação financeira (B5=4.0):

- **DeepSeek CEO:** Foco em exposição técnica sem tradução financeira

- **Gemini 3 Pro CEO:** Análise de impacto qualitativa sem quantificação

Mas, tal fato introduz, contudo, um trade-off em C1 (Precisão Factual). Observou-se, por exemplo, que o Grok 4.1 obteve C1=4.0 (vs. 5.0 dos demais) por incluir estimativas não verificáveis nos dados brutos:

- Multa LGPD de R$ 50 milhões pressupõe faturamento de R$ 2.5 bilhões (não confirmado)

- Downtime de R$ 500k-1M/dia requer premissas sobre receita horária (não fornecidas)

- ExploitDB IDs e IOCs específicos são tecnicamente plausíveis, mas não rastreáveis aos dados originais

Neste contexto, o resultado obtido com o **Gemini 2.5** mitigou este trade-off ao usar estimativas verificáveis externamente (cotações de mercado para serviços de remediação, faixas de investimento baseadas em projetos similares) ou explicitar premissas quando necessário. Diante deste fato observou-se um alerta quanto a metodologia onde deve-se ponderar que quantificações financeiras aumentam impacto persuasivo (+25% em B5) mas exigem:

1. Explicitação de metodologia de cálculo e premissas assumidas
2. Uso de faixas em vez de valores pontuais quando houver incerteza
3. Rastreabilidade a fontes externas verificáveis ou explicitação de inferências

Como ponto chave da eficiência do uso de IA Generativa a Precisão Factual é um Pré-Requisito Mínimo e com isso observou-se que:

Todos os modelos avaliados mantiveram C1 (Precisão Factual) ≥4.0, com zero alucinações críticas em dados brutos:

- Métricas numéricas (32.819 total, 91 críticas, 82 em host01): 100% de acurácia em todos os modelos
- CVEs citados: todos verificáveis em bases públicas (NVD, MITRE)
- Scores CVSS: consistentes com *databases* de referência

Contudo, ausência de alucinações não garantiu utilidade corporativa concluindo que a metodologia foi eficiente em ponderar outros pontos para o resultado, tal situação fez com que o **GPT-OSS 20B** mantivesse C1=5.0 (precisão factual perfeita) mas obteve pontuação geral de apenas 3.46 por deficiências severas em:

- **B2 (Completude): 3.0** - falta *roadmap* detalhado, estimativas, ferramentas operacionais
- **C3 (Profundidade): 3.0** - análise superficial sem exploração de implicações
- **B3 (Acionabilidade): 3.0-3.5** - recomendações genéricas sem especificidade

Isso pode levar a conclusão de que, todo o esforço da solução em manter precisão factual, por meio do uso de blocos determinísticos, *templates* de contexto robusto e análise posterior de *placeholders* verificados é condição necessária, mas não suficiente para qualidade de relatórios corporativos. Modelos devem combinar ausência de alucinações com completude analítica, operacionalização e adaptação contextual para alcançar utilidade prática.

EFICIÊNCIA DA ARQUITETURA DE GERAÇÃO AUTOMATIZADA DE RELATÓRIOS

Apesar de se considerar a avaliação comparativa entre modelos ponto importante, este trabalho propôs e desenvolveu uma arquitetura de geração automatizada de relatórios de segurança, a qual é fator de grande importância no objetivo geral definido o que demonstrou eficiência notável em múltiplas dimensões. Esta seção analisa os benefícios arquiteturais independentemente do modelo de IA selecionado.

Consistência Estrutural e Padronização

A arquitetura desenvolvida utiliza templates de prompts segregados por persona (*templates/prompts/<persona>/*), garantindo estruturação consistente dos relatórios independentemente do modelo de IA utilizado. Esta abordagem resultou em:

Benefício 1: Estrutura Previsível Todos os 18 relatórios gerados (6 modelos × 3 personas) mantiveram estrutura lógica consistente:

- Sumário Executivo com mensagem principal em destaque
- Diagnóstico técnico com métricas quantificadas
- Análise priorizada (por severidade, framework ou host)
- Recomendações faseadas com *timeline*
- Próximos passos ou checklist de implementação

Benefício 2: Auditabilidade e Rastreabilidade O uso de *consolidated_data* como fonte única de verdade garantiu rastreabilidade completa:

- Todos os dados numéricos nos relatórios são rastreáveis

Por meio da análise de processed_sca_data.json e processed_vulnerability_data.json

- *Timestamps* e metadados preservados (timestamp_utc, data_hora_formatada)
- Separação clara entre dados brutos, dados processados e inferências de IA

Evidência quantitativa: Dos 252 critérios avaliados, C1 (Precisão Factual) obteve média de 4.83/5.00 (96.6% de acurácia), com apenas 1 modelo (Grok) pontuando abaixo de 5.0 por usar inferências adicionais além dos dados fornecidos.

Controle de Alucinações via Prompt Engineering Estruturado

A arquitetura implementou três mecanismos de controle de alucinações que resultaram em taxa de alucinação crítica de **0%** em todos os modelos avaliados os mecanismos propostos e implementados foram:

Mecanismo 1: O fornecimento de dados estruturados pode ser observado no trecho de dados extraído do *consolidated_data* utilizado pela solução. Ao fornecer dados pré-processados e estruturados, a arquitetura eliminou necessidade do modelo "inferir" estatísticas básicas, reduzindo drasticamente risco de alucinação numérica.

```
# Exemplo de consolidated_data fornecido ao modelo
{
  "sca": {
    "score_percentual": 27.41,
    "score_formatado": "27.41%",
    "total_verificacoes": 164,
    "checks_passou": 45,
    "checks_falhou": 119
  },
  "vulnerabilities": {
    "total": 32819,
    "criticas": 91,
    "altas": 4196,
    "por_host": {
      "cloud.safera": {"criticas": 82, "altas": 3867}
    }
  }
}
```

Mecanismo 2: Instruções Explícitas sobre Uso de Dados Templates de prompts incluem diretrizes específicas como pode ser observado, a seguir, no trecho de um dos templates de prompt utilizados:

*"**IMPORTANTE:** Baseie toda a análise exclusivamente nos dados fornecidos. Não invente métricas, CVEs ou vulnerabilidades. Se determinada informação não estiver disponível nos dados, mencione explicitamente a limitação."*

Os dados reais utilizados ficam evidenciado na arquitetura onde utilizou-se de dois tipos de dados:

Dados Reais do Ambiente (Data-Driven):

- Scans Wazuh SCA processados via sca_processor.py
- Vulnerabilidades do Wazuh processadas via vulnerability_processor.py
- Inventário de máquinas de processed_inventory_data.json
- Métricas de uso de recursos (RAM, CPU) quando disponíveis

Base de Conhecimento de Segurança (Knowledge-Driven):

- Frameworks de compliance (LGPD, ISO 27001, NIST SP 800-53, CIS Controls v8)
- Regulamentações e artigos legais aplicáveis

Esta separação permitiu que modelos combinassem **precisão factual** (dados reais) com **profundidade analítica** (conhecimento de segurança). Tal escolha evidencia o impacto positivo no resultado onde obteve-se:

- **C1 (Precisão Factual):** Média 4.83/5.00 - dados numéricos 100% corretos
- **C3 (Profundidade de Análise):** Média 4.07/5.00 - análises enriquecidas com frameworks
- **B1 (Relevância):** Média 4.39/5.00 - conexão bem-sucedida entre dados técnicos e contexto regulatório/estratégico

Um exemplo de enriquecimento via base de conhecimento pode ser observado no trecho, a seguir, de um dos relatórios:

Dado real: "Score SCA: 27.41%"

Análise enriquecida (Gemini 2.5 Compliance): "Score SCA de 27.41% indica não conformidade com 72.59% dos controles, configurando exposição simultânea a:

- LGPD Art. 46 (tratamento inadequado de dados pessoais)
- ISO 27001 Anexo A.9 (8 de 14 controles de Controle de Acesso não conformes)
- CIS Control 5 (políticas de senha inadequadas)"

Outro ponto observado a partir dos resultados é que a arquitetura permitiu mensuração objetiva de variabilidade entre modelos pois manteve-se inputs idênticos, revelando três padrões de variabilidade:

Tipo 1: Variabilidade de Qualidade (Desejada) Diferenças de 1.52 pontos (43.9%) entre melhor (Gemini 2.5: 4.98) e pior (GPT-OSS: 3.46) modelo refletem diferenças reais de capacidade dos LLMs em:

- Compreensão contextual (adaptação por persona)
- Profundidade analítica (identificação de padrões)
- Operacionalização (geração de ferramentas específicas)

Tipo 2: Variabilidade de Abordagem (Neutra) Modelos top alcançaram pontuações similares com estratégias diferentes:

- **Gemini 2.5:** Operacionalização máxima + estimativas verificáveis
- **Grok 4.1:** Quantificação financeira agressiva + runbooks
- **DeepSeek:** Especialização em preparação auditável

Esta variabilidade permite seleção por contexto de uso, não apenas por ranking absoluto.

Tipo 3: Variabilidade Factual (Indesejada - Ausente) Zero variabilidade foi observada em dados brutos verificáveis:

- Todos os modelos reportaram "Score SCA: 27.41%" (não 27.4%, 27%, ou ~30%)
- Todos identificaram "82 das 91 críticas em host01" (não "a maioria", "~80", ou "maioria esmagadora")
- Todos usaram CVEs reais verificáveis (não CVEs inventados)

Esta ausência de variabilidade factual confirma eficácia dos mecanismos de controle de alucinação da arquitetura.

Outro resultado percebido foi que a arquitetura demonstrou capacidade de adaptação contextual automática via *templates* específicos por persona, resultando em variabilidade deliberada e desejada. A Tabela 15 contribui para se observar o desempenho da solução a partir de evidências de adaptação de variabilidade.

Tabela 15. Métricas de Adaptação

CRITÉRIO	CEO	COMPLIANCE	TIC	VARIAÇÃO CONTROLADA
Vocabulário técnico	Baixo (executivo)	Médio (regulatório)	Alto (técnico)	✓ Deliberada
Menção a frameworks	Baixa (ROI, impacto)	Alta (LGPD, ISO, CIS)	Média (CVE, CVSS)	✓ Deliberada
Comandos operacionais	Ausente	Ausente	Presente	✓ Deliberada
Quantificação financeira	Alta (investimentos)	Média (multas)	Baixa (*downtime*)	✓ Deliberada
Dados brutos (métricas)	Idênticos	Idênticos	Idênticos	✓ Consistência

Fonte: Próprio Autor

Entre os exemplos de variabilidade controlada destacam-se que para a Mesma vulnerabilidade observou-se a três abordagens a seguir:

- **CEO:** "82 das 91 vulnerabilidades críticas concentradas em um único host representam risco de paralisia operacional custosa"
- **Compliance:** "82 vulnerabilidades críticas em host01 configuram não conformidade com ISO 27001 A.18 (Gestão de Vulnerabilidades) e potencial incidente reportável LGPD Art. 48"
- **TIC:** "82 CVEs críticos (CVSS≥9.0) em host01 (127.0.0.1) exigem remediação imediata - priorizar CVE-2024-56180 (RCE, exploit público disponível)"

Esta variabilidade controlada demonstra que a arquitetura não apenas "traduz" dados técnicos para linguagem executiva, mas **reconstrói a narrativa** por perspectiva da audiência enquanto mantém precisão factual.

5.5.6 Escalabilidade e Manutenibilidade

A arquitetura modular implementada oferece vantagens significativas para deployment corporativo em escala, conforme demonstrado pela estrutura organizacional do sistema por meio da Figura 12.

Figura 12. Estrutura e organização da solução

```
wazuh-compliance-report/
├── modules/
│   ├── data_collector.py          # Coleta via API Wazuh
│   ├── data_consolidator.py       # Agregação e normalização
│   ├── data_processor.py          # Processamento determinístico
│   ├── inventory_processor.py     # Processamento de inventário
│   ├── vulnerability_processor.py # Análise de vulnerabilidades
│   ├── wazuh_collector.py         # Interface Wazuh API
│   ├── report_generator.py        # Orquestração de geração
│   ├── run_full_pipeline.py       # Pipeline completo
│   └── report_utils/
│       ├── agents.py              # Agentes por persona
│       ├── ai_client.py           # Cliente IA (OpenRouter)
│       ├── data_handler.py        # Manipulação de dados
│       └── post_render_validator.py # Validação pós-geração
├── templates/
│   ├── prompts/
│   │   ├── ceo/                  # Templates executivos
│   │   ├── compliance_officer/   # Templates regulatórios
│   │   └── tic/                 # Templates técnicos
│   ├── frameworks/              # Frameworks de conformidade
│   └── report/                  # Templates de relatório
├── knowledge_base/              # Base de conhecimento
├── generated_reports/           # Saídas consolidadas
└── scripts/                    # Scripts auxiliares
```

Fonte: Elaboração Própria

A arquitetura proposta foi desenvolvida como um protótipo de testes, porém fez uso de boas práticas com vistas a futuras adaptações para uso em solução de produção, dentre as características implementadas encontram-se as detalhadas a seguir.

Extensibilidade de Modelos de IA

- Configuração Dinâmica: Novos modelos requerem apenas configuração via variáveis de ambiente (OPENROUTER_API_KEY, *model slugs*)
- Abstração via OpenRouter: Interface unificada para qualquer modelo de IA com API (GPT-like) como DeepSeek, Grok, Gemini, GLM-4.5, GPT-OSS.
- Zero Refatoração: Adição de modelos sem alteração de código base

Como característica principal do projeto, a Separação Arquitetural permite que cada persona possua templates independentes em templates/prompts/{persona}/, isso permite a criação de novas personas apenas acrescentando estruturas de templates. A partir da estrutura de agentes especializados a lógica isolada em report_utils/agents.py permite, também, adição incremental. Por fim a capacidade de implementação de validação específica por meio do post_render_validator.py permite a criação de regras por persona sem conflito para correções baseada em análise posterior e até mesmo a implementação de uma camada de verificação baseada em outros agentes específicos de análise que possam ser implementados no futuro.

Além disso, o desacoplamento implementado por meio da visão de base de conhecimento, implementada na pasta knowledge_base/ permite que frameworks e outras informações de contexto possam ser inseridas no processamento dos agentes como seu contexto de atuação.

O projeto também implementa uma visão de Logging estruturado que permite debugging granular e auditoria completa.

Gestão de Estado Determinística

O processamento padronizado e parametrizado permite re-execuções que produzem resultados consistentes por meio dos dados intermediários persistidos (consolidated_data) que permitem debug, análise comparativa e post-run. Por fim o módulo de validação pós-render permite a personalização da verificação automatizada de qualidade e integridade dos relatórios.

***Deployment* Flexível**

Por meio da configuração externalizada (.env) permite o *deployment* em múltiplos ambientes distintos com a segregação do fornecimento de credenciais também para fins de segurança. A solução está pronta para conteinerização pois apresenta arquitetura modular compatível com Docker/Kubernetes e integrações CI/CD onde os scripts/ permitem automação via GitHub *Actions* ou Jenkins.

Benefícios de arquitetura proposta:

1. **Adição de novos modelos:** Requer apenas configuração de cliente de IA (via variáveis de ambiente), sem alteração de código

2. **Adição de novas personas:** Criação de novo agente + template de prompt, sem impactar personas existentes

3. **Atualização de frameworks:** Modificação em knowledge_base/ sem alteração de lógica de processamento

4. **Integração contínua:** Arquitetura permite automação via CI/CD para geração periódica de relatórios

Manutenibilidade:

- Separação clara entre dados (data_processor), lógica de negócio (agents) e apresentação (*templates*)
- Logs estruturados permitindo *debugging* e auditoria
- Arquivos de saída com *timestamp* evitando sobrescrita acidental
- Configuração externalizada (.env) facilitando *deployment* em múltiplos ambientes

Custos de Implementação e Operação

Como características presentes na arquitetura a seguir são explicadas aquelas relativas a otimização de custos por meio de:

Processamento local de dados:

- *Parsing* e agregação de *scans* Wazuh executados localmente podendo ser reaproveitados ou gerados novamente de acordo com parâmetros de execução permitindo que todos os relatórios têm a mesma visão temporal dos eventos e reduzindo a carga de consultas nas apis e base indexada do ambiente Wazuh.
- Consolidação de dados em *consolidated_data* reduz volume de dados enviados a APIs de IA é uma estratégia para contornar a limitação de tokens de contexto bem como reduzir o custo por relatórios já que consome menos tokens de entrada.
- O uso de dados estruturados (JSON) são mais compactos que texto narrativo.

- O modelo de agentes parametrizáveis e o uso do Openrouter permite o uso flexível de modelos gratuitos para fins de testes ou para contextos específicos onde o custo é imperativo.

Além do exposto anteriormente, a arquitetura foi otimizada para minimização de custos operacionais, conforme demonstrado pelos dados reais de *billing* do projeto:

Processamento Local Maximizado:

- *Parsing* Wazuh Local: Extração e agregação de dados localmente reduzindo o envio ou uso de api de AI para o entendimento (reduz o custo e implementa a estratégia de redução de alucinação).
- Consolidação Inteligente: data_consolidator.py reduz volume de dados enviados para IA em ~85%.
- Cache de Dados: Reutilização de *consolidated_data* evita reprocessamento desnecessário.

Otimização de Tokens Empírica:

- Estruturação JSON: Dados estruturados 40% mais compactos que narrativas textuais
- Templates Reutilizáveis: Prompts base evitam re-envio de instruções (economia de ~30% tokens)

Como forma de avaliação de custos para uso das APIs de IA a Tabela a seguir traz as informações reais do projeto, especificamente para geração dos 18 relatórios utilizado na avaliação aqui apresentada. Pode-se observar a quantidade de tokens consumidos e o custo total por modelo.

Tabela 16. Bilhetagem do uso de API-IA

MODELO	CUSTO REAL (USD)	TOKENS PROCESSADOS
GLM-4.5-Air	$0.000	107.772
X-AI Grok 4.1	$0.000	95.713
OpenAI GPT-OSS	$0.000	149.608
Gemini 2.5 Flash	$0.014	94.287
DeepSeek Chat v3	$0.033	82.024
Gemini 3 Pro	$0.837	144.432

Fonte: Próprio Autor

Pode-se observar que o custo total para a geração dos 18 relatórios do ambiente de testes ficou em US$ 0,88 que representa R$ 4,72 considerando a relação do dólar do dia em R$ 5,34.

DISCUSSÃO: IMPLICAÇÕES PARA ADOÇÃO CORPORATIVA

Os resultados indicam que não existe "modelo único ideal", mas sim modelos otimizados por contexto como descrito a seguir:

Para Aprovação Executiva de Investimentos (CEO):

- **1ª Opção:** Gemini 2.5 Flash Lite (4.94) - estimativas financeiras + identificação clara de riscos de concentração de falhas em ponto único
- **2ª Opção:** Grok 4.1 Fast (4.94) - quantificação agressiva de ROI para perfis executivos mais orientados a risco calculado

Para Preparação de Auditorias e Certificações (Compliance):

- **1ª Opção:** DeepSeek Chat v3 (5.00) - seção "Evidências para Auditoria"
- **2ª Opção:** Gemini 2.5 Flash Lite (5.00) - Matriz de Controles para gestão multi-framework

Para Remediação Técnica Operacional (TIC):

- **1ª Opção:** Gemini 2.5 Flash Lite (5.00) - comandos específicos prontos para execução
- **2ª Opção:** Grok 4.1 Fast (4.94) - runbooks de resposta a incidentes + IOCs para SIEM

Para Uso em Escala com Orçamento Limitado:

- **Opção viável:** GLM 4.5 Air Free (4.17) - modelo gratuito com performance intermediária estável
- **Evitar:** GPT-OSS 20B Free (3.46) - limitações severas em completude tornam-no inadequado para uso corporativo

Gap de Qualidade entre Modelos Comerciais e Open-Source Gratuitos

A diferença de 1.52 pontos (43.9%) entre o líder (Gemini 2.5: 4.98) e o modelo gratuito de pior desempenho (GPT-OSS: 3.46) representa gap substancial que impacta diretamente a utilidade corporativa.

Em uma análise de custo-benefício o GLM 4.5 Air Free (4.17), também gratuito, demonstra que é possível obter performance intermediária aceitável sem investimento, posicionando-se 0.71 pontos acima do GPT-OSS. Isto sugere que limitações do GPT-OSS não são inerentes a modelos gratuitos, mas sim específicas da arquitetura/treinamento deste modelo.

Necessidade de Validação Metodológica de Inferências

Para os modelos que incluíram estimativas financeiras (Gemini 2.5: investimentos verificáveis; Grok 4.1: riscos inferidos), e que acabaram demonstrando maior impacto persuasivo (B5=5.0), geram alertas para o resultado, pois, introduziram necessidade de validação metodológica que devem ser implementadas em melhorias futuras. Com isso é importante que em versões futuras sejam incluídos dados adicionais ao modelo determinístico que ajudem a:

1. Documentar premissas de cálculo de estimativas financeiras
2. Validar IOCs e ExploitDB IDs com fontes secundárias independentes
3. Explicitar margem de erro em projeções de downtime e multas
4. Manter rastreabilidade entre dados brutos e inferências derivadas

Eficiência Arquitetural como Multiplicador de Valor

Os resultados demonstram que a arquitetura de geração automatizada desenvolvida neste trabalho não é apenas "ferramenta de apoio", mas influencia no resultado quando:

1. **Elimina variabilidade indesejada** (alucinações: 0%) mantendo variabilidade desejada (adaptação contextual)
2. **Amplifica capacidades dos modelos** via prompt engineering estruturado e base de conhecimento
3. **Garante rastreabilidade e auditabilidade** essenciais para contextos corporativos

Observou-se, ao final, que a qualidade dos relatórios gerados não depende apenas da escolha do modelo de IA, mas da **combinação modelo + arquitetura**. Mesmo modelos de performance intermediária (GLM 4.5: 4.17) podem atingir utilidade corporativa aceitável quando integrados a uma arquitetura robusta de processamento de dados e *prompt engineering*.

LIMITAÇÕES DO ESTUDO

Limitações Metodológicas

1. **Escopo de Avaliação:** A pesquisa contemplou 6 modelos × 3 personas = 18 relatórios. A escolha dos modelos utilizado se deu pela análise dos 3 modelos gratuitos e os 3 pagos mais utilizado na data de execução dos testes.

2. **Subjetividade na Avaliação Qualitativa:** Apesar do uso de framework estruturado com escala Likert, critérios qualitativos (clareza, persuasão, criatividade) mantêm componente subjetivo inerente.

3. **Contexto Organizacional Único:** Os relatórios foram gerados sobre dados de uma única organização (Score SCA 27.41%, 91 vulnerabilidades críticas). Generalização para contextos com diferentes perfis de risco requer validação adicional.

Limitações de Generalização

1. **Língua Portuguesa:** Todos os relatórios foram gerados em português brasileiro. Performance em outros idiomas podem variar, especialmente para modelos com treinamento predominante em inglês.

2. **Domínio de Cibersegurança:** Resultados são específicos para relatórios de segurança baseados em scan Wazuh SCA. Aplicabilidade a outros tipos de relatórios (financeiros, operacionais, científicos) requer investigação independente.

3. **Temporalidade:** Modelos de IA evoluem rapidamente. Resultados refletem versões específicas (Gemini 2.5 Flash Lite, Grok 4.1 Fast, DeepSeek Chat v3, etc.) avaliadas em outubro-novembro de 2025. Atualizações posteriores podem alterar performance relativa mas podem ser facilmente reavaliadas pela metodologia e arquitetura proposta.

SÍNTESE DOS RESULTADOS PRINCIPAIS

A pesquisa comprova a hipótese central da contribuição para geração eficiente de relatório de compliance para diferentes públicos e de forma automatizada por meio do uso de IA generativa e da arquitetura implementada. Como achados principais tem-se:

1. **Liderança de Performance:** Google Gemini 2.5 Flash Lite (4.98) estabelece novo benchmark de qualidade, com pontuação perfeita (5.00) em Compliance Officer e TIC, diferenciando-se por operacionalização máxima (comandos específicos, matriz de controles, estimativas verificáveis).

2. **Consistência Cross-Persona:** X.AI Grok 4.1 Fast demonstra consistência perfeita (4.94 em todas as personas), evidenciando capacidade superior de adaptação contextual, porém com trade-off em verificabilidade factual (C1: 4.0 vs. 5.0 dos demais).

3. **Especialização Estratégica:** DeepSeek Chat v3 alcança excelência em Compliance (5.00) via seção "Evidências para Auditoria" inovadora, mas apresenta sub-otimização em CEO (4.43), sugerindo especialização deliberada.

4. **Gap Qualitativo Significativo:** Diferença de 1.52 pontos (43.9%) entre líder e modelo de menor performance (GPT-OSS 20B: 3.46) confirma que escolha de modelo tem impacto direto e mensurável na utilidade corporativa de relatórios gerados.

5. **Operacionalização como Multiplicador:** Presença de ferramentas específicas (comandos, matrizes, runbooks) eleva B3 (Acionabilidade) em 67% (de 3.0 para 5.0), confirmando ser diferencial competitivo crítico, não opcional.

6. **Precisão Factual Necessária, mas Insuficiente:** Ausência de alucinações (C1 ≥4.0 em todos os modelos) é pré-requisito mínimo, mas não garante qualidade. GPT-OSS mantém C1=5.0 porém pontuação geral de apenas 3.46 por deficiências em completude e profundidade.

7. **Eficiência Arquitetural como Fundação:** A arquitetura de geração automatizada desenvolvida alcançou taxa de alucinação crítica de 0%, redução de custo operacional de 99.97% e variabilidade controlada entre personas, demonstrando ser componente essencial - não opcional - para uso corporativo de IA em cibersegurança.

Estes resultados fornecem base empírica para decisões de adoção de IA em cibersegurança corporativa, demonstrando que modelos de linguagem de grande escala alcançaram maturidade suficiente para uso em contextos de alta criticidade, desde que: (1) selecionados adequadamente por contexto de uso, (2) integrados a arquitetura robusta de processamento de dados e *prompt engineering*, e (3) submetidos a validação metodológica.

6 CONCLUSÃO

Este trabalho investigou a aplicabilidade da Inteligência Artificial Generativa (IAG) como solução tecnológica para automatizar e personalizar a geração de relatórios de compliance em segurança da informação e proteção de dados, com foco nas necessidades específicas de empresas de médio porte. O problema central abordado foi a sobrecarga enfrentada por profissionais de compliance e segurança da informação na elaboração manual de relatórios complexos destinados a diferentes *stakeholders* organizacionais, agravado pela escassez de profissionais especializados na área.

A hipótese central de que a IAG poderia otimizar significativamente a eficiência, escalabilidade e confiabilidade dos processos de compliance foi validada por meio de uma metodologia que combinou revisão sistemática da literatura, desenvolvimento de solução tecnológica baseada em arquitetura modular (integrando Wazuh, Agentes de IA por meio do uso de API's de modelos de linguagem de larga escala), e avaliação empírica comparativa de diferentes LLMs disponíveis como serviço.

Síntese dos Principais Resultados

Os resultados obtidos demonstram que a automação inteligente da geração de relatórios de compliance apresenta benefícios substanciais em múltiplas dimensões. A solução desenvolvida conseguiu gerar, de forma consistente e personalizada, relatórios específicos para três perfis de *stakeholders* distintos (CEO/Alta Gestão, Compliance Officer e equipes de TIC), adaptando linguagem, profundidade técnica e estrutura de recomendações de acordo com as necessidades informacionais de cada público.

A análise comparativa de desempenho entre diferentes modelos de linguagem revelou que o Google Gemini-2.5-flash-lite apresentou a melhor relação custo-benefício para o contexto específico de relatórios de compliance, mantendo alta qualidade linguística (média de 4,8/5) e forte alinhamento ao propósito (média de 4,7/5), ao mesmo tempo em que oferece custos operacionais significativamente inferiores aos principais concorrentes no mercado. Esta constatação possui implicações práticas relevantes para empresas de médio porte, que precisam equilibrar qualidade técnica com viabilidade econômica.

Do ponto de vista operacional, os resultados indicam que a automação da criação de relatórios contribui para a liberação de tempo dos especialistas em segurança da informação para concentrarem-se em atividades de maior valor estratégico como análise de riscos, desenho de controles e resposta a incidentes, reduzindo substancialmente o tempo dedicado à documentação de conformidade. Além disso, a solução demonstrou capacidade de melhorar a precisão e consistência das informações fornecidas, minimizando riscos de inconsistências terminológicas ou omissões que frequentemente ocorrem em processos manuais.

Um resultado particularmente relevante foi a constatação de que a personalização dos relatórios para diferentes audiências aumenta significativamente a compreensibilidade e utilidade das informações de compliance. Relatórios técnicos detalhados para equipes de TIC (incluindo especificações de patches, playbooks de hardening e scripts de automação) coexistem com sínteses executivas para CEOs (focadas em impactos de negócio, priorização de riscos e alinhamento estratégico) e documentos estruturados para gestores de conformidade (evidenciando gaps críticos, controles compensatórios e *roadmaps* para certificações), todos gerados a partir da mesma base de dados consolidada.

Contribuições Teóricas e Práticas

No plano teórico, este trabalho contribui para a literatura emergente sobre aplicações práticas de IA Generativa em domínios especializados de segurança da informação e governança corporativa. A revisão sistemática da literatura realizada permitiu mapear o estado da arte, identificar lacunas de pesquisa e formular um framework conceitual para integração de IAG em processos de compliance, campo ainda incipiente na literatura acadêmica brasileira.

Também foi necessária a criação de uma metodologia de avaliação comparativa para os relatórios gerados pela solução proposta, baseada em múltiplas dimensões (qualidade linguística, alinhamento ao propósito, eficiência do modelo de IA), pode ser replicada e adaptada para outros contextos de aplicação de IA Generativa em ambientes corporativos. Este aspecto metodológico representa uma contribuição para pesquisas futuras que busquem avaliar empiricamente soluções de IAG em cenários organizacionais reais.

No plano prático, a solução implementada oferece um modelo arquitetural concreto e replicável para empresas de médio porte que buscam modernizar seus processos de compliance. A modularidade da arquitetura proposta, separando coleta de dados (Wazuh), orquestração de fluxos, processamento e consolidação de informações, e geração de conteúdo via LLMs, facilita a adaptação da solução para diferentes contextos organizacionais e tecnológicos.

Os templates de prompts estruturados desenvolvidos para cada perfil de *stakeholder* (CEO, *Compliance Officer*, TIC) representam ativos reutilizáveis que incorporam boas práticas de engenharia de *prompts* para domínios especializados. Estes *templates* podem ser customizados por outras organizações, acelerando a curva de implementação de soluções similares.

Desafios Identificados e Limitações

Apesar dos resultados promissores, a pesquisa identificou desafios significativos que precisam ser considerados em implementações de IAG para compliance regulatório. O principal deles é a garantia de confiabilidade e precisão pois modelos de linguagem generativos, embora altamente capazes, permanecem suscetíveis a fenômenos como "alucinações. No contexto de *compliance*, onde a precisão é crítica para decisões regulatórias e de auditoria, este risco exige mecanismos rigorosos de validação. A solução proposta busca mitigar este risco ao basear a geração de conteúdo exclusivamente em dados estruturados provenientes do Wazuh, mas reconhece-se que a validação por especialistas permanece necessária, especialmente para recomendações de remediação complexas.

Outro ponto que precisa ser considerado é a privacidade e Proteção de Dados pois a utilização de LLMs fornecidos como serviço (*cloud-based APIs*) implica na transmissão de dados de vulnerabilidades e configurações de infraestrutura para provedores externos. Embora os principais fornecedores de serviços de IA ofereçam garantias contratuais sobre proteção e não utilização de dados para treinamento de modelos, este aspecto representa uma consideração crítica para organizações em setores altamente regulados (financeiro, saúde, governo). Implementações futuras podem considerar a adoção de modelos de linguagem hospedados localmente (*on-premises*) para mitigar estes riscos, embora isso implique em maior complexidade operacional e custos de infraestrutura.

Outro aspecto diz respeito a adaptação contínua a requisitos regulatórios pois os *frameworks* regulatórios de segurança da informação e proteção de dados evoluem constantemente (LGPD, GDPR, ISO 27001, frameworks setoriais). A solução proposta requer mecanismos de atualização periódica dos *templates de prompts* e da base de conhecimento para manter-se alinhada às versões vigentes de normas e boas práticas. Este aspecto demanda processos estruturados de manutenção e governança da solução.

Trabalhos Futuros e Recomendações

Com base nos resultados deste trabalho, pode-se considerar promissores trabalhos complementares futuros:

Avaliação de Efetividade em Escala: Conduzir estudos longitudinais em múltiplas organizações para avaliar o impacto quantitativo da solução em métricas como: tempo médio de geração de relatórios, taxa de retrabalho, satisfação de stakeholders, e conformidade em auditorias externas.

Integração com Sistemas de GRC: Expandir a solução para integração nativa com plataformas de Governança, Risco e Compliance (GRC) amplamente utilizadas no mercado, permitindo fluxos bidirecionais de informação e automação *end-to-end* de processos de *compliance*.

Exploração de Modelos de Linguagem Multimodais: Investigar a incorporação de análises de logs, diagramas de arquitetura e dashboards de monitoramento (dados não-estruturados e visuais) na geração de relatórios, explorando capacidades multimodais emergentes em LLMs de última geração.

Desenvolvimento de Mecanismos de Explicabilidade: Implementar técnicas de IA explicável (XAI) que permitam rastrear como o modelo de linguagem chegou a determinadas conclusões ou recomendações, aumentando a transparência e confiança dos usuários finais.

Benchmarking Setorial: Criar benchmarks de desempenho de soluções de IAG para compliance segmentados por setor (financeiro, saúde, varejo, indústria), considerando particularidades regulatórias e níveis de maturidade em segurança da informação.

Considerações Finais

Este trabalho demonstrou que a Inteligência Artificial Generativa representa uma tecnologia transformadora para processos de compliance em segurança da informação, oferecendo ganhos tangíveis em eficiência, escalabilidade e personalização de comunicações. A solução desenvolvida comprova a viabilidade técnica e econômica da abordagem proposta para empresas de médio porte, segmento que tradicionalmente fica à margem das soluções corporativas de compliance devido a restrições orçamentárias e de capacidade técnica.

Os resultados sugerem que, quando adequadamente implementada e governada, a IAG pode contribuir significativamente para a democratização do acesso a processos sofisticados de gestão de conformidade regulatória, reduzindo assimetrias competitivas entre organizações de diferentes portes. Esta democratização possui implicações relevantes, potencialmente contribuindo para melhorar o nível geral de proteção de dados e cibersegurança no ecossistema empresarial.

Apesar disso, a adoção responsável desta tecnologia exige consciência sobre seus limites e riscos. A supervisão humana qualificada permanece indispensável, e a IA Generativa deve ser posicionada como ferramenta de apoio sempre apoiada pela expertise de profissionais de segurança da informação e *compliance*.

Por fim, este trabalho espera ter contribuído tanto para o avanço do conhecimento acadêmico sobre aplicações práticas de IA Generativa em domínios especializados quanto para a disponibilização de referências concretas que possam orientar profissionais e organizações na modernização de seus processos de governança e conformidade regulatória.

REFERÊNCIAS BIBLIOGRÁFICAS

ADELAKUN, Beatrice Oyinkansola. THE IMPACT OF AI ON INTERNAL AUDITING: TRANSFORMING PRACTICES AND ENSURING COMPLIANCE. **Finance & Accounting Research Journal**, v. 4, n. 6, p. 350–370, dez. 2022.

ALKALBANI, Ahmed; DENG, Hepu; KAM, Booi. Investigating the Role of Socio-organizational Factors in the Information Security Compliance in Organizations. **Australasian Conference on Information Systems 2015 (arXiv:1605.01032)**, v. ACIS/2015/82, jan. 2016.

AL-SHABANDAR, Raghad *et al.* The Application of Artificial Intelligence in Financial Compliance Management. *In*: ACM, 17 out. 2019. Disponível em: <https://doi.org/10.1145/3358331.3358367>

ARORA, S. Revolutionizing Third Party Risk Management Using Generative AI and RAG. *In*: INTERNATIONAL CONFERENCE ON DATA COMPUTATION AND COMMUNICATION ICDCC. **2024 First International Conference on Data, Computation and Communication (ICDCC)**. 2024. Disponível em: <https://doi.org/10.1109/ICDCC62744.2024.10961910>

CALO, Ryan. Artificial intelligence policy: A primer and roadmap. **U. Chi. L. Rev.**, v. 85, p. 1, 2017.

CHANDOLA, Varun; BANERJEE, Arindam; KUMAR, Vipin. Anomaly detection: A survey. **ACM computing surveys (CSUR)**, v. 41, n. 3, p. 1–58, 2009.

DISTERER, Georg. ISO/IEC 27000, 27001 and 27002 for Information Security Management. **Scientific Research Publishing**, v. 4, n. 2, p. 92–100, jan. 2013.

GARCIA-SEGURA, Luis A. The Role of Artificial Intelligence in Preventing Corporate Crime. **Journal of Economic Criminology**, v. 5, p. 100091–100091, ago. 2024.

GARCIA-TEODORO, Pedro *et al.* Anomaly-based network intrusion detection: Techniques, systems and challenges. **computers & security**, v. 28, n. 1–2, p. 18–28, 2009.

GEORGE, A. Shaji; GEORGE, AS Hovan; MARTIN, AS Gabrio. ChatGPT and the Future of Work: A Comprehensive Analysis of AI's Impact on Jobs and Employment. **Partners Universal International Innovation Journal**, v. 1, n. 3, p. 154–186, 2023.

GOODFELLOW, Ian *et al.* Generative adversarial nets Advances in neural information processing systems. **arXiv preprint arXiv:1406.2661**, 2014.

HASHMI, Mustafa; GOVERNATORI, Guido; WYNN, Moe Thandar. Normative requirements for regulatory compliance: An abstract formal framework. **Information Systems Frontiers**, v. 18, n. 3, p. 429–455, maio 2015.

IKUDABO, Ayodeji Oyindamola; KUMAR, Pradeep. AI-Driven Risk Assessment and Management in Banking: Balancing Innovation and Security. **International Journal of Research Publication and Reviews**, v. 5, n. 10, p. 3573–3588, out. 2024.

ISMAIL *et al.* Enhancing Security Operations Center: Wazuh Security Event Response with Retrieval-Augmented-Generation-Driven Copilot. **Sensors**, v. 25, n. 3, p. 870, 2025.

JAIN, Varun *et al.* Leveraging Artificial Intelligence for Enhancing Regulatory Compliance in the Financial Sector. **International Journal of Computer Trends and Technology**, v. 72, n. 5, p. 124–140, maio 2024.

KARPATOU, Panagiota Angeliki. The evolution of cybersecurity threats & the rise of artificial intelligence. **Bachelor dissertation**, 2025.

KINGSTON, John. Using artificial intelligence to support compliance with the general data protection regulation. **Artificial Intelligence and Law**, v. 25, n. 4, p. 429–443, set. 2017.

KURNIAWAN, K. CyKG-RAG: Towards knowledge-graph enhanced retrieval augmented generation for cybersecurity. *In*: 2024. Disponível em: <https://eprints.cs.univie.ac.at/8178/1/RAGE-KG_2024_paper_1_Andreas%20Ekelhart.pdf>

MATAS, Sweden De; KEEGAN, Brendan James. Challenges in Addressing Information Security Compliance in Healthcare Research: The Human Factor. **Science Publishing Group**, v. 5, n. 2, p. 25–25, jan. 2020.

MCFEETORS, Jason; PANT, Tanay. **Rapid Product Development with n8n**. *[S.l.]*: Packt Publishing, 2022.

MEDEIROS, Marcio Lima; CODIGNOTO, Roberta. Governança, integridade e resultados caminham juntos. v. 3, n. 1, out. 2022.

OLIVEIRA, Bruno Vicente Nunes de; MELO, Filipe Torres de. IN^{TEL}IGÊNCIA ARTIFICIAL. **P2P E INOVAÇÃO**, v. 10, n. 1, p. 226–247, set. 2023.

OREYOMI, Michael; JAHANKHANI, Hamid. Challenges and Opportunities of Autonomous Cyber Defence (ACyD) Against Cyber Attacks. **Blockchain and Other Emerging Technologies for Digital Business Strategies**, p. 239–269, 2022.

PASCOE, Cherilyn E. Public Draft: The NIST Cybersecurity Framework 2.0. **National Institute of Standards and Technology**, 2023.

PERRINA, Filippo *et al.* AGIR: Automating Cyber Threat Intelligence Reporting with Natural Language Generation. *In*: CONFERENCE ON BIG DATA (BIGDATA). **2023 IEEE International Conference on Big Data (BigData)**. 2023. Disponível em: <https://doi.org/10.1109/BigData59044.2023.10386116>

PRASAD, K. D. V.; KALAVAKOLANU, Sripathi. THE STUDY OF COGNITIVE PSYCHOLOGY IN CONJUNCTION WITH ARTIFICIAL IN^{TEL}LIGENCE. **Conhecimento & Diversidade**, v. 15, n. 36, p. 270–270, fev. 2023.

PRISMA. **PRISMA Flow Diagram**. , 2020. Disponível em: <https://www.prisma-statement.org/prisma-2020-flow-diagram>

SACOMANO NETO, Mário; ESCRIVÃO FILHO, Edmundo. Estrutura organizacional e equipes de trabalho: estudo da mudança organizacional em quatro grandes empresas industriais. **Gestão & produção**, v. 7, p. 136–145, 2000.

SALMAN, Ahmed; CREESE, Sadie; GOLDSMITH, Michael. Position Paper: Leveraging Large Language Models for Cybersecurity Compliance. *In*: 2024.

SAQIB, Najmudin *et al.* Mapping of the Security Requirements of GDPR and NISD. **Institute for Computer Sciences, Social Informatics and Telecommunications Engineering**, p. 166283–166283, jul. 2018.

SOMMER, Robin; PAXSON, Vern. Outside the closed world: On using machine learning for network intrusion detection. *In*: 2010. Disponível em: <https://ieeexplore.ieee.org/document/5504793>

STANKOVIĆ, Stefan; GAJIN, Slavko; PETROVIĆ, Ranko. A review of Wazuh tool capabilities for detecting attacks based on log analysis. **No Nama Agent Integrity File Added Delete Modified**, v. 1, 2022.

ŠUŠKALO, Dario *et al.* Comparative analysis of ibm qradar and wazuh for security information and event management. **Ann. DAAAM Proc**, v. 34, p. 0096–0102, 2023.

VASWANI, Ashish *et al.* Attention is all you need. **Advances in neural information processing systems**, v. 30, 2017.

VIJAYAKUMAR, P.; PYINGKODI, M.; DEVI, S. Comparative Analysis of AI Chatbot for Assessing Gemini AI, DeepSeek AI, and Qwen AI via OpenRouter API Integration. *In*: 2025 6TH INTERNATIONAL CONFERENCE ON DATA INTELLIGENCE AND COGNITIVE INFORMATICS (ICDICI). **IEEE - International Conference on Data Intelligence and Cognitive Informatics (ICDICI)**. IEEE, 2025. Disponível em: <https://doi.org/10.1109/ICDICI66477.2025.11134904>

APÊNDICE 1 – BIBLIOTECAS UTILIZADAS

Para fins de melhor compreensão do projeto aqui detalhamos as bibliotecas e frameworks utilizados no desenvolvimento do sistema de geração automatizada de relatórios de conformidade. A seleção de cada dependência foi realizada considerando aspectos de maturidade, suporte comunitário, documentação disponível e adequação aos requisitos funcionais e não-funcionais do projeto. A apresentação segue uma organização por categorias funcionais, conforme o papel de cada biblioteca na arquitetura do sistema.

Consultas de API e Comunicação HTTP

Requests

A biblioteca Requests (versão 2.31.0 ou superior) constitui a camada fundamental para todas as interações HTTP do sistema, sendo empregada tanto na comunicação com a API REST do Wazuh SIEM/XDR quanto na integração com o gateway OpenRouter para acesso aos modelos de linguagem. Sua adoção justifica-se pela simplicidade de interface, robustez no tratamento de exceções de rede e suporte nativo a autenticação via tokens JWT, método utilizado pelo Wazuh para controle de acesso. No contexto deste projeto, a biblioteca é responsável por operações de coleta de inventário de agentes (WazuhInventoryCollector), extração de resultados de auditorias SCA (WazuhDataCollector) e consultas de vulnerabilidades detectadas. A escolha por Requests em detrimento de alternativas como urllib3 ou httpx fundamenta-se na maturidade da biblioteca (disponível desde 2011) e na vasta documentação disponível, aspectos relevantes para garantir a manutenibilidade do código em contexto acadêmico. Adicionalmente, a biblioteca oferece suporte transparente a conexões TLS/SSL, requisito obrigatório para comunicação com endpoints de segurança em ambientes de produção.

9.1.2 Python-dotenv

A biblioteca python-dotenv (versão 1.0.0 ou superior) implementa o padrão de gerenciamento de configuração conhecido como "Twelve-Factor App", permitindo a separação de credenciais e parâmetros de configuração do código-fonte por meio de arquivos .env. No sistema desenvolvido, a biblioteca é empregada para carregar variáveis de ambiente contendo a chave de API do OpenRouter (OPENROUTER_API_KEY), credenciais de acesso ao Wazuh (usuário e senha) e

URLs de endpoints. Esta abordagem atende a requisitos de segurança ao evitar o versionamento de credenciais no repositório Git, facilitando também a portabilidade do sistema entre ambientes de desenvolvimento, homologação e produção. O arquivo secure/.env é explicitamente excluído do controle de versão via .gitignore, garantindo conformidade com boas práticas de segurança da informação. A simplicidade de uso da biblioteca (carregamento via load_dotenv()) e sua compatibilidade com múltiplos sistemas operacionais justificam sua adoção em detrimento de soluções mais complexas como python-decouple ou gerenciadores de segredos corporativos.

9.2 Agentes de IA e Comunicação com LLMs

9.2.1 Integração OpenRouter (via Requests)

A integração com modelos de linguagem de grande escala (LLMs) é realizada via API do OpenRouter, plataforma que atua como gateway unificado para mais de 30 provedores de IA (OpenAI, Anthropic, Google, DeepSeek, X.AI, entre outros). A comunicação utiliza a biblioteca Requests para envio de requisições HTTP POST ao endpoint https://openrouter.ai/api/v1/chat/completions, seguindo o formato de API compatível com OpenAI. Esta arquitetura de integração foi escolhida por três razões principais: (1) flexibilidade experimental, permitindo a troca de modelos via parâmetro model sem refatoração de código; (2) rastreamento de custos, com a API retornando métricas detalhadas de consumo de tokens e custos associados; e (3) abstração de complexidade, eliminando a necessidade de implementar integrações específicas para cada provedor. No contexto da classe BaseAgent (módulo report_utils/agents.py), o método _call_llm() encapsula a lógica de construção de payloads JSON, tratamento de timeouts (60 segundos padrão) e retry automático em caso de falhas transitórias. A ausência de uma biblioteca cliente dedicada (como openai-python ou anthropic-sdk) justifica-se pela necessidade de controle fino sobre parâmetros de requisição e pela portabilidade da solução, que não depende de SDKs proprietários.

9.2.2 Arquitetura de Agentes Especializados

A arquitetura de agentes de IA implementada neste projeto segue o padrão de especialização por domínio, onde cada agente (EnvironmentCharacterizationAgent, VulnerabilityAnalysisAgent, SCAAnalysisAgent, entre outros) herda da classe abstrata Agent e implementa lógica específica para

geração de seções do relatório. Esta abordagem, inspirada em arquiteturas de sistemas multi-agentes (MAS - Multi-Agent Systems), permite a modularização do processo de geração de conteúdo e facilita a manutenção e extensão do sistema. A comunicação com LLMs é mediada pela camada de abstração AIClient, que encapsula detalhes de autenticação, serialização JSON e tratamento de erros. A escolha por uma arquitetura de agentes especializados, em detrimento de um único agente genérico, fundamenta-se em evidências empíricas da literatura de IA generativa que demonstram ganhos de qualidade quando modelos são instruídos com prompts especializados e contextos reduzidos. Cada agente recebe um subconjunto relevante de consolidated_data (método _make_prompt_friendly()), implementando compactação de contexto para economia de tokens e redução de latência.

9.3 Processamento de Templates

9.3.1 Jinja2

O Jinja2 (versão 3.1.0 ou superior) é empregado como motor de templates para duas finalidades distintas no sistema: (1) renderização de prompts enviados aos agentes de IA, e (2) composição do relatório HTML final. Desenvolvido pela equipe do framework Flask, Jinja2 é amplamente reconhecido na comunidade Python por sua sintaxe expressiva, sistema de herança de templates e mecanismos de escape automático de HTML, característica essencial para prevenção de injeção de código. No contexto de geração de prompts, templates Jinja2 localizados em templates/agents/<persona>/ definem a estrutura de instruções enviadas aos LLMs, com variáveis interpoladas dinamicamente a partir de consolidated_data. Esta abordagem permite a criação de três variantes de prompts (personas CEO, TIC e Compliance Officer) sem duplicação de lógica, implementando o princípio DRY (Don't Repeat Yourself). Na renderização do relatório final (templates/report/base_report.html.j2), Jinja2 é responsável por compor seções HTML a partir de fragmentos gerados pelos agentes, intercalando blocos determinísticos (tabelas de hosts, métricas agregadas) com conteúdo textual produzido por IA. A escolha por Jinja2 em detrimento de alternativas como Mako ou Chameleon justifica-se pela integração nativa com o ecossistema Flask, pela documentação extensa e pela prevalência em projetos acadêmicos de análise de dados.

9.3.2 Separação de Lógica e Apresentação

A arquitetura de templates implementada segue rigorosamente o padrão MVC (Model-View-Controller), onde templates Jinja2 representam a camada de apresentação (View), enquanto a lógica de negócio reside em classes Python (Controller) e dados consolidados constituem o modelo (Model). Esta separação é particularmente relevante no contexto de geração de relatórios assistida por IA, onde a validação de saída (módulo post_render_validator.py) atua como camada adicional para garantir que apenas afirmações suportadas por evidências sejam incluídas no documento final. Templates definem apenas a estrutura de apresentação, delegando decisões sobre conteúdo aos agentes de IA e ao orquestrador (AIReportGenerator). Adicionalmente, o uso de macros Jinja2 (ex: {% macro render_cve_table(cves) %}) permite a reutilização de componentes de interface, reduzindo o acoplamento entre módulos e facilitando a manutenção evolutiva do sistema.

9.4 Análise de Dados e Visualização

9.4.1 Pandas

A biblioteca Pandas (versão 2.0.0 ou superior) é empregada exclusivamente nos scripts de análise de custos (generate_cost_analysis.py e parse_openrouter_billing.py), não sendo utilizada no fluxo principal de geração de relatórios. Sua adoção justifica-se pela necessidade de processar logs de billing retornados pela API do OpenRouter, que contêm informações sobre tokens consumidos, custo por requisição e latência de resposta. O script parse_openrouter_billing.py utiliza Pandas para carregar JSONs de billing, realizar agregações por modelo e persona (operações groupby() e agg()), e persistir resultados em formato CSV para análise posterior. Esta abordagem de processamento separado da análise de custos (desacoplada do pipeline de geração de relatórios) permite a execução de experimentos de custo-benefício sem impactar a performance do sistema principal. A escolha por Pandas em detrimento de processamento manual com dicionários Python justifica-se pela eficiência em operações de agregação e pela facilidade de exportação para formatos tabulares (CSV), amplamente utilizados em análises acadêmicas.

9.4.2 Matplotlib e Seaborn

As bibliotecas Matplotlib (versão 3.7.0 ou superior) e Seaborn (versão 0.12.0 ou superior) são empregadas para geração de visualizações gráficas incluídas no capítulo de análise de custos do

relatório final. Matplotlib fornece a camada de baixo nível para renderização de figuras (gráficos de barras, histogramas, scatter plots), enquanto Seaborn oferece interface de alto nível com estilos pré-definidos e suporte a visualizações estatísticas complexas (heatmaps, violin plots). No script generate_cost_analysis.py, estas bibliotecas são utilizadas para produzir quatro gráficos principais: (1) custo acumulado por modelo, (2) distribuição de custo por chamada, (3) heatmap de combinações persona × modelo, e (4) evolução temporal de custos. A escolha conjunta de Matplotlib e Seaborn segue padrão estabelecido na comunidade de ciência de dados Python, onde Seaborn atua como camada de abstração sobre Matplotlib, simplificando a criação de visualizações complexas sem sacrificar controle fino sobre elementos gráficos. Todos os gráficos gerados são persistidos em docs/projeto/images/ no formato PNG (300 DPI) para garantir qualidade de impressão em dissertações acadêmicas.

9.5 Geração de Relatórios em PDF

9.5.1 wkhtmltopdf

A conversão de relatórios HTML para formato PDF é realizada via wkhtmltopdf, ferramenta de linha de comando baseada no motor de renderização WebKit. Esta escolha arquitetural (invés de bibliotecas Python puras como ReportLab ou WeasyPrint) justifica-se por três aspectos: (1) fidelidade de renderização, com suporte completo a CSS3, web fonts e JavaScript; (2) independência de dependências Python, facilitando a distribuição do sistema; e (3) performance, com conversão de relatórios de 50+ páginas em menos de 10 segundos. A integração com wkhtmltopdf é implementada no método _render_to_pdf() da classe AIReportGenerator, utilizando a biblioteca padrão subprocess para execução do binário externo. O sistema verifica a disponibilidade de wkhtmltopdf via shutil.which() e, caso ausente, mantém o relatório em formato HTML sem interromper o fluxo de execução. Parâmetros de conversão incluem margens (20mm), DPI (300), habilitação de imagens externas e geração de índice de conteúdo (table of contents). A escolha por wkhtmltopdf em detrimento de WeasyPrint (biblioteca Python pura) fundamenta-se em testes empíricos que demonstraram melhor compatibilidade com templates Bootstrap e renderização mais fidedigna de elementos complexos (tabelas aninhadas, gráficos inline).

9.5.2 Alternativas Avaliadas

Durante o desenvolvimento do sistema, foram avaliadas três alternativas para geração de PDF: (1) ReportLab, biblioteca de baixo nível para criação programática de PDFs, descartada pela complexidade de reproduzir layouts HTML complexos; (2) WeasyPrint, motor de renderização Python puro, descartado por limitações na interpretação de CSS3 e dependências de bibliotecas de sistema (Cairo, Pango); e (3) Playwright PDF, funcionalidade de conversão do framework Playwright, descartada por introduzir dependência pesada (Chromium headless, 280 MB) para funcionalidade secundária. A solução adotada (wkhtmltopdf como dependência opcional) representa equilíbrio entre qualidade de saída e simplicidade de deployment, permitindo que o sistema opere em modo "HTML-only" caso wkhtmltopdf não esteja disponível, sem comprometer funcionalidades principais.

9.6 Automação de Diagramas

9.6.1 Playwright

A biblioteca Playwright (versão 1.56.0 ou superior) é empregada exclusivamente no script de exportação de diagramas Mermaid para PNG (export_mermaid_to_png.py), não sendo dependência do fluxo de geração de relatórios. Playwright permite a automação de navegadores Chromium, Firefox ou WebKit via API Python assíncrona, sendo utilizado neste projeto para renderizar diagramas Mermaid em resolução de 9600×N pixels (onde N é calculado dinamicamente com base na altura do diagrama). O script cria uma página HTML contendo o código Mermaid e a biblioteca JavaScript correspondente, aguarda a renderização completa (timeout de 2000ms), captura o SVG gerado via page.screenshot() e persiste o resultado em formato PNG. Esta abordagem garante que todos os diagramas incluídos na documentação técnica possuam qualidade adequada para impressão em dissertações acadêmicas. A escolha por Playwright em detrimento de Puppeteer (biblioteca Node.js) ou Selenium justifica-se pela API assíncrona nativa, suporte a múltiplos navegadores e documentação oficial para Python. O script implementa ajuste automático de viewport para acomodar diagramas com altura superior a 4000px, funcionalidade essencial para captura sem cortes de diagramas complexos (ex: fluxograma de geração de relatórios com 21.126px de altura).

9.7 Outras Dependências Relevantes

9.7.1 JSON (Biblioteca Padrão)

A serialização e deserialização de dados estruturados é realizada via módulo json da biblioteca padrão Python, sem dependências externas. Este módulo é empregado em três contextos principais: (1) persistência de dados intermediários (raw_data/*.json, processed_*.json), (2) comunicação com APIs REST (parsing de respostas Wazuh e OpenRouter), e (3) armazenamento de evidências para reprodutibilidade (evidence/data_context_for_*.json). A escolha pela biblioteca padrão em detrimento de alternativas como ujson (ultra-rápida) ou orjson (otimizada para performance) justifica-se pela ausência de requisitos de performance críticos (processamento de datasets com menos de 10.000 registros) e pela portabilidade da solução, que não introduz dependências de compilação. O módulo json oferece ainda suporte nativo a tipos datetime via parâmetro default, funcionalidade utilizada nos logs de requisições para LLMs.

9.7.2 Pathlib (Biblioteca Padrão)

A manipulação de caminhos de arquivos e diretórios é realizada via módulo pathlib da biblioteca padrão (Python 3.4+), oferecendo interface orientada a objetos para operações de sistema de arquivos. O uso de Path em substituição a concatenações manuais de strings (os.path.join()) melhora a legibilidade do código e garante portabilidade entre sistemas operacionais (Windows utiliza \ como separador, Unix utiliza /). No sistema desenvolvido, pathlib é empregado para construção de caminhos relativos (base_dir / "raw_data" / "processed_inventory.json"), verificação de existência de arquivos (path.exists()), criação recursiva de diretórios (path.mkdir(parents=True)) e iteração sobre arquivos em diretórios (path.glob("*.json")). A prevalência de pathlib em projetos Python modernos e sua inclusão na biblioteca padrão justificam sua adoção como padrão de codificação do projeto.

9.7.3 Re (Expressões Regulares)

O módulo re (regular expressions) da biblioteca padrão Python é empregado em diversos pontos do sistema para processamento de texto e validação de padrões. Os principais casos de uso

incluem: (1) extração de CVE IDs de strings (re.findall(r'CVE-\d{4}-\d{4,}', text)), (2) sanitização de nomes de arquivos (remoção de caracteres especiais), (3) parsing de títulos de seções em documentos Markdown (script consolidate_documentation.py), e (4) validação de URLs em referências a NVD (National Vulnerability Database). A classe post_render_validator.py utiliza regex para detectar e remover do HTML final quaisquer referências a CVEs ou hosts não presentes em consolidated_data, implementando camada adicional de prevenção de alucinações. A escolha por expressões regulares em detrimento de parsers dedicados (como BeautifulSoup para HTML) fundamenta-se na eficiência para operações de busca/substituição em larga escala e na portabilidade da solução.

9.8 Gestão de Dependências e Reprodutibilidade

9.8.1 Requirements.txt

O arquivo requirements.txt localizado na raiz do repositório define as dependências Python essenciais do projeto, seguindo o formato padrão pip. A versão atual especifica apenas bibliotecas de alto nível (python-dotenv, requests, pandas, matplotlib, seaborn), delegando a resolução de dependências transitivas ao pip. Esta abordagem minimalista foi adotada para facilitar a instalação em ambientes com restrições de rede ou políticas de segurança que requerem auditoria de dependências. A ausência de pinning de versões específicas (ex: requests==2.31.0) representa decisão consciente para permitir atualizações de segurança automáticas, estratégia adequada para projetos acadêmicos com ciclo de vida limitado. Para ambientes de produção, recomenda-se a geração de requirements-lock.txt via pip freeze para garantir reprodutibilidade bit-a-bit.

9.8.2 Dependências Opcionais

Três dependências são classificadas como opcionais no contexto deste projeto: (1) Playwright (necessário apenas para regeneração de diagramas Mermaid), (2) wkhtmltopdf (necessário apenas para conversão HTML→PDF), e (3) Pandoc (necessário apenas para geração de documentação consolidada em formato DOCX). Estas ferramentas não são incluídas em requirements.txt pois não impactam o fluxo principal de geração de relatórios, sendo utilizadas exclusivamente em tarefas de manutenção e documentação. Esta separação de dependências segue

princípio de modularidade, permitindo deployment do sistema em ambientes minimalistas (ex: containers Docker) sem sobrecarga de ferramentas auxiliares. Scripts que requerem dependências opcionais (export_mermaid_to_png.py, consolidate_documentation.py) implementam verificação de disponibilidade via blocos try-except, exibindo mensagens informativas caso dependências estejam ausentes.

9.9 Considerações sobre Segurança de Dependências

A seleção de bibliotecas para este projeto considerou aspectos de segurança da cadeia de suprimentos de software (supply chain security), priorizando bibliotecas com histórico de manutenção ativa, processos de disclosure de vulnerabilidades estabelecidos e ampla adoção pela comunidade. Todas as dependências principais (Requests, Jinja2, Pandas) possuem mais de 10 anos de desenvolvimento e são mantidas por organizações reconhecidas (Python Software Foundation, Pallets, NumFOCUS). A ausência de bibliotecas de nicho ou pouco conhecidas reduz a superfície de ataque e facilita a auditoria de segurança. Adicionalmente, o sistema não utiliza funcionalidades de execução de código arbitrário (ex: eval(), exec()) nem desserialização insegura (ex: pickle.load() de fontes não confiáveis), mitigando riscos de injeção de código. A gestão de credenciais via python-dotenv, com exclusão explícita de arquivos .env do controle de versão, previne vazamento acidental de segredos em repositórios públicos.

APÊNDICE 2 – METODOLOGIA DE AVALIAÇÃO

A avaliação da solução proposta para a geração automatizada de relatórios de segurança, utilizando dados reais de SIEM (Wazuh) e modelos de Linguagem de Grande Escala (LLM), foi conduzida por meio de um framework de avaliação estruturado e um questionário administrado a especialistas no assunto. Este arcabouço metodológico visou garantir uma análise abrangente e multifacetada da qualidade e eficácia dos relatórios gerados.

Framework de Avaliação Proposto

O framework de avaliação foi concebido para analisar os relatórios sob três dimensões ortogonais, consideradas cruciais para o contexto de relatórios de segurança gerados por IA: Qualidade Linguística (A), Alinhamento ao Propósito (*Persona-Fit*) (B) e Desempenho do Modelo de IA (C). Cada dimensão foi subdividida em métricas específicas, acompanhadas de descrições claras e indicadores de escala para facilitar a avaliação.

A. **Qualidade Linguística**: Esta dimensão focou na clareza, legibilidade, coerência, estrutura, adequação da terminologia e tom/formalidade dos relatórios. O objetivo foi verificar se o texto era facilmente compreensível pelo público-alvo, se apresentava um fluxo lógico entre as seções e se utilizava a linguagem apropriada para cada persona (CEO, TIC, Compliance Officer).

B. **Alinhamento ao Propósito** (Persona-Fit): Esta dimensão avaliou a relevância do conteúdo para as necessidades específicas de cada persona, a acionabilidade das recomendações, a comunicação de risco (mapeando achados técnicos para impactos de negócio/regulatórios) e a cobertura dos artefatos requeridos (tabelas de evidências, listas de CVEs, etc.). A métrica B5, especificamente, avaliou a concisão e utilidade do sumário executivo para a tomada de decisões.

C. **Desempenho do Modelo de IA**: Esta dimensão investigou a consistência dos dados factuais entre os relatórios gerados por diferentes modelos para a mesma persona, a fidelidade aos dados originais (precisão dos valores numéricos), a profundidade da análise (discussão de tendências, mapeamento de frameworks, custo-benefício), a originalidade do texto (equilíbrio entre texto padrão e narrativa específica) e a taxa de erros factuais ou gramaticais.

A pontuação geral para cada relatório foi calculada como uma soma ponderada das médias das dimensões, com as seguintes ponderações: 30% para Qualidade Linguística (A), 40% para Alinhamento ao Propósito (B) e 30% para Desempenho do Modelo de IA (C). Esta ponderação foi escolhida para enfatizar a importância do alinhamento do conteúdo às necessidades da persona, que é um dos pilares da proposta de solução.

Instrumento de Avaliação Formal (Questionário)

O instrumento de avaliação formal consistiu em um questionário detalhado. Cada item do questionário foi respondido em uma escala *Likert* de 5 pontos (1 = Discordo Totalmente, 5 = Concordo Totalmente).

Tabela 17. Framework de Avaliação
Fonte: Próprio Autor

Dimensão	*Métrica*	*Descrição*	*Escala de Avaliação (1-5)*
A. Qualidade Linguística			
A1	Clareza e Legibilidade	O texto é claro, conciso e fácil de ler?	1=Muito Confuso, 5=Muito Claro
A2	Coerência e Estrutura	A estrutura do relatório é lógica e bem organizada?	1=Desorganizado, 5=Muito Coerente
A3	Adequação da Terminologia	A terminologia utilizada é apropriada para a persona?	1=Inadequada, 5=Muito Adequada
A4	Tom e Formalidade	O tom e o nível de formalidade são adequados para a persona?	1=Inapropriado, 5=Muito Apropriado
B. Alinhamento ao Propósito (Persona-Fit)			
B1	Relevância do Conteúdo	O conteúdo é relevante para as necessidades e responsabilidades da persona?	1=Irrelevante, 5=Muito Relevante

B2	Implementabilidade	As recomendações são claras, específicas e implementáveis pela persona?	1=Não, 5=Muito Implementavel
B3	Comunicação de Risco	O risco é comunicado de forma eficaz, com impacto claro para a persona?	1=Ineficaz, 5=Muito Eficaz
B4	Cobertura de Artefatos Requeridos	O relatório inclui os artefatos (tabelas, gráficos, listas) esperados pela persona?	1=Incompleto, 5=Completo
B5	Sumário Executivo	O sumário executivo é conciso, informativo e útil para a tomada de decisão da persona?	1=Inútil, 5=Muito Útil
C. Desempenho do Modelo de IA			
C1	Consistência entre Relatórios	Os dados factuais (ex: número de vulnerabilidades) são consistentes entre relatórios gerados para a mesma persona por diferentes modelos?	1=Inconsistente, 5=Totalmente Consistente
C2	Fidelidade aos Dados Originais	Os dados numéricos e factuais apresentados no relatório são fiéis aos dados de entrada (SIEM)?	1=Incorreto, 5=Totalmente Fiel
C3	Profundidade da Análise	A análise vai além da descrição, oferecendo insights, tendências ou implicações?	1=Superficial, 5=Muito Profunda
C4	Originalidade vs. Repetição	O texto é original ou parece uma repetição de frases padrão?	1=Muito Repetitivo, 5=Muito Original
C5	Taxa de Erros	O relatório contém erros factuais, gramaticais ou de formatação?	1=Muitos Erros, 5=Sem Erros

Tabela 18. Questionário (Instrumento de Avaliação Formal)
Fonte: Próprio Autor

ID	PERGUNTA	ESCALA DE RESPOSTA
A. QUALIDADE LINGUÍSTICA		
A1	O texto do relatório é claro, conciso e fácil de ler?	1=Muito Confuso, 5=Muito Claro
A2	A estrutura do relatório (títulos, subtítulos, parágrafos) é lógica e facilita a compreensão?	1=Desorganizado, 5=Muito Coerente
A3	A terminologia utilizada no relatório é apropriada para a minha função/persona?	1=Inadequada, 5=Muito Adequada
A4	O tom e o nível de formalidade do relatório são adequados para a minha função/persona?	1=Inapropriado, 5=Muito Apropriado
B. ALINHAMENTO AO PROPÓSITO (PERSONA-FIT)		
B1	O conteúdo do relatório é relevante para as minhas necessidades e responsabilidades?	1=Irrelevante, 5=Muito Relevante
B2	As recomendações apresentadas são claras, específicas e posso agir sobre elas?	1=Não, 5=Muito Implementavel
B3	O relatório comunica os riscos de segurança de forma eficaz, com um impacto claro para o meu contexto de negócio/operação?	1=Ineficaz, 5=Muito Eficaz
B4	O relatório inclui os artefatos (tabelas, gráficos, listas de vulnerabilidades/configurações) que eu esperaria ver?	1=Incompleto, 5=Completo
B5	O sumário executivo é conciso, informativo e útil para a minha tomada de decisão?	1=Inútil, 5=Muito Útil
C. DESEMPENHO DO MODELO DE IA		
C1	Os dados factuais (ex: número de vulnerabilidades, scores) são consistentes com o que eu esperaria ver em diferentes relatórios para a mesma situação?	1=Inconsistente, 5=Totalmente Consistente
C2	Os dados numéricos e factuais apresentados no relatório são precisos e fiéis aos dados de origem (SIEM)?	1=Incorreto, 5=Totalmente Fiel
C3	A análise fornecida no relatório vai além da descrição, oferecendo insights, tendências ou implicações relevantes?	1=Superficial, 5=Muito Profunda
C4	O texto do relatório parece original e personalizado, ou é muito genérico/repetitivo?	1=Muito Repetitivo, 5=Muito Original
C5	O relatório contém erros factuais, gramaticais ou de formatação?	1=Muitos Erros, 5=Sem Erros

APÊNDICE 3 – PROMPT GERAÇÃO MATRIZ COMPARATIVA

O prompt a seguir foi fornecido ao modelo Gemini 3 pro (e fornecidos os 18 relatórios, em blocos de 6 (Cada persona x 6 modelos utilizados para sua geração). O prompt instrui o modelo a sintetizar as pontuações quantitativas e produzir uma matriz comparativa.

Prompt

Você é um analista acadêmico preparando um estudo comparativo de relatórios de segurança gerados por IA.
Usando os resultados do questionário fornecidos abaixo, construa uma **matriz comparativa** que avalie os seis modelos de IA (Google Gemini-25flashlite, OpenAI GPT-OSS20b, XAI Grok41fast, ZAI GLM45air) para cada uma das três personas (CEO, Compliance Officer, TIC).
Para cada célula modelo-persona, reporte:
- **Qualidade Linguística (A)** – média de A1-A4 (escala 1-5).
- **Alinhamento ao Propósito (B)** – média de B1-B5 (escala 1-5).
- **Desempenho do Modelo de IA (C)** – média de C1-C5 (escala 1-5).
- **Pontuação Geral** – soma ponderada (A 30 %, B 40 %, C 30 %).
- **Principais Pontos Fortes** (máx. 2 bullet points).
- **Principais Pontos Fracos** (máx. 2 bullet points).

Apresente a matriz em formato de **tabela Markdown**, com linhas = Personas e colunas = Modelos. Inclua uma **legenda** explicando a ponderação e as escalas.

Após a tabela, escreva um breve **parágrafo interpretativo** (≈150 palavras) resumindo qual modelo tem o melhor desempenho geral, qual modelo mostra a maior vantagem específica para a persona e quaisquer trade-offs notáveis.

Dados:
(Insira aqui os valores do questionário preenchido para cada par modelo-persona, ex: “CEO-GoogleGemini: A1=5, A2=4, … C5=5”, etc.)

Garanta que a saída final não contenha **números inventados** – apenas os valores fornecidos na seção de dados.
Quando o prompt for executado, o LLM produzirá uma matriz como o exemplo abaixo (os valores são apenas ilustrativos):
Persona \ Modelo	Google Gemini	OpenAI GPT-OSS	XAI Grok	ZAI GLM
CEO	A = 4.5 B = 4.2 C = 4.3 **Geral** = 4.33 • Clareza estratégica • Forte conjunto de KPIs – Profundidade limitada em custo-benefício	...
Compliance Officer	...
TIC	...
A legenda deve ser: A = média dos itens A1-A4 (máx. 5); B = média dos itens B1-B5 (máx. 5); C = média dos itens C1-C5 (máx. 5). Geral = 0.30·A + 0.40·B + 0.30·C.

APÊNDICE 4 – RESPOSTAS BRUTAS DO QUESTIONÁRIO

ANÁLISE COMPLETA - PERSONA CEO

1. CEO - Modelo: DeepSeek Chat v3 (deepseek/deepseek-chat-v3-0324)

Arquivo: CEO_deepseek_deepseekchatv30324_2025_11_24_00_57.html

DIMENSÃO A: Qualidade Linguística

A1 - Clareza e Legibilidade: 4

- **Evidência:** "A avaliação estratégica da nossa postura de cibersegurança… revela uma exposição significativa a riscos críticos."
- **Justificativa:** Linguagem executiva adequada, direta e sem jargões excessivos. Uso apropriado de termos como "exposição", "postura" e "riscos críticos" que ressoam com audiência C-level.

A2 - Coerência e Estrutura: 4

- **Evidência:** Estrutura clara em "Avaliação da Postura de Segurança Atual" → "Recomendação Estratégica" → "Próximos Passos"
- **Justificativa:** Fluxo lógico bem definido. A divisão em três fases (30/60/contínuo) organiza bem o roadmap de ação.

A3 - Adequação do Vocabulário: 5

- **Evidência:** "paralisações custosas", "danos à marca difíceis de quantificar", "alocação imediata de recursos"
- **Justificativa:** Vocabulário perfeitamente alinhado ao nível estratégico - foco em impacto financeiro, reputacional e alocação de recursos (não em detalhes técnicos).

A4 - Ausência de Erros: 5

- **Evidência:** Revisão completa do texto não identificou erros gramaticais, ortográficos ou de formatação.
- **Justificativa:** Relatório profissional, sem falhas de português ou inconsistências de dados.

Média A: 4.5

DIMENSÃO B: Alinhamento ao Propósito

B1 - Relevância para a Persona: 5

- **Evidência:** "A concentração de vulnerabilidades críticas em sistemas essenciais poderia resultar em paralisações custosas e danos à marca"
- **Justificativa:** Foco total em impacto de negócio - não em detalhes técnicos. Traduz risco técnico em linguagem de ROI e risco reputacional.

B2 - Completude: 4

- **Evidência:** Inclui métricas críticas (Score SCA 27.41%, 91 vulnerabilidades críticas), recomendações estratégicas e timeline.
- **Justificativa:** Cobre todos os pontos essenciais para tomada de decisão executiva. Perde 1 ponto por não incluir estimativa de investimento necessário.

B3 - Acionabilidade: 5

- **Evidência:** "Recomendo a alocação imediata de recursos para um programa de correção em três fases: 1) Mitigação emergencial… (30 dias)"
- **Justificativa:** Recomendações específicas com prazos definidos e priorização clara. CEO sabe exatamente o que precisa aprovar.

B4 - Objetividade: 4

- **Evidência:** Uso de métricas quantificáveis (27.41%, 91 críticas, 32.819 total)
- **Justificativa:** Baseado em dados concretos. Perde 1 ponto por não incluir benchmarks de mercado ou comparação com padrões do setor.

B5 - Persuasão: 4

- **Evidência:** "exposição significativa a riscos críticos", "paralisações custosas", "danos à marca difíceis de quantificar"
- **Justificativa:** Tom persuasivo adequado, conectando riscos técnicos a consequências de negócio. Poderia ser mais enfático sobre urgência.

Média B: 4.4

DIMENSÃO C: Desempenho do Modelo de IA

C1 - Precisão Factual: 5

- **Evidência:** Todas as métricas (Score SCA: 27.41%, Vulnerabilidades: 32.819, Críticas: 91) correspondem exatamente aos dados brutos.
- **Justificativa:** Zero alucinações identificadas. Números e hosts apresentados estão corretos.

C2 - Consistência: 5

- **Evidência:** Referências consistentes aos frameworks (NIST CSF, ISO 27001, CIS Controls v8) ao longo do documento.
- **Justificativa:** Nenhuma contradição interna. Timeline e recomendações se reforçam mutuamente.

C3 - Profundidade de Análise: 4

- **Evidência:** "A concentração de vulnerabilidades críticas em sistemas essenciais" - identifica padrão estratégico relevante.
- **Justificativa:** Vai além da apresentação de números, conectando dados a insights de negócio. Poderia explorar mais implicações de longo prazo.

C4 - Criatividade/Originalidade: 3

- **Evidência:** Estrutura padrão de sumário executivo → recomendações → próximos passos.
- **Justificativa:** Abordagem conservadora e profissional, mas sem diferenciação significativa ou insights inovadores.

C5 - Adaptação Contextual: 5

- **Evidência:** Toda linguagem ajustada para nível C-level - zero jargão técnico, foco em impacto de negócio.
- **Justificativa:** Compreensão perfeita da persona CEO e adaptação completa do conteúdo.

Média C: 4.4

PONTUAÇÃO GERAL: 4.42

Cálculo: (4.5 × 0.30) + (4.4 × 0.40) + (4.4 × 0.30) = 1.35 + 1.76 + 1.32 = **4.43**

Pontos Fortes:

• **Linguagem executiva exemplar** - vocabulário estratégico focado em impacto de negócio • **Alta precisão factual** - zero alucinações, dados 100% consistentes com fonte

Pontos Fracos:

• **Falta estimativa de investimento** - não quantifica recursos necessários para as 3 fases • **Criatividade limitada** - abordagem conservadora sem insights diferenciados

2. CEO - Modelo: Google Gemini 2.5 Flash Lite

Arquivo: CEO_google_gemini25flashlite_2025_11_24_00_52.html

DIMENSÃO A: Qualidade Linguística

A1 - Clareza e Legibilidade: 5

- **Evidência:** "Nossa análise revela uma situação crítica que demanda atenção imediata da liderança executiva."
- **Justificativa:** Abertura direta e impactante. Linguagem clara, assertiva e sem ambiguidades.

A2 - Coerência e Estrutura: 5

- **Evidência:** Estrutura em "Situação Atual" → "Impactos Potenciais" → "Recomendações Prioritárias" → "Investimento Necessário"
- **Justificativa:** Fluxo narrativo excelente - diagnóstico, consequências, soluções e custos. Arquitetura ideal para decisão executiva.

A3 - Adequação do Vocabulário: 5

- **Evidência:** "janela de exposição crítica", "investimento estratégico", "retorno tangível em redução de risco"
- **Justificativa:** Vocabulário executivo de alto nível - foco em investimento, ROI, exposição e impacto estratégico.

A4 - Ausência de Erros: 5

- **Evidência:** Texto impecável, sem erros de gramática, formatação ou inconsistências.
- **Justificativa:** Qualidade profissional editorial.

Média A: 5.0

DIMENSÃO B: Alinhamento ao Propósito

B1 - Relevância para a Persona: 5

- **Evidência:** "Perda de confiança de clientes e parceiros", "Impacto na continuidade de negócios", "Exposição regulatória (LGPD, ISO 27001)"
- **Justificativa:** Foco preciso em consequências de negócio - reputação, operação, compliance. Linguagem perfeita para CEO.

B2 - Completude: 5

- **Evidência:** Inclui situação atual, impactos, recomendações priorizadas, timeline (30/60/90 dias) E estimativa de investimento.
- **Justificativa:** Documento completo para tomada de decisão - não falta nenhum elemento crítico.

B3 - Acionabilidade: 5

- **Evidência:** "Fase 1 (30 dias): Mitigação emergencial das 91 vulnerabilidades críticas. Investimento estimado: R$ 150-200 mil"
- **Justificativa:** Ações específicas com prazos, custos e responsabilidades. CEO tem roadmap executável.

B4 - Objetividade: 5

- **Evidência:** Métricas quantificadas (27.41%, 91 críticas) + estimativas financeiras concretas (R$ 150-200k, R$ 300-400k)
- **Justificativa:** Baseado em dados + quantificação de investimento necessário. Altamente objetivo.

B5 - Persuasão: 5

- **Evidência:** "janela de exposição crítica", "investimento estratégico com retorno tangível", "situação crítica que demanda atenção imediata"
- **Justificativa:** Tom de urgência bem calibrado, uso de evidências concretas e conexão clara entre investimento e retorno.

Média B: 5.0

DIMENSÃO C: Desempenho do Modelo de IA

C1 - Precisão Factual: 5

- **Evidência:** Métricas precisas (Score SCA: 27.41%, 32.819 vulnerabilidades, 91 críticas) alinhadas aos dados brutos.
- **Justificativa:** Zero inconsistências factuais.

C2 - Consistência: 5

- **Evidência:** Frameworks mencionados (NIST CSF, ISO 27001, CIS v8) usados consistentemente nas recomendações.
- **Justificativa:** Narrativa coesa do início ao fim, sem contradições.

C3 - Profundidade de Análise: 5

- **Evidência:** "A concentração de 82 das 91 vulnerabilidades críticas em um único host (cloud. safera) representa um ponto único de falha"
- **Justificativa:** Vai além de listar números - identifica padrões críticos (single point of failure) e traduz para risco de negócio.

C4 - Criatividade/Originalidade: 4

- **Evidência:** Inclusão de seção "Investimento Necessário" com estimativas de custo - diferencial em relação a outros modelos.
- **Justificativa:** Abordagem mais completa que o padrão, mas estrutura narrativa ainda convencional.

C5 - Adaptação Contextual: 5

- **Evidência:** Toda comunicação traduzida para linguagem de negócio - investimento, ROI, continuidade, exposição regulatória.
- **Justificativa:** Perfeita compreensão e adaptação para persona CEO.

Média C: 4.8

PONTUAÇÃO GERAL: 4.92

Cálculo: (5.0 × 0.30) + (5.0 × 0.40) + (4.8 × 0.30) = 1.50 + 2.00 + 1.44 = **4.94**

Pontos Fortes:

• **Completude excepcional** - único modelo que inclui estimativas de investimento detalhadas • **Profundidade analítica superior** - identifica single point of failure e traduz para risco estratégico

Pontos Fracos:

• **Criatividade moderada** - estrutura narrativa ainda conservadora apesar de ser mais completa • poderia **incluir benchmarks** - comparação com padrões de mercado agregaria contexto

3. CEO - Modelo: Google Gemini 3 Pro Preview

Arquivo: CEO_google_gemini3propreview_2025_11_24_01_01.html

DIMENSÃO A: Qualidade Linguística

A1 - Clareza e Legibilidade: 4

- **Evidência:** "A presente análise técnica da postura de cibersegurança… identifica uma situação de risco elevado"
- **Justificativa:** Linguagem clara, mas ligeiramente mais formal/técnica que o ideal para CEO. Termo "análise técnica" pode criar distância.

A2 - Coerência e Estrutura: 4

- **Evidência:** Estrutura "Postura Atual" → "Principais Riscos" → "Recomendações Estratégicas"
- **Justificativa:** Lógica bem construída, mas transições entre seções poderiam ser mais fluidas.

A3 - Adequação do Vocabulário: 4

- **Evidência:** "risco operacional significativo", "exposição crítica", "implementação de controles compensatórios"
- **Justificativa:** Vocabulário adequado, mas com alguns termos técnicos ("controles compensatórios") que poderiam ser simplificados.

A4 - Ausência de Erros: 5

- **Evidência:** Texto sem erros gramaticais ou de formatação.
- **Justificativa:** Qualidade editorial profissional.

Média A: 4.25

DIMENSÃO B: Alinhamento ao Propósito

B1 - Relevância para a Persona: 4

- **Evidência:** "Impacto na disponibilidade de serviços críticos", "Exposição a sanções regulatórias (LGPD)"
- **Justificativa:** Foca em impacto de negócio, mas usa linguagem ligeiramente mais técnica que o ideal. Menos persuasivo que Gemini 2.5.

B2 - Completude: 4

- **Evidência:** Inclui métricas, riscos identificados e recomendações com timeline (30/60/90 dias).
- **Justificativa:** Completo nos aspectos técnicos, mas falta quantificação de investimento necessário.

B3 - Acionabilidade: 4

- **Evidência:** "Fase 1 (30 dias): Remediação imediata das 91 vulnerabilidades críticas, priorizando o host cloud. safera"
- **Justificativa:** Ações claras com prazos, mas falta especificação de recursos e responsabilidades.

B4 - Objetividade: 5

- **Evidência:** Métricas precisas (27.41%, 91 críticas, 32.819 total) e priorização baseada em dados.
- **Justificativa:** Fundamentação quantitativa sólida.

B5 - Persuasão: 3

- **Evidência:** "situação de risco elevado", "requer atenção imediata"
- **Justificativa:** Tom mais neutro/técnico, menos impactante que outros modelos. Falta senso de urgência mais forte.

Média B: 4.0

DIMENSÃO C: Desempenho do Modelo de IA

C1 - Precisão Factual: 5

- **Evidência:** Métricas corretas e alinhadas aos dados brutos.
- **Justificativa:** Sem alucinações ou inconsistências.

C2 - Consistência: 5

- **Evidência:** Referências consistentes a frameworks e métricas ao longo do documento.
- **Justificativa:** Narrativa coesa.

C3 - Profundidade de Análise: 4

- **Evidência:** "A concentração de vulnerabilidades em um único host representa risco operacional significativo"
- **Justificativa:** Identifica padrões relevantes, mas análise menos detalhada que Gemini 2.5 Flash Lite.

C4 - Criatividade/Originalidade: 3

- **Evidência:** Estrutura padrão de relatório executivo.

- **Justificativa:** Abordagem convencional sem diferenciação.

C5 - Adaptação Contextual: 4

- **Evidência:** Linguagem ajustada para executivo, mas com viés técnico residual.
- **Justificativa:** Boa adaptação, mas não perfeita - poderia ser mais "business-oriented".

Média C: 4.2

PONTUAÇÃO GERAL: 4.13

Cálculo: (4.25 × 0.30) + (4.0 × 0.40) + (4.2 × 0.30) = 1.28 + 1.60 + 1.26 = **4.14**

Pontos Fortes:

• **Precisão factual impecável** - dados 100% corretos e consistentes • **Estrutura lógica sólida** - fluxo bem organizado de diagnóstico a recomendações

Pontos Fracos:

• **Linguagem técnica residual** - termos como "controles compensatórios" podem afastar CEO • **Persuasão limitada** - tom neutro não transmite urgência adequada para situação crítica

4. CEO - Modelo: OpenAI GPT-OSS 20B Free

Arquivo: CEO_openai_gptoss20b_free_2025_11_24_00_51.html

DIMENSÃO A: Qualidade Linguística

A1 - Clareza e Legibilidade: 3

- **Evidência:** "A análise abrangente da nossa infraestrutura de TI… revela deficiências significativas"
- **Justificativa:** Linguagem adequada, mas menos impactante. Uso de "deficiências" é técnico demais - "exposições" ou "riscos" seria melhor.

A2 - Coerência e Estrutura: 4

- **Evidência:** Estrutura "Visão Geral" → "Riscos Prioritários" → "Recomendações Imediatas"
- **Justificativa:** Organização lógica, mas títulos de seções menos persuasivos que outros modelos.

A3 - Adequação do Vocabulário: 3

- **Evidência:** "deficiências significativas na postura de segurança", "infraestrutura de TI"
- **Justificativa:** Vocabulário misto - alguns termos executivos, outros técnicos. Inconsistente no tom.

A4 - Ausência de Erros: 5

- **Evidência:** Sem erros gramaticais ou de formatação.
- **Justificativa:** Qualidade editorial adequada.

Média A: 3.75

DIMENSÃO B: Alinhamento ao Propósito

B1 - Relevância para a Persona: 3

- **Evidência:** "Risco de interrupção de serviços", "Possível não conformidade com LGPD e ISO 27001"
- **Justificativa:** Menciona impactos de negócio, mas de forma genérica. Não quantifica ou contextualiza adequadamente.

B2 - Completude: 3

- **Evidência:** Inclui métricas básicas e recomendações, mas sem timeline detalhado ou estimativas de investimento.

- **Justificativa:** Faltam elementos críticos para decisão executiva (custo, prazos específicos, ROI).

B3 - Acionabilidade: 3

- **Evidência:** "Priorizar a correção das 91 vulnerabilidades críticas", "Melhorar o score de conformidade"
- **Justificativa:** Recomendações genéricas sem prazos específicos ou roadmap de implementação detalhado.

B4 - Objetividade: 4

- **Evidência:** Apresenta métricas quantificadas (27.41%, 91 críticas)
- **Justificativa:** Baseado em dados, mas sem análise aprofundada ou contextualização.

B5 - Persuasão: 3

- **Evidência:** "deficiências significativas", "atenção imediata da liderança"
- **Justificativa:** Tentativa de persuasão presente, mas linguagem menos impactante que outros modelos.

Média B: 3.2

DIMENSÃO C: Desempenho do Modelo de IA

C1 - Precisão Factual: 5

- **Evidência:** Métricas corretas e alinhadas aos dados brutos.
- **Justificativa:** Sem alucinações detectadas.

C2 - Consistência: 4

- **Evidência:** Frameworks mencionados, mas uso menos consistente ao longo do texto.
- **Justificativa:** Coeso, mas com menos integração dos frameworks nas recomendações.

C3 - Profundidade de Análise: 3

- **Evidência:** "Concentração de vulnerabilidades em cloud. safera representa risco"
- **Justificativa:** Análise superficial - identifica o problema, mas não explora implicações estratégicas profundas.

C4 - Criatividade/Originalidade: 2

- **Evidência:** Estrutura genérica de relatório executivo.
- **Justificativa:** Abordagem padrão sem diferenciação ou insights inovadores.

C5 - Adaptação Contextual: 3

- **Evidência:** Mistura de linguagem executiva e técnica.
- **Justificativa:** Adaptação parcial à persona - não mantém consistência de tom executivo.

Média C: 3.4

PONTUAÇÃO GERAL: 3.42

Cálculo: (3.75 × 0.30) + (3.2 × 0.40) + (3.4 × 0.30) = 1.13 + 1.28 + 1.02 = **3.43**

Pontos Fortes:

• **Precisão factual mantida** - dados corretos sem alucinações • **Estrutura básica adequada** - organização lógica do conteúdo

Pontos Fracos:

• **Completude limitada** - falta timeline detalhado e estimativas de investimento • **Profundidade analítica superficial** - não explora implicações estratégicas dos riscos identificados

5. CEO - Modelo: X.AI Grok 4.1 Fast

Arquivo: CEO_xai_grok41fast_2025_11_24_00_59.html

DIMENSÃO A: Qualidade Linguística

A1 - Clareza e Legibilidade: 5

- **Evidência:** "Nossa postura de cibersegurança encontra-se em estado crítico, demandando decisões executivas imediatas."
- **Justificativa:** Abertura direta, assertiva e impactante. Clareza cristalina na mensagem principal.

A2 - Coerência e Estrutura: 5

- **Evidência:** Estrutura "Estado Crítico Atual" → "Impactos ao Negócio" → "Plano de Ação Executivo" → "Investimento e ROI"
- **Justificativa:** Fluxo narrativo perfeito para CEO - problema, consequências, solução e retorno financeiro.

A3 - Adequação do Vocabulário: 5

- **Evidência:** "decisões executivas imediatas", "exposição inaceitável", "investimento defensivo com ROI mensurável"
- **Justificativa:** Vocabulário executivo de elite - foco em decisão, risco, investimento e retorno.

A4 - Ausência de Erros: 5

- **Evidência:** Texto impecável sem erros.
- **Justificativa:** Qualidade editorial profissional.

Média A: 5.0

DIMENSÃO B: Alinhamento ao Propósito

B1 - Relevância para a Persona: 5

- **Evidência:** "Risco de paralisia operacional estimada em R$ 500k-1M por dia de downtime", "Exposição regulatória (LGPD) com multas de até 2% do faturamento"
- **Justificativa:** Quantificação financeira precisa de riscos - linguagem perfeita para CEO. Traduz tudo para impacto monetário.

B2 - Completude: 5

- **Evidência:** Inclui estado atual, impactos quantificados, plano de ação faseado (30/60/90), investimento necessário (R$ 180-250k + R$ 400-600k) e projeção de ROI.
- **Justificativa:** Documento completo para decisão de investimento - inclui todos os elementos críticos.

B3 - Acionabilidade: 5

- **Evidência:** "Fase 1 (30 dias): Correção emergencial de 91 críticas. Orçamento: R$ 180-250k. Responsável: CTO + equipe de resposta rápida"
- **Justificativa:** Ações específicas com prazos, custos, responsáveis e métricas de sucesso. Totalmente executável.

B4 - Objetividade: 5

- **Evidência:** Métricas quantificadas + estimativas financeiras detalhadas + projeção de ROI ("redução de 70% do risco crítico em 90 dias")
- **Justificativa:** Máxima objetividade com dados concretos e projeções de resultado.

B5 - Persuasão: 5

- **Evidência:** "estado crítico", "exposição inaceitável", "R$ 500k-1M por dia de downtime", "investimento defensivo com ROI mensurável"
- **Justificativa:** Tom de urgência calibrado perfeitamente, uso de quantificação financeira para maximizar impacto persuasivo.

Média B: 5.0

DIMENSÃO C: Desempenho do Modelo de IA

C1 - Precisão Factual: 4

- **Evidência:** Métricas bases corretas (27.41%, 91 críticas), mas estimativas financeiras (R$ 500k-1M/dia) são inferências razoáveis, não dados brutos.
- **Justificativa:** Precisão alta nas métricas de segurança. Estimativas financeiras são plausíveis, mas não verificáveis nos dados originais.

C2 - Consistência: 5

- **Evidência:** Narrativa coesa com frameworks (NIST CSF, ISO 27001) integrados em todas as recomendações.
- **Justificativa:** Totalmente consistente do início ao fim.

C3 - Profundidade de Análise: 5

- **Evidência:** "82 das 91 críticas concentradas em cloud.safera = ponto único de falha com potencial de paralisia total de operações cloud-dependent"
- **Justificativa:** Análise estratégica profunda - identifica single point of failure e traduz para impacto operacional concreto.

C4 - Criatividade/Originalidade: 5

- **Evidência:** Inclusão de seção "Investimento e ROI" com projeção de redução de risco (70% em 90 dias) e análise custo-benefício.
- **Justificativa:** Abordagem inovadora - trata segurança como investimento defensivo com ROI mensurável. Diferenciação significativa.

C5 - Adaptação Contextual: 5

- **Evidência:** Toda comunicação traduzida para linguagem financeira e de negócio - investimento, ROI, downtime, receita em risco.
- **Justificativa:** Compreensão perfeita da persona CEO e adaptação total do conteúdo.

Média C: 4.8

PONTUAÇÃO GERAL: 4.92

Cálculo: (5.0 × 0.30) + (5.0 × 0.40) + (4.8 × 0.30) = 1.50 + 2.00 + 1.44 = **4.94**

Pontos Fortes:

• **Quantificação financeira excepcional** - traduz todos os riscos para impacto monetário direto (R$ 500k-1M/dia) • **Abordagem ROI inovadora** - trata segurança como investimento defensivo com retorno mensurável

Pontos Fracos:

• **Estimativas não verificáveis** - valores financeiros são inferências razoáveis, mas não constam nos dados brutos • **Risco de over-simplificação** - quantificação agressiva pode ser questionada sem metodologia explícita

6. CEO - Modelo: Z.AI GLM 4.5 Air Free

Arquivo: CEO_zai_glm45air_free_2025_11_24_00_55.html

DIMENSÃO A: Qualidade Linguística

A1 - Clareza e Legibilidade: 4

- **Evidência:** "A análise de cibersegurança realizada revela uma necessidade urgente de intervenção estratégica"
- **Justificativa:** Linguagem clara e direta, mas menos impactante que os melhores modelos. Abertura adequada, mas não memorável.

A2 - Coerência e Estrutura: 4

- **Evidência:** Estrutura "Diagnóstico Atual" → "Riscos Principais" → "Plano de Ação Recomendado"
- **Justificativa:** Organização lógica e bem sequenciada. Transições fluidas entre seções.

A3 - Adequação do Vocabulário: 4

- **Evidência:** "intervenção estratégica", "exposição crítica", "priorização baseada em risco"
- **Justificativa:** Vocabulário executivo adequado, mas sem o nível de sofisticação dos modelos top (falta termos como ROI, downtime quantificado).

A4 - Ausência de Erros: 5

- **Evidência:** Texto sem erros gramaticais ou de formatação.
- **Justificativa:** Qualidade editorial profissional.

Média A: 4.25

DIMENSÃO B: Alinhamento ao Propósito

B1 - Relevância para a Persona: 4

- **Evidência:** "Impacto na disponibilidade de serviços críticos", "Risco reputacional e de conformidade (LGPD, ISO 27001)"
- **Justificativa:** Foca em impactos de negócio relevantes, mas sem quantificação financeira. Menos persuasivo que modelos top.

B2 - Completude: 4

- **Evidência:** Inclui diagnóstico, riscos priorizados, plano de ação com fases (30/60/90 dias) e métricas de sucesso.
- **Justificativa:** Completo nos aspectos estratégicos, mas falta estimativa de investimento necessário.

B3 - Acionabilidade: 4

- **Evidência:** "Fase 1 (30 dias): Remediação das 91 vulnerabilidades críticas, priorizando cloud.safera (82 vulnerabilidades)"
- **Justificativa:** Ações claras com prazos e priorização, mas sem especificação de recursos ou responsáveis.

B4 - Objetividade: 4

- **Evidência:** Métricas quantificadas (27.41%, 91 críticas) e priorização baseada em severidade.
- **Justificativa:** Baseado em dados concretos, mas sem benchmarks ou contexto de mercado.

B5 - Persuasão: 4

- **Evidência:** "necessidade urgente", "exposição crítica", "priorização absoluta"
- **Justificativa:** Tom de urgência adequado, mas menos impactante que os modelos que quantificam financeiramente.

Média B: 4.0

DIMENSÃO C: Desempenho do Modelo de IA

C1 - Precisão Factual: 5

- **Evidência:** Métricas corretas e alinhadas aos dados brutos (Score SCA: 27.41%, 91 críticas, cloud.safera com 82).
- **Justificativa:** Zero alucinações ou inconsistências.

C2 - Consistência: 5

- **Evidência:** Frameworks (NIST CSF, ISO 27001, CIS v8) mencionados e usados consistentemente.

- **Justificativa:** Narrativa coesa sem contradições.

C3 - Profundidade de Análise: 4

- **Evidência:** "Concentração de 90% das vulnerabilidades críticas em cloud.safera configura ponto único de falha"
- **Justificativa:** Identifica padrão crítico (single point of failure) e prioriza adequadamente, mas não explora implicações de negócio profundamente.

C4 - Criatividade/Originalidade: 3

- **Evidência:** Estrutura padrão de relatório executivo.
- **Justificativa:** Abordagem convencional sem diferenciação significativa.

C5 - Adaptação Contextual: 4

- **Evidência:** Linguagem ajustada para nível executivo com foco em impacto estratégico.
- **Justificativa:** Boa adaptação à persona, mas sem excelência dos modelos que traduzem tudo para impacto financeiro.

Média C: 4.2

PONTUAÇÃO GERAL: 4.13

Cálculo: (4.25 × 0.30) + (4.0 × 0.40) + (4.2 × 0.30) = 1.28 + 1.60 + 1.26 = **4.14**

Pontos Fortes:

• **Precisão factual impecável** - dados 100% corretos e análise baseada em evidências • **Estrutura sólida e coesa** - organização lógica com fases de ação bem definidas

Pontos Fracos:

• **Falta quantificação financeira** - não estima custos de investimento ou impacto monetário de riscos • **Criatividade limitada** - abordagem padrão sem insights inovadores ou diferenciação

Tabela 19. Matriz comparativa – Persona CEO
Fonte: Próprio Autor

MODELO	A	B	C	GERAL	PONTOS FORTES	PONTOS FRACOS
DeepSeek Chat v3	4.5	4.4	4.4	4.43	• Linguagem executiva exemplar • Alta precisão factual (0 alucinações)	• Falta estimativa de investimento • Criatividade limitada
Gemini 2.5 Flash Lite	5.0	5.0	4.8	4.94	• Completude excepcional (inclui custos) • identifica *single point of failure*	• Estrutura narrativa conservadora • poderia incluir *benchmarks*
Gemini 3 Pro Preview	4.25	4.0	4.2	4.14	• Precisão factual impecável • Estrutura lógica sólida	• Linguagem com viés técnico residual • Persuasão limitada
GPT-OSS 20B Free	3.75	3.2	3.4	3.43	• Precisão factual mantida • Estrutura básica adequada	• Falta timeline/custos detalhados • Análise superficial
Grok 4.1 Fast	5.0	5.0	4.8	4.94	• Quantificação financeira excepcional • Abordagem ROI inovadora	• Estimativas não verificáveis • Risco de over-simplificação
GLM 4.5 Air Free	4.25	4.0	4.2	4.14	• Precisão factual impecável • Estrutura coesa e organizada	• Falta quantificação financeira • Abordagem padrão sem inovação

Legenda:
A = Qualidade Linguística (média A1-A4, escala 1-5)
B = Alinhamento ao Propósito (média B1-B5, escala 1-5)
C = Desempenho do Modelo de IA (média C1-C5, escala 1-5)
Geral = (A × 0.30) + (B × 0.40) + (C × 0.30)

ANÁLISE INTERPRETATIVA - PERSONA CEO

Empate técnico entre Gemini 2.5 Flash Lite e Grok 4.1 Fast (4.94) como melhores performers gerais, ambos destacando-se por completude excepcional e linguagem executiva impecável. O **Gemini 2.5** vence em confiabilidade ao incluir estimativas de investimento verificáveis (R$ 150-200k fase 1, R$ 300-400k fase 2), enquanto o **Grok 4.1** diferencia-se pela abordagem inovadora de ROI mensurável e quantificação financeira agressiva de riscos (R$ 500k-1M/dia de downtime). Para CEOs avessos a risco, o Gemini 2.5 é mais seguro; para perfis mais agressivos que valorizam análise custo-benefício quantificada, o Grok é superior. O **DeepSeek (4.43)** oferece excelente equilíbrio qualidade-custo sem estimativas financeiras, enquanto **GPT-OSS 20B (3.43)** mostra limitações significativas em profundidade analítica e completude. Trade-off notável: modelos que incluem quantificação financeira (Gemini 2.5, Grok) têm maior poder persuasivo, mas exigem validação metodológica das estimativas.

ANÁLISE COMPLETA - PERSONA COMPLIANCE OFFICER

1. COMPLIANCE OFFICER - Modelo: DeepSeek Chat v3

Arquivo: Compliance_deepseek_deepseekchatv30324_2025_11_24_00_57.html

DIMENSÃO A: Qualidade Linguística

A1 - Clareza e Legibilidade: 5

- **Evidência:** "A avaliação de conformidade revela gaps críticos de segurança que comprometem nossa postura regulatória diante de LGPD, ISO 27001 e frameworks internacionais."
- **Justificativa:** Linguagem precisa e direta focada em conformidade regulatória. Clareza cristalina na mensagem principal para audiência de compliance.

A2 - Coerência e Estrutura: 5

- **Evidência:** Estrutura "Gaps Críticos de Conformidade" → "Exposição Regulatória" → "Plano de Remediação" → "Evidências para Auditoria"
- **Justificativa:** Fluxo narrativo perfeito para Compliance Officer - diagnóstico de gaps, exposição legal, ações corretivas e documentação auditável.

A3 - Adequação do Vocabulário: 5

- **Evidência:** "gaps críticos", "exposição regulatória", "controles compensatórios", "evidências auditáveis", "baseline de conformidade"
- **Justificativa:** Vocabulário técnico de compliance perfeito - usa termos específicos da área (gaps, baseline, controles compensatórios) sem simplificação excessiva.

A4 - Ausência de Erros: 5

- **Evidência:** Texto impecável sem erros gramaticais ou de formatação.
- **Justificativa:** Qualidade editorial profissional.

Média A: 5.0

DIMENSÃO B: Alinhamento ao Propósito

B1 - Relevância para a Persona: 5

- **Evidência:** "Art. 46 LGPD: Falta de controles de acesso adequados pode caracterizar 'tratamento inadequado' de dados pessoais", "ISO 27001 Anexo A.9: Não conformidade com 8 dos 14 controles de Controle de Acesso"
- **Justificativa:** Foco total em exposição legal e frameworks de compliance. Mapeia cada gap técnico para artigo específico de lei/norma.

B2 - Completude: 5

- **Evidência:** Inclui gaps de conformidade mapeados por framework (LGPD, ISO 27001, NIST, CIS), exposição regulatória quantificada (Score SCA 27.41%), plano de remediação com prazos (30/60/90 dias) e seção dedicada a "Evidências para Auditoria".

- **Justificativa:** Documento completo para gestão de compliance - não falta nenhum elemento crítico para preparação de auditoria.

B3 - Acionabilidade: 5

- **Evidência:** "Fase 1 (30 dias): Implementar políticas de senha conforme CIS Control 5 (mínimo 14 caracteres, complexidade, expiração 365 dias). Evidência: GPO aplicada + relatório de conformidade."

- **Justificativa:** Ações específicas com prazos, controles a implementar E tipo de evidência a coletar para auditoria. Totalmente executável.

B4 - Objetividade: 5

- **Evidência:** "Score SCA: 27.41% (meta: ≥90%)", "8 de 14 controles de Controle de Acesso não conformes", "91 vulnerabilidades críticas = potencial de incidente reportável LGPD Art. 48"

- **Justificativa:** Máxima objetividade - quantifica gaps, mapeia para artigos legais específicos e define metas numéricas de conformidade.

B5 - Persuasão: 5

- **Evidência:** "gaps críticos", "exposição regulatória", "potencial de incidente reportável", "risco de multa de até 2% do faturamento (LGPD Art. 52)"

- **Justificativa:** Tom de urgência calibrado com consequências legais específicas. Usa linguagem de risco regulatório para maximizar atenção.

Média B: 5.0

DIMENSÃO C: Desempenho do Modelo de IA

C1 - Precisão Factual: 5

- **Evidência:** Métricas corretas (Score SCA 27.41%, 91 críticas) e mapeamento preciso para artigos legais (Art. 46, 48, 52 LGPD; ISO 27001 A.9, A.18).

- **Justificativa:** Zero alucinações. Artigos legais citados existem e são aplicáveis ao contexto.

C2 - Consistência: 5

- **Evidência:** Frameworks (LGPD, ISO 27001, NIST SP 800-53, CIS v8) usados consistentemente em gaps, exposição e remediação.

- **Justificativa:** Narrativa totalmente coesa com cross-reference entre frameworks.

C3 - Profundidade de Análise: 5

- **Evidência:** "Ausência de políticas de senha (comprimento mínimo, complexidade, expiração) viola simultaneamente: CIS Control 5.2, NIST AC-7, ISO 27001 A.9.4.3 e expõe ao Art. 46 LGPD"

- **Justificativa:** Análise multi-framework profunda - cada gap é mapeado para múltiplos controles/artigos simultaneamente, mostrando interconexões.

C4 - Criatividade/Originalidade: 5

- **Evidência:** Inclusão de seção "Evidências para Auditoria" com lista específica de documentos a coletar (GPOs, logs, relatórios de scan).

- **Justificativa:** Abordagem inovadora e altamente prática - antecipa necessidades de auditoria e prepara documentação necessária.

C5 - Adaptação Contextual: 5

- **Evidência:** Toda comunicação em linguagem de compliance - gaps, baseline, evidências auditáveis, exposição regulatória.

- **Justificativa:** Compreensão perfeita da persona Compliance Officer e adaptação total do conteúdo.

Média C: 5.0

PONTUAÇÃO GERAL: 5.00

Cálculo: (5.0 × 0.30) + (5.0 × 0.40) + (5.0 × 0.30) = 1.50 + 2.00 + 1.50 = **5.00**

Pontos Fortes:

• **Mapeamento legal perfeito** - cada gap técnico vinculado a artigos específicos de LGPD, ISO 27001, NIST • **Seção "Evidências para Auditoria" inovadora** - antecipa necessidades documentais e prepara coleta

Pontos Fracos:

• **Nenhum identificado** - relatório exemplar em todas as dimensões para persona Compliance

• **Poderia incluir templates** - fornecer modelos de políticas/procedimentos aceleraria implementação

2. COMPLIANCE OFFICER - Modelo: Google Gemini 2.5 Flash Lite

Arquivo: Compliance_google_gemini25flashlite_2025_11_24_00_52.html

DIMENSÃO A: Qualidade Linguística

A1 - Clareza e Legibilidade: 5

- **Evidência:** "Nossa análise de conformidade identifica desvios críticos que expõem a organização a riscos regulatórios mensuráveis."
- **Justificativa:** Linguagem clara focada em desvios de conformidade e exposição regulatória. Tom apropriado para Compliance Officer.

A2 - Coerência e Estrutura: 5

- **Evidência:** Estrutura "Gaps de Conformidade por Framework" → "Exposição Regulatória Quantificada" → "Roadmap de Adequação" → "Matriz de Controles"
- **Justificativa:** Fluxo lógico excelente - diagnóstico por framework, quantificação de risco legal, plano de ação e matriz de rastreabilidade.

A3 - Adequação do Vocabulário: 5

- **Evidência:** "desvios críticos", "baseline de conformidade", "matriz de rastreabilidade", "controles compensatórios", "evidências objetivas"
- **Justificativa:** Vocabulário técnico de compliance de alto nível - termos específicos e precisos da área.

A4 - Ausência de Erros: 5

- **Evidência:** Texto impecável sem erros.
- **Justificativa:** Qualidade editorial profissional.

Média A: 5.0

DIMENSÃO B: Alinhamento ao Propósito

B1 - Relevância para a Persona: 5

- **Evidência:** "LGPD Art. 46 + 48: Tratamento inadequado + obrigação de notificação em caso de incidente (91 vulnerabilidades críticas = alto risco de materialização)", "ISO 27001: Não conformidade com 72.59% dos controles avaliados"
- **Justificativa:** Foco preciso em exposição legal com quantificação de risco. Linguagem perfeita para Compliance Officer.

B2 - Completude: 5

- **Evidência:** Inclui gaps por framework (LGPD, ISO 27001, NIST, CIS), exposição quantificada (27.41% conformidade), roadmap faseado (30/60/90/180 dias) e "Matriz de Controles" cruzando frameworks.

- **Justificativa:** Documento completo com elemento diferenciado - matriz de rastreabilidade entre frameworks facilita gestão de múltiplas auditorias.

B3 - Acionabilidade: 5

- **Evidência:** "Fase 1 (30 dias): Implementar baseline mínimo de segurança: Políticas de senha (CIS 5.2), Bloqueio de conta (NIST AC-7), Renomeação de contas padrão (CIS 5.1). Evidência: Documentação de políticas + screenshots de GPO."

- **Justificativa:** Ações específicas com controles a implementar, frameworks de referência e evidências a coletar. Totalmente executável.

B4 - Objetividade: 5

- **Evidência:** "Score SCA: 27.41%", "72.59% de não conformidade", "91 vulnerabilidades críticas", "Matriz de Controles" com status por framework

- **Justificativa:** Máxima objetividade com métricas quantificadas e matriz de rastreabilidade que permite gestão visual do status de conformidade.

B5 - Persuasão: 5

- **Evidência:** "desvios críticos", "riscos regulatórios mensuráveis", "alto risco de materialização", "potencial multa de 2% faturamento"

- **Justificativa:** Tom de urgência adequado com consequências regulatórias específicas e quantificadas.

Média B: 5.0

DIMENSÃO C: Desempenho do Modelo de IA

C1 - Precisão Factual: 5

- **Evidência:** Métricas corretas e artigos legais precisos (LGPD Art. 46, 48, 52; ISO 27001 Anexo A).

- **Justificativa:** Zero alucinações ou inconsistências.

C2 - Consistência: 5

- **Evidência:** Uso consistente de frameworks em todas as seções com cross-reference na Matriz de Controles.

- **Justificativa:** Narrativa totalmente coesa.

C3 - Profundidade de Análise: 5

- **Evidência:** "Score SCA 27.41% indicam que 72.59% dos controles avaliados estão não conformes, criando múltiplos vetores de exposição legal simultâneos"

- **Justificativa:** Análise profunda que conecta métricas técnicas a implicações regulatórias múltiplas.

C4 - Criatividade/Originalidade: 5

- **Evidência:** Inclusão de "Matriz de Controles" cruzando LGPD, ISO 27001, NIST e CIS com status de conformidade por controle.

- **Justificativa:** Abordagem inovadora - matriz de rastreabilidade é ferramenta prática para gestão de múltiplas auditorias simultaneamente.

C5 - Adaptação Contextual: 5

- **Evidência:** Toda comunicação em linguagem de compliance com ferramentas específicas da área (matriz de rastreabilidade, evidências objetivas).

- **Justificativa:** Compreensão perfeita da persona e adaptação total do conteúdo.

Média C: 5.0

PONTUAÇÃO GERAL: 5.00

Cálculo: (5.0 × 0.30) + (5.0 × 0.40) + (5.0 × 0.30) = 1.50 + 2.00 + 1.50 = **5.00**

Pontos Fortes:

• **Matriz de Controles inovadora** - rastreabilidade cruzada entre frameworks facilita gestão de múltiplas auditorias • **Quantificação precisa de exposição** - traduz score técnico em percentual de não conformidade regulatória

Pontos Fracos:

• **Nenhum identificado** - relatório exemplar em todas as dimensões

• **Poderia adicionar timeline de auditorias** - calendário de certificações/revisões agregaria valor

3. COMPLIANCE OFFICER - Modelo: Google Gemini 3 Pro Preview

Arquivo: Compliance_google_gemini3propreview_2025_11_24_01_01.html

DIMENSÃO A: Qualidade Linguística

A1 - Clareza e Legibilidade: 4

- **Evidência:** "A análise de conformidade de configuração identifica desvios significativos em relação aos frameworks regulatórios adotados."
- **Justificativa:** Linguagem clara, mas menos impactante que os modelos anteriores. Termo "desvios significativos" é adequado, mas menos urgente que "gaps críticos".

A2 - Coerência e Estrutura: 5

- **Evidência:** Estrutura "Status de Conformidade" → "Gaps Prioritários por Framework" → "Exposição Regulatória" → "Plano de Adequação"
- **Justificativa:** Fluxo lógico bem construído e organizado.

A3 - Adequação do Vocabulário: 4

- **Evidência:** "desvios significativos", "baseline de conformidade", "controles não implementados", "evidências documentais"
- **Justificativa:** Vocabulário adequado, mas menos sofisticado que os modelos tops (falta termos como "matriz de rastreabilidade", "controles compensatórios").

A4 - Ausência de Erros: 5

- **Evidência:** Texto sem erros gramaticais ou de formatação.
- **Justificativa:** Qualidade editorial profissional.

Média A: 4.5

DIMENSÃO B: Alinhamento ao Propósito

B1 - Relevância para a Persona: 5

- **Evidência:** "LGPD Art. 46: Ausência de medidas de segurança adequadas configura tratamento inadequado", "ISO 27001 A.9: 8 de 14 controles de Acesso não conformes"
- **Justificativa:** Foco correto em exposição legal e gaps de framework. Mapeia para artigos específicos.

B2 - Completude: 4

- **Evidência:** Inclui status de conformidade, gaps por framework, exposição regulatória e plano de adequação com fases (30/60/90 dias).
- **Justificativa:** Completo nos aspectos essenciais, mas falta ferramentas avançadas como matriz de rastreabilidade ou seção específica de evidências para auditoria.

B3 - Acionabilidade: 4

- **Evidência:** "Fase 1 (30 dias): Implementar políticas de senha (CIS 5.2, NIST AC-7): mínimo 14 caracteres, complexidade, expiração 365 dias."
- **Justificativa:** Ações claras com prazos e controles a implementar, mas sem especificação de evidências a coletar para auditoria.

B4 - Objetividade: 5

- **Evidência:** "Score SCA: 27.41%", "72.59% de não conformidade", quantificação de gaps por framework
- **Justificativa:** Baseado em métricas quantificadas e precisas.

B5 - Persuasão: 4

- **Evidência:** "desvios significativos", "exposição regulatória", "não conformidade"
- **Justificativa:** Tom adequado, mas menos impactante que modelos que quantificam multas potenciais (ex: 2% faturamento LGPD).

Média B: 4.4

DIMENSÃO C: Desempenho do Modelo de IA

C1 - Precisão Factual: 5

- **Evidência:** Métricas corretas e artigos legais precisos.
- **Justificativa:** Sem alucinações ou inconsistências.

C2 - Consistência: 5

- **Evidência:** Frameworks usados consistentemente ao longo do documento.
- **Justificativa:** Narrativa coesa.

C3 - Profundidade de Análise: 4

- **Evidência:** "Ausência de políticas de senha viola CIS 5.2, NIST AC-7 e ISO 27001 A.9.4.3"
- **Justificativa:** Análise *multi-framework* adequada, mas menos detalhada que modelos que exploram interconexões entre gaps.

C4 - Criatividade/Originalidade: 3

- **Evidência:** Estrutura padrão de relatório de compliance.
- **Justificativa:** Abordagem convencional sem ferramentas diferenciadas (ex: matriz de rastreabilidade, seção de evidências).

C5 - Adaptação Contextual: 5

- **Evidência:** Linguagem ajustada para compliance com foco em frameworks e exposição regulatória.
- **Justificativa:** Boa adaptação à persona.

Média C: 4.4

PONTUAÇÃO GERAL: 4.44

Cálculo: (4.5 × 0.30) + (4.4 × 0.40) + (4.4 × 0.30) = 1.35 + 1.76 + 1.32 = **4.43**

Pontos Fortes:

• **Precisão factual impecável** - métricas e artigos legais corretos • **Estrutura lógica sólida** - organização clara por framework e fases de adequação

Pontos Fracos:

• **Falta ferramentas avançadas** - não inclui matriz de rastreabilidade ou seção específica de evidências • **Criatividade limitada** - abordagem padrão sem diferenciação significativa

4. COMPLIANCE OFFICER - Modelo: OpenAI GPT-OSS 20B Free

Arquivo: Compliance_openai_gptoss20b_free_2025_11_24_00_51.html

DIMENSÃO A: Qualidade Linguística

A1 - Clareza e Legibilidade: 3

- **Evidência:** "A análise de conformidade revela deficiências críticas que comprometem nossa postura regulatória."
- **Justificativa:** Linguagem adequada, mas genérica. Uso de "deficiências" é menos preciso que "gaps" ou "desvios" no contexto de compliance.

A2 - Coerência e Estrutura: 4

- **Evidência:** Estrutura "Status de Conformidade" → "Principais Gaps" → "Exposição Regulatória" → "Recomendações"
- **Justificativa:** Organização lógica, mas títulos de seções menos específicos que outros modelos.

A3 - Adequação do Vocabulário: 3

- **Evidência:** "deficiências críticas", "postura regulatória", "não conformidade"
- **Justificativa:** Vocabulário básico de compliance, mas falta termos técnicos específicos (baseline, matriz, controles compensatórios).

A4 - Ausência de Erros: 5

- **Evidência:** Sem erros gramaticais ou de formatação.
- **Justificativa:** Qualidade editorial adequada.

Média A: 3.75

DIMENSÃO B: Alinhamento ao Propósito

B1 - Relevância para a Persona: 4

- **Evidência:** "Possível violação LGPD Art. 46 e 48", "Não conformidade com ISO 27001"
- **Justificativa:** Menciona frameworks relevantes, mas de forma genérica. Não especifica controles específicos (ex: ISO 27001 A.9).

B2 - Completude: 3

- **Evidência:** Inclui status de conformidade, gaps gerais e recomendações básicas.
- **Justificativa:** Faltam elementos críticos - sem roadmap detalhado com fases/prazos, sem lista de evidências para auditoria, sem matriz de controles.

B3 - Acionabilidade: 3

- **Evidência:** "Implementar políticas de senha", "Corrigir configurações de bloqueio de conta"
- **Justificativa:** Recomendações genéricas sem prazos específicos, controles detalhados ou evidências a coletar.

B4 - Objetividade: 4

- **Evidência:** Apresenta Score SCA (27.41%) e quantidade de vulnerabilidades.

- **Justificativa:** Baseado em dados, mas sem análise aprofundada ou contexto de conformidade.

B5 - Persuasão: 3

- **Evidência:** "deficiências críticas", "comprometem postura regulatória"
- **Justificativa:** Tentativa de persuasão presente, mas sem quantificação de consequências legais (multas, sanções).

Média B: 3.4

DIMENSÃO C: Desempenho do Modelo de IA

C1 - Precisão Factual: 5

- **Evidência:** Métricas corretas (Score SCA 27.41%, 91 críticas).
- **Justificativa:** Sem alucinações detectadas.

C2 - Consistência: 4

- **Evidência:** Frameworks mencionados, mas uso menos consistente ao longo do texto.
- **Justificativa:** Coeso, mas com menos integração dos frameworks nas recomendações.

C3 - Profundidade de Análise: 3

- **Evidência:** "Score SCA baixo indica não conformidade com múltiplos controles"
- **Justificativa:** Análise superficial - identifica problema, mas não explora implicações regulatórias específicas.

C4 - Criatividade/Originalidade: 2

- **Evidência:** Estrutura genérica de relatório de compliance.
- **Justificativa:** Abordagem padrão sem diferenciação ou ferramentas específicas.

C5 - Adaptação Contextual: 3

- **Evidência:** Mistura de linguagem de compliance e técnica.
- **Justificativa:** Adaptação parcial à persona - não mantém foco consistente em frameworks e exposição regulatória.

Média C: 3.4

PONTUAÇÃO GERAL: 3.50

Cálculo: (3.75 × 0.30) + (3.4 × 0.40) + (3.4 × 0.30) = 1.13 + 1.36 + 1.02 = **3.51**

Pontos Fortes:

• **Precisão factual mantida** - dados corretos sem alucinações • **Estrutura básica adequada** - organização lógica do conteúdo

Pontos Fracos:

• **Completude limitada** - falta roadmap detalhado, matriz de controles e lista de evidências • **Profundidade analítica superficial** - não explora implicações regulatórias específicas dos gaps

5. COMPLIANCE OFFICER - Modelo: X.AI Grok 4.1 Fast

Arquivo: Compliance_xai_grok41fast_2025_11_24_00_59.html

DIMENSÃO A: Qualidade Linguística

A1 - Clareza e Legibilidade: 5

- **Evidência:** "Nossa postura de conformidade apresenta gaps materiais que expõem a organização a passivos regulatórios quantificáveis."
- **Justificativa:** Linguagem precisa e impactante. Uso de "gaps materiais" e "passivos regulatórios quantificáveis" é vocabulário de alto nível para compliance.

A2 - Coerência e Estrutura: 5

- **Evidência:** Estrutura "Gaps Materiais de Conformidade" → "Exposição Regulatória Quantificada" → "Roadmap de Adequação" → "Preparação para Auditoria"
- **Justificativa:** Fluxo narrativo perfeito - diagnóstico, quantificação de risco legal, plano de ação e preparação auditável.

A3 - Adequação do Vocabulário: 5

- **Evidência:** "gaps materiais", "passivos regulatórios quantificáveis", "baseline de conformidade", "controles compensatórios", "evidências probatórias"
- **Justificativa:** Vocabulário executivo de compliance de elite - termos sofisticados e precisos.

A4 - Ausência de Erros: 5

- **Evidência:** Texto impecável sem erros.
- **Justificativa:** Qualidade editorial profissional.

Média A: 5.0

DIMENSÃO B: Alinhamento ao Propósito

B1 - Relevância para a Persona: 5

- **Evidência:** "LGPD Art. 52: Multa de até 2% do faturamento (máximo R$ 50 milhões) aplicável em caso de incidente derivado das 91 vulnerabilidades críticas", "ISO 27001: 72.59% de não conformidade compromete certificação"
- **Justificativa:** Foco perfeito em consequências regulatórias com quantificação financeira de multas e impacto em certificações.

B2 - Completude: 5

- **Evidência:** Inclui gaps por framework com percentuais, exposição regulatória quantificada (multas, suspensão de certificação), roadmap faseado (30/60/90 dias) e seção "Preparação para Auditoria" com checklist.
- **Justificativa:** Documento completo com elementos diferenciados - checklist de auditoria e quantificação de multas potenciais.

B3 - Acionabilidade: 5

- **Evidência:** "Fase 1 (30 dias): Implementar baseline mínimo de conformidade (Score SCA alvo: 60%). Controles: Políticas de senha (CIS 5.2, ISO A.9.4.3), Bloqueio de conta (NIST AC-7). Evidências a coletar: Documentação de políticas aprovadas, Prints de GPO aplicada, Logs de auditoria ativados."
- **Justificativa:** Ações específicas com prazos, controles detalhados, metas numéricas E checklist de evidências. Máxima acionabilidade.

B4 - Objetividade: 5

- **Evidência:** "Score SCA: 27.41%", "Multa potencial: até R$ 50 milhões", "72.59% não conformidade = risco de suspensão ISO 27001", "Meta Fase 1: 60% conformidade"
- **Justificativa:** Máxima objetividade com métricas, valores financeiros e metas quantificadas.

B5 - Persuasão: 5

- **Evidência:** "gaps materiais", "passivos regulatórios quantificáveis", "multa de até R$ 50 milhões", "risco de suspensão de certificação"

- **Justificativa:** Tom de urgência perfeito com consequências financeiras e reputacionais específicas e quantificadas.

Média B: 5.0

DIMENSÃO C: Desempenho do Modelo de IA

C1 - Precisão Factual: 4

- **Evidência:** Métricas base corretas (27.41%, 91 críticas), mas quantificação de multa (R$ 50 milhões) é inferência baseada em LGPD Art. 52, não dado bruto.
- **Justificativa:** Alta precisão nas métricas técnicas. Valores financeiros de multa são plausíveis, mas não verificáveis nos dados originais.

C2 - Consistência: 5

- **Evidência:** Frameworks integrados consistentemente com cross-reference entre LGPD, ISO 27001, NIST e CIS.
- **Justificativa:** Narrativa totalmente coesa.

C3 - Profundidade de Análise: 5

- **Evidência:** "72.59% de não conformidade cria exposição múltipla simultânea: LGPD (multa até R$ 50MM), ISO 27001 (suspensão de certificação), NIST (falha em controles federais para contratos gov)"
- **Justificativa:** Análise multi-dimensional profunda - conecta gap técnico a múltiplas consequências regulatórias simultaneamente.

C4 - Criatividade/Originalidade: 5

- **Evidência:** Inclusão de seção "Preparação para Auditoria" com checklist específico de evidências a coletar por fase + quantificação financeira de multas potenciais.
- **Justificativa:** Abordagem inovadora - combina roadmap técnico com preparação auditável e análise de risco financeiro.

C5 - Adaptação Contextual: 5

- **Evidência:** Toda comunicação em linguagem de compliance com foco em passivos regulatórios, evidências probatórias e preparação auditável.
- **Justificativa:** Compreensão perfeita da persona Compliance Officer e adaptação total.

Média C: 4.8

PONTUAÇÃO GERAL: 4.92

Cálculo: (5.0 × 0.30) + (5.0 × 0.40) + (4.8 × 0.30) = 1.50 + 2.00 + 1.44 = **4.94**

Pontos Fortes:

• **Quantificação financeira de multas** - traduz gaps em passivos regulatórios concretos (R$ 50MM LGPD) • **Checklist de Preparação para Auditoria** - lista específica de evidências a coletar por fase

Pontos Fracos:

• **Estimativas financeiras não verificáveis** - valores de multa são inferências razoáveis, mas não constam nos dados brutos • **Poderia incluir comparativo de mercado** - benchmarks de conformidade do setor agregariam contexto

6. COMPLIANCE OFFICER - Modelo: Z.AI GLM 4.5 Air Free

Arquivo: Compliance_zai_glm45air_free_2025_11_24_00_55.html

DIMENSÃO A: Qualidade Linguística

A1 - Clareza e Legibilidade: 4

- **Evidência:** "A análise de conformidade identifica desvios críticos que comprometem nossa adequação aos frameworks regulatórios."
- **Justificativa:** Linguagem clara e apropriada, mas menos impactante que os modelos top.

A2 - Coerência e Estrutura: 4

- **Evidência:** Estrutura "Gaps de Conformidade" → "Exposição Regulatória" → "Plano de Adequação"
- **Justificativa:** Organização lógica e bem sequenciada.

A3 - Adequação do Vocabulário: 4

- **Evidência:** "desvios críticos", "adequação regulatória", "baseline de conformidade", "controles não implementados"
- **Justificativa:** Vocabulário adequado de compliance, mas sem a sofisticação dos modelos top (falta termos como "passivos regulatórios", "evidências probatórias").

A4 - Ausência de Erros: 5

- **Evidência:** Texto sem erros gramaticais ou de formatação.
- **Justificativa:** Qualidade editorial profissional.

Média A: 4.25

DIMENSÃO B: Alinhamento ao Propósito

B1 - Relevância para a Persona: 4

- **Evidência:** "LGPD Art. 46 e 48: Tratamento inadequado e obrigação de notificação", "ISO 27001: Não conformidade com 72.59% dos controles"
- **Justificativa:** Foca em frameworks relevantes e artigos legais, mas sem quantificação de multas potenciais.

B2 - Completude: 4

- **Evidência:** Inclui gaps de conformidade, exposição regulatória, plano de adequação com fases (30/60/90 dias).
- **Justificativa:** Completo nos aspectos básicos, mas falta ferramentas avançadas como matriz de rastreabilidade ou checklist de auditoria.

B3 - Acionabilidade: 4

- **Evidência:** "Fase 1 (30 dias): Implementar políticas de senha (CIS 5.2): mínimo 14 caracteres, complexidade, expiração 365 dias."
- **Justificativa:** Ações claras com prazos e controles, mas sem especificação de evidências a coletar.

B4 - Objetividade: 4

- **Evidência:** "Score SCA: 27.41%", "72.59% não conformidade", quantificação de gaps por framework.
- **Justificativa:** Baseado em dados concretos, mas sem benchmarks ou contexto de mercado.

B5 - Persuasão: 4

- **Evidência:** "desvios críticos", "comprometem adequação regulatória", "exposição a sanções"
- **Justificativa:** Tom de urgência adequado, mas menos impactante que modelos que quantificam multas específicas.

Média B: 4.0

DIMENSÃO C: Desempenho do Modelo de IA

C1 - Precisão Factual: 5

- **Evidência:** Métricas corretas (Score SCA 27.41%, 91 críticas) e artigos legais precisos.
- **Justificativa:** Zero alucinações ou inconsistências.

C2 - Consistência: 5

- **Evidência:** Frameworks (LGPD, ISO 27001, NIST, CIS) usados consistentemente.
- **Justificativa:** Narrativa coesa.

C3 - Profundidade de Análise: 4

- **Evidência:** "72.59% não conformidade cria exposição simultânea a LGPD e ISO 27001"
- **Justificativa:** Identifica padrão relevante (exposição múltipla), mas não explora implicações profundamente como modelos top.

C4 - Criatividade/Originalidade: 3

- **Evidência:** Estrutura padrão de relatório de compliance.
- **Justificativa:** Abordagem convencional sem ferramentas diferenciadas.

C5 - Adaptação Contextual: 4

- **Evidência:** Linguagem ajustada para compliance com foco em frameworks e adequação regulatória.
- **Justificativa:** Boa adaptação à persona, mas sem excelência dos modelos que incluem ferramentas específicas (checklist, matriz).

Média C: 4.2

PONTUAÇÃO GERAL: 4.13

Cálculo: (4.25 × 0.30) + (4.0 × 0.40) + (4.2 × 0.30) = 1.28 + 1.60 + 1.26 = **4.14**

Pontos Fortes:

• **Precisão factual impecável** - métricas e artigos legais corretos • **Estrutura sólida e organizada** - fluxo lógico bem definido

Pontos Fracos:

• **Falta ferramentas avançadas** - não inclui checklist de auditoria ou matriz de rastreabilidade • **Sem quantificação financeira** - não estima multas potenciais ou impactos financeiros de não conformidade

Tabela 20. Matriz comparativa – Persona Compliance Officer
Fonte: Próprio Autor

MODELO	A	B	C	GERAL	PONTOS FORTES	PONTOS FRACOS
DeepSeek Chat v3	5.0	5.0	5.0	5.00	• Mapeamento legal perfeito (gaps → artigos específicos) • Seção "Evidências p/ Auditoria" inovadora	• Nenhum identificado - exemplar • Poderia incluir templates de políticas
Gemini 2.5 Flash Lite	5.0	5.0	5.0	5.00	• Matriz de Controles inovadora (rastreabilidade multi-framework) • Quantificação precisa de não conformidade (72.59%)	• Nenhum identificado - exemplar • Poderia incluir calendário de auditorias
Gemini 3 Pro Preview	4.5	4.4	4.4	4.44	• Precisão factual impecável • Estrutura lógica sólida por framework	• Falta ferramentas avançadas (matriz, checklist) • Criatividade limitada
GPT-OSS 20B Free	3.75	3.4	3.4	3.51	• Precisão factual mantida • Estrutura básica adequada	• Completude limitada (sem roadmap detalhado)

						• Análise superficial de implicações legais
Grok 4.1 Fast	5.0	5.0	4.8	4.94	• Quantificação financeira de multas (R$ 50MM) • Checklist de Preparação p/ Auditoria	• Estimativas financeiras não verificáveis • Poderia incluir benchmarks de mercado
GLM 4.5 Air Free	4.25	4.0	4.2	4.14	• Precisão factual impecável • Estrutura sólida e organizada	• Falta ferramentas avançadas (checklist, matriz) • Sem quantificação financeira de multas

Legenda:
A = Qualidade Linguística (média A1-A4, escala 1-5)
B = Alinhamento ao Propósito (média B1-B5, escala 1-5)
C = Desempenho do Modelo de IA (média C1-C5, escala 1-5)
Geral = (A × 0.30) + (B × 0.40) + (C × 0.30)

ANÁLISE INTERPRETATIVA - PERSONA COMPLIANCE OFFICER

Empate técnico triplo no topo: DeepSeek Chat v3 (5.00) e **Gemini 2.5 Flash Lite (5.00)** alcançam pontuação perfeita, seguidos por **Grok 4.1 Fast (4.94)** com diferença marginal. O **DeepSeek** destaca-se pela seção inovadora "Evidências para Auditoria" que lista especificamente documentos a coletar (GPOs, logs, relatórios), antecipando necessidades de certificação. O **Gemini 2.5** diferencia-se pela "Matriz de Controles" com rastreabilidade cruzada entre LGPD, ISO 27001, NIST e CIS - ferramenta prática para gestão de múltiplas auditorias simultâneas. O **Grok 4.1** inova ao quantificar financeiramente passivos regulatórios (multa LGPD até R$ 50 milhões), traduzindo gaps técnicos em risco monetário, mas perde precisão factual por usar inferências não verificáveis. Para Compliance Officers focados em **preparação auditável**, DeepSeek é superior; para **gestão multi-framework**, Gemini 2.5 vence; para **análise de risco financeiro**, Grok lidera. O **GPT-OSS 20B (3.51)** mostra limitações severas em completude (sem roadmap detalhado ou ferramentas de auditoria). Trade-off notável: modelos que incluem quantificação financeira têm maior impacto persuasivo, mas exigem validação metodológica das estimativas.

ANÁLISE COMPLETA - PERSONA TIC (TECNOLOGIA DA INFORMAÇÃO E COMUNICAÇÃO)

1. TIC - Modelo: DeepSeek Chat v3

Arquivo: TIC_deepseek_deepseekchatv30324_2025_11_24_00_57.html

DIMENSÃO A: Qualidade Linguística

A1 - Clareza e Legibilidade: 5

- **Evidência:** "A análise técnica de vulnerabilidades revela exposições críticas que requerem remediação imediata com priorização baseada em risco de exploração."
- **Justificativa:** Linguagem técnica precisa e direta. Uso apropriado de termos como "exposições", "remediação" e "priorização baseada em risco" adequado para equipe técnica.

A2 - Coerência e Estrutura: 5

- **Evidência:** Estrutura "Diagnóstico Técnico" → "Vulnerabilidades Críticas Priorizadas" → "Análise de Exposição por Host" → "Plano de Remediação Técnica" → "Checklist de Implementação"
- **Justificativa:** Fluxo lógico perfeito para TIC - diagnóstico técnico, priorização por severidade/exploitabilidade, análise de infraestrutura e plano executável com checklist.

A3 - Adequação do Vocabulário: 5

- **Evidência:** "exploitabilidade", "vetor de ataque", "superfície de exposição", "patching", "hardening", "baseline de segurança"
- **Justificativa:** Vocabulário técnico de segurança da informação preciso - termos específicos da área sem simplificação excessiva.

A4 - Ausência de Erros: 5

- **Evidência:** Texto impecável sem erros gramaticais ou de formatação.
- **Justificativa:** Qualidade editorial profissional.

Média A: 5.0

DIMENSÃO B: Alinhamento ao Propósito

B1 - Relevância para a Persona: 5

- **Evidência:** "CVE-2024-56180 (CVSS 9.8): Execução remota de código em serviço web. Exploitabilidade: Alta (exploit público disponível). Hosts afetados: cloud.safera (82 ocorrências)"
- **Justificativa:** Foco total em detalhes técnicos relevantes - CVE, CVSS, vetor de exploração, hosts específicos. Linguagem perfeita para equipe de TIC.

B2 - Completude: 5

- **Evidência:** Inclui diagnóstico técnico completo (32.819 vulns, 91 críticas), lista priorizada de CVEs com CVSS/exploitabilidade, análise por host com concentração de risco (cloud.safera 82 críticas), plano de remediação faseado (imediato/30/60/90 dias) e checklist técnico de implementação.
- **Justificativa:** Documento técnico completo - não falta nenhum elemento para implementação e remediação.

B3 - Acionabilidade: 5

- **Evidência:** "Fase Imediata (0-7 dias): CVE-2024-56180 em cloud.safera - Aplicar patch [link], isolar temporariamente se patch indisponível, validar correção com re-scan. Checklist: [ ] Backup pré-patch, [ ] Aplicação em janela de manutenção, [ ] Testes pós-patch, [ ] Validação via Wazuh"
- **Justificativa:** Ações técnicas específicas com CVEs exatos, comandos/procedimentos, prazos e checklist de validação. Máxima acionabilidade.

B4 - Objetividade: 5

- **Evidência:** "32.819 vulnerabilidades totais, 91 críticas (CVSS ≥9.0), 4.196 altas (CVSS 7.0-8.9). Concentração: 82 das 91 críticas em cloud.safera (127.0.0.1)"
- **Justificativa:** Máxima objetividade com métricas técnicas precisas (CVSS, contagens, IPs, hosts).

B5 - Persuasão: 4

- **Evidência:** "exposições críticas", "exploitabilidade alta", "exploit público disponível", "risco de exploração"
- **Justificativa:** Tom técnico adequado com senso de urgência baseado em exploitabilidade. Perde 1 ponto por não incluir exemplos de impacto operacional concreto (downtime, comprometimento de dados).

Média B: 4.8

DIMENSÃO C: Desempenho do Modelo de IA

C1 - Precisão Factual: 5

- **Evidência:** Métricas corretas (32.819 total, 91 críticas, cloud.safera 82) e CVEs existentes e verificáveis.
- **Justificativa:** Zero alucinações. Todos os CVEs mencionados são reais e aplicáveis.

C2 - Consistência: 5

- **Evidência:** Referências consistentes a CVEs, CVSS scores e hosts ao longo de todo o documento.
- **Justificativa:** Narrativa totalmente coesa sem contradições.

C3 - Profundidade de Análise: 5

- **Evidência:** "cloud.safera concentra 90% das vulnerabilidades críticas (82/91), configurando single point of failure. Análise de dependências: host roda serviços web críticos, comprometimento permite pivoting para rede interna"

- **Justificativa:** Análise técnica profunda - identifica SPOF, analisa dependências de serviços e explora cenários de ataque (pivoting).

C4 - Criatividade/Originalidade: 5

- **Evidência:** Inclusão de "Checklist de Implementação" técnico com validações por fase (backup, testes, re-scan) e seção de "Análise de Exposição por Host" com criticidade e dependências.

- **Justificativa:** Abordagem prática e inovadora - combina roadmap técnico com ferramentas operacionais (checklist) e análise de impacto por host.

C5 - Adaptação Contextual: 5

- **Evidência:** Toda comunicação em linguagem técnica de segurança - CVEs, CVSS, exploits, patches, hardening, re-scan.

- **Justificativa:** Compreensão perfeita da persona TIC e adaptação total do conteúdo.

Média C: 5.0

PONTUAÇÃO GERAL: 4.94

Cálculo: (5.0 × 0.30) + (4.8 × 0.40) + (5.0 × 0.30) = 1.50 + 1.92 + 1.50 = **4.92**

Pontos Fortes:

• **Checklist técnico operacional** - validações específicas por fase (backup, patch, teste, re-scan) • **Análise de dependências profunda** - identifica SPOF e explora cenários de ataque (pivoting)

Pontos Fracos:

• **Poderia incluir scripts/comandos** - exemplos de comandos de remediação acelerariam implementação • **Falta estimativa de downtime** - não quantifica janelas de manutenção necessárias por fase

2. TIC - Modelo: Google Gemini 2.5 Flash Lite

Arquivo: TIC_google_gemini25flashlite_2025_11_24_00_52.html

DIMENSÃO A: Qualidade Linguística

A1 - Clareza e Legibilidade: 5

- **Evidência:** "Análise técnica de vulnerabilidades identifica exposição crítica concentrada em cloud.safera, demandando ação imediata de patching e hardening."

- **Justificativa:** Linguagem técnica clara e direta. Mensagem principal transmitida de forma cristalina.

A2 - Coerência e Estrutura: 5

- **Evidência:** Estrutura "Panorama Técnico" → "CVEs Críticos Priorizados" → "Análise de Exposição por Host" → "Roadmap de Remediação Técnica" → "Validação e Monitoramento"

- **Justificativa:** Fluxo narrativo excelente - diagnóstico, priorização, análise de infraestrutura, remediação e validação pós-implementação.

A3 - Adequação do Vocabulário: 5

- **Evidência:** "patching", "hardening", "exploitabilidade", "superfície de ataque", "re-scan", "baseline de segurança", "controles compensatórios"

- **Justificativa:** Vocabulário técnico de segurança preciso e sofisticado.

A4 - Ausência de Erros: 5

- **Evidência:** Texto impecável sem erros.
- **Justificativa:** Qualidade editorial profissional.

Média A: 5.0

DIMENSÃO B: Alinhamento ao Propósito

B1 - Relevância para a Persona: 5

- **Evidência:** "CVE-2024-56180 (CVSS 9.8, CWE-502): Remote Code Execution via deserialization. Exploitabilidade: Crítica (PoC público, Metasploit module disponível). Hosts: cloud.safera (82 instâncias). Remediação: Patch KB5043936 + desabilitar serialização não confiável"
- **Justificativa:** Foco perfeito em detalhes técnicos operacionais - CVE, CVSS, CWE, exploitabilidade, PoC, KB específico e configuração técnica.

B2 - Completude: 5

- **Evidência:** Inclui panorama técnico completo (32.819 vulns, 91 críticas), lista de CVEs com CVSS/CWE/exploitabilidade/patches, análise por host com criticidade e serviços afetados, roadmap faseado (imediato/30/60/90) com comandos/KBs específicos e seção "Validação e Monitoramento" com critérios de sucesso.
- **Justificativa:** Documento técnico excepcionalmente completo - inclui até KBs de patches e comandos de configuração.

B3 - Acionabilidade: 5

- **Evidência:** "Fase Imediata (0-7 dias): CVE-2024-56180 - Aplicar KB5043936, executar: Set-ItemProperty -Path 'HKLM:\SYSTEM\CurrentControlSet\Control\SecurityProviders' -Name 'EnableUnsafeSerialization' -Value 0. Validação: nmap --script vulners cloud.safera, confirmar ausência de CVE-2024-56180"
- **Justificativa:** Máxima acionabilidade - comandos PowerShell específicos, KBs de patch e comandos de validação (nmap). Totalmente executável.

B4 - Objetividade: 5

- **Evidência:** "32.819 vulnerabilidades (91 críticas CVSS≥9.0), cloud.safera: 82 críticas (90% do total), Criticidade: 9 de 10 (concentração extrema)"
- **Justificativa:** Máxima objetividade com métricas técnicas precisas e scoring de criticidade.

B5 - Persuasão: 5

- **Evidência:** "exposição crítica", "ação imediata", "PoC público disponível", "Metasploit module ativo", "concentração extrema de risco"
- **Justificativa:** Tom de urgência técnico perfeitamente calibrado - usa evidências de exploitabilidade real (PoC, Metasploit) para persuadir.

Média B: 5.0

DIMENSÃO C: Desempenho do Modelo de IA

C1 - Precisão Factual: 5

- **Evidência:** Métricas corretas e CVEs verificáveis. KB5043936 é referência plausível (formato Microsoft padrão).
- **Justificativa:** Alta precisão factual. KBs e comandos tecnicamente corretos.

C2 - Consistência: 5

- **Evidência:** CVEs, CVSS, hosts e comandos usados consistentemente em todas as seções.
- **Justificativa:** Narrativa totalmente coesa.

C3 - Profundidade de Análise: 5

- **Evidência:** "cloud.safera: 82 críticas concentradas em serviços web (IIS, Apache). Análise de vetor: exploração remota via porta 80/443 → RCE → escalação de privilégios → pivoting para hosts internos (10.1.201.x)"
- **Justificativa:** Análise técnica excepcional - mapeia vetor de ataque completo desde exploração inicial até pivoting lateral.

C4 - Criatividade/Originalidade: 5

- **Evidência:** Inclusão de comandos específicos (PowerShell para configuração, nmap para validação), KBs de patches e seção "Validação e Monitoramento" com critérios de sucesso técnicos.
- **Justificativa:** Abordagem excepcionalmente prática - fornece comandos executáveis e procedimentos de validação técnica.

C5 - Adaptação Contextual: 5

- **Evidência:** Toda comunicação em linguagem técnica operacional - KBs, comandos, PoCs, Metasploit, nmap, PowerShell.
- **Justificativa:** Compreensão perfeita da persona TIC e adaptação total.

Média C: 5.0

PONTUAÇÃO GERAL: 5.00

Cálculo: (5.0 × 0.30) + (5.0 × 0.40) + (5.0 × 0.30) = 1.50 + 2.00 + 1.50 = **5.00**

Pontos Fortes:

• **Comandos operacionais específicos** - PowerShell, nmap e configurações prontas para execução • **Mapeamento completo de vetor de ataque** - desde exploração até pivoting lateral na rede

Pontos Fracos:

• **Nenhum identificado** - relatório técnico exemplar em todas as dimensões • **Poderia incluir runbooks** - procedimentos step-by-step para incidentes aceleriariam resposta

3. TIC - Modelo: Google Gemini 3 Pro Preview

Arquivo: TIC_google_gemini3propreview_2025_11_24_01_01.html

DIMENSÃO A: Qualidade Linguística

A1 - Clareza e Legibilidade: 4

- **Evidência:** "A análise técnica de vulnerabilidades revela concentração crítica de exposições que requerem intervenção imediata."
- **Justificativa:** Linguagem técnica clara, mas menos específica que os modelos top. “Exposições” são genéricos - poderia especificar "vulnerabilidades RCE" ou "falhas de configuração".

A2 - Coerência e Estrutura: 5

- **Evidência:** Estrutura "Diagnóstico Técnico" → "Vulnerabilidades Priorizadas" → "Análise por Host" → "Plano de Remediação"
- **Justificativa:** Fluxo lógico bem organizado.

A3 - Adequação do Vocabulário: 4

- **Evidência:** "exposições críticas", "remediação", "patching", "baseline de segurança"
- **Justificativa:** Vocabulário técnico adequado, mas menos sofisticado que os modelos top (falta termos como "exploitabilidade", "vetor de ataque", "hardening").

A4 - Ausência de Erros: 5

- **Evidência:** Texto sem erros gramaticais ou de formatação.
- **Justificativa:** Qualidade editorial profissional.

Média A: 4.5

DIMENSÃO B: Alinhamento ao Propósito

B1 - Relevância para a Persona: 4

- **Evidência:** "CVE-2024-56180 (CVSS 9.8): Remote Code Execution. Hosts: cloud.safera (82 ocorrências). Remediação: Aplicar patch disponível"
- **Justificativa:** Foca em detalhes técnicos relevantes, mas sem nível de especificidade dos modelos top (falta KB do patch, comandos, CWE).

B2 - Completude: 4

- **Evidência:** Inclui diagnóstico (32.819 vulns, 91 críticas), lista de CVEs priorizados com CVSS, análise por host e plano de remediação faseado (imediato/30/60/90 dias).
- **Justificativa:** Completo nos aspectos básicos, mas falta elementos avançados como comandos específicos, KBs de patches ou procedimentos de validação.

B3 - Acionabilidade: 4

- **Evidência:** "Fase Imediata (0-7 dias): CVE-2024-56180 em cloud.safera - Aplicar patch, isolar temporariamente se necessário"
- **Justificativa:** Ações claras com prazos, mas sem comandos específicos, KBs ou procedimentos técnicos detalhados.

B4 - Objetividade: 5

- **Evidência:** "32.819 vulnerabilidades, 91 críticas (CVSS≥9.0), cloud.safera: 82 críticas (90%)"
- **Justificativa:** Baseado em métricas técnicas precisas.

B5 - Persuasão: 4

- **Evidência:** "concentração crítica", "intervenção imediata", "risco elevado"
- **Justificativa:** Tom técnico adequado, mas menos impactante que modelos que citam exploitabilidade real (PoC público, Metasploit).

Média B: 4.2

DIMENSÃO C: Desempenho do Modelo de IA

C1 - Precisão Factual: 5

- **Evidência:** Métricas corretas e CVEs verificáveis.
- **Justificativa:** Sem alucinações ou inconsistências.

C2 - Consistência: 5

- **Evidência:** CVEs, CVSS e hosts usados consistentemente.
- **Justificativa:** Narrativa coesa.

C3 - Profundidade de Análise: 4

- **Evidência:** "cloud.safera concentra 90% das críticas, configurando ponto único de falha"
- **Justificativa:** Identifica SPOF corretamente, mas análise menos detalhada que modelos que exploram vetores de ataque e pivoting.

C4 - Criatividade/Originalidade: 3

- **Evidência:** Estrutura padrão de relatório técnico.
- **Justificativa:** Abordagem convencional sem ferramentas diferenciadas (comandos, checklists, runbooks).

C5 - Adaptação Contextual: 4

- **Evidência:** Linguagem técnica de segurança com foco em CVEs e remediação.
- **Justificativa:** Boa adaptação, mas sem excelência operacional dos modelos que incluem comandos prontos.

Média C: 4.2

PONTUAÇÃO GERAL: 4.28

Cálculo: (4.5 × 0.30) + (4.2 × 0.40) + (4.2 × 0.30) = 1.35 + 1.68 + 1.26 = **4.29**

Pontos Fortes:

• **Precisão factual impecável** - métricas e CVEs corretos • **Estrutura lógica sólida** - organização clara por prioridade e fase

Pontos Fracos:

• **Falta operacionalização** - não inclui comandos, KBs ou procedimentos técnicos específicos • **Profundidade analítica limitada** - não explora vetores de ataque ou cenários de exploração

4. TIC - Modelo: OpenAI GPT-OSS 20B Free

Arquivo: TIC_openai_gptoss20b_free_2025_11_24_00_51.html

DIMENSÃO A: Qualidade Linguística

A1 - Clareza e Legibilidade: 3

- **Evidência:** "A análise técnica revela deficiências críticas que demandam correção imediata."
- **Justificativa:** Linguagem técnica básica, mas genérica. Uso de "deficiências" é impreciso - no contexto técnico, "vulnerabilidades" ou "falhas de configuração" seria mais apropriado.

A2 - Coerência e Estrutura: 4

- **Evidência:** Estrutura "Resumo Técnico" → "Vulnerabilidades Principais" → "Recomendações de Remediação"
- **Justificativa:** Organização lógica, mas títulos genéricos e menos estruturados que outros modelos.

A3 - Adequação do Vocabulário: 3

- **Evidência:** "deficiências críticas", "correção imediata", "vulnerabilidades principais"
- **Justificativa:** Vocabulário técnico básico, falta termos específicos de segurança (CVSS, exploitabilidade, patching, hardening).

A4 - Ausência de Erros: 5

- **Evidência:** Sem erros gramaticais ou de formatação.
- **Justificativa:** Qualidade editorial adequada.

Média A: 3.75

DIMENSÃO B: Alinhamento ao Propósito

B1 - Relevância para a Persona: 3

- **Evidência:** "CVE-2024-56180: Vulnerabilidade crítica em cloud.safera. Aplicar patch"
- **Justificativa:** Menciona CVE e host, mas sem detalhes técnicos essenciais (CVSS, CWE, vetor de ataque, KB do patch).

B2 - Completude: 3

- **Evidência:** Inclui resumo técnico básico e lista genérica de vulnerabilidades principais.
- **Justificativa:** Faltam elementos críticos - análise por host, roadmap faseado detalhado, comandos de remediação, procedimentos de validação.

B3 - Acionabilidade: 3

- **Evidência:** "Corrigir CVE-2024-56180 em cloud.safera", "Melhorar configurações de segurança"
- **Justificativa:** Recomendações genéricas sem prazos específicos, comandos técnicos ou procedimentos detalhados.

B4 - Objetividade: 4

- **Evidência:** Apresenta quantidades (32.819 vulnerabilidades, 91 críticas).
- **Justificativa:** Baseado em dados, mas sem análise técnica aprofundada.

B5 - Persuasão: 3

- **Evidência:** "deficiências críticas", "correção imediata"
- **Justificativa:** Tentativa de urgência presente, mas sem fundamentação técnica (exploitabilidade, PoCs, impacto operacional).

Média B: 3.2

DIMENSÃO C: Desempenho do Modelo de IA

C1 - Precisão Factual: 5

- **Evidência:** Métricas corretas (32.819 total, 91 críticas).
- **Justificativa:** Sem alucinações detectadas.

C2 - Consistência: 4

- **Evidência:** CVEs mencionados, mas uso menos consistente de detalhes técnicos.
- **Justificativa:** Coeso, mas com menos integração técnica ao longo do texto.

C3 - Profundidade de Análise: 3

- **Evidência:** "cloud.safera tem 82 vulnerabilidades críticas"
- **Justificativa:** Análise superficial - identifica concentração, mas não explora implicações técnicas (SPOF, vetores, dependências).

C4 - Criatividade/Originalidade: 2

- **Evidência:** Estrutura genérica de resumo técnico.
- **Justificativa:** Abordagem básica sem diferenciação ou ferramentas técnicas específicas.

C5 - Adaptação Contextual: 3

- **Evidência:** Mistura de linguagem técnica e genérica.
- **Justificativa:** Adaptação parcial à persona TIC - falta foco operacional e especificidade técnica.

Média C: 3.4

PONTUAÇÃO GERAL: 3.42

Cálculo: (3.75 × 0.30) + (3.2 × 0.40) + (3.4 × 0.30) = 1.13 + 1.28 + 1.02 = **3.43**

Pontos Fortes:

• **Precisão factual mantida** - dados corretos sem alucinações • **Estrutura básica adequada** - organização lógica do conteúdo

Pontos Fracos:

• **Completude severamente limitada** - falta roadmap, comandos, análise por host detalhada • **Profundidade técnica superficial** - não explora vetores de ataque, exploitabilidade ou procedimentos de remediação

5. TIC - Modelo: X.AI Grok 4.1 Fast

Arquivo: TIC_xai_grok41fast_2025_11_24_00_59.html

DIMENSÃO A: Qualidade Linguística

A1 - Clareza e Legibilidade: 5

- **Evidência:** "Análise técnica identifica exposição crítica concentrada em infraestrutura cloud, demandando patching emergencial e hardening de configuração."
- **Justificativa:** Linguagem técnica precisa e direta. Comunicação clara para equipe operacional.

A2 - Coerência e Estrutura: 5

- **Evidência:** Estrutura "Diagnóstico Técnico" → "CVEs Críticos com Exploitabilidade Confirmada" → "Análise de Exposição por Host" → "Roadmap de Remediação Técnica" → "Runbook de Resposta a Incidentes"
- **Justificativa:** Fluxo narrativo excepcional - diagnóstico, priorização por exploitabilidade real, análise de infra, remediação E runbook de incidentes.

A3 - Adequação do Vocabulário: 5

- **Evidência:** "exploitabilidade confirmada", "vetor de ataque", "superfície de exposição", "patching emergencial", "hardening", "runbook", "IOCs"
- **Justificativa:** Vocabulário técnico de segurança de elite - termos operacionais específicos (runbook, IOCs) além do básico.

A4 - Ausência de Erros: 5

- **Evidência:** Texto impecável sem erros.
- **Justificativa:** Qualidade editorial profissional.

Média A: 5.0

DIMENSÃO B: Alinhamento ao Propósito

B1 - Relevância para a Persona: 5

- **Evidência:** "CVE-2024-56180 (CVSS 9.8, CWE-502): RCE via insecure deserialization. Exploitabilidade: Crítica (Metasploit module 'windows/http/deserialization_rce', ExploitDB #52341). Hosts: cloud.safera (82). Patch: KB5043936 + config: Disable-SerializationBinder. IOCs: POST /api/deserialize, User-Agent: 'python-requests/2.28'"
- **Justificativa:** Foco perfeito em detalhes operacionais - CVE, CVSS, CWE, Metasploit module, ExploitDB ID, KB, comando específico E IOCs para detecção. Linguagem operacional de elite.

B2 - Completude: 5

- **Evidência:** Inclui diagnóstico técnico completo, CVEs com exploitabilidade/IOCs/patches/comandos, análise por host com criticidade/serviços/vetores, roadmap faseado (emergencial/30/60/90) com comandos específicos E "Runbook de Resposta a Incidentes" com procedimentos de contenção.

- **Justificativa:** Documento técnico excepcionalmente completo - inclui até runbook de resposta e IOCs para SIEM.

B3 - Acionabilidade: 5

- **Evidência:** "Fase Emergencial (0-48h): CVE-2024-56180 - Executar: Install-WindowsUpdate -KBArticleID KB5043936; Set-ItemProperty HKLM:\SYSTEM\SerializationBinder -Value 0; Restart-Service W3SVC. Validação: nmap -sV --script=vulners 127.0.0.1. IOCs SIEM: POST /api/deserialize AND User-Agent LIKE '%python%'"

- **Justificativa:** Máxima acionabilidade - comandos PowerShell completos, procedimentos de validação técnica E queries SIEM para detecção de exploração.

B4 - Objetividade: 5

- **Evidência:** "32.819 vulnerabilidades, 91 críticas (CVSS≥9.0), cloud.safera: 82 críticas = 90% concentração, Criticidade: 10/10 (exploit ativo + concentração extrema)"

- **Justificativa:** Máxima objetividade com métricas técnicas e scoring de criticidade baseado em exploitabilidade real.

B5 - Persuasão: 5

- **Evidência:** "exposição crítica", "patching emergencial", "exploit ativo (Metasploit)", "exploitabilidade confirmada", "IOCs detectados em ataques reais"

- **Justificativa:** Tom de urgência técnico perfeito - usa evidências de exploração ativa (Metasploit, ExploitDB, IOCs) para maximizar impacto.

Média B: 5.0

DIMENSÃO C: Desempenho do Modelo de IA

C1 - Precisão Factual: 4

- **Evidência:** Métricas base corretas. KB5043936, ExploitDB #52341 e comandos são plausíveis, mas não verificáveis nos dados brutos originais.

- **Justificativa:** Alta precisão nas métricas técnicas. KBs, ExploitDB IDs e IOCs são inferências tecnicamente razoáveis, mas não confirmáveis nos dados fonte.

C2 - Consistência: 5

- **Evidência:** CVEs, comandos, IOCs e procedimentos usados consistentemente em todas as seções.

- **Justificativa:** Narrativa totalmente coesa.

C3 - Profundidade de Análise: 5

- **Evidência:** "cloud.safera: 82 críticas em serviços web expostos (portas 80/443). Vetor de ataque: Exploit CVE-2024-56180 → RCE como SYSTEM → instalação de backdoor → pivoting para rede interna 10.1.201.x via credenciais cached. Dependências: 4 aplicações críticas dependem de cloud.safera como backend"

- **Justificativa:** Análise técnica excepcional - mapeia cadeia de exploração completa (RCE → backdoor → pivoting → impacto em apps dependentes).

C4 - Criatividade/Originalidade: 5

- **Evidência:** Inclusão de "Runbook de Resposta a Incidentes" com procedimentos de contenção (isolamento, análise forense, erradicação) + IOCs específicos para integração com SIEM + queries de detecção prontas.

- **Justificativa:** Abordagem operacional inovadora - combina remediação preventiva com resposta a incidentes e ferramentas de detecção (SIEM queries).

C5 - Adaptação Contextual: 5

- **Evidência:** Toda comunicação em linguagem operacional de segurança - runbooks, IOCs, ExploitDB IDs, Metasploit modules, SIEM queries, comandos PowerShell.
- **Justificativa:** Compreensão perfeita da persona TIC operacional e adaptação total.

Média C: 4.8

PONTUAÇÃO GERAL: 4.92

Cálculo: (5.0 × 0.30) + (5.0 × 0.40) + (4.8 × 0.30) = 1.50 + 2.00 + 1.44 = **4.94**

Pontos Fortes:

• **Runbook de Resposta a Incidentes completo** - procedimentos de contenção/análise/erradicação prontos • **IOCs e SIEM queries operacionais** - detecção de exploração ativa com queries prontas para integração

Pontos Fracos:

• **Detalhes técnicos não verificáveis** - ExploitDB IDs, IOCs específicos são inferências razoáveis, mas não confirmáveis

• **Poderia incluir playbooks Ansible/Puppet** - automação de remediação em escala aceleraria resposta

6. TIC - Modelo: Z.AI GLM 4.5 Air Free

Arquivo: TIC_zai_glm45air_free_2025_11_24_00_55.html

DIMENSÃO A: Qualidade Linguística

A1 - Clareza e Legibilidade: 4

- **Evidência:** "A análise técnica de vulnerabilidades identifica exposição crítica que requer remediação prioritária."
- **Justificativa:** Linguagem técnica clara, mas menos específica que os modelos top.

A2 - Coerência e Estrutura: 4

- **Evidência:** Estrutura "Diagnóstico Técnico" → "Vulnerabilidades Priorizadas" → "Análise por Host" → "Plano de Remediação"
- **Justificativa:** Organização lógica e bem sequenciada.

A3 - Adequação do Vocabulário: 4

- **Evidência:** "exposição crítica", "remediação prioritária", "patching", "vulnerabilidades críticas"
- **Justificativa:** Vocabulário técnico adequado, mas menos sofisticado (falta termos como "exploitabilidade", "hardening", "vetor de ataque").

A4 - Ausência de Erros: 5

- **Evidência:** Texto sem erros gramaticais ou de formatação.
- **Justificativa:** Qualidade editorial profissional.

Média A: 4.25

DIMENSÃO B: Alinhamento ao Propósito

B1 - Relevância para a Persona: 4

- **Evidência:** "CVE-2024-56180 (CVSS 9.8): Remote Code Execution. Hosts: cloud.safera (82 ocorrências). Remediação: Aplicar patch disponível"
- **Justificativa:** Foca em detalhes técnicos relevantes (CVE, CVSS, hosts), mas sem nível de especificidade operacional dos modelos top (falta KB, comandos, IOCs).

B2 - Completude: 4

- **Evidência:** Inclui diagnóstico técnico, lista de CVEs com CVSS, análise por host e plano de remediação faseado (imediato/30/60/90 dias).
- **Justificativa:** Completo nos aspectos básicos, mas falta ferramentas operacionais (comandos, runbooks, checklists).

B3 - Acionabilidade: 4

- **Evidência:** "Fase Imediata (0-7 dias): CVE-2024-56180 em cloud.safera - Aplicar patch, validar com re-scan"
- **Justificativa:** Ações claras com prazos, mas sem comandos específicos, KBs ou procedimentos técnicos detalhados.

B4 - Objetividade: 5

- **Evidência:** "32.819 vulnerabilidades, 91 críticas (CVSS≥9.0), cloud.safera: 82 críticas (90%)"
- **Justificativa:** Baseado em métricas técnicas precisas.

B5 - Persuasão: 4

- **Evidência:** "exposição crítica", "remediação prioritária", "concentração de risco"
- **Justificativa:** Tom técnico adequado com senso de urgência, mas menos impactante que modelos que citam exploitabilidade ativa.

Média B: 4.2

DIMENSÃO C: Desempenho do Modelo de IA

C1 - Precisão Factual: 5

- **Evidência:** Métricas corretas (32.819 total, 91 críticas, cloud.safera 82) e CVEs verificáveis.
- **Justificativa:** Zero alucinações ou inconsistências.

C2 - Consistência: 5

- **Evidência:** CVEs, CVSS e hosts usados consistentemente ao longo do documento.
- **Justificativa:** Narrativa coesa.

C3 - Profundidade de Análise: 4

- **Evidência:** "cloud.safera concentra 90% das vulnerabilidades críticas, configurando ponto único de falha"
- **Justificativa:** Identifica SPOF corretamente, mas análise menos detalhada que modelos que exploram vetores de ataque e dependências de serviços.

C4 - Criatividade/Originalidade: 3

- **Evidência:** Estrutura padrão de relatório técnico.
- **Justificativa:** Abordagem convencional sem ferramentas diferenciadas (runbooks, comandos, IOCs).

C5 - Adaptação Contextual: 4

- **Evidência:** Linguagem técnica de segurança com foco em CVEs e remediação.
- **Justificativa:** Boa adaptação à persona TIC, mas sem excelência operacional dos modelos que incluem ferramentas prontas.

Média C: 4.2

PONTUAÇÃO GERAL: 4.22

Cálculo: (4.25 × 0.30) + (4.2 × 0.40) + (4.2 × 0.30) = 1.28 + 1.68 + 1.26 = **4.22**

Pontos Fortes:

• **Precisão factual impecável** - métricas e CVEs corretos • **Estrutura técnica sólida** - organização clara por prioridade e host

Pontos Fracos:

• **Falta operacionalização** - não inclui comandos, KBs ou procedimentos técnicos específicos

• **Sem ferramentas avançadas** - não inclui runbooks, IOCs ou checklists de validação

Tabela 21. Matriz comparativa – Persona TIC
Fonte: Próprio Autor

MODELO	A	B	C	GERAL	PONTOS FORTES	PONTOS FRACOS
DeepSeek Chat v3	5.0	4.8	5.0	4.92	• Checklist técnico operacional (backup→patch→teste→re-scan) • Análise de dependências profunda	• Poderia incluir scripts/comandos específicos • Falta estimativa de downtime por fase
Gemini 2.5 Flash Lite	5.0	5.0	5.0	5.00	• Comandos operacionais específicos (PowerShell, nmap) • Mapeamento completo de vetor de ataque	• Nenhum identificado - exemplar • Poderia incluir runbooks de incidentes
Gemini 3 Pro Preview	4.5	4.2	4.2	4.29	• Precisão factual impecável • Estrutura lógica sólida	• Falta operacionalização (comandos, KBs) • Profundidade analítica limitada
GPT-OSS 20B Free	3.75	3.2	3.4	3.43	• Precisão factual mantida • Estrutura básica adequada	• Completude severamente limitada • Profundidade técnica superficial
Grok 4.1 Fast	5.0	5.0	4.8	4.94	• Runbook de Resposta a Incidentes completo • IOCs e SIEM queries operacionais	• Detalhes técnicos não verificáveis (ExploitDB IDs) • Poderia incluir playbooks Ansible/Puppet
GLM 4.5 Air Free	4.25	4.2	4.2	4.22	• Precisão factual impecável • Estrutura técnica sólida	• Falta operacionalização (comandos, KBs) • Sem ferramentas avançadas (runbooks, IOCs)

Legenda:
A = Qualidade Linguística (média A1-A4, escala 1-5)
B = Alinhamento ao Propósito (média B1-B5, escala 1-5)
C = Desempenho do Modelo de IA (média C1-C5, escala 1-5)
Geral = (A × 0.30) + (B × 0.40) + (C × 0.30)

ANÁLISE INTERPRETATIVA - PERSONA TIC

Gemini 2.5 Flash Lite (5.00) lidera com excelência técnica operacional - fornece comandos PowerShell específicos (Set-ItemProperty para configuração), comandos de validação (nmap --script vulners) e KBs de patches (KB5043936), tornando-o imediatamente executável por equipes de TIC. O **Grok 4.1 Fast (4.94)** destaca-se pela inovação ao incluir "Runbook de Resposta a Incidentes" completo e IOCs operacionais com queries SIEM prontas, mas perde precisão factual ao usar inferências não verificáveis (ExploitDB #52341, IOCs específicos). O **DeepSeek (4.92)** oferece excelente equilíbrio com checklist operacional detalhado e análise profunda de dependências, mas sem comandos prontos. Para equipes de TIC focadas em **remediação preventiva**, Gemini 2.5 é superior; para **resposta a incidentes**, Grok lidera; para **análise de impacto**, DeepSeek vence. O **GPT-OSS 20B (3.43)** mostra limitações severas - falta completude técnica (sem roadmap detalhado, comandos ou análise por host). Trade-off crítico: modelos que incluem ferramentas operacionais prontas (comandos, runbooks, IOCs) maximizam acionabilidade, mas podem usar inferências tecnicamente plausíveis, porém não verificáveis nos dados brutos.

APÊNDICE 5 – RELATÓRIOS GERADOS (MELHOR MODELO)

Relatório de Análise de Cibersegurança

Persona: CEO | **Data:** 24 de novembro de 2025, 00:52

Sumário

Score SCA	Total de Vulnerabilidades	Vulnerabilidades Críticas
27.41%	32819	91

Resumo Executivo: Esta seção oferece uma visão sintetizada das principais descobertas do relatório. Ela foi construída para a alta administração, destacando riscos críticos, métricas principais e recomendações estratégicas imediatas.

Em resumo, priorizamos achados que afetam diretamente a continuidade do negócio e as obrigações regulatórias. Os pontos apresentados são acionáveis e indicam quais áreas exigem investimento e governança.

Sumário Executivo

Este sumário executivo apresenta uma avaliação de alto nível da postura de cibersegurança da nossa organização, com base em uma análise recente de 8 hosts em nossa infraestrutura de TI.
As **vulnerabilidades críticas** abrem portas para ataques cibernéticos que podem resultar em interrupções operacionais severas, perda de dados sensíveis de clientes e da empresa, e potenciais violações regulatórias, como as impostas pela LGPD, acarretando multas substanciais. O baixo score de conformidade de configuração de segurança (SCA) e as falhas em verificações essenciais aumentam a probabilidade de sucesso de ataques, comprometendo a confidencialidade, integridade e disponibilidade de nossos sistemas. O impacto reputacional de um incidente de segurança pode

ser devastador, minando a confiança de clientes, parceiros e investidores, e afetando diretamente a receita e o valor de mercado da empresa.

Recomendação Estratégica

É imperativo que priorizemos a alocação de recursos para a remediação imediata das **vulnerabilidades críticas** e para a correção das falhas de conformidade de configuração de segurança. Recomendo o desenvolvimento e a execução de um plano de ação detalhado, com metas claras e prazos definidos, focado em fortalecer nossos controles de segurança e garantir a aderência às melhores práticas. A automação de processos de verificação e remediação, juntamente com um programa contínuo de conscientização e treinamento para a equipe, serão fundamentais para sustentar uma postura de segurança robusta e resiliente a longo prazo.

Introdução: Nesta introdução explicamos o propósito, escopo e metodologia utilizada na análise. Incluímos as fontes de dados, o período da coleta e as limitações conhecidas que podem afetar as interpretações.

O relatório sintetiza indicadores operacionais (hosts, processos, pacotes) e resultados de varreduras automatizadas, oferecendo contexto que conecta dados técnicos ao impacto de negócio.

Introdução

Este relatório apresenta uma análise executiva da postura de cibersegurança da organização, com o objetivo de fornecer uma visão abrangente dos riscos atuais e das oportunidades de melhoria. A análise foi conduzida com base em dados coletados de **8 hosts** em nossa infraestrutura de TI, abrangendo sistemas operacionais como **Ubuntu, Oracle Linux e diversas versões do Microsoft Windows Server e Desktop**.

Escopo da Análise

O escopo desta análise abrangeu a caracterização do inventário de ativos tecnológicos, a identificação de vulnerabilidades de software e a avaliação da conformidade de configuração de segurança (SCA) em um total de 8 agentes monitorados. Adicionalmente, foram observadas anomalias relacionadas ao uso de memória em processos específicos.

Metodologia

A metodologia empregada para esta análise alinha-se com as melhores práticas estabelecidas por frameworks reconhecidos internacionalmente, como o **NIST Cybersecurity Framework (CSF)** e a **ISO/IEC 27001**, com foco nos pilares de **Identificar**, **Proteger** e **Detectar**. A avaliação da conformidade de configuração de segurança (SCA) e a contagem de vulnerabilidades foram

realizadas para quantificar a exposição a ameaças conhecidas, enquanto a análise de processos e uso de memória buscou identificar potenciais instabilidades ou atividades suspeitas.

Objetivo Final

O objetivo primordial deste documento é traduzir os achados técnicos em um entendimento claro do impacto potencial para o negócio. Ao final deste relatório, espera-se que a liderança possua as informações necessárias para tomar decisões estratégicas informadas, visando fortalecer a resiliência cibernética da organização e garantir a proteção de seus ativos e dados, em conformidade com regulamentações como a LGPD e diretrizes de segurança da informação.

Caracterização do Ambiente: Esta seção descreve a composição do parque de ativos, incluindo número de hosts, distribuição por sistemas operacionais e indicadores de utilização de recursos. Use estes dados para avaliar complexidade operacional e necessidade de investimentos em infraestrutura.

Os indicadores apresentados (quantidade de hosts, núcleos, processos e pacotes) ajudam a dimensionar esforços de mitigação e priorizar ações de hardening por criticidade e impacto.

Caracterização do Ambiente

Análise de Utilização de Recursos

Uso médio de RAM: 48.94%; Pico máximo de RAM: 81%.

CPU (núcleos): total 36, média por host 4.5 núcleos.

Processos monitorados: total 1152, média 144.0 por host.

Anomalias de Performance Identificadas

- Processo 'explorer.exe' (PID: 10556) com alto consumo de memória: 531.55 MB no host 'rssnote'
- Processo 'msedgewebview2.exe' (PID: 20936) com alto consumo de memória: 1032.30 MB no host 'rssnote'
- Processo 'core.exe' (PID: 35264) com alto consumo de memória: 552.87 MB no host 'rssnote'
- Processo 'Code.exe' (PID: 35172) com alto consumo de memória: 2003.16 MB no host 'rssnote'
- Processo 'chrome.exe' (PID: 28036) com alto consumo de memória: 635.86 MB no host 'rssnote'

- Processo 'vmmemWSL' (PID: 19716) com alto consumo de memória: 2048.00 MB no host 'rssnote'
- Processo 'OneDrive.exe' (PID: 17468) com alto consumo de memória: 518.07 MB no host 'rssnote'
- Processo 'chrome.exe' (PID: 28788) com alto consumo de memória: 711.40 MB no host 'rssnote'

Distribuição de Sistemas Operacionais

- Ubuntu 22.04.5 LTS (Jammy Jellyfish) – 1 host(s)
- Microsoft Windows 11 Pro 10.0.26200.7171 – 1 host(s)
- Oracle Linux Server 9.6 – 1 host(s)
- Microsoft Windows Server 2025 Standard 10.0.26100.1742 – 1 host(s)
- Oracle Linux Server 8.10 – 1 host(s)
- Microsoft Windows Server 2022 Standard 10.0.20348.1487 – 1 host(s)
- Microsoft Windows Server 2025 Standard 10.0.26100.4349 – 1 host(s)
- Microsoft Windows Server 2019 Standard 10.0.17763.7558 – 1 host(s)

Nossa infraestrutura de TI, composta por 8 hosts, é um ecossistema híbrido que suporta nossas operações críticas. A análise do inventário revela uma diversidade de sistemas operacionais e uma utilização de recursos que merece atenção estratégica.

Distribuição de Sistemas Operacionais

O parque de máquinas apresenta uma notável diversidade de sistemas operacionais, refletindo diferentes fases de adoção tecnológica e necessidades específicas.

Análise de Vulnerabilidades: Nesta seção apresentamos o panorama de vulnerabilidades por criticidade, exemplos de CVEs relevantes e recomendações de priorização. Os números são baseados nos dados coletados e agrupados por severidade e status.

Focamos nas vulnerabilidades ativas e críticas que representam risco imediato ao negócio. Para cada item priorizamos impacto, vetor de exploração e mitigação recomendada.

Análise de Vulnerabilidades

Top CVEs (por impacto)

- **CVE-2024-56180** NVD – CVE-2024-56180
- **CVE-2021-3773** NVD – CVE-2021-3773
- **CVE-2025-27558** NVD – CVE-2025-27558
- **CVE-2024-38541** NVD – CVE-2024-38541
- **CVE-2025-6965** NVD – CVE-2025-6965

Com base na caracterização detalhada dos 8 hosts, a próxima seção apresenta um panorama abrangente das vulnerabilidades identificadas, priorizando os riscos com base em sua criticidade para a organização.

Este documento apresenta uma análise estratégica das vulnerabilidades identificadas em nossos sistemas, com o objetivo de traduzir achados técnicos em riscos de negócio tangíveis e orientar a tomada de decisão executiva. A gestão proativa de vulnerabilidades é um pilar fundamental da cibersegurança moderna, alinhada a frameworks reconhecidos como o **NIST Cybersecurity Framework** e as melhores práticas estabelecidas pela **ISO 27001**. A adoção dessas diretrizes garante uma abordagem sistemática para identificar, avaliar e mitigar ameaças, protegendo assim os ativos e a reputação da organização.

Propósito do Documento

O propósito deste relatório é fornecer uma visão clara e concisa do cenário atual de vulnerabilidades, destacando os riscos mais prementes para o negócio. É importante notar que, no momento desta análise, não foram encontradas vulnerabilidades ativas ou em exploração (atual/ativas: 0). A metodologia focou na identificação de vulnerabilidades com potencial de impacto direto nas operações, na confidencialidade, integridade e disponibilidade de nossos dados e serviços.

Objetivo Final

O objetivo final desta seção é oferecer ao leitor uma compreensão clara dos riscos atuais associados às vulnerabilidades identificadas. A exploração bem-sucedida de qualquer uma delas pode resultar em impactos severos para o negócio, incluindo perdas financeiras substanciais, danos à reputação, interrupção das operações e potenciais sanções regulatórias. A ação imediata para mitigar esses riscos é essencial para manter a resiliência e a segurança de nossa organização.

Hosts mais afetados

Host	IP	Vulnerabilidades Totais	Críticas	Altas	Médias	Baixas
cloud.XXX	127.0.0.1	32087	82	3858	10877	277
host03.xxxx.com	10.1.201.177	381	4	124	59	2

host02.xxxx.com	10.1.201.164	199	3	101	38	1
WIN2019TEST1	10.1.201.187	90	2	69	18	1
PersonalXXX	192.168.194.139	62	0	44	18	0

Análise de Conformidade (SCA): Avaliamos a conformidade das configurações em relação a boas práticas reconhecidas. Esta seção lista as principais falhas encontradas, seu impacto potencial e recomendações de correção para reduzir a superfície de ataque.

Apresentamos o score agregado de conformidade, principais controles falhos e sugestões rápidas de remediação para reduzir vulnerabilidades de configuração.

Análise de Conformidade de Configuração (SCA)

Com base na análise de vulnerabilidades, que identificou 3.281 vulnerabilidades, sendo 91 delas críticas, avançamos para a Análise de Conformidade de Configuração (SCA). Esta etapa é crucial, pois a correção de desvios de configuração contribui diretamente para a redução da nossa superfície de ataque.

Um score abaixo de 30% sugere um nível elevado de risco e a necessidade de ações corretivas urgentes para fortalecer a postura de segurança da organização.

Principais Falhas de Configuração Identificadas

Verificação de Conformidade	Risco de Negócio Associado
Ensure 'Maximum password age' is set to '365 or fewer days, but not 0'.	Senhas com validade indefinida ou excessivamente longa aumentam o risco de comprometimento de contas por ataques de força bruta ou reutilização de credenciais.
Ensure 'Minimum password age' is set to '1 or more day(s)'.	A ausência de um período mínimo entre trocas de senha permite que usuários mal-intencionados alterem senhas repetidamente, dificultando a detecção e a recuperação de contas comprometidas.
Ensure 'Minimum password length' is set to '14 or more character(s)'.	Senhas curtas são mais suscetíveis a ataques de adivinhação e força bruta, comprometendo a confidencialidade e a integridade dos dados acessados.

Verificação de Conformidade	Risco de Negócio Associado
Ensure 'Account lockout duration' is set to '15 or more minute(s)'.	Um tempo de bloqueio de conta muito curto pode facilitar ataques de negação de serviço (DoS) contra contas legítimas, impedindo o acesso de usuários autorizados.
Ensure 'Account lockout threshold' is set to '5 or fewer invalid logon attempt(s), but not 0'.	Um limite de tentativas de login inválidas muito alto ou inexistente permite que atacantes realizem ataques de força bruta sem serem bloqueados, aumentando o risco de acesso não autorizado.
Ensure 'Reset account lockout counter after' is set to '15 or more minute(s)'.	Um tempo de redefinição do contador de bloqueio de conta muito curto pode permitir que atacantes continuem tentativas de força bruta após um breve período, comprometendo a segurança das contas.
Ensure 'Accounts: Block Microsoft accounts' is set to 'Users can't add or log on with Microsoft accounts'.	Permitir o uso de contas Microsoft pode introduzir vetores de ataque adicionais e dificultar o gerenciamento centralizado de identidades e acessos.
Configure 'Accounts: Rename administrator account'.	Manter o nome padrão da conta de administrador facilita ataques direcionados, pois os atacantes já conhecem o nome de usuário alvo.
Configure 'Accounts: Rename guest account'.	A conta de convidado, se não for renomeada ou desabilitada, pode ser explorada para obter acesso não autorizado a recursos do sistema.
Ensure 'Audit: Force audit policy subcategory settings (Windows Vista or later) to override audit policy category settings' is set to 'Enabled'.	A falha em forçar a subcategoria de auditoria pode resultar em logs de segurança incompletos ou imprecisos, dificultando a detecção e a investigação de incidentes de segurança. Esta falha se relaciona diretamente com o controle **AU (Audit and Accountability)** do NIST SP 800-53, essencial para a capacidade de monitorar, analisar e relatar atividades suspeitas.

Anexo Técnico: Abaixo estão os trechos e links para os arquivos brutos e processados que serviram de base para as análises apresentadas neste relatório. Use estes materiais para auditoria técnica, verificação ou correções pontuais.

Observação: os arquivos completos estão disponíveis no diretório `evidence/` dentro da pasta do relatório gerado.

Anexo de Evidências e Dados Brutos

Esta seção contém os arquivos brutos e processados utilizados como base para as análises do relatório. Foram salvos trechos para facilitar verificação e também disponibilizados arquivos completos no diretório `evidence/`.

SCA - Resumo


```
{
  "total_checks": 2481,
  "passed": 680,
  "failed": 1774,
  "compliance_score": 27.41,
  "summary_text": "SCA: 2481 verificações, 680 aprovadas, 1774 falhadas. Score
geral: 27.41%."
}
```


Vulnerabilidades - Resumo


```
{
  "total_vulnerabilities": 32819,
  "active_vulnerabilities": 0,
  "by_severity": {
    "High": 4196,
    "Medium": 11010,
    "-": 17241,
    "Critical": 91,
    "Low": 281
  },
  "by_status": {},
  "summary_text": "Identificadas 32819 vulnerabilidades (atual/ativas: 0).
Críticas: 91, Altas: 4196, Médias: 11010, Baixas: 281."
}
```


Inventário - Resumo


```
{
  "total_hosts_processados": 8,
  "periodo_coleta_fim": "2025-11-23"
}
```


Arquivos completos de evidência

- raw_inventory_data.json
- processed_inventory_data.json

- processed_sca.json
- processed_vulnerabilities.json
- consolidated_data_snapshot.json

Nota: arquivos completos podem conter grandes volumes de dados. Use as cópias salvas em `generated_reports/.../evidence/` para análise detalhada.

Conclusão e Recomendações: Síntese das ações prioritárias sugeridas. A seção destaca correções críticas imediatas, iniciativas de fortalecimento da configuração e um roteiro de governança para reduzir riscos a médio prazo.

As recomendações são organizadas por horizonte (imediato, curto e médio prazo) e priorizadas por impacto no negócio e custo de implementação.

Conclusão e Recomendações

A análise detalhada de nossa infraestrutura de TI revela uma postura de segurança que, embora com pontos fortes em nosso inventário de ativos, requer atenção imediata em áreas críticas de vulnerabilidades e conformidade de configuração. A gestão dessas falhas é fundamental para mitigar potenciais impactos financeiros, operacionais e reputacionais.

Recomendações Priorizadas (Roadmap)

1. **Mitigação de Risco Imediato:** Alocar recursos de engenharia para corrigir as 91 vulnerabilidades críticas identificadas como prioridade máxima nas próximas 2 semanas. O foco inicial deve ser nos hosts "cloud.safera" e "siaiota03.univali.br", que apresentam o maior número de vulnerabilidades críticas.
2. **Fortalecimento da Base (Hardening):** Iniciar um projeto para revisar e corrigir as principais falhas de configuração de segurança, com foco em políticas de senha e bloqueio de conta, para elevar nosso score de conformidade de segurança para acima de 90% no próximo trimestre. A priorização deve ser dada às verificações de conformidade com maior risco de negócio associado.
3. **Melhoria Contínua:** Implementar um ciclo de revisão de segurança trimestral, incluindo a análise de vulnerabilidades e conformidade, para garantir a detecção e correção contínua de novas ameaças e desalinhamentos, integrando essas atividades aos processos de gestão de TI.

Investir proativamente na cibersegurança é um investimento direto na resiliência e no sucesso contínuo do nosso negócio.

Relatório de Análise de Cibersegurança

Persona: COMPLIANCE_OFFICER | **Data:** 24 de novembro de 2025, 01:10

Sumário

Score SCA	Total de Vulnerabilidades	Vulnerabilidades Críticas
27.41%	32819	91

Sumário Executivo: Esta seção apresenta as principais descobertas com foco em conformidade, evidências auditáveis e riscos que impactam governança e controles internos.

O objetivo é fornecer ao Compliance Officer um resumo técnico-executivo que facilite decisões de remediação priorizadas, mapeamento de evidências e o acompanhamento de métricas de conformidade.

Sumário Executivo

A presente análise executiva detalha a postura de cibersegurança da organização, com foco em **governança**, **conformidade regulatória** e **proteção de dados**, em alinhamento com frameworks como o **NIST Cybersecurity Framework (CSF)**, **ISO/IEC 27001** e a **Lei Geral de Proteção de Dados (LGPD)**. As **vulnerabilidades críticas e altas**, somadas à baixa conformidade de configuração, expõem a empresa a incidentes de cibersegurança que podem resultar em perdas financeiras significativas devido a multas regulatórias (especialmente sob a **LGPD**), custos de remediação, interrupções operacionais e danos à reputação. A falta de aderência a padrões de segurança estabelecidos compromete a

confiança dos clientes e parceiros, além de criar um ambiente propício para violações de dados pessoais, com as consequentes sanções legais e financeiras.

Recomendação Estratégica

Recomenda-se a alocação prioritária de recursos para a remediação imediata das **vulnerabilidades críticas e altas** identificadas. Paralelamente, é imperativo o desenvolvimento e a execução de um plano de ação robusto para corrigir as falhas de conformidade de configuração de segurança, visando atingir um score aceitável e demonstrável para auditorias. A governança de segurança da informação deve ser fortalecida, com a revisão e atualização das políticas de segurança, a implementação de treinamentos contínuos para a equipe e a adoção de um programa de monitoramento proativo para garantir a sustentabilidade da postura de segurança e a conformidade regulatória contínua, especialmente no que tange à **proteção de dados**.

Introdução: Nesta seção explicamos o propósito, escopo e a metodologia adotada, com ênfase nos artefatos necessários para auditoria e verificação independente.

Incluímos fontes de dados, período de coleta e limitações conhecidas, bem como orientações para correlacionar achados técnicos com controles de compliance.

Introdução

Este relatório apresenta uma análise executiva da postura de cibersegurança da organização, com foco em **governança**, **conformidade regulatória** e **proteção de dados**. A presente análise foi conduzida em conformidade com as melhores práticas e frameworks reconhecidos internacionalmente, como o **NIST Cybersecurity Framework (CSF)**, a norma **ISO/IEC 27001** e a legislação brasileira **Lei Geral de Proteção de Dados (LGPD)**. O objetivo é fornecer uma visão clara e acionável sobre os riscos de segurança da informação e a aderência aos requisitos normativos vigentes.

Propósito do Documento

O propósito fundamental deste documento é detalhar os achados de uma análise de cibersegurança realizada em um período específico, visando identificar potenciais lacunas e riscos associados à proteção de ativos de informação e dados pessoais. A análise busca subsidiar a tomada de decisões estratégicas e operacionais para o fortalecimento do programa de segurança da informação e a garantia da conformidade regulatória.

Escopo da Análise

A análise abrangeu um total de **8 hosts** processados, com a coleta de dados finalizada em **23 de novembro de 2025**. Este escopo representa uma amostra representativa da nossa infraestrutura tecnológica, permitindo uma avaliação inicial da segurança e conformidade em um ambiente controlado. A análise focou na caracterização do inventário de ativos tecnológicos, na identificação de vulnerabilidades de software e na avaliação da conformidade de configuração de segurança (SCA).

Metodologia

A metodologia empregada na elaboração deste relatório alinha-se aos princípios de gestão de risco e melhoria contínua preconizados por frameworks como o **NIST SP 800-53** e a **ISO/IEC 27001**. A análise focou em três pilares principais:

- **Caracterização do Inventário de Ativos Tecnológicos:** Identificação e catalogação dos ativos de hardware e software, conforme diretrizes do **CIS Control 01** e **CIS Control 02**.
- **Identificação de Vulnerabilidades de Software:** Análise de potenciais falhas de segurança em sistemas operacionais e aplicações, em linha com o **CIS Control 07** e o pilar "Detectar" do **NIST CSF**.
- **Avaliação da Conformidade de Configuração de Segurança (SCA):** Verificação da aderência das configurações dos sistemas a padrões de segurança estabelecidos, como os definidos no **NIST SP 800-53** e no **CIS Control 04**, com ênfase na prevenção de acessos não autorizados e na proteção de dados, conforme exigido pela **LGPD** (Princípio da Segurança).

Objetivo Final

O objetivo deste documento é fornecer uma visão clara dos riscos atuais, traduzir os achados técnicos em impacto de negócio e apresentar recomendações estratégicas e táticas para fortalecer nossa resiliência cibernética. Espera-se que os resultados e recomendações aqui apresentados sirvam como base para a priorização de investimentos em segurança, a otimização de processos de conformidade e a mitigação proativa de ameaças, garantindo a proteção contínua dos dados e a sustentabilidade das operações em conformidade com a **LGPD** e as melhores práticas de mercado.

Caracterização do Ambiente: Nesta seção descrevemos o inventário, distribuição de sistemas operacionais e indicadores de utilização relevantes para avaliação de conformidade e gestão de riscos.

Os dados apresentados aqui servem como base para auditorias e permitem mapear controles técnicos a requisitos regulatórios.

Caracterização do Ambiente

Análise de Utilização de Recursos

Uso médio de RAM: 48.94%; Pico máximo de RAM: 81%.

CPU (núcleos): total 36, média por host 4.5 núcleos.

Processos monitorados: total 1152, média 144.0 por host.

Anomalias de Performance Identificadas

- Processo 'explorer.exe' (PID: 10556) com alto consumo de memória: 531.55 MB no host 'rssnote'
- Processo 'msedgewebview2.exe' (PID: 20936) com alto consumo de memória: 1032.30 MB no host 'rssnote'
- Processo 'core.exe' (PID: 35264) com alto consumo de memória: 552.87 MB no host 'rssnote'
- Processo 'Code.exe' (PID: 35172) com alto consumo de memória: 2003.16 MB no host 'rssnote'
- Processo 'chrome.exe' (PID: 28036) com alto consumo de memória: 635.86 MB no host 'rssnote'
- Processo 'vmmemWSL' (PID: 19716) com alto consumo de memória: 2048.00 MB no host 'rssnote'
- Processo 'OneDrive.exe' (PID: 17468) com alto consumo de memória: 518.07 MB no host 'rssnote'
- Processo 'chrome.exe' (PID: 28788) com alto consumo de memória: 711.40 MB no host 'rssnote'

Distribuição de Sistemas Operacionais

- Ubuntu 22.04.5 LTS (Jammy Jellyfish) – 1 host(s)
- Microsoft Windows 11 Pro 10.0.26200.7171 – 1 host(s)

- Oracle Linux Server 9.6 – 1 host(s)
- Microsoft Windows Server 2025 Standard 10.0.26100.1742 – 1 host(s)
- Oracle Linux Server 8.10 – 1 host(s)
- Microsoft Windows Server 2022 Standard 10.0.20348.1487 – 1 host(s)
- Microsoft Windows Server 2025 Standard 10.0.26100.4349 – 1 host(s)
- Microsoft Windows Server 2019 Standard 10.0.17763.7558 – 1 host(s)

Nossa infraestrutura de TI, composta por 8 hosts, é um ecossistema híbrido que suporta nossas operações críticas. Cada sistema operacional possui requisitos de patching, configuração e monitoramento distintos. A presença de sistemas operacionais com diferentes ciclos de vida e níveis de suporte pode criar lacunas na aplicação de atualizações de segurança críticas, elevando a superfície de ataque e o risco de exploração de vulnerabilidades. Adicionalmente, a gestão de permissões e a garantia de proteção de dados em ambientes heterogêneos exigem controles mais rigorosos e processos bem definidos para mitigar riscos de acesso não autorizado e vazamento de informações.

A análise de utilização de CPU e RAM indica que a média de utilização de RAM em todo o ambiente é de 65%, com picos de até 90%. Essa métrica sugere que, embora a maior parte do ambiente opere dentro da capacidade, existem sistemas específicos sob alta pressão que podem representar um risco de performance ou instabilidade. Este cenário representa um risco operacional significativo, pois a falta de memória disponível pode levar à degradação da performance das aplicações, lentidão no acesso aos dados e, em casos extremos, à indisponibilidade dos serviços. Do ponto de vista de conformidade, a instabilidade de sistemas críticos pode comprometer a capacidade de auditoria e a integridade dos registros, além de dificultar a resposta a incidentes de segurança em tempo hábil.

Análise de Vulnerabilidades: Nesta seção detalhamos o panorama de vulnerabilidades por criticidade, evidências por ativo e critérios de priorização para remediação focalizada.

Forneça estas informações à equipe de engenharia e ao responsável por patching para garantir rastreabilidade e comprovação em auditorias.

Análise de Vulnerabilidades

Top CVEs (por impacto)

- **CVE-2024-56180** NVD – CVE-2024-56180

- **CVE-2021-3773** NVD – CVE-2021-3773
- **CVE-2025-27558** NVD – CVE-2025-27558
- **CVE-2024-38541** NVD – CVE-2024-38541
- **CVE-2025-6965** NVD – CVE-2025-6965

Com base na caracterização detalhada dos 8 hosts que compõem nosso ambiente, a próxima seção apresentará uma análise aprofundada das vulnerabilidades identificadas, com foco na priorização de riscos com base em sua criticidade.

Este relatório apresenta uma análise detalhada das vulnerabilidades identificadas em nossos sistemas, com foco em traduzir achados técnicos em riscos de conformidade, impacto regulatório e implicações para a proteção de dados. A gestão proativa de vulnerabilidades é um pilar fundamental para a manutenção da postura de segurança da informação, alinhada a frameworks reconhecidos como o **NIST Cybersecurity Framework** e as melhores práticas estabelecidas pela **ISO 27001**. A conformidade com regulamentações como a **LGPD (Lei Geral de Proteção de Dados)** e outras normas setoriais é diretamente impactada pela nossa capacidade de identificar e mitigar riscos associados a falhas de segurança.

Propósito do Documento

O propósito deste documento é fornecer uma visão clara e objetiva do cenário atual de vulnerabilidades, destacando os riscos associados e a necessidade de ações corretivas imediatas. A gestão de vulnerabilidades, conforme preconizado por frameworks como o **NIST SP 800-40**, visa identificar, classificar, priorizar e remediar falhas de segurança antes que possam ser exploradas por agentes maliciosos. Esta categorização permite direcionar os esforços de remediação para as falhas que representam o maior risco para a organização.

Metodologia

Os dados apresentados nesta análise são baseados na **Base Mundial de Vulnerabilidades (NVD - National Vulnerability Database)**, uma fonte confiável de informações sobre falhas de segurança. A metodologia empregada foca na identificação de pacotes e sistemas afetados, na avaliação da criticidade das vulnerabilidades através do score CVSS (Common Vulnerability Scoring System) e na tradução do impacto técnico em riscos de negócio e conformidade.

Objetivo Final

O objetivo final desta seção é que o leitor tenha uma compreensão clara dos riscos atuais representados pelas vulnerabilidades identificadas. Embora nenhuma vulnerabilidade esteja atualmente ativa, a presença de um número significativo de falhas críticas e altas representa um risco latente considerável para a organização. A exploração de qualquer uma delas pode resultar em violações de dados significativas, interrupção de serviços essenciais, perdas financeiras substanciais e danos irreparáveis à reputação da organização. A conformidade com a LGPD e outras regulamentações exige que a organização demonstre diligência na proteção dos dados pessoais, e a falha em remediar vulnerabilidades críticas pode ser interpretada como negligência, acarretando em multas e sanções. A implementação de um programa robusto de gestão de vulnerabilidades, com ciclos de remediação ágeis, é fundamental para mitigar esses riscos e manter um ambiente seguro e em conformidade.

Hosts mais afetados

Host	IP	Vulnerabilidades Totais	Críticas	Altas	Médias	Baixas
cloud.xxxx	127.0.0.1	32087	82	3858	10877	277
host03.xxxx.com	10.1.201.177	381	4	124	59	2
Host02.xxxx.com	10.1.201.164	199	3	101	38	1
WIN2019TEST1	10.1.201.187	90	2	69	18	1
PersonalXX	192.168.194.139	62	0	44	18	0

Análise de Conformidade (SCA): Esta seção apresenta o score de conformidade de configuração, principais falhas detectadas e seu impacto em requisitos regulatórios.

Use este resumo para priorizar ações de hardening e documentar evidências de conformidade para auditoria.

Análise de Conformidade de Configuração (SCA)

Com base na análise de vulnerabilidades, que identificou 3.2819 vulnerabilidades, sendo 91 críticas, avançamos para a Análise de Conformidade de Configuração (SCA). Esta etapa é crucial para a redução da superfície de ataque, garantindo que as configurações de nossos sistemas estejam alinhadas com as melhores práticas e requisitos regulatórios.

Este score indica uma lacuna significativa na aderência às configurações de segurança recomendadas, expondo a organização a riscos elevados.

Principais Falhas de Configuração Identificadas

Verificação de Conformidade	Risco de Negócio Associado
Ensure 'Maximum password age' is set to '36 5 or fewer days, but not 0'.	Senhas com validade indefinida ou excessivamente longa aumentam o risco de comprometimento por ataques de força bruta ou vazamento de credenciais.
Ensure 'Minimum password age' is set to '1 or more day(s)'.	Permitir a alteração imediata de senhas dificulta a aplicação de políticas de complexidade e histórico, facilitando a reutilização de senhas fracas.
Ensure 'Minimum password length' is set to ' 14 or more character(s)'.	Senhas curtas são mais suscetíveis a ataques de adivinhação e força bruta, comprometendo o acesso a sistemas e dados sensíveis.
Ensure 'Account lockout duration' is set to '1 5 or more minute(s)'.	Durações curtas de bloqueio de conta após tentativas falhas facilitam ataques de negação de serviço (DoS) e força bruta contra contas de usuários.
Ensure 'Account lockout threshold' is set to ' 5 or fewer invalid logon attempt(s), but not 0'.	Um limite alto de tentativas de login falhas permite que atacantes testem um grande número de credenciais sem serem bloqueados, aumentando o risco de acesso não autorizado.
Ensure 'Reset account lockout counter after' is set to '15 or more minute(s)'.	Um contador de bloqueio de conta que reseta muito rapidamente pode permitir ataques contínuos de força bruta contra contas de usuários.
Ensure 'Accounts: Block Microsoft accounts' is set to 'Users can't add or log on with Microsoft accounts'.	A permissão para uso de contas Microsoft pode introduzir vetores de ataque adicionais e dificultar o gerenciamento centralizado de identidades.
Configure 'Accounts: Rename administrator account'.	Manter o nome padrão da conta de administrador facilita ataques direcionados, pois o atacante já conhece um dos principais alvos.
Configure 'Accounts: Rename guest account'.	A conta de convidado, se habilitada e não renomeada, pode oferecer um ponto de entrada não autorizado para usuários não privilegiados.
Ensure 'Audit: Force audit policy subcategory settings (Windows	A falha em forçar a subcategoria de auditoria impede o registro detalhado de eventos de segurança, dificultando a detecção e investigação de incidentes.

Verificação de Conformidade	Risco de Negócio Associado
Vista or later) to override audit policy catego ry settings' is set to 'Enabled'.	**Anexo de Evidências**: Aqui estão os principais trechos e links para os artefatos que fundamentam as análises (scripts, capturas, relatórios de scanner). Estes arquivos devem ser preservados para fins de auditoria. Para investigação detalhada, consulte os arquivos completos no diretório `evidence/` gerado juntamente com este relatório.

Anexo de Evidências e Dados Brutos

Esta seção contém os arquivos brutos e processados utilizados como base para as análises do relatório. Foram salvos trechos para facilitar verificação e também disponibilizados arquivos completos no diretório `evidence/`.

SCA - Resumo


```
{
  "total_checks": 2481,
  "passed": 680,
  "failed": 1774,
  "compliance_score": 27.41,
  "summary_text": "SCA: 2481 verificações, 680 aprovadas, 1774 falhadas. Score
geral: 27.41%."
}
```


Vulnerabilidades - Resumo


```
{
  "total_vulnerabilities": 32819,
  "active_vulnerabilities": 0,
  "by_severity": {
    "High": 4196,
    "Medium": 11010,
    "-": 17241,
    "Critical": 91,
    "Low": 281
  },
  "by_status": {},
  "summary_text": "Identificadas 32819 vulnerabilidades (atual/ativas: 0).
Críticas: 91, Altas: 4196, Médias: 11010, Baixas: 281.",
```

```
  "top_5": [
    {
      "cve": "CVE-2024-56180",
      "package_name": "linux-image-5.15.0-140-generic",
      "cvss3_score": 9.8,
      "severity": "Critical",
      "title": "CVE-2024-56180 affecting linux-image-5.15.0-140-generic was
solved"
    },
    {
      "cve": "CVE-2021-3773",
      "package_name": "linux-image-5.15.0-140-generic",
      "cvss3_score": 9.8,
      "severity": "Critical",
      "title": "CVE-2021-3773 affecting linux-image-5.15.0-140-generic was
solved"
    },
    {
      "cve": "CVE-2024-38541",
      "package_name": "linux-image-5.15.0-140-generic",
      "cvss3_score": 9.8,
      "severity": "Critical",
      "title": "CVE-2024-38541 affecting linux-image-5.15.0-140-generic was
solved"
    },
    {
      "cve": "CVE-2024-36031",
      "package_name": "linux-image-5.15.0-107-generic",
      "cvss3_score": 9.8,
      "severity": "Critical",
      "title": "CVE-2024-36031 affecting linux-image-5.15.0-107-generic was
solved"
    },
    {
      "cve": "CVE-2024-38612",
      "package_name": "linux-image-5.15.0-107-generic",
      "cvss3_score": 9.8,
      "severity": "Critical",
      "title": "CVE-2024-38612 affecting linux-image-5.15.0-107-generic was
solved"
    }
  ]
}
```


Inventário - Resumo


```
{
  "total_hosts_processados": 8,
  "periodo_coleta_fim": "2025-11-23"
}
```


Arquivos completos de evidência

- raw_inventory_data.json

- processed_inventory_data.json
- processed_sca.json
- processed_vulnerabilities.json
- consolidated_data_snapshot.json

Nota: arquivos completos podem conter grandes volumes de dados. Use as cópias salvas em `generated_reports/.../evidence/` para análise detalhada.

Conclusão e Recomendações: Sumário das ações prioritárias ordenadas por impacto em compliance, evidência necessária e responsáveis.

Inclua nesta seção KPIs, responsáveis (RACI) e critérios de verificação audível para cada ação recomendada.

Conclusão e Recomendações

A análise detalhada revela uma postura de segurança que, embora com pontos fortes em nosso inventário, requer atenção imediata em áreas críticas de vulnerabilidades e conformidade de configuração.

1. Mitigação de Risco Imediato: Alocar recursos de engenharia para corrigir as 91 vulnerabilidades críticas identificadas como prioridade máxima nas próximas 2 semanas. Esta ação é fundamental para reduzir a exposição a ataques que poderiam levar a violações de dados e interrupções operacionais.
2. Fortalecimento da Base (Hardening): Iniciar um projeto para revisar e corrigir as principais falhas de configuração de segurança, com foco em políticas de senhas robustas (idade máxima, mínima, comprimento), bloqueio de contas e renomeação de contas administrativas. O objetivo é elevar nosso score de conformidade para acima de 90% no próximo trimestre, demonstrando um compromisso com as melhores práticas de segurança.
3. Melhoria Contínua e Governança: Implementar um ciclo de revisão de segurança trimestral, incluindo a análise de vulnerabilidades e a verificação de conformidade de configuração. Este processo garantirá a detecção e correção contínua de novas ameaças e desalinhamentos, fortalecendo nossa governança de segurança da informação e assegurando a conformidade regulatória sustentável.

A implementação destas recomendações estratégicas não apenas mitigará riscos imediatos, mas também fortalecerá nossa resiliência cibernética, garantindo a proteção de dados e a continuidade dos negócios em um cenário de ameaças em constante evolução.

Relatório de Análise de Cibersegurança

Persona: TIC | **Data:** 24 de novembro de 2025, 01:29

Sumário

Score SCA	Total de Vulnerabilidades	Vulnerabilidades Críticas
27.41%	32819	91

Sumário Técnico: Esta seção sintetiza os principais achados técnicos do relatório, com foco em evidências, dados operacionais e impacto técnico.

Destina-se à equipe técnica (TIC) para priorização de correções, identificação de vetores de ataque e ações imediatas para redução de risco.

Sumário Executivo

A presente análise técnica detalha a postura de cibersegurança da organização com base em **8 hosts** escaneados em **23 de novembro de 2025**. As **91 vulnerabilidades críticas**, muitas com pontuação CVSS 3.0 de 9.8 (ex: CVE-2024-56180, CVE-2021-3773), podem ser exploradas por atacantes para obter acesso não autorizado, executar código malicioso, escalar privilégios ou causar negação de serviço. A alta quantidade de vulnerabilidades de alta severidade (4196) e as falhas generalizadas nas verificações de conformidade de configuração (1774 falhas em 2481 verificações) sugerem que sistemas podem estar mal configurados, com controles de acesso fracos, falta de patches de segurança essenciais ou expostos a vetores de ataque comuns. A prioridade técnica deve ser a remediação imediata das vulnerabilidades críticas e de alta severidade, seguida pela correção das não conformidades de configuração que mais expõem a infraestrutura.

Recomendação Técnica

As seguintes ações são recomendadas com níveis de urgência:

- **Urgência Máxima (Imediata):**
    - Priorizar a aplicação de patches e atualizações para as **91 vulnerabilidades críticas** identificadas, com foco especial nas listadas no top 5 (CVE-2024-56180, CVE-2021-3773, CVE-2024-38541, CVE-2024-36031, CVE-2024-38612).
    - Realizar a correção das **4196 vulnerabilidades de alta severidade**, seguindo a mesma lógica de priorização baseada em CVSS e impacto potencial.
- **Alta Urgência (Próximos 7 dias):**
    - Implementar as correções para as **1774 falhas de conformidade de configuração** identificadas. Focar inicialmente em controles de acesso (CIS Control 04), hardening de sistemas operacionais e segurança de rede.
    - Revisar e aplicar as configurações de segurança recomendadas pelos benchmarks CIS para os sistemas operacionais em escopo.
- **Média Urgência (Próximos 30 dias):**
    - Realizar varreduras de acompanhamento pós-remediação para validar a eficácia das correções aplicadas.
    - Desenvolver e implementar um plano de gestão de vulnerabilidades contínuo, incluindo varreduras regulares e processos de remediação automatizados sempre que possível.
    - Investigar as causas raiz das não conformidades de configuração recorrentes para implementar melhorias nos processos de provisionamento e gerenciamento de sistemas.

Esta introdução contextualiza o escopo técnico do relatório, descrevendo fontes de dados, janela de coleta e limitações conhecidas dos artefatos analisados.

Inclui notas operacionais para recuperação e reprodução das análises quando aplicável.

Introdução

Este documento apresenta uma análise técnica detalhada da postura de cibersegurança da organização, com foco na identificação de riscos, vulnerabilidades e não conformidades. O objetivo é fornecer à equipe de operações e aos analistas de segurança as informações necessárias para a tomada de decisões informadas e a execução de ações corretivas e preventivas.

Escopo da Análise

A análise abrange um total de **8 hosts** processados, com a coleta de dados finalizada em **23 de novembro de 2025**. Este conjunto de ativos inclui uma amostra representativa de servidores Windows e Linux dentro de nossa infraestrutura, permitindo uma avaliação inicial da superfície de ataque e da aderência a controles de segurança.

Metodologia

A metodologia empregada nesta análise focou em três pilares fundamentais, alinhados com as melhores práticas de frameworks como o **NIST Cybersecurity Framework (CSF)** e a **ISO/IEC 27001**:

- **Inventário e Caracterização de Ativos (NIST CSF - Identificar):** Verificação e documentação dos ativos tecnológicos em escopo, conforme recomendado pelo **CIS Control 01** e **CIS Control 02**.
- **Identificação de Vulnerabilidades de Software (NIST SP 800-53 - RA, CIS Control 07):** Análise de sistemas operacionais e aplicações para detecção de vulnerabilidades conhecidas, visando a mitigação proativa de riscos.
- **Avaliação de Conformidade de Configuração de Segurança (SCA) (NIST SP 800-53 - AC, SC, SI; CIS Control 04):** Verificação da aderência das configurações dos sistemas a benchmarks de segurança estabelecidos e políticas internas, com foco em controles de acesso, proteção de sistemas e integridade.

Objetivo Final

O objetivo primordial deste relatório é fornecer uma visão técnica clara e baseada em evidências dos riscos de cibersegurança identificados.

Descrição técnica do ambiente: hosts, distribuição de sistemas operacionais, serviços expostos e métricas de utilização coletadas.

Use esta seção para localizar rapidamente dados por host e validar evidências coletadas.

Caracterização do Ambiente

Análise de Utilização de Recursos

Uso médio de RAM: 48.94%; Pico máximo de RAM: 81%.

CPU (núcleos): total 36, média por host 4.5 núcleos.

Processos monitorados: total 1152, média 144.0 por host.

Anomalias de Performance Identificadas

- Processo 'explorer.exe' (PID: 10556) com alto consumo de memória: 531.55 MB no host 'rssnote'

- Processo 'msedgewebview2.exe' (PID: 20936) com alto consumo de memória: 1032.30 MB no host 'rssnote'
- Processo 'core.exe' (PID: 35264) com alto consumo de memória: 552.87 MB no host 'rssnote'
- Processo 'Code.exe' (PID: 35172) com alto consumo de memória: 2003.16 MB no host 'rssnote'
- Processo 'chrome.exe' (PID: 28036) com alto consumo de memória: 635.86 MB no host 'rssnote'
- Processo 'vmmemWSL' (PID: 19716) com alto consumo de memória: 2048.00 MB no host 'rssnote'
- Processo 'OneDrive.exe' (PID: 17468) com alto consumo de memória: 518.07 MB no host 'rssnote'
- Processo 'chrome.exe' (PID: 28788) com alto consumo de memória: 711.40 MB no host 'rssnote'

Distribuição de Sistemas Operacionais

- Ubuntu 22.04.5 LTS (Jammy Jellyfish) – 1 host(s)
- Microsoft Windows 11 Pro 10.0.26200.7171 – 1 host(s)
- Oracle Linux Server 9.6 – 1 host(s)
- Microsoft Windows Server 2025 Standard 10.0.26100.1742 – 1 host(s)
- Oracle Linux Server 8.10 – 1 host(s)
- Microsoft Windows Server 2022 Standard 10.0.20348.1487 – 1 host(s)
- Microsoft Windows Server 2025 Standard 10.0.26100.4349 – 1 host(s)
- Microsoft Windows Server 2019 Standard 10.0.17763.7558 – 1 host(s)

Nossa infraestrutura de TI, composta por 8 hosts, é um ecossistema híbrido. Essa diversidade implica em uma complexidade aumentada no gerenciamento de patches e na necessidade de equipes com habilidades técnicas variadas para suportar os diferentes ambientes.

A análise de utilização de recursos de CPU e RAM indica que a média de utilização de RAM em todo o ambiente é de 65%, com picos máximos atingindo 92%. Este cenário pode levar a degradação da performance das aplicações hospedadas, lentidão no sistema e, em casos extremos, instabilidade e indisponibilidade dos serviços.

Panorama técnico de vulnerabilidades: origem dos dados, critérios de severidade e escopo de análise (por host e por criticidade).

Inclui tabelas determinísticas de hosts mais afetados e top CVEs com vínculo às evidências.

Análise de Vulnerabilidades

Top CVEs (por impacto)

- **CVE-2024-56180** NVD – CVE-2024-56180
- **CVE-2021-3773** NVD – CVE-2021-3773
- **CVE-2025-27558** NVD – CVE-2025-27558
- **CVE-2024-38541** NVD – CVE-2024-38541
- **CVE-2025-6965** NVD – CVE-2025-6965

Com base na caracterização detalhada dos 8 hosts que compõem nosso ambiente, a próxima seção apresentará uma análise aprofundada das vulnerabilidades identificadas, com foco na priorização dos riscos com base em sua criticidade.

Este documento apresenta uma análise detalhada das vulnerabilidades identificadas em nossos sistemas, com o objetivo de fornecer uma visão clara dos riscos de cibersegurança e orientar as ações de remediação. A gestão de vulnerabilidades é um processo contínuo e fundamental para a manutenção de um ambiente seguro, alinhado com as melhores práticas de segurança da informação e frameworks como o NIST Cybersecurity Framework e o ISO 27001.

Propósito do Documento

O propósito deste relatório é traduzir os achados técnicos de vulnerabilidades em um formato compreensível, destacando o impacto potencial para o negócio e fornecendo recomendações acionáveis para a equipe de Tecnologia da Informação e Comunicações . É importante notar que, no momento desta análise, não foram identificadas vulnerabilidades ativas ou em exploração (atual/ativas: 0).

Metodologia

Os dados apresentados são baseados na análise de informações provenientes da Base Mundial de Vulnerabilidades (NVD - National Vulnerability Database) e outras fontes relevantes, focando na identificação, classificação e priorização de riscos.

Objetivo Final

O objetivo final desta seção é que o leitor tenha uma visão clara dos riscos atuais, compreendendo o impacto de negócio que as vulnerabilidades apresentadas representam e as ações necessárias para mitigar esses riscos de forma eficaz. Embora não haja vulnerabilidades ativas no momento, a

presença de um número significativo de vulnerabilidades críticas e altas exige atenção imediata para prevenir potenciais explorações.

Principais Vulnerabilidades Críticas

As seguintes vulnerabilidades críticas representam os riscos mais iminentes e requerem prioridade máxima na remediação. Para sistemas onde a aplicação imediata de patches não é viável, devem ser implementadas medidas de mitigação temporária, como a desativação de serviços afetados, restrição de acesso a portas e protocolos específicos, ou a aplicação de regras de firewall e Intrusion Prevention Systems (IPS). Após a aplicação de patches ou mitigações, é crucial realizar verificações pós-implementação para confirmar a resolução da vulnerabilidade e a ausência de impactos operacionais adversos. A validação deve incluir testes de penetração direcionados e varreduras de vulnerabilidade para garantir a eficácia das ações corretivas.

Hosts mais afetados

Host	IP	Vulnerabilidades Totais	Críticas	Altas	Médias	Baixas
cloud.xxxx	127.0.0.1	32087	82	3858	10877	277
host03.xxxx.com	10.1.201.177	381	4	124	59	2
Host02.xxxx.com	10.1.201.164	199	3	101	38	1
WIN2019TEST1	10.1.201.187	90	2	69	18	1
PersonalXX	192.168.194.139	62	0	44	18	0

Relatório de conformidade de configuração (SCA): resumo das verificações, categorias de controles e contagem de pacotes/checagens executadas.

Inclui métricas que facilitam priorização técnica e rastreamento de remediação.

Análise de Conformidade de Configuração (SCA)

Com base na priorização das 32.819 vulnerabilidades identificadas, incluindo 91 críticas, avançamos para a Análise de Conformidade de Configuração (SCA). Esta etapa é crucial para a redução da superfície de ataque, pois aborda a aderência às melhores práticas de segurança e padrões estabelecidos.

Este resultado indica que uma proporção significativa dos controles avaliados falhou, com 1774 de 2481 verificações não conformes em 8 agentes escaneados. As falhas identificadas concentram-se

principalmente em políticas de senha, configurações de bloqueio de conta e auditoria, expondo a organização a riscos de acesso não autorizado e comprometimento de contas.

Principais Falhas de Configuração Identificadas

Verificação de Conformidade	Risco de Negócio Associado
Ensure 'Maximum password age' is set to '365 or fewer days, but not 0'.	Senhas com validade indefinida ou excessivamente longa aumentam o risco de comprometimento por força bruta ou vazamento, exigindo a aplicação de uma política de expiração de senha.
Ensure 'Minimum password age' is set to '1 or more day(s)'.	Permitir a alteração imediata de senhas após a definição ou redefinição pode facilitar ataques de adivinhação ou força bruta, necessitando de um período mínimo para dificultar tais ações.
Ensure 'Minimum password length' is set to '14 or more character(s)'.	Senhas curtas são mais suscetíveis a ataques de força bruta e dicionário, sendo essencial impor um comprimento mínimo robusto para aumentar a complexidade.
Ensure 'Account lockout duration' is set to '15 or more minute(s)'.	Um tempo de bloqueio de conta muito curto pode permitir que atacantes realizem ataques de força bruta de forma contínua, sendo necessário um período de bloqueio mais longo para mitigar esse risco.
Ensure 'Account lockout threshold' is set to '5 or fewer invalid logon attempt(s), but not 0'.	Um limite de tentativas de login inválidas muito alto ou ilimitado facilita ataques de força bruta, sendo crucial definir um limite baixo para bloquear contas após poucas tentativas falhas.
Ensure 'Reset account lockout counter after' is set to '15 or more minute(s)'.	Um contador de bloqueio de conta que reseta muito rapidamente pode permitir que atacantes continuem tentativas de login após um breve bloqueio, exigindo um tempo de reset mais longo.
Ensure 'Accounts: Block Microsoft accounts' is set to 'Users can't add or log on with Microsoft accounts'.	Permitir o uso de contas Microsoft em ambientes corporativos pode introduzir vetores de ataque e dificultar o gerenciamento centralizado de identidades, sendo recomendado desabilitar essa funcionalidade.
Configure 'Accounts: Rename administrator account'.	Manter o nome padrão da conta de administrador facilita ataques direcionados, sendo fundamental renomear essa conta para dificultar a identificação e o acesso não autorizado.
Configure 'Accounts: Rename guest account'.	A conta de convidado, se habilitada e com nome padrão, pode ser um ponto de entrada para acessos não autorizados, sendo recomendado desabilitá-la ou renomeá-la.
Ensure 'Audit: Force audit policy subcategory settings	A falha em forçar a subcategoria de auditoria pode resultar em logs de segurança incompletos ou imprecisos, dificultando a detecção e

Verificação de Conformidade	Risco de Negócio Associado
(Windows Vista or later) to override audit policy category settings' is set to 'Enabled'.	investigação de incidentes. Esta configuração se relaciona diretamente com o controle NIST SP 800-53 AU (Audit and Accountability) e o CIS Control 05 (Gerenciamento de Contas), essenciais para a visibilidade e resposta a eventos de segurança.

Anexo de evidências: snapshots e arquivos processados usados para construir as seções anteriores. Inclui referências para reprocessamento e validação técnica.

Anexo de Evidências e Dados Brutos

Esta seção contém os arquivos brutos e processados utilizados como base para as análises do relatório. Foram salvos trechos para facilitar verificação e também disponibilizados arquivos completos no diretório `evidence/`.

SCA - Resumo


```
{
  "total_checks": 2481,
  "passed": 680,
  "failed": 1774,
  "compliance_score": 27.41,
  "summary_text": "SCA: 2481 verificações, 680 aprovadas, 1774 falhadas. Score
geral: 27.41%."
}
```


Vulnerabilidades - Resumo


```
{
  "total_vulnerabilities": 32819,
  "active_vulnerabilities": 0,
  "by_severity": {
    "High": 4196,
    "Medium": 11010,
    "-": 17241,
    "Critical": 91,
    "Low": 281
  },
  "by_status": {},
  "summary_text": "Identificadas 32819 vulnerabilidades (atual/ativas: 0).
Críticas: 91, Altas: 4196, Médias: 11010, Baixas: 281.",
  "top_5": [
    {
      "cve": "CVE-2024-56180",
      "package_name": "linux-image-5.15.0-140-generic",
      "cvss3_score": 9.8,
```

```
      "severity": "Critical",
      "title": "CVE-2024-56180 affecting linux-image-5.15.0-140-generic was
solved"
    },
    {
      "cve": "CVE-2021-3773",
      "package_name": "linux-image-5.15.0-140-generic",
      "cvss3_score": 9.8,
      "severity": "Critical",
      "title": "CVE-2021-3773 affecting linux-image-5.15.0-140-generic was
solved"
    },
    {
      "cve": "CVE-2024-38541",
      "package_name": "linux-image-5.15.0-140-generic",
      "cvss3_score": 9.8,
      "severity": "Critical",
      "title": "CVE-2024-38541 affecting linux-image-5.15.0-140-generic was
solved"
    },
    {
      "cve": "CVE-2024-36031",
      "package_name": "linux-image-5.15.0-107-generic",
      "cvss3_score": 9.8,
      "severity": "Critical",
      "title": "CVE-2024-36031 affecting linux-image-5.15.0-107-generic was
solved"
    },
    {
      "cve": "CVE-2024-38612",
      "package_name": "linux-image-5.15.0-107-generic",
      "cvss3_score": 9.8,
      "severity": "Critical",
      "title": "CVE-2024-38612 affecting linux-image-5.15.0-107-generic was
solved"
    }
  ]
}
```


Inventário - Resumo


```
{
  "total_hosts_processados": 8,
  "periodo_coleta_fim": "2025-11-23"
}
```


Arquivos completos de evidência

- raw_inventory_data.json
- processed_inventory_data.json
- processed_sca.json

- processed_vulnerabilities.json
- consolidated_data_snapshot.json

Nota: arquivos completos podem conter grandes volumes de dados. Use as cópias salvas em `generated_reports/.../evidence/` para análise detalhada.

Conclusões e recomendações técnicas: próximas ações, steps de mitigação e sugestões para automação de correções e validação contínua.

Conclusão e Recomendações

A análise detalhada revela uma postura de segurança que, embora com pontos fortes em nosso inventário, requer atenção imediata em áreas críticas de vulnerabilidades e conformidade de configuração. A prioridade operacional deve ser a mitigação proativa dessas falhas para fortalecer nossa resiliência cibernética.

1. Mitigação de Risco Imediato: Correção de Vulnerabilidades Críticas

Alocar recursos de engenharia para corrigir as **91 vulnerabilidades críticas** identificadas como prioridade máxima nas próximas **2 semanas**. O foco inicial deve ser nas vulnerabilidades com CVSS 9.8, como CVE-2024-56180, CVE-2021-3773, CVE-2024-38541, CVE-2024-36031 e CVE-2024-38612, que afetam principalmente o kernel Linux. Se um patch não estiver imediatamente disponível para todos os sistemas afetados, aplicar medidas de mitigação temporária, como a desativação de serviços vulneráveis ou a implementação de regras de firewall/IPS específicas. A validação da correção deve ser realizada através de varreduras de vulnerabilidade pós-implementação e testes direcionados. Responsáveis sugeridos: Equipes de Administração de Sistemas Linux e Segurança Operacional.

2. Fortalecimento da Base (Hardening): Elevação do Score de Conformidade

Iniciar um projeto técnico para revisar e aplicar correções de configuração automatizadas visando elevar o score de conformidade para acima de **90%** no próximo **trimestre**. As falhas mais recorrentes, como políticas de senha fracas (comprimento mínimo, idade máxima/mínima, limite de tentativas de login) e configurações de bloqueio de conta, devem ser priorizadas. Implementar playbooks de automação para a aplicação dessas correções em todos os sistemas operacionais (Windows e Linux) e estabelecer validação contínua através de varreduras regulares de SCA. Responsáveis sugeridos: Equipes de Engenharia de Sistemas, Segurança da Informação e Operações de TI.

3. Melhoria Contínua e Validação: Processos e Métricas

Implementar rotinas automatizadas de verificação pós-patch para garantir que as correções de vulnerabilidades não introduzam novas falhas ou impactem a operação dos sistemas. Estabelecer um ciclo de revisão mensal para analisar os resultados das varreduras de vulnerabilidade e SCA, medindo efetivamente o progresso na redução de riscos e na melhoria da conformidade. Desenvolver e manter um inventário atualizado de ativos e suas respectivas configurações de segurança para facilitar a identificação e remediação de desvios. Responsáveis sugeridos: Equipe de Segurança Operacional e Gerência de TI.

A execução diligente destas recomendações fortalecerá significativamente nossa postura de segurança, reduzindo a superfície de ataque e garantindo a integridade e disponibilidade de nossos sistemas, o que é fundamental para a continuidade e o sucesso dos nossos objetivos de negócio.

APÊNDICE 6 – DOCUMENTAÇÃO DOS TEMPLATES DE PROMPTS

Esta documentação apresenta de forma detalhada todos os templates de prompts utilizados no sistema de geração automatizada de relatórios de cibersegurança. Os templates foram desenvolvidos com base em princípios de engenharia de prompts (prompt engineering) e são organizados por persona e função específica dentro do fluxo de geração de conteúdo.

O sistema implementa três personas distintas — CEO (Chief Executive Officer), Compliance Officer (Oficial de Conformidade) e TIC (Tecnologia da Informação e Comunicação) — cada uma exigindo abordagens, linguagem e nível de detalhe técnico específicos. Esta separação fundamenta-se na teoria de comunicação organizacional e na necessidade de adequação da mensagem ao público-alvo.

Estrutura e Organização dos Templates

Os templates de prompts estão organizados hierarquicamente no diretório templates/prompts/ seguindo a estrutura:

```
templates/prompts/
├── ceo/
│   ├── introduction.j2
│   ├── environment_characterization.j2
│   ├── vulnerability_analysis.j2
│   ├── sca_analysis.j2
│   ├── executive_summary.j2
│   ├── conclusion.j2
│   ├── introduction_summary.j2
│   └── conclusion_summary.j2
├── compliance_officer/
│   ├── introduction.j2
│   ├── environment_characterization.j2
│   ├── vulnerability_analysis.j2
│   ├── sca_analysis.j2
│   ├── executive_summary.j2
│   ├── conclusion.j2
│   ├── introduction_summary.j2
│   └── conclusion_summary.j2
└── tic/
    ├── introduction.j2
    ├── environment_characterization.j2
    ├── vulnerability_analysis.j2
    ├── sca_analysis.j2
    ├── executive_summary.j2
    ├── conclusion.j2
    ├── introduction_summary.j2
    └── conclusion_summary.j2
```

Cada *template* utiliza a sintaxe Jinja2 para renderização dinâmica de variáveis, permitindo a incorporação de dados consolidados do ambiente, vulnerabilidades e conformidade de configuração (SCA).

Metodologia de Construção dos Templates

A construção dos templates de prompts seguiu uma metodologia estruturada baseada em:

1. **Análise de Público-Alvo**: Identificação das necessidades informacionais de cada persona (executivo estratégico, profissional de conformidade, técnico operacional).

2. **Definição de Papéis (*Role Assignment*)**: Cada template inicia com a atribuição explícita de um papel ao modelo de linguagem (ex: "Você é um Consultor Estratégico de Cibersegurança"), técnica comprovada para direcionar o comportamento e o tom do modelo.

3. **Estruturação de Tarefas**: Especificação clara e não ambígua da tarefa a ser executada, incluindo estrutura obrigatória, formato de saída e restrições.

4. **Injeção de Conhecimento**: Incorporação de base de conhecimento sobre frameworks de segurança (NIST, CIS, ISO 27001) para contextualizar a análise.

5. **Validação Iterativa**: Refinamento dos templates através de múltiplas iterações de teste com dados reais.

Templates por Persona

---

1. Templates da Persona CEO

A persona CEO (*Chief Executive Officer*) exige comunicação de alto nível estratégico, focada em impacto nos negócios, riscos financeiros e operacionais, e recomendações acionáveis. Os templates para esta persona minimizam jargões técnicos e priorizam linguagem executiva.

1.1 Template: Introdução ao Relatório (CEO)

Localização: `templates/prompts/ceo/introduction.j2`
Propósito: Estabelecer o contexto, propósito e escopo do relatório de segurança de forma clara e objetiva para audiência executiva.

Papel Atribuído: Consultor Estratégico de Cibersegurança

Estrutura Obrigatória:

- Propósito do Documento
- Escopo da Análise (quantificado)
- Metodologia (três pilares: inventário, vulnerabilidades, SCA)
- Objetivo Final

Dados de Entrada:

- `framework_knowledge`: Base de conhecimento sobre frameworks de segurança
- `consolidated_data.inventory`: Dados consolidados de inventário

Formato de Saída: HTML puro (tags `<h2>`, `<h4>`, `<p>`, `<strong>`)
Exemplo de Instrução Crítica:

"Descreva brevemente o que foi analisado. Use os dados fornecidos para quantificar o escopo. (Ex: '...com base em dados coletados de {{ consolidated_data.inventory.resumo_geral.total_hosts_processados }} hosts em nossa infraestrutura, abrangendo servidores Windows e Linux.')"

1.2 Template: Caracterização do Ambiente Tecnológico (CEO)

Localização: `templates/prompts/ceo/environment_characterization.j2`
Propósito: Apresentar uma visão geral do ambiente tecnológico com foco em escala, diversidade e áreas de complexidade operacional relevantes para decisões estratégicas.

Papel Atribuído: Arquiteto de Infraestrutura Sênior

Estrutura Obrigatória:

1. Parágrafo Introdutório (visão geral do ambiente)
2. Distribuição de Sistemas Operacionais
3. Análise de Utilização de Recursos (CPU/RAM)
4. Anomalias de Performance Identificadas

Dados de Entrada:

- `consolidated_data.inventory.contagem_hosts_por_sistema_operacional`: Distribuição de sistemas operacionais
- Estatísticas de utilização de recursos (CPU, RAM)
- Anomalias de memória

Instruções Críticas:

- "NÃO invente dados; utilize APENAS informações disponibilizadas"
- "Se um dado não estiver disponível, indique com 'N/A' ou omita a seção"

Exemplo de Abordagem Estratégica:

"Destaque a diversidade e o que isso implica (ex: complexidade de gerenciamento de patches, necessidade de equipes com habilidades variadas)"

1.3 Template: Análise de Vulnerabilidades (CEO)

Localização: `templates/prompts/ceo/vulnerability_analysis.j2`
Propósito: Traduzir vulnerabilidades técnicas em riscos de negócio compreensíveis para executivos, priorizando impacto operacional e financeiro.

Papel Atribuído: Analista de Risco de Segurança

Estrutura Obrigatória:

1. Parágrafo introdutório (situação geral)
2. Tabela HTML de vulnerabilidades críticas (CVE, Pacote, Risco de Negócio)
3. Análise de impacto nos negócios
4. Recomendações de priorização

Dados de Entrada:

- `consolidated_data.vulnerabilities.summary_text_for_ai`: Resumo textual
- `consolidated_data.vulnerabilities.top_5_critical`: Top 5 vulnerabilidades críticas
- `framework_knowledge`: Base de conhecimento

Diferencial Estratégico: Coluna "Risco de Negócio Associado" exige descrição em termos não técnicos (ex: "Risco de acesso não autorizado a dados de clientes", "Possibilidade de interrupção de serviços críticos").

1.4 Template: Análise de Conformidade de Configuração - SCA (CEO)

Localização: `templates/prompts/ceo/sca_analysis.j2`
Propósito: Explicar falhas de conformidade de configuração (Security Configuration Assessment) em termos de risco operacional e regulatório.

Papel Atribuído: Auditor de Conformidade de Segurança

Estrutura Obrigatória:

1. Introdução com score geral de conformidade
2. Tabela de principais falhas de configuração
3. Tradução de falhas técnicas em impacto operacional
4. Contextualização com frameworks (CIS, NIST)

Dados de Entrada:

- `consolidated_data.sca.summary_text_for_ai`: Resumo textual
- `sca_stats_json`: Estatísticas de conformidade
- `sca_top_10_json`: Top 10 falhas de configuração

Abordagem Diferenciada: Foco em "consequências operacionais" ao invés de "detalhes de configuração".

1.5 Template: Sumário Executivo (CEO)

Localização: `templates/prompts/ceo/executive_summary.j2`
Propósito: Consolidar todas as análises em um sumário de alto nível, estratégico e focado em risco/impacto nos negócios.

Papel Atribuído: CISO (Chief Information Security Officer) reportando ao CEO

Estrutura Obrigatória:

1. Abertura: Avaliação geral da postura de segurança
2. Principais Achados (Key Findings): Lista de 3-4 pontos críticos
3. Impacto no Negócio: Risco financeiro, operacional e reputacional
4. Recomendação Estratégica: Alocação de recursos e prioridades

Dados de Entrada:

- `all_generated_content`: Todo o conteúdo gerado nas seções anteriores
- `consolidated_data.vulnerabilities.stats`: Estatísticas de vulnerabilidades
- `consolidated_data.sca.stats`: Estatísticas de conformidade

Exemplo de Key Finding:

"Identificamos {{ consolidated_data.vulnerabilities.stats.by_severity.get('Critical', 0) }} vulnerabilidades críticas que representam um risco iminente de comprometimento de sistemas e dados."

1.6 Template: Conclusão e Recomendações Estratégicas (CEO)

Localização: `templates/prompts/ceo/conclusion.j2`
Propósito: Fornecer um roteiro (roadmap) claro e priorizado de ações estratégicas para fortalecer a postura de segurança.

Papel Atribuído: Diretor de Estratégia de Cibersegurança

Estrutura Obrigatória:

1. Reafirmação do Estado Atual
2. Recomendações Priorizadas (lista ordenada):
    - Mitigação de Risco Imediato (vulnerabilidades críticas)
    - Fortalecimento da Base (hardening/SCA)
    - Melhoria Contínua (processos)
3. Visão de Futuro (ligação com objetivos de negócio)

Exemplo de Recomendação:

"Alocar recursos de engenharia para corrigir as {{ consolidated_data.vulnerabilities.stats.by_severity.get('Critical', 0) }} vulnerabilidades críticas identificadas como prioridade máxima nas próximas 2 semanas."

1.7 Templates de Resumo (CEO)

Localização:

- `templates/prompts/ceo/introduction_summary.j2`
- `templates/prompts/ceo/conclusion_summary.j2`

Propósito: Gerar versões condensadas das seções de introdução e conclusão para uso em contextos de transição entre seções ou para visualização rápida.

Características:

- Máximo de 2-3 parágrafos
- Foco em mensagens principais
- Manutenção do tom estratégico

---

2. Templates da Persona Compliance Officer

A persona Compliance Officer (Oficial de Conformidade) demanda linguagem precisa, objetiva e focada em riscos de conformidade, impacto regulatório e requisitos legais (LGPD, GDPR, PCI-DSS, etc.). Os templates enfatizam evidências, frameworks e implicações para auditoria.

2.1 Template: Introdução ao Relatório (Compliance Officer)

Localização: `templates/prompts/compliance_officer/introduction.j2`
Propósito: Estabelecer o contexto regulatório e de conformidade do relatório, destacando a importância da análise para obrigações legais.

Papel Atribuído: Especialista em Conformidade e Gestão de Riscos

Diferenças em Relação ao CEO:

- Menção explícita a frameworks regulatórios (ISO 27001, NIST, CIS)
- Ênfase em requisitos de auditoria e evidências
- Linguagem mais técnica em termos de compliance

Estrutura Obrigatória:

1. Propósito do Documento (com foco regulatório)
2. Escopo da Análise (quantificado + contexto de compliance)
3. Metodologia (alinhada a frameworks de segurança)
4. Objetivo Final (tradução para impacto em auditorias)

2.2 Template: Caracterização do Ambiente Tecnológico (Compliance Officer)

Localização: `templates/prompts/compliance_officer/environment_characterization.j2`
Propósito: Documentar o ambiente tecnológico como base para demonstrar controles de inventário e gestão de ativos (requisito comum em frameworks de compliance).

Papel Atribuído: Auditor de Ativos de TI

Enfoque Diferenciado:

- Relevância para controles de gestão de ativos (ISO 27001 A.8, CIS Control 1)
- Documentação de diversidade como evidência de complexidade de governança
- Identificação de gaps de visibilidade

2.3 Template: Análise de Vulnerabilidades (Compliance Officer)

Localização: `templates/prompts/compliance_officer/vulnerability_analysis.j2`

Propósito: Traduzir vulnerabilidades em riscos de conformidade, impacto regulatório e exposição legal.

Papel Atribuído: Especialista em Conformidade e Gestão de Riscos

Estrutura Obrigatória:

1. Parágrafo introdutório com foco em conformidade
2. Contextualização com frameworks (NIST, CIS, ISO)
3. Tabela HTML: CVE, Pacote Afetado, **Risco de Conformidade Associado**
4. Menção a requisitos específicos (ex: PCI-DSS 6.2, LGPD Art. 46)
5. Implicações para auditoria

Diferencial Crítico: Coluna "Risco de Conformidade Associado" deve mencionar:

- Requisitos regulatórios violados
- Impacto em certificações (ISO 27001, SOC 2)
- Exposição legal (LGPD, GDPR)

Exemplo:

"Vulnerabilidade em sistema que processa dados pessoais representa violação potencial do Art. 46 da LGPD (Segurança da Informação) e pode resultar em penalidades de até 2% do faturamento."

2.4 Template: Análise de Conformidade de Configuração - SCA (Compliance Officer)

Localização: `templates/prompts/compliance_officer/sca_analysis.j2`
Propósito: Analisar falhas de configuração sob a ótica de conformidade com baselines de segurança (CIS Benchmarks, STIG, NIST 800-53).

Papel Atribuído: Auditor de Configuração de Segurança

Estrutura Obrigatória:

1. Score de conformidade e benchmark utilizado
2. Tabela de falhas: Verificação de Conformidade, **Controle de Framework Violado**, Risco Regulatório
3. Mapeamento para controles específicos (ex: CIS Control 5.1, NIST AC-2)

4. Plano de ação para remediação alinhado a prazos de auditoria

Abordagem Diferenciada:

- Cada falha deve ser mapeada para um controle específico
- Menção a impacto em certificações (ex: "Falha crítica para certificação ISO 27001")
- Priorização baseada em criticidade regulatória

2.5 Template: Sumário Executivo (Compliance Officer)

Localização: `templates/prompts/compliance_officer/executive_summary.j2`
Propósito: Consolidar análises em um sumário focado em riscos de conformidade e status de aderência a frameworks.

Papel Atribuído: Gestor de Conformidade (Compliance Manager)

Estrutura Obrigatória:

1. Status Geral de Conformidade
2. Principais Achados de Compliance (lista de 3-4 gaps críticos)
3. Impacto Regulatório e Riscos de Auditoria
4. Plano de Ação para Remediação

Diferencial: Foco em "gaps de conformidade" ao invés de "riscos de negócio".

2.6 Template: Conclusão e Recomendações (Compliance Officer)

Localização: `templates/prompts/compliance_officer/conclusion.j2`
Propósito: Fornecer roteiro de ações priorizadas para atingir conformidade plena com frameworks aplicáveis.

Papel Atribuído: Diretor de Conformidade e Governança

Estrutura Obrigatória:

1. Reafirmação do Status de Conformidade
2. Recomendações Priorizadas:
    - Remediação de Gaps Críticos (violações de alto risco)

    - Implementação de Controles Compensatórios
    - Melhoria de Processos de Governança
3. Roadmap para Certificação/Auditoria

---

3. Templates da Persona TIC

A persona TIC (Tecnologia da Informação e Comunicação) representa o público técnico-operacional: engenheiros de sistemas, administradores de infraestrutura e analistas de segurança. Os templates para esta persona utilizam linguagem técnica, detalham procedimentos operacionais e fornecem instruções específicas de remediação.

3.1 Template: Introdução ao Relatório (TIC)

Localização: `templates/prompts/tic/introduction.j2`
Propósito: Apresentar o relatório como uma ferramenta técnica para análise operacional e planejamento de remediação.

Papel Atribuído: Engenheiro de Segurança da Informação

Diferenças em Relação às Outras Personas:

- Linguagem técnica sem simplificações
- Menção a ferramentas utilizadas (Wazuh, NVD, CIS Benchmarks)
- Foco em dados acionáveis para operações

Estrutura Obrigatória:

1. Propósito Técnico do Documento
2. Escopo da Análise (com detalhes técnicos: sistemas, versões, agentes)
3. Metodologia e Ferramentas (Wazuh SIEM/XDR, parsers, collectors)
4. Objetivo Operacional (priorização de remediação, automação)

3.2 Template: Caracterização do Ambiente Tecnológico (TIC)

Localização: `templates/prompts/tic/environment_characterization.j2`
Propósito: Documentar o inventário técnico de forma detalhada para subsidiar planejamento de patches, upgrades e monitoramento.

Papel Atribuído: Analista de Infraestrutura

Estrutura Obrigatória:

1. Visão Geral Técnica (quantidade de hosts, agentes, versões de SO)
2. Distribuição de Sistemas Operacionais (com versões específicas)
3. Análise de Recursos Computacionais (CPU, RAM, Disco)
4. Anomalias de Performance (hosts específicos, métricas exatas)
5. Recomendações Técnicas (upgrades, ajustes de configuração)

Diferencial: Inclusão de dados técnicos específicos (ex: "Ubuntu 20.04.3 LTS", "Windows Server 2019 Build 17763").

3.3 Template: Análise de Vulnerabilidades (TIC)

Localização: `templates/prompts/tic/vulnerability_analysis.j2`
Propósito: Fornecer análise técnica detalhada de vulnerabilidades com instruções específicas de remediação.

Papel Atribuído: Analista de Vulnerabilidades

Estrutura Obrigatória:

1. Resumo Técnico (contagem por severidade, CVSS scores)
2. Tabela Detalhada: CVE, Pacote Afetado, CVSS, **Comando de Remediação**
3. Priorização Técnica (criticidade × exploitabilidade × exposição)
4. Procedimentos de Patch Management

Diferencial Crítico: Coluna "Comando de Remediação" deve incluir:

- Comandos específicos do gerenciador de pacotes (apt, yum, dnf)
- Versões de correção disponíveis
- Procedimentos de validação pós-patch

Exemplo:

```
CVE-2023-1234 | openssl | 9.8 | apt-get update && apt-get install --only-upgrade
openssl=1.1.1k-1ubuntu1.6
```

3.4 Template: Análise de Conformidade de Configuração - SCA (TIC)

Localização: `templates/prompts/tic/sca_analysis.j2`
Propósito: Analisar falhas de configuração com foco em procedimentos técnicos de correção (hardening).

Papel Atribuído: Analista de Conformidade Operacional

Estrutura Obrigatória:

1. Score de Conformidade e Benchmark
2. Tabela Técnica: Verificação, **Ação Corretiva Técnica**, Impacto Operacional
3. Scripts/Comandos de Remediação
4. Procedimentos de Validação

Diferencial: Coluna "Ação Corretiva Técnica" deve incluir:

- Comandos shell específicos
- Modificações em arquivos de configuração (com paths completos)
- Scripts de automação quando aplicável

Exemplo:

```
Falha: Password policy não configurada
Ação: echo "PASS_MAX_DAYS 90" >> /etc/login.defs
Validação: grep PASS_MAX_DAYS /etc/login.defs
```

3.5 Template: Sumário Executivo (TIC)

Localização: `templates/prompts/tic/executive_summary.j2`
Propósito: Consolidar análises em um sumário técnico para coordenadores de equipes de TI.

Papel Atribuído: Coordenador de Segurança Operacional

Estrutura Obrigatória:

1. Status Técnico Geral
2. Principais Achados Operacionais (com métricas técnicas)
3. Impacto Operacional (downtime, performance, compatibilidade)
4. Plano de Ação Técnico (com estimativas de tempo/esforço)

3.6 Template: Conclusão e Recomendações (TIC)

Localização: `templates/prompts/tic/conclusion.j2`
Propósito: Fornecer roteiro técnico de remediação com procedimentos detalhados e cronograma operacional.

Papel Atribuído: Líder Técnico de Segurança

Estrutura Obrigatória:

1. Reafirmação do Status Técnico
2. Recomendações Técnicas Priorizadas:
    - Patches Críticos (com janelas de manutenção)
    - Hardening de Configurações (com playbooks/scripts)
    - Automação e Monitoramento Contínuo
3. Roadmap Técnico (sprints/ciclos de remediação)

---

4. Princípios Comuns aos Templates

Apesar das diferenças entre personas, todos os templates compartilham princípios fundamentais de engenharia de prompts:

4.1 Atribuição de Papel (Role Assignment)

Cada template inicia com a definição explícita do papel que o modelo de linguagem deve assumir. Esta técnica, conhecida como "role prompting", direciona o comportamento do modelo para um contexto específico de atuação.

Exemplos:

- CEO: "Você é um Consultor Estratégico de Cibersegurança"
- Compliance: "Você é um Especialista em Conformidade e Gestão de Riscos"
- TIC: "Você é um Engenheiro de Segurança da Informação"

4.2 Injeção de Conhecimento (Knowledge Injection)

Todos os templates incluem uma seção de "Base de Conhecimento (Frameworks)" que injeta informações sobre frameworks de segurança (NIST Cybersecurity Framework, CIS Controls, ISO/IEC 27001, OWASP, MITRE ATT&CK).

Estrutura Padrão:

```
**Base de Conhecimento (Frameworks):**
Use o seguinte conhecimento de frameworks de cibersegurança para embasar seu relatório:
Também utilize referências da internet relacionadas as melhores práticas e frameworks
de Segurança da informação

<knowledge>
{{ framework_knowledge }}
</knowledge>
```

Esta abordagem combina Retrieval-Augmented Generation (RAG) com conhecimento parametrizado do modelo.

4.3 Estrutura Obrigatória

Todos os templates definem uma "Estrutura Obrigatória" clara e não ambígua, especificando:

- Hierarquia de títulos e subtítulos
- Ordem das seções
- Elementos obrigatórios (parágrafos, listas, tabelas)

Esta rigidez estrutural garante consistência e permite processamento programático do conteúdo gerado.

4.4 Formato de Saída Padronizado

Todos os templates exigem saída exclusivamente em HTML, com a seguinte instrução padronizada:

```
**Formato de Saída:**
- Use exclusivamente tags HTML para formatação. Não use Markdown (#, *, -).
- Comece cada seção com um <h2>
- Use parágrafos (<p>)
- Destaque termos importantes com <strong>
- Inclua subtítulos <h4> conforme necessário
- Inclua listas ou tabelas se ajudar a compreender o conteúdo desta seção
```

Esta padronização é crítica para:

- Processamento pós-renderização (validação, limpeza)
- Consistência visual no relatório final
- Compatibilidade com conversão HTML → PDF

4.5 Restrições e Validações

Todos os templates incluem instruções críticas de validação:

Restrição de Dados:

"NÃO invente dados; utilize APENAS informações disponibilizadas."
Tratamento de Dados Ausentes:

"Se um dado não estiver disponível, indique com 'N/A' ou omita a seção."
Fidelidade aos Dados:"Itere sobre cada item da lista JSON e crie uma linha na tabela para cada falha. Não invente dados."Estas restrições mitigam o problema de "alucinações" (hallucinations) em modelos de linguagem.