

AUDIT.AI: UM SISTEMA DE INTELIGÊNCIA ARTIFICIAL PARA FORTALECIMENTO DA GOVERNANÇA E COMBATE À CORRUPÇÃO

Giselle da Costa Batista^{1,2}, Ruan Vieira Rolim de Sá¹, Pedro Bacelar Moreira¹, Marcela Alves de Souza^{1,2}, Hugo Pereira Kuribayashi^{1,2}.

¹ ManivaLab - Grupo de Pesquisa em Transformação Digital na Amazônia Sul Oriental
Universidade Federal do Sul e Sudeste do Pará (Unifesspa) - Marabá, PA – Brasil

² Programa de Pós-Graduação em Ciências Forenses
Universidade Federal do Sul e Sudeste do Pará (Unifesspa) - Marabá, PA – Brasil

{ruan.vieira,gisellebatista,pedrobacelar}@unifesspa.edu.br
{marcela.alves,hugo}@unifesspa.edu.br

RESUMO

Este estudo apresenta o Audit.AI, um sistema de inteligência artificial desenvolvido para aprimorar a governança pública e a transparência institucional. O objetivo do trabalho é analisar e identificar padrões em Processos Administrativos Disciplinares (PADs) de Instituições Federais de Ensino Superior (IFES) para otimizar a prevenção e detecção de irregularidades. Para isso, foram solicitados PADs concluídos entre 2015 e 2022, possibilitando a criação de um *dataset* analisado com técnicas de Processamento de Linguagem Natural (PLN) e aprendizado de máquina para extrair informações relevantes e gerar sumarizações automatizadas. Os resultados indicam que o Audit.AI pode apoiar a tomada de decisões em auditorias, mitigar riscos institucionais e fortalecer práticas de integridade no setor público. Conclui-se que o sistema apresenta potencial de aplicação em outras instituições da administração pública, desde que haja suporte governamental e institucional adequado.

Palavras-chave: Inteligência artificial. Governança pública. Processamento de linguagem natural. Detecção de padrões. Combate à corrupção.

1. INTRODUÇÃO

Segundo o *United Nations Office on Drugs and Crime* – UNODC (2024), a corrupção é um fenômeno social, político e econômico complexo que afeta todos os países do mundo. Em diferentes contextos, ela prejudica as instituições democráticas, freia o desenvolvimento econômico e contribui para a instabilidade política. Os impactos sociais da corrupção em larga escala a tornam uma preocupação internacional (Transparência internacional, 2025a). Estudos indicam que países com baixos índices de controle da corrupção tendem a apresentar também deficiências na qualidade da administração pública (Topchii et al., 2021).

Diante da relevância internacional deste tema, as políticas de prevenção e combate à corrupção devem priorizar a redução de espaços de impunidade e correção de falhas institucionais. Para isso, é fundamental o aprimoramento contínuo da capacidade estatal de detecção de ilícitos (BRASIL, 2024, p. 11).

O combate à corrupção e o fortalecimento das instituições públicas são prioridades globais, refletidas na Agenda 2030 da Organização das Nações Unidas (ONU), que estabeleceu o Objetivo de Desenvolvimento Sustentável (ODS) 16 – Paz,

Justiça e Instituições Eficazes (ONU, 2015). Esse objetivo estabelece metas como a redução substancial da corrupção (meta 16.5) e o fortalecimento da transparência e da responsabilidade institucional. A inclusão desse tema entre os ODS evidencia seu impacto na governança pública, tornando essencial a implementação de mecanismos que aprimorem a fiscalização, a integridade administrativa e a responsabilização de agentes públicos.

A Transparência Internacional é um movimento global que criou o Índice de Percepção da Corrupção (IPC), considerado o principal indicador de corrupção do mundo, que avalia 180 países desde 1995. Esse índice atribui pontuações de 0 a 100, onde notas mais altas indicam uma percepção maior de integridade do país. (Transparência Internacional, 2025b)

A ausência de medidas anticorrupção eficazes nas Américas tem promovido a impunidade e ampliado a influência de elites e grupos criminosos na definição de políticas, fragilizando as instituições democráticas e os direitos da população. Em 2024, o Brasil registrou 34 pontos e a 107ª posição, entre 180 países, no Índice de Percepção da Corrupção. Estas são a pior nota e a pior colocação do país na série histórica do índice, iniciada em 2012 (Transparência Internacional, 2025b).

Esta pesquisa está sendo conduzida em Instituições Federais de Ensino Superior (IFES). É reconhecido que para a busca por soluções eficazes, é necessário um olhar atento para as especificidades de cada contexto em que a corrupção ocorre, visto que suas causas variam de um lugar para outro, dependendo de fatores culturais, históricos, econômicos e políticos, além da estrutura e qualidade das instituições e regramentos (BRASIL, 2024). O banco de dados utilizado na construção do produto desta pesquisa foi criado especificamente para este fim, uma vez que as ações formuladas para promover mudanças sejam o máximo possível baseadas em situações concretas, permitindo análises contextualizadas que identifiquem causas, fatores facilitadores, áreas de maior vulnerabilidade e consequências das práticas corruptas, garantindo que as intervenções sejam mais direcionadas e eficientes, conforme indicado por BRASIL (2024). Assim, o *dataset* é composto por Processos Administrativos Disciplinares (PADs) que resultaram em expulsões de agentes públicos vinculados às IFES, no período de 2015 a 2022. No período mencionado, a CGU informa o total de 228 aplicações de penalidades com essas características.

Esse banco de dados serve como base para o Audit.AI, um sistema de inteligência artificial voltado ao aprimoramento das ferramentas de prevenção e detecção de ilícitos. Diante deste cenário, este artigo apresenta o Audit.AI como uma iniciativa inovadora que alia análise histórica e aprendizado contínuo para fortalecer a gestão da integridade e mitigar riscos. O estudo evidencia suas funcionalidades iniciais de sumarização e análise de PADs, além de delinear os próximos passos para seu aprimoramento e aplicação no combate à corrupção no serviço público federal.

O artigo está estruturado da seguinte forma: a Seção 2 apresenta a revisão da literatura; a Seção 3 detalha a metodologia aplicada; a Seção 4 discute os resultados esperados; a Seção 5 aborda os desafios e limitações da implementação; e a Seção 6 traz as considerações finais e perspectivas futuras.

2. REVISÃO DA LITERATURA

A literatura apresenta diferentes abordagens para o enfrentamento desse problema, destacando a importância da transparência, da governança eficiente e do uso de tecnologias avançadas para aprimorar a fiscalização e a detecção de irregularidades.

Esta seção aborda estudos sobre corrupção em instituições de ensino, o papel da inteligência artificial no controle interno e as abordagens existentes para a detecção de fraudes administrativas, além de discutir trabalhos correlatos que utilizam técnicas inovadoras para promover maior eficiência e transparência na gestão pública.

2.1. Estudos sobre Corrupção em Instituições de Ensino

A corrupção é amplamente reconhecida como um dos principais entraves ao desenvolvimento econômico e social, afetando organizações de diferentes naturezas e setores. Sua presença pode ser observada tanto no setor público quanto no privado, impactando negativamente a eficiência, a confiança pública e os resultados institucionais. Klitgaard (1998, apud Macedo e Valadares, 2021) define a corrupção como um fenômeno multidimensional que envolve diferentes tipos de organizações e coloca em risco a integridade das instituições. No Brasil, isso se manifesta de forma evidente, impactando várias áreas do setor público.

A corrupção nas instituições públicas é um problema recorrente, amplamente documentado por órgãos de controle e relatórios internacionais. Ela prejudica a transparência, enfraquece a confiança nas instituições e desestimula a participação popular em processos de fiscalização e controle social. Como destacam Santos e Ferrer Silva (2021, apud Barbosa et al., 2023), a transparência é um instrumento essencial no combate à corrupção e, quando ausente, a corrupção torna-se um inibidor da participação popular, ao ferir o processo democrático e a justiça social, o que leva o cidadão a perder a confiança nos governantes e nas instituições públicas, comprometendo o desenvolvimento local.

O setor educacional é considerado particularmente vulnerável à corrupção devido à sua complexa estrutura administrativa e ao alto volume de recursos que movimenta. Segundo a Controladoria-Geral da União (CGU, 2016, apud Rodrigues et al., 2020), essa área enfrenta desafios expressivos na fiscalização e no controle dos recursos públicos. No caso das Instituições Federais de Ensino Superior (IFES), essa preocupação se acentua ainda mais. Como entidades da administração indireta vinculadas ao Ministério da Educação (MEC), são responsáveis tanto pela gestão de consideráveis montantes financeiros quanto pela formação acadêmica e cívica dos estudantes. Diante disso, a adoção de elevados padrões de governança é essencial para garantir a integridade e a transparência de suas atividades.

Apesar da relevância do tema, há uma lacuna significativa na produção acadêmica sobre a corrupção nas IFES. Oliveira e Cruz (2021) destacam que a forma como o fenômeno da corrupção se manifesta no contexto da educação superior brasileira permanece pouco explorada na literatura nacional. Essa lacuna dificulta o desenvolvimento de soluções específicas que atendam às necessidades das IFES, especialmente no que diz respeito à compreensão, prevenção e combate à corrupção. Estudos recentes, como o de Rodrigues et al. (2020), defendem que a integração de políticas públicas eficazes com tecnologias avançadas, como inteligência artificial, pode fortalecer significativamente os mecanismos de controle e promover maior eficiência e transparência.

Somente com uma base acadêmica sólida e integrada às práticas de governança será possível combater efetivamente a corrupção nesse contexto, garantindo a integridade das instituições de ensino superior. Para isso, é essencial inovar e aprimorar continuamente as ferramentas de detecção e prevenção, utilizando tecnologias avançadas para fortalecer a transparência e a eficiência dos processos institucionais.

2.2. Tecnologias de Inteligência Artificial no Controle Interno

A automatização de processos no setor público é uma ferramenta indispensável para lidar com a complexidade e a amplitude das responsabilidades institucionais, especialmente diante da limitação de recursos. Nesse contexto, ferramentas baseadas em IA não apenas otimizam a análise de grandes volumes de dados, mas também promovem maior transparência e confiança nos processos administrativos. A utilização de tais tecnologias também pode ajudar a prevenir a corrupção, ao identificar padrões suspeitos de atividades ilegais antes que causem danos significativos (Gilson, 2023). Essa capacidade de identificar padrões complexos e gerar previsões torna o aprendizado de máquina uma ferramenta estratégica no fortalecimento da governança e na prevenção de irregularidades.

Sabemos que a auditoria governamental é uma atividade de fiscalização, cujo escopo é amplo e requer um esforço significativo dos agentes públicos para desempenhar suas competências. Por outro lado, os recursos disponíveis são limitados para o cumprimento dos deveres constitucionais (A. L. Silva et al., 2024). Assim, a incorporação de soluções tecnológicas pode contribuir consideravelmente para aumentar a eficiência operacional e aprimorar a capacidade de atender às demandas crescentes com maior eficácia.

O avanço da tecnologia tem transformado significativamente a forma como governos e instituições públicas enfrentam desafios complexos, como a corrupção. A IA, em especial, emergiu como uma ferramenta poderosa para promover transparência, eficiência e responsabilização no setor público. Técnicas como o Processamento de Linguagem Natural (PLN) e aprendizado de máquina permitem trabalhar com análises de grandes volumes de dados nos mais diversos processos, o que pode ser usado para diminuir brechas para práticas corruptas. Neste contexto, o uso de IA não apenas fortalece os mecanismos de controle interno, mas também contribui para restaurar a confiança pública e otimizar a gestão de recursos, tornando-se um aliado indispensável na busca por uma administração mais íntegra e eficiente.

A aplicação de modelos avançados, como o ChatGPT, demonstra um impacto significativo na eficiência operacional e na transparência de processos complexos, conforme destacado por E. C. Silva et al. (2024). Esses modelos permitem a extração e organização de informações críticas de forma eficiente, e, ao automatizar tarefas que antes dependiam exclusivamente de esforços humanos, eles reduzem o tempo necessário para análises detalhadas e aumentam a clareza dos dados disponíveis, promovendo maior confiança e agilidade nas operações. Tecnologia como a segmentação pode ser utilizada, pois é baseada em sentenças ou seções permite dividir textos extensos, como pareceres preliminares, em partes menores e mais manejáveis, facilitando a aplicação de algoritmos de sumarização. (Silva, M., 2024).

De acordo com Deisenroth, Faisal e Ong (2019, apud Gilson, 2023), o Aprendizado de Máquina (*Machine Learning*) é uma área que pode ser definida como o conjunto de métodos capazes de detectar automaticamente padrões nos dados e, a partir desses padrões, prever dados futuros ou realizar outros tipos de decisões sob condições de incerteza. Essa definição enfatiza a relevância do aprendizado de máquina em contextos que exigem análise preditiva e decisões informadas, especialmente em ambientes complexos.

2.3. Abordagens Existentes para Detecção de Fraudes Administrativas

Quando o assunto é combate à fraude e corrupção temos vários órgãos, com papéis diversos atuando nesse sentido. Aqui vamos destacar dois protagonistas em questão de regulação no executivo federal, uma vez que nosso grupo amostral para construção do banco de dados e planejamento de caso de uso são as IFES. Assim, os principais órgãos reguladores são a Controladoria-Geral da União (CGU) e o Tribunal de Contas da União (TCU).

A CGU exerce a função de controle interno do Governo Federal, com a responsabilidade de proteger o patrimônio público e promover a transparência na gestão. Suas atividades abrangem auditorias públicas, correições, ações preventivas e de combate à corrupção, além de atuar como ouvidoria (Controladoria-Geral da União, 2021). Já o TCU exerce o controle externo do governo federal, com a meta de promover uma administração pública eficiente e responsável, fiscalizando aspectos contábeis, financeiros, orçamentários, operacionais e patrimoniais, sempre observando a legalidade, legitimidade e economicidade (TRIBUNAL DE CONTAS DA UNIÃO, [s.d.]).

Atualmente, o documento governamental que orienta sobre assuntos relacionados à fraude e corrupção foi elaborado pelo Governo Federal, no Comitê Interministerial de Combate à Corrupção, sendo aplicável a órgãos e entidades da administração pública (BRASIL, 2024). O Plano de Integridade e Combate à Corrupção 2025 - 2027 estabelece, entre suas premissas, a necessidade de uma abordagem contextualizada e baseada em situações concretas para a promoção da integridade e o combate à corrupção.

Essa diretriz evidencia a importância de construir um dataset específico para esta pesquisa, garantindo que as análises sejam fundamentadas em dados concretos e representativos.

Hoje, todos os órgãos do Executivo Federal seguem às orientações da CGU quanto à construção do Plano de Integridade, que é obrigatório em todas as instituições por força do Decreto nº 9.203, de 22 de novembro de 2017:

Art. 19. Os órgãos e as entidades da administração direta, autárquica e fundacional instituirão programa de integridade, com o objetivo de promover a adoção de medidas e ações institucionais destinadas à prevenção, à detecção, à punição e à remediação de fraudes e atos de corrupção, estruturado nos seguintes eixos:

I - Comprometimento e apoio da alta administração;

II - Existência de unidade responsável pela implementação no órgão ou na entidade;

III - Análise, avaliação e gestão dos riscos associados ao tema da integridade;

e

IV - Monitoramento contínuo dos atributos do programa de integridade. (BRASIL, 2017)

Como parte do Programa de Integridade, as universidades e institutos federais de ensino, elaboram e aprovam o Plano de Integridade, documento este que detalham a execução e resultados quanto aos incisos III e IV mencionados no Decreto nº 9.203.

Hoje, as ferramentas utilizadas para atender a essas demandas consistem, em sua maioria, em planilhas eletrônicas extensas, cujo conteúdo é generalista e atualizado periodicamente pelos gestores das unidades que compõem a estrutura das IFES. Além disso, a detecção e a mitigação de riscos relacionados ao controle interno das instituições podem ser realizadas por meio de auditorias, que podem ser solicitadas pela CGU ou iniciadas pela própria unidade de auditoria interna, como no

caso das chamadas “auditorias especiais”. Nesse contexto, a implementação de uma ferramenta específica, capaz de direcionar e otimizar esses processos, tem o potencial de aumentar significativamente a eficiência das ações de integridade e o sucesso das auditorias realizadas.

2.4. Trabalhos Correlatos

Os avanços recentes no uso de IA no setor público têm gerado soluções promissoras em áreas como auditorias, processamento de documentos e identificação de irregularidades administrativas. Trabalhos correlatos têm explorado abordagens inovadoras baseadas em IA, com destaque para aplicação de técnicas como PLN, aprendizado de máquina e Grandes Modelos de Linguagem (GML).

Os autores Silva, Medeiros, Menezes e Kamikawachi (2024) exploram segmentação e sumarização automatizada para extrair informações de editais governamentais, trazendo eficiência ao processamento de dados relacionados a equipamentos de TI. Os autores aplicam técnicas de PLN, incluindo segmentação e sumarização de textos, utilizando modelos como o GPT-3.5, para extrair informações relevantes sobre produtos, quantidades e valores em editais de licitação de produtos tecnológicos. Os resultados obtidos ser promissora a metodologia proposta, evidenciando uma melhoria significativa na precisão e na eficiência da extração de dados críticos, facilitando assim a tomada de decisões informadas no contexto das licitações públicas.

Já o estudo de M. Silva et al (2024) utiliza LLMs e PLN para simplificar a compreensão de documentos oficiais, promovendo acessibilidade e transparência. Os autores propõem o uso do modelo GPT-4, para a simplificação e sumarização automática de opiniões preliminares emitidas pelo Tribunal de Contas do Estado de Pernambuco, visando facilitar o acesso e a compreensão por parte da população. Os resultados obtidos demonstram que os textos simplificados apresentam uma linguagem acessível, destacando os pontos principais das decisões, como valores monetários e percentuais, o que contribui para a promoção da transparência e do exercício da cidadania.

No contexto das auditorias, o trabalho de A. L. Silva et al. (2024) aborda a análise de editais com foco na precificação de medicamentos, otimizando processos específicos por meio de IA. O trabalho contempla a implementação de um sistema automatizado que utiliza algoritmos de agrupamento, como o K-Means, para analisar e comparar preços de medicamentos a partir de Notas Fiscais Eletrônicas (NF-e) e dados da Câmara de Regulação do Mercado de Medicamentos (CMED). Os resultados obtidos demonstraram uma taxa de sucesso de 83% na avaliação correta dos preços de referência de 150 itens de um edital, evidenciando a eficácia da solução proposta na identificação de discrepâncias de preços no contexto de licitações públicas.

Outro exemplo é o estudo de Gilson e Bramili (2023) que aplica ferramentas de IA para detectar padrões associados à corrupção em corregedorias municipais. Em particular, a abordagem técnica do trabalho consiste na implementação de algoritmos de aprendizado de máquina para a detecção de padrões de fraude e corrupção em dados financeiros. Os algoritmos utilizados no projeto incluem técnicas de aprendizado supervisionado e não supervisionado, como regressão logística e clusterização, para identificar tipologias de irregularidades e padrões de comportamento em dados financeiros, permitindo a detecção de anomalias associadas à corrupção. Os resultados preliminares indicam um desempenho

promissor na identificação de riscos, evidenciando a eficácia da inteligência artificial como ferramenta de apoio à governança pública.

O Quadro 1 apresenta os principais trabalhos correlatos, destacando suas semelhanças e diferenças em relação ao presente estudo.

Quadro 1 - Trabalhos Correlatos

Artigo	Semelhança	Diferença
<i>Segmentation and Summarization for Extracting Information about Information Technology Equipment from Government Procurement Notice</i>	Processamento de linguagem natural e aprendizado de máquina para segmentação e sumarização automatizada de textos	Uso não relacionado ao combate à corrupção
<i>Using Generative AI for Simplifying Official Documents in the Public Accounts Domain</i>	Uso de GML, PLN, segmentação e sumarização	Embora possa ser usado para combater a corrupção, não é focado em observar padrões
Ferramenta para Auxílio à Auditoria de Editais Municipais para Compra de Medicamentos	Utilizado pelo setor de auditoria	Relacionado a precificação de produtos em processos de licitação
Inteligência Artificial no Combate à Fraude e Corrupção: A Experiência da Controladoria Geral do Município do Rio de Janeiro	Detecção de casos de corrupção através de padrões e anomalias	Aplicado em corregedoria municipal

Fonte: Elaborado pelos autores (2025).

Neste contexto promissor, o presente trabalho vem somar esforços, trazendo uma iniciativa voltada inicialmente para o fortalecimento da governança e a promoção da transparência nas IFES. O sistema Audit.AI combina técnicas de PLN, aprendizado de máquina, segmentação e análise de padrões para identificar características associadas à corrupção em PADs. Com sua flexibilidade, o Audit.AI possui um potencial de aplicação que vai além das IFES, podendo ser adaptado para diferentes instituições e contextos no setor público, ampliando significativamente seu potencial impacto na sociedade.

Este sistema busca fortalecer o papel da IA como uma aliada indispensável na busca por uma administração pública mais ética, eficiente e transparente. Seu caráter versátil demonstra que a tecnologia pode ser um catalisador para avanços em diferentes níveis de gestão pública, consolidando-se como uma ferramenta essencial para o combate à corrupção e o fortalecimento da integridade institucional.

3. METODOLOGIA

Figura 1 - Fluxo da Metodologia



Fonte: Elaborado pelos autores (2025).

3.1. Definição do Escopo da Pesquisa

Reconhece-se que a aplicação dessas tecnologias exige um planejamento cuidadoso, especialmente no que diz respeito à qualidade e integridade dos dados utilizados. Conforme destacado por Gilson e Bramili (2023), um dos desafios mais relevantes de qualquer projeto de IA com certeza reside nas bases de dados que são utilizadas. Neste contexto, garantir que os dados sejam representativos e confiáveis é essencial para que as soluções de IA cumpram seu papel de maneira eficaz e segura, potencializando seus benefícios no fortalecimento das instituições públicas.

Portanto, a criação de um banco de dados específico foi essencial para parametrizar o Audit.AI, garantindo que o sistema fosse treinado de maneira adequada para suas funções. Para isso, foi necessário definir previamente o escopo da pesquisa e as principais características que o banco deveria conter. Desta maneira, entre as principais informações coletadas destacam-se:

1. Tipo de informação coletada: Processos Administrativos Disciplinares que resultaram em expulsões de servidores por corrupção;
2. Tipo de instituição: PADs de universidades e institutos federais de ensino. Essas instituições são regidas pelos mesmos órgãos reguladores, vinculadas ao mesmo ministério, e possuem semelhantes finalidades, portanto compartilham de estruturas administrativas, carreiras e processos de trabalho correlatos. Então, devido às similaridades entre essas instituições, consideramos como um só grupo;
3. Lapso temporal das informações coletadas: 2015 a 2022, pois a partir da publicação do Decreto nº 8.539/2015 em 2015, a utilização de processos eletrônicos no âmbito dos órgãos e das entidades da administração pública federal direta, autárquica e fundacional tornou-se obrigatória. (BRASIL, 2015);

Essa mudança não apenas trouxe consigo uma nova obrigatoriedade, mas também uma série de benefícios, tais como eficiência operacional, transparência e economia de recursos. Diante desse contexto, a escolha de analisar os dados a partir

de 2015 é crucial, ao entender que as instituições teriam menos dificuldade em localizar e encaminhar processos que estão em formato eletrônico, torna-se evidente que essa data marca um ponto de partida significativo para nossa pesquisa.

Escolhemos como fim do período analisado o ano de 2022, garantindo que os processos concluídos até este ano estão arquivados internamente e no CGU-PAD, não cabendo mais edições, excetuando se houver alguma judicialização posterior.

3.2. Criação do Dataset

Após a definição do escopo, o procedimento de coleta e aquisição dos dados foi realizado por meio de uma pesquisa documental, seguindo a ordem descrita a seguir para a criação do *dataset*.

- a) Localização dos documentos;
- b) Solicitação das informações (incluindo acompanhamento e recursos necessários);
- c) Pré-processamento dos dados.

3.2.1. Localização

Os dados de correição do Executivo Federal estão disponíveis no Painel Correição em Dados da CGU (CONTROLADORIA-GERAL DA UNIÃO, [s.d.]). Como se trata de um painel dinâmico, com números que podem ser atualizados diariamente, informamos que os dados sobre a quantidade de casos em cada instituição foram obtidos na pesquisa realizada em 11 de março de 2024.

Iniciamos buscando a quantidade e as instituições que possuíram casos de servidores punidos com sanções expulsivas em IFES, em PADs concluídos entre os anos de 2015 e 2022, conforme mencionado. No Painel da CGU aplicamos os seguintes filtros disponíveis para obter as informações necessárias:

- a) Poder Executivo;
- b) Esfera Federal;
- c) Descrição Grupo Fundamento é Corrupção (4878-43-LXI e IX, 8112-117-IX, XII, XIII e XVI, 8112-132-IV, X e XI);
- d) Descrição Natureza Sanção é Expulsiva;
- e) Descrição Tipo Pessoa é Agente Público;
- f) Descrição Ministério é Ministério da Educação;
- g) Ano 2015 a 2022.

Essa pesquisa identificou um total de 228 casos distribuídos em 76 IFES. Com essas informações, que detalham a quantidade de sanções e os locais onde ocorreram, foi possível avançar para a etapa de solicitação dos PADs.

3.2.2. Coleta de Documentos

As solicitações foram realizadas com base na Lei nº 12.527, de 18 de novembro de 2011, que regula o acesso à informação pública no Brasil (BRASIL, 2011). Encaminhamos os pedidos de acesso à informação por meio do Fala.Br, uma plataforma integrada de Ouvidoria e Acesso à Informação gerida pela CGU, que promove a transparência administrativa. Dado que os PADs são documentos extensos e detalhados, contendo desde denúncias, documentos comprobatórios e

depoimentos, até defesas, recursos, relatórios e análises externas, foi solicitado, na impossibilidade de envio do processo completo, que fossem disponibilizadas ao menos as peças essenciais para a análise. Essas peças incluem o documento que motivou a investigação, o relatório final da comissão do PAD e o parecer da procuradoria federal vinculada à instituição, uma vez que esses documentos são elaborados de forma detalhada e apresentam informações suficientes para compreender os principais aspectos de cada caso.

3.2.3. Pré-processamento dos Dados

A partir da solicitação de informações realizada conforme detalhado na seção anterior, recebemos muitos processos que culminaram na expulsão de agentes públicos. Esta coleta resultou em um volume significativo de dados, conforme esperado. As respostas vieram por diversos meios: diretamente na plataforma Fala.BR, via e-mail e disponibilização em nuvem. Algumas instituições mandaram processos na íntegra, outras mandaram peças separadas dos processos, pois na solicitação também colocamos esta opção. Então foi necessário organizar estes dados, seguindo estas etapas:

- a) Download e Centralização: Todos os documentos recebidos foram baixados e concentrados em uma única pasta na nuvem, identificada com as siglas das instituições que enviaram as informações.
- b) Verificação Inicial: Em seguida, todos os documentos foram abertos para verificar o conteúdo enviado. Nos casos em que as peças estavam separadas, foi verificado se incluíam todos os itens solicitados, como o relatório final e o parecer da procuradoria.
- c) Separação dos Documentos: Para os processos completos, foram separados apenas os documentos específicos (relatório final e parecer da procuradoria).
- d) Organização dos Dados Pré-Processados: Os documentos pré-processados, incluindo o relatório final da comissão de PAD e o parecer da procuradoria de cada processo, foram organizados em outra pasta na nuvem, divididos por instituição.
- e) Ajustes e complementos: Pedidos de recursos ou novas solicitações de informações foram enviadas às instituições que não haviam enviado todos os documentos necessários, na tentativa de garantir que os dados estivessem completos para a análise.

Essas etapas visaram garantir a organização e a completude dos dados, preparando-os para a análise quali-quantitativa subsequente. Além disso, as tarefas de pré-processamento dos dados coletados incluíram não apenas a organização e verificação das informações, mas também a transformação dos dados utilizando *embeddings*, que são representações vetoriais de palavras ou frases em um espaço multidimensional. Essa técnica permite capturar o significado semântico das informações, facilitando a análise e a interpretação dos dados. Ao aplicar *embeddings*, conseguimos melhorar a qualidade da análise, extraindo contexto semânticos e tornando-a mais eficiente, o que é fundamental para a identificação de padrões relevantes nos dados analisados.

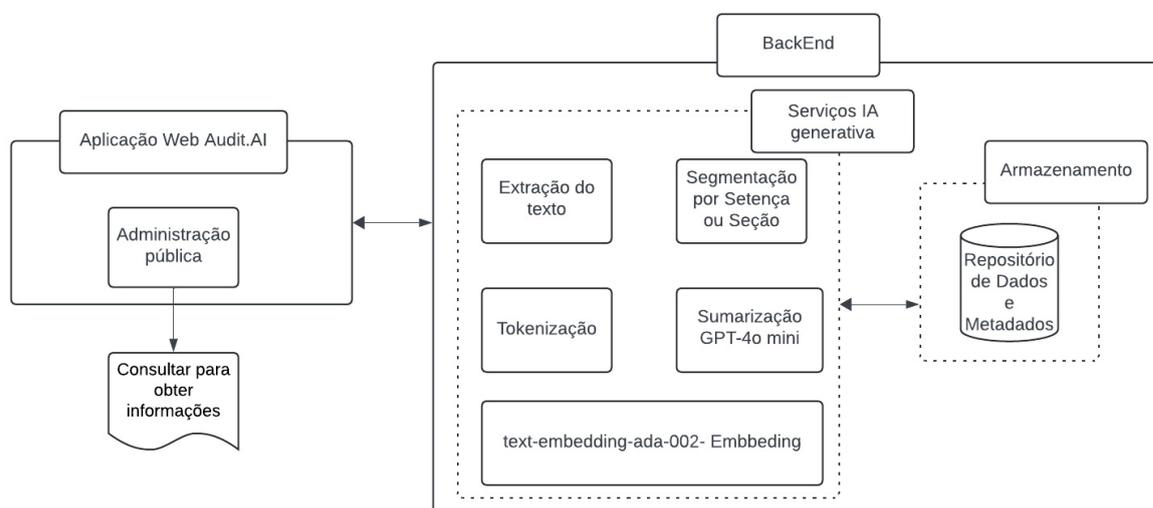
O modelo text-embedding-ada-002 da OpenAI foi utilizado para a geração dos *embeddings*, destacando-se por sua capacidade de produzir representações vetoriais de alta qualidade que capturam nuances semânticas complexas. É importante

destacar que o modelo também oferece eficiência em termos de custo e desempenho, apresentando razoável viabilidade financeira frente a outras opções técnicas disponíveis.

3.3. Plataforma Audit.AI

A Figura 2 apresenta uma representação geral dos componentes da plataforma Audit.AI. A aplicação é composta por uma aplicação web, segmentada em duas partes principais: (1) *Frontend* que representa as interfaces da aplicação; e (2) *Backend* que encapsula as principais funcionalidades da plataforma, com especial atenção aos módulos destinados a extração de texto, tokenização e sumarização de informações. Além disso, a aplicação conta com um banco de dados para armazenamento dos dados estruturados e seus metadados, assim como para armazenamento dos dados brutos coletadas de fontes externas.

Figura 2 – Componentes da plataforma



Fonte: Elaborado pelos autores (2025)

3.3.1. Seleção de Modelo de GML

É importante destacar que o escopo do trabalho incluiu uma busca intensiva por plataformas de GML, considerando as necessidades específicas do projeto e as características dos dados coletados. A pesquisa focou na identificação de soluções que pudessem ser utilizadas como serviço, levando em conta questões de viabilidade técnica e financeira. Essa abordagem permitiu avaliar diferentes opções disponíveis no mercado, garantindo que a escolha do modelo fosse alinhada com os objetivos do Audit.AI e as limitações orçamentárias do projeto.

O modelo GPT-4 Mini foi escolhido para o Audit.AI em detrimento a outras opções técnicas devido às suas capacidades avançadas de compreensão e geração de texto, que auxiliam tarefas complexas de PLN. O modelo GPT-4 apresenta melhorias significativas em relação ao seu antecessor, incluindo uma maior capacidade de lidar com contextos mais longos e uma compreensão mais profunda de nuances linguísticas, o que é crucial para a análise de PADs. Além disso, o modelo GPT-4 mini foi selecionado para o Audit.AI também devido ao seu menor custo no processamento de texto em comparação com versões anteriores, como o 3.5-Turbo,

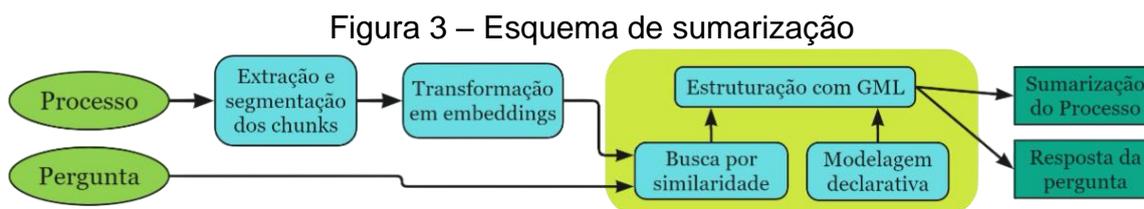
por exemplo. Essa característica torna o GPT-4 Mini uma opção economicamente viável, permitindo que o sistema realize análises de grandes volumes de dados de forma mais acessível, sem comprometer a qualidade das respostas geradas.

A arquitetura de software do Audit.AI foi concebida de maneira a permitir a substituição do modelo de linguagem utilizado de forma flexível e eficiente. Essa abordagem modular garante que, caso surjam novas versões de modelos de linguagem ou alternativas mais avançadas no futuro, a integração dessas tecnologias possa ser realizada com facilidade, sem a necessidade de reestruturar todo o sistema. Além disso, a arquitetura foi projetada para facilitar a utilização de diferentes modelos de linguagem, de forma a potencialmente explorar os pontos fortes de cada modelo. Esta abordagem permite a utilização de múltiplos modelos de linguagem, oferecendo uma abordagem versátil e adaptável às diversas necessidades do sistema. Essa flexibilidade possibilita que modelos que se destacam na sumarização de informações sejam empregados em tarefas específicas, enquanto outros, mais adequados para a extração de dados, possam ser utilizados em diferentes contextos.

3.3.2. Sumarização dos Dados

Entre as principais funcionalidades do Audit.AI, destaca-se a capacidade de sumarização de informações dos PADs. Essa funcionalidade permite que o sistema analise grandes volumes de dados textuais e extraia os pontos mais relevantes de forma concisa e clara, facilitando a compreensão e a tomada de decisões por parte dos gestores. Sob a perspectiva de aprendizado de máquina, sumarização é uma técnica de PLN que visa condensar um texto extenso em uma versão mais curta, mantendo as informações essenciais e a coerência do conteúdo original. Essa prática pode ser dividida em duas abordagens principais: a sumarização extrativa, que seleciona e combina frases significativas do texto original para formar um resumo, e a sumarização abstrativa, que gera um novo texto que sintetiza as ideias principais, podendo incluir reescritas e reformulações. A sumarização é amplamente utilizada em diversas áreas, e sua eficácia depende da identificação de entidades nomeadas, análise de partes do discurso e compreensão semântica, que juntas contribuem para a produção de resumos que não apenas informam, mas também preservam a intenção comunicativa do texto original.

A Figura 3 apresenta um esquema do processo de análise textual adotado no Audit.AI. O fluxo envolve a extração e segmentação de trechos de um processo, seguidos pela conversão desses trechos em representações vetoriais. Em seguida, as informações passam por um processo de estruturação com GML, que inclui busca por similaridade e modelagem declarativa. Como resultado, o sistema gera tanto a sumarização do processo quanto respostas específicas para perguntas formuladas pelo usuário.



Fonte: Elaborado pelos autores (2025)

4. ANÁLISE PROSPECTIVA DOS RESULTADOS ESPERADOS

Atualmente, o sistema Audit.AI já possui Certificado de Registro de Programa de Computador, registrado sob o processo nº BR512025000158-4, opera como um Produto Mínimo Viável - *Minimum Viable Product* (MVP) e encontra-se em fase de refinamento.

A Figura 4 apresenta uma das telas da interface do sistema Audit.AI, onde são exibidos os detalhes do processo analisado, incluindo informações como número do processo, ano da infração, recomendação de penalidade e conclusão. Além disso, a tela permite ao usuário realizar consultas e visualizar o histórico de perguntas e respostas geradas pelo sistema.

Figura 4 – Interface do sistema Audit.AI para consulta de processos

The screenshot displays the 'Detalhes dos processos' (Process Details) page in the Audit.AI system. The interface includes a sidebar with navigation options like 'Lista de Processos', 'Carregar Processo', 'Manual', 'Histórico', 'Suporte', and 'Sobre'. The main content area shows the following details for a process:

- Nup:** 23208.001039/2017-70
- ID do processo:** 3cea5d0e-6419-41a4-8f90-65ad329c77ca
- Ano da Infração:** 2019
- Região:** Amapá
- Quantidade de servidores penalizados:** 5
- Recomendação de penalidade:** Arquivamento do processo.
- Conclusão:** A comissão concluiu que não foram encontradas evidências suficientes para comprovar as irregularidades denunciadas.
- Descrição:** Supostas irregularidades na fiscalização do contrato 12/2013.

Below the details is a 'Histórico de perguntas' (History of questions) section with a table:

Id	Perguntas	Respostas
1	Qual o número único do processo (nup)?	23208.00103
2	Quais os fatos utilizados para embasar as acusações aos servidores indiciados?	As acusações

The interface also features a 'Consultar' button and a footer with the text 'Todos os direitos reservados'.

Fonte: Audit.AI - Plataforma Anticorrupção baseada em Inteligência Artificial, 2025.

Simultaneamente, estão sendo realizados os trâmites burocráticos necessários para iniciar a aplicação prática. A análise prospectiva apresentada neste capítulo delinea os resultados esperados com base no planejamento e nos objetivos traçados, antecipando os impactos e benefícios que o Audit.AI pode proporcionar às IFES e à gestão pública em geral.

4.1. Aplicação Prática Simulada

Embora a aplicação prática do sistema em cenários reais ainda não tenha sido realizada, os trâmites burocráticos para obtenção de autorização institucional estão em andamento. A previsão é de que o Audit.AI seja aplicado em setores voltados para a gestão da integridade e auditoria, explorando suas capacidades preventivas e de detecção.

Na gestão da integridade, o sistema busca prever processos ou áreas de trabalho com risco potencial de ocorrência de ilícitos relacionados à corrupção,

permitindo ações preventivas mais eficazes. Já na auditoria, o Audit.AI será utilizado para nortear decisões relativas à realização de auditorias especiais, ajudando a priorizar casos com maior probabilidade de irregularidades, otimizando assim o uso de recursos e esforços investigativos.

Além disso, os testes em ambientes simulados serão essenciais para:

- **Avaliar o desempenho do sistema:** Medindo sua precisão, escalabilidade e adaptabilidade em diferentes contextos institucionais.
- **Identificar ajustes necessários:** Corrigindo limitações e promovendo melhorias antes da implementação em larga escala.
- **Validar os modelos de inteligência artificial:** Garantindo a confiabilidade e a consistência dos resultados gerados, alinhados aos objetivos institucionais de gestão de riscos e auditoria.

4.2. Potencial de Expansão para Outras Instituições

Além das IFES, o Audit.AI possui potencial de aplicação em outros órgãos da administração pública, especialmente em áreas que demandam alta eficiência na análise de processos administrativos e na identificação de padrões. A perspectiva de expansão inclui:

- a) Adaptação a diferentes contextos: Personalizando os modelos para atender às especificidades de outras instituições públicas.
- b) Fortalecimento das linhas de defesa: Funcionando como norteador para gestores, auditores e áreas de *compliance*, ampliando a capacidade de detecção de irregularidades.

5. DISCUSSÃO

Esta seção apresenta os desafios identificados na coleta e análise dos dados, destacando as dificuldades enfrentadas na obtenção de PADs e na padronização do dataset. Além disso, discute as limitações para a implementação do Audit.AI, como o acesso às informações e os custos operacionais, ressaltando a necessidade de planejamento estratégico para garantir sua viabilidade e expansão.

5.1. Aspectos gerais construção do dataset

A investigação realizada, com base nas características previamente definidas, revelou a existência de 228 sanções expulsivas relacionadas à corrupção, distribuídas entre 76 instituições de ensino. A análise dos dados coletados revelou uma disparidade significativa entre as informações fornecidas pelas Instituições Federais de Ensino Superior (IFES) e os registros disponíveis no Painel de Correição da Controladoria-Geral da União (CGU). Esse desalinhamento foi evidenciado pelas respostas das instituições, muitas das quais justificaram a não disponibilização das informações sob a alegação de que não haviam ocorrido PADs dessa natureza em suas estruturas. Essa situação é preocupante, pois sugere potenciais lacunas na gestão e registro dessas informações.

Além disso, foram identificadas inconsistências relevantes durante a coleta, que dificultaram a padronização do *dataset*. Os problemas mais recorrentes incluíram:

- a) Variedade de formatos de arquivos: A heterogeneidade dos documentos recebidos, com diferentes padrões de digitalização e organização.
- b) Qualidade das digitalizações: Muitos documentos apresentavam baixa qualidade, dificultando a leitura e análise.
- c) Incompletude dos processos: Alguns arquivos estavam fragmentados ou careciam de informações essenciais, prejudicando a análise detalhada.

A análise evidencia desafios estruturais e administrativos na disponibilização de processos disciplinares pelas IFES. A variação na taxa de resposta entre as regiões sugere desigualdades institucionais, enquanto os atrasos e dificuldades operacionais refletem a ausência de padronização e recursos tecnológicos adequados. A implementação de ferramentas automatizadas, a padronização de procedimentos e o fortalecimento da capacidade administrativa são medidas essenciais para aprimorar a transparência e a eficiência na gestão de informações disciplinares.

1

5.2. Desafios e Limitações da Implementação

A implementação de um sistema de análise automatizada de PADs apresenta desafios relacionados ao acesso às informações e aos custos operacionais. Uma das principais dificuldades identificadas é a dificuldade de acesso às informações necessárias para a atualização do banco de dados. Atualmente, a obtenção de PADs depende de solicitações manuais realizadas às IFES, um processo burocrático, demorado e suscetível a atrasos e negativas de envio. Essa limitação compromete a abrangência dos dados e a manutenção de uma base de informações sempre atualizada. Uma possível solução seria a integração direta com sistemas da Controladoria-Geral de União (CGU), permitindo a disponibilização automática dos PADs para análise, eliminando a dependência de solicitações individuais. No entanto, essa integração dependeria de autorizações institucionais para assegurar a conformidade com normas de transparência e proteção de dados.

Outro fator a ser considerado é o custo associado ao uso do ChatGPT, ferramenta central para o processamento automatizado de grandes volumes de dados. Embora o modelo ofereça valores acessíveis para análise de dados em escala, é importante reconhecer que, ainda que relativamente pequeno, há um custo envolvido no uso contínuo da tecnologia. Esse fator deve ser levado em consideração no planejamento financeiro, especialmente em projetos de longo prazo que exijam uma aplicação intensiva do sistema.

Esses desafios destacam a necessidade de planejamento estratégico, considerando não apenas os avanços tecnológicos, mas também os aspectos administrativos e financeiros envolvidos. Com o devido alinhamento institucional e investimento em soluções adequadas, o sistema tem grande potencial para ser mantido e expandido, tornando-se uma ferramenta eficaz para a análise e prevenção de irregularidades em diferentes órgãos públicos.

6. CONSIDERAÇÕES FINAIS

O desenvolvimento do sistema Audit.AI reafirma a importância da aplicação de tecnologias emergentes como ferramenta para fortalecer a governança pública e aprimorar os mecanismos institucionais de controle. O sistema foi projetado para apoiar a análise de PADs, oferecendo uma abordagem inovadora para a identificação

de padrões relacionados à corrupção nas IFES e, potencialmente, em outros órgãos públicos, utilizando técnicas de PLN e similaridade semântica.

Durante a execução do projeto, dificuldades institucionais por parte das IFES comprometeram a celeridade na obtenção das informações necessárias. Essas limitações ressaltam a necessidade de políticas institucionais mais robustas e de autorizações específicas para o compartilhamento seguro de dados entre os órgãos públicos envolvidos.

Apesar desses desafios, a pesquisa confirmou que o Audit.AI tem potencial para promover uma análise automatizada eficiente e escalável. O sistema demonstrou capacidade de processar grandes volumes de dados com custo relativamente acessível, considerando sua complexidade técnica e os benefícios esperados. Sua implementação prática dependerá de autorizações institucionais e de investimentos contínuos para garantir sua manutenção e expansão para outros órgãos públicos além das IFES.

Além disso, a aplicação de um sistema como o Audit.AI contribui diretamente para o fortalecimento da governança, ao possibilitar análises orientadas por dados e promover decisões mais transparentes, responsáveis e eficazes. O uso de inteligência artificial para análise de PADs aprimora os mecanismos de controle interno, reduzindo riscos operacionais e institucionais.

Conclui-se que o Audit.AI oferece uma abordagem inovadora para modernizar a gestão pública, promovendo transparência, eficiência e responsabilidade administrativa. Com uma implementação adequada e respaldada por políticas claras, espera-se que a tecnologia desenvolvida contribua para o aprimoramento dos mecanismos de controle e para a prevenção e detecção de práticas ilícitas na administração pública.

7. REFERÊNCIAS

BRASIL. **Plano de Integridade e Combate à Corrupção 2025-2027**. Controladoria-Geral da União, Brasília, 2024.

BRASIL. **Decreto nº 8.539, de 8 de outubro de 2015**. Dispõe sobre o uso do meio eletrônico para a realização do processo administrativo no âmbito dos órgãos e das entidades da administração pública federal direta, autárquica e fundacional. Casa Civil. Disponível em: https://www.planalto.gov.br/ccivil_03/ato2015-2018/2015/decreto/d8539.htm. Acesso em: 25 jan. 2025.

BRASIL. **Decreto nº 9.203, de 22 de novembro de 2017**. Dispõe sobre a política de governança da administração pública federal direta, autárquica e fundacional. Casa Civil. Disponível em: http://www.planalto.gov.br/ccivil_03/ato2015-2018/2017/decreto/d9203.htm. Acesso em: 30 jan. 2025.

BRASIL. **Lei nº 12.527, de 18 de novembro de 2011**. Dispõe sobre o acesso a informações públicas e dá outras providências. Casa Civil. Disponível em: https://www.planalto.gov.br/ccivil_03/ato2011-2014/2011/lei/l12527.htm. Acesso em: 22 jan. 2025.

Controladoria-Geral da União (CGU). **Painel de Correções**. Disponível em: <https://centralpaineis.cgu.gov.br/visualizar/corregedorias>. Acesso em: 5 fev. 2025.

Gilson, D. H. M. I.; Bramili, G. A. **Inteligência artificial no combate à fraude e corrupção: a experiência da Controladoria Geral do Município do Rio de Janeiro**. Revista da CGU, [online], 2023. Disponível em: https://revista.cgu.gov.br/Revista_da_CGU/article/view/531. Acesso em: 28 jan. 2025.

Macedo, S. V.; Valadares, J. L. **A produção acadêmica brasileira sobre corrupção: uma revisão sistemática**. REAd. Revista Eletrônica de Administração, [online], v. 27, n. 2, p. 400-429, 2021. DOI: <https://doi.org/10.1590/1413-2311.319.101302>. Acesso em: 3 fev. 2025.

Oliveira, J. C.; Cruz, M. V. G. **Corrupção na educação superior pública? Estudo das IFES de Minas Gerais**. Revista de Administração Pública, [online], v. 55, n. 6, p. 1295-1311, 2021. DOI: <https://doi.org/10.1590/0034-761220210100>. Acesso em: 20 jan. 2025.

Organização das Nações Unidas (ONU). **Transformando nosso mundo: a Agenda 2030 para o Desenvolvimento Sustentável**. 2015. Disponível em: <https://sdgs.un.org/2030agenda>. Acesso em: 10 fev. 2025.

Rodrigues, D. S.; Faroni, W.; Santos, N. A.; Ferreira, M. A. M.; Diniz, J. A. **Corrupção e má gestão nos gastos com educação: fatores socioeconômicos e políticos**. Revista de Administração Pública, [online], v. 54, n. 2, p. 301-320, 2020. DOI: <http://dx.doi.org/10.1590/0034-761220190103>. Acesso em: 27 jan. 2025.

Silva, A. L.; Sampaio, V. G. R. C. A.; Lima, A. M. A.; Cabral, G. G.; Valença, G. **Ferramenta para auxílio à auditoria de editais municipais para compra de medicamentos**. Anais do XX Simpósio Brasileiro de Sistemas de Informação (SBSI) [online], 2024. Disponível em: https://sol.sbc.org.br/index.php/sbsi_estendido/article/view/28623/28427. Acesso em: 1 fev. 2025.

Silva, E. C.; Medeiros, I. P.; Menezes, M. V.; Kamikawachi, D. S. L. **Segmentation and summarization for extracting information about information technology equipment from government procurement notice**. In: Symposium on Knowledge Discovery, Mining and Learning (KDMILE). Sociedade Brasileira de Computação, 2024.

Silva, M.; Santos, E.; Alves, K.; Silva, H.; Pedrosa, F.; Valença, G.; Brito, K. **Using generative AI for simplifying official documents in the public accounts domain**. Anais do Workshop de Computação Aplicada em Governo Eletrônico (WCGE) [online], 2024. Disponível em: <https://sol.sbc.org.br/index.php/wcge/article/view/29543>. Acesso em: 8 fev. 2025.

Topchii, V.; Zadereiko, S.; Didkivska, G.; Bodunova, O.; Shevchenko, D. **International anti-corruption standards**. Baltic Journal of Economic Studies, v. 7, n. 5, p. 277–286, 2021. DOI: <https://doi.org/10.30525/2256-0742>.

Transparência Internacional. **Índice de Percepção da Corrupção 2023**. Berlim: Transparência Internacional, 2024. Disponível em: <https://comunidade.transparenciainternacional.org.br/ipc-indice-de-percepcao-da-corrupcao-2023>. Acesso em: 29 jan. 2025.

Transparência Internacional. **Índice de Percepção da Corrupção 2024**. Berlim: Transparência Internacional, 2025a. Disponível em: <https://comunidade.transparenciainternacional.org.br/indice-de-percepcao-da-corrupcao-2024>. Acesso em: 12 fev. 2025.

Transparência Internacional. **Página oficial**. 2025b. Disponível em: <https://transparenciainternacional.org.br/ipc/2024>. Acesso em: 13 fev. 2025.

Tribunal de Contas da União (TCU). **Competências do TCU**. Disponível em: <https://portal.tcu.gov.br/institucional/conheca-o-tcu/competencias/>. Acesso em: 4 fev. 2025.

UNODC - *United Nations Office on Drugs and Crime*. **UNODC and Corruption**. 2024. Disponível em: <https://www.unodc.org/lpo-brazil/pt/corrupcao/index.html>. Acesso em: 6 fev. 2025.