



## PAES, UMA CLASSIFICADOR DE EDITAIS LICITATÓRIOS SEGUNDO CRITÉRIOS DE SUSTENTABILIDADE

Felipe dos Santos Silva<sup>1</sup>

Juliano de Araújo Mota Mesquita<sup>2</sup>

Bruno César Barreto de Figueirêdo<sup>3</sup>

**Resumo:** O presente trabalho tem por objetivo propor um analisador e classificador de editais de compras públicas sustentáveis, denominado PAES (Proposta de Analisador de Editais Sustentáveis). A sua performance será comparada com o desempenho do modelo de linguagem Gemini em diferentes configurações. Como estudo de caso foram utilizados vinte cinco editais de aquisição de materiais elétricos, área de elevada relevância para a sustentabilidade devido ao seu impacto no consumo energético e na geração de resíduos. A metodologia de avaliação consistiu na definição de um Padrão Ouro, contra o qual foram avaliadas as notas geradas tanto pelo PAES quanto pelo Gemini. O PAES atribui notas aos editais com base em dez critérios de sustentabilidade ponderados por pesos previamente definidos (eficiência energética, normas técnicas, certificações, resíduos, responsabilidade social, entre outros). Para mensuração da performance, utilizaram-se métricas contínuas de erro (MAE e RMSE) e métricas de classificação (Precisão, Recall, Acurácia e F1-score). Os resultados indicaram que o PAES apresentou desempenho consistentemente superior, com valores constantes de MAE (0,125) e RMSE (0,156), independentemente do número de editais analisados. Em contraste, o Gemini demonstrou degradação de desempenho à medida que a quantidade de documentos aumentou, além de maior variabilidade entre cenários. Nas métricas de classificação, o PAES alcançou F1-score de 0,8462, superando significativamente as variantes do Gemini, que oscilaram entre 0 e 0,5263. Conclui-se que o PAES oferece maior robustez, previsibilidade e auditabilidade, características fundamentais para a aplicação de inteligência artificial em auditoria e controle de contratações públicas.

**Palavras-chave:** sustentabilidade; análise de editais; inteligência artificial; classificação automática; auditoria pública.

**Abstract:** This paper aims to propose an analyzer and classifier for sustainable public procurement tenders, called PAES (Proposal for a Sustainable Tender Analyzer). Its performance will be compared with the performance of the Gemini language model in different configurations. As a case study, twenty-five tenders for the acquisition of electrical materials were used, an area of high relevance for sustainability due to its impact on energy consumption and waste generation. The evaluation methodology consisted of defining a Gold Standard, against which the

---

<sup>1</sup> Discente da Universidade Estadual de Roraima, [felipe.silva@alunos.uerr.edu.br](mailto:felipe.silva@alunos.uerr.edu.br);

<sup>2</sup> Discente da Universidade Estadual de Roraima, [juliano.mesquita@alunos.uerr.edu.br](mailto:juliano.mesquita@alunos.uerr.edu.br);

<sup>3</sup> Docente da Universidade Estadual de Roraima, Servidor Público do Tribunal de Contas do Estado de Roraima, [brunocesar@tcerr.tc.br](mailto:brunocesar@tcerr.tc.br)





scores generated by both PAES and Gemini were evaluated. PAES assigns scores to tenders based on ten sustainability criteria weighted by predefined weights (energy efficiency, technical standards, certifications, waste, social responsibility, among others). To measure performance, continuous error metrics (MAE and RMSE) and classification metrics (Precision, Recall, Accuracy, and F1-score) were used. The results indicated that PAES consistently outperformed the system, with constant MAE (0.125) and RMSE (0.156) values, regardless of the number of tenders analyzed. In contrast, Gemini demonstrated performance degradation as the number of documents increased, as well as greater variability between scenarios. In the classification metrics, PAES achieved an F1-score of 0.8462, significantly surpassing the Gemini variants, which ranged from 0 to 0.5263. It is concluded that PAES offers greater robustness, predictability, and auditability, fundamental characteristics for the application of artificial intelligence in auditing and controlling public procurement.

**Keywords:** sustainability; bidding process analysis; artificial intelligence; automatic classification; public audit.





## 1. Introdução

A crescente preocupação com a sustentabilidade nas contratações públicas reflete a consolidação de um paradigma em que critérios ambientais e sociais passam a ser incorporados de forma sistemática nos processos de compras governamentais. Esse movimento não é apenas tendência internacional, mas obrigação normativa no Brasil, a partir de marcos como a **Política Nacional de Resíduos Sólidos (Lei nº 12.305/2010)** e a **Lei nº 14.133/2021**, que estabelece em seu art. 11 o princípio do **desenvolvimento nacional sustentável** como norteador das licitações e contratos administrativos.

No campo específico de materiais elétricos, o atendimento a requisitos como **eficiência energética, normas técnicas de segurança, rotulagem de desempenho e certificações ambientais** é particularmente relevante. O setor elétrico está diretamente ligado ao consumo energético e, portanto, às emissões de gases de efeito estufa, tema central em compromissos multilaterais como o **Acordo de Paris (2015)** e os **Objetivos de Desenvolvimento Sustentável da Agenda 2030 da ONU**, notadamente os ODS 7 (Energia limpa e acessível), ODS 12 (Consumo e produção responsáveis) e ODS 13 (Ação contra a mudança global do clima) (ONU, 2015).

Nesse contexto, a adoção de **ferramentas automatizadas** para análise e classificação de editais de compra pode apoiar os órgãos de controle na verificação de conformidade e na indução de boas práticas. Modelos de linguagem como o **Gemini** oferecem flexibilidade na interpretação textual, mas apresentam limitações em consistência e estabilidade quando submetidos a grandes volumes de documentos. Por outro lado, soluções estruturadas como o **PAES (Plataforma de Avaliação de Editais Sustentáveis)**, ao se basear em critérios objetivos e pesos predefinidos, tendem a oferecer maior previsibilidade, auditabilidade e robustez metodológica, qualidades essenciais no âmbito da Administração Pública (Carvalho et al., 2020; Pacheco & Sampaio, 2021).

Assim, o presente trabalho busca comparar o desempenho do PAES com o Gemini em diferentes cenários de análise de editais de materiais elétricos, avaliando métricas de erro (MAE e RMSE), bem como indicadores clássicos de classificação (precisão, recall, acurácia e F1-score). A partir dessa comparação, pretende-se demonstrar a relevância de sistemas parametrizados como o PAES para garantir maior alinhamento





às normas e políticas públicas de sustentabilidade, oferecendo subsídios técnicos à gestão e ao controle externo.

Este artigo está estruturado em cinco seções, além desta introdução. A segunda seção apresenta o **referencial teórico**, destacando a literatura sobre contratações públicas sustentáveis e a aplicação de técnicas de inteligência artificial em auditoria e análise documental. Na terceira seção, descreve-se a **metodologia** adotada, explicitando os critérios utilizados, seus pesos relativos e as métricas estatísticas e de classificação aplicadas. A quarta seção reúne os **resultados e discussões**, com destaque para a comparação entre o PAES e as variantes do Gemini, apoiada em gráficos e tabelas de desempenho. Por fim, a quinta seção traz as **considerações finais**, discutindo implicações práticas, limitações do estudo e perspectivas de pesquisa futura.

## 2. Referencial Teórico

### 2.1. Compras públicas sustentáveis e o papel normativo

A incorporação de critérios de sustentabilidade nas contratações públicas deixou de ser mera “boa prática” e passou a constituir **diretriz legal e política pública obrigatória**. No Brasil, esse movimento é ancorado por marcos como a **Política Nacional de Resíduos Sólidos (PNRS)**, que institui responsabilidade compartilhada e metas de redução, reutilização e reciclagem (Lei nº 12.305/2010), e pela **Lei nº 14.133/2021**, que consagra o **desenvolvimento nacional sustentável** como princípio das licitações (art. 11). No plano infralegal, o **Decreto nº 7.746/2012** reforça a obrigatoriedade de inserir critérios de sustentabilidade nas compras públicas federais.

Estudos como os de **Carvalho, Oliveira e Santos (2020)** e **Pacheco e Sampaio (2021)** mostram que a adoção de compras públicas sustentáveis funciona como instrumento de inovação regulatória, ampliando a eficiência e a transparência. No âmbito internacional, pesquisas de **Thai (2001)** e **Brammer & Walker (2011)** destacam que a função estratégica da contratação pública não se restringe ao preço, mas deve incorporar dimensões sociais e ambientais.

No plano global, a **Agenda 2030** e os **Objetivos de Desenvolvimento Sustentável (ODS)** — especialmente ODS 7, 12 e 13 —, assim como o **Acordo de Paris**





(2015), reforçam a centralidade do consumo e da produção responsáveis como eixos estruturantes das políticas públicas (ONU, 2015).

## 2.2.IA generativa e análise documental em governo

A evolução dos Modelos de Linguagem de Grande Porte (LLMs) ampliou a capacidade de processamento automático de documentos, permitindo localizar cláusulas específicas em editais, verificar a presença de critérios normativos e sintetizar indicadores de conformidade. No entanto, como apontam Dwivedi et al. (2023) e Sun & Medaglia (2019), o uso direto de LLMs em governo ainda enfrenta barreiras como instabilidade, vieses e falta de explicabilidade.

Por isso, a literatura recente tem defendido o uso de arquiteturas híbridas, em que a LLM desempenha funções de extração, mas a decisão final é governada por critérios normativos auditáveis (Wirtz et al., 2019; Gil-Garcia, Dawes & Pardo, 2018). O PAES insere-se nessa lógica, funcionando como um modelo que combina a flexibilidade das LLMs com a previsibilidade de pesos e critérios objetivos. Por isso, a literatura recente e as boas práticas em governo têm recomendado **pipelines estruturados** que combinem LLMs com **regras de validação, esquemas de dados (JSON)**, verificações semânticas por páginas e **agregação ponderada** dos achados. Em outras palavras, **a LLM desempenha o papel de extrator/rastreador**, enquanto a decisão normativa é ancorada em **critérios explícitos e pesos transparentes**. É exatamente essa a filosofia do **PAES**: um arranjo “LLM-inside” com **camadas de controle**, reduzindo a variabilidade e oferecendo **auditabilidade ex-post** (possibilidade de reexecutar, rastrear e explicar cada decisão).

## 2.3. Critérios, pesos e materialidade regulatória

A literatura nacional já discute a importância de alinhar critérios de sustentabilidade a pesos que reflitam a sua materialidade regulatória. Estudos como o de Santos et al. (2019) analisam a aplicação de critérios de sustentabilidade em licitações municipais, enquanto Souza e Gomes (2021) destacam que pesos diferenciados podem reforçar a eficácia da política pública.





No caso de materiais elétricos, a ênfase em eficiência energética (0,20) é coerente com metas de descarbonização (Lei nº 14.133/2021; ODS 7). Normas técnicas (0,15) e ensaios/verificações (0,10) estão associados à segurança e conformidade regulatória (Lei nº 9.933/1999). Rotulagem/selos (0,10) e certificações ambientais (0,08) funcionam como evidências auditáveis (ISO 14001), em linha com o Decreto nº 7.746/2012. Critérios relacionados a resíduos (logística reversa, recicláveis e redução) derivam da PNRs, enquanto origem sustentável e responsabilidade social se ligam ao princípio constitucional da função social da contratação pública (CF/88, art. 170, III).

#### **2.4. Explicabilidade, accountability e risco regulatório**

No contexto das contratações públicas, não basta que um sistema automatizado acerte em média: é preciso que as decisões sejam explicáveis e auditáveis. O conceito de IA Explicável (XAI) (Doshi-Velez & Kim, 2017; Arrieta et al., 2020) é essencial para assegurar transparência em setores sensíveis. Em auditoria, isso significa permitir que gestores e órgãos de controle compreendam e reproduzam os passos que levaram a uma decisão.

A literatura de governança algorítmica (Yeung, 2018; Zuboff, 2019) também reforça que sistemas aplicados a decisões públicas devem oferecer previsibilidade, refutabilidade e accountability, mitigando riscos de arbitrariedade. O PAES, ao estruturar saídas intermediárias (JSON) e notas ponderadas, materializa esse tipo de governança.

#### **2.5. Métricas de avaliação e validade de uso**

As métricas utilizadas neste estudo (MAE, RMSE, Precisão, Recall, Acurácia e F1-score) têm suporte na literatura de avaliação de sistemas de informação. Pereira et al. (2022) mostram como indicadores objetivos podem revelar padrões ocultos em licitações fraudulentas, reforçando a importância de métricas robustas em análises públicas. Em IA, medidas contínuas e classificatórias são recomendadas para capturar tanto a proximidade numérica quanto os padrões de acerto/erro em classificação (Sokol & Flach, 2020).

No presente trabalho, o PAES mostrou-se mais estável, mantendo valores constantes de erro e maior F1, corroborando estudos que apontam para a superioridade de modelos explicáveis em cenários de auditoria digital (Dwivedi et al., 2023).





### 3. Metodologia

A avaliação comparativa entre o PAES e o Gemini foi conduzida em três etapas principais: (i) definição da base de referência (Padrão Ouro), (ii) cálculo das métricas de erro absoluto e quadrático médio, e (iii) aplicação das métricas clássicas de classificação binária.

**Padrão Ouro:** Inicialmente, foram selecionados aleatoriamente 25 editais de licitações do Ocorridas em 2025 no Estado de Roraima, cujo objeto era a aquisição de material elétrico.

Como métrica, adotou-se dez critérios de sustentabilidade, com pesos calibrados para refletir sua materialidade técnica e regulatória no contexto de materiais elétricos: **eficiência energética (0,20); normas técnicas (0,15); rotulagem/selos (0,10); verificação/ensaio (0,10); embalagem e logística reversa (0,10); materiais recicláveis ou renováveis (0,10); certificações ambientais (0,08); origem sustentável dos insumos (0,07); política de redução de resíduos (0,05); e responsabilidade social (0,05)**. A ênfase em **eficiência energética (0,20)** alinha-se ao objetivo de redução do consumo e das emissões ao longo do ciclo de vida dos equipamentos, em consonância com o princípio do **desenvolvimento nacional sustentável** previsto na **Lei nº 14.133/2021** e com os programas oficiais de rotulagem e desempenho energético (**PBE/INMETRO** e **Selo Procel**), amplamente utilizados como referência técnica em contratações.

A exigência de conformidade a **normas técnicas (0,15)** e de **verificação/ensaio (0,10)** decorre do dever de garantir segurança, qualidade e interoperabilidade, observando **ABNT NBRs** e **Regulamentos de Avaliação da Conformidade do INMETRO** (Lei nº 9.933/1999), o que reduz assimetria de informação e riscos de “greenwashing”.

Os critérios de **rotulagem/selos (0,10)** e de **certificações ambientais (0,08)** (p.ex., ISO 14001) funcionam como evidências auditáveis de desempenho ambiental e de gestão, estimuladas por **Decreto nº 7.746/2012** (critérios de sustentabilidade nas aquisições públicas federais) e pelas diretrizes de compras públicas sustentáveis. No eixo de resíduos, **embalagem e logística reversa (0,10)**, **materiais recicláveis/renováveis (0,10)** e **política de redução de resíduos (0,05)** estão diretamente amparados pela **Política Nacional de Resíduos Sólidos — PNRS (Lei nº 12.305/2010 e regulamentos)**,





que instituem responsabilidade compartilhada, metas de redução e retorno de pós-consumo.

A **origem sustentável dos insumos (0,07)** reforça a rastreabilidade e a conformidade socioambiental na cadeia de suprimentos, aderindo às diretrizes de sustentabilidade da **Lei nº 14.133/2021** e aos instrumentos federais de compras sustentáveis (p. ex., guias e catálogos de referência). Por fim, **responsabilidade social (0,05)** materializa a dimensão social da contratação pública, coerente com a finalidade constitucional de promoção do desenvolvimento e com o vetor de sustentabilidade explicitado na nova lei de licitações. Em conjunto, tais critérios — e seus respectivos pesos — traduzem, de forma operacional e auditável, os comandos normativos que orientam a Administração a **internalizar requisitos de desempenho ambiental e social** nas aquisições, garantindo comparabilidade objetiva entre propostas e rastreabilidade de conformidade em auditorias. A aplicação desse critério aos Editais constitui o baseline de referência, em relação ao qual todas as predições foram comparadas.

**Cálculo das métricas de erro:** Para medir a proximidade entre as notas preditas e o Padrão Ouro, foram adotadas duas métricas clássicas: o Erro Absoluto Médio (MAE) e a Raiz do Erro Quadrático Médio (RMSE), conforme Equações (1) e (2).

$$(1) MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|$$

$$(2) RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}$$

Onde  $\hat{y}_i$  representa a nota predita pelo modelo,  $y_i$  a nota de referência do Padrão Ouro, e  $n$  o número total de editais avaliados.

Para aprofundar a análise, as notas contínuas foram convertidas em classificações binárias por meio de um limiar fixado em **0,5** (Equação 3):

$$(3) \hat{y}_i^{bin} = \begin{cases} 1, & \hat{y}_i \geq 0.5 \\ 0, & \hat{y}_i < 0.5 \end{cases}$$

Com base nessa binarização, foram construídas as matrizes de confusão, considerando Verdadeiros Positivos (TP), Verdadeiros Negativos (TN), Falsos Positivos





(FP) e Falsos Negativos (FN). A partir delas, foram derivadas as métricas clássicas (Equação 4):

$$(4) \text{ Precision} = \frac{TP}{TP+FP}, \text{ Recall} = \frac{TP}{TP+FN}, \text{ Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} \text{ e } F1 = 2 \cdot \frac{\text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}}$$

Essas métricas permitem identificar não apenas a proximidade numérica entre as notas (Equações 1 e 2), mas também o equilíbrio entre **detectar editais sustentáveis (Recall, Equação 4)** e **evitar falsos alarmes (Precisão, Equação 4)**.

Além do limiar fixo de 0,5 (Equação 3), foram realizados testes de sensibilidade variando o *threshold* ( $t \in \{0.3, 0.4, 0.5, 0.6, 0.7\}$ ), com o objetivo de observar a estabilidade dos modelos frente a diferentes critérios de decisão.

**Fluxo de Funcionamento do PAES:** O fluxo de funcionamento do PAES (Imagem 1) compreende cinco etapas sequenciais. Inicialmente, ocorre o upload dos editais em formato PDF, os quais são armazenados em servidor para posterior processamento. Em seguida, na fase de análise da IA, cada documento é submetido a um prompt estruturado que orienta a LLM na extração das informações relevantes; como saída, a LLM retorna um arquivo em JSON, no qual as evidências identificadas são organizadas de forma padronizada. Na etapa de validação, o sistema verifica a consistência dos dados extraídos, confrontando-os com o conteúdo original do edital, de modo a garantir a fidedignidade das evidências. Posteriormente, na fase de consolidação, os critérios de sustentabilidade previamente definidos são ponderados por seus respectivos pesos normativos e transformados em uma nota final para cada edital. Por fim, na etapa de resultado, o usuário recebe um relatório consolidado contendo não apenas as notas atribuídas, mas também as evidências textuais que fundamentaram a avaliação, assegurando transparência, reprodutibilidade e auditabilidade do processo.

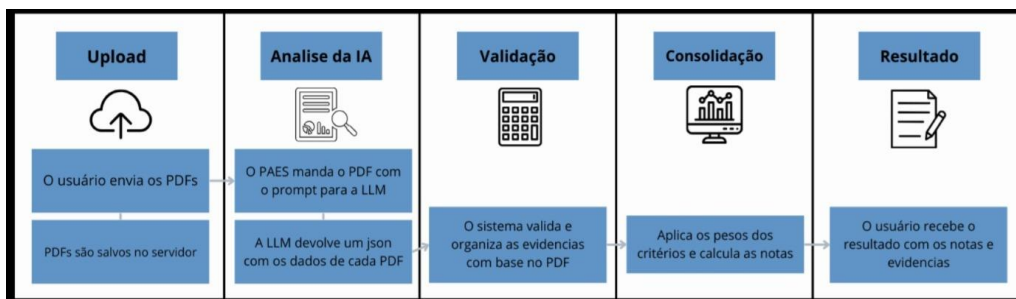


Imagem 1 – Fluxo de Funcionamento do PAES





#### 4. Resultados e Discussão

Importa esclarecer que, para fins de comparação, foram testadas duas variações do Gemini.

**Gemini com JSON:** técnica de *encadeamento de prompts*, em que inicialmente se solicita ao modelo a geração de um arquivo JSON com a presença ou ausência dos critérios, e em seguida pede-se que a LLM atribua a nota com base nesse JSON.

**Gemini sem JSON:** prompt único, em que o modelo deve identificar os critérios e atribuir diretamente a nota em uma mesma etapa.

Como o **PAES gera JSON por construção**, a variante *com JSON* buscou aproximar o comportamento do Gemini ao pipeline proposto. Entretanto, mesmo com esse encadeamento, o Gemini apresentou maior instabilidade e degradação de desempenho, confirmando que o **PAES é mais preciso, estável e escalável**.

A seguir serão abordadas análises comparativas de performance entre os modelos.

##### 4.1. Escalabilidade do PAES frente ao Gemini

A análise dos resultados evidencia a superioridade do sistema **PAES** em relação às variantes do Gemini. Como se observa no Gráfico 1 (a), os valores de **MAE** do Gemini (tanto na versão *com JSON* quanto *sem JSON*) apresentam tendência de crescimento à medida que aumenta o número de editais analisados, chegando a níveis de erro superiores a 0,35 em grupos maiores (20 e 25 editais). Em contraste, o **PAES mantém constante seu MAE em 0,125**, independentemente da quantidade de documentos processados.

No Gráfico 1 (b), que ilustra o **RMSE**, o mesmo comportamento se repete: enquanto o Gemini apresenta erros quadráticos médios crescentes, atingindo até 0,471 na versão *com JSON*, o **PAES mantém-se estável em 0,156**. Essa estabilidade é um indicativo de maior robustez metodológica, pois demonstra que a acurácia do PAES não se degrada quando mais editais são incluídos na análise.

(a)

(b)



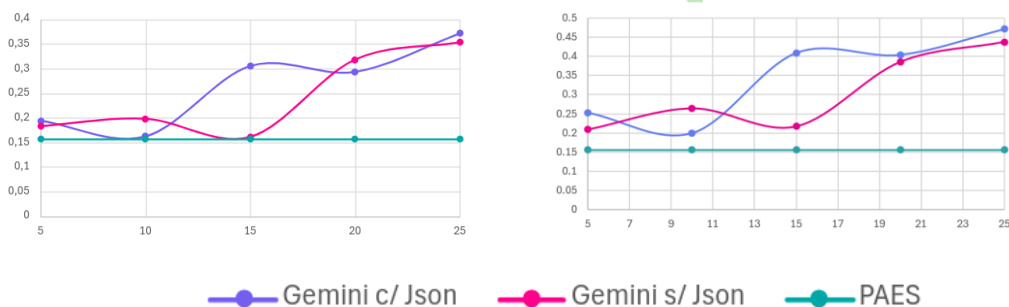


Gráfico 1 – Análise de Escalabilidade do Modelo PAES

Em termos práticos, o Gemini, quando exposto a cargas maiores de documentos, torna-se mais propenso a oscilações e erros, comprometendo a confiabilidade de seus resultados. Já o PAES, por operar com regras de validação e pesos fixos por critério, assegura **consistência e previsibilidade**, características fundamentais para aplicação em ambientes de auditoria pública.

#### 4.2.Avaliação por métricas clássicas de classificação

Para além das métricas de erro médio (MAE e RMSE), que oferecem uma noção global de proximidade em relação ao Padrão Ouro (item 4.1), é fundamental observar como os modelos se comportam quando analisados sob a ótica das métricas clássicas de classificação: **precisão, recall, acurácia e F1-score**. Essas medidas complementares permitem compreender a natureza dos erros cometidos por cada sistema e, principalmente, o impacto prático dessas falhas no contexto de auditoria de editais.

**Precisão** indica quantos editais identificados como sustentáveis pelo modelo realmente o eram. Alta precisão significa que o sistema quase não gera “falsos alarmes”, isto é, não confere sustentabilidade a editais que não a possuem.

**Recall (ou sensibilidade)** mede a capacidade do modelo de encontrar todos os editais que de fato deveriam ser classificados como sustentáveis. Recall elevado significa que poucos editais corretos ficam de fora.

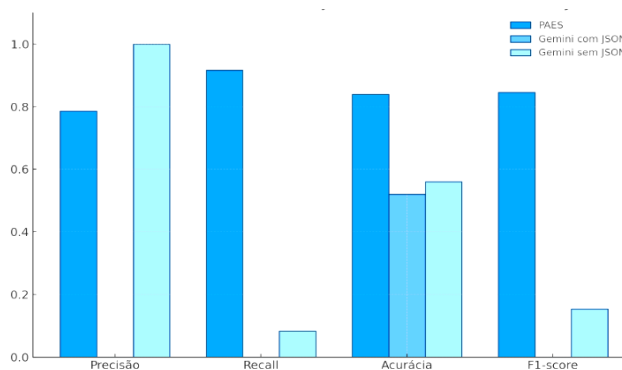
**Acurácia** avalia o percentual de classificações corretas considerando tanto positivos quanto negativos, fornecendo uma visão global do desempenho.

**F1-score** combina precisão e recall em uma única métrica equilibrada, sendo especialmente importante quando falsos positivos e falsos negativos trazem riscos relevantes.





O Gráfico 2 sintetiza esses resultados de forma comparativa. Observa-se que o **PAES alcançou os melhores valores em todas as métricas**, com precisão de 0,79, recall de 0,92, acurácia de 0,84 e F1-score de 0,85. Isso demonstra não apenas um alinhamento elevado ao Padrão Ouro, mas também um equilíbrio raro: o PAES consegue tanto evitar falsos positivos quanto capturar a grande maioria dos editais que realmente contemplam critérios de sustentabilidade. Essa combinação de desempenho torna-o uma ferramenta confiável para apoiar a tomada de decisão em auditorias, onde a consistência metodológica é indispensável.



**Gráfico 2 – Precisão, Recall, Acurácia e F1-Score dos Modelos**

O desempenho das variantes do **Gemini**, por outro lado, mostrou-se significativamente inferior. O Gemini com JSON, mesmo estruturado para se aproximar do pipeline do PAES, não conseguiu identificar corretamente os casos positivos, resultando em recall e F1-score iguais a zero — um desempenho que inviabiliza sua aplicação prática. Já o Gemini sem JSON atingiu precisão perfeita (1,0), mas com recall ínfimo (0,08), revelando um comportamento extremamente conservador: acerta quando atribui sustentabilidade, mas praticamente nunca a atribui, deixando de reconhecer quase todos os casos corretos.

Esses achados possuem relevância prática direta. Em auditoria pública, **falsos positivos** podem gerar validações indevidas de editais que não cumprem critérios normativos, comprometendo a legitimidade das compras. Por outro lado, **falsos negativos** podem levar à exclusão de editais que efetivamente atendem aos requisitos de sustentabilidade, inibindo boas práticas. O **PAES**, ao apresentar um desempenho elevado e equilibrado em todas as métricas, revela-se mais confiável e mais aderente às necessidades de controle externo.





Em síntese, a avaliação por métricas de classificação complementa as análises de erro médio e evidencia de forma ainda mais clara a **superioridade do PAES frente ao Gemini**. Enquanto o Gemini oscila entre extremos e apresenta inconsistências graves, o PAES se mostra robusto, previsível e escalável, características essenciais para a institucionalização de soluções baseadas em inteligência artificial no âmbito da auditoria pública.

### 4.3.Comparação da Proximidade com o Padrão Ouro

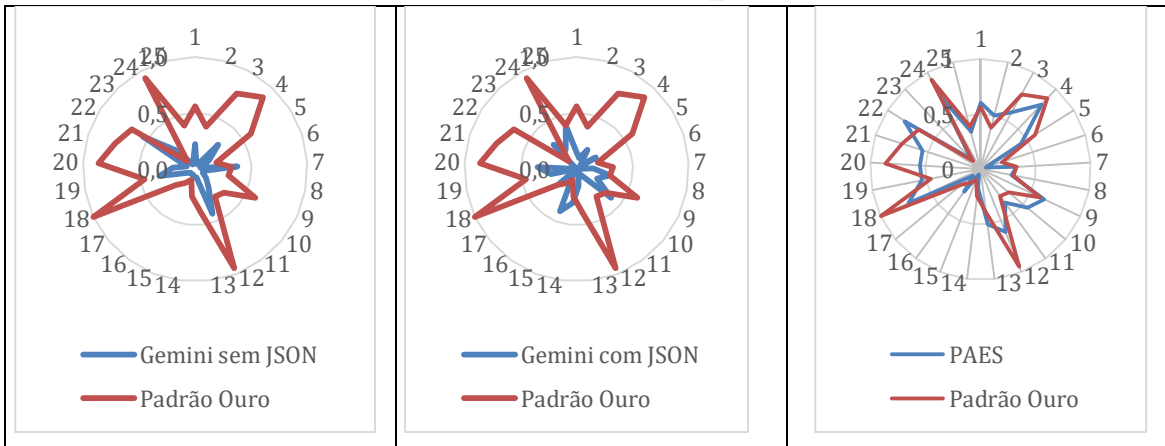
As três análises apresentadas (Gráfico 3) ilustram o **distanciamento bruto** entre as notas atribuídas pelos diferentes métodos e o **Padrão Ouro**, considerado como baseline de referência. Essa métrica é particularmente relevante porque permite avaliar não apenas o valor médio do erro, mas também a consistência do desempenho ao longo dos diferentes editais.

Observa-se que o **PAES apresenta uma trajetória muito mais próxima do Padrão Ouro** (Gráfico 3 (c)), com flutuações que acompanham de maneira mais fiel a linha de referência. Isso demonstra que o sistema, ao aplicar regras de validação e ponderações fixas, consegue manter uma aproximação estável e previsível em relação ao julgamento humano.

Por outro lado, as duas variantes do **Gemini** (com e sem JSON) revelam **maior dispersão e instabilidade** (Gráfico 3 (a,b)). Em diversos editais, especialmente nos grupos mais extensos, o Gemini se distancia de forma significativa do Padrão Ouro, evidenciando dificuldades em manter consistência. O modelo *com JSON* buscava replicar a lógica do PAES ao gerar um arquivo estruturado antes da atribuição da nota; entretanto, mesmo com esse artifício, os resultados mostraram-se mais erráticos e menos aderentes à linha de referência. O modelo *sem JSON*, embora em alguns casos se aproxime mais do Padrão Ouro, apresenta oscilações abruptas, comprometendo sua confiabilidade.

(a)	(b)	(c)
-----	-----	-----





**Gráfico 3 – Análise do distanciamento das notas dos modelos frente ao padrão ouro**

Esses resultados são relevantes porque confirmam que o **PAES não apenas atinge menor erro médio (MAE e RMSE (item 4.1)), como também entrega maior estabilidade e previsibilidade**, qualidades fundamentais em processos de auditoria e análise documental. A proximidade visual com o Padrão Ouro, como ilustrado nos gráficos, reforça a noção de que o PAES é o método que melhor traduz em escala os critérios normativos definidos, garantindo robustez mesmo quando a base de editais é expandida.

Em suma, a comparação gráfica fornece uma evidência empírica clara da **superioridade do PAES sobre as variantes do Gemini**, tanto em termos de acurácia quanto de consistência, o que o torna a ferramenta mais adequada para aplicações práticas na análise de editais de compras públicas.

## 5. Considerações Finais

Os resultados obtidos neste estudo demonstram que o **PAES** constitui uma alternativa metodologicamente mais robusta e confiável do que o uso isolado do modelo de linguagem **Gemini** para a análise de editais de compras públicas de materiais elétricos. Ao ancorar sua avaliação em **critérios objetivos e pesos normativamente fundamentados**, o PAES apresentou **desempenho estável** em termos de erro absoluto e quadrático médio (MAE e RMSE), além de superioridade consistente nas métricas de classificação (precisão, recall, acurácia e F1-score).

Enquanto o Gemini mostrou-se sensível ao aumento do número de documentos e apresentou variação significativa de resultados, o PAES manteve **uniformidade e**





**previsibilidade**, qualidades essenciais em ambientes de auditoria e controle externo. Essa característica garante maior **auditabilidade** e reduz os riscos de inconsistência, alucinação ou decisões não reprodutíveis, desafios recorrentes no emprego de LLMs em contextos públicos.

Do ponto de vista institucional, a adoção de soluções como o PAES pode contribuir para a **efetivação do princípio do desenvolvimento nacional sustentável** previsto na Lei nº 14.133/2021, bem como para a consolidação de práticas de governança alinhadas aos **Objetivos de Desenvolvimento Sustentável da ONU**. Além disso, evidencia-se que a integração entre inteligência artificial e critérios normativos representa uma estratégia promissora para fortalecer a **transparência, a integridade e a eficiência** das contratações públicas.

Por fim, propõe-se que pesquisas futuras aprofundem a aplicação do PAES em outros setores além dos materiais elétricos, ampliando o espectro de critérios avaliados e investigando a possibilidade de integração com sistemas de monitoramento em tempo real, de modo a consolidar um ecossistema de auditoria digital explicável e escalável. De forma complementar a comparação de performance com outros sistemas baseados em IA e mesmo outras IAs Generativas além do Gemini apresenta-se como objetivos viáveis.

Os algoritmos e bases de dados utilizadas neste artigo encontram-se disponíveis para consulta em <https://1drv.ms/f/c/1f8ab80c0c1306f5/Eo4-Ot3XTOJKt7z94zRvQsMBLfk5SrOa-drQJ6cYPYHTSg?e=je6Eft>





## Referências

ARRIETA, A. B. et al. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges. *Information Fusion*, v. 58, p. 82-115, 2020.

BRAMMER, S.; WALKER, H. Sustainable procurement in the public sector: an international comparative study. *International Journal of Operations & Production Management*, v. 31, n. 4, p. 452-476, 2011.

BRASIL. Decreto nº 7.746, de 5 de junho de 2012. Regulamenta critérios de sustentabilidade nas contratações públicas federais.

BRASIL. Lei nº 12.305, de 2 de agosto de 2010. Institui a Política Nacional de Resíduos Sólidos.

BRASIL. Lei nº 14.133, de 1º de abril de 2021. Lei de Licitações e Contratos Administrativos.

BRASIL. Lei nº 9.933, de 20 de dezembro de 1999. Dispõe sobre as competências do Conmetro e do Inmetro.

CARVALHO, F.; OLIVEIRA, T.; SANTOS, L. Compras públicas sustentáveis: uma análise das práticas no Brasil. *Revista de Administração Pública*, v. 54, n. 3, p. 650-672, 2020.

DOSHI-VELEZ, F.; KIM, B. Towards a rigorous science of interpretable machine learning. *arXiv preprint*, arXiv:1702.08608, 2017.

DWIVEDI, Y. K. et al. Generative AI for public sector: opportunities, challenges, and research agenda. *Government Information Quarterly*, v. 40, n. 4, p. 101794, 2023.

GIL-GARCIA, J. R.; DAWES, S.; PARDO, T. Digital government and public management research: finding the crossroads. *Public Management Review*, v. 20, n. 5, p. 633-646, 2018.

ONU. Transformando Nosso Mundo: A Agenda 2030 para o Desenvolvimento Sustentável. Nova Iorque: ONU, 2015.

PACHECO, R.; SAMPAIO, R. Sustentabilidade e inovação em contratações públicas. *Revista do Serviço Público*, v. 72, n. 1, p. 45-63, 2021.

PEREIRA, A. K. S. et al. Usando redes complexas na identificação de empresas fraudulentas em licitações públicas. In: *Anais do X Workshop de Computação Aplicada em Governo Eletrônico (WCGE 2022)*. Porto Alegre: SBC, 2022. p. 13-24.



SANTOS, L. M.; MORAES, J. A.; FERREIRA, M. Compras públicas sustentáveis em municípios brasileiros: avanços e desafios. *Revista do Serviço Público*, v. 70, n. 2, p. 201-220, 2019.

SOKOL, K.; FLACH, P. Explainability fact sheets: a framework for systematic assessment of explainable approaches. In: *Proceedings of the AAAI Conference on Artificial Intelligence*, v. 34, n. 4, p. 5602-5609, 2020.

SOUZA, A. P.; GOMES, R. C. Sustentabilidade e pesos diferenciados em compras públicas: um estudo exploratório. *Revista de Políticas Públicas*, v. 25, n. 1, p. 98-115, 2021.

SUN, T. Q.; MEDAGLIA, R. Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare. *Government Information Quarterly*, v. 36, n. 2, p. 368-383, 2019.

THAI, K. V. Public procurement re-examined. *Journal of Public Procurement*, v. 1, n. 1, p. 9-50, 2001.

WIRTZ, B. W. et al. Artificial Intelligence and the public sector—applications and challenges. *International Journal of Public Administration*, v. 42, n. 7, p. 596-615, 2019.

YEUNG, K. Algorithmic regulation: a critical interrogation. *Regulation & Governance*, v. 12, n. 4, p. 505-523, 2018.

ZUBOFF, S. *The Age of Surveillance Capitalism*. New York: PublicAffairs, 2019.

