

UNIVERSIDADE FEDERAL DE SANTA CATARINA
CAMPUS REITOR JOÃO DAVID FERREIRA LIMA
PROGRAMA DE GRADUAÇÃO EM SISTEMAS DE INFORMAÇÃO

Francisco Camerini

FERRAMENTA PARA ANÁLISE AUTOMÁTICA DE POLÍTICAS DE PRIVACIDADE COM BASE NA LEI GERAL DE PROTEÇÃO DE DADOS

Florianópolis, Santa Catarina – Brasil
2026

Francisco Camerini

FERRAMENTA PARA ANÁLISE AUTOMÁTICA DE POLÍTICAS DE PRIVACIDADE COM BASE NA LEI GERAL DE PROTEÇÃO DE DADOS

Trabalho de Conclusão de Curso submetido ao Programa de Graduação em Sistemas de Informação da Universidade Federal de Santa Catarina para a obtenção do Grau de Bacharel em Sistemas de Informação.
Orientador: José Eduardo de Lucca, Dr.

Florianópolis, Santa Catarina – Brasil
2026

Catalogação na fonte pela Biblioteca Universitária da Universidade Federal de Santa Catarina.
Arquivo compilado às 23:20h do dia 10 de julho de 2026.

Francisco Camerini
Ferramenta para análise automática de políticas de privacidade com base na Lei Geral de Proteção de Dados / Francisco Camerini; Orientador, José Eduardo de Lucca, Dr. – Florianópolis, Santa Catarina – Brasil, 30 de junho de 2026.
82 p.

Trabalho de Conclusão de Curso – Universidade Federal de Santa Catarina, INE – Departamento de Informática e Estatística, CTC – Centro Tecnológico, Programa de Graduação em Sistemas de Informação.

Inclui referências

1. Lei Geral de Proteção de Dados, 2. Políticas de Privacidade, 3. Inteligência Artificial, 4. Modelos de Linguagem, 5. Extensão de Navegador, I. José Eduardo de Lucca, Dr. II. Programa de Graduação em Sistemas de Informação III. Ferramenta para análise automática de políticas de privacidade com base na Lei Geral de Proteção de Dados

CDU 02:141:005.7

Francisco Camerini

FERRAMENTA PARA ANÁLISE AUTOMÁTICA DE POLÍTICAS DE PRIVACIDADE COM BASE NA LEI GERAL DE PROTEÇÃO DE DADOS

Este(a) Trabalho de Conclusão de Curso foi julgado adequado(a) para obtenção do Título de Bacharel em Sistemas de Informação, e foi aprovado em sua forma final pelo Programa de Graduação em Sistemas de Informação do INE – Departamento de Informática e Estatística, CTC – Centro Tecnológico da Universidade Federal de Santa Catarina.

Florianópolis, Santa Catarina – Brasil, 30 de junho de 2026.

José Eduardo de Lucca, Dr.
Coordenador(a) do Programa de
Graduação em Sistemas de Informação

Banca Examinadora:

José Eduardo de Lucca, Dr.
Orientador
Universidade Federal de Santa
Catarina – UFSC

Patrícia Vilain, Dra.
Universidade Federal de Santa Catarina –
UFSC

Elder Rizzon Santos, Dr.
Universidade Federal de Santa Catarina –
UFSC

Dedico este trabalho ao meu avô, Sergio.

AGRADECIMENTOS

Agradeço todo o corpo docente da UFSC, em especial meu professor orientador José de Lucca, por todo aprendizado transmitido ao longo desse anos.

Agradeço todos os meus colegas de curso e amigos pelo companheirismo e momentos de leveza muito necessários.

Agradeço meus pais Sergio e Claudia por proporcionarem o privilégio de uma educação de qualidade desde sempre; minhas irmãs Fernanda e Eduarda por terem trilhado esse caminho antes de mim, minhas maiores fontes de inspiração.

Por fim, agradeço a vida por me presentear diariamente com a oportunidade de aprender algo novo.

“Quando partires em viagem para Ítaca
faz votos para que seja longo o caminho,
pleno de aventuras, pleno de conhecimentos”
Konstantinos Kaváfis

RESUMO

Este trabalho desenvolveu uma ferramenta automatizada capaz de analisar políticas de privacidade de sites e notificar seus usuários sobre potenciais inconformidades com a Lei Geral de Proteção de Dados Pessoais. Para isso, foi projetada uma extensão para navegadores web que realiza a coleta automática desses documentos, enviando-os a uma *API* responsável pelo processamento do texto utilizando um modelo de linguagem de grande escala. O método adotado combinou técnicas de *web scraping*, processamento de linguagem natural e *prompt engineering* para reconhecer omissões, ambiguidades e desconformidades com os princípios da legislação brasileira. Foi construído um dataset original com 200 políticas de privacidade brasileiras, utilizado tanto em um benchmark comparativo entre oito modelos de linguagem quanto no ajuste fino supervisionado, com a técnica LoRA, do modelo Qwen 2.5 14B. O modelo resultante atingiu F1 macro de 0,8409 no conjunto de validação, superando o critério de aceitação estabelecido a priori e todos os modelos proprietários avaliados no benchmark. Os resultados demonstram a viabilidade da automatização da análise de conformidade legal, contribuindo para a promoção de maior transparência nas práticas de tratamento de dados.

Palavras-chaves: Lei Geral de Proteção de Dados. Políticas de Privacidade. Inteligência Artificial. Modelos de Linguagem. Extensão de Navegador.

ABSTRACT

This work developed an automated tool capable of analyzing website privacy policies and notifying users of potential nonconformities with the Brazilian General Data Protection Law. To achieve this, a web browser extension was designed to automatically collect these documents and send them to an API responsible for processing the text using a large language model. The adopted method combined web scraping, natural language processing, and prompt engineering techniques to identify omissions, ambiguities, and inconsistencies with the principles established by Brazilian legislation. An original dataset of 200 Brazilian privacy policies was built and used both in a comparative benchmark across eight language models and in the supervised fine-tuning, with the LoRA technique, of the Qwen 2.5 14B model. The resulting model achieved a macro F1 score of 0.8409 on the validation set, surpassing the acceptance criterion established a priori and every proprietary model evaluated in the benchmark. The results demonstrate the feasibility of automating legal compliance analysis, contributing to greater transparency in data processing practices.

Keywords: General Data Protection Law. Privacy Policies. Artificial Intelligence. Language Models. Browser Extension.

LISTA DE FIGURAS

Figura 1 – Exemplo de opção padrão invasiva 19
Figura 2 – Exemplo de obscuridade informacional 19
Figura 3 – Fluxo de dados entre extensão e API 53
Figura 4 – Pipeline de validação e processamento da API 55
Figura 5 – Estado inicial do popup, aguardando ação do usuário 58
Figura 6 – Estados do popup após acionamento pelo usuário 59

LISTA DE TABELAS

Tabela 1 – Estados da Matriz de Confusão aplicados à Análise de Políticas de Privacidade 32
Tabela 2 – Distribuição do dataset por setor 42
Tabela 3 – Resultados do benchmark com prompt simples 46
Tabela 4 – Resultados do benchmark com few-shot prompting 47
Tabela 5 – Evolução da *loss* de validação por época nos experimentos de ajuste fino 49
Tabela 6 – Hiperparâmetros da configuração final de ajuste fino 50
Tabela 7 – Evolução do desempenho do Qwen 2.5 14B ao longo dos experimentos 51

LISTA DE ABREVIATURAS E SIGLAS

API	Interface de Programação de Aplicações
DOM	Modelo de Objeto de Documento
HTTP	Protocolo de Transferência de Hipertexto
IA	Inteligência Artificial
LGPD	Lei Geral de Proteção de Dados Pessoais
LLM	Modelo de Linguagem de Grande Escala
TCU	Tribunal de Contas da União

SUMÁRIO

1	**INTRODUÇÃO**	**15**
1.1	OBJETIVO GERAL	15
1.2	OBJETIVOS ESPECÍFICOS	15
1.3	JUSTIFICATIVA	16
2	**FUNDAMENTAÇÃO TEÓRICA**	**17**
2.1	PRIVACIDADE E PROTEÇÃO DE DADOS	17
2.1.1	**Vazamentos de Dados**	**17**
2.1.2	**Capitalismo de Vigilância**	**18**
2.2	LEI GERAL DE PROTEÇÃO DE DADOS	20
2.2.1	**Contexto Histórico**	**20**
2.2.2	**Definição de Papéis**	**21**
2.2.3	**Dado Pessoal vs Dado Sensível**	**21**
2.2.4	**Princípios**	**22**
2.2.5	**Impactos**	**23**
2.2.6	**Críticas**	**24**
2.3	POLÍTICAS DE PRIVACIDADE	24
2.4	MODELOS DE LINGUAGEM DE GRANDE ESCALA	25
2.4.1	**Limitações**	**26**
2.4.2	**Engenharia de Prompt**	**27**
2.4.3	**Fine-Tuning**	**28**
2.4.3.1	Modalidades de Ajuste Fino	29
2.4.3.2	Dataset e Origem dos Dados	29
2.4.3.3	Infraestrutura de Treinamento	30
2.4.3.4	Riscos e Mitigações	30
2.4.4	**Análise de Texto com PLN Tradicional vs LLMs**	**30**
2.4.5	**Métricas de Avaliação**	**31**
3	**TRABALHOS RELACIONADOS**	**33**
3.0.1	**P3P**	**33**
3.0.2	**ToS;DR**	**33**
3.0.3	**Abordagem Baseada em Grafos de Conhecimento**	**34**
3.0.4	**Abordagem Baseada em LLM**	**34**
3.1	EXTENSÕES DE NAVEGADOR	35
3.2	ENGENHARIA DE PROMPT EM LLMS: UMA REVISÃO SISTEMÁTICA	35
3.2.1	**LoRA: Adaptação Eficiente de Modelos de Linguagem**	**36**
3.3	SÍNTESE	36

4 METODOLOGIA 38
4.1 REQUISITOS DE TRANSPARÊNCIA AVALIADOS 38
4.2 CONSTRUÇÃO DO DATASET 40
4.2.1 Curadoria das Políticas de Privacidade 40
4.2.2 Silver Labeling com Claude Opus 4.6 41
4.2.3 Revisão Manual e Validação dos Rótulos 42
4.3 DEFINIÇÃO OPERACIONAL DE CONFORMIDADE 42
4.4 CRITÉRIO DE ACEITAÇÃO 43
4.5 EXPERIMENTOS E BENCHMARK COMPARATIVO 44
4.5.1 Modelos Avaliados 44
4.5.2 Protocolo de Avaliação 45
4.5.3 Resultados do Benchmark 46
4.5.4 Seleção do Modelo para Fine-Tuning 47
4.6 PROCESSO DE AJUSTE FINO 48
4.6.1 Infraestrutura e Configuração do Ambiente 48
4.6.2 Configuração do LoRA e Exploração de Hiperparâmetros . . . 48
4.6.3 Formato dos Dados de Treinamento 50
4.6.4 Resultados do Ajuste Fino 51

5 DESENVOLVIMENTO 52
5.1 ARQUITETURA DO SISTEMA 52
5.2 BACKEND 53
5.2.1 Endpoints 53
5.2.2 Pipeline de Processamento 54
5.2.3 Módulo de Análise 55
5.2.4 Cache e Persistência 56
5.3 EXTENSÃO DE NAVEGADOR 56
5.3.1 Detecção Automática de Políticas 56
5.3.2 Service Worker 57
5.3.3 Interface do Popup 58

6 CONCLUSÃO 60
6.1 RESULTADOS 60
6.2 LIMITAÇÕES 61
6.3 TRABALHOS FUTUROS 62
6.3.1 Padronização da Localização de Políticas de Privacidade . . . 62
6.3.2 Expansão do Dataset 62
6.3.3 Validação Jurídica dos Rótulos 63
6.3.4 Revisão do Esquema de Categorização 63
6.3.5 API como Apoio à Elaboração de Políticas de Privacidade . . . 63

REFERÊNCIAS 64

APÊNDICE A – PROMPT DO CENÁRIO DE PROMPT SIMPLES 70

APÊNDICE B – PROMPT DO CENÁRIO FEW-SHOT 71

APÊNDICE C – INSTRUÇÕES DE ACESSO AO CÓDIGO-FONTE 73

APÊNDICE D – ARTIGO PRODUZIDO A PARTIR DESTE TRABALHO 75

1 INTRODUÇÃO

A crescente digitalização de serviços e o uso intensivo de dados pessoais em plataformas online tornaram a privacidade um importante tema de debate na sociedade contemporânea. A expansão do volume de informações sensíveis sobre indivíduos geradas, exige mecanismos cada vez mais robustos de proteção e governança. Nesse cenário, as políticas de privacidade surgem como instrumentos por meio dos quais organizações informam aos usuários como seus dados são coletados, armazenados e utilizados.

Entretanto, apesar da obrigatoriedade e da importância desses documentos, como é constatato por Fabian, Ermakova e Lentz (2017), muitos deles ainda são redigidos de forma ambígua, incompleta ou excessivamente técnica, dificultando a compreensão por parte do usuário. Essa realidade evidencia a necessidade de soluções que permitam verificar de maneira prática e automatizada se uma política de privacidade atende aos princípios legais de transparência, adequação e necessidade.

Com o avanço das *LLM*s, tornou-se possível processar textos complexos e interpretar linguagem natural em alto nível, o que viabiliza a análise automatizada de documentos jurídicos e normativos.

Diante desse contexto, este trabalho propõe o desenvolvimento de uma extensão para navegadores web que realize o *web scraping* automático de políticas de privacidade, enviando-as a uma *API* que processa o texto utilizando uma *LLM* a fim de identificar potenciais disconformidades com a LGPD. A ferramenta busca oferecer uma solução prática e tecnicamente fundamentada para apoiar o cumprimento da legislação de proteção de dados pessoais no Brasil.

1.1 OBJETIVO GERAL

Desenvolver uma ferramenta computacional, na forma de uma extensão para navegadores web, capaz de analisar automaticamente políticas de privacidade de sites e indicar possíveis inconformidades com a LGPD.

1.2 OBJETIVOS ESPECÍFICOS

- Implementar uma extensão compatível com navegadores modernos, capaz de localizar e extrair políticas de privacidade de páginas web por meio de técnicas de *web scraping*.
- Desenvolver uma *API HTTP* para comunicação entre a extensão e a *LLM* responsável pela análise, garantindo processamento seguro e eficiente dos textos coletados.

- Realizar a análise semântica das políticas de privacidade utilizando uma *LLM*, baseando-se nos princípios e obrigações definidos pela LGPD.
- Avaliar a eficácia da solução por meio de testes com diferentes políticas de privacidade, analisando a capacidade do sistema em reconhecer inconsistências legais.

1.3 JUSTIFICATIVA

A promulgação da LGPD representou um avanço significativo na regulação do uso de informações pessoais no Brasil. Contudo, a mera existência da lei não garante sua efetividade. Diversas organizações, dos setores público e privado, ainda enfrentam dificuldades em implementar práticas compatíveis com as exigências legais. Segundo TCU, 28,2% das organizações públicas ainda não adotaram medidas básicas de adequação (Tribunal de Contas da União, 2025). No setor privado, estima-se que apenas 32% das organizações tenham implementado planos de *compliance* com a LGPD (Cetic, 2024).

Nesse contexto, uma ferramenta de análise automatizada pode desempenhar papel relevante na disseminação da cultura de privacidade e na fiscalização do cumprimento da LGPD. Além de seu valor prático, o projeto contribui para o avanço da pesquisa em aplicações de IA no campo jurídico, explorando o potencial dos *LLM*s para verificação de conformidade documental.

2 FUNDAMENTAÇÃO TEÓRICA

Esta revisão teórica fornece o suporte conceitual necessário para compreender o problema e fundamentar as escolhas metodológicas adotadas na construção da ferramenta. A análise de políticas de privacidade por meio de inteligência artificial situa-se na interseção de duas grandes áreas, a esfera jurídica, que define as obrigações de proteção de dados, e a esfera computacional, que provê os avanços tecnológicos necessários para processar volumes massivos de texto legal.

2.1 PRIVACIDADE E PROTEÇÃO DE DADOS

Para compreender a relevância de uma ferramenta de análise de conformidade, é necessário primeiro entender o objeto de proteção, a privacidade. O conceito evoluiu significativamente desde a definição clássica de Warren e Brandeis (1890) como o direito de ser deixado em paz. Na sociedade da informação, a privacidade se transformou com o conceito de autodeterminação informativa, entendido como o direito do cidadão de controlar o fluxo de suas próprias informações (Doneda, 2011).

A privacidade no ambiente digital é constantemente ameaçada, seja por políticas internas de tratamento de dados desconformes com a legislação, impulsionadas pelo conceito do “Capitalismo de Vigilância", ou por ataques de terceiros à infraestrutura de controladores de dados para usos indevidos com fins econômicos. Estas duas problemáticas serão exploradas nos tópicos subsequentes.

2.1.1 Vazamentos de Dados

Um vazamento de dados, é qualquer incidente de segurança que resulte na exposição, acesso não autorizado, destruição, perda ou divulgação indevida de informações pessoais. Vazamentos geram consequências como fraudes, discriminação, extorsão, danos reputacionais ou uso indevido para fins econômicos.

Além de erros técnicos, muitos vazamentos ocorrem devido a falhas organizacionais, como configurações inadequadas de servidores, ausência de criptografia ou permissões de acesso excessivamente amplas. O maior vazamento de dados da história, conhecido como *Mother of all breaches*, expôs 26 bilhões de registros de múltiplas plataformas digitais; as informações vieram principalmente de três fontes: malwares, ataques de força bruta com senhas reutilizadas e reaproveitamento de dados de vazamentos anteriores (Petkauskas, 2024). Um caso emblemático no contexto brasileiro ocorreu em 2021, quando uma base contendo nome completo, CPF, dados demográficos e informações financeiras de mais de 223 milhões de brasileiros foi disponibilizada para venda na *deep web* (Belli, 2021).

Esse fenômeno evidencia a necessidade de uma legislação que exija políticas de

segurança robustas, auditorias frequentes e mecanismos de resposta a incidentes.

2.1.2 Capitalismo de Vigilância

A internet comercial moderna opera sob uma lógica econômica denominada como "Capitalismo de Vigilância", (Zuboff, 2023). Neste modelo, a experiência humana é reivindicada como matéria-prima gratuita para tradução em dados comportamentais. Esses dados não são usados apenas para melhoria de serviços, mas para a criação de produtos preditivos que antecipam e, muitas vezes, manipulam o comportamento do usuário.

Essa manipulação não é teórica, mas documentada. Um exemplo emblemático é o uso de *targeting* político, evidenciado no escândalo da Cambridge Analytica, onde perfis comportamentais construídos a partir de dados de redes sociais foram utilizados para explorar vieses cognitivos e modular a intenção de voto de eleitores indecisos nas eleições presidenciais americanas de 2016 (Cadwalladr; Graham-Harrison, 2018). Outra exemplo é a exploração de estados emocionais, relatórios vazados do Facebook indicaram a capacidade da plataforma de identificar momentos em que adolescentes se sentiam "inseguros", "ansiosos"ou "inúteis", permitindo que anunciantes otimizassem o momento exato para influenciar decisões de consumo baseadas em vulnerabilidade psicológica (Tiku, 2017).

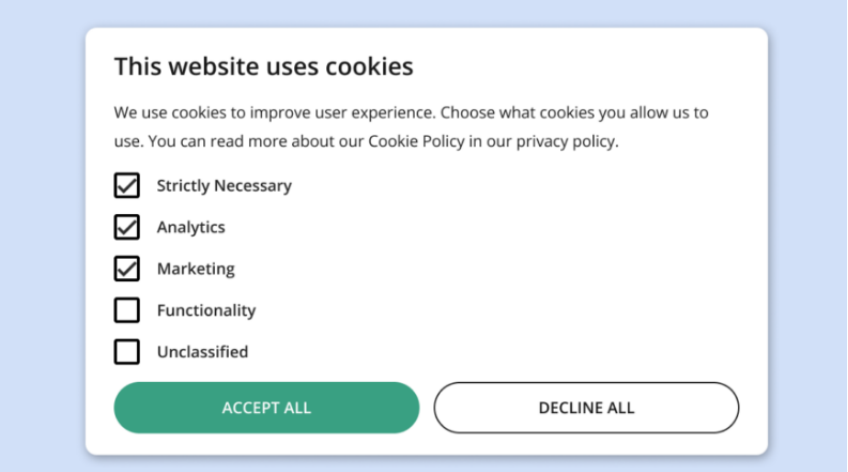
Uma prática que contribui para a intensificação desse conceito são os chamados *Dark Patterns*. Esses são um conjunto de técnicas de design de interface deliberadamente elaboradas para induzir ou enganar o usuário a tomar decisões que favorecem os interesses da plataforma, muitas vezes em detrimento de sua autonomia e privacidade. Diferentemente de falhas de usabilidade, os *Dark patterns* exploram vieses cognitivos conhecidos, como aversão à perda, pressão temporal, sobrecarga informacional e tendência à opção padrão, para conduzir o comportamento do indivíduo sem que este perceba a extensão da influência exercida (Brignull, 2015).

Na esfera da proteção de dados, isso é especialmente relevante porque escolhas aparentemente simples podem resultar na concessão involuntária de permissões amplas de coleta, no compartilhamento de informações sensíveis ou na dificuldade artificial de revogar consentimentos.

Como exemplos clássicos de *Dark Patterns* podemos citar:

- **Opção padrão invasiva**: caixas de seleção previamente marcadas ou botões primários autorizando rastreamento, marketing direcionado ou compartilhamento com terceiros, exigindo que o usuário ativamente remova permissões que deveriam ser concedidas de forma livre e informada, como ilustrado na Figura 1.

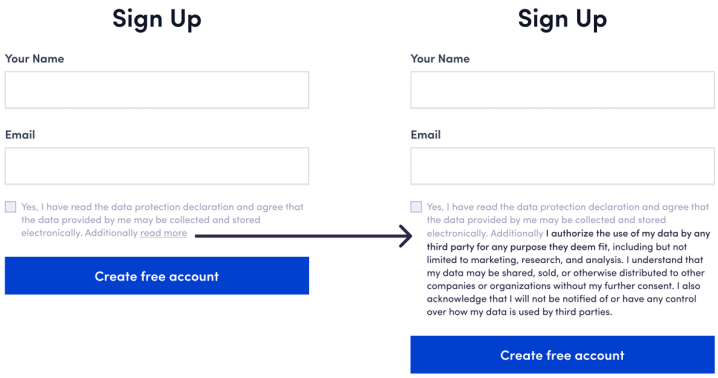
Figura 1 – Exemplo de opção padrão invasiva



Cookies de rastreamento pré selecionados
(Imagem capturada pelo autor)

- **Obscuridade informacional**: informações essenciais, como compartilhamento com parceiros comerciais, finalidade secundária ou período de retenção, são disfarçadas em textos longos, pouco legíveis ou em links com baixa visibilidade, como exemplificado na Figura 2.

Figura 2 – Exemplo de obscuridade informacional



Informação sobre autorização de uso de dados pessoais escondida
(Reprodução/Stefani (2023))

- **Labirintos de cancelamento**: configurações de privacidade fáceis de ativar, mas extremamente difíceis de desativar, distribuídas em múltiplos menus ou formulários extensos, criando barreiras de fricção comportamental.

Essas práticas levam a um conceito chamado Paradoxo da Informação, que descreve a aparente contradição entre o comportamento declarado e o comportamento efetivo dos usuários no que diz respeito à proteção de seus dados pessoais. Embora pesquisas mostrem que usuários expressam elevada preocupação com sua privacidade e com o uso indevido de suas informações, suas ações cotidianas, como aceitar

termos extensos sem leitura, compartilhar dados sensíveis em troca de conveniência ou permitir rastreamento contínuo em aplicativos muitas vezes não refletem esse nível de preocupação (Kokolakis, 2017).

Tudo isso contribui para um cenário de assimetra de informação, onde as empresas sabem tudo sobre um determinado usuário, enquanto o usuário sabe pouco ou nada sobre como seus dados estão sendo processados.

2.2 LEI GERAL DE PROTEÇÃO DE DADOS

A Lei nº 13.709/2018, conhecida como Lei Geral de Proteção de Dados Pessoais, (Brasil, 2018), estabelece as diretrizes para o tratamento de dados pessoais no Brasil. Inspirada principalmente no Regulamento Geral de Proteção de Dados da União Europeia (GDPR), (União Europeia, 2016), a LGPD surgiu em resposta ao aumento do uso de informações pessoais por organizações públicas e privadas, e à necessidade de assegurar transparência, segurança e respeito aos direitos dos titulares de dados.

2.2.1 Contexto Histórico

Antes da promulgação da LGPD, a legislação brasileira não carecia totalmente de proteção a privacidade, mas se caracterizava por uma abordagem consideravlmente fragmentada. A Constituição Federal de 1988 já assegurava, em seu artigo 5º, a inviolabilidade da intimidade e da vida privada, servindo como o princípio base para a proteção de dados (Brasil, 1988).

Fora do âmbito constitucional, o Código de Defesa do Consumidor (Lei nº 8.078/1990) foi responsável por regular a formação de bancos de dados de consumo, garantindo ao consumidor o direito de acesso e retificação de informações a seu respeito (Brasil, 1990). Outras normas, como a Lei de Acesso à Informação (Lei nº 12.527/2011), (Brasil, 2011b), e a Lei do Cadastro Positivo (Lei nº 12.414/2011), (Brasil, 2011a), tangenciavam o tema criando um espécie de *framework* legal que, embora existente, com o aumento e a complexidade crescente do tratamento massivo de dados na era digital, tornou-se insuficiente com o passar do tempo.

O grande marco para a regulação do ambiente digital ocorreu com a aprovação do Marco Civil da Internet (Lei nº 12.965/2014). Impulsionado pelas revelações de espionagem global feitas por Edward Snowden, o Marco Civil estabeleceu princípios, garantias, direitos e deveres para o uso da internet no Brasil, definindo a proteção da privacidade e dos dados pessoais como condição para o pleno exercício da cidadania digital (Brasil, 2014).

Por mais essencial que tenha sido no momento de sua promulgação, o Marco Civil da Internet definiu princípios a serem seguidos, porém não detalhou na prática como o tratamento de dados deveria ocorrer. A ausência de uma autoridade fiscalizadora e de regras específicas para o setor privado como um todo gerava insegurança jurídica.

Com esse cenário, somado ao aumento dos debates internacionais sobre o tema, haja vista a promulgação da GDPR na Europa, acelerou-se a discussão do tema entre sociedade civil e legislativo, culminando na sanção da LGPD em 18 de agosto de 2018.

2.2.2 Definição de Papéis

A LGPD define papéis específicos dentro do ecossistema de tratamento de dados. O objetivo dessa estrutura é delimitar responsabilidades, identificar obrigações legais e garantir a responsabilização adequada em caso de incidentes ou violações. Compreender estes papéis é essencial para o entendimento da legislação.

- **Titular**: Pessoa natural cujos dados pessoais são objeto de tratamento. A lógica da lei é construída visando a proteção dos direitos desse indivíduo.
- **Controlador**: Pessoa natural ou jurídica responsável pelas decisões relativas ao tratamento dos dados pessoais. É o principal responsável pela conformidade com a LGPD e deve demonstrar que adota medidas administrativas, técnicas e organizacionais adequadas.
- **Operador**: Pessoa natural ou jurídica que realiza o tratamento de dados em nome do controlador.

2.2.3 Dado Pessoal vs Dado Sensível

A LGPD estabelece, em seu artigo 5º, uma distinção entre diferentes categorias de informações, especificando obrigações e salvaguardas apropriadas para cada tipo de dado. Essa diferenciação é essencial para a interpretação da lei e para a elaboração de políticas de privacidade, uma vez que define níveis distintos de risco e exigências de proteção.

O conceito de dado pessoal é definido como qualquer informação relacionada à pessoa natural identificável. Inclui desde elementos diretos de identificação, como nome, CPF, RG e endereço, até identificadores indiretos, como dados de localização, hábitos de navegação ou características comportamentais. A amplitude dessa definição reflete a ideia de que a individualização de um titular de dados pode ocorrer mesmo a partir de fragmentos de informação, especialmente em ambientes digitais em que técnicas de correlação e perfilamento permitem a identificação de usuários.

Por outro lado, a lei prevê uma categoria especial de dados, conhecidos como dado pessoal sensível, englobando informações que, por sua natureza, apresentam maior potencial de discriminação ou violação de direitos fundamentais. Entre esses dados estão informações sobre origem racial ou étnica, convicção religiosa, opinião política, filiação sindical, saúde, vida sexual, biometria e genética. A razão para essa classificação diferenciada está associada ao impacto potencial que o uso inadequado

desses dados pode gerar na dignidade, autonomia e igualdade do titular, justificando um regime jurídico mais rigoroso.

No contexto do trabalho, essa distinção implica que o controlador de dados deve informar de forma clara nas políticas de privacidade não apenas quais dados são coletados, mas também se algum deles pertence à categoria sensível, especificando finalidade, justificativa legal e mecanismos de proteção associados.

2.2.4 Princípios

A LGPD rompeu com a lógica na qual os dados pessoais eram tratados como ativos de propriedade exclusiva das organizações que os coletavam. A legislação inaugurou um regime de direitos fundamentais, consolidando o entendimento de que o dado pessoal é uma projeção da personalidade do indivíduo, sobre a qual ele deve ter o direito de exercer controle.

Para operacionalizar esse controle, a lei é estruturada em torno de dez princípios definidos no seu artigo 6º:

- **Finalidade**: Realização do tratamento para propósitos legítimos, específicos, explícitos e informados ao titular, sem possibilidade de tratamento posterior de forma incompatível com essas finalidades.
- **Adequação**: Compatibilidade do tratamento com as finalidades informadas ao titular, de acordo com o contexto do tratamento.
- **Necessidade**: Limitação do tratamento ao mínimo necessário para a realização de suas finalidades, com abrangência dos dados pertinentes, proporcionais e não excessivos.
- **Livre acesso**: Garantia, aos titulares, de consulta facilitada e gratuita sobre a forma e a duração do tratamento, bem como sobre a integralidade de seus dados pessoais.
- **Qualidade dos dados**: Garantia, aos titulares, de exatidão, clareza, relevância e atualização dos dados, de acordo com a necessidade e para o cumprimento da finalidade de seu tratamento.
- **Transparência**: Garantia, aos titulares, de informações claras, precisas e facilmente acessíveis sobre a realização do tratamento e os respectivos agentes de tratamento, observados os segredos comercial e industrial.
- **Segurança**: Utilização de medidas técnicas e administrativas aptas a proteger os dados pessoais de acessos não autorizados e de situações acidentais ou ilícitas de destruição, perda, alteração, comunicação ou difusão.

- **Prevenção**: Adoção de medidas para prevenir a ocorrência de danos em virtude do tratamento de dados pessoais.
- **Não discriminação**: Impossibilidade de realização do tratamento para fins discriminatórios ilícitos ou abusivos.
- **Responsabilização e prestação de contas**: Demonstração, pelo agente, da adoção de medidas eficazes e capazes de comprovar a observância e o cumprimento das normas de proteção de dados pessoais e, inclusive, da eficácia dessas medidas.

No contexto deste trabalho, destacam-se mais os princípios da finalidade, necessidade e transparência. A transparência é o pilar que justifica a existência e a obrigatoriedade de Políticas de Privacidade claras.

A violação desses princípios não é apenas uma falha ética, mas uma infração legal. Para assegurar o cumprimento da norma, a legislação instituiu a Autoridade Nacional de Proteção de Dados (ANPD).

A ANPD é o órgão da administração pública federal responsável por zelar pela proteção de dados pessoais, fiscalizar e aplicar sanções em caso de tratamento de dados realizado em descumprimento com a legislação. Essas sanções podem variar de advertências a multas de até 2% do faturamento da empresa, limitadas a R$ 50 milhões por infração. (Autoridade Nacional de Proteção de Dados, 2023)

2.2.5 Impactos

Embora a LGPD represente um avanço significativo para a segurança jurídica e a privacidade no Brasil, impondo uma mudança cultural corporativa, onde a privacidade é incorporada desde a concepção do produto ou serviço digital, sua aplicação prática enfrenta desafios operacionais. Entre eles podemos citar os custos relacionados à conformidade, com a contratação de ferramentas de segurança da informação e de profissionais especializados, como o Encarregado de Dados/*DPO*, que onera desproporcionalmente micro e pequenas empresas. Podendo gerar maior informalidade de processos em negócios menores, resultando em uma conformidade potencialmente apenas documental, sem eficácia prática na segurança dos dados.

Além do impacto econômico, a implementação da lei desencadeou transformações estruturais na forma como organizações públicas e privadas tratam dados pessoais. Houve a necessidade de mapear fluxos informacionais, revisar contratos com fornecedores e parceiros, estabelecer políticas de retenção e descarte, bem como criar mecanismos de registro e prestação de contas. Essa reconfiguração institucional, ainda que positiva para a transparência e para a responsabilização dos agentes de tratamento, exige tempo, treinamento contínuo das equipes e adaptação de sistemas legados, o que gera complexidades adicionais.

Outro efeito é a intensificação das demandas dos titulares. Com a consolidação dos direitos de acesso, correção, exclusão e portabilidade, empresas passaram a receber solicitações que exigem respostas rápidas, precisas e fundamentadas. Impondo rotinas administrativas adicionais e maior sobrecarga operacional. (Oliveira, 2022)

2.2.6 Críticas

Análises recentes apontam que ANPD permanece em fase de consolidação e enfrenta limitações relacionadas à sua capacidade fiscalizatória, ao volume de demandas, ao quadro reduzido de servidores e à necessidade de aprimorar procedimentos internos. Como consequência, a ausência de padrões técnicos claros e a percepção de baixa efetividade sancionadora são sintomas de um sistema que ainda necessita de mais fortalecimento institucional a fim de garantir a aplicação da LGPD. (Oliveira Rocha Loures, 2024)

A LGPD foi desenhada em um contexto prévio ao advento das IAs generativas, o que cria zonas cinzentas de aplicação. Em julho de 2024 a ANPD determinou a suspensão cautelar de uma alteração na política de privacidade da Meta que definia a autorização para o uso de dados pessoais publicados em suas plataformas para treinamento de sistemas de IA (Brasil. Autoridade Nacional de Proteção de Dados, 2024a). No mês seguinte, a ANPD autorizou a retomada da prática com a exigência de melhor notificação aos titulares dos dados sob a utilização e a facilitação da recusa do uso de dados para essa finalidade. (Brasil. Autoridade Nacional de Proteção de Dados, 2024b). Esse acontecimento ilustra uma postura reativa pragmática da ANPD, diante da ausência de regras específicas para IA definidas na LGPD, optou-se por uma solução de compromisso baseada no aumento da transparência, evitando o bloqueio da inovação tecnológica, mas expondo a insuficiência da legislação atual para regular na raiz a coleta indiscriminada de dados para fins de aprendizado de máquina.

2.3 POLÍTICAS DE PRIVACIDADE

As políticas de privacidade são instrumentos informativos que estabelecem as diretrizes, princípios e responsabilidades relacionados ao tratamento de dados pessoais em uma organização. Sua função central é assegurar transparência e previsibilidade nas operações de coleta, uso, compartilhamento, armazenamento e descarte de informações pessoais, de modo a explicitar aos titulares dos dados o cumprimento dos princípios definidos pela LGPD.

De acordo com Fabian, Ermakova e Lentz (2017), muitas dessas políticas apresentam problemas como excesso de termos técnicos e textos extensos e pouco transparentes, o que dificulta a leitura e limita a capacidade do titular de compreender plenamente o tratamento de seus dados.

O Modelo de Política de Proteção de Dados Pessoais, (Ministério da Gestão e da Inovação em Serviços Públicos, 2024), orienta a estruturação dessas políticas definindo os seguintes fundamentos:

Propósito: explicitação dos objetivos da política e das razões que justificam sua adoção.

Escopo: definição de quem está sujeito às regras da política e quais sistemas, incluindo eventuais exclusões.

Termos e definições: glossário de conceitos técnicos e jurídicos utilizados.

Declarações da política: conjunto de diretrizes que orientam o tratamento dos dados. É neste fundamento que devem estar especificados quais dados pessoais são coletados, junto às demais diretrizes de finalidade, base legal, retenção e compartilhamento, e não no propósito, no escopo ou no glossário de termos, que cumprem função meramente introdutória ou terminológica.

Em termos de forma e linguagem, o modelo orienta que as políticas sejam redigidas de maneira objetiva, em formato normativo, utilizando linguagem técnica precisa e evitando ambiguidades. O documento reforça ainda a importância de revisões periódicas da política.

2.4 MODELOS DE LINGUAGEM DE GRANDE ESCALA

Os Modelos de Linguagem de Grande Escala (*Large Language Models*, LLMs) são sistemas baseados em redes neurais profundas, treinados em imensos volumes de texto, capazes de compreender, gerar e manipular linguagem natural de maneira contextualizada. Seu funcionamento é orientado pela identificação de padrões linguísticos complexos, relações semânticas e estruturas discursivas, permitindo que realizem tarefas como tradução, sumarização, classificação e extração de informações em nível comparável, em certos cenários, ao desempenho humano.

A base conceitual que possibilitou a ascensão desses modelos está na arquitetura *Transformer*, introduzida por Vaswani et al. (2017). O mecanismo de *self-attention* permite que o modelo avalie a relevância relativa de cada palavra em um enunciado considerando todas as demais, sem depender de estruturas sequenciais rígidas. Essa capacidade de capturar relações de longo alcance é particularmente valiosa na análise de documentos complexos, como políticas de privacidade, que frequentemente apresentam condicionantes e dependências semânticas distribuídas ao longo do texto.

Com o avanço dessa arquitetura, emergiram modelos com bilhões ou trilhões de parâmetros, como GPT (OpenAI), Gemini (Google), Claude (Anthropic) e LLaMA (Meta). Tais modelos apresentam uma notável capacidade de generalização e podem ser adaptados a domínios específicos por meio de técnicas como *fine-tuning* e *prompt engineering*. Ainda assim, é importante destacar que, apesar de sua sofisticação, esses sistemas não possuem entendimento jurídico ou conhecimento normativo;

em essência, eles produzem predições estatísticas que soam plausíveis, mas não necessariamente corretas.

2.4.1 Limitações

A aplicação de LLMs ao domínio da proteção de dados e à análise de conformidade normativa exige uma discussão que vá além da descrição técnica de sua arquitetura. Esses modelos carregam limitações fundamentais que decorrem de sua natureza estatística, sendo suscetíveis a erros que podem comprometer tanto a precisão quanto a confiabilidade dos resultados.

Uma limitação crítica é a ocorrência de alucinações, fenômeno em que o modelo gera informações factualmente incorretas, inventa trechos de leis, cita dispositivos inexistentes ou estabelece relações normativas sem respaldo jurídico (Sun et al., 2024). Esses erros não são falhas esporádicas, mas consequências estruturais do funcionamento preditivo dos modelos, que buscam a resposta linguística mais provável, não a mais verdadeira.

Outro desafio relevante diz respeito ao treinamento. Embora esses modelos aprendam padrões linguísticos gerais durante o pré-treinamento, o domínio jurídico exige conhecimento especializado sobre conceitos normativos, terminologia técnica e estruturas lógicas próprias do Direito. Se o modelo for treinado majoritariamente em textos jurídicos estrangeiros, é comum que reproduza raciocínios alinhados ao GDPR, ao direito norte-americano ou à jurisprudência internacional, o que pode gerar interpretações equivocadas da LGPD. Essa assimetria exige o uso de técnicas adicionais de contextualização, como o fornecimento explícito de trechos legais ou o desenvolvimento de prompts cuidadosamente elaborados.

Além disso, LLMs apresentam sensibilidade significativa à formulação das instruções. Pequenas alterações no enunciado podem produzir respostas substancialmente diferentes, revelando fragilidade semântica e falta de consistência. Essa dependência reforça a necessidade de padronização metodológica na formulação de prompts, especialmente no contexto de uma ferramenta de análise, onde se busca comparabilidade e reprodutibilidade dos resultados.

Há também limitações técnicas relacionadas à janela de contexto, isto é, ao tamanho máximo de texto que o modelo consegue analisar simultaneamente. Políticas de privacidade extensas podem exceder esse limite, exigindo fragmentação do texto. Essa divisão pode romper relações lógicas importantes e dificultar a identificação de omissões, incoerências ou contradições internas.

Por fim, é importante notar que LLMs não realizam raciocínio jurídico formal. Mesmo quando não produzem alucinações, podem interpretar dispositivos legais de forma literal, ignorar condicionantes, aplicar conceitos de maneira excessivamente ampla ou não captar nuances estruturantes como finalidade, boa-fé ou proporcionalidade.

2.4.2 Engenharia de Prompt

A *Prompt Engineering*, ou engenharia de prompt, é o conjunto de técnicas voltadas à construção sistemática de instruções que orientam o comportamento de um LLM durante a fase de inferência. Diferentemente de abordagens que modificam os parâmetros internos do modelo, a engenharia de prompt opera exclusivamente sobre a entrada fornecida ao sistema, sem qualquer alteração nos pesos da rede neural. Por essa razão, trata-se da forma mais ágil e economicamente viável de especializar uma LLM para tarefas específicas, sendo o ponto de partida natural para qualquer aplicação baseada nesses modelos (White et al., 2023).

Do ponto de vista estrutural, um prompt eficaz geralmente é composto por camadas de informação que delimitam progressivamente o espaço de respostas do modelo. A primeira camada é o *system prompt*, uma instrução de alto nível que define o papel, o escopo e as restrições do modelo para toda a sessão. Em seguida, é possível injetar no contexto informações externas relevantes, como trechos da legislação, definições normativas ou exemplos de análises anteriores, que o modelo utilizará como referência. Por fim, a instrução de tarefa especifica com precisão o que deve ser produzido, o formato esperado e os critérios de qualidade da resposta (Sahoo et al., 2024).

A literatura de processamento de linguagem natural consolidou um conjunto de estratégias de prompting que variam conforme a quantidade de exemplos fornecidos ao modelo e a forma como o raciocínio é guiado (Sahoo et al., 2024):

- **Zero-shot prompting**: O modelo recebe apenas a descrição da tarefa, sem exemplos de entrada e saída. Adequado para tarefas genéricas, mas tende a produzir resultados menos consistentes em domínios especializados, como o jurídico.

- **Few-shot prompting**: São fornecidos ao modelo de dois a cinco exemplos concretos da tarefa antes da solicitação real (Brown et al., 2020). Essa estratégia reduz ambiguidade e calibra o formato e o nível de detalhe esperados nas respostas, sendo especialmente eficaz em tarefas de classificação e extração estruturada de informações.

- **Chain-of-Thought (CoT)**: O modelo é instruído a explicitar seu raciocínio passo a passo antes de apresentar uma conclusão (Wei et al., 2022). Essa técnica melhora significativamente o desempenho em tarefas que exigem inferências encadeadas, como a verificação de conformidade de cláusulas com múltiplos requisitos legais simultâneos. Kojima et al. (2022) demonstraram que mesmo sem exemplos, a simples adição da instrução “pense passo a passo” ao prompt induz o modelo a adotar raciocínio estruturado, o que os autores denominaram *zero-shot chain-of-thought*.

- **Role prompting**: O modelo recebe uma persona definida, por exemplo, a de um especialista em conformidade com a LGPD, antes de executar a tarefa. Essa estratégia favorece a adoção de vocabulário técnico adequado ao domínio e reduz a tendência do modelo de recorrer a respostas genéricas ou baseadas em legislações estrangeiras.

No contexto deste trabalho, a engenharia de prompt desempenha um papel estrutural. O *system prompt* da aplicação incorporará a persona de especialista em LGPD, o texto dos artigos mais relevantes da lei e uma instrução explícita para que o modelo restrinja sua análise ao ordenamento jurídico brasileiro. A estratégia *few-shot* será utilizada para calibrar o formato das respostas, garantindo que a análise de cada cláusula seja apresentada de forma estruturada e comparável entre diferentes políticas de privacidade. A técnica *Chain-of-Thought* será aplicada nas tarefas que exigem verificação de conformidade multidimensional, onde o modelo deve confrontar diferentes trechos do documento com múltiplos princípios da LGPD simultaneamente.

Apesar de sua flexibilidade, a engenharia de prompt apresenta limitações intrínsecas. O tamanho finito da janela de contexto impõe um teto para a quantidade de informação que pode ser injetada em cada requisição, tornando inviável incluir toda a legislação e jurisprudência relevante em um único prompt. Além disso, a técnica não altera a distribuição de probabilidades interna do modelo: termos e conceitos específicos da LGPD que não estavam adequadamente representados nos dados de pré-treinamento continuarão subrepresentados, independentemente da qualidade do prompt. Nesses cenários, torna-se necessário recorrer ao *fine-tuning*.

2.4.3 Fine-Tuning

O *fine-tuning*, ou ajuste fino, é o processo de continuar o treinamento de um LLM pré-treinado utilizando um conjunto de dados menor, curado e específico ao domínio de interesse. Enquanto o pré-treinamento expõe o modelo a vastos corpora textuais gerais com o objetivo de adquirir competência linguística ampla, o ajuste fino direciona o aprendizado para padrões, vocabulário e estruturas lógicas características de um domínio particular, atualizando os pesos da rede neural para que as distribuições probabilísticas das saídas reflitam esse conhecimento especializado (Howard; Ruder, 2018).

Do ponto de vista técnico, o processo segue o paradigma de aprendizado supervisionado: o modelo recebe pares de entrada e saída esperada, por exemplo, o texto de uma cláusula de política de privacidade e a análise de sua conformidade com a LGPD, e os parâmetros são ajustados iterativamente para minimizar a diferença entre as saídas do modelo e as saídas de referência. Após o ajuste fino, o modelo passa a responder de maneira mais consistente e precisa para as tarefas do domínio,

com menor necessidade de instruções elaboradas no prompt e menor incidência de alucinações sobre terminologia jurídica específica.

2.4.3.1 Modalidades de Ajuste Fino

O custo computacional do ajuste fino de modelos com dezenas ou centenas de bilhões de parâmetros é proibitivo quando todos os pesos da rede são atualizados simultaneamente. Para contornar essa limitação, a literatura propôs métodos de *Parameter-Efficient Fine-Tuning* (PEFT), que atualizam apenas um subconjunto reduzido dos parâmetros do modelo, mantendo os demais congelados (Ding et al., 2023).

O método mais amplamente adotado nessa categoria é o *Low-Rank Adaptation* (LoRA), proposto por Hu et al. (2022). A premissa central do LoRA é que as atualizações de pesos necessárias para adaptar um modelo a um domínio específico tendem a ter baixa dimensionalidade intrínseca, isto é, podem ser representadas de forma compacta. Em vez de modificar diretamente as matrizes de pesos do modelo original, o LoRA injeta matrizes de baixo posto (*low-rank matrices*) em paralelo a cada camada relevante da arquitetura *Transformer*. Durante o treinamento, apenas essas matrizes auxiliares de dimensão reduzida são atualizadas; os pesos originais permanecem congelados. Ao final do processo, as matrizes adaptadoras podem ser fundidas aos pesos originais sem adicionar latência na inferência.

Essa abordagem reduz a quantidade de parâmetros treináveis em até cem vezes em comparação ao ajuste fino completo, viabilizando a execução do treinamento em infraestruturas com recursos de GPU mais modestos ou em plataformas de nuvem com menor custo operacional. Para o presente trabalho, o LoRA representa a modalidade de ajuste fino mais adequada, dado o tamanho do corpus disponível e as restrições de infraestrutura.

2.4.3.2 Dataset e Origem dos Dados

A qualidade do ajuste fino depende diretamente da qualidade e representatividade do conjunto de dados utilizado. Para o domínio específico deste trabalho, análise de conformidade de políticas de privacidade com a LGPD, o dataset será composto por duas categorias de dados:

A primeira categoria é formada por políticas de privacidade reais de empresas que operam no Brasil, coletadas publicamente a partir de seus sites. Esses documentos serão anotados manualmente, cláusula por cláusula, com rótulos que indicam conformidade, não conformidade ou ausência de informação em relação a cada um dos dez princípios da LGPD e aos requisitos mínimos de transparência estabelecidos pela ANPD. O processo de anotação seguirá um protocolo definido com base nas orientações publicadas pela própria autoridade reguladora (Autoridade Nacional de Proteção de Dados, 2023), garantindo consistência entre os anotadores.

A segunda categoria é composta por documentos normativos de referência, o texto integral da LGPD, resoluções e guias publicados pela ANPD e decisões administrativas relevantes, formatados como exemplos de instrução e resposta para o treinamento supervisionado. Essa camada tem como objetivo fazer com que o modelo internalize não apenas padrões textuais de políticas de privacidade, mas também a lógica interpretativa da autoridade reguladora brasileira.

2.4.3.3 Infraestrutura de Treinamento

O treinamento será conduzido em plataforma de computação em nuvem, sem necessidade de infraestrutura local dedicada. Duas alternativas principais serão avaliadas: a API de fine-tuning da OpenAI, que oferece ajuste fino supervisionado do modelo GPT mediante envio de arquivo de dados no formato JSONL, abstraindo toda a gestão de infraestrutura; e o Hugging Face com aceleração via GPU em nuvem (Google Colab Pro ou AWS SageMaker), que permite maior controle sobre hiperparâmetros, arquitetura LoRA e processo de avaliação, sendo compatível com modelos de código aberto como LLaMA e Mistral.

A escolha definitiva entre essas alternativas será orientada pelos resultados da fase experimental de prompt engineering: caso o desempenho com engenharia de prompt atinja métricas de F1 satisfatórias, métrica detalhada na Seção 2.4.5, o ajuste fino pode ser utilizado como etapa de refinamento; caso contrário, passa a ser a estratégia central de adaptação do modelo.

2.4.3.4 Riscos e Mitigações

O ajuste fino introduz riscos específicos que devem ser gerenciados. O principal é o *catastrophic forgetting*, fenômeno em que o modelo perde competências linguísticas gerais adquiridas durante o pré-treinamento ao ser exposto excessivamente a um corpus especializado e pequeno. A utilização de LoRA mitiga parcialmente esse risco, pois os pesos originais permanecem intocados. Adicionalmente, o processo de treinamento utilizará uma taxa de aprendizado (*learning rate*) conservadora e será monitorado por métricas de validação em um conjunto de dados reservado, com interrupção antecipada (*early stopping*) para evitar overfitting.

Outro risco é o viés introduzido pelo processo de anotação manual do dataset. Para mitigá-lo, uma amostra das anotações será revisada por um segundo avaliador, calculando-se o coeficiente de concordância entre anotadores (*inter-annotator agreement*) como indicador da qualidade e objetividade dos rótulos.

2.4.4 Análise de Texto com PLN Tradicional vs LLMs

Antes da popularização das LLMs, a análise automática de documentos textuais dependia de metodologias tradicionais de Processamento de Linguagem Natural (PLN).

Essas abordagens tratam o texto como uma coleção de palavras ou expressões descontextualizadas, utilizando técnicas para identificar padrões estatísticos superficiais. Embora eficientes para detectar termos específicos, essas técnicas têm pouca capacidade de compreensão semântica profunda, sendo incapazes de interpretar relações complexas entre cláusulas, captar ambiguidade ou inferir sentido a partir de contexto amplo.

Políticas de privacidade, entretanto, exigem exatamente esse tipo de análise contextualizada. Muitas vezes, sua conformidade depende não da presença de palavras-chave, mas do modo como elas se relacionam com outras seções do documento. Termos como “compartilhamento”, “consentimento” ou “retenção”, isolados, dizem pouco sobre a adequação à LGPD; o que importa é a forma como são definidos, condicionados e conectados a obrigações específicas.

É nesse cenário que as LLMs oferecem vantagens substanciais. Diferentemente das técnicas de PLN tradicional, esses modelos operam com representações distribuídas de palavras e frases, capazes de captar nuances semânticas e dependências de longo alcance. Assim, conseguem identificar quando uma política é excessivamente vaga, quando omite informações essenciais ou quando utiliza linguagem incompatível com os princípios da legislação. Além disso, a capacidade de operar em múltiplas tarefas, sumarização, explicação, classificação e detecção de incoerências, sem a necessidade de modelos separados torna o processo mais integrado e eficiente.

Dessa forma, enquanto abordagens tradicionais de PLN fornecem análises fragmentadas e baseadas em padrões superficiais, as LLMs permitem uma leitura mais contextualizada, flexível e consistente, aproximando-se da complexidade interpretativa exigida por documentos jurídicos.

2.4.5 Métricas de Avaliação

Para mensurar a eficácia de modelos na tarefa de classificação, a literatura tradicionalmente utiliza uma Matriz de Confusão definida por Fawcett (2006) que define quatro estados. Adaptando-os ao escopo do trabalho:

Tabela 1 – Estados da Matriz de Confusão aplicados à Análise de Políticas de Privacidade

Estado	**Definição no Contexto**	**Impacto e Consequência**
Verdadeiro Positivo (VP)	O modelo identifica corretamente a presença de uma cláusula ou prática de tratamento de dados que realmente consta no documento.	Representa o sucesso da ferramenta em alertar o usuário.
Verdadeiro Negativo (VN)	O modelo afirma corretamente que uma determinada característica (ex: venda de dados) não está presente na política.	Garante a confiabilidade da ferramenta ao não inventar riscos inexistentes, essencial para a credibilidade da análise.
Falso Positivo (FP)	O modelo sinaliza erroneamente a existência de uma prática ou risco que não consta no texto original.	Gera excesso de alarmes falsos e consequente desconfiança no usuário. Pode levar a interpretações equivocadas sobre a idoneidade da empresa analisada.
Falso Negativo (FN)	O modelo falha em identificar um elemento presente, omitindo uma informação relevante ou um risco.	Considerado o erro mais crítico em *compliance*. Deixa o titular dos dados exposto a riscos jurídicos ou de privacidade não detectados.

Fonte: Elaborado pelo autor.

A partir destes estados é possível calcular o F1-Score do modelo. A fórmula é dada por:

$$F1 = \frac{2 \cdot VP}{2 \cdot VP + FP + FN} \tag{1}$$

Onde:

- VP: Verdadeiros Positivos
- FP: Falsos Positivos
- FN: Falsos Negativos

O resultado varia de 0 a 1, onde valores próximos de 1 indicam um modelo que consegue, simultaneamente, poucos alarmes falsos e poucas omissões.

3 TRABALHOS RELACIONADOS

A busca por automatizar a compreensão e a avaliação de políticas de privacidade é um desafio que acompanha a evolução da própria web. Diferentes abordagens foram propostas ao longo das últimas duas décadas, cada uma explorando perspectivas tecnológicas e metodológicas distintas. Os trabalhos apresentados a seguir foram identificados por meio de uma revisão bibliográfica não sistemática conduzida na base Google Scholar, combinando os termos “privacy policy analysis”, “LGPD compliance”, “large language models” e “prompt engineering”, priorizando estudos amplamente citados ou publicados em veículos de relevância reconhecida na área. Nesta seção, apresentam-se os principais trabalhos e ferramentas que antecedem ou dialogam com a proposta deste estudo.

3.0.1 P3P

Uma das primeiras iniciativas de grande escala foi o *Platform for Privacy Preferences Project* (P3P), desenvolvido pelo W3C no início dos anos 2000 (Cranor et al., 2002). Seu objetivo era criar um protocolo padronizado e legível por máquina que permitisse a descrição formal das práticas de coleta e uso de dados de cada site. A ideia era que navegadores pudessem interpretar automaticamente as políticas, alertando o usuário quando um site adotasse práticas incompatíveis com suas preferências.

Apesar da inovação conceitual, o P3P fracassou em sua adoção prática. A principal resistência veio das próprias empresas, que viam a transparência estrita como um risco estratégico, especialmente em um cenário no qual modelos de negócio baseados em coleta de dados ainda se consolidavam. Além disso, a rigidez do formato dificultava representar nuances textuais presentes em políticas reais, levando muitas organizações a fornecer declarações incompletas ou imprecisas. (TechRepublic, 2002)

3.0.2 ToS;DR

O projeto *Terms of Service; Didn’t Read* (ToS;DR), surgiu como uma iniciativa colaborativa de *crowsourcing* que utiliza voluntários para ler, interpretar e classificar termos de uso e políticas de privacidade. Ao invés de tentar automatizar a leitura, o projeto aposta na curadoria humana como mecanismo de qualidade, publicando resumos de fácil compreensão e atribuindo notas a serviços populares com base em critérios éticos e de transparência. (TOS. . . , 2025)

Apesar de seu impacto social positivo, a abordagem esbarra no limite estrutural da falta de escalabilidade. A vastidão da web e a constante atualização das políticas e o volume de serviços tornam inviável manter uma cobertura ampla apenas com trabalho manual. Assim, o projeto serve mais como um recurso informativo de nicho do que como uma solução abrangente para análise sistemática de conformidade.

3.0.3 Abordagem Baseada em Grafos de Conhecimento

O projeto POLIGRAPH, proposto por Cui et al. (2023), representa uma abordagem que analisa o texto completo de uma política de privacidade de forma automática baseando-se na construção de *Knowledge Graphs*, estruturas que representam o conhecimento através de entidades e os seus relacionamentos, a fim de mapear semanticamente as declarações de uma política. Nesta estrutura, termos como tipos de dados e entidades são conectados por relações de "recolha"ou "subsunção", permitindo inferências sobre práticas de dados que não estão explicitamente declaradas numa única frase.

Apesar do avanço metodológico, os autores reconhecem na conclusão do trabalho as limitações inerentes às abordagens de PLN baseadas em pipelines sintáticos tradicionais. É mencionada a dificuldade de cobrir todos os padrões sintáticos possíveis, o que pode levar a erros de interpretação. Devido a estas limitações, o estudo ressalta que os resultados da análise automática, especialmente quando aplicados a políticas individuais, devem ser validados por especialistas humanos para garantir a precisão.

O trabalho sugere como direção para trabalhos futuros e exploração do uso de LLMs. Os autores observam que os recentes avanços nestas tecnologias aumentaram muito a capacidade de compreensão de textos complexos, reduzindo a necessidade de combinar múltiplos componentes e heurísticas sofisticadas.

3.0.4 Abordagem Baseada em LLM

O estado da arte neste domínio é representado pelo trabalho de Xie, Ramakrishnan e Li (2025), que também propôs uma estrutura automatizada baseada em uma LLM para avaliar políticas de privacidade em escala. O estudo considerou requisitos de 10 leis de privacidade modernas, incluindo a GDPR da União Europeia e nove leis estaduais dos EUA.

A metodologia dividiu a análise em duas tarefas principais:

- **Tarefas de Cobertura:** Verificar se uma política contém o conteúdo exigido por uma cláusula específica, como a divulgação de categorias de dados coletados ou métodos para exercer direitos do usuário.
- **Tarefas de Inspeção:** Extrair detalhes granulares sobre as práticas de informação pessoal, como propósitos de coleta e categorias de terceiros com quem os dados são compartilhados.

Os resultados demonstraram que a abordagem com LLMs alcançou uma pontuação F1 média de 0,94 em todas as tarefas, superando significativamente os métodos anteriores, que geralmente apresentavam pontuações abaixo de 0,80. Além da alta

precisão, o estudo comprovou a viabilidade de medições em larga escala, analisando políticas de mais de 100 mil websites.

A aplicação destas técnicas ao contexto da LGPD no Brasil apresenta um potencial promissor, embora ainda pouco explorado. Na revisão bibliográfica conduzida para este trabalho, não foi identificado nenhum estudo publicado que aplique LLMs especificamente à avaliação de conformidade com a LGPD, tampouco um corpus público anotado de políticas de privacidade brasileiras sob os critérios dessa legislação, o que evidencia a lacuna que este trabalho se propõe a preencher. O estudo de Xie et al. explicitamente sugere que trabalhos futuros devem expandir a análise para leis de privacidade de outras regiões do mundo e desenvolver suporte para análise multilíngue, visto que os LLMs possuem capacidades de análise de texto que vão além do inglês/.

3.1 EXTENSÕES DE NAVEGADOR

As extensões para navegadores web são pequenos programas desenvolvidos para modificar, ampliar ou personalizar as funcionalidades de um navegador, permitindo a integração direta com páginas da internet e a execução de tarefas específicas durante a navegação. Elas desempenham papel relevante na automação de processos, na coleta e análise de informações e na melhoria da experiência do usuário, sem a necessidade de software adicional instalado no sistema operacional (Mozilla, 2025).

Essas extensões funcionam como camadas intermediárias entre o navegador e o conteúdo das páginas, utilizando *API*s que permitem interagir com o DOM, realizar requisições em segundo plano e acessar recursos locais limitados. O desenvolvimento de extensões envolve linguagens e tecnologias da web, como HTML, CSS e JavaScript, além do uso de *API*s específicas para manipulação de abas, histórico, armazenamento local, notificações e requisições HTTP.

Entre as principais aplicações das extensões para navegadores podemos citar automação de tarefas repetitivas, como preenchimento de formulários; bloqueio ou filtragem de informações, como anúncios e rastreadores de navegação; funcionalidades de leitura, tradução, categorização ou sumarização de texto em tempo real.

3.2 ENGENHARIA DE PROMPT EM LLMS: UMA REVISÃO SISTEMÁTICA

O trabalho de Sahoo et al. (2024) oferece uma revisão abrangente das principais técnicas de engenharia de prompt aplicadas a LLMs, organizando-as em uma taxonomia que abrange desde estratégias básicas até abordagens avançadas de raciocínio encadeado. Os autores analisam sistematicamente métodos como *zero-shot prompting*, *few-shot prompting* e *Chain-of-Thought*, detalhando para cada técnica a metodologia subjacente, os modelos avaliados, os conjuntos de dados utilizados e as limitações observadas. O estudo também investiga o impacto da formulação das

instruções na qualidade das respostas, evidenciando que variações semânticas sutis no enunciado de uma mesma tarefa podem produzir resultados substancialmente distintos.

Embora o escopo do survey não se restrinja a nenhum domínio de aplicação específico, as conclusões são diretamente transferíveis ao contexto deste trabalho. A análise de políticas de privacidade compartilha das mesmas características que tornam a engenharia de prompt indispensável em outros domínios especializados: a necessidade de precisão terminológica, a exigência de raciocínio estruturado sobre múltiplos critérios simultâneos e a sensibilidade do modelo à forma como a tarefa é apresentada. O levantamento de Sahoo et al. (2024) fundamenta, portanto, as escolhas metodológicas adotadas na construção dos prompts deste projeto, em particular o uso combinado de *few-shot prompting* e *Chain-of-Thought* para tarefas de verificação de conformidade com os princípios da LGPD.

3.2.1 LoRA: Adaptação Eficiente de Modelos de Linguagem

O trabalho de Hu et al. (2022) propõe o *Low-Rank Adaptation* (LoRA), um método de ajuste fino que resolve o principal gargalo prático da especialização de LLMs: o custo computacional proibitivo de atualizar todos os parâmetros de modelos com dezenas ou centenas de bilhões de pesos. A premissa central do método é que a mudança necessária nos pesos de um modelo durante a adaptação a um domínio específico possui baixa dimensionalidade intrínseca e pode, portanto, ser aproximada por matrizes de posto reduzido. Em vez de modificar diretamente os pesos pré-treinados, o LoRA injeta pares de matrizes treináveis em paralelo a cada camada relevante da arquitetura *Transformer*, mantendo os pesos originais congelados. Os experimentos conduzidos pelos autores demonstraram que o método alcança desempenho comparável ao ajuste fino completo em tarefas de geração de texto e resposta a perguntas, reduzindo em até dez mil vezes a quantidade de parâmetros treináveis e em até três vezes o consumo de memória GPU em relação ao ajuste fino tradicional do GPT-3.

A relevância desse trabalho para o presente estudo é direta: o LoRA viabiliza a execução do ajuste fino em infraestrutura de nuvem de custo acessível, sem exigir hardware dedicado de alto desempenho. Embora os experimentos originais de Hu et al. (2022) não envolvam domínios jurídicos ou análise de conformidade regulatória, o método foi amplamente adotado para especialização de modelos em áreas técnicas com terminologia própria, cenário análogo ao da adaptação ao vocabulário e à lógica normativa da LGPD.

3.3 SÍNTESE

A revisão bibliográfica apresentada neste capítulo e no anterior mostra que proteger a privacidade no ambiente digital é um desafio complexo, situado na interseção

entre a vulnerabilidade dos usuários, a atuação do Estado e o avanço das tecnologias de processamento de informação.

Observa-se que o modelo de negócios dominante na internet, marcado pelo chamado capitalismo de vigilância, produz uma forte assimetria de poder. O usuário, frequentemente preso ao chamado paradoxo da privacidade, permanece exposto diante de práticas de coleta de dados pouco transparentes e presentes em praticamente todos os serviços digitais.

Como resposta a esse cenário, a LGPD instituiu um arcabouço jurídico sólido, baseado em princípios como transparência e autodeterminação informativa. Na prática, porém, o principal instrumento voltado à comunicação com o titular, as políticas de privacidade, tendem a ser longos, técnicos e pouco acessíveis, sem muito impacto real para o usuário.

Historicamente, diversas iniciativas buscaram resolver esse problema, como o protocolo P3P ou projetos colaborativos como o ToS;DR, mas acabaram limitadas por baixa adoção ou dificuldades de escala. Abordagens computacionais mais recentes, baseadas em PLN tradicional, como o POLIGRAPH, também avançaram na estruturação e classificação das informações, mas demonstraram limitações para lidar com nuances semânticas e com a grande variação textual das políticas de privacidade atuais.

É nesse contexto que as LLMs representam uma mudança significativa. Com a capacidade de analisar contexto de forma aprofundada, esses modelos superam limitações das técnicas anteriores e permitem uma interpretação automatizada mais próxima da leitura humana. Estudos recentes, como o de Xie, Ramakrishnan e Li (2025), mostram que LLMs têm se destacado na análise de conformidade em grande escala, oferecendo mais precisão e abrangência do que métodos tradicionais.

Apesar disso, o uso de LLMs especificamente no contexto da LGPD, com seus princípios, requisitos e terminologias próprias, ainda é pouco explorado: nenhum dos trabalhos revisados neste capítulo propõe ou avalia uma solução de conformidade voltada à legislação brasileira, o que caracteriza uma lacuna de pesquisa ainda não preenchida na literatura consultada. Dessa forma, reforça-se o objetivo central deste trabalho: fortalecer a posição do titular de dados por meio de uma ferramenta que aproveite a capacidade interpretativa de LLMs, entregue como uma extensão de navegador, para transformar políticas de privacidade complexas em informações claras, acessíveis e compatíveis com a legislação brasileira.

4 METODOLOGIA

A análise conduzida no capítulo anterior evidenciou que, embora a LGPD ofereça um arcabouço jurídico robusto, a aplicação prática de seus princípios em documentos de políticas de privacidade ainda é marcada por inconsistências, ambiguidades e omissões. Além disso, tais documentos frequentemente apresentam redação excessivamente técnica ou extensa, dificultando sua compreensão pelo usuário comum e ampliando a assimetria informacional entre titulares e controladores. Esse conjunto de limitações abre espaço para o desenvolvimento de um mecanismo automatizado de avaliação que não apenas identifique o grau de conformidade desses documentos, mas também facilite sua interpretação por parte do usuário, no contexto da língua portuguesa e da legislação brasileira, visto que soluções já se mostraram efetivas em outros contextos legislativos, conforme demonstrado por Xie, Ramakrishnan e Li (2025). Assim, a estratégia de implementação deste trabalho parte dessa necessidade.

4.1 REQUISITOS DE TRANSPARÊNCIA AVALIADOS

A avaliação automatizada de políticas de privacidade exige a definição prévia das obrigações de transparência que serão verificadas e dos critérios objetivos que determinam o grau de atendimento de cada uma. Este trabalho avalia oito requisitos derivados dos dispositivos da LGPD que estabelecem deveres de informação ao titular. A seleção partiu da premissa de que uma política funcionalmente completa deve, ao menos, permitir que o titular identifique quem trata seus dados, com que propósito, durante quanto tempo, com base em qual hipótese legal, com quais garantias de segurança, com quem os dados são eventualmente compartilhados e de que forma pode exercer seus direitos. Para cada requisito, o modelo classifica a política em um de três níveis: **presente**, quando a obrigação é atendida de forma clara e suficiente; **parcial**, quando é endereçada de forma incompleta ou ambígua; e **ausente**, quando o tema é inteiramente omitido. Os oito requisitos são os seguintes:

Finalidade Identificação explícita dos propósitos que justificam a coleta e o tratamento dos dados pessoais (art. 6º, I; art. 9º, I). É *presente* quando a política declara as finalidades de forma clara e determinada; *parcial* quando a descrição é vaga ou genérica (ex.: “melhorar nossos serviços”); e *ausente* quando nenhuma finalidade é enunciada.

Base legal Indicação da hipótese legal que autoriza o tratamento, conforme os artigos 7º e 11º. É *presente* quando a política menciona ao menos uma base legal prevista na LGPD (ex.: consentimento, execução de contrato, legítimo interesse); *parcial* quando há referência genérica à lei sem identificação da hipótese específica; e *ausente* quando a base legal não é mencionada.

Retenção Especificação do prazo ou critério durante o qual os dados serão mantidos antes de sua eliminação ou anonimização (art. 9º, I; art. 16º). É *presente* quando a política informa prazos ou critérios concretos de retenção; *parcial* quando menciona o tema sem especificar duração ou critério; e *ausente* quando a questão não é abordada.

Direitos Informação ao titular sobre os direitos que pode exercer em relação a seus dados (arts. 17–22), como acesso, correção, exclusão, portabilidade e revogação do consentimento. É *presente* quando a política enumera os direitos disponíveis; *parcial* quando os menciona de forma incompleta; e *ausente* quando o tema é omitido.

Canal de exercício Indicação dos meios pelos quais o titular pode contatar o controlador para exercer seus direitos ou apresentar solicitações (art. 9º, III; art. 41º). É *presente* quando a política informa um canal concreto (e-mail, formulário, encarregado); *parcial* quando indica a possibilidade de contato sem fornecer o meio; e *ausente* quando não há qualquer indicação.

Controlador Identificação da pessoa física ou jurídica responsável pelo tratamento dos dados (art. 5º, VI; art. 9º). É *presente* quando a política identifica o controlador com nome e dados de contato; *parcial* quando a identificação é incompleta (ex.: nome sem contato); e *ausente* quando o responsável pelo tratamento não é identificado.

Segurança Descrição das medidas técnicas e organizacionais adotadas para proteger os dados pessoais contra acessos não autorizados e situações acidentais ou ilícitas (art. 6º, VII; art. 46º). É *presente* quando a política descreve as medidas adotadas; *parcial* quando afirma adotar medidas de segurança sem descrevê-las; e *ausente* quando o tema não é mencionado.

Compartilhamento Indicação de se os dados pessoais são transferidos a terceiros e, em caso afirmativo, identificação de quem são esses terceiros e para quais finalidades (art. 9º, I; art. 7º). Adota-se uma regra específica para este requisito: é classificado como *presente* sempre que a política identifica os terceiros e as finalidades do compartilhamento, independentemente de declarar que o realiza ou de afirmar que não o faz; é classificado como *parcial* quando a política menciona a existência de compartilhamento sem identificar os terceiros ou as finalidades envolvidas; e o rótulo *ausente* é reservado exclusivamente para documentos que omitem o assunto por completo.

Os oito requisitos recebem peso idêntico na rotulagem e na definição de conformidade descrita na Seção 4.3. Essa uniformidade é deliberada e reflete o caráter

cumulativo dos princípios da LGPD: o artigo 6º os enumera sem estabelecer hierarquia entre eles, de modo que o cumprimento exemplar de sete requisitos não compensa a omissão de um oitavo. A única gradação de severidade admitida pela regra de conformidade é a que distingue *ausente* de *parcial*: um requisito *parcial* evidencia tentativa de cumprimento com execução incompleta, situação que a regra penaliza com menos rigor do que a omissão total, como detalhado na Seção 4.3.

4.2 CONSTRUÇÃO DO DATASET

Embora a literatura internacional disponha de corpora voltados à análise automatizada de políticas de privacidade, esses recursos concentram-se predominantemente no idioma inglês e em contextos legislativos distintos da LGPD. O *OPP-115 Corpus* (Wilson et al., 2016), apresentado na ACL 2016, reúne 115 políticas de privacidade de sítios norte-americanos anotadas manualmente segundo um esquema de dez categorias de práticas informacionais, tornando-se referência amplamente utilizada em pesquisas de aprendizado de máquina para análise de privacidade. Em escala distinta, Amos et al. (2021) compilaram um dataset longitudinal com mais de um milhão de políticas em inglês coletadas ao longo de duas décadas, possibilitando análises históricas sobre a evolução das práticas de transparência. Apesar de sua relevância, esses recursos não contemplam os requisitos específicos da legislação brasileira nem documentos redigidos em português, o que os torna inadequados como base de treinamento ou avaliação para o contexto deste trabalho.

A inexistência de um corpus público anotado de políticas de privacidade brasileiras avaliadas sob os critérios da LGPD tornou necessária a construção de um dataset original para este trabalho. O processo foi conduzido em três etapas sequenciais: curadoria das políticas, rotulagem automática por *Silver Labeling* e revisão manual de uma amostra representativa.

4.2.1 Curadoria das Políticas de Privacidade

O corpus é composto por 200 políticas de privacidade de organizações que atuam no Brasil, coletadas diretamente de seus sítios eletrônicos oficiais. A seleção buscou representatividade setorial, cobrindo oito segmentos da economia e da administração pública: educação (22 políticas), financeiro (22), governo (31), indústria (23), saúde (21), serviços (22), tecnologia (35) e varejo (24). A diversidade setorial é relevante porque as práticas de coleta e tratamento de dados variam significativamente entre esses contextos, e um modelo treinado apenas em um segmento tenderia a generalizar mal para os demais.

As políticas foram acessadas por meio de suas URLs públicas e armazenadas em formato texto após extração do conteúdo relevante, com remoção de elementos de na-

vegação, rodapés, scripts e outros artefatos não textuais. Documentos disponibilizados em formato PDF foram igualmente processados com extração de texto.

4.2.2 Silver Labeling com Claude Opus 4.6

A anotação manual de 200 documentos por especialistas jurídicos seria inviável dentro do escopo e prazo deste trabalho. Para contornar essa limitação sem comprometer a qualidade dos rótulos, adotou-se a estratégia de *Silver Labeling*: geração automática de rótulos por um modelo de linguagem de alta capacidade, utilizado como anotador substituto. Essa abordagem é estabelecida na literatura de NLP como alternativa viável quando a anotação humana em escala não é praticável, produzindo rótulos de qualidade inferior à anotação especializada, porém superiores à ausência de supervisão (Tan et al., 2024).

O modelo utilizado como anotador foi o Claude Opus 4.6 (Anthropic), ocupante das primeiras posições do *Chatbot Arena* (Chiang et al., 2024), benchmark de referência da comunidade para comparação de LLMs por preferência humana, no momento da coleta. Sua escolha como anotador é deliberada: ao utilizar o modelo de maior capacidade disponível para gerar os rótulos de referência, minimiza-se o risco de que erros sistemáticos do anotador contaminem o dataset. Um modelo de menor capacidade tenderia a produzir classificações inconsistentes em requisitos com formulação textual ambígua, como base legal e retenção, justamente os mais relevantes para a avaliação de conformidade. O benchmark comparativo conduzido na Seção 4.5 confirma que todos os modelos utilizados nos experimentos de treinamento e avaliação operam com desempenho significativamente inferior ao Claude Opus, o que reforça sua adequação como referência de qualidade para o processo de anotação. Cada política foi submetida ao modelo com um prompt estruturado instruindo a avaliação individual dos oito requisitos definidos na Seção 4.1. Para cada requisito, o modelo produziu um dos três rótulos possíveis: *presente*, *ausente* ou *parcial*. A partir dessa avaliação granular, um rótulo binário de conformidade geral foi derivado segundo o critério definido na Seção 4.3: uma política é classificada como conforme quando nenhum dos oito requisitos é avaliado como ausente e, no máximo, um é avaliado como parcial.

O resultado da rotulagem revelou um desbalanceamento pronunciado entre as classes: das 200 políticas, 41 foram classificadas como conformes (20,5%) e 159 como não conformes (79,5%). Esse desbalanceamento, longe de ser um artefato do processo de anotação, reflete a realidade do mercado brasileiro, onde a maioria das organizações ainda não atende plenamente às exigências de transparência da LGPD. O setor governamental apresentou o resultado mais crítico: nenhuma das 31 políticas coletadas foi classificada como conforme. O setor financeiro apresentou a maior taxa de conformidade relativa, com 8 das 22 políticas (36,4%) atingindo o limiar estabelecido. A distribuição por setor é apresentada na Tabela 2.

Tabela 2 – Distribuição do dataset por setor

Setor	**Total**	**Conformes**	**Não conformes**
Educação	22	2	20
Financeiro	22	8	14
Governo	31	0	31
Indústria	23	5	18
Saúde	21	7	14
Serviços	22	6	16
Tecnologia	35	11	24
Varejo	24	2	22
Total	**200**	**41**	**159**

Fonte: Elaborado pelo autor.

4.2.3 Revisão Manual e Validação dos Rótulos

Para verificar a qualidade dos rótulos gerados automaticamente, foi conduzida uma revisão manual de uma amostra estratificada do dataset. A amostra compreendeu 16 políticas, sendo dois documentos por setor, selecionados para cobrir os dois extremos do espectro de qualidade em cada segmento.

O protocolo de revisão não consistiu em uma reanotação independente e cega, com produção de um segundo conjunto de rótulos para cálculo formal de concordância entre anotadores, mas em uma verificação de plausibilidade: o revisor tinha acesso simultâneo ao texto da política e ao rótulo já atribuído pelo Claude Opus 4.6, avaliando requisito a requisito se essa classificação correspondia ao que o documento efetivamente declarava. A revisão confirmou a concordância entre os rótulos automáticos e a avaliação manual para os requisitos com critérios objetivos, como controlador e canal de exercício, e maior variabilidade nos requisitos que admitem formulações ambíguas, como base legal e retenção, onde a presença de linguagem vaga nas políticas torna a classificação inerentemente subjetiva.

O dataset final, com os rótulos revisados, foi dividido em conjunto de treino (75%, 150 políticas) e conjunto de validação (25%, 50 políticas), mantendo a proporção de classes de cada setor em ambas as partições por meio de amostragem estratificada.

4.3 DEFINIÇÃO OPERACIONAL DE CONFORMIDADE

A avaliação granular dos oito requisitos descrita na seção anterior produz, para cada política, um vetor de classificações individuais que é diretamente apresentado ao usuário pela extensão de navegador. Para fins de comunicação, a extensão exibe

também um veredito agregado, conforme ou não conforme, que sintetiza o resultado da análise em uma única mensagem compreensível ao titular de dados. Esse veredito não é utilizado como rótulo nos experimentos de treinamento ou avaliação, cujas métricas operam sobre as classificações individuais de cada requisito; sua função é exclusivamente visual, como indicador de alto nível na interface do *popup*. Para derivá-lo de forma consistente com os princípios da LGPD, é necessário estabelecer uma regra de agregação que traduza o vetor de avaliações por requisito em um único veredito.

Essa tradução não é trivial, pois a LGPD não hierarquiza seus princípios nem atribui pesos distintos às diferentes obrigações de transparência. O artigo 6º da lei estabelece os princípios que regem o tratamento de dados pessoais de forma cumulativa: a organização deve observá-los em conjunto, não de maneira alternativa ou proporcional. Isso significa que o cumprimento exemplar de sete requisitos não compensa a omissão integral de um oitavo. A ausência de qualquer obrigação de transparência configura, por si só, descumprimento da lei, independentemente do nível de aderência aos demais.

Com base nesse entendimento, adota-se a seguinte definição operacional: uma política de privacidade é classificada como **conforme** somente quando nenhum dos oito requisitos for avaliado como ausente e, no máximo, um for avaliado como parcial, tolerância que torna a análise menos estrita diante de ambiguidades textuais pontuais, sem abandonar o caráter cumulativo, e não proporcional, dos princípios da LGPD. Nos demais casos, a política é classificada como **não conforme**. Formalmente, a regra de agregação é:

$$\text{conforme} = \begin{cases} \text{verdadeiro} & \text{se } n_{\text{ausente}} = 0 \text{ e } n_{\text{parcial}} \leq 1 \\ \text{falso} & \text{caso contrário} \end{cases} \tag{2}$$

onde n_{ausente} é o número de requisitos classificados como ausentes e n_{parcial} é o número de requisitos classificados como parciais.

A tolerância de um requisito parcial reconhece que certas obrigações admitem formulações textuais genuinamente ambíguas, nas quais a intenção de cumprimento está presente mas a execução é incompleta. A intransigência com qualquer ausência, por outro lado, reflete a natureza cumulativa dos princípios da LGPD: um documento que não identifica o controlador ou que omite o canal pelo qual o titular pode exercer seus direitos impede, na prática, o exercício dos demais direitos garantidos pela lei, independentemente de quão bem os outros requisitos estejam atendidos.

4.4 CRITÉRIO DE ACEITAÇÃO

O critério de aceitação estabelecido para este trabalho tem como finalidade determinar o ponto em que o modelo atinge um desempenho satisfatório na tarefa de

identificar a presença ou ausência de cada requisito nas políticas de privacidade. Considerando o contexto discutido na Fundamentação Teórica, especialmente a relevância dos erros de classificação expressos pela Matriz de Confusão, opta-se por utilizar uma métrica que reflita de maneira equilibrada o impacto de Falsos Positivos (FP) e, principalmente, Falsos Negativos (FN).

Adicionalmente, o dataset utilizado nos experimentos apresenta desbalanceamento de classes: das 200 políticas coletadas, 159 são classificadas como não conformes e 41 como conformes. Em cenários com distribuição assimétrica, métricas como acurácia tendem a ser enganosas, pois um modelo que sempre prediz a classe majoritária atingiria acurácia elevada sem qualquer capacidade discriminativa real. Por essa razão, adota-se o **F1-Score macro**, que calcula o F1 individualmente para cada classe e faz a média aritmética entre elas, conferindo peso igual à classe minoritária independentemente de sua frequência.

Sua escolha se justifica por combinar precisão e revocação em um único indicador, penalizando situações em que o modelo alerta para riscos inexistentes (FP), prejudicando a confiança do usuário, ou deixa de detectar elementos relevantes verdadeiramente presentes na política (FN), comprometendo a segurança jurídica do titular.

Considerando o valor de referência de 0,94 atingido por Xie, Ramakrishnan e Li (2025) em contexto distinto, com legislações em inglês e modelos de maior escala, e reconhecendo as particularidades do corpus brasileiro em português, estabelece-se como critério de aceitação para o presente projeto um F1-Score macro binário de 0,80, inferior ao valor de referência internacional em razão das limitações deste trabalho, notadamente a escala reduzida do dataset e a ausência de validação jurídica especializada em toda a sua extensão, avaliado sobre o conjunto de validação.

4.5 EXPERIMENTOS E BENCHMARK COMPARATIVO

Antes de conduzir o ajuste fino, foi realizado um benchmark comparativo entre múltiplos modelos de linguagem com o objetivo de identificar qual apresentava o melhor desempenho de base na tarefa de avaliação de políticas de privacidade e, consequentemente, qual seria o candidato mais adequado para a fase de especialização. Todos os modelos foram avaliados sobre o mesmo conjunto de validação, com os mesmos prompts e os mesmos critérios de métricas, garantindo comparabilidade direta entre os resultados.

4.5.1 Modelos Avaliados

Foram incluídos no benchmark oito modelos de diferentes famílias, abrangendo tanto soluções proprietárias acessíveis via API quanto modelos de código aberto. O critério de seleção priorizou modelos de menor porte e custo operacional reduzido,

adequados a cenários de produção com volume elevado de requisições, excluindo deliberadamente modelos de maior escala e custo, como o GPT-4.1 e o GPT-5.4 da OpenAI, cujo custo por inferência os torna economicamente inviáveis para implantação em escala e que, por hipótese, deveriam aproximar-se do desempenho do Claude Opus utilizado na anotação. Os modelos proprietários avaliados foram o GPT-4.1 mini e o GPT-4.1 nano, da OpenAI; o Gemini 2.5 Flash Lite e o Gemini 3.1 Flash Lite, do Google; o Claude Haiku 4.5, da Anthropic, modelo de menor custo da família Claude; e o Sabiá 4 Small, da Maritaca AI, único modelo do benchmark desenvolvido por empresa nacional, com pré-treinamento em corpus jurídico brasileiro e otimização nativa para o português do Brasil, incluído para avaliar se a especialização linguística e jurídica produz vantagem mensurável na tarefa de avaliação de conformidade com a LGPD. Entre os modelos de código aberto, foram avaliados o Llama 3.3 70B, da Meta, acessado gratuitamente via OpenRouter, e o Qwen 2.5 14B, da Alibaba, executado localmente no ambiente Google Colab com quantização de 4 bits.

4.5.2 Protocolo de Avaliação

Cada modelo foi submetido a dois cenários de prompting distintos, aplicados sobre o conjunto de validação de 50 políticas. Ambos os cenários compartilharam o mesmo *system prompt*:

```
Você é um especialista em conformidade com a LGPD (Lei 13.709/2018).
Analise o texto de uma política de privacidade e classifique cada requisito.
Responda APENAS com JSON válido. Sem explicações. Sem texto antes ou depois do JSON.
```

O primeiro cenário utilizou um **prompt simples**, sem exemplos. A instrução de usuário descrevia os oito requisitos a serem avaliados, definia os três rótulos possíveis (*presente*, *ausente* ou *parcial*), especificava o esquema exato do JSON de saída e delimitava o texto da política com tags `<policy>`. Critérios de desambiguação foram incluídos diretamente na definição dos requisitos, em especial para o campo *compartilhamento*, cuja regra instrui o modelo a classificar como *presente* qualquer menção a terceiros, independentemente de a política declarar ou não que realiza compartilhamento, reservando o rótulo *ausente* apenas para documentos que omitem o tema inteiramente.

O segundo cenário utilizou ***few-shot prompting***, inserindo cinco exemplos anotados de políticas reais, com seus respectivos rótulos JSON de referência, imediatamente antes da política alvo, seguindo o formato de conversação descrito na Seção 4.6.3. A hipótese do cenário few-shot é que demonstrações concretas da tarefa auxiliam o modelo a calibrar o limiar decisório para cada rótulo, reduzem ambiguidades em requisitos com formulação textual subjetiva como *base legal* e *retenção*, e diminuem erros de formato na geração do JSON (Brown et al., 2020). Esperava-se, portanto, que o

cenário few-shot produzisse resultados superiores ao prompt simples na maioria dos modelos, com ganhos proporcionais à dificuldade do modelo em inferir o comportamento esperado apenas a partir das instruções textuais. O texto integral de cada prompt, incluindo as instruções de classificação e os exemplos anotados utilizados no cenário few-shot, é reproduzido nos Apêndices A e B.

As métricas registradas para cada combinação de modelo e cenário foram o F1-Score macro binário, utilizado como métrica principal de comparação por ser a métrica de aceitação definida na Seção 4.4, e o custo em dólares americanos por execução completa sobre o conjunto de validação. Esse custo foi calculado a partir do preço por token de entrada e por token de saída publicado por cada provedor, multiplicado pelo número total de tokens efetivamente consumidos pelo modelo ao processar as 50 políticas do conjunto de validação naquele cenário, registrado para os modelos proprietários como indicador de viabilidade econômica em escala.

4.5.3 Resultados do Benchmark

Os resultados obtidos nos dois cenários de prompting são apresentados nas Tabelas 3 e 4.

Tabela 3 – Resultados do benchmark com prompt simples

Modelo	**F1 macro binário**	**Custo (USD)**
Gemini 3.1 Flash Lite	**0,7718**	$0,0368
Qwen 2.5 14B baseline	0,7577	-
Gemini 2.5 Flash Lite	0,7212	$0,0142
GPT-4.1 mini	0,7157	$0,0543
Claude Haiku 4.5	0,7005	$0,1472
Sabiá 4 Small	0,6879	$0,0322
Llama 3.3 70B	0,6149	$0,00
GPT-4.1 nano	0,5695	$0,0136

Fonte: Elaborado pelo autor.

Tabela 4 – Resultados do benchmark com few-shot prompting

Modelo	**F1 macro binário**	**Custo (USD)**
Qwen 2.5 14B baseline	**0,7937**	-
Gemini 3.1 Flash Lite	0,7316	$0,0432
GPT-4.1 mini	0,7269	$0,0640
Claude Haiku 4.5	0,7267	$0,1684
Gemini 2.5 Flash Lite	0,7141	$0,0167
Sabiá 4 Small	0,6860	$0,0378
Llama 3.3 70B	0,6340	$0,00
GPT-4.1 nano	0,5541	$0,0160

Fonte: Elaborado pelo autor.

No cenário de prompt simples (Tabela 3), o Gemini 3.1 Flash Lite obteve o melhor resultado com F1 macro de 0,7718, seguido pelo Qwen 2.5 14B com 0,7577. O Sabiá 4 Small, único modelo com especialização em corpus jurídico brasileiro, atingiu F1 de 0,6879, abaixo de todos os modelos proprietários multilíngues avaliados, apesar de ter superado o Llama 3.3 70B (0,6149) e o GPT-4.1 nano (0,5695), o que sugere que a especialização linguística e jurídica não compensou, neste cenário, diferenças de capacidade geral de seguimento de instruções em relação aos modelos proprietários de maior porte.

No cenário few-shot (Tabela 4), o Qwen 2.5 14B assumiu a liderança com F1 de 0,7937, beneficiando-se mais dos exemplos contextuais do que os demais modelos. O Gemini 3.1 Flash Lite, que liderou no cenário simples, caiu para 0,7316 com a adição dos exemplos, evidenciando que o few-shot nem sempre produz ganhos uniformes e que, para modelos com forte capacidade de instrução, os exemplos podem introduzir ruído ou conflito com padrões já internalizados. O Sabiá 4 Small apresentou variação marginal entre os dois cenários (0,6879 simples vs. 0,6860 few-shot), indicando que os exemplos adicionais não alteraram substantivamente sua calibração. Esses resultados em conjunto evidenciam que escala e custo não são garantias de desempenho superior em tarefas de domínio específico, e que a distância entre o melhor modelo do benchmark e o Claude Opus utilizado na anotação justifica a separação metodológica entre anotador e modelos avaliados.

4.5.4 Seleção do Modelo para Fine-Tuning

Com base nos resultados do benchmark, o Qwen 2.5 14B foi selecionado como modelo base para a fase de ajuste fino. Três fatores orientaram essa decisão. O primeiro é o desempenho de base: entre todos os modelos avaliados, o Qwen 2.5 14B apresentou o melhor F1 macro no cenário de few-shot sem qualquer especializa-

ção adicional, sugerindo que sua distribuição de probabilidades já está relativamente alinhada com a tarefa. O segundo fator é a natureza de código aberto do modelo: diferentemente dos modelos proprietários, o Qwen 2.5 14B tem seus pesos publicamente disponíveis, o que viabiliza a aplicação de fine-tuning supervisionado e garante controle total sobre o processo de especialização, sem dependência de infraestrutura de terceiros para treinar. O terceiro fator é a viabilidade computacional: com 14 bilhões de parâmetros e suporte a quantização de 4 bits via biblioteca Unsloth, o modelo pode ser carregado e treinado inteiramente na GPU disponível no Google Colab Pro, com 102 GB de VRAM, o que não seria possível com modelos open-source de maior porte, como o Llama 3.3 com 70 bilhões de parâmetros. Adicionalmente, o Qwen 2.5 foi treinado em dados multilíngues com representação substantiva do português, o que reduz a assimetria linguística que modelos treinados majoritariamente em inglês tendem a apresentar ao analisar documentos jurídicos brasileiros.

4.6 PROCESSO DE AJUSTE FINO

Com o modelo base selecionado, a fase de ajuste fino teve como objetivo especializar o Qwen 2.5 14B no domínio específico da análise de conformidade de políticas de privacidade com a LGPD, internalizando o vocabulário jurídico, o formato de saída estruturado e os padrões de raciocínio esperados para a tarefa.

4.6.1 Infraestrutura e Configuração do Ambiente

O treinamento foi conduzido no ambiente Google Colab Pro, utilizando o runtime G4, equipado com GPU NVIDIA RTX PRO 6000 Blackwell Server Edition, com 102 GB de VRAM e 180 GB de RAM. O modelo foi carregado com quantização de 4 bits (*load_in_4bit*), reduzindo o consumo de memória a aproximadamente um quarto do que seria necessário em precisão total, viabilizando tanto o carregamento do modelo quanto a manutenção das matrizes de adaptação LoRA em memória simultaneamente durante o treinamento. A biblioteca Unsloth foi utilizada para otimizar o carregamento e o processo de treinamento, com suporte a *gradient checkpointing* para redução adicional do consumo de memória sem comprometer a propagação dos gradientes.

4.6.2 Configuração do LoRA e Exploração de Hiperparâmetros

O ajuste fino foi conduzido com a técnica LoRA, descrita na Seção 2.4.3, com matrizes de adaptação injetadas nas camadas de atenção do modelo, especificamente nas projeções de query, key, value e output (`q_proj`, `k_proj`, `v_proj`, `o_proj`). Foram realizados três experimentos variando o posto das matrizes LoRA, a fim de identificar a configuração que maximizava o F1 macro no conjunto de validação sem incorrer em sobreajuste.

O primeiro experimento utilizou posto $r = 16$, com taxa de aprendizado de 1×10^{-5} e três épocas de treinamento, concluído em 118,5 segundos. A curva de *loss* de validação apresentou convergência consistente ao longo das épocas, decaindo de 1,252 na primeira época para 1,201 na segunda e 1,184 na terceira. Apesar da convergência aparentemente saudável, o F1 macro binário no conjunto de validação foi de 0,7685, abaixo do baseline few-shot sem ajuste fino (0,7937). O resultado revelou um padrão importante: a *loss* de validação não é um indicador suficiente do desempenho na tarefa de classificação, pois o modelo pode aprender a reproduzir o formato e a distribuição geral dos rótulos sem necessariamente melhorar sua capacidade discriminativa entre as classes conforme e não conforme.

O segundo experimento elevou o posto para $r = 32$ e estendeu o treinamento para cinco épocas, concluído em 199,2 segundos. A *loss* de validação continuou decrescendo ao longo de todas as épocas, de 1,663 na primeira até 1,458 na quinta, sem dar sinais de estabilização. Esse comportamento é característico de sobreajuste progressivo: o modelo continuava ajustando os pesos às particularidades do corpus de treinamento, mas sem ganho de generalização. O F1 macro binário de 0,6366 confirmou o diagnóstico, representando queda de 0,1571 pontos em relação ao baseline, com degradação pronunciada em requisitos de menor frequência no corpus, como base legal e direitos. A evolução da *loss* de validação nos três experimentos é apresentada na Tabela 5.

Tabela 5 – Evolução da *loss* de validação por época nos experimentos de ajuste fino

Configuração	**Ép. 1**	**Ép. 2**	**Ép. 3**	**Ép. 4**	**Ép. 5**	**F1 final**
$r = 16$, 3ep	1,252	1,201	1,184	-	-	0,7685
$r = 32$, 5ep	1,663	1,568	1,507	1,471	1,458	0,6366
$r = 8$, 4ep	1,264	1,214	1,170	1,154	-	**0,8409**

Fonte: Elaborado pelo autor.

O padrão observado é consistente com o que a literatura de PEFT descreve para corpus de tamanho reduzido (Ding et al., 2023): postos maiores ampliam a capacidade de memorização das matrizes de adaptação, o que pode ser benéfico com datasets grandes mas prejudicial quando o conjunto de treinamento não é suficientemente diverso para sustentar essa capacidade adicional. A *loss* de validação do experimento $r = 32$ nunca estabilizou, indicando que o modelo não encontrou um mínimo generalizável dentro das cinco épocas, e o colapso do F1 macro de 0,7937 para 0,6366 confirmou o sobreajuste.

A configuração final vencedora utilizou posto $r = 8$, com *lora_alpha* igual a 16, *dropout* de 0,05 e taxa de aprendizado de 1×10^{-5}. O treinamento foi conduzido por quatro épocas com tamanho de lote efetivo de oito exemplos, obtido via acumulação

de gradiente em quatro passos com dois exemplos por dispositivo. A estratégia de *early stopping* com carregamento automático do melhor *checkpoint* com base na *loss* de validação foi adotada para mitigar o risco de sobreajuste residual. Os principais hiperparâmetros da configuração final são sumarizados na Tabela 6.

Tabela 6 – Hiperparâmetros da configuração final de ajuste fino

Hiperparâmetro	**Valor**
Posto LoRA (r)	8
lora_alpha	16
lora_dropout	0,05
Camadas alvo	q_proj, k_proj, v_proj, o_proj
Taxa de aprendizado	1×10^{-5}
Épocas	4
Tamanho de lote efetivo	8
Warmup ratio	0,10
Weight decay	0,01
Max gradient norm	0,3
Comprimento máximo de sequência	32.768 tokens
Quantização	4 bits

Fonte: Elaborado pelo autor.

4.6.3 Formato dos Dados de Treinamento

Os exemplos de treinamento foram formatados no padrão de conversação do Qwen 2.5, com três turnos: o *system prompt* definindo a persona de especialista em LGPD; a mensagem do usuário contendo as instruções de classificação em formato few-shot, com cinco exemplos anotados, seguidas do texto da política a ser avaliada delimitado por tags `<policy>`; e a resposta esperada do assistente em JSON compacto com os oito requisitos. Esse formato instrui o modelo não apenas sobre o conteúdo da tarefa, mas também sobre o padrão exato de saída esperado, reduzindo erros de formatação durante a inferência.

Os rótulos de saída foram simplificados em relação ao formato de anotação original: cada requisito foi representado por um único valor de texto (*presente*, *ausente* ou *parcial*), sem campos adicionais de justificativa, mantendo o JSON compacto e reduzindo o risco de o modelo gerar conteúdo fora do esquema esperado.

4.6.4 Resultados do Ajuste Fino

O modelo ajustado com a configuração final atingiu F1 macro binário de **0,8409** no conjunto de validação, representando um ganho de 0,0472 pontos em relação ao baseline few-shot do mesmo modelo (0,7937) e superando todos os modelos proprietários avaliados no benchmark, independentemente do cenário de prompting e do custo por inferência. A Tabela 7 consolida a evolução do desempenho ao longo dos experimentos.

Tabela 7 – Evolução do desempenho do Qwen 2.5 14B ao longo dos experimentos

Configuração	**F1 macro binário**
Baseline, prompt simples	0,7577
Baseline, few-shot (5 exemplos)	0,7937
Fine-tuning $r = 16$, lr=1×10^{-5}, 3 épocas	0,7685
Fine-tuning $r = 32$, lr=1×10^{-5}, 5 épocas	0,6366
Fine-tuning $r = 8$, lr=1×10^{-5}, 4 épocas	**0,8409**
Referência, melhor resultado proprietário do benchmark	
Gemini 3.1 Flash Lite, prompt simples	0,7718

Fonte: Elaborado pelo autor.

O resultado final de 0,8409 supera o critério de aceitação de 0,80 estabelecido na Seção 4.4, validando a abordagem de especialização adotada. O modelo fine-tunado supera o melhor resultado individual de qualquer modelo proprietário no benchmark, 0,7718 do Gemini 3.1 Flash Lite no cenário de prompt simples, e supera todos os proprietários em ambos os cenários de prompting, incluindo modelos de custo operacional muito superior, como o Claude Haiku 4.5 ($0,17 por execução completa no conjunto de validação). A trajetória dos experimentos intermediários reforça que a seleção de hiperparâmetros tem impacto determinante no resultado: um posto LoRA elevado em corpus de escala moderada não apenas deixa de melhorar o desempenho, como pode degradá-lo substancialmente em relação ao baseline sem ajuste fino. O resultado é alinhado com as recomendações de Hu et al. (2022) para tarefas de classificação com datasets de escala moderada, que indicam postos baixos como suficientes para capturar as atualizações de peso necessárias sem comprometer a generalização do modelo base.

5 DESENVOLVIMENTO

Com a estratégia metodológica definida e os experimentos de ajuste fino concluídos, a etapa de desenvolvimento compreendeu a construção do sistema de software que disponibiliza a capacidade de análise ao usuário final. O sistema é composto por dois componentes principais: uma API de backend responsável pelo processamento das políticas e pela comunicação com o modelo de linguagem, e uma extensão de navegador que serve como interface de interação com o titular de dados durante sua navegação na web. O código-fonte completo do sistema, incluindo a API, a extensão de navegador e os scripts de curadoria do dataset, benchmark e ajuste fino, está disponível publicamente (Camerini, 2026).

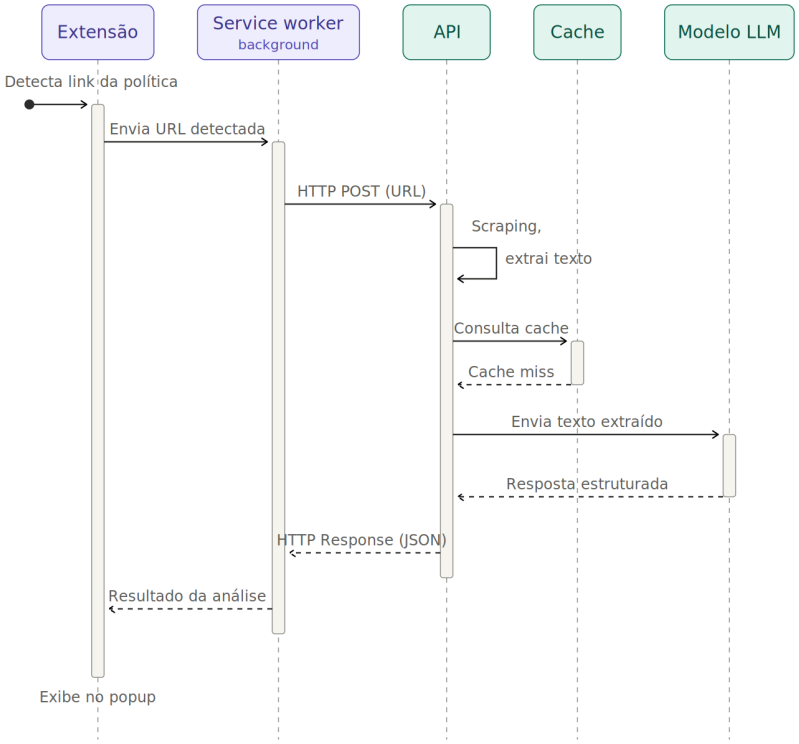
5.1 ARQUITETURA DO SISTEMA

A arquitetura adotada segue o padrão cliente-servidor, com separação clara de responsabilidades entre os dois componentes. A extensão de navegador atua exclusivamente como camada de apresentação e detecção, sem realizar qualquer processamento do conteúdo das políticas. Toda a lógica de negócio, incluindo o scraping, a preparação do texto, a comunicação com o modelo de linguagem e o gerenciamento de cache, é centralizada na API de backend.

Essa separação oferece duas vantagens práticas. Primeiramente, mantém a extensão leve e com superfície de ataque reduzida, uma vez que ela não armazena nem processa dados sensíveis das políticas. Em segundo lugar, permite que a API seja atualizada independentemente, possibilitando a substituição do modelo de linguagem ou ajustes no pipeline de análise sem que o usuário precise reinstalar ou atualizar a extensão.

O fluxo de uma requisição percorre as seguintes etapas: o script de conteúdo da extensão detecta um link de política de privacidade na página visitada pelo usuário e o envia ao *background service worker*; o *service worker* encaminha a URL à API via requisição HTTP POST; a API realiza o *scraping* da página, extrai o texto, consulta o cache e, se necessário, envia o conteúdo ao modelo de linguagem; a resposta estruturada retorna à extensão, que a exibe ao usuário no *popup*. A Figura 3 ilustra esse fluxo de forma esquemática.

Figura 3 – Fluxo de dados entre extensão e API



Fonte: Elaborado pelo autor.

5.2 BACKEND

A API de backend foi desenvolvida com o framework FastAPI para Python, escolhido por seu suporte nativo a operações assíncronas, pela geração automática de documentação OpenAPI e pela integração direta com Pydantic para validação de entradas. O servidor ASGI utilizado é o Uvicorn. A aplicação foi implantada na plataforma Railway, que gerencia o ambiente de execução, as variáveis de ambiente e o certificado TLS da aplicação sem necessidade de configuração de infraestrutura manual.

5.2.1 Endpoints

A API expõe os seguintes endpoints públicos principais:

- `GET /health`: verificação de disponibilidade do serviço, utilizada pela extensão para determinar se deve rotear as requisições para a instância local ou para o servidor em produção.
- `POST /privacy-policy`: endpoint principal. Recebe um objeto JSON com o campo `url`, realiza o *scraping* da política, executa a análise e retorna o resultado estruturado.

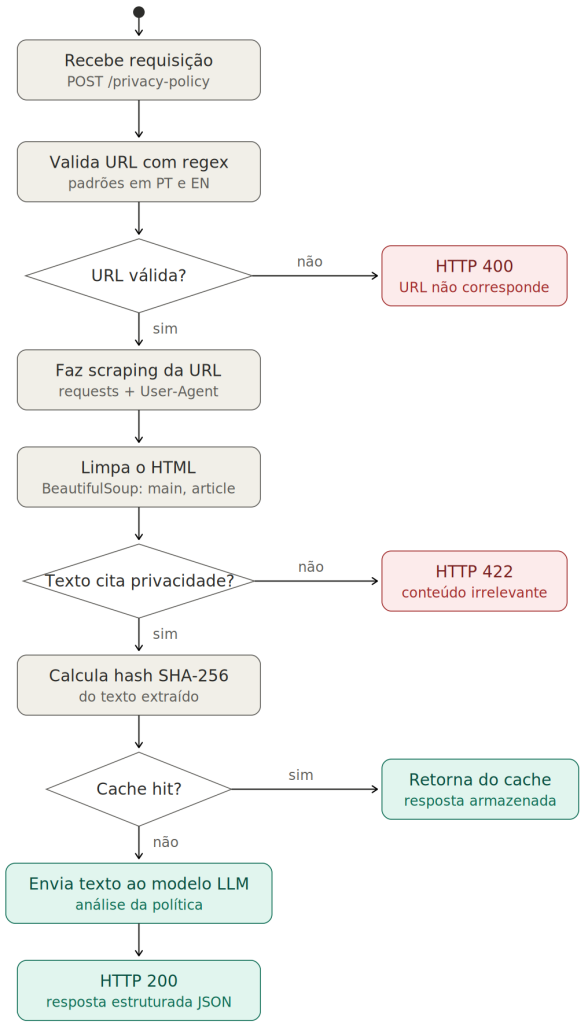
5.2.2 Pipeline de Processamento

Ao receber uma requisição no endpoint `POST /privacy-policy`, a API executa um pipeline de validação e processamento em etapas sequenciais. Primeiro, a URL é verificada por uma expressão regular que identifica padrões lexicais associados a políticas de privacidade em português e inglês, como *política de privacidade*, *aviso de privacidade* e *privacy policy*. Requisições com URLs que não correspondam a esses padrões são rejeitadas imediatamente com código HTTP 400, evitando o processamento de páginas arbitrárias.

Em seguida, a API realiza o *scraping* da URL com a biblioteca `requests`, utilizando um *User-Agent* de navegador para evitar bloqueios por parte de servidores que filtram requisições automatizadas. O conteúdo HTML retornado é processado com BeautifulSoup, que remove elementos não textuais como scripts, estilos, elementos de navegação e rodapés, preservando o conteúdo semântico principal a partir da hierarquia `<main>`, `<article>` ou `<body>` do documento. O texto limpo é então verificado quanto ao seu conteúdo: se não contiver referências lexicais à privacidade de dados, a requisição é rejeitada com HTTP 422, filtrando páginas que contenham a palavra privacidade na URL mas não no corpo do documento.

Um hash SHA-256 do texto extraído é calculado e consultado no banco de dados de cache antes de qualquer chamada ao modelo. Se um resultado recente para aquele conteúdo já estiver armazenado, ele é retornado diretamente, reduzindo o tempo de resposta e o custo operacional para políticas consultadas repetidamente. A Figura 4 sintetiza as etapas de validação e processamento descritas nesta subseção.

Figura 4 – Pipeline de validação e processamento da API



Fonte: Elaborado pelo autor.

5.2.3 Módulo de Análise

O módulo de análise (`llm.py`) implementa uma classe abstrata `LLMBackend` que desacopla a lógica de análise do modelo específico utilizado. A implementação padrão em produção é a `AnthropicBackend`, que se comunica com o modelo Claude Opus 4.6 via SDK oficial da Anthropic, o mesmo modelo utilizado na geração dos silver labels durante a construção do dataset. A arquitetura suporta a adição de backends alternativos sem alteração do pipeline principal, o que viabiliza a substituição futura pelo modelo Qwen 2.5 14B fine-tunado descrito na Seção 4.6.

Para políticas de privacidade extensas, cujo texto exceda o limiar de 600.000 caracteres, o módulo implementa uma estratégia de fragmentação (*chunking*): o texto é dividido em segmentos de 60.000 caracteres com sobreposição de 500 caracteres entre fragmentos consecutivos, a fim de preservar o contexto em trechos que cruzem os limites de divisão. Cada fragmento é analisado independentemente, e os resultados

são mesclados segundo a estratégia de maximização: para cada requisito, é mantida a avaliação mais favorável entre todos os fragmentos, na ordem ausente $<$ parcial $<$ presente.

A conformidade geral é calculada pela função `_is_conformant`, que aplica exatamente o critério definido na Seção 4.3: a política é conforme se nenhum requisito for ausente e no máximo um for parcial. Adicionalmente, um escore numérico percentual é calculado com base nos pesos 1,0 para presente, 0,5 para parcial e 0,0 para ausente, oferecendo uma medida contínua de aderência além do veredito binário.

5.2.4 Cache e Persistência

O cache é implementado sobre um banco de dados PostgreSQL provisionado pelo Railway. A tabela `policy` armazena, para cada par de hash de conteúdo e modelo utilizado, o resultado da análise, a URL de origem, a data de processamento e o veredito de conformidade. A validade das entradas é de sete dias: requisições com conteúdo idêntico dentro desse período são respondidas diretamente a partir do banco, sem nova chamada ao modelo. O mecanismo de *upsert* garante que políticas atualizadas, cujo conteúdo altere o hash, sejam reanalisadas automaticamente.

5.3 EXTENSÃO DE NAVEGADOR

A extensão foi desenvolvida seguindo o padrão Manifest V3 da API de extensões do Chrome, compatível com os principais navegadores baseados no motor Chromium. É composta por três componentes: um *content script*, um *background service worker* e um *popup*.

5.3.1 Detecção Automática de Políticas

O *content script* (`script/popup.js`) é injetado em todas as páginas visitadas pelo usuário após o carregamento completo do DOM (*document_idle*), conforme declarado no `manifest.json`. Sua função principal é identificar o link de política de privacidade mais relevante na página e encaminhá-lo ao *service worker*.

A detecção opera em duas etapas sequenciais. Na primeira, o script verifica se a URL da página corrente corresponde a um padrão lexical de política de privacidade por meio da função `isPrivacyPage()`. Caso afirmativo, a URL corrente é imediatamente enviada ao *service worker* sem varredura de links: o usuário já se encontra na página da política. Na segunda etapa, executada apenas quando a página corrente não é identificada como uma política, o script varre todos os elementos `<a href>` do DOM e atribui uma pontuação a cada um por meio da função `getPrivacyScore()`.

A pontuação é calculada com base em dois padrões de expressão regular. O primeiro, `privacyPattern`, é um padrão fraco que reconhece um conjunto amplo de

termos relacionados à privacidade em português e inglês, incluindo variantes ortográficas:

```
/(?:privacy(?:\s+policy|\s+notice)?|pol[ií]tica(?:\s+de)?\s+privacidade
  |privacidade|aviso\s+de\s+privacidade)/i
```

O segundo, `strongPrivacyPattern`, é um padrão restrito que exige a forma completa da expressão, como *política de privacidade* ou *privacy policy*:

```
/(?:privacy\s+policy|privacy\s+notice|pol[ií]tica(?:\s+de)?\s+privacidade
  |aviso\s+de\s+privacidade)/i
```

Para cada link, o algoritmo concatena o texto âncora e o atributo `href` em uma cadeia de avaliação. Se o padrão fraco não encontrar correspondência nessa cadeia, o link é descartado com pontuação -1. Caso contrário, a pontuação inicial é 1, acrescida de 2 pontos se o padrão forte corresponder à cadeia, e de 1 ponto adicional se o padrão fraco corresponder especificamente ao texto âncora (independentemente do `href`). A pontuação máxima alcançável é 4, atribuída a um link cujo texto âncora contenha a forma completa da expressão, como “Política de Privacidade”. Ao final da varredura, o link com maior pontuação é selecionado como candidato; em caso de empate, prevalece o primeiro encontrado no DOM. A URL do candidato é enviada ao *service worker* via `chrome.runtime.sendMessage`; na ausência de qualquer candidato elegível, uma mensagem do tipo `privacy-policy-not-found` é transmitida para que o *popup* informe o usuário.

5.3.2 Service Worker

O *background service worker* (`background.js`) é o componente central de orquestração da extensão. Ao receber a URL detectada pelo *content script*, verifica primeiro o cache de sessão do navegador (`chrome.storage.session`) para o identificador da aba atual. Caso um resultado recente esteja disponível, ele é retransmitido diretamente ao *popup* sem nova requisição à API.

O *service worker* também implementa uma lógica de seleção automática de ambiente: ao inicializar, tenta alcançar a instância local da API em `http://localhost:8080/health` com um *timeout* de 500 milissegundos. Se a instância local responder, ela é utilizada para as requisições subsequentes, o que facilita o desenvolvimento e os testes locais. Caso contrário, as requisições são roteadas para o servidor de produção em `https://policy-check.up.railway.app`. Quando o usuário navega para uma nova URL em uma aba, o cache de sessão daquela aba é automaticamente invalidado, garantindo que a análise exibida sempre corresponda ao conteúdo atual da página.

5.3.3 Interface do Popup

O *popup* é a interface visível ao usuário, ativada pelo clique no ícone da extensão na barra do navegador. Ao abri-lo, o usuário visualiza um botão de ação principal, *Verificar Política de Privacidade*, e deve clicar nele para que a extensão encaminhe a URL detectada pelo *content script* para análise (Figura 5). A interação é deliberadamente explícita: nenhuma requisição à API é disparada sem ação do usuário, o que preserva a privacidade de navegação e evita chamadas desnecessárias em páginas sem política identificável.

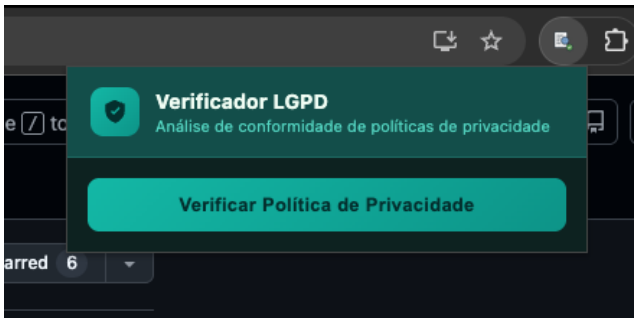
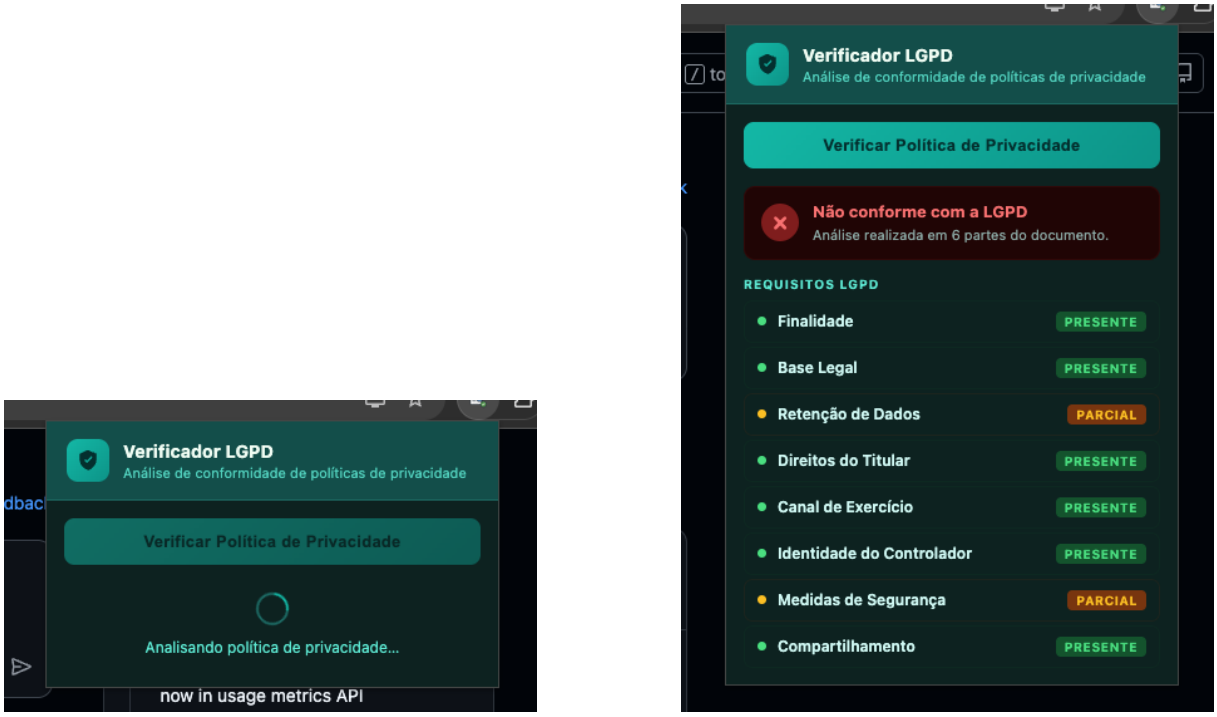


Figura 5 – Estado inicial do popup, aguardando ação do usuário

Fonte: Elaborado pelo autor.

Após o clique, a extensão entra em estado de carregamento enquanto a API processa a política. Concluída a análise, o *popup* exibe dois elementos principais: um indicador de veredito global informando conformidade ou não conformidade com a LGPD; e uma lista dos oito requisitos avaliados, cada um com um indicador visual de status codificado por cor (*presente*, *parcial* ou *ausente*). O usuário pode expandir individualmente cada requisito para ler o raciocínio que embasou a classificação, tornando a avaliação auditável e não apenas um veredito opaco. A Figura 6 ilustra os dois estados intermediários desse fluxo.



(a) Análise em andamento (b) Resultado da análise

Figura 6 – Estados do popup após acionamento pelo usuário

Fonte: Elaborado pelo autor.

6 CONCLUSÃO

A assimetria informacional entre titulares de dados e controladores constitui um dos problemas centrais da era digital: embora a LGPD estabeleça obrigações claras de transparência, a efetividade dessas obrigações depende da capacidade do usuário de compreender os documentos que as materializam. Políticas de privacidade são, em sua maioria, extensas, redigidas em linguagem técnico-jurídica e raramente lidas por quem deveria ser seu destinatário principal. O resultado prático é que o consentimento informado, pressuposto jurídico do tratamento de dados em diversas hipóteses da lei, frequentemente se reduz a um clique em “aceitar” sem que o titular tenha condições reais de avaliar o que está autorizando.

Este trabalho propôs uma resposta técnica a esse problema: um sistema automatizado de avaliação de conformidade de políticas de privacidade com a LGPD, acessível ao usuário comum por meio de uma extensão de navegador. A solução foi construída em três etapas complementares. Primeiro, foi elaborado um dataset original composto por 200 políticas de privacidade brasileiras, com representação de oito setores da economia e da administração pública, anotadas automaticamente por um modelo de linguagem de alta capacidade e parcialmente validadas por revisão manual. Segundo, foi conduzido um benchmark comparativo entre oito modelos de linguagem menores em dois cenários de prompting, com o duplo objetivo de identificar o modelo base mais adequado para especialização e de estabelecer uma linha de referência de desempenho sem ajuste fino. Terceiro, o modelo selecionado, o Qwen 2.5 14B, foi submetido a ajuste fino supervisionado com a técnica LoRA, com exploração sistemática de hiperparâmetros para identificar a configuração que maximizava a generalização sem sobreajuste.

6.1 RESULTADOS

Os resultados obtidos validam a abordagem. A configuração final de ajuste fino, com posto $r = 8$ e quatro épocas de treinamento, atingiu F1 macro binário de 0,8409 no conjunto de validação, superando o critério de aceitação de 0,80 estabelecido a priori e ultrapassando todos os modelos proprietários avaliados no benchmark, incluindo modelos de custo operacional substancialmente superior. A trajetória dos experimentos intermediários revelou um padrão relevante para trabalhos futuros na área: em corpus de escala moderada, postos LoRA elevados ampliam a capacidade de memorização além do que o conjunto de treinamento pode sustentar, degradando o desempenho abaixo da linha de base sem ajuste fino, o experimento com $r = 32$ e cinco épocas produziu F1 de 0,6366, queda de 0,1571 pontos em relação ao baseline. A sensibilidade do resultado ao posto LoRA reforça a necessidade de avaliação empírica dos hiperparâmetros de PEFT mesmo em tarefas aparentemente bem delimitadas.

Do ponto de vista das contribuições, o trabalho produziu três artefatos de uso independente. O dataset anotado de 200 políticas de privacidade brasileiras constitui, até onde se sabe, o primeiro corpus público voltado especificamente à avaliação de conformidade com a LGPD, e pode servir de base para pesquisas futuras em análise automatizada de privacidade no contexto do direito brasileiro. O sistema composto pela API de backend e pela extensão de navegador é o único resultado imediatamente utilizável pelo público não técnico: permite que qualquer usuário, sem conhecimento jurídico ou computacional, obtenha em segundos uma avaliação granular dos oito requisitos de transparência da LGPD para a política de privacidade do site que está visitando. O benchmark comparativo, por sua vez, oferece uma referência quantitativa sobre o desempenho de modelos de linguagem multilíngues e nacionais na tarefa de avaliação de conformidade em português, evidenciando que a especialização linguística e jurídica do Sabiá 4 Small não produziu vantagem mensurável sobre modelos proprietários multilíngues de porte comparável.

Em comparação com as iniciativas discutidas no Capítulo 3, o sistema desenvolvido supera limitações que impediram a adoção em escala de propostas anteriores: o P3P (Seção 3.0.1) fracassou por depender de um formato rígido e da adesão voluntária das próprias empresas, o ToS;DR (Seção 3.0.2) permanece limitado pela dependência de curadoria humana não escalável, e o POLIGRAPH (Seção 3.0.3) esbarra nas limitações de pipelines sintáticos tradicionais de PLN diante de nuances semânticas. Ao acoplar uma LLM especializada a uma extensão de navegador, o sistema aqui proposto combina a automação em escala buscada por essas iniciativas com a robustez interpretativa que apenas modelos de linguagem oferecem, aproximando-se da metodologia de Xie, Ramakrishnan e Li (2025) (Seção 3.0.4), porém voltado especificamente ao ordenamento jurídico brasileiro, ainda não coberto por aquele estudo.

6.2 LIMITAÇÕES

O trabalho apresenta limitações que devem ser consideradas na interpretação dos resultados. A qualidade dos rótulos depende do Claude Opus 4.6 como anotador de referência: embora sua superioridade em relação aos demais modelos avaliados justifique sua escolha, os rótulos de *silver labeling* não possuem a autoridade de uma anotação conduzida por especialistas em direito digital, e a revisão manual realizada cobriu apenas 8% do dataset. O tamanho do corpus, com 200 políticas e desbalanceamento pronunciado entre classes (79,5% não conformes), limita a capacidade de generalização do modelo fine-tunado para setores e perfis de empresa sub-representados, como startups, organizações do terceiro setor e empresas de pequeno porte. Por fim, o modelo implantado em produção na API é atualmente o Claude Opus 4.6, não o Qwen 2.5 14B fine-tunado: a arquitetura do backend suporta a substituição sem alteração no pipeline, mas a transição ainda não foi realizada, de modo que os ganhos

de especialização obtidos no ajuste fino ainda não beneficiam os usuários finais da extensão.

Apesar dessas limitações, o conjunto dos resultados demonstra que é viável construir, com recursos computacionais disponíveis em serviços de nuvem com custo razoável e sem anotação especializada em escala, um sistema capaz de avaliar automaticamente a aderência de políticas de privacidade aos principais requisitos de transparência da LGPD com desempenho superior ao de modelos proprietários de uso geral. A extensão de navegador torna essa capacidade acessível de forma não intrusiva, sem exigir que o usuário copie textos ou navegue por interfaces externas, aproximando a promessa de transparência da lei da experiência concreta de quem a ela deveria ser protegido.

6.3 TRABALHOS FUTUROS

A partir dos resultados e limitações identificados, cinco direções se destacam como extensões naturais deste trabalho.

6.3.1 Padronização da Localização de Políticas de Privacidade

Durante a curadoria do dataset, a ausência de um padrão estruturado para a publicação desses documentos exigiu o acesso manual às URLs, inviabilizando a coleta automatizada em grande escala. Uma contribuição de infraestrutura relevante seria a proposta de um URI padronizado seguindo a convenção `.well-known` da RFC 8615 (Nottingham, 2019), por exemplo `/.well-known/privacy-policy`, que permitiria a qualquer agente automatizado localizar a política de privacidade de um domínio sem análise do HTML da página principal. Essa convenção, se adotada amplamente, também reduziria a carga de heurísticas de detecção necessárias na extensão e na API.

6.3.2 Expansão do Dataset

As 200 políticas atuais, embora suficientes para a fase experimental deste trabalho, são insuficientes para treinar modelos com alta capacidade de generalização para perfis de organização não representados, como pequenas e médias empresas, startups e entidades do terceiro setor, que tendem a apresentar políticas com menor elaboração jurídica e maior variabilidade de formato. Ampliar o corpus para 500 ou mais políticas, com amostragem mais equilibrada entre classes e setores, permitiria reduzir o desbalanceamento atual e avaliar o impacto da escala do dataset no desempenho do modelo fine-tunado.

6.3.3 Validação Jurídica dos Rótulos

A revisão manual realizada neste trabalho foi conduzida sem formação jurídica especializada, o que limita a confiabilidade dos julgamentos em requisitos com maior carga interpretativa, como base legal e retenção. A participação de advogados ou especialistas em direito digital na revisão ou anotação de uma amostra representativa permitiria estimar o grau de concordância entre o modelo anotador e a interpretação jurídica qualificada, bem como identificar sistematicamente os requisitos em que o *silver labeling* introduz maior ruído. Essa estimativa é necessária para determinar até que ponto melhorias adicionais no modelo são limitadas pela qualidade do anotador de referência.

6.3.4 Revisão do Esquema de Categorização

O esquema ternário adotado para cada requisito (*presente*, *parcial*, *ausente*) e o veredito binário de conformidade derivado dele, descrito na Seção 4.3, simplificam a comunicação ao titular de dados, mas também comprimem gradações relevantes: duas políticas podem receber o mesmo rótulo de *não conforme* mesmo apresentando graus de aderência à LGPD substancialmente diferentes. A API já calcula, como subproduto do módulo de análise, um escore contínuo de aderência a partir de pesos atribuídos a cada rótulo (Seção 5.2), mas esse indicador não é atualmente exposto ao usuário nem validado como alternativa às categorias discretas. Um trabalho futuro poderia investigar formas alternativas de categorização, como uma pontuação contínua ou uma escala ordinal com mais níveis, avaliando empiricamente se essas representações comunicam de forma mais fiel o grau de conformidade de uma política do que o veredito binário atualmente adotado.

6.3.5 API como Apoio à Elaboração de Políticas de Privacidade

O sistema desenvolvido foi concebido para uso pelo titular de dados, que consulta a análise de uma política já publicada. Uma direção complementar de trabalho futuro é disponibilizar a mesma API de análise como ferramenta de apoio a quem elabora políticas de privacidade, permitindo que equipes jurídicas e de produto submetam rascunhos do documento durante sua redação e recebam, em tempo real, a avaliação granular dos oito requisitos de transparência antes da publicação. Essa aplicação inverteria o público-alvo do sistema, do titular de dados para o controlador, e poderia contribuir para reduzir, na origem, a proporção de políticas não conformes identificada no dataset construído neste trabalho (Seção 4.2).

REFERÊNCIAS

AMOS, Ryan et al. Privacy Policies over Time: Curation and Analysis of a Million-Document Dataset. In: PROCEEDINGS of the Web Conference 2021. [S.l.]: ACM, 2021. p. 2165–2176. DOI: 10.1145/3442381.3450048. Disponível em: <https://arxiv.org/abs/2008.09159>. Citado na p. 40.

AUTORIDADE NACIONAL DE PROTEÇÃO DE DADOS. **ANPD publica Regulamento de Dosimetria**. [S.l.*:* s.n.], fev. 2023. Acessado em: 21 Nov. 2025. Disponível em: <https://www.gov.br/anpd/pt-br/assuntos/noticias/anpd-publica-regulamento-de-dosimetria>. Citado nas pp. 23, 29.

BELLI, Luca. **O maior vazamento de dados pessoais na história brasileira e quais lições devemos aprender**. [S.l.*:* s.n.], fev 2021. Portal FGV. Acedido em: 29 Nov. 2025. Disponível em: <https://portal.fgv.br/artigos/maior-vazamento-dados-pessoais-historia-brasileira-e-quais-licoes-devemos-aprender>. Citado na p. 17.

BRASIL. **Constituição da República Federativa do Brasil de 1988**. [S.l.*:* s.n.], 1988. Diário Oficial da União. Acesso em: 21 Nov. 2025. Disponível em: <http://www.planalto.gov.br/ccivil_03/constituicao/constituicao.htm>. Citado na p. 20.

BRASIL. **Lei n° 13.709, de 14 de agosto de 2018: Lei Geral de Proteção de Dados Pessoais (LGPD)**. [S.l.*:* s.n.], 2018. http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2018/Lei/L13709.htm. Diário Oficial da União, Brasília, DF, 15 ago. 2018. Citado na p. 20.

BRASIL. **Lei nº 12.414, de 9 de junho de 2011. Disciplina a formação e consulta a bancos de dados com informações de adimplemento, de pessoas naturais ou jurídicas, para formação de histórico de crédito (Lei do Cadastro Positivo)**. [S.l.*:* s.n.], 2011. Diário Oficial da União. Acesso em: 21 Nov. 2025. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12414.htm>. Citado na p. 20.

BRASIL. **Lei nº 12.527, de 18 de novembro de 2011. Regula o acesso a informações previsto no inciso XXXIII do art. 5º, no inciso II do § 3º do art. 37 e no § 2º do art. 216 da Constituição Federal (Lei de Acesso à Informação)**. [S.l.*:* s.n.], 2011. Diário Oficial da União. Acesso em: 21 Nov. 2025. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2011/lei/l12527.htm>. Citado na p. 20.

BRASIL. **Lei nº 12.965, de 23 de abril de 2014 (Marco Civil da Internet)**. [S.l.*:* s.n.], 2014. Diário Oficial da União. Acesso em: 21 Nov. 2025. Disponível em: <http://www.planalto.gov.br/ccivil_03/_ato2011-2014/2014/lei/l12965.htm>. Citado na p. 20.

BRASIL. **Lei nº 8.078, de 11 de setembro de 1990. Código de Defesa do Consumidor**. [S.l.*:* s.n.], 1990. Diário Oficial da União. Acesso em: 21 Nov. 2025. Disponível em: <http://www.planalto.gov.br/ccivil_03/leis/l8078.htm>. Citado na p. 20.

BRASIL. AUTORIDADE NACIONAL DE PROTEÇÃO DE DADOS. **ANPD determina suspensão cautelar do tratamento de dados pessoais para treinamento da IA da Meta**. Brasília: [s.n.], jul. 2024. Disponível em: <https://www.gov.br/anpd/pt-br/assuntos/noticias/anpd-determina-suspensao-cautelar-do-tratamento-de-dados-pessoais-para-treinamento-da-ia-da-meta>. Citado na p. 24.

BRASIL. AUTORIDADE NACIONAL DE PROTEÇÃO DE DADOS. **Meta cumpre exigências da ANPD e poderá retomar, com restrições, o uso de dados pessoais para treinamento de inteligência artificial**. Brasília: [s.n.], ago. 2024. Disponível em: <https://www.gov.br/anpd/pt-br/assuntos/noticias/meta-cumpre-exigencias-da-anpd-e-podera-retomar-com-restricoes-o-uso-de-dados-pessoais-para-treinamento-de-inteligencia-artificial>. Citado na p. 24.

BRIGNULL, Harry. **Deceptive Patterns: User Interfaces Designed to Trick People**. [S.l.*:* s.n.], 2015. https://www.darkpatterns.org. Acessado em 2025. Citado na p. 18.

BROWN, Tom B. et al. Language Models are Few-Shot Learners. In: ADVANCES in Neural Information Processing Systems. [S.l.]: Curran Associates, Inc., 2020. p. 1877–1901. Disponível em: <https://proceedings.neurips.cc/paper_files/paper/2020/hash/1457c0d6bfcb4967418bfb8ac142f64a-Abstract.html>. Citado nas pp. 27, 45.

CADWALLADR, Carole; GRAHAM-HARRISON, Emma. Revealed: 50 million Facebook profiles harvested for Cambridge Analytica in major data breach. **The Guardian**, Guardian News & Media London, v. 17, p. 22, 2018. Disponível em: <https://www.theguardian.com/news/2018/mar/17/cambridge-analytica-facebook-influence-us-election>. Citado na p. 18.

CAMERINI, Francisco. **tcc-policy-analysis**. 2026. Disponível em: <https://codigos.ufsc.br/francisco.camerini/tcc-policy-analysis>. Acesso em: 10 jul. 2026. Citado nas pp. 52, 73.

CETIC. **Privacidade e Proteção de Dados Pessoais 2023**. [S.l.], 2024. Apenas 32% das empresas desenvolveram política de privacidade que informa o tratamento de dados. Disponível em: <https://cetic.br/media/docs/publicacoes/2/20240901120340/privacidade-e-protecao-de-dados-2023.pdf>. Citado na p. 16.

CHIANG, Wei-Lin et al. **Chatbot Arena: An Open Platform for Evaluating LLMs by Human Preference**. [S.l.*:* s.n.], 2024. arXiv: 2403.04132. Disponível em: <https://arxiv.org/abs/2403.04132>. Citado na p. 41.

CRANOR, Lorrie et al. The platform for privacy preferences 1.0 (P3P1. 0) specification. **W3C recommendation**, v. 16, 2002. Disponível em: <https://www.w3.org/P3P/>. Citado na p. 33.

CUI, Hao et al. PoliGraph: Automated privacy policy analysis using knowledge graphs. In: 32ND USENIX Security Symposium (USENIX Security 23). [S.l.*:* s.n.], 2023. p. 1037–1054. Disponível em: <https://www.usenix.org/conference/usenixsecurity23/presentation/cui>. Citado na p. 34.

DING, Ning et al. Parameter-efficient fine-tuning of large-scale pre-trained language models. **Nature Machine Intelligence**, Nature Publishing Group, v. 5, p. 220–235, 2023. DOI: 10.1038/s42256-023-00626-4. Disponível em: <https://www.nature.com/articles/s42256-023-00626-4>. Citado nas pp. 29, 49.

DONEDA, Danilo. A proteção dos dados pessoais como um direito fundamental. **Espaço Jurídico: Journal of Law**, Unoesc, v. 12, n. 2, p. 91–108, 2011. Disponível em: <https://dialnet.unirioja.es/servlet/articulo?codigo=4555153>. Citado na p. 17.

FABIAN, Benjamin; ERMAKOVA, Tatiana; LENTZ, Tino. Large-scale readability analysis of privacy policies. In: PROCEEDINGS of the International Conference on Web Intelligence. Leipzig, Germany: Association for Computing Machinery, 2017. (WI '17), p. 18–25. DOI: 10.1145/3106426.3106427. Disponível em: <https://doi.org/10.1145/3106426.3106427>. Citado nas pp. 15, 24.

FAWCETT, Tom. An introduction to ROC analysis. **Pattern Recognition Letters**, v. 27, n. 8, p. 861–874, 2006. ROC Analysis in Pattern Recognition. ISSN 0167-8655. DOI: https://doi.org/10.1016/j.patrec.2005.10.010. Disponível em: <https://www.sciencedirect.com/science/article/pii/S016786550500303X>. Citado na p. 31.

HOWARD, Jeremy; RUDER, Sebastian. Universal Language Model Fine-tuning for Text Classification. In: PROCEEDINGS of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Melbourne, Australia: Association for Computational Linguistics, 2018. p. 328–339. DOI: 10.18653/v1/P18-1031. Disponível em: <https://aclanthology.org/P18-1031/>. Citado na p. 28.

HU, Edward J et al. LoRA: Low-Rank Adaptation of Large Language Models. In: INTERNATIONAL Conference on Learning Representations. [S.l.*:* s.n.], 2022. Disponível em: <https://openreview.net/forum?id=nZeVKeeFYf9>. Citado nas pp. 29, 36, 51.

KOJIMA, Takeshi et al. Large Language Models are Zero-Shot Reasoners. In: ADVANCES in Neural Information Processing Systems. [S.l.*:* s.n.], 2022. p. 22199–22213. Disponível em: <https://proceedings.neurips.cc/paper_files/

paper/2022/hash/8bb0d291acd4acf06ef112099c16f326-Abstract-Conference.html>. Citado na p. 27.

KOKOLAKIS, Spyros. Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. **Computers & security**, Elsevier, v. 64, p. 122–134, 2017. Disponível em: <https://www.sciencedirect.com/science/article/abs/pii/S0167404815001017>. Citado na p. 20.

MINISTÉRIO DA GESTÃO E DA INOVAÇÃO EM SERVIÇOS PÚBLICOS. **Modelo de Política de Proteção de Dados Pessoais**. [S.l.*:* s.n.], 2024. Acesso em: 20 nov. 2025. Disponível em: <https://www.gov.br/governodigital/pt-br/privacidade-e-seguranca/ppsi/modelo_ppdp.pdf>. Citado na p. 25.

MOZILLA. **What are extensions?** Última modificação em 25 de agosto de 2025. 2025. Disponível em: <https://developer.mozilla.org/en-US/docs/Mozilla/Add-ons/WebExtensions/What_are_WebExtensions>. Citado na p. 35.

NOTTINGHAM, Mark. **Well-Known Uniform Resource Identifiers (URIs)**. [S.l.]: Internet Engineering Task Force, maio 2019. RFC 8615. DOI: 10.17487/RFC8615. Disponível em: <https://www.rfc-editor.org/rfc/rfc8615>. Citado na p. 62.

OLIVEIRA, R. M. de. LGPD e seu impacto nas pequenas empresas. **WorkTec, edições Fatec RP**, v. 2022-1, 2022. Trabalho publicado em Fatec RP - versão PDF disponível online. Disponível em: <http://www.fatecrp.edu.br/WorkTec/edicoes/2022-1/trabalhos/GNI/artigo8.pdf>. Citado na p. 24.

OLIVEIRA ROCHA LOURES, Matheus de. **A eficácia da fiscalização e sanção na lei geral de proteção de dados: propostas de melhoria para a atuação da Autoridade Nacional de Proteção de Dados baseadas no modelo francês**. 2024. Trabalho de Conclusão de Curso (Pós-graduação em Direito Digital) – Centro Universitário de Brasília (UniCEUB), Instituto CEUB de Pesquisa e Desenvolvimento. Acesso em 23 Nov. 2025. Disponível em: <https://repositorio.uniceub.br/jspui/handle/prefix/17597>. Citado na p. 24.

PETKAUSKAS, Vilius. **Mother of All Breaches: a Historic Data Leak Reveals 26 Billion Records**. [S.l.*:* s.n.], jan. 2024. Cybernews. Acedido em: 29 Nov. 2025. Disponível em: <https://cybernews.com/security/billions-passwords-credentials-leaked-mother-of-all-breaches/>. Citado na p. 17.

SAHOO, Pranab et al. A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications. **arXiv preprint arXiv:2402.07927**, 2024. Disponível em: <https://arxiv.org/abs/2402.07927>. Citado nas pp. 27, 35, 36.

STEFANI, Wolfgang. **Dark patterns examples - recognising manipulative marketing and design**. Raidboxes Blog. 2023. Disponível em:

<https://blog.raidboxes.io/en/security/dark-patterns/>. Acesso em: 30 nov. 2025. Citado na p. 19.

SUN, Yujie et al. AI hallucination: towards a comprehensive classification of distorted information in artificial intelligence-generated content. **Humanities and Social Sciences Communications**, v. 11, n. 1, p. 1278, 2024. DOI: 10.1057/s41599-024-03811-x. Disponível em: <https://doi.org/10.1057/s41599-024-03811-x>. Citado na p. 26.

TAN, Zhen et al. Large Language Models for Data Annotation and Synthesis: A Survey. **arXiv preprint arXiv:2402.13446**, 2024. Accepted at EMNLP 2024. Disponível em: <https://arxiv.org/abs/2402.13446>. Citado na p. 41.

TECHREPUBLIC. **Despite big-name support, new privacy standard slow to catch on**. [S.l.*:* s.n.], 2002. Acessado em 20 Nov 2025. Disponível em: <https://archive.today/20120707050136/http://articles.techrepublic.com.com/5100-10878-1059571.html>. Citado na p. 33.

TIKU, Nitasha. Facebook told advertisers it can identify teens feeling 'insecure' and 'worthless'. **The Guardian**, v. 1, 2017. Acesso em: 20 nov. 2025. Disponível em: <https://www.theguardian.com/technology/2017/may/01/facebook-advertising-data-insecure-teens>. Citado na p. 18.

TOS;DR - Terms of Service; Didn't Read. Disponível em: <https://tosdr.org/en>. Acesso em: 24 nov. 2025. Citado na p. 33.

TRIBUNAL DE CONTAS DA UNIÃO. **Auditoria sobre a aplicação da LGPD por organizações federais**. [S.l.], 2025. Quase um terço dos órgãos federais sem medidas básicas de adequação. Disponível em: <https://portal.tcu.gov.br/imprensa/noticias/auditoria-aponta-falhas-na-aplicacao-da-lgpd-por-organizacoes-federais>. Citado na p. 16.

UNIÃO EUROPEIA. **Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation)**. [S.l.*:* s.n.], 2016. Official Journal of the European Union, L119, 4 May 2016. Disponível em: <https://eur-lex.europa.eu/eli/reg/2016/679/oj>. Citado na p. 20.

VASWANI, Ashish et al. Attention Is All You Need. In: CURRAN ASSOCIATES, INC. ADVANCES in Neural Information Processing Systems. [S.l.*:* s.n.], 2017. p. 5998–6008. Disponível em: <https://arxiv.org/abs/1706.03762>. Citado na p. 25.

WARREN, Samuel D.; BRANDEIS, Louis D. The Right to Privacy. **Harvard Law Review**, The Harvard Law Review Association, v. 4, n. 5, 1890. Disponível em: <https://groups.csail.mit.edu/mac/classes/6.805/articles/privacy/Privacy_brand_warr2.html>. Citado na p. 17.

WEI, Jason et al. Chain-of-Thought Prompting Elicits Reasoning in Large Language Models. In: ADVANCES in Neural Information Processing Systems. [S.l.: s.n.], 2022. p. 24824–24837. Disponível em: <https://arxiv.org/abs/2201.11903>. Citado na p. 27.

WHITE, Jules et al. A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT. **arXiv preprint arXiv:2302.11382**, 2023. Disponível em: <https://arxiv.org/abs/2302.11382>. Citado na p. 27.

WILSON, Shomir et al. The Creation and Analysis of a Website Privacy Policy Corpus. In: PROCEEDINGS of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). Berlin, Germany: Association for Computational Linguistics, 2016. p. 1330–1340. DOI: 10.18653/v1/P16-1126. Disponível em: <https://aclanthology.org/P16-1126/>. Citado na p. 40.

XIE, Qinge; RAMAKRISHNAN, Karthik; LI, Frank. Evaluating privacy policies under modern privacy laws at scale: An {LLM-Based} automated approach. In: 34TH USENIX Security Symposium (USENIX Security 25). [S.l.: s.n.], 2025. p. 5797–5816. Disponível em: <https://www.usenix.org/conference/usenixsecurity25/presentation/xie>. Citado nas pp. 34, 37, 38, 44, 61.

ZUBOFF, Shoshana. The age of surveillance capitalism. In: SOCIAL theory re-wired. [S.l.]: Routledge, 2023. p. 203–213. Disponível em: <https://www.taylorfrancis.com/chapters/edit/10.4324/9781003320609-27/age-surveillance-capitalism-shoshana-zuboff>. Citado na p. 18.

APÊNDICE A – PROMPT DO CENÁRIO DE PROMPT SIMPLES

Este apêndice reproduz o prompt completo utilizado no cenário de **prompt simples** do benchmark comparativo descrito na Seção 4.5, exatamente como enviado à *API* de cada modelo avaliado.

SYSTEM PROMPT

```
Você é um especialista em conformidade com a LGPD (Lei Geral de Proteção de
  Dados).
Analise o texto da política de privacidade e classifique cada requisito.
Responda APENAS com JSON válido. Sem explicações. Sem texto antes ou depois
  do JSON.
```

INSTRUÇÃO DE USUÁRIO

```
Classifique cada requisito como exatamente um dos valores: "presente",
  "ausente", "parcial"

Requisitos:
- finalidade: a finalidade do tratamento dos dados é declarada
  explicitamente
- base_legal: a base legal é indicada (consentimento, contrato, legítimo
  interesse, etc.)
- retencao: o prazo de retenção dos dados é informado
- direitos: os direitos do titular são listados (acesso, correção, exclusão,
  portabilidade, revogação)
- canal_exercicio: como exercer os direitos é descrito (e-mail, formulário,
  link)
- controlador: nome e contato do controlador estão presentes
- seguranca: medidas de segurança são mencionadas
- compartilhamento: QUALQUER menção a terceiros, "presente" se a política
  compartilha OU declara que não compartilha; "ausente" apenas se nunca
  mencionado

Responda com APENAS este JSON:
{"finalidade":"...","base_legal":"...","retencao":"...","direitos":"...",
  "canal_exercicio":"...","controlador":"...","seguranca":"...",
  "compartilhamento":"..."}
```

Ao final da instrução de usuário, o texto da política avaliada é delimitado pelas tags `<policy>` e `</policy>` e anexado à mensagem.

APÊNDICE B – PROMPT DO CENÁRIO FEW-SHOT

Este apêndice reproduz o prompt completo utilizado no cenário ***few-shot*** do benchmark comparativo descrito na Seção 4.5 e no formato de treinamento do ajuste fino do Qwen 2.5 14B descrito na Seção 4.6.3. O *system prompt* é idêntico ao utilizado no cenário de prompt simples, reproduzido no Apêndice A. A instrução de usuário insere cinco exemplos anotados de políticas reais antes do texto da política a ser avaliada.

INSTRUÇÃO DE USUÁRIO

```
Classifique cada requisito como exatamente um dos valores: "presente",
  "ausente", "parcial"

Requisitos:
- finalidade: a finalidade do tratamento dos dados é declarada
  explicitamente
- base_legal: a base legal é indicada (consentimento, contrato, legítimo
  interesse, etc.)
- retencao: o prazo de retenção dos dados é informado
- direitos: os direitos do titular são listados (acesso, correção, exclusão,
  portabilidade, revogação)
- canal_exercicio: como exercer os direitos é descrito (e-mail, formulário,
  link)
- controlador: nome e contato do controlador estão presentes
- seguranca: medidas de segurança são mencionadas
- compartilhamento: QUALQUER menção a terceiros, "presente" se a política
  compartilha OU declara que não compartilha; "ausente" apenas se nunca
  mencionado

Exemplos:

<policy>Utilizamos seus dados para envio de comunicações promocionais. Entre
  em contato pelo e-mail privacidade@empresa.com para cancelar.</policy>
{"finalidade":"presente","base_legal":"ausente","retencao":"ausente",
  "direitos":"parcial","canal_exercicio":"presente","controlador":"ausente",
  "seguranca":"ausente","compartilhamento":"ausente"}

<policy>Seus dados poderão ser compartilhados com processadores de pagamento
  para operação do serviço. Não vendemos seus dados a terceiros.</policy>
{"finalidade":"presente","base_legal":"ausente","retencao":"ausente",
  "direitos":"ausente","canal_exercicio":"ausente","controlador":"ausente",
```

```
  "seguranca":"ausente","compartilhamento":"presente"}

<policy>Não compartilhamos suas informações pessoais com empresas terceiras,
  exceto quando necessário por exigência legal.</policy>
{"finalidade":"ausente","base_legal":"ausente","retencao":"ausente",
  "direitos":"ausente","canal_exercicio":"ausente","controlador":"ausente",
  "seguranca":"ausente","compartilhamento":"presente"}

<policy>Coletamos nome e CPF com base no seu consentimento (Art. 7, I, LGPD)
  para prestação dos serviços contratados. Dados mantidos por 5 anos. Você
  tem direito a acesso, correção, exclusão, portabilidade e revogação via
  dpo@empresa.com. Empresa S.A., CNPJ 00.000.000/0001-00. Utilizamos
  criptografia AES-256. Não compartilhamos com terceiros.</policy>
{"finalidade":"presente","base_legal":"presente","retencao":"presente",
  "direitos":"presente","canal_exercicio":"presente",
  "controlador":"presente","seguranca":"presente",
  "compartilhamento":"presente"}

<policy>Esta política descreve como tratamos suas informações.</policy>
{"finalidade":"ausente","base_legal":"ausente","retencao":"ausente",
  "direitos":"ausente","canal_exercicio":"ausente","controlador":"ausente",
  "seguranca":"ausente","compartilhamento":"ausente"}

Responda com APENAS um objeto JSON em uma única linha:
{"finalidade":"...","base_legal":"...","retencao":"...","direitos":"...",
  "canal_exercicio":"...","controlador":"...","seguranca":"...",
  "compartilhamento":"..."}
```

Como no cenário simples, o texto da política avaliada é delimitado pelas tags `<policy>` e `</policy>` e anexado ao final da mensagem do usuário.

APÊNDICE C – INSTRUÇÕES DE ACESSO AO CÓDIGO-FONTE

Este apêndice descreve como acessar e explorar o repositório de código-fonte do sistema desenvolvido neste trabalho (Camerini, 2026).

ACESSO AO REPOSITÓRIO

O código-fonte completo está hospedado no GitLab institucional da UFSC e pode ser acessado publicamente pela URL:

`https://codigos.ufsc.br/francisco.camerini/tcc-policy-analysis`

Para obter uma cópia local do repositório, é necessário ter o Git instalado e executar o seguinte comando em um terminal:

```
git clone https://codigos.ufsc.br/francisco.camerini/tcc-policy-analysis.git
```

ESTRUTURA DO REPOSITÓRIO

O repositório está organizado em três diretórios principais, correspondentes aos componentes descritos neste trabalho:

`api/` Código-fonte da API de backend (Seção 5.2), desenvolvida em Python com o *framework* FastAPI. O arquivo `main.py` define os *endpoints* HTTP, `llm.py` implementa o módulo de análise e a comunicação com o modelo de linguagem, e `cache.py` gerencia a persistência dos resultados.

`extension/` Código-fonte da extensão de navegador (Seção 5.3), no padrão Manifest V3, incluindo o *content script*, o *background service worker* e a interface do *popup*.

`model/` Scripts e artefatos relacionados à construção do dataset (Seção 4.2) e aos experimentos de benchmark (Seção 4.5) e ajuste fino (Seção 4.6), incluindo as políticas anotadas, os *notebooks* utilizados e os resultados brutos de cada experimento.

EXECUÇÃO LOCAL

Para executar a API localmente, é necessário instalar as dependências Python listadas em `api/requirements.txt` e iniciar o servidor com o comando `uvicorn main:app --port 8080`, conforme descrito na Seção 5.2. A extensão de navegador pode ser carregada em modo de desenvolvedor em navegadores baseados em Chromium, apontando para o diretório `extension/`. Por padrão, a extensão tenta se conectar à API local em

`http://localhost:8080` e recorre ao servidor de produção caso essa instância não esteja disponível, conforme descrito na Seção 5.3.

APÊNDICE D – ARTIGO PRODUZIDO A PARTIR DESTE TRABALHO

Este apêndice reproduz, na íntegra, o artigo científico produzido a partir dos resultados deste trabalho, redigido conforme o modelo de formatação da Sociedade Brasileira de Computação (SBC).

Análise Automática de políticas de privacidade com base na Lei Geral de Proteção de Dados

Francisco Camerini[1]

[1]Instituto de Informática e Estatística (INE), Centro Tecnológico, Universidade Federal de Santa Catarina (UFSC), Florianópolis, SC, Brasil

francisco.camerini@ufsc.br

Abstract. *Privacy policies are the primary mechanism through which organizations inform data subjects about personal data processing under Brazil's General Data Protection Law (LGPD), yet they are typically long, technical and rarely read. This paper presents a tool that automates the compliance assessment of privacy policies against eight LGPD transparency requirements using a fine-tuned large language model. We built an original dataset of 200 Brazilian privacy policies, silver-labeled by Claude Opus and partially validated by manual review, benchmarked eight smaller language models under two prompting scenarios, and fine-tuned the best performing open-weight model, Qwen 2.5 14B, with LoRA. The fine-tuned model reached a macro F1 of 0.8409 on the validation set, above the acceptance criterion and every proprietary model in the benchmark. The system is delivered as a browser extension backed by an API, giving end users an on-the-fly compliance assessment of any privacy policy they visit.*

Resumo. Este trabalho apresenta uma ferramenta que automatiza a avaliação de conformidade de políticas de privacidade com oito requisitos de transparência da Lei Geral de Proteção de Dados (LGPD), utilizando um modelo de linguagem de grande escala especializado por ajuste fino. Foi construído um dataset original com 200 políticas de privacidade brasileiras, rotuladas automaticamente pelo Claude Opus e parcialmente validadas por revisão manual, utilizado em um benchmark comparativo entre oito modelos de linguagem e no ajuste fino, com a técnica LoRA, do modelo Qwen 2.5 14B. O modelo resultante atingiu F1 macro de 0,8409 no conjunto de validação, superando o critério de aceitação e todos os modelos proprietários avaliados. O sistema é disponibilizado como uma extensão de navegador apoiada por uma API, oferecendo ao usuário uma avaliação de conformidade em tempo real de qualquer política de privacidade visitada.

1. Introdução

A crescente digitalização de serviços e o uso intensivo de dados pessoais em plataformas online tornaram a privacidade um tema central de debate na sociedade contemporânea. As políticas de privacidade são o instrumento pelo qual organizações informam aos usuários como seus dados são coletados, armazenados e utilizados,

obrigação reforçada pela Lei Geral de Proteção de Dados Pessoais (LGPD, Lei nº 13.709/2018). Apesar dessa obrigatoriedade, esses documentos costumam ser redigidos de forma ambígua, incompleta ou excessivamente técnica, o que dificulta a compreensão pelo titular dos dados (Fabian et al., 2017). Segundo o TCU, quase um terço das organizações públicas federais ainda não adotou medidas básicas de adequação à LGPD (TCU, 2025), e apenas 32% das empresas brasileiras desenvolveram uma política de privacidade que de fato informa o tratamento de dados (Cetic, 2024).

O avanço dos modelos de linguagem de grande escala (LLMs) tornou viável processar textos jurídicos complexos e interpretar linguagem natural em alto nível, abrindo caminho para a análise automatizada de conformidade documental. Este trabalho propõe uma ferramenta computacional, na forma de uma extensão para navegadores web apoiada por uma API, capaz de analisar automaticamente políticas de privacidade de sites e indicar possíveis inconformidades com a LGPD. O restante deste artigo apresenta a fundamentação teórica (Seção 2), os trabalhos relacionados (Seção 3), a metodologia (Seção 4), o sistema desenvolvido (Seção 5), os resultados obtidos (Seção 6) e a conclusão (Seção 7).

2. Fundamentação Teórica

2.1. Privacidade, Vigilância e a LGPD

A internet comercial moderna opera sob a lógica do chamado capitalismo de vigilância, em que a experiência humana é traduzida em dados comportamentais explorados para fins preditivos e comerciais (Zuboff, 2023). Práticas de design que induzem o usuário a conceder permissões amplas de coleta, os chamados *dark patterns*, agravam a assimetria de informação entre titulares e controladores de dados. A LGPD, inspirada na GDPR europeia (União Europeia, 2016), estabelece dez princípios para o tratamento de dados pessoais no Brasil, dos quais finalidade, necessidade e transparência são os mais relevantes para a existência e a obrigatoriedade das políticas de privacidade.

2.2. Modelos de Linguagem de Grande Escala

LLMs são redes neurais baseadas na arquitetura *Transformer* (Vaswani et al., 2017), capazes de compreender e gerar linguagem natural de forma contextualizada. A engenharia de prompt, isto é, a construção sistemática de instruções que orientam o comportamento do modelo (Sahoo et al., 2024), inclui técnicas como o *few-shot prompting*, em que exemplos concretos da tarefa são fornecidos antes da solicitação real (Brown et al., 2020). Quando a engenharia de prompt não é suficiente para especializar o modelo em um domínio, recorre-se ao ajuste fino (*fine-tuning*). O *Low-Rank Adaptation* (LoRA) (Hu et al., 2022) é o método de ajuste fino eficiente em parâmetros mais adotado, injetando matrizes de baixo posto em paralelo às camadas da arquitetura *Transformer* e mantendo os pesos originais congelados, o que reduz drasticamente o custo computacional do treinamento.

3. Trabalhos Relacionados

Diversas iniciativas buscaram automatizar a análise de políticas de privacidade. O *Platform for Privacy Preferences* (P3P) propunha um protocolo legível por máquina para descrever formalmente as práticas de um site (Cranor et al., 2002), mas fracassou por rigidez de formato e baixa adesão das empresas. O ToS;DR aposta na curadoria humana colaborativa, mas esbarra na falta de escalabilidade. O PoliGraph analisa o texto de uma política por meio de grafos de conhecimento (Cui et al., 2023), porém depende de pipelines sintáticos de PLN tradicional com limitações conhecidas para lidar com nuances semânticas. O corpus OPP-115, com 115 políticas norte-americanas anotadas (Wilson et al., 2016), é referência para pesquisas em inglês, mas não cobre a legislação brasileira.

O estado da arte é representado por Xie, Ramakrishnan e Li (2025), que propuseram uma estrutura baseada em LLM para avaliar políticas de privacidade em escala frente a dez leis modernas, incluindo a GDPR e leis estaduais dos EUA, alcançando F1 médio de 0,94. Os próprios autores sugerem, como trabalho futuro, a expansão da análise para outras legislações e idiomas. Na revisão bibliográfica conduzida para este trabalho, não foi identificado nenhum estudo que aplique LLMs especificamente à avaliação de conformidade com a LGPD, tampouco um corpus público anotado de políticas de privacidade brasileiras, lacuna que este trabalho se propõe a preencher.

4. Metodologia

Este trabalho avalia oito requisitos de transparência derivados da LGPD: (i) finalidade, a explicitação dos propósitos da coleta e do tratamento dos dados; (ii) base legal, a indicação da hipótese legal que autoriza o tratamento; (iii) retenção, o prazo ou critério de eliminação dos dados; (iv) direitos, a informação sobre os direitos que o titular pode exercer; (v) canal de exercício, os meios pelos quais o titular pode exercer esses direitos; (vi) controlador, a identificação de quem é responsável pelo tratamento; (vii) segurança, a descrição das medidas técnicas e organizacionais adotadas; e (viii) compartilhamento, a indicação de eventual transferência dos dados a terceiros. Para cada requisito, o modelo classifica a política como presente, parcial ou ausente. Uma política é considerada conforme quando nenhum requisito é avaliado como ausente e, no máximo, um como parcial, regra que reflete o caráter cumulativo, e não proporcional, dos princípios da LGPD.

4.1. Construção do Dataset

Foi construído um corpus original com 200 políticas de privacidade de organizações que atuam no Brasil, distribuídas em oito setores da economia e da administração pública. Diante da inviabilidade de anotação manual especializada em escala, adotou-se a estratégia de *Silver Labeling* (Tan et al., 2024): a geração automática dos rótulos pelo Claude Opus, modelo então entre os melhores colocados do *Chatbot Arena* (Chiang et al., 2024). Uma amostra estratificada de 16 políticas (8% do dataset) foi submetida a uma revisão manual de plausibilidade, confirmando maior concordância nos requisitos

objetivos e maior variabilidade nos requisitos que admitem formulações ambíguas, como base legal e retenção. O resultado da rotulagem revelou desbalanceamento pronunciado: apenas 20,5% das políticas foram classificadas como conformes, com o setor governamental apresentando o pior resultado (nenhuma das 31 políticas coletadas). O dataset final foi dividido em treino (75%, 150 políticas) e validação (25%, 50 políticas), com amostragem estratificada por setor.

4.2. Benchmark Comparativo

Oito modelos de linguagem foram avaliados sob dois cenários de prompting, um prompt simples e um *few-shot* com cinco exemplos anotados (Brown et al., 2020), usando o F1-Score macro binário como métrica principal. O critério de aceitação foi fixado em 0,80, inferior ao valor de referência de 0,94 obtido por Xie et al. (2025) em razão da escala reduzida do corpus e da ausência de validação jurídica especializada em toda a sua extensão. Foram avaliados GPT-4.1 mini e nano, Gemini 2.5 e 3.1 Flash Lite, Claude Haiku 4.5, Sabía 4 Small, Llama 3.3 70B e Qwen 2.5 14B, priorizando modelos de menor porte e custo operacional, adequados a cenários de produção em escala.

4.3. Ajuste Fino com LoRA

O Qwen 2.5 14B foi selecionado para a fase de especialização por apresentar o melhor desempenho de base no cenário *few-shot*, por ser um modelo de pesos abertos e por viabilizar o treinamento com quantização de 4 bits em GPU disponível via Google Colab Pro. O ajuste fino foi conduzido com a técnica LoRA (Hu et al., 2022), com exploração sistemática de hiperparâmetros: um posto elevado (r=32) produziu sobreajuste e queda de desempenho abaixo do baseline, enquanto a configuração final, com posto r=8 e quatro épocas de treinamento, maximizou a generalização sem comprometer o desempenho.

5. Sistema Desenvolvido

O sistema é composto por uma API de backend, desenvolvida em FastAPI, e uma extensão de navegador no padrão Manifest V3. A extensão detecta automaticamente o link de política de privacidade mais relevante na página visitada e o encaminha à API mediante ação explícita do usuário. A API executa um pipeline de validação da URL, extração do texto da página, verificação de cache por hash SHA-256 e, quando necessário, envio do conteúdo a um modelo de linguagem, retornando um veredito agregado de conformidade e a avaliação granular dos oito requisitos, exibidos ao usuário no *popup* da extensão de forma auditável. Esse fluxo pode ser evidenciado na Figura 1. O código-fonte completo do sistema está disponível publicamente (Camerini, 2026).

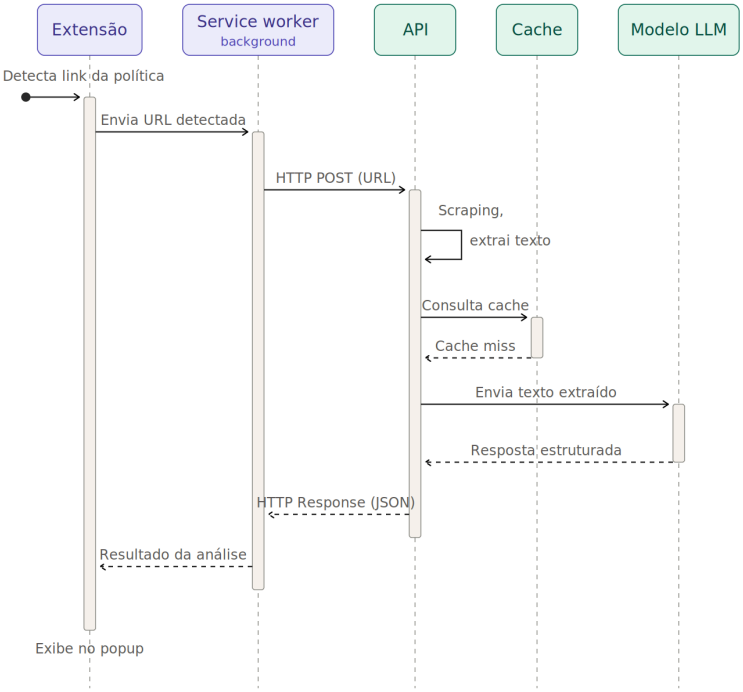


Figura 1 - Arquitetura do Sistema

6. Resultados

No benchmark comparativo, o Gemini 3.1 Flash Lite obteve o melhor resultado no cenário de prompt simples (F1 de 0,7718), enquanto o Qwen 2.5 14B liderou o cenário *few-shot* (F1 de 0,7937). Após o ajuste fino, o Qwen 2.5 14B atingiu F1 macro binário de 0,8409 no conjunto de validação, superando o critério de aceitação de 0,80 e todos os modelos proprietários avaliados, inclusive aqueles de custo operacional substancialmente superior, como o Claude Haiku 4.5. A trajetória dos experimentos intermediários revelou que um posto LoRA elevado (r=32) degrada o desempenho para 0,6366, evidenciando que a seleção de hiperparâmetros tem impacto determinante em corpus de escala moderada.

7. Conclusão

Este trabalho demonstrou ser viável construir, com recursos computacionais modestos e sem anotação jurídica especializada em escala, um sistema capaz de avaliar automaticamente a aderência de políticas de privacidade à LGPD com desempenho superior ao de modelos proprietários de uso geral. O trabalho produziu três artefatos de

uso independente: o dataset anotado de 200 políticas brasileiras, até onde se sabe o primeiro corpus público voltado especificamente à conformidade com a LGPD; o sistema composto pela API e pela extensão de navegador, imediatamente utilizável pelo público não técnico; e o benchmark comparativo entre modelos de linguagem multilíngues e nacionais na tarefa em português.

Como limitações, destacam-se a dependência de um único modelo anotador, a escala reduzida e o desbalanceamento do dataset, e o fato de o modelo em produção ainda não ser o Qwen 2.5 14B ajustado. Trabalhos futuros incluem a padronização da localização de políticas de privacidade via convenção *.well-known*, a expansão do dataset, a validação jurídica dos rótulos, a revisão do esquema de categorização adotado e a disponibilização da API como ferramenta de apoio à elaboração de políticas de privacidade pelos próprios controladores de dados.

Referências

Brasil. (2018) “Lei nº 13.709, de 14 de agosto de 2018: Lei Geral de Proteção de Dados Pessoais (LGPD)”, Diário Oficial da União, Brasília, DF.

Brown, T. B. et al. (2020) “Language Models are Few-Shot Learners”, Advances in Neural Information Processing Systems, v. 33, p. 1877-1901.

Camerini, F. (2026) “tcc-policy-analysis”, disponível em: https://codigos.ufsc.br/francisco.camerini/tcc-policy-analysis.

Cetic. (2024) “Privacidade e Proteção de Dados Pessoais 2023”, Comitê Gestor da Internet no Brasil.

Chiang, W. et al. (2024) “Chatbot Arena: An Open Platform for Evaluating LLMs by Human Preference”, arXiv preprint arXiv:2403.04132.

Cranor, L., Langheinrich, M., Marchiori, M., Presler-Marshall, M. and Reagle, J. (2002) “The Platform for Privacy Preferences 1.0 (P3P1.0) Specification”, W3C Recommendation.

Cui, H., Trimananda, R., Markopoulou, A. and Jordan, S. (2023) “PoliGraph: Automated Privacy Policy Analysis using Knowledge Graphs”, 32nd USENIX Security Symposium, p. 1037-1054.

Fabian, B., Ermakova, T. and Lentz, T. (2017) “Large-scale Readability Analysis of Privacy Policies”, Proceedings of the International Conference on Web Intelligence, p. 18-25.

Hu, E. J. et al. (2022) “LoRA: Low-Rank Adaptation of Large Language Models”, International Conference on Learning Representations.

Sahoo, P., Singh, A. K., Saha, S., Jain, V., Mondal, S. and Chadha, A. (2024) “A Systematic Survey of Prompt Engineering in Large Language Models: Techniques and Applications”, arXiv preprint arXiv:2402.07927.

Tan, Z. et al. (2024) “Large Language Models for Data Annotation and Synthesis: A Survey”, arXiv preprint arXiv:2402.13446.

TCU. (2025) “Auditoria sobre a Aplicação da LGPD por Organizações Federais”, Tribunal de Contas da União.

União Europeia. (2016) “Regulation (EU) 2016/679 (General Data Protection Regulation)”, Official Journal of the European Union, L119.

Vaswani, A. et al. (2017) “Attention Is All You Need”, Advances in Neural Information Processing Systems, v. 30, p. 5998-6008.

Wilson, S. et al. (2016) “The Creation and Analysis of a Website Privacy Policy Corpus”, Proceedings of the 54th Annual Meeting of the ACL, p. 1330-1340.

Xie, Q., Ramakrishnan, K. and Li, F. (2025) “Evaluating Privacy Policies under Modern Privacy Laws at Scale: An LLM-Based Automated Approach”, 34th USENIX Security Symposium, p. 5797-5816.

Zuboff, S. (2023) “The Age of Surveillance Capitalism”, Social Theory Re-Wired, Routledge, p. 203-213.