Compliance Verification in Administrative Sanctioning Procedure Flows: a Semantic Frame-based Approach

Wuldson F. Fernandes da Silva[1], Paulo Ribeiro Lins Júnior[1,2], Euler P. Sobral[2], Wesley W. A. Araujo[2], Edmar Candeia Gurjão[3]

[1]Instituto Federal da Paraíba (IFPB)
Programa de Pós Graduação em Tecnologia da Informação (PPGTI)
João Pessoa – PB

[2]Instituto Federal da Paraíba (IFPB)
Grupo de Pesquisa em Comunicações e Processamento de Informação (GComPI)
Campina Grande – PB

[3]Unidade Acadêmica de Engenharia Elétrica (UAEE)
Universidade Federal de Campina Grande (UFCG)
Campina Grande – PB.

wuldson.franco@academico.ifpb.edu.br, paulo.lins@ifpb.edu.br, ecg@dee.ufcg.edu.br

Resumo. *Este artigo propõe o modelo **CAFÉ** (Compliance Analysis in Administrative Flows using semantic framEs), um pipeline de três camadas para verificação automática de conformidade em fluxos de processos administrativos sancionatórios, avaliado sobre 104 Procedimentos para Apuração de Descumprimento de Obrigações (PADO) da Anatel, cobrindo integridade documental, regularidade do fluxo processual e conformidade semântica por semantic frames. Na camada semântica, comparam-se duas abordagens: a Abordagem A (expressões regulares e sentence embeddings), com 94,2% de conformidade detectada, e a Abordagem B (LSA), com 74,1% e comportamento distinto em documentos históricos. Os resultados demonstram que a verificação em fluxos multidocumento em português jurídico é viável com métodos estatísticos, e que a Abordagem B, baseada em LSA, dispensa embeddings neurais.*

Abstract. *This article proposes the **CAFÉ** (Compliance Analysis in Administrative Flows using semantic framEs) model, a pipeline for the automatic verification of compliance in administrative sanctioning process flows, evaluated on 104 Anatel Procedures for the Investigation of Non-compliance with Obligations (PADO), covering document integrity, procedural regularity and semantic compliance via semantic frames. In the semantic layer, two approaches are compared: Approach A (regular expressions and sentence embeddings), with 94.2% of compliance detected, and Approach B (LSA), with 74.1% and distinct behavior in historical documents. The results demonstrate that verification across multi-document flows in legal Portuguese is feasible using statistical methods, and that Approach B, based on LSA, does not require neural embeddings.*

1. Introdução

No âmbito do Poder Executivo, a regulamentação setorial de IA avançou de forma significativa em 2026. A Portaria RFB nº 647/2026

[Brasil. Secretaria Especial da Receita Federal do Brasil 2026] estabeleceu que sistemas de IA não podem substituir nem vincular decisões de agentes públicos, exigindo supervisão humana e proibindo o compartilhamento de dados sigilosos com ferramentas de IA generativa. A Portaria MGI nº 3.485/2026 [Brasil. Ministério da Gestão e da Inovação em Serviços Públicos 2026] instituiu a Política de Governança de IA no âmbito do Ministério da Gestão, definindo princípios de transparência, conformidade legal e responsabilidade. Essas iniciativas evidenciam que a questão central não é mais apenas a adoção de IA no setor público, mas como garantir que esses sistemas operem de modo verificável, rastreável e juridicamente responsável em contextos nos quais a conformidade normativa é constitucionalmente exigida.

Em processos administrativos sancionatórios, essa exigência de verificabilidade assume contornos ainda mais críticos: cada etapa do fluxo processual é determinada por marcos normativos precisos e qualquer desvio compromete a validade jurídica dos atos produzidos e a responsabilização dos agentes envolvidos.

No entanto, a verificação manual da conformidade desse fluxo é onerosa, suscetível a inconsistências e não escalável diante do volume crescente de processos em tramitação simultânea [Cejas et al. 2023], motivação direta para o desenvolvimento de abordagens automatizadas de verificação de conformidade.

A verificação automatizada de conformidade em documentos textuais é um campo estabelecido [Cejas et al. 2023, Hassani 2024], mas ainda apresenta três limitações: trata, em geral, documentos isolados, sem verificar fluxos multidocumento; concentra-se em inglês e em contextos privados, de *common law* ou de regulação europeia, sem transposição para o português jurídico-regulatório brasileiro [Li and Goel 2025]; e privilegia similaridade lexical ou modelagem de tópicos, sem verificar os papéis argumentativos dos atos normativos. Uma notificação pode mencionar um prazo e, ainda assim, não atribuí-lo explicitamente ao autuado. Essa lacuna motiva o presente trabalho.

A abordagem proposta representa conformidade normativa como frames semânticos (*semantic frames*) [Fillmore 1982, Baker et al. 1998]: estruturas que descrevem um evento obrigatório e seus participantes por papéis semânticos definidos (sujeito obrigado, ação exigida, objeto da ação, restrição temporal). Frames semânticos não são concorrentes de modelos de tópicos ou similaridade vetorial: verificam se os elementos normativos estão nos papéis argumentativos corretos, distinção detalhada na Seção 2. [Cejas et al. 2023] demonstraram que essa representação é significativamente mais precisa do que similaridade de *embeddings* para verificação de requisitos normativos, e [Hassani 2024] confirmaram ganhos de até quarenta pontos percentuais ao operar em nível de parágrafo em vez de frase.

Este trabalho se alinha na interseção de três dimensões da literatura, frequentemente tratadas de forma isolada:

- modelagem de tópicos e representação semântica [Blei et al. 2003, Deerwester et al. 1990];
- verificação de conformidade normativa por PLN [Cejas et al. 2023, Hassani 2024];
- mineração de processos e conformidade de fluxo [van der Aalst 2016, Berti and et al. 2019].

Na revisão realizada para este estudo, não foram identificados trabalhos que articulem simultaneamente essas três dimensões em português, sobre processos administrativos sancionatórios e no contexto regulatório brasileiro.

Este artigo propõe, implementa e avalia o **CAFÉ** (*Compliance Analysis in Administrative Flows using semantic framEs*), um *pipeline* de verificação automática de conformidade em fluxos de processos administrativos sancionatórios, estruturado em três camadas sucessivas: a Camada 1 verifica a integridade dos documentos e metadados presentes em cada fase; a Camada 2 verifica a regularidade do fluxo processual, confirmando que os atos ocorreram na sequência normativa prescrita; e a Camada 3 verifica a conformidade semântica do conteúdo, comparando *semantic frames* extraídos automaticamente de cada documento com os requisitos normativos formalizados por etapa processual. A avaliação empírica é conduzida sobre corpus real de 104 PADOs da Anatel, garantindo aderência ao cenário concreto e verificabilidade dos resultados. A arquitetura em três camadas combina dimensões complementares da conformidade: integridade documental, regularidade do fluxo e conformidade semântica. Essa integração permite um diagnóstico mais abrangente do processo do que qualquer camada isolada, hipótese examinada empiricamente neste estudo.

O restante do artigo está organizado da seguinte forma. A Seção 2 apresenta a representação semântica de documentos, posicionando frames semânticos em relação a LDA e Transformers. A Seção 3 descreve o pipeline proposto e o cenário de avaliação. A Seção 4 reporta os resultados sobre os processos do corpus. A Seção 5 discute limitações e trabalhos futuros.

2. Representação Semântica de Documentos

Modelos de tópicos como o LDA [Blei et al. 2003] e a Análise Semântica Latente (LSA) [Deerwester et al. 1990] estabeleceram as bases da representação vetorial de documentos jurídicos: o LDA modela cada documento como mistura de tópicos latentes, permitindo classificação tipológica não supervisionada; a LSA captura sinonímia de domínio por decomposição matricial, sem dependência de recursos lexicais externos. Ambos operam sobre co-ocorrência estatística e, por isso, são eficazes em corpora com vocabulário padronizado, exatamente o perfil dos documentos processuais da Anatel. A estrutura BERT-LDA [Paul et al. 2025, Li and Chen 2025, Zhu and et al. 2025] avançou ao combinar *embeddings* contextuais do modelo *Bidirectional Encoder Representations from Transformers* (BERT) com modelagem de tópicos, produzindo representações mais coerentes em domínios especializados.

Contudo, [Costa et al. 2023] documentaram que, em corpora de textos jurídicos brasileiros com vocabulário altamente padronizado, o TF-IDF (*Term Frequency–Inverse Document Frequency*) supera técnicas baseadas em BERT nas métricas de classificação, hipótese explicada pela baixa variação semântica entre classes — perfil característico de documentos processuais em contextos regulatórios, nos quais todos os atos compartilham a linguagem jurídico-normativa do órgão emissor. Essas evidências justificam o emprego de LDA e LSA, em vez de *embeddings* neurais, como camadas de representação no pipeline proposto.

Semantic frames [Fillmore 1982, Baker et al. 1998] complementam LDA e LSA ao verificar se os elementos normativos ocupam os papéis argumentativos corretos: sujeito

obrigado, ação exigida, objeto e restrição temporal, e não apenas se as palavras estão presentes. [Cejas et al. 2023] demonstraram que essa representação é mais precisa do que a similaridade de *embeddings* para requisitos normativos, e [Hassani 2024] obtiveram ganhos ao operar em nível de parágrafo, unidade adequada a requisitos distribuídos por múltiplas sentenças.

Table 1. Comparação dos trabalhos relacionados com o CAFÉ.

Trabalho	Domínio e objeto	Contribuição principal
Cejas et al. (2023)	Acordos de processamento de dados e GDPR	Verificação automatizada de conformidade por PLN
Hassani (2024)	Conformidade jurídica e análise regulatória	Investigação do uso de modelos de linguagem em tarefas jurídicas
Li e Goel (2025)	Sistemas de inteligência artificial	Revisão sistemática de medidas e estruturas de auditabilidade
van der Aalst (2016)	Mineração e conformidade de processos	Modelagem e análise de registros de eventos
CAFÉ	PADOs da Anatel em português brasileiro	Integração de classificação documental, conformidade de fluxo e frames semânticos

Os trabalhos relacionados abordam dimensões isoladas do problema. O CAFÉ diferencia-se por integrar classificação documental, verificação de sequência processual e análise semântica em fluxo multidocumento aplicado a processos administrativos sancionatórios em português brasileiro, embora sua avaliação ainda esteja restrita ao corpus de PADOs da Anatel.

3. O Modelo CAFÉ

O modelo proposto, denominado **CAFÉ** (*Compliance Analysis in Administrative Flows using semantic framEs*), é um *pipeline* de verificação automática de conformidade documental estruturado em três camadas funcionais e sucessivas, ilustradas na Figura 1. Cada camada recebe como entrada o *output* da anterior, de modo que as verificações se acumulam em um indicador unificado de conformidade por processo. O modelo é aplicado ao fluxo de instauração do PADO, mas sua arquitetura é extensível a outros processos administrativos sancionatórios com requisitos normativos formalizáveis.

3.1. Camada 1 — Verificação de Integridade Documental

A Camada 1 verifica a presença, o tipo e a consistência dos documentos do processo. Na configuração principal da Abordagem A, a identificação do tipo documental é realizada por regras e expressões regulares. Na configuração alternativa da Abordagem B, o tipo documental é inferido por um modelo de Alocação Latente de Dirichlet (LDA), implementado por meio da biblioteca *scikit-learn* e treinado de forma não supervisionada sobre o corpus, dispensando anotação por tipo documental, recurso escasso no domínio regulatório [Blei et al. 2003]. Um documento com alta carga do tópico associado à instauração e à competência regulatória é classificado como Despacho de Instauração; essa classificação ativa os requisitos normativos correspondentes na Camada 3. A verificação de consistência compara o conteúdo textual com os metadados do Sistema Eletrônico de Informações (SEI), detectando divergências de datas, partes e referências processuais

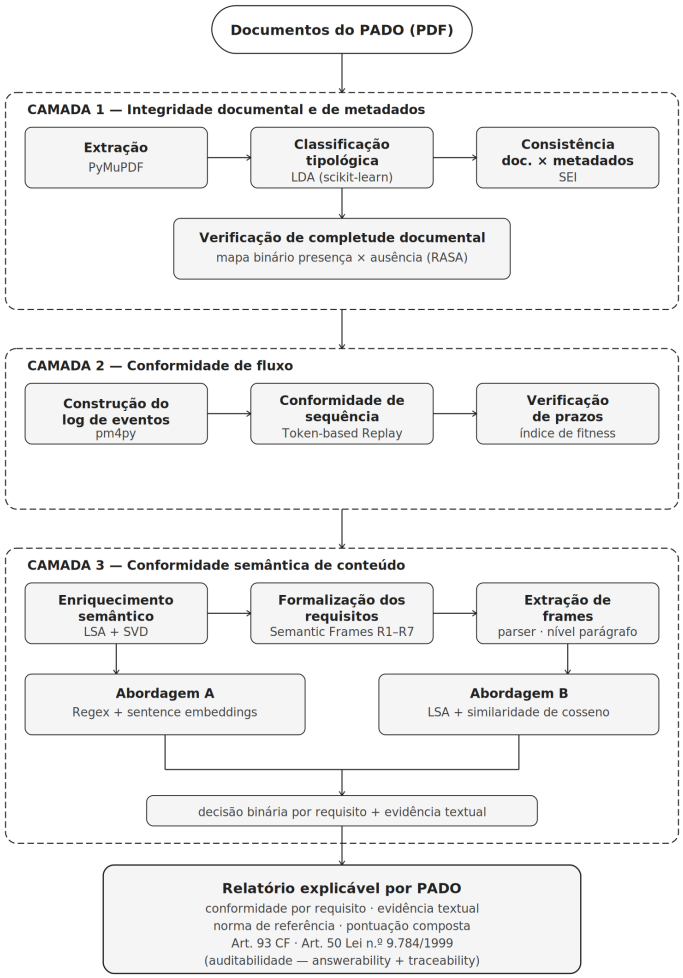


Figure 1. Fluxograma do *pipeline* de verificação automática de conformidade documental.

obrigatórias, com rastreabilidade ao tipo de divergência no *log* de auditoria. A verificação de completude compara os tipos documentais identificados com a lista de documentos obrigatórios para a etapa processual conforme o Regulamento de Aplicação de Sanções Administrativas (RASA) [Agência Nacional de Telecomunicações (Anatel) 2012], registrando para cada documento esperado se foi encontrado ou não, com rastreabilidade ao dispositivo normativo violado.

3.2. Camada 2 — Verificação de Conformidade de Fluxo

A Camada 2 verifica sequência e prazos processuais sobre os metadados validados pela Camada 1. Os documentos tipificados são convertidos em *log* de eventos compatível com a biblioteca pm4py [Berti and et al. 2019], onde cada evento registra o tipo do ato processual, a data de produção e o responsável. A conformidade de sequência é verificada pelo algoritmo *Token-based Replay* [van der Aalst 2016], escolhido por permitir a comparação do comportamento observado com um modelo normativo representado como rede de Petri derivada do RASA e da Lei nº 9.784/1999. O *Token-based Replay* simula a execução do processo sobre o modelo normativo e produz um índice de *fitness* entre 0 e 1,

que quantifica o grau de aderência à sequência prescrita e permite identificar desvios específicos, atos fora de ordem, etapas puladas ou atos duplicados. A verificação de prazos calcula, para cada par de eventos consecutivos, o intervalo de tempo transcorrido e o compara ao prazo normativo correspondente, registrando cada violação com rastreabilidade ao dispositivo normativo.

3.3. Camada 3 — Verificação de Conformidade Semântica

A Camada 3 verifica se o conteúdo de cada documento contém os elementos normativos prescritos, considerando não apenas a presença das palavras corretas, mas a estrutura argumentativa em que os elementos semânticos ocorrem. Os sete requisitos normativos da fase de instauração foram formalizados como *semantic frames* [Fillmore 1982, Baker et al. 1998]: R1 (documento motivador presente), R2 (análise prévia realizada), R3 (despacho de instauração formal), R4 (notificação ao autuado), R5 (base legal citada), R6 (prazo para defesa estabelecido) e R7 (identificação clara do processo). Cada frame define os papéis semânticos obrigatórios para o requisito: agente, ação, objeto e restrição temporal, de modo que a verificação exige não apenas que as palavras corretas estejam presentes, mas que os elementos semânticos estejam nos papéis argumentativos adequados. A extração de *frames* utiliza um *parser* de dependências sintáticas operando em nível de parágrafo [Hassani 2024], unidade mais adequada para requisitos que se distribuem por múltiplas sentenças. O *matching* é implementado em duas abordagens comparadas em paralelo: **Abordagem A** (linha de base) por expressões regulares combinadas com similaridade via *sentence embeddings*; e **Abordagem B** (proposta) por similaridade de cosseno no espaço semântico LSA [Deerwester et al. 1990], capturando sinonímia de domínio, termos como *notificar* e *cientificar*, ou *prazo* e *período*, sem dependência de dicionários externos em português jurídico. Para cada requisito, ambas produzem uma decisão de conformidade com evidência textual rastreável ao trecho do documento.

4. Resultados

4.1. Caso de Estudo

Neste trabalho aplicou-se o modelo CAFÉ ao Procedimento para Apuração de Descumprimento de Obrigações (PADO) da Anatel, processo administrativo sancionatório composto por múltiplos atos documentais interdependentes: Auto de Infração, Notificação, Despacho de Instauração, Defesa, Decisão e Recurso, cada qual com requisitos normativos prescritos pela Lei Geral de Telecomunicações (LGT) [Brasil 1997], pelo RASA [Agência Nacional de Telecomunicações (Anatel) 2012] e pela Lei nº 9.784/1999 [Brasil 1999]. O corpus de avaliação é composto por 104 PADOs reais extraídos do SEI público da Anatel [Agência Nacional de Telecomunicações (Anatel) 2026], cobrindo o período de 2000 a 2026, 19 unidades gestoras e volume documental de 3 a 63 arquivos PDF por processo; 73 PADOs (70,2%) apresentaram ao menos uma página digitalizada sem texto extraível, sinalizada para tratamento por Reconhecimento Óptico de Caracteres (OCR), e 31 (29,8%) foram integralmente textuais. O código do pipeline CAFÉ está disponível em [Silva 2026].

4.2. Camada 1 — Integridade Documental

O modelo LDA foi configurado com $k = 10$ tópicos, escolhido como configuração operacional exploratória para representar a heterogeneidade dos tipos documentais presentes

no corpus, e treinado com 20 iterações pelo método *batch* sobre uma matriz documento-termo de 2.000 características (*CountVectorizer*, *min_df*=2, *max_df*=0,99). Os tópicos foram interpretados por inspeção manual das 10 palavras mais representativas de cada componente. O modelo identificou 10 tópicos latentes com distribuição interpretável por superintendência: tópicos 2 e 8 predominam em autuações de clandestinidade; tópicos 5 e 6 em grandes operadoras; e o tópico 1 identifica processos históricos convertidos do formato físico para o SEI.

A métrica de *score* LDA refere-se à distribuição de Dirichlet (γ), que indica a proporção de cada tópico presente em um documento; ela foi utilizada para conferir objetividade estatística à classificação documental, permitindo a identificação de categorias temáticas de forma não supervisionada. Em 96 dos 104 PADOs (92,3%), o maior componente temático apresentou proporção superior a 0,5, indicando predominância relativa de um tópico no documento. Esse resultado não comprova, isoladamente, a correção da classificação tipológica, pois a distribuição de tópicos não equivale a uma anotação de referência. Os oito PADOs restantes apresentam estrutura documental atípica ou alto volume de páginas digitalizadas sem texto extraível. A verificação de consistência documento-metadados revelou que as divergências mais frequentes são: datas do texto incompatíveis com os campos de data do SEI (presente em mais da metade dos processos com inconsistência), identificação de partes divergente entre corpo do documento e campos de metadados, e ausência de número de processo ou unidade gestora nos campos obrigatórios.

4.3. Camada 2 — Conformidade de Fluxo

A verificação foi realizada sobre os 57 PADOs (54,8%) com datas extraíveis dos metadados de data sistêmica ("Data de Geração" no SEI) consistentes para a reconstrução do fluxo. A impossibilidade de construir o *log* de eventos com datas confiáveis para os 47 PADOs restantes (45,2%) decorre de inconsistências nos metadados do SEI, como datas no texto incompatíveis com os campos de registro e ausência de identificação de unidade gestora, o que demonstra ser este o principal fator limitante do pipeline na configuração atual. A conformidade de fluxo foi avaliada por *Token-based Replay* (pm4py), técnica que reproduz cada *trace* do *log* sobre uma rede de Petri e contabiliza os *tokens* produzidos, consumidos, faltantes e remanescentes; a medida de *fitness* resultante varia de 0 a 1, em que 1 indica aderência total do caso ao modelo normativo [van der Aalst 2016]. O *Token-based Replay* sobre o modelo normativo do RASA produziu *fitness* médio de 0,527 (11 PADOs plenamente conformes); o modelo manual com dois fluxos legítimos – Tipo A (A2→A3→A4) e Tipo B (A1→A3→A4), produziu 0,740 (39 PADOs conformes), sugerindo que o modelo normativo rígido derivado do RASA não representa as variantes de fluxo observadas no corpus. A perspectiva temporal detectou atrasos extremos: 3.982 dias entre despacho e notificação (PADO 53500.030037/2023-92) e 5.683 dias entre documento motivador e despacho (PADO 53542.001110/2000-36). Foram identificadas ainda 9 inconsistências em que a data de notificação ao autuado é anterior ao despacho de instauração, indício de irregularidade relevante frente ao devido processo legal (CF, Art. 5º, Inc. LIV), sujeito a validação jurídica, pois indica notificação registrada antes da existência formal do processo. A perspectiva de recursos identificou 61 PADOs com ao menos um ato sem responsável identificado [Li and Goel 2025].

4.4. Camada 3 — Conformidade Semântica

O critério de julgamento entre as abordagens é a concordância por requisito, definida como o percentual dos 104 PADOs em que as Abordagens A e B emitiram o mesmo veredito (conforme ou não conforme) para um dado requisito (Figura 4). Não se trata da taxa de conformidade, e sim da taxa de acordo entre os dois métodos: requisitos com concordância inferior a 80% indicam pontos sensíveis à representação semântica adotada, em que as duas abordagens divergem quanto à presença do requisito no documento. A taxa de conformidade de cada abordagem é reportada apenas como caracterização descritiva do comportamento de cada método, não como medida de qualidade, uma taxa mais alta não implica abordagem superior, pois a Abordagem A, mais restritiva, tende a ser mais precisa, enquanto a Abordagem B pode produzir falsos positivos em requisitos cujo vocabulário é compartilhado entre tipos documentais distintos. Por esse motivo, a vantagem relativa entre os métodos é avaliada pela análise dos pontos de divergência e pela inspeção qualitativa desses casos, e não pela comparação direta das taxas de conformidade.

A escolha dessas duas abordagens não é arbitrária: A e B representam os dois extremos do espectro de verificação semântica, sem e com espaço de domínio, permitindo contrastar o efeito do enriquecimento por LSA sobre a verificação.

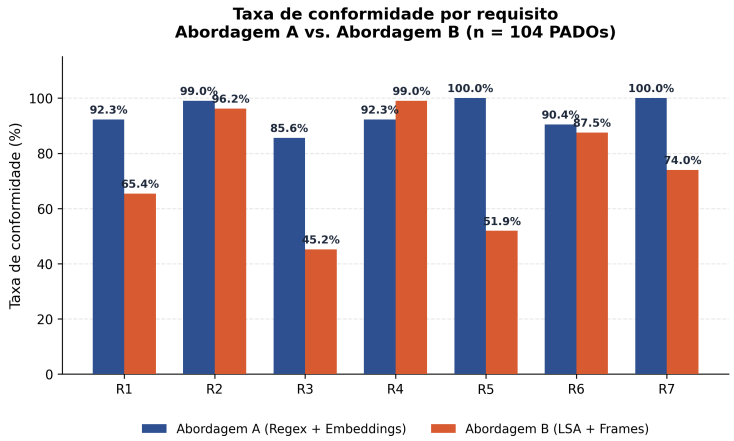


Figure 2. Taxa de conformidade por requisito - Abordagem A vs. Abordagem B (n = 104 PADOs).

Considerando os sete requisitos individualmente, a Abordagem A apresentou taxa média de conformidade detectada de 94,2%. No nível de processo, 87 dos 104 PADOs (83,7%) foram classificados como plenamente conformes segundo os critérios operacionais adotados. Para a Abordagem B, a taxa média de conformidade detectada foi de 74,1%, enquanto 18 PADOs (17,3%) foram classificados como plenamente conformes.

A concordância entre as abordagens foi alta em R2 (95,2%) e R4 (91,3%), requisitos de vocabulário padronizado e baixa ambiguidade; moderada em R6 (79,8%) e R7 (74,0%); e baixa em R1 (64,4%), R3 (32,7%) e R5 (51,9%). A divergência em R1 relaciona-se à heterogeneidade dos documentos motivadores (Auto de Infração, Relatório de Fiscalização, Informe, Denúncia), muitos deles legados e digitalizados sem texto extraível; em R3, a variações na forma de referência à instauração entre tipos documentais; e em R5, à diversidade na citação da fundamentação normativa entre superintendências.

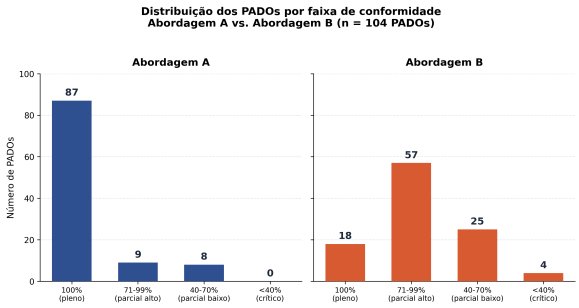


Figure 3. Distribuição por faixa de conformidade (n = 104).

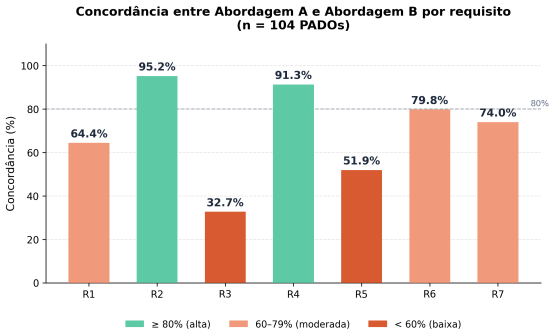


Figure 4. Concordância entre A e B por requisito; barras abaixo de 80% indica sensibilidade semântica (n = 104).

Esses valores expressam concordância entre os métodos, não medidas de acurácia ou superioridade, na ausência de *ground truth* independente. A análise comparativa sugere uma possível regra de decisão prática: a Abordagem A pode ser mais adequada quando o corpus apresenta vocabulário padronizado e contemporâneo, enquanto a Abordagem B pode ser útil em processos históricos ou com maior variação vocabular. Essa hipótese deve ser confirmada por avaliação independente em corpus anotado. Uma arquitetura híbrida, com a Abordagem A como verificador primário e a Abordagem B como verificador complementar nos casos em que a primeira falha, representa uma direção possível para a evolução do **CAFÉ**.

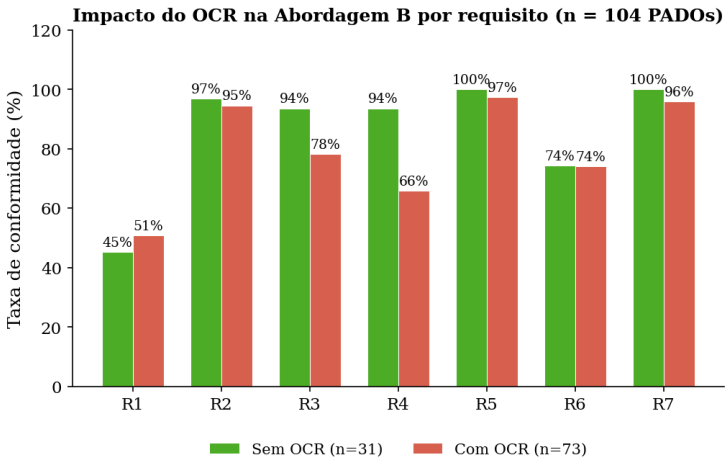


Figure 5. Impacto do OCR na Abordagem B por requisito — processos com e sem OCR (n = 104 PADOs).

Nos 31 PADOs sem OCR, a Abordagem B apresentou maior taxa de conformidade detectada, indicando que parte de sua desvantagem média pode estar associada à qualidade do corpus, e não necessariamente à abordagem em si.

O score composto é definido como $S = 0{,}50 \cdot C_1 + 0{,}35 \cdot C_2 + 0{,}15 \cdot C_3$. Nessa expressão, C_1 é a taxa de conformidade da Camada 1, $C_2 \in \{0; 0{,}5; 1\}$ codifica o fluxo como conforme, parcial ou não conforme, e $C_3 = 1 - \frac{P_{\text{sem texto}}}{P_{\text{total}}}$, sendo $P_{\text{sem texto}}$ o número de páginas sem texto extraível e P_{total} o número total de páginas do processo. Quando o

registro de eventos não pode ser reconstruído, atribui-se $C_2 = 0$. As faixas são: ALTO ($S \geq 0{,}85$), MÉDIO ($0{,}65 \leq S < 0{,}85$), BAIXO ($0{,}40 \leq S < 0{,}65$) e CRÍTICO ($S < 0{,}40$). Os pesos foram definidos por julgamento de especialista, sem calibração formal.

O score composto integrado classificou 44 PADOs (42,3%) como ALTO: processos que satisfazem integridade documental, fluxo normativo e requisitos semânticos com alta cobertura, 46 (44,2%) como MÉDIO: processos com conformidade parcial, geralmente por limitações de metadados ou OCR, e 14 (13,5%) como BAIXO: concentrados em processos com alto volume de páginas digitalizadas e metadados incompletos. Na prática, PADOs com score BAIXO sinalizam processos que requerem revisão humana prioritária antes de qualquer aproveitamento automatizado dos dados, enquanto os de score ALTO apresentam menos inconsistências, embora ainda exijam supervisão humana antes de uso decisório.

5. Discussão e Trabalhos Futuros

Os resultados caracterizam a verificação automática de conformidade em fluxos de processos administrativos sancionatórios em português jurídico como tarefa tecnicamente viável no corpus avaliado; em particular, a Abordagem B, baseada em LSA, realiza a verificação sem depender de *embeddings* neurais. A Abordagem A apresentou taxa média de conformidade detectada de 94,2% e a Abordagem B, 74,1%. Esses valores são descritivos do comportamento dos métodos e, na ausência de anotação independente, não devem ser interpretados como estimativas definitivas de acurácia nem como evidência de superioridade entre as abordagens. A Abordagem B revelou vantagem comparativa nos nove PADOs históricos com variação vocabular, sugerindo que uma arquitetura híbrida, com a Abordagem A como verificador primário e a Abordagem B como complementar nos casos de falha, pode ampliar a cobertura do sistema, possibilidade a avaliar em experimentos futuros com gabarito de referência.

O CAFÉ, como sistema integrado, oferece cobertura complementar às camadas isoladas: a Camada 1 detecta inconsistências de metadados, mas não verifica sequência nem conteúdo; a Camada 2 verifica o fluxo, mas depende de metadados confiáveis; e a Camada 3 verifica o conteúdo, mas não detecta atos fora de ordem. Dos 104 PADOs avaliados, 90 (86,5%) receberam score ALTO ou MÉDIO no indicador integrado, resultado que demonstra a capacidade operacional do pipeline no corpus heterogêneo estudado, sem implicar generalização automática para outras bases.

Na Camada 2, a diferença de 21,3 pontos percentuais entre o modelo normativo rígido derivado do RASA (52,7%) e o modelo manual que reconhece duas variantes de instauração (74,0%) sugere que o modelo rígido não representa explicitamente todos os fluxos observados no corpus. A identificação dessas variantes como legítimas requer análise normativa e validação institucional adicional, mas o resultado indica que a representação computacional do fluxo pode influenciar diretamente a avaliação de conformidade. De forma análoga, as inconsistências temporais detectadas devem ser lidas como indícios relevantes para a análise da regularidade processual; sua caracterização como violação do devido processo legal requer validação jurídica específica, pois diferenças entre datas podem decorrer de problemas de registro ou da distinção entre a produção do ato e seu cadastramento.

A principal limitação metodológica é a ausência de um *ground truth* anotado de forma independente por especialistas. As taxas reportadas descrevem o comportamento de cada abordagem, mas não constituem medidas de acurácia frente a um gabarito jurídico; nos casos de divergência entre as Abordagens A e B, não é possível afirmar qual está correta sem validação humana formal. Durante o desenvolvimento, uma amostra dos resultados foi inspecionada manualmente pela equipe, o que auxiliou a interpretação qualitativa, mas não substitui uma avaliação formal com métricas de precisão, revocação e *F1-score*. Somam-se limitações de cobertura: base única (PADOs da Anatel), dependência da qualidade do OCR em 73 processos, impossibilidade de reconstruir o fluxo de 47 processos por inconsistências de metadados, e escopo restrito à fase de instauração. Todas essas limitações são detectáveis e rastreáveis pelo próprio pipeline, o que contribui para a explicabilidade e para a identificação dos casos que requerem revisão humana.

Três direções de trabalho futuro emergem diretamente dos achados. A primeira é a construção de um corpus anotado por juristas do domínio, que permita mensurar o desempenho por precisão, revocação e *F1-score* frente a um gabarito independente, além de estender a avaliação às fases de instrução, defesa, decisão e recurso do PADO. A segunda é a investigação da arquitetura híbrida entre as Abordagens A e B e a substituição do analisador de dependências pelo BERTimbau [Souza et al. 2020] para extração de papéis semânticos em construções sintáticas complexas. A terceira é a formalização das variantes observadas no fluxo do PADO como modelos normativos alternativos no pm4py [Berti and et al. 2019], permitindo avaliar a conformidade frente a mais de uma trajetória processual admissível.

References

Agência Nacional de Telecomunicações (Anatel) (2012). Regulamento de aplicação de sanções administrativas (RASA). Resolução Anatel n.º 589/2012.

Agência Nacional de Telecomunicações (Anatel) (2026). SEI público - sistema eletrônico de informações. Disponível em: `https://sei.anatel.gov.br/sei/modulos/pesquisa/md_pesq_processo_pesquisar.php?acao_externa=protocolo_pesquisar&acao_origem_externa=protocolo_pesquisar&id_orgao_acesso_externo=0`. Acesso em: junho de 2026.

Baker, C. F., Fillmore, C. J., and Lowe, J. B. (1998). The Berkeley FrameNet project. In *Proceedings of the 17th International Conference on Computational Linguistics (COLING)*, pages 86–90.

Berti, A. and et al. (2019). Process mining for Python (pm4py): A library for process discovery and conformance checking. In *Proceedings of the ICPM Demo Track*.

Blei, D. M., Ng, A. Y., and Jordan, M. I. (2003). Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022.

Brasil (1997). Lei n.º 9.472, de 16 de julho de 1997 - lei geral de telecomunicações (LGT). Diário Oficial da União.

Brasil (1999). Lei n.º 9.784, de 29 de janeiro de 1999 - regula o processo administrativo no âmbito da administração pública federal. Diário Oficial da União. Art. 50.

Brasil. Ministério da Gestão e da Inovação em Serviços Públicos (2026). Portaria MGI nº 3.485, de 24 de abril de 2026 - institui a política de governância de inteligência artificial no âmbito do MGI. Diário Oficial da União.

Brasil. Secretaria Especial da Receita Federal do Brasil (2026). Portaria RFB nº 647, de 5 de fevereiro de 2026 - dispõe sobre a política de inteligência artificial no âmbito da RFB. Diário Oficial da União.

Cejas, O. A., Azeem, M. I., Abualhaija, S., and Briand, L. C. (2023). NLP-based automated compliance checking of data processing agreements against GDPR. *IEEE Transactions on Software Engineering*, 49(9):4282–4303.

Costa, J. A. F., Dantas, N. C. D., and Silva, E. D. d. S. A. (2023). Evaluating text classification in the legal domain using BERT embeddings. In *Intelligent Data Engineering and Automated Learning – IDEAL 2023*, volume 14404 of *Lecture Notes in Computer Science*. Springer.

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., and Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, 41(6):391–407.

Fillmore, C. J. (1982). Frame semantics. In *Linguistics in the Morning Calm*, pages 111–137. Hanshin Publishing Co., Seoul.

Hassani, S. (2024). Enhancing legal compliance and regulation analysis with large language models. arXiv preprint arXiv:2404.17522. Doctoral Symposium, 32nd IEEE International Requirements Engineering Conference (RE'24).

Li, X. and Chen, Y. (2025). Domain-specific fine-tuning of sentence-BERT for enhanced topic coherence. *Journal of Natural Language Engineering*.

Li, Y. and Goel, S. (2025). Making it possible for the auditing of AI: A systematic review of AI audits and AI auditability. *Information Systems Frontiers*, 27(3):1121–1151.

Paul, P. C., Ahmed, M. T., Chakraborty, D., and Rahman, M. S. (2025). Combining BERT with LDA: Improved topic modeling in Bengali language. *IAENG International Journal of Computer Science*, 52(2). Article IJCS_52_2_11.

Silva, W. F. F. d. (2026). CAFÉ - pipeline de verificação de conformidade em PADOs. Disponível em: `https://github.com/wuldson-franco/cafe-frw-semantic-frame`. Acesso em: junho de 2026.

Souza, F., Nogueira, R., and Lotufo, R. (2020). BERTimbau: Pretrained BERT models for Brazilian Portuguese. In *9th Brazilian Conference on Intelligent Systems (BRACIS)*.

van der Aalst, W. M. P. (2016). *Process Mining: Data Science in Action*. Springer, Berlin, 2 edition.

Zhu, L. and et al. (2025). LDA as a semantic enrichment module for BERT-based models. In *Proceedings of the 2025 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.