

pesquisa_projeto

2025-05-18

Conteúdo

1	Dados Abertos Governamentais	2
2	Dados Abertos Governamentais	2
2.1	Conceituação e Princípios dos Dados Abertos Governamentais	2
2.2	Metadados e sua Importância para os Dados Abertos Governamentais	2
2.3	Desafios na Disponibilização e Utilização de Dados Abertos Governamentais	3
2.4	Potencial dos LLMs na Análise de Dados Abertos Governamentais	3
2.5	Perspectivas para a Adoção de LLMs na Análise de Dados Abertos Governamentais	3
3	Large Language Models e IA Generativa	4
4	Large Language Models e IA Generativa	4
4.1	Fundamentos e Evolução dos Modelos de Linguagem de Grande Escala	4
4.2	Características e Capacidades dos LLMs	4
4.3	Aplicações dos LLMs em Análise de Dados	4
4.4	LLMs na Análise de Documentos e Relatórios	5
4.5	Desafios Éticos e Limitações	5
4.6	Perspectivas Futuras e Implicações para a Pesquisa	6
5	Aplicações de LLMs no Setor Público	6
6	Aplicações de Modelos de Linguagem de Grande Escala no Setor Público	6
6.1	LLMs como Ferramentas de Auditoria Cidadã	6
6.2	Aprimorando a Descoberta de Dados Governamentais Abertos	7
6.3	Combate à Desinformação no Setor Público	7
6.4	Desafios e Considerações Éticas	7
6.5	Perspectivas para a Pesquisa	8
7	Metodologias e Frameworks para Análise de Dados	8
8	Metodologias e Frameworks para Análise de Dados	8
8.1	Ecossistemas de Dados Públicos e sua Evolução	8
8.2	Análise de Texto para Co-criação no Setor Público	8
8.3	Frameworks para Avaliação de Riscos e Conformidade Legal	9
8.4	Integração de Metodologias para Análise de Dados Governamentais com LLMs	9
8.5	Considerações Finais	9
9	Transformação Digital no Setor Público	10
10	Transformação Digital no Setor Público	10
10.1	Fundamentos da Transformação Digital Governamental	10
10.2	Ciência de Dados como Catalisador da Transformação Digital	10
10.3	Desafios e Oportunidades na Era dos Dados Abertos e LLMs	10

10.4	Implicações para a Pesquisa e Prática	11
10.5	Considerações Finais	11
11	Síntese da Fundamentação Teórica	11

Fundamentação Teórica

Este capítulo apresenta a fundamentação teórica que sustenta esta pesquisa sobre adoção de llms na análise de dados abertos governamentais. A revisão da literatura foi conduzida seguindo os critérios de inclusão e exclusão estabelecidos, resultando na seleção de 141 artigos que abordam especificamente a intersecção entre Large Language Models (LLMs) e dados abertos governamentais.

O objetivo desta fundamentação é analisar a adoção dos llms com o objetivo de caracterizar do ponto de vista dos pesquisadores no contexto da análise de dados abertos governamentais, organizando o conhecimento existente em categorias temáticas que permitam uma compreensão abrangente do estado atual da pesquisa nesta área emergente.

1 Dados Abertos Governamentais

2 Dados Abertos Governamentais

Os Dados Abertos Governamentais (DAG) representam um paradigma fundamental na era da transparência e da governança digital, constituindo-se como informações produzidas ou custodiadas pelo poder público que são disponibilizadas à sociedade sem restrições de acesso, permitindo sua reutilização em diferentes contextos e para diversos fins. Este conceito está intrinsecamente ligado aos princípios de governo aberto, transparência pública e participação cidadã, elementos essenciais para o fortalecimento democrático na sociedade contemporânea.

2.1 Conceituação e Princípios dos Dados Abertos Governamentais

Os Dados Abertos Governamentais fundamentam-se em princípios que visam garantir o acesso irrestrito e a utilização efetiva das informações públicas. Conforme apontam Maratsi et al. [?], a crescente quantidade de dados disponibilizados através de portais de dados abertos acompanha o impulso e a proliferação de iniciativas de Dados Abertos que, além dos numerosos benefícios oferecidos, apresenta uma realidade que traz consigo diversos desafios. Entre estes princípios, destacam-se a completude, a primariedade, a atualidade, a acessibilidade, a processabilidade por máquinas, o acesso não discriminatório, a utilização de formatos não proprietários e o licenciamento livre.

A implementação efetiva destes princípios visa não apenas disponibilizar informações, mas fazê-lo de maneira que promova a interoperabilidade, a descoberta e, conseqüentemente, o consumo de dados abertos pela sociedade. Neste contexto, a qualidade e a gestão de metadados emergem como aspectos fundamentais, cuja ausência pode comprometer significativamente o potencial de utilização dos dados disponibilizados [?].

2.2 Metadados e sua Importância para os Dados Abertos Governamentais

Os metadados, compreendidos como "dados sobre dados", desempenham papel crucial no ecossistema de Dados Abertos Governamentais. Eles fornecem informações estruturadas sobre os conjuntos de dados, facilitando sua descoberta, compreensão e utilização. A análise realizada por Maratsi et al. [?] sobre aspectos semânticos de portais internacionais de dados e seus esquemas de metadados mapeados para as principais classes do esquema DCAT (Data Catalog Vocabulary) evidencia a importância deste elemento para a efetividade das iniciativas de dados abertos.

A qualidade dos metadados impacta diretamente na interoperabilidade entre diferentes sistemas e portais de dados, permitindo que informações de diversas fontes possam ser integradas e analisadas conjuntamente. Além disso, metadados bem estruturados facilitam a descoberta dos dados pelos usuários, potencializando seu consumo e aplicação em diferentes contextos, desde a pesquisa acadêmica até o desenvolvimento de aplicações e serviços inovadores para a sociedade.

2.3 Desafios na Disponibilização e Utilização de Dados Abertos Governamentais

Apesar dos avanços significativos na área, a disponibilização e utilização efetiva de Dados Abertos Governamentais ainda enfrentam desafios consideráveis. Conforme destacado por Maratsi et al. [?], a quantidade crescente de dados disponíveis através de portais de dados abertos, embora represente um avanço importante, também traz consigo complexidades que precisam ser adequadamente endereçadas.

Entre os principais desafios, destacam-se:

1. ****Qualidade e padronização dos metadados****: A ausência de padrões consistentes para metadados compromete a interoperabilidade e a descoberta dos dados, limitando seu potencial de utilização.
2. ****Fragmentação de iniciativas****: A multiplicidade de portais e plataformas, muitas vezes com abordagens distintas para a disponibilização de dados, dificulta a integração e análise conjunta das informações.
3. ****Barreiras técnicas e semânticas****: Diferenças nos formatos, estruturas e terminologias utilizadas podem criar obstáculos significativos para a compreensão e utilização dos dados por diferentes públicos.
4. ****Sustentabilidade das iniciativas****: Garantir a continuidade, atualização e evolução das plataformas de dados abertos representa um desafio importante para as instituições públicas.

Estes desafios evidenciam a necessidade de abordagens inovadoras que possam potencializar a análise e utilização dos Dados Abertos Governamentais, contexto no qual os Modelos de Linguagem de Grande Escala (LLMs) emergem como ferramentas promissoras.

2.4 Potencial dos LLMs na Análise de Dados Abertos Governamentais

Os Modelos de Linguagem de Grande Escala (LLMs) representam uma evolução significativa no campo da Inteligência Artificial, oferecendo capacidades avançadas de processamento e compreensão de linguagem natural. No contexto dos Dados Abertos Governamentais, estas tecnologias apresentam potencial transformador para superar alguns dos desafios previamente identificados.

A capacidade dos LLMs de processar e interpretar grandes volumes de informações textuais não estruturadas ou semi-estruturadas pode contribuir significativamente para a análise de conjuntos de dados governamentais, especialmente quando estes apresentam inconsistências ou limitações em seus metadados. Como apontado por Maratsi et al. [?], a qualidade dos metadados é um aspecto crítico que afeta a interoperabilidade, a descoberta e o consumo de dados abertos. Neste cenário, os LLMs podem atuar como facilitadores, auxiliando na interpretação e contextualização das informações disponíveis.

Além disso, os LLMs podem potencializar a análise integrada de dados provenientes de diferentes fontes e formatos, contribuindo para superar a fragmentação das iniciativas de dados abertos. Sua capacidade de compreender contextos e estabelecer relações semânticas entre diferentes conjuntos de informações representa um avanço significativo para a extração de insights relevantes a partir dos dados governamentais.

2.5 Perspectivas para a Adoção de LLMs na Análise de Dados Abertos Governamentais

A adoção de LLMs para a análise de Dados Abertos Governamentais representa um campo de pesquisa emergente, com potencial para transformar significativamente a forma como as informações públicas são utilizadas e interpretadas. Esta abordagem alinha-se ao objetivo de caracterizar, do ponto de vista dos pesquisadores, a aplicação destas tecnologias no contexto da análise de dados governamentais.

A perspectiva dos pesquisadores sobre a adoção dos LLMs neste contexto é fundamental para compreender não apenas as possibilidades técnicas, mas também os desafios metodológicos, éticos e práticos envolvidos. A análise desta adoção pode contribuir para o desenvolvimento de frameworks e metodologias

que potencializem o uso responsável e efetivo destas tecnologias para a extração de valor dos dados governamentais.

Considerando os desafios identificados por Maratsi et al. [?] relacionados à qualidade dos metadados e seus impactos na interoperabilidade e descoberta dos dados, a adoção de LLMs pode representar uma abordagem complementar que amplie as possibilidades de análise e utilização das informações públicas, mesmo em cenários onde a estruturação e padronização dos dados apresentam limitações.

Em síntese, a fundamentação teórica sobre Dados Abertos Governamentais evidencia tanto o potencial transformador destas iniciativas quanto os desafios significativos para sua efetiva implementação e utilização. Neste contexto, a análise da adoção de LLMs representa uma contribuição relevante para o avanço do campo, explorando abordagens inovadoras que possam potencializar o valor e o impacto dos dados governamentais para a sociedade.

3 Large Language Models e IA Generativa

4 Large Language Models e IA Generativa

4.1 Fundamentos e Evolução dos Modelos de Linguagem de Grande Escala

Os Modelos de Linguagem de Grande Escala (Large Language Models - LLMs) representam um avanço significativo no campo da Inteligência Artificial (IA), particularmente na área de Processamento de Linguagem Natural (PLN). Estes modelos, baseados em arquiteturas de redes neurais profundas, são treinados com vastos conjuntos de dados textuais e têm demonstrado capacidades notáveis de compreensão e geração de texto em diversos contextos e aplicações [?]. A evolução destes modelos tem transformado significativamente a forma como interagimos com sistemas computacionais e como extraímos conhecimento de grandes volumes de dados não estruturados.

Os LLMs fazem parte do que se convencionou chamar de IA Generativa, um ramo da inteligência artificial capaz de criar conteúdo novo a partir do aprendizado obtido em dados de treinamento. Diferentemente dos sistemas tradicionais de IA, que são predominantemente analíticos ou classificatórios, os modelos generativos podem produzir textos, imagens, códigos e outros tipos de conteúdo com características que emulam a criatividade humana [?]. Esta capacidade generativa tem implicações profundas para diversos campos, incluindo a análise de dados governamentais abertos, foco desta pesquisa.

4.2 Características e Capacidades dos LLMs

Uma característica fundamental dos LLMs é sua capacidade de compreender contextos complexos e gerar respostas coerentes e contextualmente apropriadas. Esta habilidade deriva de sua arquitetura baseada em transformers e da técnica de atenção, que permite ao modelo estabelecer relações entre diferentes partes do texto [?]. Além disso, os LLMs demonstram capacidades emergentes, ou seja, habilidades que não foram explicitamente programadas, mas que surgem como resultado do treinamento em larga escala e da complexidade do modelo.

Lin et al. [?] destacam que os LLMs têm se mostrado instrumentais no avanço de tarefas de engenharia de software, demonstrando eficácia não apenas na geração de código, mas também na detecção de defeitos, no aprimoramento de medidas de segurança e na melhoria da qualidade geral do software. Esta versatilidade sugere um potencial significativo para aplicações em análise de dados governamentais, onde a interpretação de informações complexas e a geração de insights são cruciais.

4.3 Aplicações dos LLMs em Análise de Dados

A aplicação de LLMs na análise de dados tem ganhado destaque em diversos domínios. Jamil [?] propõe um modelo para um interpretador de LLM, denominado ProAb, que visa responder consultas

científicas em linguagem natural utilizando uma linguagem de consulta estruturada. Esta abordagem ilustra como os LLMs podem servir como interface entre usuários e bases de conhecimento estruturadas, facilitando o acesso e a interpretação de dados complexos.

No contexto específico de dados governamentais abertos, os LLMs oferecem potencial para transformar a forma como pesquisadores e cidadãos interagem com informações públicas. A capacidade destes modelos de processar e interpretar grandes volumes de texto não estruturado pode facilitar a extração de insights relevantes de documentos governamentais, relatórios e conjuntos de dados que, de outra forma, seriam difíceis de analisar manualmente [?].

Cabral et al. [?] investigam a aplicação de LLMs para Extração de Informação Aberta (EIA) em língua portuguesa, destacando que, embora a maioria dos métodos de EIA tenha sido desenvolvida para a língua inglesa, há um crescente interesse em métodos para o português. Este aspecto é particularmente relevante para a análise de dados governamentais no contexto brasileiro, onde a capacidade de processar informações em português é essencial.

4.4 LLMs na Análise de Documentos e Relatórios

Uma aplicação promissora dos LLMs é a análise de documentos e relatórios complexos. Bronzini et al. [?] exploram o uso de modelos de linguagem de grande escala para derivar insights estruturados de relatórios de sustentabilidade. Os autores destacam que informações sobre práticas de sustentabilidade são frequentemente divulgadas em documentação diversa, não estruturada e multimodal, o que representa um desafio para a coleta e alinhamento eficientes dos dados em um framework unificado.

De forma similar, Kazakov et al. [?] apresentam o "ESGify", um modelo de Processamento de Linguagem Natural baseado na arquitetura MPNetbase, destinado a classificar textos no âmbito de riscos ambientais, sociais e de governança corporativa (ESG). Esta aplicação demonstra como os LLMs podem ser adaptados para domínios específicos, oferecendo ferramentas eficazes para a análise de informações complexas e multifacetadas.

Estas aplicações têm paralelos diretos com a análise de dados governamentais abertos, que frequentemente envolve a interpretação de documentos complexos, relatórios e conjuntos de dados heterogêneos. A capacidade dos LLMs de processar e extrair informações relevantes destes materiais pode significativamente aumentar a acessibilidade e utilidade dos dados governamentais para pesquisadores e cidadãos.

4.5 Desafios Éticos e Limitações

Apesar de seu potencial transformador, a adoção de LLMs na análise de dados governamentais apresenta desafios significativos, particularmente no que diz respeito a questões éticas. Germani et al. [?] discutem a natureza dual da IA na disseminação de informações, destacando que, embora os sistemas de IA tenham o potencial de criar campanhas de informação valiosas com repercussões positivas para a saúde pública e a democracia, existem preocupações quanto ao potencial uso destes sistemas para gerar desinformação convincente.

As consequências desta natureza dual da IA, capaz tanto de iluminar quanto de obscurecer o panorama informacional, são complexas e multifacetadas. Os autores argumentam que a rápida integração da IA na sociedade exige uma compreensão abrangente de suas implicações éticas e o desenvolvimento de estratégias para aproveitar seu potencial para o bem maior [?].

No contexto da análise de dados governamentais, estas considerações éticas são particularmente relevantes, dado o potencial impacto das interpretações e conclusões derivadas destes dados nas políticas públicas e na percepção cidadã. A transparência, a responsabilidade e a mitigação de vieses são aspectos cruciais a serem considerados na adoção de LLMs para este fim.

4.6 Perspectivas Futuras e Implicações para a Pesquisa

O campo dos LLMs e da IA Generativa está em rápida evolução, com novos modelos e aplicações emergindo continuamente. Zhao et al. [?] apresentam uma análise prospectiva das lojas de aplicativos de LLM, focando em aspectos-chave como mineração de dados, identificação de riscos de segurança, assistência ao desenvolvimento e dinâmicas de mercado. Esta visão abrangente estende-se às relações intrincadas entre vários stakeholders e aos avanços tecnológicos, oferecendo insights valiosos para pesquisadores e desenvolvedores.

Para a pesquisa sobre a adoção de LLMs na análise de dados abertos governamentais, estas perspectivas futuras sugerem um campo fértil de investigação, com potencial para desenvolver metodologias inovadoras que aproveitem as capacidades dos LLMs enquanto abordam seus desafios e limitações. A caracterização das percepções e experiências dos pesquisadores neste contexto, objetivo central desta pesquisa, pode fornecer insights valiosos para orientar o desenvolvimento e a implementação de soluções baseadas em LLMs que sejam eficazes, éticas e alinhadas com as necessidades dos usuários.

Em suma, os LLMs e a IA Generativa representam uma fronteira promissora para a análise de dados governamentais abertos, oferecendo capacidades sem precedentes para extrair insights de informações complexas e não estruturadas. No entanto, sua adoção efetiva requer uma compreensão profunda de suas capacidades, limitações e implicações éticas, bem como o desenvolvimento de abordagens que maximizem seu potencial enquanto mitigam seus riscos.

5 Aplicações de LLMs no Setor Público

6 Aplicações de Modelos de Linguagem de Grande Escala no Setor Público

A emergência dos Modelos de Linguagem de Grande Escala (Large Language Models - LLMs) representa um avanço significativo no campo da inteligência artificial, com potencial transformador para diversos setores, incluindo a administração pública. Esta seção explora as aplicações desses modelos no contexto governamental, com ênfase particular na análise de dados abertos governamentais, alinhando-se ao objetivo central desta pesquisa.

6.1 LLMs como Ferramentas de Auditoria Cidadã

O conceito de "auditores de poltrona" (armchair auditors) refere-se a cidadãos que utilizam dados abertos para investigar e monitorar atividades governamentais, empregando análises e outras abordagens para promover a transparência e a responsabilização [?]. Esses atores desempenham um papel fundamental no ecossistema de governança aberta, contribuindo para a fiscalização das ações governamentais e para o fortalecimento da democracia participativa.

O'Leary Jr. [?] destaca que, historicamente, os auditores cidadãos enfrentavam diversos desafios em suas análises, incluindo dificuldades no acesso, interpretação e processamento de grandes volumes de dados governamentais. No entanto, o surgimento dos LLMs apresenta um potencial significativo para mitigar essas limitações, fornecendo suporte substancial tanto no processamento de dados quanto nos processos analíticos necessários para a auditoria cidadã eficaz.

Os LLMs podem auxiliar os auditores cidadãos de diversas formas, incluindo a interpretação de documentos complexos, a identificação de padrões em grandes conjuntos de dados e a geração de insights a partir de informações dispersas. Essa capacidade de processamento e análise pode democratizar o acesso à fiscalização governamental, permitindo que cidadãos sem expertise técnica específica possam participar ativamente do processo de accountability pública.

6.2 Aprimorando a Descoberta de Dados Governamentais Abertos

Um dos desafios críticos no ecossistema de dados abertos governamentais (Open Government Data - OGD) é a descoberta eficiente de informações relevantes. Klimask e Nikiforova [?] observam que, à medida que mais conjuntos de dados são publicados em portais de OGD, torna-se progressivamente mais difícil encontrar informações específicas, levando à sobrecarga informacional e ao fenômeno conhecido como "dados obscuros" (dark data) - informações valiosas que permanecem inacessíveis devido à dificuldade de localização.

Em sua análise do Portal de Dados Abertos da Estônia, os autores identificaram deficiências significativas nas práticas de etiquetagem de dados: entre 1.787 conjuntos de dados publicados, 11% não possuíam etiquetas associadas, enquanto 26% tinham apenas uma etiqueta atribuída [?]. Essas descobertas evidenciam os desafios na localização e acessibilidade dos dados dentro dos portais governamentais.

Para enfrentar essa problemática, Klimask e Nikiforova [?] propõem o TAGIFY, uma interface de etiquetagem baseada em LLMs projetada para melhorar a descoberta de dados em portais de OGD. Esta solução utiliza o poder dos modelos de linguagem avançados, como GPT-4 e GPT-3.5, para automatizar e aprimorar o processo de etiquetagem, facilitando a documentação completa e precisa dos conjuntos de dados.

A implementação de sistemas como o TAGIFY representa um avanço significativo na utilização de LLMs para aprimorar a infraestrutura de dados abertos governamentais. Ao melhorar a qualidade das etiquetas associadas aos conjuntos de dados, essas soluções aumentam a descoberta e a acessibilidade das informações governamentais, potencializando seu valor para pesquisadores, jornalistas, formuladores de políticas e cidadãos interessados.

6.3 Combate à Desinformação no Setor Público

Outro domínio promissor para a aplicação de LLMs no setor público é o combate à desinformação. Donner et al. [?] argumentam que a detecção rápida de desinformação generalizada constitui um dos problemas mais prementes de política pública, exigindo pesquisa transdisciplinar no campo da ciência de dados.

Os autores propõem o TRUExT (Trustworthiness Regressor Unified Explainable Tool), uma ferramenta que utiliza LLMs e processamento de linguagem natural para classificar informações quanto à sua confiabilidade [?]. Esta abordagem reconhece que, embora a ciência de dados ofereça potencial para resolver o problema da desinformação em grande escala de forma automatizada, ela requer insights dos campos da comunicação e do jornalismo para definir características quantificáveis que possam auxiliar em previsões mais precisas.

No contexto governamental, ferramentas como o TRUExT podem ser particularmente valiosas para agências públicas que precisam monitorar e responder rapidamente à desinformação relacionada a políticas, serviços ou crises. Ao integrar LLMs em seus processos de análise de informação, as instituições públicas podem desenvolver capacidades mais robustas para identificar, classificar e responder a conteúdos enganosos, fortalecendo a integridade do discurso público.

6.4 Desafios e Considerações Éticas

Apesar do potencial transformador dos LLMs no setor público, sua implementação não está isenta de desafios e considerações éticas. Questões relacionadas à privacidade, vieses algorítmico, transparência e responsabilidade devem ser cuidadosamente abordadas para garantir que essas tecnologias sejam utilizadas de maneira ética e responsável.

A aplicação de LLMs na análise de dados governamentais exige um equilíbrio delicado entre inovação tecnológica e salvaguardas éticas. É fundamental que os desenvolvedores, pesquisadores e formuladores de políticas trabalhem colaborativamente para estabelecer diretrizes e padrões que orientem o uso responsável dessas tecnologias no contexto da administração pública.

6.5 Perspectivas para a Pesquisa

As aplicações de LLMs no setor público, particularmente na análise de dados abertos governamentais, representam um campo de pesquisa emergente e promissor. Conforme evidenciado pelos estudos discutidos nesta seção, essas tecnologias oferecem potencial significativo para transformar a maneira como os dados governamentais são acessados, analisados e utilizados.

No contexto específico desta pesquisa, que visa analisar a adoção dos LLMs na análise de dados abertos governamentais, compreender as perspectivas dos pesquisadores sobre essas aplicações é fundamental. As percepções, expectativas e preocupações da comunidade acadêmica em relação ao uso de LLMs nesse domínio podem fornecer insights valiosos para orientar o desenvolvimento e a implementação dessas tecnologias de maneira alinhada às necessidades e valores da sociedade.

7 Metodologias e Frameworks para Análise de Dados

8 Metodologias e Frameworks para Análise de Dados

A análise de dados tem se tornado um componente fundamental para a tomada de decisões baseadas em evidências, especialmente no contexto de dados governamentais abertos. Com o advento de tecnologias emergentes como os Modelos de Linguagem de Grande Escala (LLMs), novas metodologias e frameworks têm sido desenvolvidos para lidar com os desafios específicos relacionados à análise, compartilhamento e utilização desses dados. Esta seção apresenta uma revisão das principais abordagens metodológicas e frameworks que fundamentam a análise de dados no contexto da pesquisa sobre a adoção de LLMs para análise de dados abertos governamentais.

8.1 Ecossistemas de Dados Públicos e sua Evolução

Os ecossistemas de dados públicos têm evoluído significativamente nas últimas décadas, impulsionados por avanços tecnológicos e mudanças nas demandas sociais. Nikiforova et al. [?] discutem essa evolução, destacando como os ecossistemas de dados públicos estão se transformando em resposta às tecnologias emergentes, incluindo a Inteligência Artificial (IA) e, mais especificamente, os LLMs. Os autores argumentam que o setor público enfrenta desafios multifacetados de natureza social, regulatória e técnica devido ao crescente volume de dados disponíveis e à necessidade de otimizar sua utilização para maior eficiência e eficácia.

A transição para o que os autores chamam de "ecossistema inteligente de dados públicos" representa uma mudança paradigmática na forma como os dados governamentais são coletados, processados e disponibilizados. Este novo paradigma incorpora tecnologias emergentes como IA generativa e LLMs para transformar dados brutos em insights acionáveis, permitindo uma tomada de decisão mais informada e baseada em evidências. Esta evolução é particularmente relevante para nossa pesquisa, uma vez que fornece o contexto mais amplo no qual a adoção de LLMs para análise de dados governamentais abertos ocorre.

8.2 Análise de Texto para Co-criação no Setor Público

A co-criação de serviços públicos tem emergido como uma alternativa aos modelos tradicionais de prestação de serviços, permitindo maior participação cidadã e melhor adequação às necessidades da população. Neste contexto, as técnicas de análise de texto (Text Analytics - TA) desempenham um papel crucial ao permitir a extração de insights valiosos a partir de grandes volumes de dados textuais não estruturados [?].

A revisão sistemática da literatura apresentada em [?] identifica como as técnicas de análise de texto podem apoiar a co-criação de serviços públicos, destacando seu potencial para fomentar soluções baseadas em dados para o setor público. Esta abordagem é particularmente relevante quando consideramos o uso de LLMs, que são fundamentalmente ferramentas avançadas de processamento e geração de texto, capazes de analisar e sintetizar grandes volumes de informação textual.

A aplicação de técnicas de análise de texto, potencializadas por LLMs, pode transformar significativamente a forma como as organizações do setor público interagem com os cidadãos e utilizam seus feedbacks para melhorar serviços. Isso se alinha diretamente com o objetivo de nossa pesquisa de caracterizar a adoção de LLMs na análise de dados abertos governamentais, especialmente considerando a perspectiva dos pesquisadores que trabalham neste domínio.

8.3 Frameworks para Avaliação de Riscos e Conformidade Legal

Um aspecto crítico da análise de dados, especialmente quando envolve dados pessoais ou sensíveis, é a gestão de riscos e a conformidade com regulamentações de proteção de dados. Baloukas et al. [?] propõem um framework para avaliação de riscos e conformidade legal que apoia o compartilhamento de dados pessoais com preservação de privacidade para pesquisa científica.

Os autores destacam que, embora o acesso a grandes volumes de dados seja essencial para pesquisas de ponta, como o treinamento de modelos de IA (incluindo LLMs), os provedores de dados frequentemente relutam em compartilhar seus dados devido a preocupações com a privacidade e os riscos associados. Além disso, a necessidade de conformidade com regulamentações como o GDPR (Regulamento Geral de Proteção de Dados) cria uma sobrecarga adicional que dificulta o compartilhamento de dados.

O framework proposto por Baloukas et al. [?] oferece uma abordagem estruturada para avaliar e mitigar riscos associados ao compartilhamento de dados, garantindo simultaneamente a conformidade legal. Esta abordagem é particularmente relevante no contexto da adoção de LLMs para análise de dados governamentais abertos, onde questões de privacidade, ética e conformidade legal são preocupações primordiais.

8.4 Integração de Metodologias para Análise de Dados Governamentais com LLMs

A integração efetiva de LLMs nas metodologias existentes para análise de dados governamentais requer uma abordagem holística que considere tanto os aspectos técnicos quanto os sociais, éticos e legais. Com base nas contribuições discutidas anteriormente, podemos identificar alguns elementos-chave para esta integração:

1. **Evolução do ecossistema de dados:** Conforme destacado por Nikiforova et al. [?], é essencial compreender como os ecossistemas de dados públicos estão evoluindo em resposta às tecnologias emergentes, incluindo os LLMs, e como essa evolução afeta as metodologias de análise de dados.

2. **Técnicas de análise de texto avançadas:** As técnicas de análise de texto discutidas em [?] podem ser significativamente aprimoradas com a incorporação de LLMs, permitindo análises mais sofisticadas e insights mais profundos a partir de dados textuais governamentais.

3. **Gestão de riscos e conformidade:** O framework proposto por Baloukas et al. [?] oferece uma base sólida para abordar questões de privacidade e conformidade legal, que são particularmente importantes quando se utilizam LLMs para analisar dados governamentais que podem conter informações sensíveis.

A integração desses elementos em um framework metodológico coerente pode fornecer uma base sólida para pesquisadores e profissionais que buscam adotar LLMs para análise de dados abertos governamentais, permitindo maximizar os benefícios dessas tecnologias enquanto mitigam os riscos associados.

8.5 Considerações Finais

As metodologias e frameworks discutidos nesta seção fornecem uma base teórica sólida para compreender como os LLMs podem ser efetivamente adotados para análise de dados abertos governamentais. A evolução dos ecossistemas de dados públicos, as técnicas avançadas de análise de texto e os frameworks para gestão de riscos e conformidade legal são elementos complementares que, quando integrados, podem potencializar o uso de LLMs neste contexto.

Esta fundamentação teórica é essencial para nossa pesquisa sobre a adoção de LLMs na análise de dados abertos governamentais, fornecendo o arcabouço conceitual necessário para caracterizar as perspectivas dos pesquisadores que trabalham neste domínio e identificar os fatores que influenciam a adoção dessas tecnologias emergentes.

9 Transformação Digital no Setor Público

10 Transformação Digital no Setor Público

A transformação digital no setor público representa uma mudança paradigmática na forma como os governos operam, prestam serviços e interagem com os cidadãos. Este fenômeno transcende a mera informatização de processos existentes, constituindo uma reconfiguração fundamental das estruturas, processos e culturas organizacionais governamentais por meio da incorporação estratégica de tecnologias digitais [?]. No contexto contemporâneo, caracterizado pela crescente disponibilidade de dados e pelo avanço de tecnologias como os Grandes Modelos de Linguagem (LLMs), a transformação digital no setor público adquire novas dimensões e possibilidades.

10.1 Fundamentos da Transformação Digital Governamental

A transformação digital no setor público pode ser compreendida como um processo multidimensional que envolve não apenas a adoção de novas tecnologias, mas também mudanças organizacionais, culturais e estratégicas. Kopanaki e Stamoulis [?] argumentam que, diferentemente do setor privado, onde sistemas integrados como ERPs são comuns, o setor público frequentemente opera com bases de dados isoladas e ferramentas de software dispersas, o que limita significativamente a capacidade de formulação de políticas baseadas em evidências.

Esta fragmentação representa um desafio fundamental para a efetiva transformação digital governamental, uma vez que "recursos de informação, não importa quão extensos sejam, não constituem sistemas de informação de suporte à formulação de políticas"[?]. A integração desses recursos em sistemas coerentes e interoperáveis emerge, portanto, como um imperativo para governos que buscam modernizar suas operações e aprimorar a qualidade de suas decisões.

10.2 Ciência de Dados como Catalisador da Transformação Digital

No cenário atual, a ciência de dados surge como um componente crucial da transformação digital no setor público. Adebamowo et al. [?] destacam que a pesquisa em saúde baseada em ciência de dados "promete benefícios tremendos", embora sua implementação enfrente "riscos substanciais de governança ética que poderiam frustrar a entrega desses benefícios antecipados". Esta observação, embora feita no contexto específico da pesquisa em saúde na África, ilustra um princípio mais amplo aplicável à transformação digital governamental: o potencial transformador da ciência de dados é acompanhado por desafios significativos de governança e ética.

A incorporação da ciência de dados nas operações governamentais possibilita a transição de um modelo de tomada de decisão baseado em intuição ou precedentes para um paradigma fundamentado em evidências empíricas. Kopanaki e Stamoulis [?] enfatizam que sistemas de informação no setor público são "absolutamente necessários para a formulação de políticas baseada em evidências e orientada por dados". Esta abordagem data-driven tem o potencial de aumentar a eficácia e a eficiência das políticas públicas, ao mesmo tempo em que promove maior transparência e responsabilização.

10.3 Desafios e Oportunidades na Era dos Dados Abertos e LLMs

A convergência entre a transformação digital governamental, dados abertos e tecnologias emergentes como os LLMs apresenta tanto desafios quanto oportunidades significativas. Por um lado, a disponibilização de dados governamentais em formatos abertos e acessíveis amplia as possibilidades de análise, colaboração

e inovação. Por outro lado, a extração de valor desses dados requer capacidades analíticas avançadas, frequentemente escassas no setor público.

Neste contexto, os LLMs emergem como ferramentas potencialmente transformadoras para a análise de dados abertos governamentais. Sua capacidade de processar e interpretar grandes volumes de texto não estruturado pode facilitar a descoberta de padrões, tendências e insights que permaneceriam ocultos em abordagens analíticas tradicionais. No entanto, como sugerem Adebamowo et al. [?] em relação à ciência de dados em geral, a implementação dessas tecnologias deve ser acompanhada por "esforços emergentes para construir estruturas de governança ética" adequadas.

A adoção de LLMs para análise de dados abertos governamentais se alinha com a visão de Kopanaki e Stamoulis [?] sobre a necessidade de sistemas de informação que suportem a formulação de políticas baseada em evidências. Estas tecnologias podem potencialmente preencher a lacuna entre a abundância de dados disponíveis e a capacidade limitada de extrair insights acionáveis desses dados, contribuindo para uma transformação digital mais profunda e impactante no setor público.

10.4 Implicações para a Pesquisa e Prática

A intersecção entre transformação digital governamental, dados abertos e LLMs constitui um campo fértil para pesquisa e inovação. Compreender como os pesquisadores percebem e utilizam os LLMs na análise de dados abertos governamentais, objetivo central da presente dissertação, pode fornecer insights valiosos sobre os desafios, oportunidades e implicações desta convergência tecnológica.

Adebamowo et al. [?] destacam a importância de "investimentos por governos e instituições africanas, organizações internacionais de financiamento e colaborações para pesquisa e desenvolvimento de capacidades" no contexto da ciência de dados para pesquisa em saúde. De forma análoga, o avanço da transformação digital no setor público, particularmente no que diz respeito à adoção de LLMs para análise de dados abertos, requer investimentos coordenados em infraestrutura, capacitação e estruturas de governança.

A caracterização das percepções e práticas dos pesquisadores neste domínio, conforme proposto nesta dissertação, pode contribuir significativamente para informar tais investimentos e orientar o desenvolvimento de políticas e estratégias que maximizem os benefícios da transformação digital governamental enquanto mitigam seus riscos potenciais.

10.5 Considerações Finais

A transformação digital no setor público representa um processo complexo e multifacetado que transcende a mera adoção de tecnologias. Como argumentam Kopanaki e Stamoulis [?], sistemas de informação adequados são "absolutamente necessários" para uma formulação de políticas efetiva e baseada em evidências. No contexto atual, caracterizado pela crescente disponibilidade de dados abertos governamentais e pelo surgimento de tecnologias analíticas avançadas como os LLMs, esta necessidade torna-se ainda mais premente.

Compreender como os pesquisadores estão adotando e utilizando LLMs para análise de dados abertos governamentais constitui, portanto, um passo importante para avançar nossa compreensão sobre a transformação digital no setor público e suas implicações para a sociedade contemporânea. Esta investigação pode não apenas contribuir para o avanço do conhecimento acadêmico neste campo, mas também informar práticas e políticas que promovam uma transformação digital governamental mais efetiva, inclusiva e ética.

11 Síntese da Fundamentação Teórica

A análise da literatura existente sobre a adoção de LLMs na análise de dados abertos governamentais revela um campo de pesquisa emergente e promissor. Os 141 artigos selecionados demonstram diferentes perspectivas e abordagens para a integração dessas tecnologias no contexto do setor público.

Esta fundamentação teórica estabelece as bases conceituais necessárias para compreender os desafios e oportunidades associados à adoção de llms na análise de dados abertos governamentais, fornecendo o embasamento científico para o desenvolvimento desta pesquisa.

->