Relatório de Impacto à Proteção de Dados Pessoais (RIPD)

LUCAS SILVA BICALHO

MARCELO RODRIGUES JÚNIOR

MARIANA GABRIELE FERREIRA SALES

RENAN RIBEIRO PEREIRA

LAVRAS - MG

2025

SUMÁRIO

1 INTRODUÇÃO

O presente relatório tem como objetivo analisar os riscos à proteção de dados pessoais associados à implementação de uma Inteligência Artificial (IA) Distribuída desenvolvida no âmbito da disciplina de Sistemas Distribuídos (GCC129) do curso de Ciência da Computação. Como parte do desenvolvimento, foram analisadas possíveis vulnerabilidades e ameaças que podem comprometer a segurança e privacidade dos usuários. A metodologia STRIDE foi empregada para identificar e avaliar os riscos, bem como propor medidas que poderiam ser adotadas para mitigar essas ameaças, utilizando a modelagem de ameaças baseada na metodologia de Torr (2005).

A solução desenvolvida consiste em uma interface *web* que permite que o usuário insira um texto e escolha entre gerar uma imagem ou um GIF (Graphics Interchange Format), utilizando duas IAs distintas, cada uma responsável por um tipo de mídia. O processamento ocorre de forma distribuída, onde os dados são enviados ao back-end, que os encaminha a agentes de IA especializados na geração de cada tipo de conteúdo. Os modelos utilizados, Stable Diffusion v1-5 e AnimateDiff-Lightning, são executados em contêineres Docker, garantindo isolamento e escalabilidade. No entanto, essa arquitetura distribuída introduz vulnerabilidades relacionadas ao envio e armazenamento de dados, comunicação entre serviços e integridade dos resultados gerados. Diante disso, este relatório detalha os principais riscos identificados, bem como as estratégias que poderiam ser adotadas para mitigação dessas ameaças.

2 RISCOS IDENTIFICADOS NO SISTEMA

A análise de riscos do sistema foi conduzida com base no modelo **STRIDE**, que identifica seis categorias principais de ameaças: *Spoofing*, *Tampering*, *Repudiation*, *Information Disclosure*, *Denial of Service* e *Elevation of Privilege*. Considerando a arquitetura do nosso sistema, que envolve um *front-end* para entrada de texto e escolha de tipo de mídia a ser gerada, bem como agentes de IA especializados na criação de imagens e GIFs, essas ameaças representam riscos concretos e específicos.

Sendo assim, anteriormente ao processo de identificação dos riscos, foi elaborado um Diagrama de Fluxo de Dados (DFD) do sistema (Figura 1), para melhor visualização da interação entre os componentes. O processo se inicia com o **Usuário Externo**, que insere um texto e escolhe entre a geração de uma imagem ou um GIF. Essa informação é enviada ao **Front-end**, desenvolvido com HTML, JavaScript e Nginx, que realiza uma requisição via Fetch API para o **Back-end**. O **Back-end**, implementado com FastAPI e CUDA, processa a requisição e encaminha os dados para o modelo de IA correspondente. Se a escolha for a conversão de **Texto para Imagem**, os dados são enviados para a **IA Stable Diffusion**, que retorna um arquivo de imagem. Caso a opção seja **Texto para GIF**, a requisição é processada pela **IA AnimateDiff**, que gera um arquivo GIF. O conteúdo gerado é então enviado de volta ao **Back-end**, que o repassa ao **Front-end** para exibição ao usuário na interface. Esse fluxo destaca a separação clara entre as camadas do sistema e a comunicação entre elas, possibilitando, assim, a realização da análise de riscos descrita a seguir.

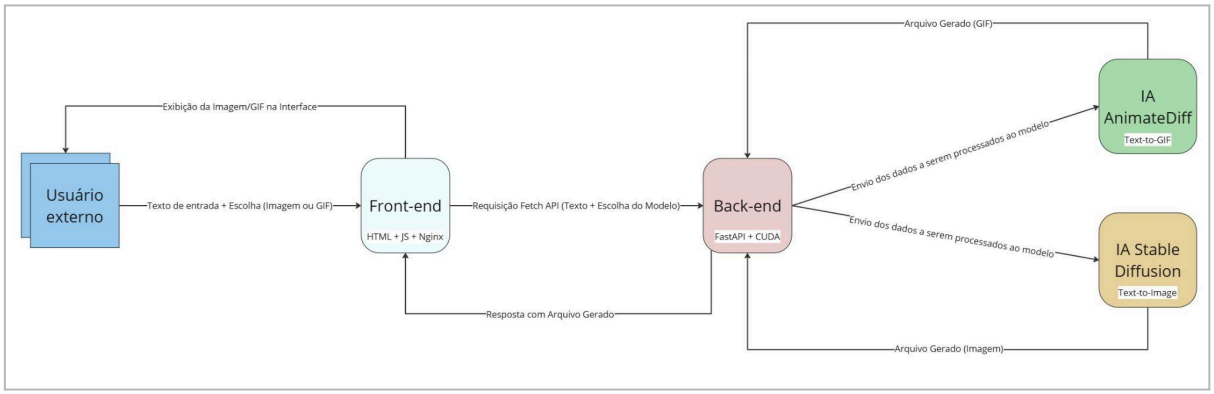

Figura 1 - Diagrama de Fluxo de Dados do Sistema

No que se refere a ***spoofing***, o sistema apresenta vulnerabilidades significativas, uma vez que não há um mecanismo de autenticação formal para validar a identidade dos usuários. Isso significa que um agente mal-intencionado pode facilmente se passar por outro usuário para explorar possíveis falhas ou inserir comandos maliciosos com o objetivo de influenciar a

geração de mídia. Como não há restrições sobre as entradas enviadas ao sistema, um atacante pode, por exemplo, enviar solicitações em nome de outro usuário, gerando conteúdos indesejados ou potencialmente abusivos sem que haja um meio confiável de rastrear a autoria dessas ações.

O risco de ***tampering*** se manifesta na comunicação entre o *front-end* e os agentes de IA, uma vez que os dados inseridos pelo usuário trafegam pela rede até atingirem os modelos responsáveis pela geração de mídia. Caso essa comunicação não esteja devidamente protegida, existe a possibilidade de interceptação e manipulação dos dados durante a transmissão. Um atacante poderia, por exemplo, modificar o texto inserido pelo usuário, alterando o *prompt* de geração de imagem ou GIF para produzir um conteúdo inesperado e potencialmente prejudicial. Além disso, sem um canal seguro de comunicação, um intermediário mal-intencionado poderia adulterar as respostas recebidas dos agentes de IA antes de entregá-las ao usuário final, comprometendo a integridade da informação gerada pelo sistema.

A ameaça de ***repudiation*** é particularmente relevante no nosso cenário, pois o sistema não implementa nenhum mecanismo de armazenamento de *logs* detalhados sobre as interações dos usuários. Isso significa que qualquer ação realizada dentro da plataforma não pode ser posteriormente verificada ou auditada com precisão. Dessa forma, um usuário mal-intencionado poderia realizar atividades prejudiciais, como o envio de conteúdos proibidos ou a tentativa de sobrecarregar o sistema, e posteriormente negar qualquer envolvimento. A ausência de registros confiáveis dificulta a aplicação de medidas corretivas e torna praticamente impossível atribuir responsabilidades em caso de uso indevido da plataforma.

O risco de ***information disclosure*** surge principalmente devido à possibilidade de armazenamento indevido dos dados fornecidos pelos usuários. Embora o sistema não colete informações sensíveis explicitamente, ele ainda processa entradas textuais que podem conter dados pessoais ou contextos sensíveis. Se esses dados forem armazenados sem a devida anonimização ou expostos a terceiros de maneira não intencional, a privacidade dos usuários poderá ser comprometida. Além disso, caso o sistema registre metadados como endereços IP ou padrões de uso sem o devido controle, tais informações poderiam ser utilizadas para rastrear ou identificar usuários específicos, criando riscos relacionados à exposição de dados pessoais.

A ameaça de ***denial of service*** é uma preocupação fundamental, pois o sistema depende da infraestrutura de serviços externos para processar as requisições e gerar mídia.

Como o processamento de imagens e GIFs pode ser computacionalmente intensivo, um volume excessivo de requisições simultâneas pode sobrecarregar os servidores, tornando o serviço indisponível para usuários legítimos. Além disso, sem mecanismos de controle, o sistema se torna vulnerável a ataques distribuídos de negação de serviço (DDoS), nos quais um invasor pode explorar falhas na arquitetura para forçar a interrupção temporária do funcionamento da plataforma. Caso isso ocorra, tanto a experiência do usuário quanto a confiabilidade do serviço seriam severamente prejudicadas.

Por fim, a ameaça de ***elevation of privilege*** representa um risco crítico relacionado à possibilidade de exploração de vulnerabilidades na infraestrutura do sistema. Como os agentes de IA são executados em containers Docker, qualquer configuração inadequada pode permitir que um invasor obtenha permissões administrativas ou acesso indevido aos serviços internos. Caso um atacante consiga explorar falhas no isolamento dos containers, ele poderia comprometer toda a infraestrutura, modificando os modelos de IA, acessando dados internos do sistema ou até mesmo assumindo o controle do ambiente de execução. Essa vulnerabilidade se torna ainda mais preocupante na ausência de restrições rígidas de acesso e controle sobre os processos em execução dentro dos containers.

3 ESTRATÉGIAS DE MITIGAÇÃO DOS RISCOS

Diante das ameaças identificadas na análise de riscos do sistema, diversas estratégias de mitigação poderiam ser adotadas para fortalecer a segurança, integridade e disponibilidade da plataforma. Considerando a arquitetura do sistema, é essencial a implementação de medidas que protejam tanto a comunicação entre os componentes quanto os próprios processos internos da aplicação.

Para mitigar o risco de *spoofing*, seria necessário implementar um mecanismo de autenticação de usuários antes que qualquer requisição seja processada pelo sistema. Uma abordagem eficaz seria a adoção de um sistema de autenticação baseado em *tokens*, como JSON Web Tokens (JWT), onde cada usuário receberia um *token* único após um processo de verificação, garantindo que apenas usuários legítimos possam interagir com o serviço. Além disso, para evitar a falsificação de requisições, seria fundamental a implementação de medidas como restrição de CORS (Cross-Origin Resource Sharing) e proteção contra ataques de injeção de *scripts* (XSS), garantindo que somente requisições legítimas sejam processadas pelo *front-end*.

O risco de *tampering* poderia ser mitigado por meio da criptografia de ponta a ponta na comunicação entre o *front-end* e os agentes de IA. O uso do protocolo HTTPS (Hypertext Transfer Protocol Secure) garantiria que os dados trafegassem de forma segura pela rede, evitando que fossem interceptados e adulterados por terceiros. Além disso, a implementação de assinaturas digitais nas requisições poderia assegurar que qualquer tentativa de modificação dos dados no meio do caminho seja detectada e rejeitada pelo sistema. Outra medida essencial seria a validação rigorosa das entradas dos usuários, aplicando técnicas como listas de permissões (*whitelisting*) para garantir que apenas comandos válidos sejam processados pelos modelos de IA, prevenindo assim a inserção de conteúdo inesperado ou prejudicial.

Para reduzir a vulnerabilidade associada à ameaça de *repudiation*, seria fundamental a implementação de um sistema de *logs* detalhado e seguro, armazenando todas as interações realizadas pelos usuários de maneira que possam ser auditadas posteriormente. Esses *logs* deveriam conter metadados essenciais, como identificadores das requisições, carimbos de data e hora e identificadores anônimos dos usuários, garantindo que todas as ações possam ser rastreadas sem comprometer a privacidade dos indivíduos. Além disso, os *logs* deveriam ser protegidos contra modificações e exclusões não autorizadas, utilizando um sistema de

controle de acesso baseado em privilégios e armazenamento criptografado para evitar manipulações indevidas dos registros.

A mitigação do risco de *information disclosure* exigiria a implementação de medidas rigorosas de proteção de dados. Primeiramente, seria crucial evitar o armazenamento desnecessário de informações dos usuários e, caso fosse necessário manter registros temporários para fins operacionais, esses dados deveriam ser anonimizados antes do armazenamento. A aplicação de técnicas de minimização de dados garantiria que apenas informações essenciais fossem processadas, reduzindo a superfície de exposição a ataques. Além disso, a configuração de controles de acesso apropriados nos servidores garantiria que apenas componentes autorizados do sistema pudessem acessar determinados dados, prevenindo acessos indevidos. A utilização de técnicas de encriptação, tanto em repouso quanto em trânsito, complementaria essa estratégia, garantindo que, mesmo em caso de comprometimento da infraestrutura, os dados permaneçam protegidos.

Para reduzir a ameaça de *denial of service*, seria necessário adotar medidas de controle de carga e proteção contra requisições excessivas. A implementação de um sistema de r*ate limiting* permitiria restringir a quantidade de requisições feitas por um único usuário em um determinado período, prevenindo a sobrecarga da infraestrutura. Além disso, a introdução de um balanceador de carga distribuiria eficientemente as requisições entre diferentes instâncias do sistema, garantindo uma melhor escalabilidade e reduzindo o risco de indisponibilidade. Outra estratégia importante seria a implementação de mecanismos de detecção de ataques DDoS, como *firewalls* de aplicação *web* (WAFs) e serviços especializados que possam identificar padrões suspeitos de tráfego e bloquear automaticamente requisições maliciosas antes que afetem a disponibilidade do sistema.

A mitigação do risco de *elevation of privilege* exigiria um controle rigoroso sobre a configuração dos containers Docker responsáveis pela execução dos agentes de IA. Para evitar a exploração de vulnerabilidades no ambiente de execução, os containers deveriam ser configurados com permissões mínimas, impedindo que processos tenham acesso desnecessário ao sistema operacional ou a outros serviços internos. A implementação do princípio de menor privilégio (PoLP) garantiria que cada componente do sistema operasse apenas com as permissões estritamente necessárias para sua funcionalidade. Além disso, a adoção de mecanismos de isolamento, como *namespaces* e *cgroups* no Docker, ajudaria a evitar que um ataque em um container comprometesse a integridade de outros serviços. Também seria essencial a utilização de imagens de containers verificadas e constantemente atualizadas, prevenindo a exploração de falhas conhecidas.

Dessa forma, a implementação dessas estratégias de mitigação fortaleceria significativamente a segurança e confiabilidade do sistema, garantindo que ele possa operar de maneira resiliente diante de ameaças diversas. A combinação de autenticação robusta, comunicação segura, controle de acesso rigoroso, proteção contra ataques e boas práticas na configuração da infraestrutura reduziria consideravelmente os riscos identificados na análise de ameaças, permitindo um ambiente mais seguro e eficiente para os usuários e operadores do sistema.

4 CONCLUSÃO

A análise de riscos realizada neste Relatório de Impacto à Proteção de Dados Pessoais permitiu identificar ameaças significativas associadas à Inteligência Artificial Distribuída desenvolvida. A metodologia STRIDE possibilitou uma avaliação detalhada das vulnerabilidades do sistema, abrangendo aspectos como *spoofing*, *tampering*, *repudiation*, *information disclosure*, *denial of service* e *elevation of privilege*. E a implementação de medidas de segurança seria essencial para garantir a proteção dos dados dos usuários e a confiabilidade da solução.

A ausência de um mecanismo de autenticação adequada expõe o sistema a ataques de falsificação de identidade (*spoofing*), enquanto a comunicação entre os serviços, se não for devidamente protegida, pode permitir alterações não autorizadas nos dados em trânsito (*tampering*). A falta de registros auditáveis dificulta a rastreabilidade de ações no sistema (*repudiation*), e o armazenamento indevido de informações pode resultar na exposição de dados sensíveis dos usuários (*information disclosure*). Ademais, a plataforma precisa estar preparada para resistir a ataques de negação de serviço (*denial of service*) e evitar que falhas na infraestrutura permitam a obtenção de acessos privilegiados indevidos (*elevation of privilege*).

Diante desse contexto, foram propostas estratégias eficazes para mitigar tais riscos, incluindo a implementação de autenticação baseada em *tokens* (JWT), criptografia de comunicação via HTTPS, armazenamento seguro e auditável de *logs*, anonimização de dados sensíveis e controle de acessos rigoroso. Além disso, mecanismos como *rate limiting*, balanceamento de carga e monitoramento ativo de ataques DDoS seriam fundamentais para manter a disponibilidade e o desempenho do sistema.

A segurança da informação é um desafio contínuo, exigindo revisão e aprimoramento constante das medidas implementadas. A adoção das soluções apresentadas neste relatório contribuiria para a maior proteção aos dados processados pelo sistema, minimizando riscos e assegurando a conformidade com princípios de privacidade e segurança digital.

5 REFERÊNCIAS

OWASP. **Threat Modeling Process**. Disponível em: https://owasp.org/www-community/Threat_Modeling_Process. Acesso em: 05 fev. 2025.