

Inteligência Artificial Generativa (IAG): perspectivas inéditas no controle público

Alessandro Marinho de Albuquerque

Doutorando em Ciência da Computação pelo Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Santa Catarina (PPGCC/UFSC). Auditor fiscal de Controle Externo do Tribunal de Contas do Estado de Santa Catarina (TCE/SC). *E-mail:* alessandro.marinho@tcesc.tc.br.

Igor May Wensing

Bacharel em Ciência da Computação pela Universidade Federal de Santa Catarina (UFSC). Cientista de dados. *E-mail:* igormaywensing@gmail.com.

Nelson Luiz Joppi Filho

Bacharel em Ciência da Computação pela Universidade Federal de Santa Catarina (UFSC). Cientista de dados. *E-mail:* nelsonluizjoppifilho@gmail.com.

Resumo: Na esteira de um governo cada vez mais digital, o volume, a variedade e a velocidade das informações somente aumenta, originando um fenômeno chamado dados escuros. Tal fenômeno ocorre quando a capacidade de produção/custódia dessas informações superam a capacidade de curadoria ou análise. As consequências de um dado escuro afetam a tomada de decisão de gestores públicos, bem como representam riscos de ordem de não conformidade com normas. Na função de controle público, os riscos são maiores ainda, pois como orientar ou detectar irregularidades em um cenário em que necessita de rápida resposta decorrente da velocidade dos dados, do grande volume e da alta variedade de estrutura de dados? Para essa resposta, este trabalho apresenta uma perspectiva inovadora do uso de Inteligência Artificial Generativa (IAG) para ser aplicado no controle público em vários cenários mapeados no Tribunal de Contas de Santa Catarina (TCE/SC).

Palavras-chave: Controle público. Dados escuros. Inteligência Artificial Generativa (IAG).

Sumário: 1 Introdução – 2 Dados escuros – 3 Inteligência Artificial Generativa (IAG) – 4 Perspectiva de uso de IAG no controle externo – 5 Considerações finais – Referências

1 Introdução

De acordo com relatório intitulado *2022 GovTech Maturity Index Update*, do Banco Mundial, que versa acerca o grau de maturidade de governo digital no mundo, o Brasil tem se destacado como um dos líderes de governo digital (“Grupo A”), que reúne 43 países (Dener *et al.*, 2021, p. 24).

Nesse mesmo relatório, o Brasil se destaca com boas práticas relacionadas ao Portal Gov.br, com mais de 80 milhões de usuários com acesso a serviços públicos digitais, e à plataforma Fala.BR, que permite uma interação entre cidadãos e governo. Destaca-se também com a nova estratégia de governo digital do Brasil, focada em políticas e serviços orientados ao cidadão (Dener *et al.*, 2021, p. 56).

O volume e a variedade de informações por meio de uma sociedade cada vez mais digitalizada são crescentes. Esse cenário também se repete para os governos, em especial o do Brasil, em que continua fortemente investindo em digitalização.

Diante desse cenário, a capacidade de aumento de curadoria das informações não acompanha a sua produção, ocasionando o fenômeno de dados escuros (Munot *et al.*, 2019, p. 3). Os dados escuros, para o governo, representam diversos riscos como a falha de suporte à decisão (Munot *et al.*, 2019, p. 4) e o vazamento de dados (Moumeni *et al.*, 2021, p. 218).

Destaca-se que o desconhecimento de dados, seja qual for a sua proporção, desencadeia aumento de risco associado à execução do ato de governo ou da política pública (Gertler *et al.*, 2016, p. 3). Cumpre que essas ações sejam detectadas e corrigidas pela função de controle de governo.

O controle, quando considerado do ponto de vista contábil, representa um sistema abrangente composto por quatro componentes integrados: (1) o elemento de detecção, responsável por identificar e reagir a eventos dentro do processo controlado; (2) o elemento de avaliação, encarregado de determinar a importância desses eventos, comparando-os a padrões predefinidos; (3) o componente executor, capaz de ajustar o comportamento do processo conforme as ocorrências relevantes identificadas pelo avaliador; e (4) o elemento de comunicação, que interliga os elementos de detecção, avaliação e execução por meio do compartilhamento de informações (Anthony; Govindarajan, 2008, p. 2).

Na gestão pública do Brasil, conforme estipulado pela Constituição Federal de 1988, o controle desempenha dois papéis distintos: o interno e o externo (Di Pietro, 1999, p. 1027). O controle interno é essencial para o funcionamento adequado da entidade. Sua ausência ou fragilidade expõe a entidade a desvios, malversação de recursos e práticas fraudulentas. Em sua essência, o controle interno, conforme definido na Constituição, desempenha um papel fundamental na construção da confiança nos controles externo e social.

Por outro lado, o controle externo exercido pelo Poder Legislativo e por órgãos como os tribunais de contas opera sob a perspectiva dos “riscos de auditoria”, com ênfase no chamado “risco de controle”. Este risco diz respeito à possibilidade de os sistemas contábeis e de controle interno falharem em prevenir, detectar e corrigir prontamente distorções que possam causar danos aos ativos da entidade (Instituto Rui Barbosa, 2022, p. 17).

Assim, a função de controle em um governo digital deve ser repensada para que suas ações mitiguem o surgimento de dados escuros, com o objetivo de detectar irregularidades e orientar o gestor público para melhor execução da política pública. Soma-se a isso a necessidade de agir de forma concomitante mediante a disponibilidade de dados governamentais de grande volume e a ampla diversidade de atividades de governo.

Nesse contexto, as tecnologias analíticas emergem como uma ferramenta crucial para impulsionar a atividade de controle do governo, tais como a Big Data, permitindo integração de múltiplas fontes em diferentes formatos, bem como a coleta de tais dados (Kim; Trimi; Chung, 2014, p. 80). Com mais agências de governo utilizando ferramentas de Big Data para transformar políticas públicas e órgãos, os líderes de governo se perguntam: Que descobertas podemos fazer com Big Data? Que análises fazemos com os dados? Como confiar nos resultados? (Joseph; Johnson, 2013, p. 47).

Recentemente a IAG, com a evolução dos modelos *transformers* e o surgimento do ChatGPT da OpenAI, revolucionou o mundo, promovendo mudanças em todas as escalas, incluindo o próprio governo digital. Os efeitos da IAG são promissores e expandem inúmeras possibilidades em todas as facetas da sociedade, e isso inclui a função do controle.

Diante disso, este artigo propõe trazer elementos do impacto de modelos de IAG no controle externo, citando inclusive exemplos de uso. A metodologia utilizada foi exploratória, elencando possíveis casos de uso a serem investigados para contribuição desse tipo de tecnologia no controle externo e também de forma científica.

2 Dados escuros para o controle

Uma das primeiras definições de dados escuros surgiu em 2008. Dados escuros são aqueles que não são previamente indexados e armazenados, o que os torna de difícil acesso para cientistas e outros usuários em potencial. Como resultado, esses dados são mais propensos a serem subutilizados e eventualmente perdidos (Heidorn, 2008, p. 282).

Outro trabalho associa dados escuros a dados que foram “cortados”. Tais dados são escuros por conta de que seu conhecimento não alcança o âmbito científico e a comunidade (Goetz, 2007, p. 1).

Alguns estudos investigam a relevância única que dados escuros podem ter para avanços científicos (Goetz, 2007). No contexto organizacional, Moumeni *et al.* (2021) destacam a importância de extrair novos *insights* a partir de dados escuros, os quais podem agregar valor aos negócios das organizações e contribuir para uma tomada de decisão mais eficaz.

A tomada de decisão é um aspecto em que dados escuros podem desempenhar um papel crucial. De fato, há um potencial ganho de valor em explorá-los, assim como um risco significativo em não o fazer. Há relato de associação de dados escuros ao risco de falhas no suporte à decisão, ressaltando que esses dados podem conter informações relevantes para o processo decisório. Quando permanecem no estado de dados escuros, essas informações ficam ocultas para a organização, o que representa uma oportunidade perdida e um risco potencial (Munot *et al.*, 2019, p. 4).

Em recente mapeamento acerca de dados escuros para o controle, Albuquerque e Dornelles (2022, p. 84) elaboraram um mapa conceitual que traz as principais características: sem curadoria atual; não documentados; não gerenciados; coletados para uma finalidade mas não para outras; inexplorados ou não analisados e dados estruturados ou não.

Os autores supracitados pontuam que o governo tem uma enorme capacidade de criar sistemas e coletar dados; contudo, boa parte desses dados permanecem inexplorados, o que se atribui à falta de curadoria, de tecnologia analítica e de mão de obra especializada. Outra característica dada pelos autores se dá na falta de capacidade de integrar esses dados para geração de um novo conhecimento, a exemplo do caso de “falecidos na folha de pagamento”, por eles citado.

Quanto à natureza dos dados escuros e seu impacto, Albuquerque e Dornelles (2022, p. 83) pontuam:

O fato de uma parcela significativa dos dados escuros serem não estruturados reflete basicamente a grande capacidade do governo realizar seus atos a partir da geração de documentos eletrônicos (contratos, atos de admissão, empenhamento, emissão de nota fiscal, memorandos, ofícios etc.), os inúmeros dados decorrentes de interações sociais por meio de aplicativos de mensagens corporativos. Especialmente no contexto da continuidade do trabalho remoto, iniciado em decorrência da pandemia de COVID-19 em meados de março de 2020. Tais documentos não estruturados tem um potencial enorme para o exercício de controle público. Sua natureza não estruturada, volume e variabilidade, dificulta o alcance do controle. Nesse cenário, artefatos probatórios para fins de investigação que poderiam ser o elo para responsabilização ou expansão de atividades de fiscalização/investigação permanecem escuros ao controle.

Quanto a exemplos de dados escuros no contexto de controle público, Albuquerque e Dornelles demonstram os principais na Tabela 1:

Tabela 1 – Tipos de dados escuros para controle público

Tipo de dados	Formato
Documentos de comunicação oficial	PDFs, arquivos texto
Documentos de admissão de pessoal, despesa e comprovação, processos licitatórios	PDFs, imagens, arquivos e textos
<i>E-mails</i> e aplicativos de comunicação oficial	Arquivos texto, imagens e vídeos
Sistemas legados	Tabelas relacionais
Documentos de arrecadação	Arquivos PDF e texto
Logs de sistemas	Arquivos texto
Informações obtidas de cruzamentos de dados	Arquivos PDF, tabelas relacionais
Tabelas relacionais legadas	Tabelas relacionais

Fonte: Albuquerque e Dornelles (2022, p. 84).

Diante das várias características apresentadas, a literatura busca responder a uma questão comum: como reduzir a “longa cauda” para que dados escuros se tornem conhecidos e bem utilizados? As possíveis soluções para lançar luz sobre esses dados começaram com a proposta de criar centros de ciência segregados por disciplinas, bibliotecas e museus (Heidorn, 2008, p. 293).

Posteriormente, surgiram as primeiras propostas de ferramentas, como as apresentadas por Shukla (2015), De Sa (2016) e Schembera (2021). Também há a sugestão de criar um papel específico de dados escuros nas organizações (Schembera; Dúran, 2020). Outras iniciativas envolvem o uso de algoritmos de aprendizado de máquina e de processamento de linguagem natural (Munot *et al.*, 2019, p. 6).

3 Inteligência Artificial Generativa (IAG)

As tecnologias de IAG representam uma mudança de paradigma na IA, avançando além do simples processamento e da análise de dados existentes para a criação de conteúdo totalmente novo (Vrabie, 2023, p. 1).

Esses sistemas utilizam técnicas avançadas de aprendizado de máquina, especialmente grandes modelos de linguagem treinados (*large language models*) em vastos conjuntos de dados, para gerar saídas semelhantes às humanas em várias modalidades, como texto, imagem, áudio e código.

Ao detectar padrões nos dados de treinamento, os modelos de IAG conseguem produzir conteúdo coerente e contextualmente relevante, a partir de entradas recebidas (*prompts*) (Loukis *et al.*, 2023, p. 109).

Sistemas como o ChatGPT exemplificam essa capacidade generativa no campo do processamento de linguagem natural. Com arquiteturas baseadas em transformadores, esses modelos de linguagem conseguem realizar trocas escritas com fluência, compreendendo e respondendo a consultas com saídas textuais e contextualmente apropriadas, que muitas vezes exibem características de comunicação semelhantes às humanas (Dwivedi *et al.*, 2023, p. 1).

A habilidade de produzir conteúdo sob demanda que imita de perto a expressão humana abre possibilidades transformadoras em diversos setores. Na esfera de governo, a IAG pode revolucionar a forma como os cidadãos interagem com o governo, acessam informações e participam dos processos de tomada de decisão, através de interfaces conversacionais intuitivas e de geração personalizada de conteúdo (Vrabie, 2023, p. 6)

Ainda no contexto de governo, os grandes modelos de linguagem (LLMs) são ideais para auxiliar atores de integridade na automação de certas atividades de detecção de fraudes, como a consulta de documentos e de fontes de dados para identificar riscos potenciais (Ugale; Hall, 2024, p. 15).

Eles também podem ajudar auditores e investigadores a realizar diversas tarefas operacionais que, embora não sejam exclusivas dos atores de integridade, são especialmente promissoras devido ao grande volume de documentação e de dados que os órgãos de auditoria, anticorrupção e investigação geralmente processam. Por exemplo, os LLMs podem organizar grandes volumes de texto para facilitar a priorização e o consumo, além de auxiliar em análises de causas raiz ou no reconhecimento de padrões (Ugale; Hall, 2024, p. 15).

Alguns países, como a Suécia, estão desenvolvendo assistentes virtuais para uso em toda a Administração Pública, para otimizar essas tarefas operacionais (Ugale; Hall, 2024, p. 15).

As eficiências obtidas com essas técnicas podem reduzir tanto o esforço quanto os erros, permitindo que auditores e investigadores se concentrem em tarefas de alto valor que exigem julgamento e expertise humanos, algo que a IAG ainda não substitui. Ao tornar as atividades de combate à corrupção e à fraude mais eficazes, a IAG pode também fortalecer a integridade e a transparência (Ugale; Hall, 2024, p. 15).

4 Perspectivas de uso de IAG no controle

4.1 Licitações

O caso de uso de licitações consagra o uso de modelos generativos de IA como principal cenário da atualidade de dados escuros – grande volume (centenas

de licitações diariamente no país), alta variedade (compras de diversos objetos: obras, tecnologia da informação, material permanente, serviços etc.) – e requer uma velocidade de análise rápida por parte do controle público, de preferência antes da data de abertura.

Para o controle, analisar o inteiro teor do edital em conjunto com as peças (termo de referência, estudo técnico preliminar etc.) em busca de irregularidades é uma tarefa de leitura extensiva, que consome horas de trabalho de agentes de controle, tarefa que, além disso, é impossível de escalar para dezenas ou centenas de licitações que acontecem diariamente.

Nesse cenário, automatizar a leitura e fazer os modelos generativos ir atrás de irregularidades é o objetivo para auxiliar na tomada de decisão, diminuindo o custo de obtenção e conclusão de irregularidades pelo agente de controle. Para esse contexto, iniciativa do TCE/SC chamada VigIA permite a análise de centenas de editais de licitação.

O VigIA é um arcabouço tecnológico composto de ferramentas de Big Data e IAG capaz integradas por meio de dados.

O VigIA funciona a partir da criação de questões que podem ser divididas em categorias. Tais questões são executadas de forma automática em cada licitação que é enviada ao TCE/SC, em um *pipeline* automático sem intervenção humana, a partir de uma integração com um modelo generativo de IA que faz uso do modelo GPT 4o. As perguntas de auditoria cadastradas pelos auditores são executadas de forma automática em cada edital, gerando respostas pelo modelo generativo, conforme demonstra a Figura 1:

Figura 1 – Funcionamento básico do VigIA



Fonte: Dados da pesquisa (2024).

Em dois meses de implantação, o VigIA depurou mais de 2.000 licitações com mais de 10.000 perguntas executadas, apontando indícios de irregularidades

e regulares. Alguns desses resultados foram confirmados por auditores fiscais e encaminhados ao jurisdicionado responsável pela licitação, retificando ou suspendendo ou anulando o edital.

O VigIA demonstrou de forma inédita a capacidade de automação de análise de editais de licitação, e alguns desafios de ordem científica e técnica surgiram, como a formação de uma base consolidada de perguntas de regularidade, com *prompts* associados de orientação de resposta do modelo, e de métricas para computação de acurácia de cada pergunta.

Outro caso de uso que o VigIA habilitou foi outro ineditismo que a possibilidade de análise de política pública de transporte escolar questionando em cada licitação de edital de transporte escolar as aderências às normas de Código de Trânsito Brasileiro (CTB) e ano de fabricação dos veículos.

Para além desses desafios, abriu oportunidades e questionamentos tais como: É possível classificar objeto de licitação? O VigIA pode analisar planilhas de custos para detectar sobrepreços? É possível que o próprio modelo faça perguntas de irregularidades? Certamente são questões que os órgãos de controle irão explorar e que contribuirão para compreender os limites e as possibilidades do uso desses modelos generativos de IA.

4.2 Concurso público

De forma inédita, modelos de IAG podem realizar a automação de análises de textos de editais de concurso público. Não há trabalhos científicos que aplicam técnicas computacionais especificamente nesse objeto, tampouco projetos ligados ao controle que exploram especificamente o dado escuro contido no edital de concurso público.

Trabalhos ligados a automação de análise de pessoal são feitos pelo Tribunal de Contas do Estado do Santa Catarina (TCE/SC),¹ Tribunal de Contas do Estado do Ceará (TCE/CE),² Tribunal de Contas do Distrito Federal (TCDF),³ Tribunal de Contas do Estado do Paraná (TCE/PR)⁴ e Tribunal de Contas do Estado do Pará (TCE/PA),⁵ além do Tribunal de Contas da União (TCU).⁶ Há relatos também

¹ Disponível em: https://www.tcesc.tc.br/sites/default/files/leis_normas/PORTARIA%20CONJUNTA%20N.%20TC-PRES-GCG-MPTC-01-2023%20CONSOLIDADA.pdf. Acesso em: 3 set. 2024.

² Disponível em: <https://www.tce.ce.gov.br/comunicacao/noticias/2230-tce-conclui-primeira-fase-do-projeto-de-automatizacao-dos-atos-de-pessoal>. Acesso em: 3 set. 2024.

³ Disponível em: <https://atrimon.org.br/praticas-exitosas-com-novas-tecnologias-sao-apresentadas-no-2-labtcs/>. Acesso em: 3 set. 2024.

⁴ Disponível em: <https://atrimon.org.br/praticas-exitosas-com-novas-tecnologias-sao-apresentadas-no-2-labtcs/>. Acesso em: 3 set. 2024.

⁵ Disponível em: <https://atrimon.org.br/praticas-exitosas-com-novas-tecnologias-sao-apresentadas-no-2-labtcs/>. Acesso em: 3 set. 2024.

⁶ Disponível em: <https://portal.tcu.gov.br/imprensa/noticias/uso-de-inteligencia-artificial-aprimora-processos-internos-no-tcu.htm>. Acesso em: 3 set. 2024.

de projetos ligados à análise automática de folha de pagamento realizados por alguns tribunais.

Não há relatos de que esses projetos permeiem o dado não estruturado do edital do concurso público. Tudo indica que parte dos dados do concurso público e dos registros de admissão é no formato estruturado, e a base desses trabalhos, parte de dados estruturados. Estruturá-los a partir de um documento não estruturado é uma tarefa normalmente declaratória por um ser humano, sendo passível de erros, e demanda processos de intervenção manual à medida que o dado do edital for sendo atualizado, o que aumenta o risco também de o dado estruturado não refletir o dado não estruturado.

Outro indicativo importante é que os trabalhos de automação mapeados atuam com ênfase no processo de admissão, ou seja, após a homologação do concurso. A tarefa de automação de análise de editais de concurso público em seu formato não estruturado permanece um desafio.

Modelos grandes de linguagem possuem potencial de analisar esses documentos buscando informações sobre aderência de vagas PCD/Cotas com a ampla concorrência, exigência de comprovação excessiva ou inadequada de tempo de exercício de profissão para determinados cargos, requisitos desnecessários para a posse, tais como idade ou graduação incompatível com a área do cargo, prazos exíguos das fases do edital.

Os modelos de linguagem permitem, assim, explorar irregularidades que, mesmo em dados estruturados, são difíceis de se detectarem, além de abrir várias outras oportunidades para explorar o texto original do edital do concurso público.

4.3 Peças orçamentárias

As peças orçamentárias são instrumentos fundamentais para a gestão pública, desempenhando um papel importante na administração dos recursos financeiros de um governo. As principais peças orçamentárias são o Plano Plurianual (PPA), a Lei de Diretrizes Orçamentárias (LDO) e a Lei Orçamentária Anual (LOA). Cada uma dessas peças possui características e finalidades específicas que contribuem para a eficiência e a transparência na alocação de recursos públicos.

O PPA estabelece as diretrizes, objetivos e metas da administração pública para um período de quatro anos. É um instrumento de planejamento de médio prazo, que orienta a elaboração dos orçamentos anuais, visando à continuidade dos programas e dos projetos governamentais. A LDO orienta a elaboração e a execução do orçamento anual, estabelecendo as metas e prioridades do governo para o exercício financeiro seguinte. A LOA, por sua vez, estima as receitas e fixa as despesas para o exercício financeiro anual. Ela é a peça orçamentária mais detalhada e operacional, sendo essencial para a execução das políticas públicas.

Trabalhos de automação de análise dessas peças orçamentárias não são triviais, dado o conteúdo detalhado e diverso. Por essa razão, não foram encontrados projetos ligados à automação de análise dessas peças.

Contudo, a IAG pode apoiar a atividade de controle na análise dessas peças, desde a aplicação de análises de sentimentos de *feedback* da população sobre os objetivos e as metas dessas peças, identificando áreas maiores de preocupação ou interesse público, auxiliando na aderência dessas peças ao interesse público.

A possibilidade de a IAG extrair informações relevantes de documentos externos, como as peças orçamentárias, permite identificar prioridades e objetivos principais sem leitura extensiva e fazer a comparação automática entre essas peças de diferentes períodos ou diferentes regiões, identificando tendências e mudanças ao longo do tempo. A identificação de indícios de irregularidades a partir de contradições nos objetivos e metas também é uma linha de atuação a ser explorada.

4.4 Reconhecimento de entidade nomeada

O reconhecimento de entidades nomeadas Named Entity Recognition (NER) é de uma subárea de estudo no campo de extração de informações, cujo objetivo é identificar entidades nomeadas, bem como classificá-las dentro de um conjunto de categorias pré-definidas, tais como pessoa, organização e local, as quais remetem a um referente específico (Fonseca *et al.*, 2015, p. 1).

Dentro desse contexto, a entrada para um sistema de extração de entidades nomeadas é um texto em sua forma livre, e sua saída é um conjunto de textos anotados, ou seja, uma representação estruturada a partir da entrada de um texto não estruturado (Fonseca *et al.*, 2015, p. 1).

Os métodos de reconhecimento de entidade nomeada atuais são predominantemente supervisionados. No aprendizado supervisionado, os modelos são treinados usando um conjunto de dados anotados, em que as entidades nomeadas estão explicitamente marcadas nas sentenças. Esse processo de anotação fornece ao modelo exemplos claros do que constitui uma entidade em várias categorias, permitindo que o modelo aprenda a reconhecer e classificá-la em novos textos (Ehrmann *et al.*, 2023, p. 10).

Os métodos supervisionados são mais custosos devido à necessidade de construir uma base de dados rotulada para treinamento. Em contraste, os métodos não supervisionados, que não requerem rotulação, têm despertado um interesse científico crescente (Ehrmann *et al.*, 2023, p. 10).

No âmbito do controle, identificar pessoa física e jurídica em textos não estruturados é um passo importante para a aplicação de regras, como a verificação de sanção que impediria que tal pessoa de ter vínculo com a Administração Pública, ou até mesmo a construção estruturada de relação de vínculo da pessoa

identificada com o documento para fins de responsabilização ou investigação. Padrões de telefones, endereços e placas de veículos também são aplicáveis para a finalidade de construção de vínculos.

Para além da necessidade específica de controle, essa detecção de entidades pode ser bastante útil para o mapeamento da Lei Geral de Proteção de Dados (LGPD).

Modelos de linguagem possuem capacidade de identificar pessoas físicas, jurídicas (Jiang *et al.*, 2024, p. 1), locais (Wang *et al.*, 2023, p. 1), telefones ou outras classificações, que podem ser realizados no contexto de negócio do controle, como a classificação de documentos tais como notas fiscais, editais de licitação ou documentos pessoais.

A grande vantagem do uso de modelos de linguagem nessa situação é que não há necessidade de supervisão ou de preparação de base de treinamento. Um indicativo de estudo para a abordagem é o uso de *few-shot prompts* para fazer com que o modelo entenda os exemplos associados de entidades nomeadas.

4.5 Análise de atos de diário oficial

Os atos administrativos são publicados em plataformas de diário oficial, de domínios de informações variadas que vão desde atos de pessoal (nomeação, exoneração, posse etc.), atos jurídicos (contratos, licitações) e até decretos, leis. As informações são de caráter não estruturado e não possuem padrão de escrita, podendo variar de ente para ente, o que dificulta o processamento de dados.

Há iniciativas por parte de alguns tribunais de contas para explorar essa base de dados: do TCE/SC,⁷ do TCU,⁸ do Tribunal de Contas do Estado do Rio Grande do Sul (TCE-RS)⁹ e dos tribunais de contas dos municípios do Pará.¹⁰ Há diversas utilidades dos referidos tribunais para varredura, tais como a validação de informações remetidas em suas bases internas, a identificação de indícios de irregularidades, a verificação de produtos de uso incomum pela administração e a identificação de pessoas físicas e jurídicas.

Não há trabalhos científicos mapeados relacionados ao uso de modelos de linguagem em bases de dados de diários oficiais; contudo, dado seu potencial generalista, o seu uso certamente será explorado pelos órgãos de controle para não somente acima, mas em conjunto com a possibilidade de classificação da publicação e q conseqüente extração de informação estruturada.

⁷ Disponível em: <https://www.tcsc.tc.br/robo-desenvolvido-pelo-tcsc-cruza-dados-e-alerta-orgaos-publicos-sobre-pessoas-fisicas-e-juridicas>. Acesso em: 3 set. 2024.

⁸ Disponível em: <https://www.tce.go.gov.br/Noticia/Detalha?noticia=16251>. Acesso em: 3 set. 2024.

⁹ Disponível em: <https://siga.tce.rs.gov.br/issues/233>. Acesso em: 3 set. 2024.

¹⁰ Disponível em: <https://atricao.org.br/tcmpa-implanta-robo-para-aumentar-fiscalizacao-em-licitacoes-relacionadas-ao-combate-a-covid-19/>. Acesso em: 3 set. 2024.

Esse tipo de análise em saídas estruturadas feita pelo modelo de IA abre oportunidades de cruzamento de dados com fontes internas do órgão de controle, para geração de novos conhecimentos e consequente ação. Um exemplo é a publicação de um edital de concurso público no Diário Oficial, com o modelo de IA identificando que se trata de um ente em particular, disparando, assim, uma checagem para verificar se aquele ente não está no limite de alerta e assim já notificar o órgão de controle.

5 Considerações finais

O advento do governo digital transformou o cenário do gerenciamento de informação no governo. Contudo, uma significativa porção dessas informações permanecem inutilizadas, escondidas ou fora do conhecimento do próprio governo. Essas características fazem parte do fenômeno de dados escuros.

Dados escuros representam riscos e oportunidades para o governo. A tomada de decisão baseada nesses dados pode melhor guiar uma política pública ou uma determinada operação realizada pelo agente governamental, além de aumentar a transparência e a mitigação de riscos, como o vazamento de dados, e de classificar as informações de maneira mais adequada.

Para a função de controle do governo, lançar luz sobre dados escuros pode ser determinante para um bom planejamento de fiscalização, detecção e prevenção de irregularidades/não conformidades.

Logo, torna-se necessário investimento em tecnologias analíticas para explorar dados escuros e transformá-los em ativos de conhecimento do órgão de controle. Nessa esteira, os modelos de IA generativos se destacam pela alta capacidade de compreensão textual.

Muito embora os primeiros casos de uso de tais modelos no âmbito do controle foram baseados em *chats* para busca de respostas de forma mais célere, o caso do TCE/SC, com seu sistema VigIA, chama atenção por ser o primeiro capaz de operar em um nível de automação considerando alto volume, velocidade e variedade de dados.

O VigIA foi aplicado no contexto de editais de licitação; contudo, a margem de atuação dele se expande para outros casos de uso, como para os editais de concurso público, com capacidade de detectar irregularidades nesses documentos, assim como para a análise de peças orçamentárias como Plano Plurianual (PPA), Lei de Diretrizes Orçamentárias (LDO) e Lei Orçamentária Anual (LOA).

O VigIA também possui capacidade de detectar entidades nomeadas em textos não estruturados como nomes de pessoas, CPFs, CNPJs e endereços.

A leitura de diários oficiais também é um caso de uso que pode ser aplicado para detecção de irregularidades. Toda essa análise com suporte de agentes de controle em busca de irregularidades, e quando tais irregularidades são apontadas pelo modelo, a confirmação pelo agente de controle ainda assim é necessária.

O uso de modelos de IAG ainda está na fase experimental, de forma que mais casos de uso aplicado ao controle irão aparecer, permitindo que uma evolução em várias estruturas de controle pelo país surja. Com a evolução do mercado global de tais produtos, mais oportunidades surgirão, consolidando a tecnologia como a última geração de análise de IA.

Generative Artificial Intelligence (GAI): Perspectives in public control

Abstract: In the wake of an increasingly digital government, the volume, variety, and velocity of information are only growing, giving rise to a phenomenon called dark data. This phenomenon occurs when the capacity to produce/store this information exceeds the capacity for curation or analysis. The consequences of dark data affect the decision-making of public managers and represent risks of non-compliance with regulations. In the function of public control, the risks are even greater, as how can one guide or detect irregularities in a scenario that requires a quick response due to the speed of data, the large volume, and the high variety of data structures? To address this issue, this work presents an innovative perspective on the use of generative artificial intelligence to be applied in public control in various scenarios mapped by the Court of Auditors of Santa Catarina – “TCE/SC.

Keywords: Public control. Dark data. Generative Artificial Intelligence (GAI).

Referências

ALBUQUERQUE, A. M. de; DORNELES, C. F. Dados escuros à luz do controle público. *In: SIMPÓSIO BRASILEIRO DE BANCOS DE DADOS*, 37., [S. l.]. *Anais (...)*. [S. l.]: SBC, 2022. p. 78-89.

ANTHONY, R.; GOVINDARAJAN, V. *Sistemas de controle gerencial*. São Paulo: AMGH Editora, 2008.

CAPRARO, V. *et al.* The impact of Generative Artificial Intelligence on socioeconomic inequalities and policy making. *PNAS Nexus*, [S. l.], v. 3, n. 6, 2024.

DE SA, C. *et al.* Deepdive: Declarative knowledge base construction. *ACM SIGMOD Record*, [S. l.], v. 45, n. 1, p. 60-67, 2016.

DENER, C.; NII-APONSAH, H.; GHUNNEY, L. E.; JOHNS, K. *GovTech maturity index: The state of public sector digital transformation*. [S. l.]: World Bank Publications, 2021.

DI PIETRO, M. S. Z. *Direito Administrativo*. São Paulo: Atlas, 1999.

DWIVEDI, Y. *et al.* So what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. *International Journal of Information Management*, [S. l.], v. 71, 2023.

EHRMANN, M. *et al.* Named entity recognition and classification in historical documents: A survey. *ACM Computing Surveys*, [S. l.], v. 56, n. 2, p. 1-47, 2023.

FONSECA, E.; CHIELE, G.; VANIM, A.; VIEIRA, R. Reconhecimento de entidades nomeadas para o português usando o opennlp. *In: ENIAC. Anais (...)*. [S. l.]: ENIAC, 2015.

GERTLER, P. *et al.* *Impact evaluation in practice*. [S. l.]: World Bank Group, 2016. Disponível em: <https://openknowledge.worldbank.org/server/api/core/bitstreams/4659ef23-61ff-5df7-9b4e-89fda12b074d/content>. Acesso em: 5 jul. 2024.

GOETZ, T. Freeing the dark data of failed scientific experiments. *Wired Magazine*, [S. l.], v. 15, n. 10, p. 15-10, 2007.

HEIDORN, P. B. Shedding light on the dark data in the long tail of science. *Library Trends*, [S. l.], v. 57, n. 2, p. 280-299, 2008.

INSTITUTO RUI BARBOSA. *NBASP 4000: Norma para Auditoria de Conformidade*. Rio de Janeiro: IRB, 2022. Disponível em: <https://nbasp.irbcontas.org.br/wp-content/uploads/2022/11/NBASP-4000-Norma-de-Auditoria-de-Conformidade.pdf>. Acesso em: 5 jul. 2024.

JIANG, G. *et al.* *P-ICL: Point In-Context Learning for Named Entity Recognition with Large Language Models*. arXiv preprint arXiv:2405.04960. [S. l.]: [s. n.]: 2024.

JOSEPH, R. C.; JOHNSON, N. A. Big Data and transformational government. *IT Professional*, [S. l.], v. 15, n. 6, p. 43-48, 2013.

KIM, G.-H.; TRIMI, S.; CHUNG, J.-H. Big Data applications in the government sector. *Communications of the ACM*, [S. l.], v. 57, n. 3, p. 78-85, 2014.

LOUKIS, E. *et al.* ChatGPT Application vis-a-vis Open Government Data (OGD): Capabilities, Public Values, Issues and a Research Agenda. *In: INTERNATIONAL CONFERENCE ON ELECTRONIC GOVERNMENT. Proceedings (...)*. Cham: Springer Nature Switzerland, 2023. p. 95-110.

MUNOT, K. *et al.* Importance of dark data and its applications. *In: INTERNATIONAL CONFERENCE ON SYSTEM, COMPUTATION, AUTOMATION AND NETWORKING (ICSCAN). Proceedings (...)*. [S. l.]: IEEE, 2019. p. 1-6.

MOUMENI, L. *et al.* Dark data as a new challenge to improve business performances: Review and perspectives. *In: INTERNATIONAL CONFERENCE ON DIGITAL AGE & TECHNOLOGICAL ADVANCES FOR SUSTAINABLE DEVELOPMENT. Proceedings (...)*. [S. l.]: IEEE, 2021. p. 216-220.

SCHEMBERA, B.; DÚRAN, J. Dark data as the new challenge for Big Data Science and the introduction of the scientific data officer. *Philosophy & Technology*, [S. l.], v. 33, n. 1, p. 93-115, 2020.

SCHEMBERA, B. Like a rainbow in the dark: Metadata annotation for HPC applications in the age of dark data. *The Journal of Supercomputing*, [S. l.], v. 77, n. 8, p. 8946-8966, 2021.

SHUKLA, M. *et al.* POSTER: WinOver enterprise dark data. *In: ACM SIGSAC CONFERENCE ON COMPUTER AND COMMUNICATIONS SECURITY, 22., [S. l.]. Proceedings (...)*. [S. l.]: ACM, 2015. p. 1674-1676.

UGALE, G.; HALL, C. *Generative AI for anti-corruption and integrity in government: Taking stock of promise, perils and practice*. [S. l.]: [s. n.], 2024.

VRABIE, C. E-Government 3.0: An AI model to use for enhanced local democracies. *Sustainability*, [S. l.], v. 15, n. 12, p. 9572, 2023.

WANG, S. *et al.* *Gpt-ner*: Named entity recognition via large language models. arXiv preprint arXiv:2304.10428. [S. l.]: [s. n.], 2023.

Informação bibliográfica deste texto, conforme a NBR 6023:2018 da Associação Brasileira de Normas Técnicas (ABNT):

ALBUQUERQUE, Alessandro Marinho de; WENSING, Igor May; JOPPI FILHO, Nelson Luiz. Inteligência Artificial Generativa (IAG): perspectivas inéditas no controle público. *Revista do Tribunal de Contas do Estado de Santa Catarina – RTCE/SC*, Belo Horizonte, ano 2, n. 3, p. 155-169, maio/out. 2024. DOI: 10.52028/tce-sc.v02.i03.ART08.SC
