Information Retrieval on Public Notices for Medicines' Purchasing: A Comparison Between an Ad Hoc approach and the GPT-4 LLM

Arthur L. Silva[1], **George G. Cabral**[1]

[1]Departamento de Computação – Universidade Federal Rural de Pernambuco (UFRPE)
Rua Manoel de Medeiros, S/N – CEP: 52171-900 – Recife – PE – Brasil

{arthur.limas,george.gcabral}@ufrpe.br

Abstract. *Auditing is an essential task when dealing with public expenses. Despite its importance, frequently auditing efforts must prioritize few targets due to the lack of human resources. When auditing public medicines acquisition processes, one may identify overpricing cases and (or) tackle ill-formed documents, for example. This work introduces a new method for precisely identify medicines given non standardized descriptions on public notices documents. Experiments conducted to evaluate the effectiveness of the proposed approach showed that it is, in average, 2.85 times more effective than a ChatGPT-4o assistant based on the same data. Yet, the proposed approach is not subject to LLMs problems such as hallucination.*

Resumo. *A auditoria é uma tarefa essencial no que se refere ao controle de despesas públicas. Apesar de sua importância, frequentemente tais esforços priorizam poucos alvos por conta da falta de recursos. A auditoria da compra pública de medicamentos, pode identificar sobrepreços e necessidade de ajustes em editais, por exemplo. Esse trabalho introduz um novo método para a identificação precisa de medicamentos dadas descrições não padronizadas em editais. Experimentos mostraram que a abordagem proposta é, em média, 2.85 vezes mais efetiva que um assistente usando o ChatGPT-4o utilizando os mesmos dados. Não obstante, a abordagem proposta não sofre de problemas como alucinação, inerentes a modelos LLM.*

1. Introduction

Auditing is a specific problem and inherently human task, however, in many cases, it can extremely benefit from computational tools [Silva et al. 2024, Emmanuel et al. 2023]. Despite their high level of specific characteristics, a common aspect for this task domain is the need of analyzing a large amount of data. Public expenses consist of a wide range of different classes of products, such as medical equipment, food supply, flight tickets, etc.. Usually, such expenses can only be performed through a formal procedure where a document (public notice) containing the desired items is publicly released so that a fair bidding procedure can happen. This is the case of the medicines purchasing by public agents.

A public notice for medicines purchasing procedure usually contains hundreds of different medicines descriptions. A human auditor needs to analyze whether each description is correct and, more than that, needs to check whether the proposed price of the medicine is acceptable or not. This is an exceedingly time consuming task since each medicine description may be produced by a number of different laboratories with different prices. I.e., for each medicine the auditor needs to know all corresponding products available in the market. With the unique identifiers of these products (i.e., a bar code or EAN) at hand, the auditor could, for example, automatically search past purchases of the same medicine to statistically identify whether the proposed price in the public notice is valid or not, for example.

This works proposes an ad-hoc method, referred to as IR-Med, for information retrieval in public notices for medicines purchasing.The outcome of the method can drastically reduce the human labor in sharply finding the corresponding medicines in the market and consequently scaling up the auditing capability of a human auditor. In addition, a LLM based method was also developed in order to provide a comparison of the results from both methods.

The remainder of the work is structured as follows: Section 2 presents some related works; Section 3 better describes the problem; Section 4 introduces the proposed approach; Section 5 presents the experiments and results discussion; and finally, Section 6 brings the conclusions and future works.

2. Related Works

Recent works tackled the problem of processing data from public notices automatically [Velasco et al. 2021, Brandão et al. 2023, Silva et al. 2024]. Velasco et al. presented a DDS (Decision Support System) that quantified dozens of risk patterns present in databases of biding processes. Brandão et al. presented PLUS (a semi-automated **P**ipe**L**ine for Fraud Detection in P**U**blic Bid**S**), a pipeline that detect documents in a bid with irregularities.

Lists of products, including medicines, are often available in a tabular format. Extracting the information from tabular data is still the topic of recent works and surveys [Zhang and Balog 2020, Liu et al. 2023]. Liu et al. classified approaches to deal with tabular data into three categories: (1) heuristic methods, that are algorithmic straightforward and don't require much effort in engineering or learning [Abdelmageed and Schindler 2021, Alobaid and Corcho 2022]; (2) Feature engineering based, that extract statistical and lexical features to use with machine learning models [Neumaier et al. 2016, Kacprzak et al. 2018]; and (3) Deep learning based [Zhou et al. 2021, Liu et al. 2022].

Recently, in the last few years, a high number of solutions for dolmain specific problems are using Large Language Models - LLMs [Chang et al. 2024, Wang et al. 2024b]. Furthermore, specific LLMs are being generated for specific domains such as software defect prediction [Wang et al. 2023] and software testing [Wang et al. 2024a]. Nonetheless, LLMs are also acknowledged by some key problems such as hallucination [Barman et al. 2024, de Wynter et al. 2023].

Despite of its weaknesses, by applying LLMs to auditing, it is possible to automate the screening of large volumes of data, identifying patterns and anomalies with greater accuracy and efficiency than traditional methods [Gu et al. 2024]. These models can process financial documents, identify discrepancies and even predict areas of risk based on historical trends [Abdullah and Almaqtari 2024]. In addition, LLMs can assist in regulatory analysis by supporting the verification of adherence to standards and policies, and contribute to a more agile audit with a reduced margin of error.

3. Problem Description

Brazilian municipalities produce and release public notices in order to purchase medicines for public hospitals and other public health services. Nonetheless, all public entities, as a rule, must acquire goods and services by means of a public bid proceeding that ensures equal conditions for all bidders[1].

Definition 1 (Active Pharmaceutical Ingredient) - a.k.a., API, are the raw material to produce medicines. It is the main substance in a medicine and gives its pharmaceutical characteristic. Nonetheless, many medicines are produced as a merging of different active ingredients and in different dosages. These cases pose a challenge to an automatic information retrieval method.

Definition 2 (Pharmaceutical Form) - the pharmaceutical form consists of the form a medicine is presented, e.g. tablet, capsule, solution for injection, cream, etc.

Predominantly, a public notice for medicines' purchasing contains a long list of medicines to be acquired. Each item of this list must be specified such that the bidders can undoubtedly identify the item, nonetheless, this specification must not trace the item to an unique supplier company. For example, the item acetaminophen is available through a variety of different brands, dosages (325 mg, 500 mg, etc.) and pharmaceutical forms (tablet, chewable tablet, liquid oral, etc.). So, the active ingredient (e.g., acetaminophen), dosage, pharmaceutical form and any other information to distinguish the item from other similar medicine must be present without citing any supplier brand or company. This is a rule that can be disrespected in few cases, however, this is not in the scope of this work.

Producing the list of medicines descriptions contained in the public notice is a human task. In addition, the way how the item is specified in the document does not follow a rigid standardized procedure. Therefore, the author of the public notice can split the information in many columns in a table, or worse, omit information. Auditing a public notice document consists, among other things, in evaluating whether or not each item is satisfactorily described. Unfortunately, frequently the auditing procedure is not conducted by an experienced medicine practitioner or a pharmacist. It is important to emphasize that each public notice contains hundreds of items to be purchased and the amount of notices to be audited, per state of the country, is often incompatible with the number of human auditors.

Given the problem of identifying a proper medicine description for each item in the public notice, a machine learning (or data science) practitioner may be promptly tempted to model a conventional classifier (e.g., a deep learning network or a random forest) in order to, given an input (the medicine description in the public notice) to produce an output containing the proper medicine description. This approach contains some challenges: (i) the lack of an annotated corpus; (ii) the extremely high number of potential classes (if we think in the number of ac-

[1]https://www.lexology.com/library/detail.aspx?g=6f266055-86ba-4849-8dd1-85e07e85b397

tive ingredients as the number of classes, this number is currently 2072 items); (iii) frequently an item in the public notice consists in a combination of many active ingredients, as aforementioned (i.e., in this case, the way the medicine is described hugely affects the classifier output); and (iv) other information such as dosage and form are often not standardized.

Figure 1 depicts some examples of how the medicines' items are described in a public notice (*in Portuguese*). Notice that there isn't a pattern on the information in each table cell. This challenges a suitable identification of a set of corresponding available products that match the description. In addition, item 2 of public notice 2 shows a case where a medicine is formed by more than one active ingredients.

Public Notice 1

ITEM	CATMAT		DESCRIÇÃO	UNIDADE DE FORNECIMENTO	QUANTIDADE	PREÇO DE REFERÊNCIA	VALOR TOTAL
1	BR	0271689	ÁCIDO ASCÓRBICO concentração/dosagem 200 mg/mL, forma farmacêutica Solução Oral - (Gotas) via de administração oral	Frasco 20 mL	198.500	R$ 1,34	265.990,00
2	BR	0278489	ÁCIDO FÓLICO concentração/dosagem 0,2 mg/mL, forma farmacêutica Solução, via de administração oral.	FRASCO 30 mL	213.400	R$ 4,30	917.620,00
3	BR	0315056	ÁGUA PARA INJEÇÃO	AMPOLA 10 mL	2.311.400	R$ 0,28	647.192,00
4	BR	0267509	ALOPURINOL concentração/dosagem 300 mg, forma farmacêutica Comprimido, via de administração oral.	COMPRIMIDO	377.200	R$ 0,30	113.160,00

Public Notice 2

ITEM	DESCRIÇÃO	VOLUME	QUANT.	VALOR UNITARIO ESTIMADO	VALOR TOTAL ESTIMADO
FARMÁCIA BÁSICA					
1	AMOXICILINA, CONCENTRAÇÃO: 500MG (ITEM EXCLUSIVO – ME/EPP)	COM	54000	R$ 0,78	R$ 42.120,00
2	AMOXICILINA + CLAVULANATO DE POTÁSSIO, CONCENTRAÇÃO: 50 MG,ML + 12,5 MG,ML, SUSPENSÃO ORAL - FRASCO 100 ML (ITEM EXCLUSIVO – ME/EPP)	FRA	1800	R$ 19,19	R$ 34.542,00
3	ACICLOVIR, DOSAGEM: 200 MG(ITEM EXCLUSIVO – ME/EPP)	FRA	4500	R$ 0,25	R$ 1.125,00
4	AMOXICILINA, CONCENTRAÇÃO: 50MG, ML, APRESENTAÇÃO: PÓ PARA SUSPENSÃO ORAL - FRASCO 60ML (ITEM EXCLUSIVO – ME/EPP)	FRA	2700	R$ 11,13	R$ 30.051,00

Public Notice 3

ITEM	CÓDIGO BPS	DESCRIÇÃO DO PRODUTO	QUANT.	UND	Média	V. Uni.	COTA
1	BR0268370	*ACICLOVIR, DOSAGEM: 200 MG*	187200	COMPRIMIDO	R$ 0,1800	R$ 33.696,00	EXCLUSIVO PARA ME/EPP
2	BR0274918	*ACETATO DE RETINOL 10.000UI/G + AMINOÁCIDOS 25MG/G + METIONINA 5 MG/G + CLORANFENICOL 5MG/G*	13	BISNAGA	R$ 11,6700	R$ 151,71	EXCLUSIVO PARA ME/EPP
3	BR0268375	*ACICLOVIR, DOSAGEM: 50 MG,G, USO: CREME. EMBALAGEM 10G*	14400	BISNAGA	R$ 2,1500	R$ 30.960,00	EXCLUSIVO PARA ME/EPP

Figure 1. Examples of how medicines' items are displayed in public notices.

3.1. The CMED list of available medicines

The CMED[2] (*Câmara de Regulação do Mercado de Medicamentos*) is a workgroup under the supervision of the national agency ANVISA (*Agência Nacional de Vigilância Sanitária*). Among others, this group is responsible by the inspection of the prices of medicines in Brazil. All medicines, represented by their structured information, available at the market are cataloged in a monthly updated report. This report releases a table where each row is formed by columns containing information such as: active ingredient; dosage; pharmaceutical form; bar code (i.e., each product available at the market, even similar medicines produced by different companies, have an unique bar code); name of the supplier company; etc. Figure 2

[2]https://www.gov.br/anvisa/pt-br/assuntos/medicamentos/cmed/precos

shows a range of rows of the CMED list. In these rows it is possible to find medicines having the same specifications (i.e., *substância* and *apresentação*) but produced by different laboratories.

SUBSTÂNCIA	LABORATÓRIO	EAN 1	PRODUTO	APRESENTAÇÃO
21-ACETATO DE DEXAMETASO	BAYER S.A.	7891106000956	BAYCUTEN N	10 MG/G + 0,443 MG/G CREM DERM CT BG AL X 40 G
ABATACEPTE	BRISTOL-MYERS SQUIBB	7896016806469	ORENCIA	250 MG PO LIOF SOL INJ CT 1 FA + SER DESCARTÁVEL
ABATACEPTE	BRISTOL-MYERS SQUIBB	7896016807442	ORENCIA	125 MG/ML SOL INJ SC CT SER PREENCHIDA
ABATACEPTE	BRISTOL-MYERS SQUIBB	7896016808197	ORENCIA	125 MG/ML SOL INJ SC CT 4 SER PREENC VD TRANS + DISPOSITIVO
ABCIXIMABE	JANSSEN-CILAG FARMA(	7896212452453	REOPRO	2 MG/ML SOL INJ CT FA VD INC X 5 ML
ABCIXIMABE	ELI LILLY DO BRASIL LTDA	7896382701801	REOPRO	2 MG/ML SOL INJ CT FA VD INC X 5 ML
ABEMACICLIBE	ELI LILLY DO BRASIL LTDA	7896382708442	VERZENIOS	50 MG COM REV CT BL AL AL X 30
ABEMACICLIBE	ELI LILLY DO BRASIL LTDA	7896382708459	VERZENIOS	50 MG COM REV CT BL AL AL X 60
ABEMACICLIBE	ELI LILLY DO BRASIL LTDA	7896382708466	VERZENIOS	100 MG COM REV CT BL AL AL X 30
ABEMACICLIBE	ELI LILLY DO BRASIL LTDA	7896382708473	VERZENIOS	100 MG COM REV CT BL AL AL X 60
ABEMACICLIBE	ELI LILLY DO BRASIL LTDA	7896382708480	VERZENIOS	150 MG COM REV CT BL AL AL X 30
ABEMACICLIBE	ELI LILLY DO BRASIL LTDA	7896382708497	VERZENIOS	150 MG COM REV CT BL AL AL X 60
ABEMACICLIBE	ELI LILLY DO BRASIL LTDA	7896382708503	VERZENIOS	200 MG COM REV CT BL AL AL X 30
ABEMACICLIBE	ELI LILLY DO BRASIL LTDA	7896382708510	VERZENIOS	200 MG COM REV CT BL AL AL X 60
ABROCITINIBE	PFIZER BRASIL LTDA	7891045164542	CIBINQO	50 MG COM REV CT FR PLAS PEAD OPC X 30
ABROCITINIBE	PFIZER BRASIL LTDA	7891045164559	CIBINQO	100 MG COM REV CT FR PLAS PEAD OPC X 30
ABROCITINIBE	PFIZER BRASIL LTDA	7891045164566	CIBINQO	200 MG COM REV CT FR PLAS PEAD OPC X 30
ACALABRUTINIBE	ASTRAZENECA DO BRASI	5000456031998	CALQUENCE	100 MG CAP DURA CT BL AL AL X 60
ACARBOSE	EMS SIGMA PHARMA LT	7894916503754	AGLUCOSE	50 MG COM CT BL AL AL X 30

Figure 2. Part of the CMED list of medicines available in the Brazilian market.

4. IR-Med - An Ad Hoc Information Retrieval Approach for Medicines' Purchasing Public Notices

Problem definition - Given a poorly standardized description of a medicine in a public notice, return all the rows and respective EANs (i.e., bar codes) in the CMED table corresponding to this description.

4.1. IR-Med - Modeling Phase

The modeling phase consists in: (i) the pre-processing of the CMED columns *substância* (active ingredients) and *apresentação* (a column comprising the remainder information of a medicine, i.e., form, dosage, etc.); (ii) clustering CMED rows belonging to the same active ingredient; and (iii) relevant tokens extraction of the CMED columns.

4.1.1. Words pre-processing

This phase handles data from both columns of the CMED list. The following treatments are performed: (i) converting the words to lowercase; (ii) removal of the accents; (iii) removval of special characters (i.e., a character not in the characters intervals a-z, A-Z, 0-9, including _); (iv) inserting blank space between numbers and words; (v) based on the ANVISA vocabulary [3], abbreviating the forms (e.g., *comprimido* to *com*); (vi) removal of numbers from the active ingredients descriptions; (vii) removal of stopwords; (viii) removal of repeated words; and (ix) removal of ions and associated chemical compounds such as *cloreto*, *permanganato*, *sulfato*, *brometo*, etc (these terms have shown a detrimental effect in the medicine identification).

[3] https://www.gov.br/anvisa/pt-br/centraisdeconteudo/publicacoes/medicamentos/publicacoes-sobre-medicamentos/vocabulario-controlado.pdf

4.1.2. Clustering CMED items by their Active Ingredients

Once the initial pre-processing is carried out, the rows are grouped by their active ingredients. In case of more than one active ingredient, the words are sorted based on their lexical order. Table 1 shows an example of active ingredients and their respective of bar codes (i.e., EANs). This information is then stored in a hash table structure.

Table 1. Stored information on the grouped-cmed hash table.

Active Ingredient(s)	Sorted Active Ingredient(s)	List of rows indexes
zidovudina	zidovudina	[29369, 29370, 29371]
zidovudina lamivudina	zidovudina lamivudina	[29372, 29373, 29374, 29375]
zinco	zinco	[411, 27728, 27729, ...]
zinco nitrato nafazolina	nafazolina nitrato zinco	[27742, 27743]

4.1.3. Identifying relevant words

Once the grouped-cmed hash table is built, two set are created: (i) **cmed-ai-words** containing all different words in the pre-processed active ingredients column of the CMED table; and (ii) **cmed-pr-words** containing all different words in the presentation column of the CMED table. The words in these sets are used to precisely identify the information on the items descriptions in a public notice.

As an example, consider that after pre-processing a item description of a public notice the following sentence is returned: "amoxicilina 500 mg clavulanato potassio 125 com". The terms amoxicilina and potassio will be associated to the active ingredient and 500 mg 125 com will be acknowledged as the medicine form, dosage, etc. (i.e., present in the *apresentação* column of the CMED table).

4.2. Information Retrieval Phase

Given a medicine description in a public notice, the words in the description are pre-processed according to Section 4.1.1 and compared to the words contained in the sets **cmed-ai-words** and **cmed-pr-words** in order to separate the active ingredients from the remaining information. This description is then splited in two sentences: (i) **desc-ai** - information regarding the active ingredient and (ii) **desc-pr** - information regarding the dosage, form, etc.

Next, **desc-ai** is compared to each active ingredient in **grouped-cmed** hash table and the most similar one is retrieved. This similarity is computed according to the Jaro-Wrinkler similarity (Eq. 1).

$$sim_j(w_1, w_2) \begin{cases} 0 & \text{if } m = 0 \\ \frac{1}{3}\left(\frac{m}{|w_1|} + \frac{m}{|w_2|} + \frac{m-t}{m}\right) & \text{otherwise} \end{cases} \tag{1}$$

where m is the number of matching characters, t is the number of transpositions, and $|w_1|$ and $|w_2|$ are, respectively, the size of the words w_1 and w_2.

Finally, the words in **desc-pr** are then compared to the set (R) of rows associated to the previously found active ingredient. For each row in R, the intersection between the set of words in the column *apresentação* and the set of words in **desc-pr** is computed and the rows in R with higher intersection are then retrieved. At this stage, returning no rows from the CMED list is a high evidence of poorly written description in the public notice.

5. Experiments and Analysis of the Results

5.1. Experimental Setup

For the experiments, ten public notices for medicines' purchasing from municipalities of Pernambuco state were analyzed. For short, we will refer to these public notices as PN_1 to PN_{10}. These documents have 200, 105, 330, 356, 290, 238, 242, 169, 293, 70 medicines' descriptions, respectively.

The results obtained by the IR-Med were compared to an assistant model based on the GPT-4o with 128k context length. This assistant was developed to search in the same CMED table (i.e., the document containing the CMED items was uploaded and processed by the GPT-4o model) of our proposed approach. The following prompt was used to retrieve the set of medicines corresponding to each description in the public notice (in portuguese):

> forneça uma tabela indexada contendo todos as substâncias, apresentações e códigos de barras de todas as linhas do arquivo cujo conteúdo inicial seja semelhante ao medicamento: "***desc_med***". As palavras na descrição do medicamento e nas linhas do arquivo podem estar embaralhadas, sem o espaçamento adequado, abreviadas ou conterem erros gramaticais. Desconsidere letras maiúsculas e minúsculas.

desc_med consists of the description of the medicine in the public notice.

The performance of both methods will be assessed based on the following criteria: (i) accuracy on the information retrieval (i.e., percentage of elements in a public notice whose all retrieved CMED rows truly correspond to the element description) and (ii) how reliable is the GPT-4o in terms of hallucination for this task?

The value for the first criterion was obtained based on a manual evaluation (carried out by two persons) of the results.

5.2. Analysis of the Results

Table 2 depicts the results for each public notice document and information retrieval method. The IR-Med approach successfully retrieved valid CMED rows for over 80% of the public notices items' descriptions (for 7 out of 10 documents). Still, for the remaining documents, IR-Med yielded performances above 70%. On the other hand, the ChatGPT-4o assistant model retrieved an average accuracy of 30.9%. Nevertheless, the proposed method performed, in average, 2.85 times better than the LLM approach.

It is important to mention that the category of ill-formed medicines descriptions (i.e., items in the public notices where it is not possible to find a corresponding list of medicines in the CMED list) is included in the IR-Med error cases. This is

Table 2. Overall results of the IR-Med versus ChatGPT-4o.

Public Notice	IR-MED (acc)	ChatGPT 4o (acc)	GPT-4o Halluc. rate
PN_1	79.00 %	26.53 %	6.12 %
PN_2	86.92 %	36.92 %	6.92 %
PN_3	80.98 %	41.30 %	7.85 %
PN_4	86.52 %	39.19 %	6.63 %
PN_5	72.41 %	24.82 %	3.19 %
PN_6	81.51 %	22.08 %	6.06 %
PN_7	75.62 %	30.21 %	5.96 %
PN_8	84.62 %	38.75 %	8.13 %
PN_9	81.91 %	33.57 %	4.64 %
PN_{10}	84.29 %	15.94 %	7.25 %

the case, for example of medicines associated to dosages or forms that do not exist at the market. These cases comprise 6.00%, 1.54%, 3.61%, 1.69%, 5.17%, 4.20%, 4.13%, 3.55%, 4.10% and 2.86% of the items from public notices 1 to 10, respectively. These cases were also manually identified.

Table 2 also shows the percentage of items in each public notice affected by the LLMs hallucination phenomenon [Barman et al. 2024, de Wynter et al. 2023]. The rates range from 3.19% to 8.13%. This is a particularly problematic aspect since the lack of confidence in the results may affect the reputation of the solution. Another observed problem, however in smaller scale, is the return of duplicate rows. Therefore, a LLM based approach for the present problem clearly needs an extra processing phase in order to validate the returned data.

6. Conclusion and Future Works

Auditing activities are inherently manual, i.e., non automated. However, parts of an auditing process can be automated in order to tackle the usual large amount of data. In this direction, examining public notices for medicines purchasing is a time consuming and error prone task. Since that for this problem a long list of medicines must be thoroughly examined, a human auditor can be affected by fatigue consequently lowering the quality of his/her work.

This work introduced an ad hoc method, referred to as IR-Med, capable of, given a non standardized description of a medicine, correctly identify a list of correspondent medicines in a public catalog of medicines available at the market and maintained by the national regulatory health agency. The results show that the proposed approach is able to correctly identify, in most of cases, more than 80% of the medicines described in the public notice document. In addition, it also identifies cases were there is the need of further improvement in the medicine description. The work also showed that an LLM based approach might not be suitable since it consistently mix different medicines for a same item description. Still, it also suffers from complex problems such as hallucination and duplicated items retrieval.

A number of future works can be derived from the present study. For example: (i) investigating the public purchase of items other than medicines; (ii) improving the results achieved in the work and for a larger number of public notices; (iii) incorporate other phases of an auditing process such as the investigation of overpricing; and (iv) combine different approaches such as LLMs and other methods in order to obtain more reliable and autonomous solutions.

References

Abdelmageed, N. and Schindler, S. (2021). Jentab meets semtab 2021's new challenges. In *SemTab@ ISWC*, pages 42–53.

Abdullah, A. A. H. and Almaqtari, F. A. (2024). The impact of artificial intelligence and industry 4.0 on transforming accounting and auditing practices. *Journal of Open Innovation: Technology, Market, and Complexity*, 10(1):100218.

Alobaid, A. and Corcho, O. (2022). Balancing coverage and specificity for semantic labelling of subject columns. *Knowledge-Based Systems*, 240:108092.

Barman, D., Guo, Z., and Conlan, O. (2024). The dark side of language models: Exploring the potential of llms in multimedia disinformation generation and dissemination. *Machine Learning with Applications*, 16:100545.

Brandão, M. A., Reis, A. P., Mendes, B. M., de Almeida, C. A. B., Oliveira, G. P., Hott, H., Gomide, L. D., Costa, L. L., Silva, M. O., Lacerda, A., et al. (2023). Plus: A semi-automated pipeline for fraud detection in public bids. *Digital Government: Research and Practice*.

Chang, Y., Wang, X., Wang, J., Wu, Y., Yang, L., Zhu, K., Chen, H., Yi, X., Wang, C., Wang, Y., Ye, W., Zhang, Y., Chang, Y., Yu, P. S., Yang, Q., and Xie, X. (2024). A survey on evaluation of large language models. 15(3).

de Wynter, A., Wang, X., Sokolov, A., Gu, Q., and Chen, S.-Q. (2023). An evaluation on large language model outputs: Discourse and memorization. *Natural Language Processing Journal*, 4:100024.

Emmanuel, Y., Silva, F., Cabral, G., and Valença, G. (2023). Inovação na contabilidade pública - uma solução que analisa atrasos de pagamentos em municípios pernambucanos. pages 123–125.

Gu, H., Schreyer, M., Moffitt, K., and Vasarhelyi, M. (2024). Artificial intelligence co-piloted auditing. *International Journal of Accounting Information Systems*, 54:100698.

Kacprzak, E., Giménez-García, J. M., Piscopo, A., Koesten, L., Ibáñez, L.-D., Tennison, J., and Simperl, E. (2018). Making sense of numerical data-semantic labelling of web tables. In *Knowledge Engineering and Knowledge Management: 21st International Conference, EKAW 2018, Nancy, France, November 12-16, 2018, Proceedings 21*, pages 163–178. Springer.

Liu, J., Chabot, Y., Troncy, R., Huynh, V.-P., Labbé, T., and Monnin, P. (2023). From tabular data to knowledge graphs: A survey of semantic table interpretation tasks and methods. *Journal of Web Semantics*, 76:100761.

Liu, J., Huynh, V.-P., Chabot, Y., and Troncy, R. (2022). Radar station: Using kg embeddings for semantic table interpretation and entity disambiguation. In *International Semantic Web Conference*, pages 498–515. Springer.

Neumaier, S., Umbrich, J., Parreira, J. X., and Polleres, A. (2016). Multi-level semantic labelling of numerical values. In *The Semantic Web–ISWC 2016: 15th International Semantic Web Conference, Kobe, Japan, October 17–21, 2016, Proceedings, Part I 15*, pages 428–445. Springer.

Silva, A., Sampaio, V., Lima, A., Cabral, G., and Valença, G. (2024). Ferramenta para auxílio à auditoria de editais municipais para compra de medicamentos. In *Anais Estendidos do XX Simpósio Brasileiro de Sistemas de Informação*, pages 265–268. SBC.

Velasco, R. B., Carpanese, I., Interian, R., Paulo Neto, O. C., and Ribeiro, C. C. (2021). A decision support system for fraud detection in public procurement. *International Transactions in Operational Research*, 28(1):27–47.

Wang, J., Huang, Y., Chen, C., Liu, Z., Wang, S., and Wang, Q. (2024a). Software testing with large language models: Survey, landscape, and vision. *IEEE Transactions on Software Engineering*, 50(4):911–936.

Wang, L., Ma, C., Feng, X., Zhang, Z., Yang, H., Zhang, J., Chen, Z., Tang, J., Chen, X., Lin, Y., Zhao, W. X., Wei, Z., and Wen, J. (2024b). A survey on large language model based autonomous agents. *Frontiers of Computer Science*, 18(6).

Wang, Y., Le, H., Gotmare, A., Bui, N., Li, J., and Hoi, S. (2023). CodeT5+: Open code large language models for code understanding and generation. In Bouamor, H., Pino, J., and Bali, K., editors, *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 1069–1088.

Zhang, S. and Balog, K. (2020). Web table extraction, retrieval, and augmentation: A survey. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 11(2):1–35.

Zhou, Y., Singh, S., and Christodoulopoulos, C. (2021). Tabular data concept type detection using star-transformers. In *Proceedings of the 30th ACM International Conference on Information & Knowledge Management*, pages 3677–3681.