

**Pró-Reitoria Acadêmica
Programa de Pós-Graduação Stricto Sensu em
Governança, Tecnologia e Inovação**

**ATOS DE CONCENTRAÇÃO ECONÔMICA: GERAÇÃO DE
MINUTA DE PARECER DECISÓRIO PARA PROCESSOS DE
RITO SUMÁRIO DO CADE**

**Autora: Ana Paula Pessoa Mello
Orientador: Prof. Dr. Edilson Ferneda
Coorientadores: Prof. Dr. Hércules Antonio do Prado
M.Sc. Ricardo Silva Carvalho**

**Brasília – DF
2023**

ANA PAULA PESSOA MELLO

**ATOS DE CONCENTRAÇÃO ECONÔMICA: GERAÇÃO DE MINUTA DE
PARECER DECISÓRIO PARA PROCESSOS DE RITO SUMÁRIO DO CADE**

Dissertação apresentada ao Programa de Pós-Graduação *Stricto Sensu* em Governança, Tecnologia e Inovação da Universidade Católica de Brasília, como requisito parcial para obtenção do Título de Mestre em Governança, Tecnologia e Inovação.

Orientador: Prof. Dr. Edilson Fereda
Coorientadores: Prof. Dr. Hércules Antonio do Prado
e M.Sc. Ricardo Silva Carvalho

Brasília – DF
2023

M527a Mello, Ana Paula Pessoa.
Atos de concentração econômica : geração de minuta de parecer decisório para processos de rito sumário do CADE / Ana Paula Pessoa Mello. – 2023.
63 f. : il. ; 30 cm.

Dissertação (mestrado) – Universidade Católica de Brasília, Programa de Pós-Graduação Stricto Sensu em Governança, Tecnologia e Inovação, 2023.
Orientação: Prof. Dr. Edilson Ferneda.
Coorientação: Prof. Dr. Hércules Antonio do Prado e M.Sc. Ricardo Silva Carvalho.

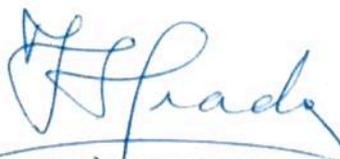
1. Inteligência artificial. 2. Recuperação de informação legal. 3. Decisão automatizada. 4. Geração de minuta. I. Ferneda, Edilson. II. Prado, Hércules Antonio do. III. Carvalho, Ricardo Silva. IV. Título.

CDU 658

Dissertação de autoria de Ana Paula Pessoa Mello, intitulada “Ato de concentração econômica: geração de minuta de parecer decisório para processos de rito sumário do CADE”, apresentada como requisito parcial para obtenção do grau de Mestre em Governança, Tecnologia e Inovação da Universidade Católica de Brasília, em 24 de fevereiro de 2023 defendida e aprovada pela banca examinadora abaixo assinada:



Prof. Dr. Edilson Feneda
Orientador
Universidade Católica de Brasília – UCB



Prof. Dr. Hércules Antonio do Prado
Coorientador
Universidade Católica de Brasília – UCB



Prof. Dr. Thiago Christiano Silva
Avaliador interno
Universidade Católica de Brasília – UCB



Prof. Dr. Ricardo Marcos Chaim
Avaliador externo
Universidade de Brasília – UnB

Dedico este trabalho a Deus, por sempre me guiar. Ao meu marido e aos meus filhos, pelo amor, pelo apoio e pelo companheirismo. Ao meu pai (*in memoriam*) e à minha mãe por terem sempre apoiado as escolhas que me fizeram quem sou. Aos meus orientadores e ao pesquisador Ricardo Carvalho que me deram todo o suporte necessário para a conquista deste título. Ao CADE, que o tornou possível.

*Parte da ausência de humanidade do computador deve-se
a que, competentemente programado e trabalhando bem,
é completamente honesto. — Isaac Asimov*

RESUMO

Referência: MELLO, Ana Paula Pessoa. Atos de concentração econômica: geração de minuta de parecer decisório para processos de rito sumário do CADE. 2023. 63 f. Dissertação do Mestrado em Governança, Tecnologia e Inovação - Universidade Católica de Brasília, Brasília, 2023.

Em uma realidade globalizada, a informação é considerada o ativo principal de uma organização. Ao considerar o avanço das tecnologias de Inteligência Artificial (IA) em conjunto com os grandes volumes de dados gerados diariamente em taxas cada vez mais exponenciais, tem-se um cenário favorável para a expansão dessas tecnologias em várias áreas da vida cotidiana, incluindo a área jurídica. O Conselho Administrativo de Defesa Econômica (CADE) é um tribunal administrativo e sua responsabilidade é investigar e punir administrativamente, em instância única, pessoas físicas e jurídicas que pratiquem infrações à ordem econômica. Além disso, o CADE também analisa atos de concentração, de modo a minimizar possíveis efeitos negativos no ambiente concorrencial de determinado mercado. Ainda que seja um tribunal administrativo, as decisões administrativas do CADE têm o peso de decisão judicial em sua área de competência. A utilização das tecnologias de IA na área judicante traz resultados promissores tanto na rapidez como na precisão das decisões. Neste cenário, foi proposta uma prova de conceito de um processo inteligente para a produção de minutas de pareceres de processos de rito sumário no âmbito do CADE. Foram utilizadas três abordagens de sumarização, sendo duas extrativas e uma abstrativa. Em avaliação dessas abordagens foi constatado que as extrativas apresentaram melhores resultados para os objetivos desse projeto de pesquisa do que a abstrativa e, dessa forma, atestou-se a validade dos resultados apresentados para a área de negócio e para o CADE.

Palavras-chave: Inteligência Artificial, Recuperação de Informação Legal, Decisão Automatizada, Geração de Minuta.

ABSTRACT

In a globalized reality, information is considered the main asset of an organization. When considering the advancement of Artificial Intelligence (AI) technologies together with the large volumes of data generated daily at increasingly exponential rates, there is a favorable scenario for the expansion of AI technologies in several areas of everyday life, including the legal area. The Administrative Council for Economic Defense (CADE) is an administrative court and its responsibility is to investigate and administratively punish, in a single instance, individuals and legal entities that commit violations of the economic order. In addition, CADE also analyzes mergers in order to minimize possible negative effects on the competitive environment of a certain market. Even though it is an administrative court, CADE's administrative decisions have the weight of a judicial decision in its area of competence. The use of AI technologies in the adjudication area brings promising results in terms of both speed and accuracy of decisions. In this scenario, a proof of concept of an intelligent process for the production of draft opinions of summary proceedings within the scope of CADE was proposed. Three summarization approaches were used, two extractive and one abstractive. In evaluating these approaches, it was verified that the extractive ones presented better results for the objectives of this research project than the abstractive one and, in this way, the validity of the results presented for the business area and for CADE was attested.

Keywords: Artificial Intelligence, Legal Information Retrieval, Automated Decision, Draft Generation.

LISTA DE QUADROS

Quadro 1: Artigos mais citados mais aderentes à temática jurídica	18
Quadro 2: Artigos mais recentes mais aderentes à temática jurídica	20
Quadro 3: Sumarização obtida com as ferramentas PyTextRank e spaCy.....	40
Quadro 4: Sumarização obtida com as ferramentas NLTK e spaCy	42
Quadro 5: Sumarização obtida com a ferramenta <i>transformers</i> e o modelo <i>ptt5-base-summ-temario</i>	45
Quadro 6: Sumarização obtida com a ferramenta <i>transformers</i> e o modelo <i>ptt5-base-summ-xlsum</i>	45
Quadro 7: Sumarização obtida com a ferramenta <i>transformers</i> e o modelo <i>ptt5-base-summ-wikilingua</i>	45
Quadro 8: Sumarização obtida com a ferramenta <i>transformers</i> e o modelo <i>ptt5-base-summ-cstnews</i>	45

LISTA DE FIGURAS

Figura 1: Gráfico de densidade da rede de coocorrência de palavras-chave	25
Figura 2: Gráfico da rede de coocorrência de palavras-chave	26
Figura 3: Os estágios do CRISP-DM	27
Figura 4: Planejamento da POC	29
Figura 5: Etapas do processamento de Text Mining para a geração da sumarização de texto	31
Figura 6: Fluxo da análise de rito sumário	36
Figura 7: Abordagens de sumarização de texto	38
Figura 8: Reconhecimento de entidades nomeadas.....	41
Figura 9: Nuvem de palavras	44
Figura 10: Modelos de sumarização em português.....	45

LISTA DE TABELAS

Tabela 1: Termos mais relevantes	26
Tabela 2: Metodologia CRISP-DM adaptada a este trabalho de pesquisa	29
Tabela 3: Indicadores de validação da POC	46

LISTA DE SIGLAS

- CADE – Conselho Administrativo de Defesa Econômica
- CGAA5 – Coordenação-Geral de Análise Antitruste 5
- CRISP-DM – *Cross Industry Standard Process for Data Mining*
- DM – *Data Mining*
- FN – Formulário de Notificação
- IA – Inteligência Artificial
- KDD – *Knowledge Discovery in Databases*
- KDT – *Knowledge Discovery in Textual Databases*
- ML – *Machine Learning*
- OCDE – Organização para Cooperação e Desenvolvimento Econômico
- PDF – *Portable Document Format*
- PLN – Processamento de Linguagem Natural
- POC – *Proof of Concept*
- PoS – *Part of Speech*
- TF-IDF – *Term Frequency-Inverse Document Frequency*
- TIC – Tecnologia da Informação e Comunicação
- TM – *Text Mining*

SUMÁRIO

1. Introdução	12
2. A Inteligência Artificial na área judicial	14
3. Revisão bibliográfica	17
3.1 Síntese dos artigos mais citados	17
3.2 Síntese dos artigos mais recentes	20
3.3 Rede de coocorrência de palavras-chave.....	23
4. Abordagem metodológica	27
4.1 Metodologia empregada.....	27
4.2 Etapas da pesquisa	29
5. Materiais e métodos	33
6. Resultados	35
6.1 Resultados obtidos.....	35
6.1.1 Compreensão do negócio.....	35
6.1.2 Compreensão dos dados	38
6.1.3 Preparação dos dados	39
6.1.4 Modelagem	39
6.1.5 Avaliação	46
7. Conclusão e Trabalhos Futuros	48
Referências	50
Anexo A: Anexo II da resolução nº 33, de 14 de abril de 2022	54
Anexo B: Texto “limpo” utilizado nos modelos de sumarização	59

1. Introdução

A tecnologia de Inteligência Artificial (IA) tem crescido de forma exponencial em todo o mundo e pretende mudar a realidade de várias atividades realizadas por humanos, em maior ou menor grau. Nesse sentido, a área judicial não ficará de fora dessa tendência. Vários estudos, a exemplo de Von Hohendorff, Lazzaretti (2020), Sousa (2019), Freitas e Barddal (2019), Tacca e Rocha (2018), Higasi (2020), Engelmann e Fröhlich (2020), mostram que a utilização da tecnologia de IA na área judicante traz resultados promissores tanto na rapidez como na precisão das decisões.

Esta pesquisa trata da utilização dessas técnicas de IA para a geração de pareceres de técnicos para a gestão de organizações. O trabalho é instanciado no caso específico do Conselho Administrativo de Defesa Econômica (CADE)¹. Busca-se propor um processo inteligente para a produção de minutas de pareceres nos processos de atos de concentração econômica. O objetivo é reduzir o tempo de resposta dos processos de rito sumário analisados pelo CADE para promover maior celeridade e precisão às decisões. Como delimitação da pesquisa, o presente trabalho focaliza os processos classificados como de rito sumário, no âmbito do CADE. A análise ficará restrita à base de dados do SEI, o repositório de processos do CADE.

Em maio de 2012, com a entrada em vigor da nova Lei de Defesa da Concorrência (Lei nº 12.529/2011), o CADE passou a ser uma autoridade de defesa da concorrência responsável por instruir os processos administrativos de apuração de infrações à ordem econômica, assim como os processos de análise de atos de concentração econômica (doravante designados atos de concentração).

Os atos de concentração são de notificação obrigatória ao CADE para serem analisados. Os ritos processuais do CADE são divididos em dois tipos, consoante a complexidade do processo, e cada tipo tem um tempo máximo para ser julgado: (i) rito sumário: até 30 dias; e (ii) rito ordinário: até 240 dias.

Os estudos analisados (VON HOHENDORFF; LAZZARETTI, 2020; SOUSA, 2019; FREITAS; BARDDAL, 2019; TACCA; ROCHA, 2018; HIGASI, 2020; ENGELMANN; FRÖHLICH, 2020) apontam que as principais técnicas de IA aplicadas à área judicante são: (i) *machine learning* (aprendizado de máquina), em projetos de classificação de documentos e previsões de decisões; (ii) *deep learning* (aprendizado profundo), para organização e divisão de

¹ <https://www.gov.br/cade/pt-br/aceso-a-informacao/institucional>. Acesso em 20 mai. 2022.

tarefas; e *(iii)* processamento de linguagem natural (PLN), na sumarização e *drafting* de documentos, e indexação para busca de casos semelhantes (jurisprudência).

O CADE é uma entidade técnica, reconhecida por sua excelência, e é considerada uma das dez melhores agências antitruste do mundo, além de ser membro do Comitê de Concorrência da OCDE desde 2019². Assim, espera-se que este projeto de pesquisa traga mais um reconhecimento ao CADE por sua busca por inovação e oferta de serviços públicos de qualidade para a sociedade. Justifica-se, então, o investimento em tecnologias como a IA na busca por agilidade à geração de pareceres decisórios, fundamentados na legislação pertinente, no âmbito do CADE.

Além disso, o uso de dados não estruturados para a tomada de decisão traz mais oportunidades de *insights*, pois pode-se extrair conhecimento de inúmeras fontes, mas a maior delas são os dados não estruturados (ISE, 2016). No âmbito desse projeto de pesquisa, os formulários de notificação a serem analisados estão em formato PDF. Dessa forma, outras iniciativas na área jurídica que venham a utilizar dados não estruturados, poderão se beneficiar dos resultados desse projeto de pesquisa.

Por fim, a base de dados a ser trabalhada nesse projeto de pesquisa é única. Os formulários de notificação em PDF estão no Sistema Eletrônico de Informações (SEI)³, facilitando as análises e as avaliações dos modelos a serem estudados.

Serão buscados os seguintes objetivos específicos: *(i)* estudo sobre o estado da arte em geração automática de pareceres jurídicos; *(ii)* identificação das técnicas de IA, mais especificamente de PLN, aplicáveis no presente caso; *(iii)* proposição de um modelo de solução e *(iv)* elaboração de uma prova de conceito.

² <https://cdn.cade.gov.br/Portal/Noticias/2020/caderno-cade-ocde-2020.pdf>. Acesso em 15 dez. 2022.

³ <https://www.gov.br/economia/pt-br/acesso-a-informacao/sei/sobre>. Acesso em 6 dez. 2022.

2. A INTELIGÊNCIA ARTIFICIAL NA ÁREA JUDICIAL

A tecnologia de IA tem crescido de forma exponencial em todo o mundo e pretende mudar a realidade de várias atividades realizadas por humanos, em maior ou menor grau. E a área judicial não ficará de fora dessa tendência. Uma das aplicações de destaque da IA na área judicial é a Recuperação de Informação Legal (*Legal Information Retrieval*), que se serve de Processamento de Linguagem Natural (*Natural Language Processing*), Aprendizado de Máquina (*Machine Learning*) e Extração de Conhecimento (*Knowledge Extraction*) (SANSONE; SPERLÍ, 2021).

Medina (2020) define a IA como um ferramental computacional “que emula a inteligência humana ou que executa funções a ela relacionadas, tais como raciocínio, aprendizagem, adaptabilidade, percepção e interação com o meio físico etc”. Nesse sentido, os sistemas computacionais seriam dotados de variadas técnicas que permitiriam uma flexibilidade necessária para realizar as tarefas de IA, como “*machine learning* (aprendizagem de máquina) e *deep learning* (aprendizagem profunda), por meio de redes neurais artificiais, processamento da linguagem natural (*natural language processing*) e análise de grandes conjuntos de dados (*big data*)”.

Buocz (2018) apresenta três tipos de papéis que a IA pode assumir no Direito: (i) pesquisador (“Librarian”); (ii) advogado-geral (“Advocate General”); e (iii) funcionário com poderes judiciais limitados (“Official with Limited Judicial Powers”). O primeiro papel tem como função a busca de materiais úteis que possam fundamentar as decisões (Recuperação de Informação). Esse papel ajuda a acelerar a busca de processos pertinentes ao tema disposto. O segundo papel deve conseguir extrair e consolidar informações relevantes sobre o tema disposto, de forma imparcial, para assistir o juiz no exercício de suas funções (Sumarização de Documentos). O terceiro papel separa o trabalho do juiz humano e o trabalho da IA. A IA teria o poder judiciário para decidir os casos de forma autônoma (Decisões Automatizadas). No entanto, a interposição de recurso contra a sua decisão a invalidaria automaticamente. O juiz humano, então, daria a decisão final.

Neste projeto de pesquisa, o foco em cada um dos papéis é: (i) como pesquisador, buscar notificações similares; (ii) como advogado-geral, sumarizar as notificações; e (iii) como funcionário com poderes judiciais limitados, gerar uma minuta do parecer decisório e não um parecer final com a própria decisão. O analista responsável prepara o parecer decisório para revisão da instância superior a partir da minuta gerada de forma automática.

Recuperação de informação legal

A área jurídica se beneficiou muito do avanço das tecnologias de informação e comunicação (TIC), em particular da IA, sobretudo para a recuperação de documentos similares por meio da análise semântica (SANSONE; SPERLÍ, 2021). Existem diferentes métodos para medir a similaridade entre documentos como a modelagem de tópico (*topic modeling*) e os modelos de rede neurais como as *word embeddings* e os *document embeddings* (MANDAL et al., 2017).

Independentemente do método utilizado, é preciso definir uma forma de medir e comparar a similaridade para que a máquina possa apontar quais documentos são mais ou menos similares. Também é preciso representar o texto dos documentos de forma quantificável (como um objeto matemático, que geralmente é uma forma vetorial), para ser possível realizar os cálculos de similaridade desses documentos.

Modelagem de tópico usa modelos estatísticos para descobrir tópicos em um *corpus* (conjunto de documentos), gerando um vetor de tamanho fixo para cada documento. Esse vetor reflete a distribuição de tópicos no documento e a similaridade.

Word embedding é um processo que cria um vetor para cada palavra do documento observando-se as ocorrências das outras palavras em relação à palavra sendo analisada. Esse método pode fornecer contextos significativos das palavras. Palavras semanticamente semelhantes tendem a ficar mais próximas no espaço vetorial.

Document embedding é um processo para converter um determinado documento em uma representação vetorial, que captura seus aspectos semânticos. A partir desse vetor numérico, é possível fazer o cálculo de similaridade para comparação com outros documentos.

Sumarização de documentos

A sumarização de documentos no domínio legal difere da sumarização em outros domínios (TRIVEDI et al. 2021). Isso em função do tamanho e da estrutura de seus documentos, do vocabulário próprio e do volume de citações em seus documentos. Existem duas abordagens de sumarização textual: (i) a extrativa, que identifica as sentenças mais importantes de um documento e as extrai como elas são, e (ii) a abstrativa, que se utiliza dos conceitos principais do documento e do conhecimento do domínio para gerar um texto diferente do original, mas com o mesmo sentido. Ambas as abordagens utilizam técnicas de PLN, porém, por também utilizar inferência e semântica, a sumarização abstrativa tem um desempenho menor.

Decisões automatizadas

A predição legal quantitativa refere-se a homens e máquinas trabalhando em conjunto de forma mais produtiva do que cada um separadamente (KATZ, 2013). O potencial preditivo das tecnologias de reconhecimento de padrões em grandes conjuntos de dados possibilita que estas se tornem uma ferramenta viável na tomada de decisão jurídica.

Esse potencial preditivo na tomada de decisão foi possibilitado pela evolução do poder computacional (capacidade de armazenar e analisar imensos conjuntos de dados) ocorrida nas últimas décadas. As técnicas de IA (PLN, *machine learning* e *deep learning*, por exemplo) e análises preditivas permitem aos tomadores de decisão extrair valor de grandes conjuntos de dados e, a partir da descoberta de padrões nesses dados, construir perfis (*profiling*) que podem ser utilizados no apoio à tomada de decisões automatizadas (DIEGA, 2018).

Ao conjecturar sobre as tecnologias de *big data* na prática jurídica, Pugliesi e Brandão (2015) argumentam que, no modelo da sociedade de controle e informação, “o uso das tecnologias de *big data* gera impactos sobre os mais diversos campos, que são tocados pelo fenômeno jurídico, como a privacidade, discriminação, direito público à informação e até automação da tomada de decisão (jurídica), dentre muitos outros”. Assim, o processo de automatização de decisões passa a ser uma realidade em que os algoritmos vêm sendo aplicados “para dar mais racionalidade a decisões judiciais”, gerando “impactos diretos sobre a sociedade” (DONEDA; ALMEIDA, 2016).

Segundo Cobbe (2019), decisões automatizadas podem ser usadas de duas formas: (i) decisão totalmente automatizada, sem intervenção humana, e (ii) decisão em que o sistema é um guia ou uma ferramenta para ajudar alguma pessoa a tomar uma decisão. No âmbito desse projeto de pesquisa, será utilizada a segunda abordagem, pois o produto final será uma minuta de parecer que ajudará o analista a fundamentar sua decisão final.

3. REVISÃO BIBLIOGRÁFICA

O presente trabalho realizou uma pesquisa bibliométrica⁴ no portal *Scopus*⁵ com a seguinte *string* de pesquisa:

```
TITLE-ABS-KEY
  (((legal OR juridic)
  AND ("document* similarity") OR ("paragraph* similarity") OR ("sentence* similarity")
  OR ("knowledge extraction") OR ("information retrieval"))))
AND (LIMIT-TO (PUBYEAR, 2022) OR LIMIT-TO (PUBYEAR, 2021)
  OR LIMIT-TO (PUBYEAR, 2020) OR LIMIT-TO (PUBYEAR, 2019)
  OR LIMIT-TO (PUBYEAR, 2018) OR LIMIT-TO (PUBYEAR, 2017))
```

Foram consideradas publicações entre 2017 e 2022, sendo obtidos 541 documentos. A pesquisa foi limitada às seguintes áreas de interesse: *Computer Science, Engineering, Business, Management and Accounting, Decision Sciences, Mathematics* e *Social Sciences* e aos seguintes tipos de documentos: *article, review, conference paper* e *conference review*, totalizando 457 documentos. Foi gerado um arquivo com os metadados para posteriores análises dos dados da pesquisa.

Uma rede de coocorrência de palavras-chave foi gerada para investigar os temas próximos ao foco da pesquisa. Foi elaborado o tesauro para controle do vocabulário e normalização dos termos relevantes, aplicados para um refinamento da rede de coocorrência e do mapa de densidade. Em seguida, foi realizada uma análise da rede (grau médio, diâmetro da rede, modularidade, centralidade de autovetor e coeficiente de *clustering* médio), identificando-se os termos mais relevantes.

3.1. SÍNTESE DOS ARTIGOS MAIS CITADOS

No Quadro 1 são apresentados os artigos mais citados, seus respectivos temas e o número de citações.

⁴ A pesquisa foi realizada em 13 jun. 2022.

⁵ <https://www.scopus.com/home.uri>.

Quadro 1: Artigos mais citados mais aderentes à temática jurídica

Documento	Tema	Nº de Citações
Chalkidis e Kampas (2019)	<i>Deep learning</i> em análise de textos legais	48
Chalkidis, Androutsopoulos e Michos (2017)	Automatização de extração de elementos contratuais	47
Van Opijnen e Santos (2017)	Relevância na recuperação de textos legais	45
Wei et al. (2019)	<i>Deep learning</i> para classificação de textos legais	40
Mandal et al. (2017)	Similaridade em documentos legais	30
Cardellino et al. (2017)	Extração de informação em documentos legais	26
Sugathadasa et al. (2019)	Processamento de linguagem natural em recuperação de informações em textos legais	24
Undavia, Meyers e Ortega (2018)	Redes Neurais em classificação de documentos de pareceres legais	24
Bhattacharya et al. (2019)	Algoritmos de sumarização aplicados aos julgamentos legais	23
Araújo et al. (2018)	Sistemas de reconhecimento de entidade nomeada aplicada a documentos legais brasileiros	22

Fonte: Portal *Scopus*

Chalkidis e Kampas (2019) observam que a aplicação de *Deep Neural Networks* para análises jurídicas aumentou significativamente nos últimos anos. Os autores focaram nas aplicações em classificação de texto, extração e recuperação de informações. Eles utilizaram representações de recursos semânticos (*e.g.*, *word embeddings*), que vem se mostrando um instrumento fundamental na aplicação de *Deep Learning* em PLN nas seguintes perspectivas: (i) representação de recursos de texto, (ii) arquitetura de rede neural e (iii) desempenho e resultados.

Chalkidis, Androutsopoulos e Michos (2017) estudam a automatização da extração de elementos de contratos (*e.g.*, datas de rescisão, referências de legislação, partes contratantes, pagamentos acordados). Para tal, analisam um conjunto de dados rotulados com anotações manuais de elementos de contrato e outro conjunto de dados não rotulados para pré-treinar as PoS (*part-of-speech*) *tag embeddings* e as *word embeddings*. Além disso, os autores definem um conjunto de características particulares (*hand-crafted features*) e o significado de cada característica (o que cada característica representa) para aplicar aos conjuntos de dados. Dessa forma, os autores aplicam métodos de extração dos elementos de contrato como algoritmos de *machine learning* com as características particulares, as *word* e PoS *tag embeddings* e comparam com outros métodos em que as regras são definidas manualmente (usadas em vez dos classificadores de *machine learning* por meio de expressões regulares).

Van Opijnen e Santos (2017) discutem um *framework* conceitual da “relevância” na recuperação de informação jurídica. Relevância, no contexto legal, merece atenção específica,

devido ao rápido crescimento dos repositórios, às características distintas dos objetos da informação jurídica e às complicadas tarefas dos profissionais do direito. O *framework* é baseado em uma tipologia de dimensões de relevância usadas na ciência da recuperação da informação geral, mas adaptada às características específicas da informação jurídica. E pode ser usado para desenvolver ou melhorar os sistemas de recuperação de informação jurídica.

Wei et al. (2019) relatam estudos preliminares sobre o uso de *Deep Learning* na classificação de documentos jurídicos. *Deep Learning* possibilitou um avanço significativo em PLN, como reconhecimento de fala e tradução de idiomas, sendo adaptada para a classificação de texto. Os autores comparam os resultados de um algoritmo supervisionado com os resultados de um algoritmo de *Deep Learning* e mostram que os resultados de *Deep Learning* têm melhor desempenho.

Mandal et al. (2017) apresentam um estudo para melhorar as metodologias baseadas em texto (como TF-IDF) para calcular a similaridade entre documentos jurídicos. Os autores usam medidas avançadas de similaridade (como modelagem de tópicos) e modelos de rede neural (como *word embeddings* e *document embeddings*) para comparação e demonstram que as metodologias propostas baseadas em *embeddings* superam significativamente outras abordagens.

Cardellino et al. (2017) criam um Reconhecedor de Entidade Nomeada, Classificador e Vinculador para a área jurídica. Os autores mostram que, com essa ferramenta, podem identificar partes relevantes de textos e vinculá-las a uma representação estruturada do conhecimento, a ontologia LKIF (*Legal Knowledge Interchange Format*), que é baseada na ontologia YAGO⁶.

Sugathadasa et al. (2019) desenvolvem três novos modelos, comparados com o *golden standard*, por meio dos repositórios *online* fornecidos, especificamente para o domínio jurídico. Os autores apresentam a pesquisa que está sendo realizada no processo de representação de documentos de casos jurídicos em diferentes espaços vetoriais, incorporando medidas semânticas de palavras (técnicas de similaridades e *embeddings*) e técnicas de PLN (TF-IDF, *Bag of Words*). Também comprovam a necessidade de incorporação de medidas de similaridade semântica de domínio específico no processo de recuperação da informação e o impacto da variação da distribuição das medidas de similaridade de palavras, em relação a diferentes dimensões de vetores de documentos, para melhorar o processo de recuperação de informações jurídicas.

Undavia, Meyers e Ortega (2018) apresentam o classificador da Suprema Corte americana (*Supreme Court Classifier - SCC*⁷), um sistema que aplica as técnicas de redes neurais ao problema de classificação documental de pareceres judiciais. Os autores comparam o método

⁶ Disponível em <https://yago-knowledge.org/>. Acesso em 27 jun. 2022.

⁷ Disponível em https://github.com/samir1/web_of_law_sctus_classification. Acesso em 28 jun. 2022.

tradicional de aprendizado de máquina com os métodos recentes de redes neurais aplicados a PLN. O objetivo do SCC é determinar o melhor método para aplicar a classificação automatizada de documentos a textos jurídicos com o intuito de facilitar essa classificação de documentos judiciais para os especialistas jurídicos.

Bhattacharya et al. (2019) realizam a primeira comparação sistemática de algoritmos de sumarização aplicados a julgamentos de casos legais. A sumarização desses julgamentos é bastante complexa em função do tamanho e da complexidade de tais documentos. Foram experimentados conjuntos de dados da Suprema Corte indiana e uma grande variedade de algoritmos de sumarização, incluindo não supervisionados e supervisionados, e os desempenhos foram avaliados considerando abordagens de sumarização de documentos em geral e considerando, também, abordagens específicas para sumarização de documentos legais em outros países.

Araújo et al. (2018) apresentam um conjunto de dados para reconhecimento de entidades nomeadas em documentos legais brasileiros, composto inteiramente por documentos legais, diferentemente de outros conjuntos de dados em língua portuguesa. Os sistemas de reconhecimento de entidades nomeadas têm o potencial não explorado de extrair informações de documentos legais, o que pode aprimorar a recuperação de informações e os processos de tomada de decisão no contexto legal.

3.2. SÍNTESE DOS ARTIGOS MAIS RECENTES

No Quadro 2 são apresentados os artigos mais recentes e seus respectivos temas.

Quadro 2: Artigos mais recentes mais aderentes à temática jurídica

Documento	Tema
Bi et al. (2022)	Similaridade em documentos legais com <i>graph embedding</i>
Šavelka e Ashley (2022)	Extração de argumentação em documentos legais
Li et al. (2022)	Classificação e modelagem de tópicos em documentos legais
Fernandes et al. (2022)	Extração de valor das decisões judiciais brasileiras
Sansone e Sperlí (2022)	Inteligência artificial no domínio jurídico
Castano et al. (2022)	Exploração e recuperação de documentos legais
Song, D. et al. (2022)	Classificação multilabel de documentos legais
Chen et al. (2022)	Classificação automatizada de textos legais
Song, Z. et al. (2022)	Classificação multilabel de textos legais
Sposito et al. (2022)	Extração de informação em documentos legais com o uso de expressões regulares

Fonte: Portal *Scopus*

Bi et al. (2019) apresentam um modelo genérico não supervisionado baseado em grafos chamado *Legal HETerogeneous Graph Representation Learning* (L-HetGRL) para medir a similaridade entre documentos jurídicos no contexto chinês. O modelo L-HetGRL incorpora a aprendizagem de informações de documentos legais e o conhecimento externo do domínio jurídico (por meio de enciclopédias). Além disso, possui um módulo de alinhamento semântico sensível ao domínio jurídico que combina efetivamente entidades jurídicas (representadas por *word embeddings*) e suas semânticas em documentos, melhorando assim a representação das entidades. O L-HetGRL é utilizado para obter os *embeddings* e medir a semelhança entre documentos legais e garantir que as mesmas situações sejam tratadas de forma idêntica, na prática judicial.

Šavelka e Ashley (2022) propõem uma nova tarefa de descoberta de sentenças para argumentação sobre o significado de termos jurídicos (*statutory terms*). A tarefa envolve analisar o tratamento passado aos termos jurídicos, que os advogados realizam rotineiramente usando uma combinação de abordagens manuais e computacionais. A análise é um componente-chave na atividade geral de argumentar um determinado ponto jurídico que envolva incerteza sobre o significado de uma frase jurídica (*statutory phrase*). A tarefa modela a análise do tratamento anterior desses termos jurídicos e trata a descoberta de sentenças como um caso especial de recuperação de documentos *ad hoc*. As recuperações específicas incluem a recuperação de textos curtos (frases), tipos de documentos especializados (textos de casos jurídicos) e, essencialmente, a definição única de relevância documental fornecida em diretrizes de anotação detalhada.

Li et al. (2022) propõem um novo modelo NMF (*Non-negative Matrix Factorization*), denominado *Guided Semi-Supervised Non-negative Matrix Factorization* (GSSNMF), que realiza tanto a classificação quanto a modelagem de tópicos, ao usar informações de rótulos adicionais para melhorar a coerência dos resultados de modelagem de tópicos enquanto incorpora palavras-chave para uma classificação mais precisa. O método proposto melhora tanto a precisão da classificação quanto a coerência do tópico em comparação com métodos anteriores, como *Semi-Supervised Non-negative Matrix Factorization* (SSNMF), *Guided Non-negative Matrix Factorization* (*Guided NMF*), e *Topic Supervised NMF*.

Fernandes et al. (2022) propõem uma metodologia para extrair valor das decisões do Tribunal brasileiro para apoiar juízes e advogados em suas tomadas de decisão. A metodologia utilizou um sistema de informação desenvolvido pelos autores que (i) extrai as ações judiciais e cada disposição específica sobre os pareceres emitidos pelos Tribunais de Primeira Instância e de Apelação, e (ii) conecta cada ação judicial com a disposição judicial correspondente. Os

corpora utilizados compreendem decisões de primeira e de segunda instâncias, nas quais é buscado um conjunto de ações e providões judiciais comumente julgadas pelo Tribunal. A metodologia é composta pela anotação de decisões do Tribunal, a criação de modelos de *Deep Learning* para extração de informações e a visualização das informações agregadas extraídas das decisões.

Sansone e Sperlí (2022) analisam o estado da arte das abordagens de inteligência artificial para o domínio jurídico, com foco em sistemas de Recuperação de Informação Jurídica (*Legal Information Retrieval*) baseados em técnicas de PLN, Aprendizado de Máquina e Extração de Conhecimento (*Knowledge Extraction*). Primeiro, os autores descrevem o fator de relevância da informação jurídica de acordo com seu volume, suas estruturas complexas e sua linguagem específica do domínio. Em segundo, os autores analisam as principais características de um sistema de Recuperação de Informação, fornecendo uma visão geral da recuperação de informações entre idiomas no domínio jurídico. E por fim, os autores analisam as técnicas baseadas em *Deep Learning* que, embora apresentem resultados promissores, normalmente utilizam modelo pré-treinado de *embeddings*, devido ao pequeno número de amostras de treinamento e à dificuldade de capturar as informações semânticas de documentos e termos legais específicos.

Castano et al. (2022) propõem o *framework* baseado em conhecimento CRIKE (*CRIME Knowledge Extraction*) concebido para apoiar a extração de conhecimento jurídico de uma coleção de documentos legais, com base em uma ontologia jurídica de referência chamada LATO (*Legal Abstract Term Ontology*). Apresentam o modelo de conhecimento LATO-KM (*LATO Knowledge Model*) aplicado ao CRIKE empregando (i) técnicas de extração, para detectar aplicações/interpretações de conceitos ao longo de documentos com base no conhecimento ontológico modelado no LATO, e (ii) técnicas de enriquecimento, para adicionar iterativamente o conhecimento terminológico à ontologia LATO a ser explorada para as etapas subsequentes de exploração/recuperação de informações.

D. Song. et al. (2022) apresentam o POSTURE50K, um novo conjunto de dados jurídicos de classificação multiclasse como resposta ao desafio da falta de conjuntos de dados de alta qualidade rotulados por humanos. O novo conjunto de dados contém 50.000 pareceres jurídicos e suas decisões processuais legais rotuladas manualmente. Os autores propõem, ainda, uma arquitetura de *Deep Learning* para classificação multiclasse de documentos, realizando um pré-treinamento adicional em conjuntos de dados específicos de domínio e adotando um mecanismo de atenção de *labels* para classificação multiclasse de documentos.

Chen et al. (2022) investigam a classificação de texto jurídico com uma grande coleção de documentos de casos norte-americanos rotulados comparando a eficácia de diferentes técnicas de classificação de texto. Os autores propõem um algoritmo de aprendizado de máquina usando conceitos de domínio como características e *random forest* como classificador, fornecendo uma referência para selecionar técnicas de aprendizado de máquina para construir sistemas de classificação de texto de alto desempenho no domínio jurídico ou em outras áreas.

Z. Song et al. (2022) apresentam um método de classificação multirrotulo de textos jurídicos com fusão de relações de rótulo (*multi-label classification method of legal texts with fusion of label relations* – MLC-FLR). Um modelo de rede de convolução de grafos é utilizado para capturar a relação de dependência entre rótulos construindo a matriz de coocorrência de rótulos. O mecanismo de atenção ao rótulo é empregado para calcular os graus de correlação entre um texto jurídico e cada palavra do rótulo, podendo-se obter a representação semântica do texto jurídico de um rótulo específico. Por fim, a representação abrangente de um texto para classificação multirrotulo é realizada, combinando a relação de dependência e a representação semântica do texto jurídico de um rótulo específico.

Sposito et al. (2022) descrevem uma proposta de extração de informações e identificação de datas e referências a entidades nomeadas, tais como N° Arquivo, N° Resolução, N° artigo de Lei, que se referem à norma legal vigente e são amplamente utilizadas em diversos documentos judiciais. Para o reconhecimento de tais entidades nomeadas, o processo emprega a definição de padrões utilizando expressões regulares, uma forma de representar uma linguagem de forma sintética, aplicando um conjunto de regras e os termos obtidos são armazenados em uma matriz de termos/documentos.

3.3. REDE DE COCORRÊNCIA DE PALAVRAS-CHAVE

Foi gerada a análise de coocorrência de palavras-chave, utilizando-se o arquivo de metadados gerado pelo portal Scopus. Os metadados foram lidos, sendo selecionada a opção de coocorrência de palavras-chave do autor. O objetivo da análise de redes é a obtenção de mapas conceituais para a análise de um domínio temático. A rede de coocorrência de palavras-chave mostra a estrutura cognitiva de um tema, identifica temas de pesquisa em ascensão, em declínio ou emergentes.

Um arquivo texto foi criado para realizar o controle do vocabulário (*Thesaurus*), permitindo a substituição e a exclusão de termos. Com o controle de vocabulário é possível fazer uma

interpretação qualitativa das palavras-chave nos metadados de cada documento, como, por exemplo, corrigir duplicidades ou erros ortográficos.

Para a primeira análise não foi utilizado o controle do vocabulário e a rede resultante, com um mínimo de três ocorrências para cada palavra-chave, resultou em uma rede de coocorrência com 220 termos, 9 comunidades e 1.213 arestas.

Na segunda análise, já com o controle do vocabulário, a nova rede resultante, com um mínimo de três ocorrências para cada palavra-chave, resultou em uma rede de coocorrência com 166 termos, 17 comunidades e 898 arestas.

Na sequência, foram feitos alguns ajustes na visualização e gerado o gráfico de densidade (mapa de calor). A Figura 1 apresenta o gráfico com a visualização do mapa de calor da rede de coocorrência de palavras-chave com controle de vocabulário. No centro do gráfico, as seguintes palavras-chave se destacam: *information retrieval*, *natural language processing*, *information extraction*, *legal ontology*, *legal information retrieval*, *deep learning*, entre outras. Entretanto, aparecem na periferia do gráfico alguns termos com baixa intensidade, mas que podem representar oportunidades de pesquisa: *intelligent justice*, *knowledge graph*, *legal question answering*, *ontology-based information text*, etc. Este gráfico permite verificar a aderência do estudo bibliométrico ao tema de pesquisa.

Após a geração dos gráficos, a rede foi exportada para serem calculadas as métricas de redes: grau médio, diâmetro da rede, modularidade, centralidade de autovetor e coeficiente de *clustering* médio. O grau médio representa o número de conexões que, em média, os nós de uma rede possuem. Esse valor pode ser útil para se comparar diferentes redes. Uma interpretação desta métrica é a popularidade.

O diâmetro de uma rede representa a maior distância existente entre dois nós nessa rede. Por distância, entende-se o número de conexões intermediárias existentes entre esses nós. É uma métrica para comparar redes e mostrar o grau de articulação existente entre seus nós. Determina o número de triângulos de cada nó.

A modularidade mensura a força da divisão da rede em módulos (também chamados de grupos, *clusters* ou comunidades). Redes com alta modularidade possuem conexões densas entre os nós das comunidades e baixa densidade entre os nós de diferentes comunidades. É uma medida de vizinhança, considerando o quanto um nó tende a aparecer em determinado grupo.

A ideia básica da métrica centralidade de autovetor é observar os nós vizinhos e avaliar quão conectados eles são. Nós bem conectados (alta centralidade de grau) a nós também bem conectados possuem maior centralidade de autovetor. Uma interpretação desta métrica é a influência.

4. ABORDAGEM METODOLÓGICA

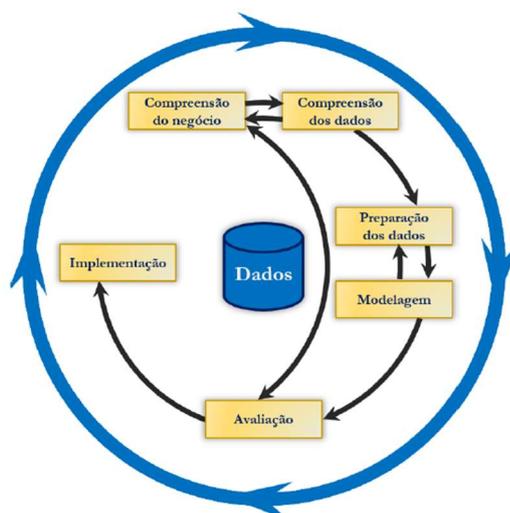
4.1. METODOLOGIA EMPREGADA

Para Fayyad, Piatetsky-Shapiro e Smyth (1996), o processo de Descoberta de Conhecimento em Bancos de Dados (DCBD) envolve a preparação dados, a geração de modelos e sua validação. Mineração de Dados (MD) concerne à etapa específica de aplicação de algoritmos para extrair padrões nos dados e geração de modelos.

Uma das subáreas de MD é a Mineração de Texto (MT) ou descoberta de conhecimento a partir de texto que, segundo Hotho, Nürnberger e Paaß (2005), trata da análise textual apoiada pela máquina. A MT utiliza técnicas de recuperação de informações, de extração de informações, bem como de PLN e as relaciona aos algoritmos e métodos de DCBD, MD, aprendizado de máquina e estatística. Dessa forma, tem-se um procedimento semelhante ao do processo DCBD, em que o foco da análise não são os dados em geral, mas os documentos de texto.

Uma metodologia amplamente utilizada em projetos de MD é o *Cross Industry Standard Process for Data Mining* (CRISP-DM). O CRISP-DM é uma metodologia utilizada para apoiar o processo de planejamento e execução de um projeto de mineração de dados, de modo flexível. Foi desenvolvido pelas empresas Daimler Chrysler, SPSS e NCR em 2000 (Chapman et al., 2000). A metodologia é flexível para mostrar que o processo não ocorre de maneira sequencial, mas sim de forma cíclica, em que pode ocorrer o retorno a um estágio, de modo a chegar a um resultado melhor e sempre focado no entendimento do negócio. A Figura 3 sintetiza esse processo.

Figura 3: Os estágios do CRISP-DM.



Os seis estágios do CRISP-DM são:

1. Compreensão do Negócio: Fase inicial do processo, focada nos problemas, objetivos e especificações com uma perspectiva de negócio. A partir desse conhecimento é definido um problema de Ciência dos Dados para ser solucionado. É de suma importância, pois o sucesso da solução depende da importância e elaboração do problema. Apesar de ser no início, essa visão deve acompanhar o processo, e é um dos principais motivos dele ser cíclico, pois as fases não podem perder o foco no objetivo final.
2. Compreensão dos Dados: Os dados são a matéria-prima para a solução. Nessa fase é analisada a qualidade dos dados segundo o problema, os dados são explorados, os primeiros *insights* são gerados e as hipóteses formuladas.
3. Preparação dos Dados: Consiste em gerar um conjunto de dados final a partir dos dados não tratados. O que inclui etapas como a seleção dos dados (instâncias e atributos), integração, limpeza e transformação.
4. Modelagem: É nessa fase em que as técnicas de Mineração de Dados são aplicadas, com intuito de descobrir padrões nos dados. Diversas técnicas são utilizadas e a preparação dos dados tem forte influência no resultado obtido, por isso, muitas vezes é necessário voltar uma etapa.
5. Avaliação: O objetivo dessa fase é estimar os resultados do modelo construído anteriormente, e os principais focos são escolher o melhor modelo e validar se os resultados atendem as expectativas. A avaliação tenta ser o mais próximo possível da realidade, para isso é necessário cuidado na seleção dos dados que serão utilizados no teste e nas métricas utilizadas.
6. Implantação: Geralmente o modelo gerado não é o final do projeto. Mas sim, parte dele, quando são colocados em uso real. Essa etapa não é necessariamente realizada por um cientista de dados.

Para o presente trabalho, a metodologia CRISP-DM será customizada e utilizada da forma apresentada na Tabela 2, visando à mineração de textos.

Tabela 2: Metodologia CRISP-DM adaptada a este trabalho de pesquisa

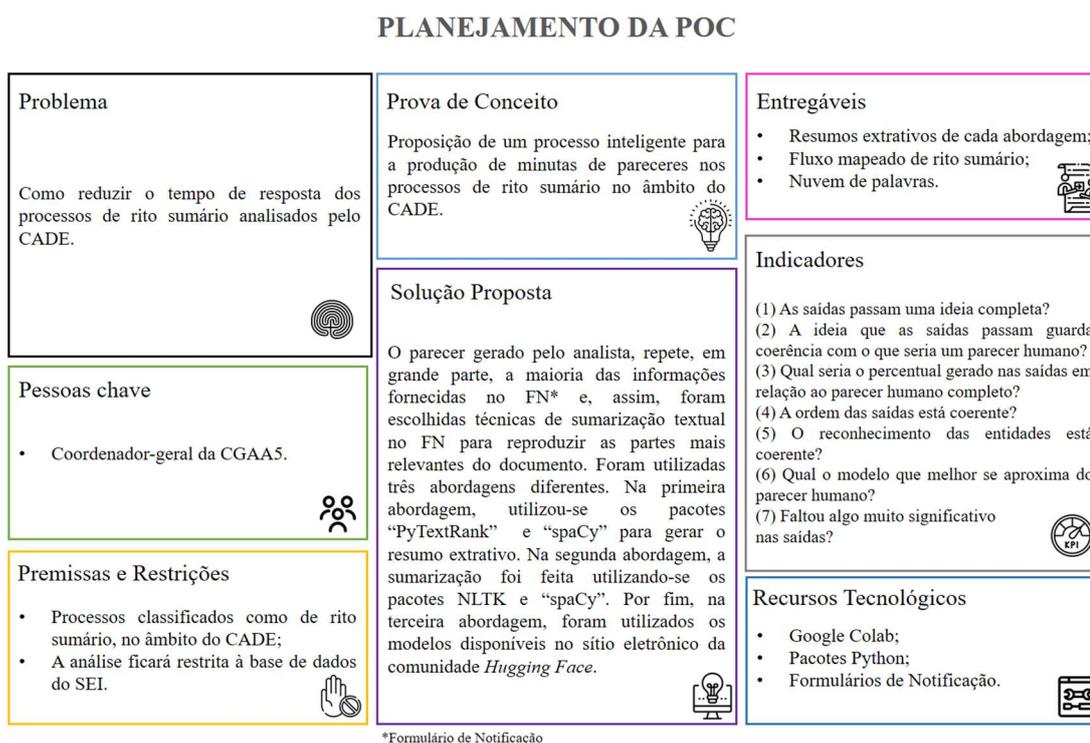
Fase	Atividades
Compreensão do Negócio	Fazer um estágio na área de negócio para compreensão das necessidades e entendimento do problema de negócio.
Compreensão dos Dados	Analisar a qualidade dos dados conforme o problema, explorar os dados, gerar os primeiros <i>insights</i> e formular as hipóteses. Nesse caso específico, os dados serão os documentos em PDF dos formulários de notificação.
Preparação dos Dados	Preparar os documentos em PDF, o que inclui gerar um documento PDF livre de formatação, livre das partes trazidas do Anexo II da Resolução n° 33/2022 e dos títulos e subtítulos.
Modelagem	Aplicar os algoritmos de sumarização e escolher o(s) melhor(es) para a construção da prova de conceito.
Avaliação	Verificar os indicadores definidos com os responsáveis e validar se os resultados atendem às expectativas, como prova de conceito.
Implantação	Este projeto de pesquisa não fará a implantação da solução, pois será ainda uma prova de conceito.

Fonte: Autoria Própria

4.2. ETAPAS DA PESQUISA

O produto final deste projeto de pesquisa é desenvolver uma Prova de Conceito (POC – *Proof of Concept*) para validar o processo inteligente de produção de minutas de pareceres de rito sumário no âmbito do CADE. O planejamento da POC é mostrado na Figura 4.

Figura 4: Planejamento da POC



Fonte: Autoria própria

O problema de pesquisa é reduzir o tempo de resposta (tempo de análise) nos processos de atos de concentração econômica de rito sumário no âmbito do CADE. A POC tratará da utilização das técnicas de IA para propor um processo inteligente para a produção de minutas de pareceres nos processos de rito sumário.

O CADE foi criado pela Lei nº 4.137/1962 como um órgão do Ministério da Justiça e transformado em autarquia pela Lei nº 8.884/1994. Em maio de 2012, com a entrada em vigor da nova Lei de Defesa da Concorrência (Lei nº 12.529/2011), o CADE passou a ser uma autoridade de defesa da concorrência responsável por instruir os processos administrativos de apuração de infrações à ordem econômica, assim como os processos de análise de atos de concentração. Sua responsabilidade é investigar e punir administrativamente, em instância única, pessoas físicas e jurídicas que pratiquem infrações à ordem econômica. Além disso, o CADE também analisa atos de concentração, de modo a minimizar possíveis efeitos negativos no ambiente concorrencial de determinado mercado.

Entende-se como ato de concentração, segundo Silveira (2021), toda fusão ou aquisição de empresas, que pode ocorrer pela obtenção do controle societário, de influência nos negócios de concorrentes e pela compra de ativos relevantes, além das *joint ventures*¹⁰, os consórcios e os contratos associativos.

Os atos de concentração são de notificação obrigatória ao CADE para serem analisados. Esses atos estão descritos no art. 88 da Lei nº 12.529/2011. Ademais, o CADE é uma entidade judicante, pois julga condutas nocivas à livre concorrência e cria jurisprudência nas matérias de sua competência, porém atua como um tribunal administrativo e não como um tribunal judiciário. O Tribunal Administrativo é o órgão judicante do CADE e suas atribuições estão previstas no art. 10 da Lei 12.529/2011. O Tribunal é composto por um presidente e seis conselheiros, de notório saber jurídico, para mandatos de quatro anos. Os conselheiros são os responsáveis pelo voto nas decisões judiciais do CADE e o voto do presidente é destinado a pôr fim a eventuais empates.

Sendo assim, as decisões administrativas do CADE, em sua área de competência, têm o peso de decisão judicial, visto que o CADE possui prerrogativa legal para decidir mesmo sendo um tribunal administrativo.

Uma das formas de se oferecer um serviço de qualidade é buscar uma diminuição do tempo de resposta dos atos e das consultas pelo rito sumário, procedimentos considerados mais simples. A tecnologia de IA na área judicante, mais especificamente de PLN, traz resultados

¹⁰ https://www.ipea.gov.br/desafios/index.php?option=com_content&id=2110:catid=28&Itemid=. Acesso em 11 jan. 2023

promissores tanto na rapidez como na precisão das decisões. Dessa forma, é possível que o CADE possa se beneficiar desta tecnologia para reduzir o tempo das decisões e permitir, em consequência, uma maior produtividade de seus analistas.

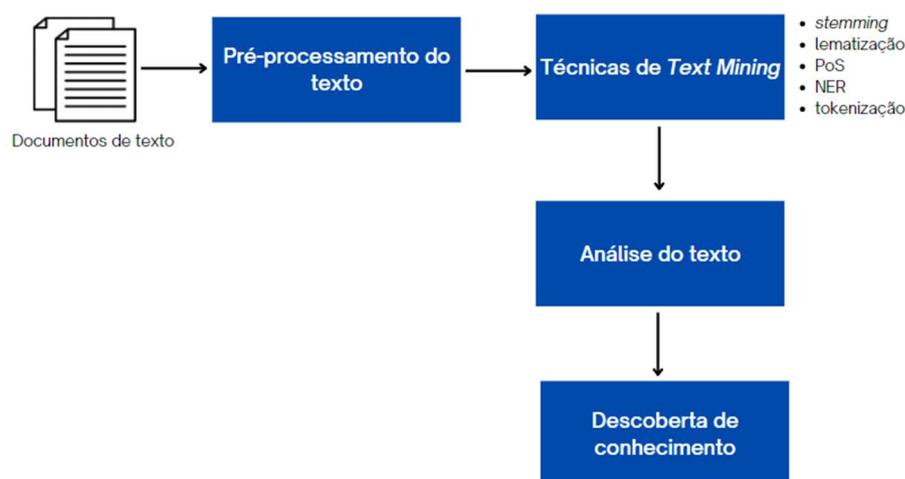
Desde 2021, o CADE já utiliza PLN em seu Sistema de Busca de Jurisprudência¹¹, que reúne seis coleções de conhecimento: jurisprudência, pareceres técnicos, legislação, guias e publicações, acórdãos e notícias.

Conforme Hotho, Nürnberger e Paaß (2005), mineração de textos (*Text Mining* – TM) é a aplicação de algoritmos e métodos das áreas de *machine learning* e estatística para descobrir padrões válidos em bases textuais.

Textos fazem parte da categoria de dados não-estruturados. Isso significa que, antes de aplicar os algoritmos e métodos, é necessário estruturá-los. Isto é, realizar algum procedimento que transforme a sequência de caracteres em objetos relacionados entre si, como um objeto matemático, que geralmente é uma forma vetorial.

O processamento de linguagem natural (PLN) é uma técnica chave para mineração de textos. Ainda segundo Hotho, Nürnberger e Paaß (2005), a extração de informações se divide em uma série de etapas de processamento, geralmente abrangendo tokenização, segmentação de sentença, atribuição de *part of speech* (PoS) e reconhecimento de entidades nomeadas, i.e., nomes de pessoas, de locais e de organizações. Em um nível mais alto, frases e sentenças devem ser analisadas, interpretadas semanticamente e integradas. A Figura 5 mostra as etapas de processamento do TM.

Figura 5: Etapas do processamento de *Text Mining* para a sumarização de texto



Fonte: Autoria própria

¹¹ <https://jurisprudencia.cade.gov.br/pesquisa>. Acesso em 1 jun. 2022.

Como solução proposta, foram utilizadas três abordagens diferentes. Na primeira abordagem, utilizou-se as ferramentas de software PyTextRank e spaCy para gerar o resumo extrativo. Na segunda abordagem, a sumarização foi feita utilizando-se as ferramentas NLTK e spaCy. Por fim, na terceira abordagem, foram utilizados os modelos disponíveis no sítio eletrônico da comunidade *Hugging Face*. Como o parecer gerado pelo analista, repete, em grande parte, a maioria das informações fornecidas no Formulário de Notificação (FN), foram escolhidas técnicas de sumarização textual no FN para reproduzir as partes mais relevantes do documento.

As premissas e restrições da POC são duas: (i) serão utilizados apenas os processos classificados como de rito sumário, no âmbito do CADE; e (ii) a análise ficará restrita à base de dados do SEI, onde estão os FNs.

Os recursos utilizados na POC são: (i) Google Colab; (ii) a linguagem *Python*; e (iii) Formulários de Notificação.

A POC se propõe a gerar os seguintes entregáveis: (i) resumos extrativos de cada abordagem; (ii) fluxo mapeado de rito sumário; e (iii) nuvem de palavras.

Por fim, para validar a POC foram propostos alguns indicadores a serem ratificados pelo coordenador-geral da CGAA5, que é a pessoa responsável pela área que faz a triagem dos processos de AC. Os indicadores propostos são:

1. As saídas passam uma ideia completa?
2. A ideia que as saídas passam guarda coerência com o que seria um parecer humano?
3. Qual seria o percentual gerado nas saídas em relação ao parecer humano completo?
4. A ordem das saídas está coerente?
5. O reconhecimento das entidades está coerente?
6. Qual o modelo que melhor se aproxima do parecer humano?
7. Faltou algo muito significativo nas saídas?

5. FERRAMENTAS UTILIZADAS

São apresentadas a seguir o conjunto de ferramentas utilizadas para sumarização, identificação e visualização de entidades nomeadas, e geração de nuvens de palavras.

Sumarização com *spaCy*, *PyTextRank* E *summary*

A ferramenta *spaCy* é uma biblioteca gratuita e de código aberto, desenvolvido na linguagem *Python*, para PLN avançado. Foi projetado especificamente para processar e “entender” grandes volumes de texto e pode ser usado para extrair informações de textos escritos em linguagem natural, ou para pré-processar textos para *deep learning*.

PyTextRank é uma extensão *spaCy* usada para extrair as frases mais bem classificadas e realizar a sumarização extrativa de documentos de texto. Para a sumarização extrativa é utilizado o método *summary*, que se baseia na distância vetorial para identificar cada uma das sentenças melhor classificadas. Como parâmetros, foram utilizados o *limit_phrases* que é o número máximo de sentenças melhor classificadas para usar nos vetores de distância, o *limit_sentences*, o número total de sentenças a serem geradas para a sumarização extrativa e o *preserve_order*, que indica se é para preservar a ordem das sentenças conforme elas ocorreram originalmente no texto fonte, sendo que o *default* é falso.

Sumarização com *NLTK*, *heapq* E *nlargest*

A ferramenta *NLTK* (*Natural Language ToolKit*) gera um resumo extrativo com base no TF-IDF e no ranqueamento das sentenças a partir de um pré-processamento envolvendo tokenização, classificação, limpeza de *stopwords* e pontuação, *tagging*, *stemming* e lematização). Para gerar o resumo extrativo, pode-se escolher o percentual das sentenças melhor classificadas utilizando-se a função *nlargest* da ferramenta *heapq*¹², que retorna uma lista com os *n* maiores elementos de um conjunto de dados.

Sumarização com *Transformers* E *T5Tokenizer*

Transformers é uma arquitetura de rede neural, introduzida pelo Google¹³, que, ao invés

¹² <https://docs.python.org/3/library/heapq.html>.

¹³ <https://www.datacamp.com/tutorial/an-introduction-to-using-transformers-and-hugging-face>.

de processar os dados em partes separadas, como era feito em outras arquiteturas de redes neurais, processa sequências de entrada inteiras de uma só vez. Além disso, fornece ferramentas para manipular modelos pré-treinados. Esses modelos suportam tarefas comuns em diferentes aplicações em PLN, como classificação de texto, reconhecimento de entidades nomeadas, perguntas e respostas, modelagem de linguagem, sumarização, tradução e geração de texto.

O *Tokenizer* é o tradutor de um texto em linguagem natural para *tokens* inteligíveis por máquinas. O T5¹⁴ (ou *Text-to-Text Transfer Transformer*) é uma arquitetura baseada em *Transformers* que usa uma abordagem de texto para texto. Cada tarefa de PLN, incluindo sumarização, é projetada para fornecer o texto do modelo como entrada e treiná-lo para gerar um texto de saída.

Reconhecimento e visualização de entidades nomeadas com *spaCy* E *displaCy*¹⁵

displaCy é uma funcionalidade da ferramenta *spaCy* que permite reconhecer as entidades nomeadas de um determinado texto, rotulando-as e destacando-as com cores distintas.

Geração de nuvens de palavras com *wordcloud*¹⁶ E *WordCloud*

Uma nuvem de palavras é uma representação visual de dados em forma textual, baseada na frequência de ocorrência de cada palavra no texto. A frequência com que cada palavra aparece é computada e utilizada para definir seus tamanhos e cores, evidenciando sua importância relativa no texto.

¹⁴ https://huggingface.co/docs/transformers/model_doc/t5.

¹⁵ <https://spacy.io/usage/visualizers>.

¹⁶ <https://pypi.org/project/wordcloud/>.

6. RESULTADOS

6.1. RESULTADOS OBTIDOS

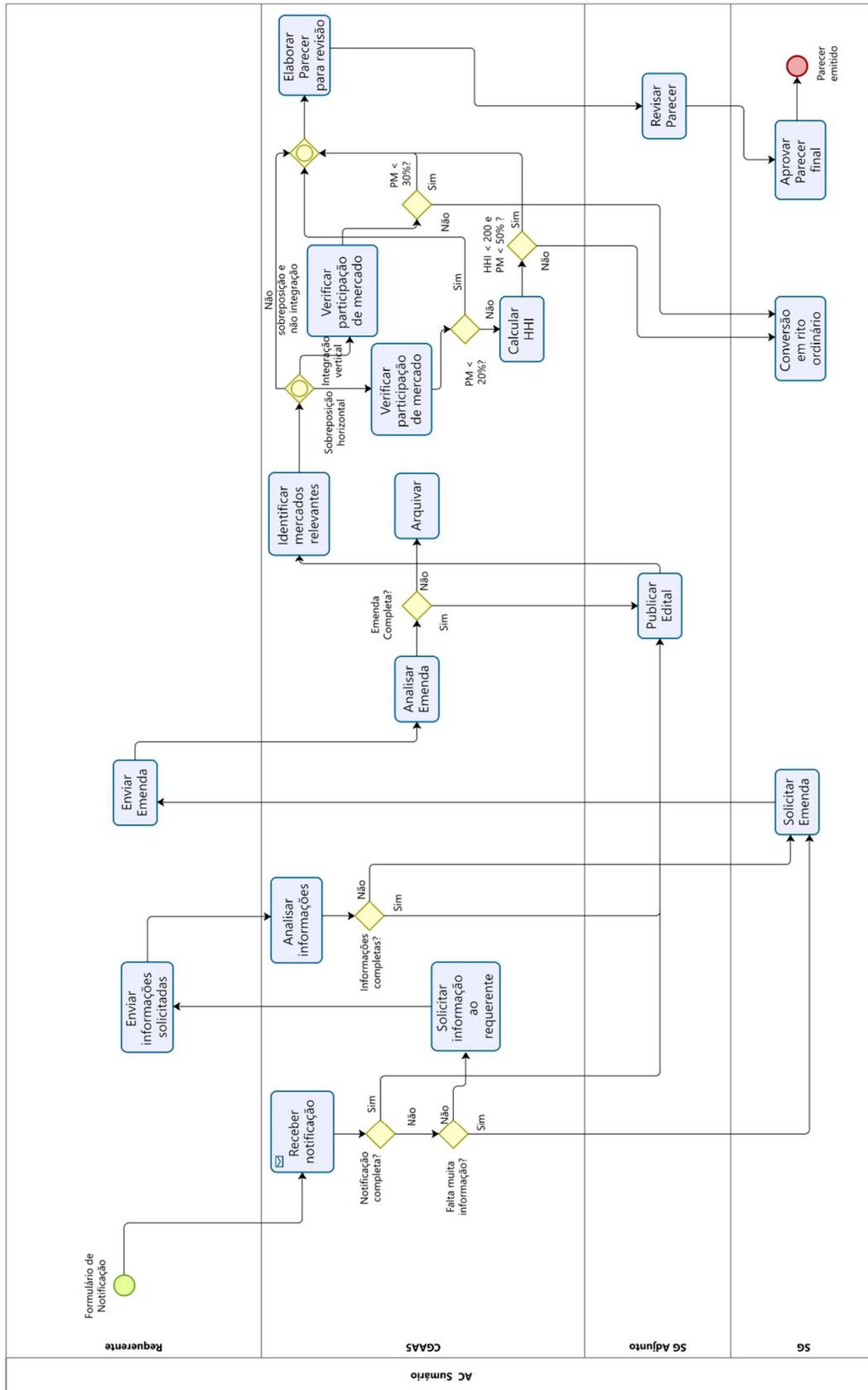
6.1.1. Compreensão do negócio

A Lei nº 12.529/2011 introduziu o controle prévio de atos de concentração, considerado pelos especialistas a mais desafiadora inovação do novo marco legal, pois existiam dúvidas sobre se “o tempo de análise seria célere, e suficientemente eficaz, em comparação ao tempo econômico das empresas” (CADE, 2013). Segundo Silva e Souza (2016), o novo marco legal possibilitou que o CADE respondesse de forma mais ágil e célere às demandas do ambiente de negócios.

Outra inovação, como consequência do controle prévio, foi a criação de um departamento de triagem no CADE, que identifica quais atos de concentração serão analisados pelo rito sumário e quais deverão seguir o rito ordinário. Dessa forma, os casos mais simples (de rito sumário) são analisados por uma área diferente da que analisará os casos mais complicados (de rito ordinário), não deixando os casos mais simples represados, caso fossem tratados pela mesma área.

Para aplicar uma melhoria no fluxo de rito sumário utilizando PLN para geração de uma minuta do parecer decisório e, com isso, reduzir o tempo de análise, foi realizado um estágio de três meses (dezembro/2021 a fevereiro/2022) na área de negócio para compreensão das necessidades e entendimento do problema de negócio. Como resultado desse tempo de estágio, foi mapeado todo o fluxo da análise de rito sumário, conforme Figura 6.

Figura 6: Fluxo da análise de rito sumário



Fonte: Autoria própria com o aval da Superintendência-Geral do CADE

Conforme mostrado, o procedimento se inicia com o recebimento do Formulário de Notificação (FN) pela Coordenação Geral de Análise Antitruste 5 (CGAA5) enviado pelas Requerentes (empresas envolvidas no ato de concentração). A CGAA5 verifica se o FN está completo. Se não estiver completo e com poucas informações faltantes, a CGAA5 solicita informações complementares para as Requerentes.

Se as informações não estiverem completas e faltarem muitas informações ou se as Requerentes não enviarem as informações adicionais solicitadas pela CGAA5, o Superintendente Geral (SG) publica uma emenda no Diário Oficial da União (DOU) solicitando as informações formalmente. A emenda suspende o prazo de análise até a resposta pelas Requerentes.

De posse das informações completas, o Superintendente Geral Adjunto (SG Adjunto) publica o edital no DOU, que marca o início do prazo de análise. Se as informações não forem satisfeitas pela emenda, o processo é arquivado. Nesse caso, se as Requerentes quiserem continuar o processo, deverão pagar uma nova taxa e enviar um novo FN instituindo, assim, um novo processo de análise.

Após a publicação do edital, a análise do ato de concentração começa pela identificação dos mercados relevantes. Em seguida, é verificado se há sobreposição horizontal e/ou integração vertical no ato de concentração em análise. Conforme Pereira Neto e Casagrande (2016), a sobreposição horizontal acontece quando as empresas envolvidas no ato de concentração atuam no mesmo mercado relevante e a integração vertical acontece entre agentes econômicos atuando em mercados adjacentes em uma mesma cadeia produtiva.

Se houver sobreposição horizontal e a participação de mercado for maior ou igual a 20% ou a participação de mercado for maior que 50% e o HHI¹⁷ for maior que 200 e/ou houver integração vertical com participação de mercado maior ou igual a 30%, o SG converterá o processo de análise em rito ordinário.

Se houver sobreposição horizontal e a participação de mercado for menor que 20%, ou se a participação de mercado for maior que 20%, mas o HHI for menor que 200 e/ou houver

¹⁷ Segundo o Guia para Análise de Atos de Concentração Horizontal, “o Índice Herfindahl-Hirschman (HHI) pode ser utilizado para o cálculo do grau de concentração dos mercados. O HHI é calculado com base no somatório do quadrado das participações de mercado de todas as empresas de um dado mercado. O HHI pode chegar até 10.000 pontos, valor no qual há um monopólio, ou seja, em que uma única empresa possua 100% do mercado. Compreende-se que os mercados são: (i) Mercados não concentrados: com HHI abaixo de 1500 pontos; (ii) Mercados moderadamente concentrados: com HHI entre 1.500 e 2.500 pontos; e (iii) Mercados altamente concentrados: com HHI acima de 2.500.”. Disponível em <https://www.gov.br/cade/pt-br/centrais-de-conteudo/publicacoes-institucionais/guias-do-cade>. Acesso em 16 jan. 2023.

integração vertical com participação de mercado menor que 30% ou não houver nem sobreposição horizontal e nem integração vertical, a CGAA5 elabora o parecer para revisão pelo SG Adjunto.

O SG Adjunto revisa o parecer e o envia para aprovação final pelo SG. Uma vez aprovado, o parecer final é emitido, assinado e o processo de análise termina.

Com a introdução do processo inteligente de geração de pareceres, o fluxo de rito sumário ganhará um novo ator (Processo Inteligente) e novos processos após a publicação do edital de início da análise. Esses novos processos incluem a geração automática de pareceres, a análise desse parecer, e a verificação de que todos os cenários do mercado relevante para o ato de concentração foram contemplados.

6.1.2. Compreensão dos dados

Os dados trabalhados neste projeto são não estruturados, visto que a fonte principal dos dados é o Formulário de Notificação (FN) dos atos de concentração de rito sumário, recebido pelo CADE no formato PDF.

Toda a análise de ato de concentração pelo CADE se inicia com o recebimento do FN. Os textos em PDF, de forma geral, reproduzem o Anexo II da Resolução CADE nº 33, de 14 de abril de 2022 (Anexo A), sendo preenchidos com as informações solicitadas sem nenhuma padronização/estruturação.

O parecer gerado pelo analista, repete, em grande parte, a maioria das informações fornecidas no FN e, assim, foram escolhidas técnicas de sumarização textual no FN para reproduzir as partes mais relevantes do documento.

De acordo com Trivedi et al. (2021), existem duas abordagens de sumarização textual, conforme mostrado na Figura 7: a extrativa, que identifica as sentenças mais importantes de um documento e as extrai como elas são, e a abstrativa, que se utiliza dos conceitos principais do documento e também do conhecimento do domínio para gerar um texto diferente do original, mas com o mesmo sentido.

Figura 7: Abordagens de sumarização de texto



Fonte: Autoria própria

Dessa forma, uma primeira minuta do parecer gerado por máquina se baseará na sumarização do FN, utilizando a abordagem extrativa.

Por fim, a fonte de dados onde se encontram os FNs em PDF é o Sistema Eletrônico de Informações (SEI), o sistema que realiza a administração dos processos eletrônicos no CADE.

6.1.3. Preparação dos dados

A fase de preparação dos dados pode ser considerada um ponto crucial do projeto de pesquisa, pois é nessa fase que os dados serão preparados para testar os modelos que resultam na minuta de parecer esperada.

Os textos utilizados para a pesquisa são reais, mas foi utilizada a versão pública dos documentos para fins de confidencialidade e sigilo.

Em um primeiro momento, foi utilizado o documento na forma em que é enviado ao CADE com todos os cabeçalhos da Resolução nº 33/2022 (Anexo A), títulos e formatação (elementos não relevantes). Identificou-se que o resultado não foi satisfatório, pois os elementos não relevantes influenciam na extração das informações mais importantes do texto.

Sendo assim, foi gerada uma versão do FN sem esses elementos não relevantes (Anexo B), sendo observada uma melhora significativa na qualidade das sumarizações. A geração desse texto “limpo” exigiu um esforço de retirada, do texto em PDF, das partes textuais relevantes para a sumarização. Esse problema será resolvido com o novo projeto do Formulário Digital de Notificação de AC, que se encontra em desenvolvimento, e permitirá que as informações sejam estruturadas em banco de dados e utilizadas apenas concatenando as informações relevantes em um único texto.

6.1.4. Modelagem

O texto limpo foi utilizado em três abordagens diferentes, cada uma com algoritmos diferentes selecionados para avaliação como POC. Na primeira, utilizaram-se as ferramentas *PyTextRank*¹⁸ e *spaCy*¹⁹ para gerar o resumo extrativo. As duas ferramentas, utilizados em con-

¹⁸ <https://pypi.org/project/pytextrank/2.0.0/#:~:text=PyTextRank%20is%20a%20Python%20implementation,extractive%20summarization%20of%20text%20documents>. Acesso em 12 dez. 2022.

¹⁹ <https://explosion.ai/blog/spacy-v2-pipelines-extensions>. Acesso em 12 dez. 2022.

junto, fornecem uma extração rápida e eficaz de frases de textos. O algoritmo funciona independente de uma linguagem natural específica e não requer conhecimento de domínio e nem de pré-processamento.

Para fazer a sumarização com essa primeira abordagem, foram utilizados os seguintes parâmetros na função *summary*:

```
doc._.textrank.summary(limit_phrases=1000, limit_sentences=25,
preserve_order=True)
```

O resultado obtido é apresentado no Quadro 3.

Quadro 3: Sumarização obtida com as ferramentas *PyTextRank* e *spaCy*

Trata -se de aquisição, pela Termogás S.A. (" Termogás "), de participações societárias nas distribuidoras locais de gás natural: (i) Companhia de Gás do Amapá - GASAP ("GASAP "), (ii) Companhia de Gás do Piauí - GASPIISA (" GASPIISA "), (iii) (" GOIASGÁS ") e (iv) Companhia Rondoniense de Gás - RONGÁS (" RONGÁS " e em conjunto com GASAP, GAPISA e GOIASGÁS " Empresas -Alvo "), atualmente detidas pela Commit Gás S.A. (" Commit "), atual denominação da Petrobrás Gás S.A. - Gaspetro (" Gaspetro "), empresa controlada pela Compass Gás e Energia S.A. (" Compass ") (a " Operação ").

A Commit é uma holding de distribuição de gás natural com participações em diversas distribuidoras de gás natural no Brasil.

O fechamento da referida operação ocorreu em 11 de julho de 2022, com a efetiva transferência de 51% das ações da Commit antes detidas pela Petrobras para a Compass, mesma data em que houve a formalização da mudança da denominação social de Petrobras Gás S.A. - Gaspetro para Commit Gás S.A.

A Compass é uma sociedade do Grupo Cosan que tem operações em comercialização de energia elétrica e na distribuição de gás natural por meio da Companhia de Gás de São Paulo - Comgás ("Comgás") e Companhia de Gás do Estado do Rio Grande do Sul - Sulgás (" Sulgás "), além de outras atividades em estágio pré-operacional ou de projeto.

Com a aquisição da Commit, a Compass passou a ser titular de participação societária indireta em sociedades que atuam na distribuição de gás canalizado em diversos Estados do Brasil, incluindo as Empresas -Alvo. As Empresas -Alvo são sociedades de economia mista, criadas através de legislações estaduais, cuja principal função é explorar, com exclusividade, o serviço público de distribuição de gás canalizado, podendo também explorar outras formas de distribuição de gás natural ou manufaturado nos estados nos quais foram constituídas.

A Operação será realizada no segmento de distribuição de gás natural canalizado nos estados do Amapá, Piauí, Goiás e Rondônia.

Assim não há que se falar em concorrência durante a vigência da concessão do Estado para a prestação do serviço de distribuição de gás natural via dutos.

Trata -se de aquisição, pela Termogás, de participações societárias nas distribuidoras locais de gás natural Companhia de Gás do Amapá - GASAP, Companhia de Gás do Piauí - GASPIISA, Agência Goiana de Gás Canalizado - GOIASGÁS e Companhia Rondoniense de Gás - RONGÁS, atualmente detidas pela Commit.A Operação representa, para a Termogás, a oportunidade de consolidar a sua participação em distribuidoras locais de gás natural e, assim, poder colaborar para que elas possam iniciar a prestação do serviço público de distribuição de gás canalizado e atividades correlatas.

Além das Empresas -Alvo, o Grupo Termogás, através da Termogás, possui participação societária nas seguintes empresas operacionais atuantes na distribuição de gás natural: (i) Companhia Maranhense de Gás - GASMAR, (ii) Companhia de Gás do Amazonas - CIGÁS 12 e (iii) Companhia Brasileira de Gás - CEBGÁS. A companhia tem por objetivo a exploração, com exclusividade, do serviço público de distribuição de gás canalizado, podendo também explorar outras formas de distribuição de gás natural ou manufaturado, de produção no estado do Amapá.

Ela é uma empresa de capital misto que tem por objeto social a exploração e distribuição de gás natural canalizado, e de outras origens, por vias terrestres e fluviais; A Commit é uma holding que detém participação em distribuidoras de gás natural em diversos Estados brasileiros, quais sejam, além das Empresas - Alvo, Rio Grande do Sul (Sulgás), Florianópolis (Scgás), Paraná (Compagás), Mato Grosso do Sul (Msgás), São Paulo (GasBrasiliiano), Rio de Janeiro (CEG Rio), Alagoas (Algás), Ceará (Cegás), Rio Grande do Norte (Potigás), Sergipe (Sergás) e Pernambuco (Copergás). A Compass tem operações nacionalmente em comercialização de energia elétrica e na prestação de serviços locais de gás canalizado em parte do território do Estado de SP por meio da Comgás e no Rio Grande do Sul, por meio da Sulgás, além de outras atividades em estágio pré-operacional ou de projeto.

Conforme mencionado acima, a Operação se refere a aquisição de participações societárias nas distribuidoras locais de gás natural GASAP, GASPISA, GOIASCÁS e RONGÁS, atualmente detidas pela Commit.

A Operação diz respeito, portanto, às atividades de distribuição de gás natural nos Estados do Amapá, Piauí, Goiás e de Rondônia.

Não obstante a ausência efeitos concorrenciais da Operação, para fins completude e com o intuito de apresentar as melhores informações disponíveis, as Partes fornecem a seguir uma breve explicação sobre a cadeia do gás natural e a jurisprudência deste E. Cade a respeito das atividades de distribuição do produto. Destaca-se que a Superintendência -Geral do CADE ("SG") já analisou o mercado downstream de distribuição de gás natural em diversas ocasiões 14, afirmando que não há necessidade de subdivisão do mercado relevante na sua dimensão produto - sendo este o gás natural - contudo, reconhecendo que existem diferenças entre os modais de distribuição do produto que pode ser feita de duas formas: por meio de dutos locais, de média e baixa pressão, conectados diretamente aos dutos de transporte; ou a granel, através de veículos como trens, navios e caminhões dotados de tanques onde o gás é armazenado e transportado até a destinação final. As Partes entendem que o mesmo raciocínio se aplica à atividade de distribuição de gás natural, ou seja, dado que esta é uma atividade caracterizada pelo monopólio natural, não há concorrência no mercado, que é restrito à área de alcance da malha dutoviária. Como informado, a Operação não gera qualquer sobreposição haja vista que as Empresas -Alvo não são ativas. Termogás não atua na distribuição de gás natural através de outros modais nos Estados nos quais as Empresas -Alvo foram constituídas e estão autorizadas a operar, quais sejam Amapá, Piauí, Rondônia e Goiás.

Em vista de todo o exposto, resta clara a ausência de sobreposição horizontal decorrente da Operação e a baixa participação das distribuidoras nas quais a Termogás é acionistas na demanda nacional por gás natural. Em primeiro lugar, nota-se que com exceção da participação nas Empresas -Alvo, nem a Termogás, tampouco o Grupo Termogás, atuam na cadeia do gás natural nos Estados do Amapá, Piauí, Goiás e Rondônia.

Portanto, resta claro que os demais investimentos do Grupo Termogás na cadeia de gás não resultam em qualquer integração vertical com as atividade de distribuição de gás via dutos as quais as Empresas -Alvo estão autorizadas a realizar.

A ferramenta *spaCy* possui diversas funções para PLN. Uma delas é o reconhecimento de entidades nomeadas que pode ser apresentada de forma visual utilizando-se a função *display*. Uma parte do texto com as entidades identificadas é apresentada na Figura 8. Observe-se que foram identificadas entidades como "ORG" (Organização), "LOC" (Localização) e "MISC" (Miscelânea), que indica que o algoritmo não conseguiu identificar uma classe exata para a entidade.

Figura 8: Reconhecimento de entidades nomeadas

A Companhia de Gás do Amapá – GASAP ORG é uma sociedade de economia mista, autorizada pela Lei Estadual MISC nº 705 de 05 de julho de 2002. A companhia tem por objetivo a exploração, com exclusividade, do serviço público de distribuição de gás canalizado, podendo também explorar outras formas de distribuição de gás natural ou manufaturado, de produção no estado do Amapá LOC. A companhia não é operacional.

Companhia de Gás do Piauí ORG – GASPISA LOC é uma sociedade de economia mista, criada por deliberação do Governo do Estado do Piauí LOC através da Lei Estadual MISC nº 5.192, de 25 de maio de 2001. A companhia detém os direitos de exploração, com exclusividade, do serviço de distribuição de gás canalizado, podendo também explorar outras formas de distribuição de gás natural e manufaturado, em todo o território do Estado do Piauí LOC. Não obstante, a companhia não é operacional.

A Agência Goiana de Gás Canalizado S/A – GOIASGÁS ORG, foi criada por meio da Lei Estadual MISC nº 13.641/2000, para a valorização e preservação de nossas riquezas naturais. Ela é uma empresa de capital

Na segunda abordagem, a sumarização foi feita utilizando-se as ferramentas NLTK (*Natural Language ToolKit*) e *spaCy*. A sumarização na segunda abordagem foi feita utilizando-se 15% de percentual das sentenças mais bem classificadas e o resultado é apresentado no Quadro 4.

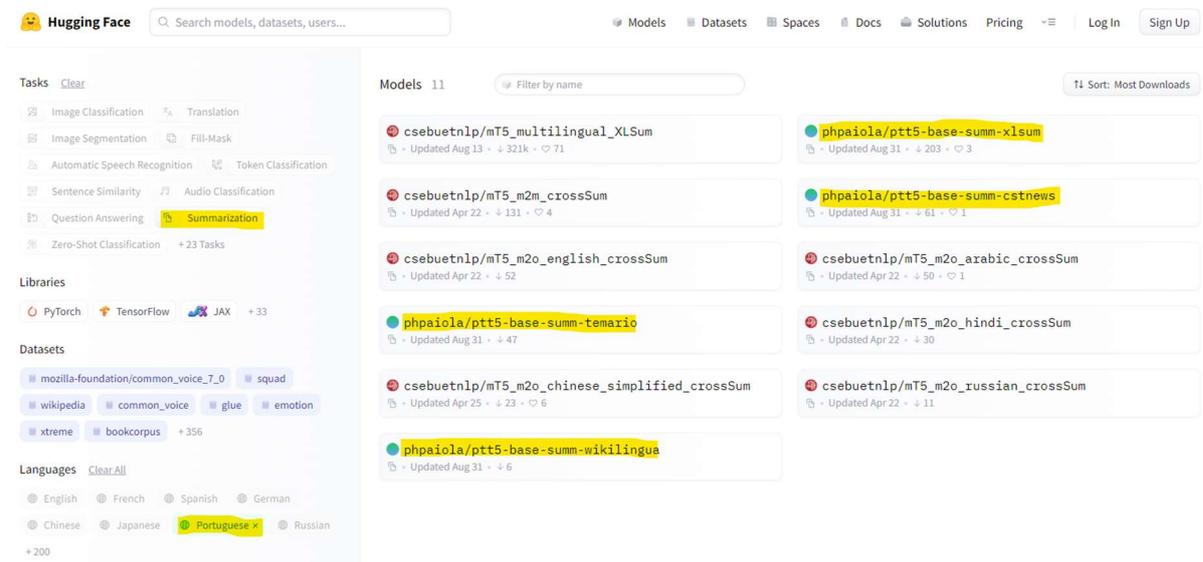
Quadro 4: Sumarização obtida com as ferramentas NLTK e *spaCy*

Trata-se de aquisição, pela Termogás, de participações societárias nas distribuidoras locais de gás natural Companhia de Gás do Amapá – GASAP, Companhia de Gás do Piauí – GASPISA, Agência Goiana de Gás Canalizado – GOIASGÁS e Companhia Rondoniense de Gás – RONGÁS, atualmente detidas pela Comit.Destaca-se que a Superintendência-Geral do CADE (“SG”) já analisou o mercado downstream de distribuição de gás natural em diversas ocasiões 14, afirmando que não há necessidade de subdivisão do mercado relevante na sua dimensão produto – sendo este o gás natural – contudo, reconhecendo que existem diferenças entre os modais de distribuição do produto que pode ser feita de duas formas: por meio de dutos locais, de média e baixa pressão, conectados diretamente aos dutos de transporte; ou a granel, através de veículos como trens, navios e caminhões dotados de tanques onde o gás é armazenado e transportado até a destinação final. Naquela ocasião, a SG entendeu que se tratava de concessão de serviços públicos (serviços de iluminação pública), cuja dimensão do mercado relevante, na jurisprudência do CADE, é delimitada de forma a considerar cada concessão como um mercado relevante específico, configurando um monopólio natural, em que a concorrência se dá pelo mercado e não no mercado. Nesse modo, concluiu-se que aquela operação não resultaria em qualquer sobreposição horizontal. As Partes entendem que o mesmo raciocínio se aplica à atividade de distribuição de gás natural, ou seja, dado que esta é uma atividade caracterizada pelo monopólio natural, não há concorrência no mercado, que é restrito à área de alcance da malha dutoviária. Como informado, a Operação não gera qualquer sobreposição haja vista que as Empresas-Alvo não são ativas. Em um cenário ainda mais conservador, seria possível concluir que há competição potencial entre distribuição canalizada e distribuição a granel para clientes ainda não conectados à malha de distribuição (i.e., localidades fora da zona de monopólio natural) e cuja demanda individualmente justificaria a expansão da malha, pois somente essas localidades estariam inseridas na competição pelo mercado. Na ausência de malha dutoviária, o GNC é considerado a melhor opção quando do transporte de pequenos volumes de gás natural a curtas distâncias e o GNL, por sua vez, é mais competitivo no transporte de gás natural em maiores volumes e longas distâncias. A companhia detém os direitos de exploração, com exclusividade, do serviço de distribuição de gás canalizado, podendo também explorar outras formas de distribuição de gás natural e manufaturado, em todo o território do Estado do Piauí. A companhia tem por objetivo na exploração, com exclusividade, do serviço público de distribuição de

gás canalizado, podendo também explorar outras formas de distribuição de gás natural ou manufaturado, onde produção no estado do Amapá. As Empresas -Alvo são sociedades de economia mista, criadas através de legislações estaduais, cuja principal função é explorar, com exclusividade, o serviço público de distribuição de gás canalizado, podendo também explorar outras formas de distribuição de gás natural ou manufaturado nos estados nos quais foram constituídas. Nos termos do referido Termo de Compromisso, as partes acordaram que a Petrobras alienaria sua participação nacional indireta em companhias distribuidoras de gás natural, seja pelo desinvestimento na Gaspetro (atual Commit), ou alienando a participação da Gaspetro (atual Commit) nas companhias distribuidoras. A Operação não suscita preocupações concorrenciais, uma vez que: (i) a Termogás já detém participação nas Empresas -Alvo; e (ii) nenhuma das Empresas -Alvo se encontra operacional de modo que não é possível verificar qualquer sobreposição ou integração decorrente da proposta aquisição de participação societária nas Empresas -Alvo. Por fim, as Partes sugerem a seguinte redação para o edital de publicação da Operação no Diário Oficial da União: "Requerentes: Termogás S.A. e Commit Gás S.A.. Portanto, não há que se falar em concorrência durante a vigência da concessão do Estado para a prestação do serviço de distribuição de gás natural via dutos. Com relação à competição entre duto e granel, o CADE afirmou que " dada a estrutura de custos inerente ao processamento para viabilizar a distribuição, o gás distribuído a granel (GNC e GNL) normalmente só atua nas localidades em que a distribuidora local de gás canalizado não atua. Além das Empresas -Alvo, o Grupo Termogás, através da Termogás, possui participação societária nas seguintes empresas operacionais atuantes na distribuição de gás natural: (i) Companhia Maranhense de Gás - GASMAR, (ii) Companhia de Gás do Amazonas - CIGÁS 12 e (iii) Companhia Brasileira de Gás - CEBGÁS. No hipotético cenário em que os demais modais de distribuição concorressem com a possibilidade de expansão dos dutos, portanto dilatando o escopo produto do mercado irrelevante e, ainda assim não seria verificada qualquer sobreposição, uma vez que o Grupo. Portanto, resta claro que os demais investimentos do Grupo Termogás na cadeia de gás não resultam em qualquer integração vertical com as atividades de distribuição de gás via dutos as quais as Empresas -Alvo estão autorizadas a realizar. Todavia, o Grupo Termogás informa que dentre seus investimentos em geração de energia elétrica existem duas usinas termelétricas, as quais estão localizadas em Porto Velho/RO, aptas a utilizar óleo e gás natural. A Compass é uma sociedade do Grupo Cosan que tem operações em comercialização de energia elétrica e na distribuição de gás natural por meio da Companhia de Gás de São Paulo - Comgás (" Comgás ") e Companhia de Gás do Estado do Rio Grande do Sul - Sulgás (" Sulgás "), além de outras atividades em estágio pré-operacional ou de projeto. Ela é uma empresa de capital misto que tem por objeto social a exploração e distribuição de gás natural canalizado, e de outras origens, por vias terrestres e fluviais; e a Companhia Hidroelétrica do São Francisco não geraria qualquer sobreposição horizontal, considerando a restrição da atividade de transmissão de energia elétrica à área concedida, por tratar-se de monopólio natural e setor regulado pela ANEEL e objeto de concessão. Tais ações representam de 19,5% do capital social votante e 30,46% do capital social total da GOIASGÁS \uf0b7 RONGÁS: 558.446 (quinhentos e cinquenta e oito mil, quatrocentas e quarenta e seis) ações ordinárias e 2.279.370 (dois milhões, duzentos e setenta e nove mil, trezentas e setenta) ações preferenciais. Dessa forma, as Partes requerem dispensa de apresentar as informações referentes a CEBGÁS na presente notificação, porém permanecem à disposição deste E. Cade para prestar os esclarecimentos que se façam necessários. Como já informado, a Termogás destaca que o contrato firmado entre ela e a Compass, posteriormente cedido pela Compass à Commit, também formaliza a aquisição de 24% do capital social total da Companhia Brasileira de Gás - CEBGÁS. A Operação representa, para a Termogás, a oportunidade de consolidar a sua participação em distribuidoras locais de gás natural e, assim, poder colaborar para que elas possam iniciar a prestação do serviço público de distribuição de gás canalizado e atividades correlatas. Assim, onde não há duto, há espaço para que o granel dispute mercado, mas, uma vez que a companhia local de distribuição por duto realize o investimento para conectar determinada localidade à malha, o gás a granel deixa de ser competitivo ". Por outro lado, a distribuição a granel pressupõe a transformação do gás natural por meio de processo de liquefação ou de compressão (GNL e GNC).

A Figura 10 apresenta os onze modelos de sumarização disponíveis para a língua portuguesa na comunidade *Hugging Face*²².

Figura 10: Modelos de sumarização em português



Dos onze modelos disponíveis para sumarização em português foram utilizados apenas quatro, os quais são especificamente para o português. Os outros modelos são multilíngues que permitem a utilização da língua portuguesa como uma opção de língua natural.

Os modelos da comunidade *Hugging Face* escolhidos são baseados na sumarização abstrativa e os resultados são apresentados nos Quadros 5, 6, 7 e 8.

Quadro 5: Sumarização obtida com a ferramenta *transformers* e o modelo *ptt5-base-summ-temario*

A Termogás S.A. adquiriu participações societárias nas distribuidoras locais de gás natural: (i) Companhia de Gás do Amapá, (GASAP), (ii) Companhia do Piauí - GASPISA ("GASPESA"), (iv) Agência Goiana de Gás Canalizado S/A ("GOIASGÁS"), e, em conjunto com GASAP, GAPISA e GOIASGOÁS ("Empresas -Alvo"), atualmente detidas pela Commit Gás, atual denominação da Petrobrás Gás e Energia, Gaspetro. A operação foi notificada ao CADE sob o Ato de Concentração no 08700.005956/2021-55, quando do exercício de preferência da Termgás, relacionada à participação na época da Gaspetra (atual Commit) 2 na distribuidora local. O contrato foi aprovado sem restrições pelo CADE em 24 de novembro de 2021 e sua conclusão ocorreu em 21 de julho de 2022. A termogás possui diversos empreendimentos nos setores de energia, infraestrutura e gás natural, incluindo participação nas Empresas

Quadro 6: Sumarização obtida com a ferramenta *transformers* e o modelo *ptt5-base-summ-xlsum*

O CADE (Conselho Nacional de Energia e Desenvolvimento Econômico) emitiu nesta sexta-feira um comunicado informando sobre a aquisição de participações societárias na Companhia Brasileira de Gás, RONGÁS e GOIAS GÁS.

²² https://huggingface.co/models?language=pt&pipeline_tag=summarization&sort=downloads. Acesso em 6 dez. 2022.

Quadro 7: Sumarização obtida com a ferramenta *transformers* e o modelo *ptt5-base-summ-wikilingua*

Entenda o que é a Transferência de Participações Societárias. Reconheça o contrato de Concentração de Participação (CEBGÁS) com a Termogás. Aprenda a identificar a Commit (Commit) e a Compass (Compass).

Quadro 8: Sumarização obtida com a ferramenta *transformers* e o modelo *ptt5-base-summ-cstnews*

O objeto da presente notificação trata de aquisição de participação societária nas distribuidoras locais de gás natural. A operação foi notificada ao CADE sob Ato de Concentração no 08700.005956/2021 -55, quando do exercício de preferência, pela Termogás, relacionada à participação nas Empresas -Alvo. A Commit é uma holding que possui diversos empreendimentos.

Como pode-se observar, a sumarização abstrativa apresentou uma sumarização bastante reduzida em relação à abordagem de sumarização extrativa. Esse resultado comprova que, para o caso dos processos de rito sumário do CADE, a sumarização extrativa apresenta um resultado mais interessante para os objetivos desse projeto de pesquisa. Pode-se pensar, porém, em outros usos para esta abordagem.

6.1.5. Avaliação

Os resultados obtidos nessa pesquisa foram submetidos ao especialista da área de negócio do CADE responsável pelos processos de ato de concentração pelo rito sumário (Coordenação-Geral de Análise Antitruste 5 – CGAA5). Este especialista entendeu que os resultados apresentados são relevantes para os trabalhos e para o cumprimento dos objetivos estratégicos da área de negócio. Apontou que já se poderia fazer um piloto com o mercado de energia utilizando as três abordagens apresentadas na POC.

O especialista sugeriu que a sumarização seja feita pelas etapas que compõem o Formulário de Notificação (Anexo A) e não se utilizando o texto completo do documento. Dessa forma, é possível fazer uma sumarização mais focada em cada etapa, tornando a sumarização mais próxima do parecer humano.

Sugeriu também, que a abordagem abstrativa seja utilizada por etapas do Formulário de Notificação para verificar se a sumarização de textos menores não traria uma sumarização mais detalhada e que pudesse ser utilizada nas análises da CGAA5. O especialista considerou que todas as abordagens apresentaram potencial de aproveitamento e que devem ser exploradas para verificar sua real aplicabilidade na área de negócio.

Os indicadores propostos no Planejamento da POC, conforme a Figura 4, foram respondidos segundo a Tabela 3.

Tabela 3: Indicadores de validação da POC

Indicador	Resposta
1. As saídas passam uma ideia completa?	Os dois primeiros resultados consolidam bem o que foi apresentado e parecem completos.
2. A ideia que as saídas passam guarda coerência com o que seria um parecer humano?	Sim, tem proximidade com uma análise descritiva realizada por humano, com exceção dos <i>typos</i> .
3. Qual seria o percentual gerado nas saídas em relação ao parecer humano completo?	Vai depender muito mais se as partes consideraram todos os precedentes e cenários possíveis. De toda forma, nesse caso, comparando com o parecer humano, chegaria a 80%.
4. A ordem das saídas está coerente?	Em geral, sim.
5. O reconhecimento das entidades está coerente?	Em geral, sim.
6. Qual o modelo que melhor se aproxima do parecer humano?	O modelo da primeira abordagem.
7. Faltou algo muito significativo nas saídas?	Em geral, não. Para reduzir as redundâncias, talvez seja o caso de gerar o processamento do texto por tópicos (etapas do formulário), caso seja possível isolar os textos.

Fonte: Autoria própria

Sendo assim, o especialista ratificou o resultado atestando a validade das abordagens apresentadas para a área de negócio e para o CADE. Dessa forma, o objetivo principal desse projeto de pesquisa foi atingido.

7. CONCLUSÃO E TRABALHOS FUTUROS

O objetivo principal deste projeto de pesquisa foi a proposição de um processo inteligente para a produção de minutas de pareceres utilizando a metodologia CRISP-DM para demarcar as etapas envolvidas. A pesquisa partiu da necessidade de reduzir o tempo de resposta dos processos de rito sumário analisados pelo CADE, por meio das tecnologias de PLN, para apoiar o CADE na melhoria contínua na oferta dos seus serviços. Os resultados obtidos ratificam o objetivo principal, atestando a validade das abordagens apresentadas para a área de negócio e para o CADE. Os modelos testados já podem entrar em produção como piloto de algum dos mercados relevantes.

Conforme definido no objetivo específico *(i)*, foi elaborada a revisão bibliográfica e a construção da rede de coocorrência de palavras-chave para identificar o estado da arte em geração automática de pareceres jurídicos e os temas emergentes relacionados (item 3). O objetivo específico *(ii)* tratou da identificação das técnicas de IA, mais especificamente de PLN, aplicáveis no projeto de pesquisa (item 2), o que permitiu atingir os objetivos específicos *(iii)* propondo um modelo de solução (item 4.2) e *(iv)* elaborando uma prova de conceito (item 6.1.4).

Durante a pesquisa foram vislumbradas algumas lacunas ou limitações que podem ser desdobradas em trabalhos futuros, melhorando os resultados do próprio tema da dissertação, assim como, a abertura de novas linhas de pesquisas. Por exemplo, a dificuldade em trabalhar o texto original dos Formulários de Notificação em PDF. Ainda não existem técnicas para tratar a formatação, os títulos e subtítulos, cabeçalhos e rodapés de um arquivo PDF.

A evolução dos métodos para tratamentos de arquivos textuais tem potencial de grande impacto para quem utiliza dados não estruturados como entrada para os algoritmos de PLN. A área judicial poderá se beneficiar com essa evolução, pois o insumo para o treinamento dos algoritmos são textos produzidos nos processos judiciais.

Para o CADE, esse problema será abordado com o novo projeto do Formulário Digital de Notificação de AC, voltado para a estruturação das informações da organização em um banco de dados. O objetivo do projeto é reunir em um único documento as informações relevantes sobre um ato de concentração em análise.

No caso do CADE, esse processo de sumarização é possível, haja vista que o FN tem uma estruturação definida pela Resolução nº 33 (Anexo A). Na área judicial, no entanto, isto nem sempre ocorre. Nesses casos, novas classes de problemas deverão ser identificadas para o possível emprego de técnicas de IA e, em particular, de PLN.

REFERÊNCIAS

- ARAÚJO, P. H. L.; CAMPOS, T. E.; OLIVEIRA, R. R. R.; STAUFFER, M.; COUTO, S.; BERMEJO, P. LeNER-Br: A Dataset for Named Entity Recognition in Brazilian Legal Text. In: **Computational Processing of the Portuguese Language (PROPOR)**. Lecture Notes in Computer Science, vol. 11122. Springer, Cham, 2018. https://doi.org/10.1007/978-3-319-99722-3_32.
- BHATTACHARYA, P.; HIWARE, K.; RAJGARIA, S.; POCHHI, N.; GHOSH, K.; GHOSH, S. A Comparative Study of Summarization Algorithms Applied to Legal Case Judgments. In: L. Azzopardi, et al. (eds) **Advances in Information Retrieval (ECIR)**. Lecture Notes in Computer Science, vol. 11437. Springer, Cham, 2019. https://doi.org/10.1007/978-3-030-15712-8_27
- BI, S.; ALI, Z.; WANG, M.; WU, T.; QI, G. Learning heterogeneous graph embedding for Chinese legal document similarity. **Knowledge-Based Systems**, v. 250, 2022. <https://doi.org/10.1016/j.knosys.2022.109046>.
- BUOCZ, T. J. Artificial Intelligence in Court: Legitimacy Problems of AI Assistance in the Judiciary. **Retskraft – Copenhagen Journal of Legal Studies**, v. 2, n. 1, p. 41-59, 2018.
- CADE. Defesa da concorrência no Brasil: 50 anos, p. 155-156, 2013. Disponível em: <https://cdn.cade.gov.br/Portal/centrais-de-conteudo/publicacoes/livro-50-anos/livro-defesa-da-concorrenca-no-brasil-50-anos.pdf>. Acesso em 16 jan. 2023.
- CARDELLINO, C.; ALEMANY, L. A.; TERUEL, M.; VILLATA, S. A low-cost, high-coverage legal named entity recognizer, classifier and linker. In: INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND LAW, 16, **Proceedings [...]**. 2017. <https://doi.org/10.1145/3086512.3086514>.
- CASTANO, S.; FALDUTI, M.; FERRARA, A.; MONTANELLI, S. A knowledge-centered framework for exploration and retrieval of legal documents. **Information Systems**, v. 106, 2022. <https://doi.org/10.1016/j.is.2021.101842>.
- CHALKIDIS, I.; ANDROUTSOPOULOS, I.; MICHOS, A. Extracting contract elements. In: INTERNATIONAL CONFERENCE ON ARTIFICIAL INTELLIGENCE AND LAW, 16, **Proceedings [...]**. 2017. <https://doi.org/10.1145/3086512.3086515>
- CHALKIDIS, I.; KAMPAS, D. Deep learning in law: early adaptation and legal word embeddings trained on large corpora. **Artificial Intelligence and Law**, v. 27, n. 2, p. 171–198, 2019. <https://doi.org/10.1007/s10506-018-9238-9>
- CHAPMAN, P.; CLINTON, J.; KERBER, R.; KHABAZA, T.; REINARTZ, T.; SHEARER, C.; WIRTH, R. **CRISP-DM 1.0: Step-by-step data mining guide**. SPSS Inc., v. 9, p. 13, 2000. Disponível em: <https://docplayer.net/202628-Crisp-dm-1-0-step-by-step-data-mining-guide.html>

CHEN, H.; WU, L.; CHEN, J.; LU, W.; DING, J. A comparative study of automated legal text classification using random forests and deep learning. **Information Processing & Management**, v. 59, n. 2, p. 102798, mar. 2022. <https://doi.org/10.1016/j.ipm.2021.102798>.

COBBE, J. Administrative law and the machines of government: judicial review of automated public-sector decision-making. **Legal Studies**. v. 39, n. 4, p. 636–655, 2019. <https://doi.org/10.1017/lst.2019.9>.

DIEGA, G. N. Against the dehumanisation of decision-making: algorithmic decisions at the crossroads of intellectual property, data protection, and freedom of information. **Journal of Intellectual Property, Information Technology and E-Commerce Law**, 2018. Disponível em: https://www.jipitec.eu/issues/jipitec-9-1-2018/4677/JIPITEC_9_1_2018_3-34

DONEDA, D.; ALMEIDA V. A.F. O que é a governança de algoritmos? **Politics. Uma publicação do Instituto Nupef**. Outubro de 2016. Recuperado de <https://www.politics.org.br/edicoes/o-que-%C3%A9-governan%C3%A7a-de-algoritmos>.

ENGELMANN, W.; FRÖHLICH, A. V. K. Inteligência artificial aplicada à decisão judicial: o papel dos algoritmos no processo de tomada de decisão. **Revista Jurídica (FURB)**, v. 24, n. 54, p. 8274, 2020.

FAYYAD, U.; PIATETSKY-SHAPIRO, G.; SMYTH, P. **From Data Mining to Knowledge Discovery in Databases**. American Association for Artificial Intelligence, 1996. <https://doi.org/10.1609/aimag.v17i3.1230>.

FERNANDES, W. P. D.; FRAJHOF, I. Z.; ALMEIDA, G. F. C. F.; RODRIGUES, A. M. B.; BARBOSA, S. D. J.; KONDER, C. N.; NASSER, R. B.; CARVALHO, G. R.; LOPES, H. C. V. Extracting value from Brazilian Court decisions. **Information Systems**, v. 106, p. 101965, maio 2022. <https://doi.org/10.1016/j.is.2021.101965>.

FREITAS, C. O. A.; BARDDAL, J. P. Análise preditiva e decisões judiciais: controvérsia ou realidade? **Revista Democracia Digital e Governo Eletrônico**, v. 1, n.18, p. 107-126, 2019.

HIGASI, P. K. B. C. **O direito aplicável às decisões produzidas por software e machine learning**. Dissertação (Mestrado em Tecnologias da Inteligência e Design Digital) – Pontificia Universidade Católica de São Paulo, São Paulo, 2020.

HOTH, A.; NÜRNBERGER, A.; PAAß, G. A brief survey of text mining. **LDV forum**, v. 20, n. 1, p. 19-62, 2005.

ISE, O. A. Integration and analysis of unstructured data for decision making: Text analytics approach. **International Journal of Open Information Technologies**, v. 4 n. 10, p. 82-88, 2016.

KATZ, D. M. Quantitative Legal Prediction – or – How I Learned to Stop Worrying and Start Preparing for the Data Driven Future of the Legal Services Industry. **Emory Law Journal**, v. 62, n. 2, 2013.

LI, P.; TSENG, C.; ZHENG, Y.; CHEW, J. A.; HUANG, L.; JARMAN, B.; NEEDELL, D. Guided Semi-Supervised Non-Negative Matrix Factorization. **Algorithms**, v. 15, n. 5, p. 136, 2022. <https://doi.org/10.3390/a15050136>.

MANDAL, A.; CHAKI, R.; SAHA, S.; GHOSH, K.; PAL, A.; GHOSH, S. Measuring similarity among legal court case documents. In: ANNUAL ACM INDIA COMPUTE CONFERENCE – COMPUTE '17, 10, **Proceedings [...]**. 2017. <https://doi.org/10.1145/3140107.3140119>.

MEDINA, J. M. G.; MARTINS, J. P. N. P. A era da inteligência artificial: as máquinas poderão tomar decisões judiciais? **Revista dos Tribunais**, v. 1020/2020, p. 311-338, 2020.

PEREIRA NETO, C. M. S.; CASAGRANDE, P. L. **Direito concorrencial**: doutrina, jurisprudência e legislação. São Paulo: Saraiva, p. 53, 2016.

PUGLIESI, M.; BRANDÃO, A. M. Uma conjectura sobre as tecnologias de big data na prática jurídica. **Revista da Faculdade de Direito da UFMG**, n. 67, p. 453-482, 2015.

SANSONE, C.; SPERLÍ, G. Legal Information Retrieval systems: State-of-the-art and open issues. **Information Systems**, v. 106, 2022. <https://doi.org/10.1016/j.is.2021.101967>.

ŠAVELKA, J.; ASHLEY, K. D. Legal information retrieval for understanding statutory terms. **Artificial Intelligence and Law**, v. 30, n. 2, p. 245–289, 2021. <https://doi.org/10.1007/s10506-021-09293-5>.

SILVA, L.H.S.; SOUZA, S.A.D. **Análise econômica da defesa da concorrência no Brasil**, 2016.

SILVEIRA, P. B. **Direito da Concorrência**. Rio de Janeiro: Forense, pg.13, 2021.

SONG, D.; VOLD, A.; MADAN, K.; SCHILDER, F. Multi-label legal document classification: A deep learning-based approach with label-attention and domain-specific pre-training. **Information Systems**, v. 106, 2021. <https://doi.org/10.1016/j.is.2021.101718>.

SONG Z.; LI Y.; LI D.; WANG S. Multi-label Classification of Legal Text with Fusion of Label Relations. **Science Press**, v. 35, n. 2, p. 185–192, fev. 2022. <https://doi.org/10.16451/j.cnki.issn1003-6059.202202009>.

SOUSA, R. N. **MINERJUS**: solução de apoio à classificação processual com uso de Inteligência Artificial. 2019. Dissertação (Mestrado em Modelagem Computacional e Sistemas) – Universidade Federal do Tocantins, Programa de Pós-graduação em Modelagem Computacional e Sistemas, Palmas, 2019.

SPOSITTO, O. M.; BOSSERO, J. C.; MORENO, E. J.; LEDESMA, V. A.; MATTEO, L. R. Lexical Analysis Using Regular Expressions for Information Retrieval from a Legal Corpus. **Communications in Computer and Information Science**, p. 312–324, 2022. https://doi.org/10.1007/978-3-031-05903-2_21.

SUGATHADASA, K.; AYESHA, B.; SILVA, N.; PERERA, A. S.; JAYAWARDANA, V.; LAKMAL, D.; PERERA, M. Legal Document Retrieval Using Document Vector Embeddings and Deep Learning. **Advances in Intelligent Systems and Computing**, p. 160–175, 2018. https://doi.org/10.1007/978-3-030-01177-2_12.

TACCA, A.; ROCHA, L. S. Inteligência artificial: reflexos no sistema do direito. **NOMOS: Revista do Programa de Pós-Graduação em Direito da UFC**, v. 38, n. 2, p. 53-68, 2018.

TRIVEDI, A.; TRIVEDI, A.; VARSHNEY, S.; JOSHIPURA, V.; MEHTA, R.; DHANANI, J. Similarity Analysis of Legal Documents: A Survey. In: Fong, S., Dey, N., Joshi, A. (eds) **ICT Analysis and Applications**, Lecture Notes in Networks and Systems, vol. 154. Springer, Singapore. https://doi.org/10.1007/978-981-15-8354-4_49

UNDAVIA, S.; MEYERS, A.; ORTEGA, J. A Comparative Study of Classifying Legal Documents with Neural Networks. In: FEDERATED CONFERENCE ON COMPUTER SCIENCE AND INFORMATION SYSTEMS, **Proceedings [...]**. 2018. <https://doi.org/10.15439/2018F227>.

VON HOHENDORFF, R.; LAZZARETTI, B. K. O uso de inteligência artificial na tomada de decisões judiciais: uma análise sob a perspectiva da crítica hermenêutica do direito. **RDUno: Revista do Programa de Pós-Graduação em Direito da Unochapecó**, v. 3, n. 4, p. 15-32, 2020.

VAN OPIJNEN, M.; SANTOS, C. On the concept of relevance in legal information retrieval. **Artificial Intelligence and Law**, v. 25, n. 1, p. 65–87, 2017. <https://doi.org/10.1007/s10506-017-9195-8>.

WEI, F.; QIN, H.; YE, S.; ZHAO, H. Empirical Study of Deep Learning for Text Classification in Legal Document Review. In: IEEE INTERNATIONAL CONFERENCE ON BIG DATA, **Proceedings [...]**. 2018. <https://doi.org/10.1109/BigData.2018.8622157>.

ANEXO A – ANEXO II DA RESOLUÇÃO Nº 33, DE 14 DE ABRIL DE 2022

ANEXO II

FORMULÁRIO PROCEDIMENTO SUMÁRIO

Instruções gerais de preenchimento:

- (i) As respostas a este formulário devem ser justificadas e as informações exatas e completas, com a indicação da fonte para os cálculos e a apresentação de documentos comprobatórios disponíveis. Na impossibilidade de fornecer informações exatas, as partes devem fornecer estimativas com a indicação das respectivas fontes e metodologia de cálculo utilizada;
- (ii) As informações devem ser fornecidas em unidades padronizadas para todas as respostas deste formulário. Os dados relativos a unidades monetárias devem ser fornecidos em reais (R\$), com a indicação da taxa de câmbio utilizada para a respectiva conversão, quando for o caso;
- (iii) Sempre que possível, apresente fontes alternativas de informações relevantes para a análise da presente operação (sítios eletrônicos, publicações diversas, associações, órgãos de controle, pesquisas realizadas por terceiros, jurisprudência nacional e internacional etc.);
- (iv) Na ausência de especificação em sentido contrário, o termo “parte” é utilizado neste formulário de forma a abranger as entidades diretamente envolvidas na operação (referidas como “partes diretamente envolvidas”), e os respectivos grupos econômicos, conforme definição do item II.5 deste Anexo;
- (v) Sempre que as partes diretamente envolvidas na operação sejam entidades sem atividades operacionais ou veículos especialmente criados para a realização da operação, devem ser fornecidas as informações relativas à parte concorrencialmente relevante;
- (vi) A Etapa VI desse formulário deve ser preenchida apenas nos casos em que haja sobreposição horizontal e/ou integração vertical decorrentes da operação em análise.

ETAPA I - DESCRIÇÃO DA OPERAÇÃO

Apresentar um resumo da operação (até 500 palavras), especificando as partes na concentração, as respectivas áreas de atividade, a natureza da concentração (por exemplo, fusão, aquisição, joint venture etc.), os mercados em que a concentração produzirá algum impacto (destacando-se os principais mercados envolvidos), e a justificativa estratégica e econômica para a operação. O resumo deve ser elaborado de forma a não conter quaisquer informações confidenciais ou segredos comerciais.

ETAPA II - INFORMAÇÕES RELATIVAS ÀS PARTES

II.1. Informe os nomes (razão social, denominação, nome do estabelecimento, nome de fantasia), CNPJ e forma legal (S.A. de capital aberto, S.A. de capital fechado, Ltda. etc.) das partes diretamente envolvidas na operação.

II.2. Apresente os endereços completos das partes diretamente envolvidas na operação (incluindo o CEP e o sítio eletrônico), e nome, números de telefone e fax, endereço eletrônico e cargo ocupado por representante técnico das empresas a ser contatado. Sempre que possível, informar o endereço da sede/representação das partes no Brasil.

II.3. Indique o(s) representante(s) legal(is) e endereço completo (incluindo telefone, fax, CEP, e endereço eletrônico).

II.4. Apresente os faturamentos brutos das partes diretamente envolvidas na operação, e de seus respectivos grupos econômicos (segundo definição do art. 4º da Resolução 02/2012), no Brasil e em todo o mundo (incluindo o Brasil), no ano fiscal anterior à presente operação.

II.5. Indique os grupos econômicos a que pertencem as partes diretamente envolvidas na operação e forneça uma lista de todas as pessoas físicas ou jurídicas de direito público ou privado pertencentes aos grupos econômicos, com atividades no território nacional, informando:

- a) Organograma com a estrutura societária das partes diretamente envolvidas na operação;
- b) Organograma com a estrutura societária do grupo econômico a que tais partes pertencem.

II.5.1. Para fins de resposta a este e aos demais itens deste Anexo, considera-se grupo econômico, cumulativamente:

- a) As empresas que estejam sob controle comum, interno ou externo; e
- b) As empresas nas quais qualquer das empresas da alínea “a” seja titular, direta ou indiretamente, de pelo menos 20% (vinte por cento) do capital social ou votante.

II.5.2. No caso dos fundos de investimento, são considerados integrantes do mesmo grupo econômico para fins de resposta a este e aos demais itens deste Anexo, cumulativamente:

- a) O fundo envolvido na operação;
- b) Os fundos que estejam sob a mesma gestão do fundo envolvido na operação;
- c) O gestor;
- d) Os grupos dos cotistas, conforme definidos no item II.5.1., que detenham direta ou indiretamente participação igual ou superior a 20% das cotas do fundo envolvido na operação;
- e) As empresas controladas pelo fundo envolvido na operação e as empresas nas quais o referido fundo detenha direta ou indiretamente participação igual ou superior a 20% (vinte por cento) do capital social ou votante; e
- f) As empresas controladas pelos fundos que estejam sob a mesma gestão do fundo envolvido na operação e as empresas nas quais esses fundos detenham direta ou indiretamente participação igual ou superior a 20% (vinte por cento) do capital social ou votante.

Observação: No que diz respeito aos agentes incluídos nas alíneas “b” e “f” do item II.5.2., fornecer listagem e demais informações somente dos fundos e empresas que sejam horizontal ou verticalmente relacionados às atividades objeto da operação, segundo CNAE 2.0 a 7 dígitos ou versão mais atual.

II.6. Informe a nacionalidade de origem dos grupos econômicos indicados no item II.5.

II.7. Informe as operações realizadas durante os últimos cinco anos, pelas pessoas listadas no item II.5, e as respectivas decisões do Cade, quando for o caso.

II.8. Informe todas as atividades econômicas desempenhadas pelas partes diretamente envolvidas na operação, no Brasil, indicando o faturamento bruto obtido com cada uma das atividades no ano fiscal anterior ao da apresentação da notificação. Classifique-as segundo a CNAE 2.0 a 7 dígitos ou versão mais atual.

II.9. Informe todas as atividades econômicas desempenhadas pelas demais empresas que fazem parte dos grupos econômicos envolvidos na operação, no Brasil. Classifique-as segundo a CNAE 2.0 a 7 dígitos ou versão mais atual.

II.10. Forneça uma lista de todas as empresas com atividades no território nacional, segundo CNAE 2.0 a 7 dígitos ou versão mais atual, que sejam horizontal ou verticalmente relacionadas às atividades objeto da operação, nas quais pelo menos um dos integrantes do grupo detenha participação igual ou superior a 10% no capital social ou votante. Na resposta, apresente também o organograma da estrutura societária das empresas que se enquadram nesse critério.

II.11. No que diz respeito às empresas referidas nas respostas aos itens II.5 e II.10, forneça uma lista dos membros dos seus órgãos de gestão que sejam igualmente membros dos órgãos de gestão ou de fiscalização de quaisquer outras empresas atuantes nas mesmas atividades econômicas, conforme CNAE 2.0 a 7 dígitos, indicando tais empresas.

ETAPA III - ELEMENTOS RELATIVOS À OPERAÇÃO

III.1. Informe se esta notificação refere-se à primeira apresentação. Caso negativo, informe o motivo pelo qual o ato está sendo reapresentado/emendado.

III.2. Descreva a operação notificada, indicando:

a) Se a operação projetada consiste em um(a):

1. fusão;
2. aquisição de controle;
3. aquisição de quotas/ações sem aquisição de controle;
4. consolidação de controle;
5. aquisição de ativos;
6. incorporação;
7. joint venture clássica (criação de empresa para explorar outro mercado);
8. joint venture concentracionista (criação de empresa para explorar mercado já explorado pelas empresas associadas);
9. outra forma de operação não coberta pelas alternativas anteriores (especificar);

b) Se a operação abrange, total ou parcialmente, as atividades das partes;

c) O valor da operação e a forma de pagamento;

d) No caso de aquisição de ativos, todos esses ativos, tangíveis e/ou intangíveis. No caso dos ativos tangíveis, indicar também suas localizações (endereço e CEP);

e) No caso de aquisição de participação societária, o dispositivo da Seção III desta Resolução no qual a operação se enquadra;

f) A estrutura societária da empresa alvo, antes e após a realização da operação, ou da nova empresa formada. As informações deste item devem ser ilustradas com a utilização de mapas, organogramas ou diagramas.

III.3. Informe as demais jurisdições em que a presente operação foi ou será apresentada, bem como a(s) data(s) da(s) notificação(ões).

III.4. Especifique se a operação está sujeita a aprovação de outros órgãos reguladores no Brasil ou no exterior.

III.5. Descreva a justificativa econômica e/ou estratégica para a operação.

III.6. Informe se a operação contempla cláusulas restritivas à concorrência (por exemplo, cláusulas de não concorrência ou de exclusividade). Caso afirmativo, apresentar tais cláusulas, indicando sua localização nos documentos relativos à operação, bem como sua justificativa econômica.

ETAPA IV - DOCUMENTAÇÃO

IV.1. Apresente cópia da versão final ou mais recente de todos os instrumentos contratuais relativos à realização da operação, listando os respectivos anexos relevantes para a análise antitruste.

IV.2. Apresente cópias de acordos de não concorrência e de acionistas, se houver.

IV.3. Apresente uma lista contendo a relação de todos os demais documentos que tenham sido criados em decorrência da operação.

IV.4. Apresente cópia mais recente do relatório anual e/ou das demonstrações financeiras auditadas das partes diretamente envolvidas na operação e dos respectivos grupos econômicos.

ETAPA V - DEFINIÇÃO DO(S) MERCADO(S) RELEVANTE(S)

V.1. Informe e descreva todas as linhas de produtos comercializados e/ou serviços prestados no Brasil pelas partes diretamente envolvidas na operação.

V.2. Informe todas as linhas de produtos comercializados e/ou serviços prestados no Brasil pelas demais empresas que fazem parte dos grupos econômicos envolvidos na operação.

V.3. Identifique todas as linhas de produtos e/ou serviços em que poderiam ser verificadas sobreposições horizontais, verticais e/ou complementares decorrentes da presente operação.

V.4. Indique todas as áreas geográficas atendidas por cada uma das empresas que ofertam as linhas de produtos e/ou serviços identificadas no item V.3 (caso a empresa não atenda todo o Brasil, especificar os municípios e/ou estados atendidos pela mesma).

V.5. Considerando as respostas dos itens V.3 e V.4, defina o(s) mercado(s) relevante(s) sob as dimensões produto e geográfica. Fundamente as definições adotadas, baseando-se nos seguintes fatores, na medida em que aplicáveis:

- a) substituíbilidade sob a ótica da demanda, levando em consideração, dentre os fatores considerados relevantes, a possibilidade de substituição frente a aumentos de preços entre marcas e/ou produtos ou serviços e entre outros produtos fora do mercado;
- b) substituíbilidade sob a ótica da oferta, levando em consideração, dentre outros fatores, aspectos técnico-operacionais, tecnologias, custos de instalação (inclusive custos irrecuperáveis) e possibilidade de redirecionamento de atividades;
- c) natureza e características dos produtos e/ou serviços;
- d) importância dos preços dos produtos e/ou serviços;
- e) perfil dos clientes (exemplo: renda, gênero, porte da firma etc.) e dimensionamento do mercado desses clientes (quantidade vendida);
- f) preferências dos consumidores, prestando informações sobre fidelidade a marcas, forma e momentos de consumo, dentre outros fatores considerados relevantes;
- g) custos de distribuição/transporte;
- h) diferenças nas estruturas de oferta e/ou de preços entre áreas geográficas vizinhas;
- i) possibilidade de importações;
- j) jurisprudência do CADE;
- k) jurisprudência internacional;
- l) outros fatores.

V.6. Explique a relevância da diferenciação dos produtos e/ou serviços em termos de atributos ou qualidades, e a medida em que os produtos das partes são substitutos próximos.

ETAPA VI - ESTRUTURA DE OFERTA

(responder esta etapa apenas nos casos em que haja sobreposição horizontal e/ou integração vertical decorrentes da operação em análise)

VI.1. Apresente uma estimativa da dimensão total do(s) mercado(s) relevante(s) identificado(s) no item V.5, em termos de valor (faturamento bruto) e de volume de vendas, no ano fiscal anterior à operação.

Observação: Para estimativa de mercado nacional, considere: produção total do produto/serviço no Brasil + importações – exportações.

VI.2. Apresente o total das vendas, em termos de valor (faturamento bruto) e volume, bem como uma estimativa das participações para cada uma das partes diretamente envolvidas na operação, no(s) mercado(s) relevante(s) identificado(s) no item V.5, no ano fiscal anterior à operação. Apresente dados segmentados para vendas cativas e vendas para clientes independentes[3], quando for o caso.

VI.3. Apresente o total das vendas, em termos de valor (faturamento bruto) e volume, bem como uma estimativa das participações das demais empresas integrantes dos respectivos grupos econômicos, no(s)

mercado(s) relevante(s) identificado(s) no item V.5, no ano fiscal anterior à operação. Apresente dados segmentados para vendas cativas e vendas para clientes independentes, quando for o caso.

VI.4. Apresente uma estimativa das participações de mercado, em termos de valor (faturamento bruto) e de volume de vendas, de todos os concorrentes (incluindo importadores) que detenham pelo menos 5% do(s) mercado(s) relevante(s) indicado(s) no item V.5, no ano fiscal anterior à operação. A partir dessas informações, forneça uma estimativa do HHI antes e após a concentração, bem como a diferença entre os dois (Δ HHI).

VI.5. Apresente nome, endereço completo, números de telefone e fax, e sítio eletrônico dos 5 (cinco) principais concorrentes identificados no item VI.4, para todos os mercados relevantes definidos no item V.5.

VI.6. Apresente uma estimativa do valor e volume totais das importações relativas ao(s) mercado(s) relevante(s) identificados no item V.5, informando também:

- a) o valor, volume e as proporções de tais importações provenientes das partes;
- b) a existência de barreiras tarifárias, regulatórias ou de qualquer natureza, que possam afetar essas importações. Em caso afirmativo, explicar em que medida tais barreiras interferem nas condições de importação do(s) produtos(s) e se são de caráter permanente ou temporário (nesse último caso, apresentar o período de validade de tais barreiras);
- c) uma estimativa da medida em que os custos de transporte e outros custos afetam essas importações.

VI.7. Identifique todas as associações comerciais a que pertencem as partes, no Brasil, relativas ao(s) mercado(s) relevante(s) identificado(s) no item V.5, com endereço completo, sítio na Internet, nome de contato e respectivo cargo, números de telefone e fax, e endereço eletrônico.

ETAPA VII – OBSERVAÇÕES FINAIS

VII.1. Apresente quaisquer outros comentários ou informações que julgue relevantes para a análise da presente operação.

ANEXO B – TEXTO “LIMPO” UTILIZADO NOS MODELOS DE SUMARIZAÇÃO

Trata-se de aquisição, pela Termogás S.A. (“Termogás”), de participações societárias nas distribuidoras locais de gás natural: (i) Companhia de Gás do Amapá – GASAP (“GASAP”), (ii) Companhia de Gás do Piauí – GASPIISA (“GASPIISA”), (iii) Agência Goiana de Gás Canalizado S/A – GOIASGÁS (“GOIASGÁS”) e (iv) Companhia Rondoniense de Gás – RONGÁS (“RONGÁS” e em conjunto com GASAP, GAPISA e GOIASGÁS “Empresas-Alvo”), atualmente detidas pela Commit Gás S.A. (“Commit”), atual denominação da Petrobrás Gás S.A. – Gaspetro (“Gaspetro”), empresa controlada pela Compass Gás e Energia S.A. (“Compass”) (a “Operação”).

Cumprido esclarecer que o contrato objeto da presente notificação prevê ainda a consumação da aquisição de participação societária, pela Termogás, na Companhia Brasileira de Gás (“CEBGÁS”). Essa operação foi notificada ao CADE sob Ato de Concentração nº 08700.005956/2021-55, quando do exercício de preferência, pela Termogás, relacionada à participação na época da Gaspetro (atual Commit)2 na distribuidora local em questão. A operação foi aprovada sem restrições em 24 de novembro de 2021 e sua conclusão ocorreu em 21 de julho de 2022, estando, portanto, excluída da Operação.

A Termogás possui diversos empreendimentos nos setores de energia, infraestrutura e gás natural, incluindo participação nas Empresas-Alvo.

A Commit é uma holding de distribuição de gás natural com participações em diversas distribuidoras de gás natural no Brasil. Antes denominada Petrobras Gás S.A. – Gaspetro, a Companhia teve 51% (cinquenta e um por cento) de suas ações vendidas pela Petróleo Brasileiro S.A. – Petrobras (“Petrobras”) para a Compass, tendo sido a operação aprovada sem restrições pelo CADE em 22 de junho de 2022. O fechamento da referida operação ocorreu em 11 de julho de 2022, com a efetiva transferência de 51% das ações da Commit antes detidas pela Petrobras para a Compass, mesma data em que houve a formalização da mudança da denominação social de Petrobras Gás S.A. – Gaspetro para Commit Gás S.A.

A Compass é uma sociedade do Grupo Cosan que tem operações em comercialização de energia elétrica e na distribuição de gás natural por meio da Companhia de Gás de São Paulo – Comgás (“Comgás”) e Companhia de Gás do Estado do Rio Grande do Sul – Sulgás (“Sulgás”), além de outras atividades em estágio pré-operacional ou de projeto. Com a aquisição da Commit, a Compass passou a ser titular de participação societária indireta em sociedades que atuam na distribuição de gás canalizado em diversos Estados do Brasil, incluindo as Empresas-Alvo. As demais sociedades do Grupo Cosan desenvolvem negócios nos segmentos de energia, combustíveis, lubrificantes e logística.

O restante das ações da Commit, correspondente à 49% (quarenta e nove por cento) das ações, pertencem à Mitsui Gás e Energia do Brasil Ltda. (“Mitsui”), que, no entanto, não é parte envolvida na Operação, sendo dispensada sua apresentação.

Portanto, a Commit pertence aos grupos econômicos Cosan e Mitsui.

As Empresas-Alvo são sociedades de economia mista, criadas através de legislações estaduais, cuja principal função é explorar, com exclusividade, o serviço público de distribuição de gás canalizado, podendo também explorar outras formas de distribuição de gás natural ou manufaturado nos estados nos quais foram constituídas. No entanto, nenhuma delas é operacional.

A Operação será realizada no segmento de distribuição de gás natural canalizado nos estados do Amapá, Piauí, Goiás e Rondônia.

No entanto, a Operação não resultará em qualquer impacto concorrencial, uma vez que a Termogás já detém participação nas Empresas-Alvo e nenhuma delas se encontra operacional atualmente.

Nesse contexto, cumpre lembrar que a distribuição de gás natural via dutos é um monopólio natural. Assim não há que se falar em concorrência durante a vigência da concessão do Estado para a prestação do serviço de distribuição de gás natural via dutos.

Em vista do exposto, as Partes, respeitosamente, entendem que a Operação pode ser aprovada pelo CADE sob o rito sumário, com base no disposto nos artigos 6º e 8º, inciso II, da Resolução CADE nº 33, de 14 de abril 2022.

Por fim, as Partes sugerem a seguinte redação para o edital de publicação da Operação no Diário Oficial da União:

“Requerentes: Termogás S.A. e Commit Gás S.A.. Advogados: Francisco Todorov, Isabella Giorgi, Vitor Damasio e Marília Ventura. Natureza da Operação: aquisição de participação societária. Setor econômico envolvido: distribuição de gás natural (CNAE 3520-4/02)”

A Termogás possui diversos empreendimentos nos setores de energia, infraestrutura e gás natural no Brasil. A Tabela abaixo apresenta a estrutura societária da Termogás, listando os seus acionistas e suas respectivas participações no capital social da empresa.

A Commit é uma sociedade controlada pela Compass, que por sua vez, é uma sociedade controlada pela Cosan S.A. (“Cosan”), que detém 88% das ações de sua emissão. De acordo com seu melhor conhecimento, nenhuma empresa do Grupo Termogás detém participação acionária de 10% ou mais em outras empresas fora do Grupo Termogás com atividades no Brasil que sejam horizontalmente ou verticalmente relacionada à indústria de gás natural.

Trata-se de aquisição, pela Termogás, de participações societárias nas distribuidoras locais de gás natural Companhia de Gás do Amapá – GASAP, Companhia de Gás do Piauí – GASPIISA, Agência Goiana de Gás Canalizado – GOIASGÁS e Companhia Rondoniense de Gás – RONGÁS, atualmente detidas pela Commit.

A Operação ocorre no contexto da compra da Gaspetro (atual Commit) pela Compass, que, por sua vez, decorreu do Termo de Compromisso de Cessação (Requerimento nº 08700.003136/2019-12), firmado entre o CADE e a Petrobras. Nos termos do referido Termo de Compromisso, as partes acordaram que a Petrobras alienaria sua participação acionária indireta em companhias distribuidoras de gás natural, seja pelo desinvestimento na Gaspetro (atual Commit), ou alienando a participação da Gaspetro (atual Commit) nas companhias distribuidoras.

A aquisição da Gaspetro (atual Commit) pela Compass foi notificada ao CADE e aprovada, sem restrições, pelo Tribunal em 22 de junho de 20229. Como resultado desta aquisição, a Compass passou a deter participação indireta em diversas Companhias Distribuidoras Locais, incluindo as Empresas-Alvo, por meio de sua participação direta na Commit.

A Termogás está adquirindo toda a participação da Commit nas Empresas-Alvo, mais especificamente:

- GASAP: 36.750 (trinta e seis mil, setecentas e cinquenta) ações ordinárias e 75.000 (setenta e cinco mil) ações preferenciais. Tais ações representam 24,5% do capital social votante e 37,3% do capital social total da GASAP.
- GASPIISA: 36.750 (trinta e seis mil, setecentas e cinquenta) ações ordinárias e 75.000 (setenta e cinco mil) preferenciais. Tais ações representam de 24,5% do capital social votante e 37,3% do capital social total da GASPIISA.
- GOIASGÁS: 480.351 (quatrocentos e oitenta mil, trezentos e cinquenta e uma) ações ordinárias e 1.770.619 (um milhão, setecentos e setenta mil, seiscentos e dezenove) ações preferenciais. Tais ações representam de 19,5% do capital social votante e 30,46% do capital social total da GOIASGÁS
- RONGÁS: 558.446 (quinhentos e cinquenta e oito mil, quatrocentas e quarenta e seis) ações ordinárias e 2.279.370 (dois milhões, duzentos e setenta e nove mil, trezentas e setenta) ações preferenciais. Tais ações representam 24,5% do capital social votante e 41,5% do capital social total da RONGÁS.

Ademais, como já informado, a Termogás destaca que o contrato firmado entre ela e a Compass, posteriormente cedido pela Compass à Commit, também formaliza a aquisição de 24% do capital social total da Companhia Brasileira de Gás - CEBGÁS. Em 2021 a Termogás teve a oportunidade de exercer o direito de preferência para aquisição da participação - à época - da Gaspetro na CEBGÁS exatamente em função da operação de venda da Gaspetro (atual Commit) para a Compass.

O exercício do direito de preferência da Termogás em relação a essa distribuidora foi notificado ao CADE (vide Ato de Concentração nº 08700.005921/2021-16) e aprovada sem restrições em 24 de novembro de 2021. Dessa forma, as Partes requerem dispensa de apresentar as informações referentes a CEBGÁS na presente notificação, porém permanecem à disposição deste E. Cade para prestar os esclarecimentos que se façam necessários.

A Operação representa, para a Termogás, a oportunidade de consolidar a sua participação em distribuidoras locais de gás natural e, assim, poder colaborar para que elas possam iniciar a prestação do serviço público de distribuição de gás canalizado e atividades correlatas.

A Operação não contempla cláusulas restritivas à concorrência.

O Grupo Termogás é responsável por diversos empreendimentos nos setores de energia, infraestrutura e gás natural.

Para fins da presente Operação, são relevantes as atividades do Grupo Termogás no setor de gás natural. Além das Empresas-Alvo, o Grupo Termogás, através da Termogás, possui participação societária nas seguintes empresas operacionais atuantes na distribuição de gás natural: (i) Companhia Maranhense de Gás – GASMAR, (ii) Companhia de Gás do Amazonas – CIGÁS 12 e (iii) Companhia Brasileira de Gás – CEBGÁS.

O Grupo Termogás detém ainda participação nas seguintes empresas que têm como objeto social o transporte de gás: 10%.

Além de sua atuação no setor de gás natural, conforme indicado no item II.9., existem empresas do Grupo Termogás que atuam no setor de geração de energia elétrica. A atuação do Grupo Termogás no setor é focada principalmente em usinas hidrelétricas (PCH) e energia eólica. Não obstante, o Grupo Termogás informa que dentre seus investimentos em geração de energia elétrica existem apenas duas usinas termelétricas, operadas pela Termo Norte Energia, localizadas em Porto Velho/RO que estão aptas a utilizar óleo e gás natural.

A Companhia de Gás do Amapá – GASAP é uma sociedade de economia mista, autorizada pela Lei Estadual nº 705 de 05 de julho de 2002. A companhia tem por objetivo a exploração, com exclusividade, do serviço público de distribuição de gás canalizado, podendo também explorar outras formas de distribuição de gás natural ou manufaturado, de produção no estado do Amapá. A companhia não é operacional.

A Companhia de Gás do Piauí – GASPISA é uma sociedade de economia mista, criada por deliberação do Governo do Estado do Piauí através da Lei Estadual nº 5.192, de 25 de maio de 2001. A companhia detém os direitos de exploração, com exclusividade, do serviço de distribuição de gás canalizado, podendo também explorar outras formas de distribuição de gás natural e manufaturado, em todo o território do Estado do Piauí. Não obstante, a companhia não é operacional.

A Agência Goiana de Gás Canalizado S/A – GOIASGÁS, foi criada por meio da Lei Estadual nº 13.641/2000, dando um largo passo rumo ao desenvolvimento consciente e a valorização e preservação de nossas riquezas naturais. Ela é uma empresa de capital misto que tem por objeto social a exploração e distribuição de gás natural canalizado, e de outras origens, por vias terrestres e fluviais; além de desenvolver atividades correlatas necessárias à distribuição do gás em todo o segmento consumidor para a geração termelétrica ou outras finalidades possibilitadas pelos avanços tecnológicos.

A Companhia Rondoniense de Gás – RONGÁS é uma sociedade de economia mista criada pela Lei Estadual nº 728, de 14 de julho de 1997. Seu objetivo social é promover a produção, importação, aquisição, armazenamento, distribuição, comercialização, transporte, e a prestação de serviços correlatos na área de gás natural, seja para fins de matéria prima, geração de energia elétrica, combustível além de outras finalidade e usos que os avanços tecnológicos permitirem. A companhia não é operacional.

A Commit é uma holding que detém participação em distribuidoras de gás natural em diversos Estados brasileiros, quais sejam, além das Empresas-Alvo, Rio Grande do Sul (Sulgás), Florianópolis (Segás), Paraná (Compagás), Mato Grosso do Sul (Msgás), São Paulo (GasBrasiliano), Rio de Janeiro (CEG Rio), Alagoas (Algás), Ceará (Cegás), Rio Grande do Norte (Potigás), Sergipe (Sergás) e Pernambuco (Copergás).

A Compass tem operações nacionalmente em comercialização de energia elétrica e na prestação de serviços locais de gás canalizado em parte do território do Estado de SP por meio da Comgás e no Rio Grande do Sul, por meio da Sulgás, além de outras atividades em estágio pré-operacional ou de projeto. Com a aquisição da Commit, a Compass passou a ser titular de participação societária indireta em sociedades que atuam na distribuição de gás canalizado em diversos Estados do Brasil, incluindo as Empresas-Alvo.

Ademais, o Grupo Cosan, para além das atividades de sua subsidiária Compass, desenvolve atividades relacionadas a: produção e distribuição de óleos lubrificantes e especialidades; geração e comercialização de energia elétrica; e logística ferroviária e portuária. Além disso, o Grupo Cosan, por meio de subsidiárias vinculadas ao Grupo Raízen, também atua em distribuição de combustíveis líquidos e em produção e exportação de açúcar, etanol, bioenergia e biogás.

Conforme mencionado acima, a Operação se refere a aquisição de participações societárias nas distribuidoras locais de gás natural GASAP, GASPISA, GOIASGÁS e RONGÁS, atualmente detidas pela Commit. A Operação diz respeito, portanto, às atividades de distribuição de gás natural nos Estados do Amapá, Piauí, Goiás e de Rondônia.

Tendo em vista que as Empresas-Alvo não exercem qualquer tipo de atividade econômica atualmente, não há que se falar em sobreposição ou integração decorrente da aquisição de participação societária da Termogás nas referidas distribuidoras.

Nesse sentido, a Termogás informa que as distribuidoras operacionais nas quais possui participação acionária, quais sejam CEBGÁS, GASMAR e CIGÁS, não atuam nos Estados nos quais as Empresas-Alvo foram constituídas. As referidas distribuidoras operacionais atuam no Distrito Federal, Maranhão e Amazonas, respectivamente.

Não obstante a ausência efeitos concorrenciais da Operação, para fins completude e com o intuito de apresentar as melhores informações disponíveis, as Partes fornecem a seguir uma breve explicação sobre a cadeia do gás natural e a jurisprudência deste E. Cade a respeito das atividades de distribuição do produto.

O gás natural é utilizado como combustível: na geração de energia elétrica em usinas termoelétricas, em veículos de transporte e geração de calor em processos diversos, comerciais, residenciais e industriais. O produto também é utilizado como matéria-prima de certas indústrias, especialmente na obtenção de produtos petroquímicos, embora em menor escala.

A cadeia do gás natural no Brasil pode ser dividida em três segmentos principais: upstream, midstream e downstream. O nível upstream abrange desde a exploração de áreas viáveis de extração até a produção e o escoamento para refinarias, ou, alternativamente, para importação. O midstream compreende o processamento e o transporte em dutos de alta pressão. Por fim, o downstream consiste na distribuição local de gás através de dutos de média e baixa pressão ou de caminhões de GNL ou GNC, até os clientes finais – consumidores residenciais, comerciais e industriais, ou usinas termoelétricas.

Destaca-se que a Superintendência-Geral do CADE (“SG”) já analisou o mercado downstream de distribuição de gás natural em diversas ocasiões¹⁴, afirmando que não há necessidade de subdivisão do mercado relevante na sua dimensão produto – sendo este o gás natural – contudo, reconhecendo que existem diferenças entre os modais de distribuição do produto que pode ser feita de duas formas: por meio de dutos locais, de média e baixa pressão, conectados diretamente aos dutos de transporte; ou a granel, através de veículos como trens, navios e caminhões dotados de tanques onde o gás é armazenado e transportado até a destinação final.

Tendo em vista que não há sobreposição horizontal decorrente da presente Operação e, como será visto abaixo, a única integração vertical entre as atividades das Partes é préoperacional, as Partes entendem que a definição precisa de mercado relevante em sua dimensão produto pode ser deixada em aberto.

Segundo jurisprudência do CADE, para compreender a dimensão geográfica da distribuição do gás natural, é necessário entender as diferenças entre os modais de distribuição e o efeito concorrencial destas diferenças.

Nesse sentido, em precedentes anteriores, o CADE entendeu que a distribuição por dutos é considerada um monopólio natural, em razão dos altos custos fixos e afundados envolvidos na instalação de dutos até determinada localidade e dos baixos custos marginais necessários para conectar clientes nas imediações do ramal principal. Por outro lado, a distribuição a granel pressupõe a transformação do gás natural por meio de processo de liquefação ou de compressão (GNL e GNC). Na ausência de malha dutoviária, o GNC é considerado a melhor opção quando do transporte de pequenos volumes de gás natural a curtas distâncias e o GNL, por sua vez, é mais competitivo no transporte de gás natural em maiores volumes e longas distâncias.

Com relação à competição entre duto e granel, o CADE afirmou que “dada a estrutura de custos inerente ao processamento para viabilizar a distribuição, o gás distribuído a granel (GNC e GNL) normalmente só atua nas localidades em que a distribuidora local de gás canalizado não atua. Assim, onde não há duto, há espaço para que o granel dispute mercado, mas, uma vez que a companhia local de distribuição por duto realize o investimento para conectar determinada localidade à malha, o gás a granel deixa de ser competitivo”.

Em um cenário ainda mais conservador, seria possível concluir que há competição potencial entre distribuição canalizada e distribuição a granel para clientes ainda não conectados à malha de distribuição (i.e., localidades fora da zona de monopólio natural) e cuja demanda individualmente justificaria a expansão da malha, pois somente essas localidades estariam inseridas na competição pelo mercado.

Tendo em vista que não há sobreposição horizontal decorrente da presente Operação e a única integração vertical entre as atividades das Partes é pré-operacional, as Partes entendem que a definição precisa de mercado relevante em sua dimensão produto pode ser deixada em aberto.

Conforme demonstrado acima, a distribuição por dutos é considerada um monopólio natural, em razão dos altos custos fixos e afundados envolvidos na instalação de dutos até determinada localidade e dos baixos custos marginais necessários para conectar clientes nas imediações da malha dutoviária. Adicionalmente, em conformidade com a Constituição Federal, a distribuição de gás canalizado é um serviço local de gás, considerado monopólio dos estados, os quais podem explorá-lo diretamente ou mediante concessão (art. 25, § 2º da Constituição Federal).

Em relação aos monopólios naturais, a SG recentemente reconheceu que a operação entre a Transmissora Delmiro Gouveia S/A, ATP Engenharia Ltda. e a Companhia Hidroelétrica do São Francisco não geraria qualquer sobreposição horizontal, considerando a restrição da atividade de transmissão de energia elétrica à área concedida, por tratar-se de monopólio natural e setor regulado pela ANEEL e objeto de concessão.

Este também foi um entendimento reforçado no ato de concentração entre Construtora Remo Ltda., Selt Engenharia Ltda., Construtora Barbosa de Mello S.A. e BH Iluminação Pública S.A., aprovado sem restrições no dia 11 de maio de 2020. Naquela ocasião, a SG entendeu que se tratava de concessão de serviços públicos (serviços de iluminação pública), cuja dimensão do mercado relevante, na jurisprudência do CADE, é delimitada de forma a considerar cada concessão como um mercado relevante específico, configurando um monopólio natural, em que a concorrência se dá pelo mercado e não no mercado. Desse modo, concluiu-se que aquela operação não resultaria em qualquer sobreposição horizontal.

As Partes entendem que o mesmo raciocínio se aplica à atividade de distribuição de gás natural, ou seja, dado que esta é uma atividade caracterizada pelo monopólio natural, não há concorrência no mercado, que é restrito à área de alcance da malha dutoviária. Como informado, a Operação não gera qualquer sobreposição haja vista que as Empresas-Alvo não são ativas. Ainda que fossem, conforme explicado anteriormente, a distribuição de gás natural via dutos é um monopólio natural. Portanto, não há que se falar em concorrência durante a vigência da concessão do Estado para a prestação do serviço de distribuição de gás natural via dutos

No hipotético cenário em que os demais modais de distribuição concorressem com a possibilidade de expansão dos dutos, portanto dilatando o escopo produto do mercado relevante, ainda assim não seria verificada qualquer sobreposição, uma vez que o Grupo Termogás não atua na distribuição de gás natural através de outros modais nos Estados nos quais as Empresas-Alvo foram constituídas e estão autorizadas a operar, quais sejam Amapá, Piauí, Rondônia e Goiás.

Em vista de todo o exposto, resta clara a ausência de sobreposição horizontal decorrente da Operação e a baixa participação das distribuidoras nas quais a Termogás é acionistas na demanda nacional por gás natural.

O fato das Empresas-Alvo não terem atividades econômicas seria suficiente para descartar quaisquer integrações decorrentes da Operação. Não obstante, a Termogás apresenta a seguir informações que complementam e robustecem tal conclusão.

Em primeiro lugar, nota-se que com exceção da participação nas Empresas-Alvo, nem a Termogás, tampouco o Grupo Termogás, atuam na cadeia do gás natural nos Estados do Amapá, Piauí, Goiás e Rondônia.

Portanto, resta claro que os demais investimentos do Grupo Termogás na cadeia de gás não resultam em qualquer integração vertical com as atividades de distribuição de gás via dutos as quais as Empresas-Alvo estão autorizadas a realizar.

Por fim, a Termogás esclarece que além da atuação no setor de gás natural, conforme indicado no item II.9., o Grupo Termogás atua também no setor de geração de energia elétrica. A atuação do Grupo Termogás no setor é focada principalmente em usinas hidrelétricas (PCH) e energia eólica. Todavia, o Grupo Termogás informa que dentre seus investimentos em geração de energia elétrica existem duas usinas termelétricas, as quais estão localizadas em Porto Velho/RO, aptas a utilizar óleo e gás natural.

Muito embora tais termoelétricas estejam localizadas em Rondônia, mesmo estado no qual a RONGÁS foi constituída, tal distribuidora não é operacional. Logo, não há que se falar em possível abastecimento dos empreendimentos pela RONGÁS.

A Operação não suscita preocupações concorrenciais, uma vez que: (i) a Termogás já detém participação nas Empresas-Alvo; e (ii) nenhuma das Empresas-Alvo se encontra operacional de modo que não é possível verificar qualquer sobreposição ou integração decorrente da proposta aquisição de participação societária nas Empresas-Alvo. A Operação é, portanto, benéfica a concorrência no setor de gás natural.

Com base no exposto, as Partes respeitosamente solicitam a aprovação sem restrições e sob o rito sumário da Operação.