

Utilizando o ChatGPT como ferramenta de suporte para a construção de uma base de dados textual sobre violências: apresentação e análise do processo de desenvolvimento

Using ChatGPT as a support tool for building a textual database on violence: presentation and analysis of the development

Guilherme Ataíde Dias¹



<http://lattes.cnpq.br/9553707435669429>



<https://orcid.org/0000-0001-6576-0017>

Tassyara Onofre de Oliveira²



<http://lattes.cnpq.br/5328618364332415>



<https://orcid.org/0000-0003-4048-8322>

Arthur Antônio Martins Sales Campelo³



<http://lattes.cnpq.br/2370276793833303>



<https://orcid.org/0000-0001-5907-917X>

Vivian Roberta Lima de Carvalho⁴



<http://lattes.cnpq.br/4602577185970003>



<https://orcid.org/0009-0003-6645-7660>

Resumo

Este estudo descreve a construção de uma base de dados textual de referência sobre violências, com o objetivo de aprimorar a recuperação de informações para pesquisadores da Rede Interdisciplinar de Estudos sobre Violências. A pesquisa, de caráter aplicado e abordagem qualitativa, integrou princípios da Ciência da Informação e tecnologias emergentes, utilizando o ChatGPT como ferramenta de suporte na criação de uma página web, tesouro temático, gerador de itens e imagens ilustrativas. O repositório foi

¹ Doutor em Comunicação (Ciência da Informação), Universidade de São Paulo (USP), São Paulo, Brasil. Professor Titular, Universidade Federal da Paraíba, João Pessoa, Brasil. guilhermeataide@ccsa.ufpb.br.

² Doutora em Ciência da Informação, Universidade Federal da Paraíba (UFPB), João Pessoa, Paraíba, Brasil. Pós-doutoranda júnior bolsista do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Programa de Pós-Graduação em Ciência da Informação da Universidade Federal da Paraíba (UFPB), João Pessoa, Paraíba, Brasil. Estância de pesquisa na Universitat de València (UV), València, Comunidade Valenciana, Espanha. tassyaraonofre@gmail.com.

³ Graduando em Biblioteconomia, Universidade Federal da Paraíba (UFPB), João Pessoa, Brasil. Bolsista de Iniciação Científica do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), João Pessoa, Paraíba, Brasil. arthur.campelo@academico.ufpb.br.

⁴ Graduanda em Biblioteconomia, Universidade Federal da Paraíba (UFPB), João Pessoa, Paraíba, Brasil. Bolsista de Iniciação Científica do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), João Pessoa, Paraíba, Brasil. viivacarvalho@gmail.com.

implementado no DSpace 7, organizado em coleções temáticas aderentes aos princípios FAIR, resultando em maior eficiência na recuperação de dados. Os resultados indicam que a inteligência artificial reduziu significativamente o tempo de desenvolvimento, embora ajustes humanos sejam essenciais para garantir qualidade. Conclui-se que o uso do ChatGPT viabilizou avanços no projeto, mas a automação de coleta e ingestão de metadados e a capacitação em DSpace permanecem como desafios a serem superados.

Palavras-chave: Inteligência Artificial; ChatGPT; DSpace; base de dados; violência.

Abstract

This study describes the development of a reference textual database on violence aimed at enhancing information retrieval for researchers from the Interdisciplinary Network for Violence Studies. Employing an applied research approach with a qualitative focus, it integrated Information Science principles and emerging technologies, using ChatGPT as a support tool to create a web page, thematic thesaurus, item generator, and illustrative images. The repository, implemented in DSpace 7 and organized into FAIR-compliant thematic collections, improved data retrieval efficiency. Results demonstrate that artificial intelligence significantly reduced development time, though human adjustments remain necessary to ensure quality. The findings suggest that ChatGPT facilitated project advancements, yet challenges such as automating metadata collection and ingestion, along with expertise in DSpace customization, persist. This work highlights the potential of artificial intelligence in database development while underscoring the need for ongoing refinement.

Keywords: Artificial Intelligence; ChatGPT; DSpace; database; violence.

1 INTRODUÇÃO

O desenvolvimento da inteligência artificial ocorreu de forma paulatina ao longo das últimas décadas, tendo como ponto de partida o período iniciado na Segunda Guerra Mundial. Esse avanço alternou períodos de intensa atividade de pesquisa com fases de menor produção científica, influenciados tanto pelos resultados obtidos pelos pesquisadores quanto pelos diversos interesses das indústrias e dos governos financiadores.

De acordo com Russel e Norvig (2004), o primeiro trabalho atualmente reconhecido como IA foi desenvolvido por Warren McCulloch e Walter Pitts em 1943. Nesse estudo, os autores propuseram um modelo matemático de neurônios artificiais baseado na lógica booleana. Esse trabalho foi fundamental para a criação das redes neurais artificiais e serviu como base para o desenvolvimento da IA, especialmente no campo do aprendizado de máquina. Outro pioneiro da área foi o cientista britânico Alan Turing, que, em 1950 publicou um artigo seminal intitulado de “*Computing Machinery and Intelligence*” no qual explorou a possibilidade de máquinas exibirem comportamento inteligente (Russel; Norvig, 2004). O verdadeiro marco da IA como disciplina científica formal ocorreu em 1956, na Conferência de *Dartmouth*, organizada por John McCarthy,

Marvin Minsky, Nathaniel Rochester e Claude Shannon. Nesse evento, o termo “Inteligência Artificial” foi cunhado (Russel; Norvig, 2004).

Atualmente, constata-se um grande avanço no desenvolvimento e uso da IA, impulsionado principalmente pelas redes neurais profundas e pela ampla disponibilidade de dados. A crescente proliferação de modelos de linguagem de grande porte (LLMs), como *ChatGPT* e outras ferramentas no mercado, tem alavancado tanto a pesquisa acadêmica quanto as iniciativas comerciais. Em paralelo, melhorias na infraestrutura computacional e na capacidade de processamento viabilizam treinamentos cada vez mais robustos e especializados. Entretanto, a disseminação veloz dessas tecnologias evidencia a necessidade de regulamentações adequadas, bem como de reflexões éticas e práticas de governança que assegurem sua utilização responsável.

A investigação apresentada neste estudo descreve o processo de desenvolvimento de uma base de dados textual de referência sobre a temática das “violências”, com o objetivo de apoiar pesquisadores que atuam nesse campo, especialmente aqueles vinculados à *Rede Interdisciplinar de Estudos sobre Violências* (Universidade Federal da Paraíba, 2025).

A motivação para o desenvolvimento da base de dados textual de referência sobre a temática das “violências” parte do pressuposto de que a pesquisa realizada em uma base de dados temática, cujos documentos passaram previamente por um processo criterioso de seleção dentro do domínio de interesse, tende a apresentar maior qualidade nos resultados recuperados em comparação à pesquisa realizada em bases que abrangem temáticas diversas.

Por se tratar de uma base de dados textual de referência, optamos por não armazenar os conteúdos originais na sua íntegra - sejam eles artigos, dissertações, teses, notícias ou outros tipos de documento. Essa decisão foi motivada principalmente pela necessidade de evitar conflitos relacionados aos direitos autorais. Dessa forma, a base de dados reúne apenas os elementos descritores e as referências das obras, materializados em dois arquivos distintos: um no formato texto simples e outro estruturado na linguagem XML.

Um ponto fundamental considerado no desenvolvimento da base de dados foi torná-la o mais aderente possível aos princípios FAIR (*Findable, Accessible, Interoperable e Reusable*), com o objetivo de ampliar o uso, o reuso e o compartilhamento dos dados disponibilizados (Wilkinson *et al.*, 2016). A adoção desse padrão garante maior visibilidade e facilidade na recuperação dos recursos, além de contribuir para o fortalecimento da colaboração científica. Ao seguir essas diretrizes, a base de dados proporciona aos pesquisadores condições mais adequadas para o acesso, a interpretação e a reutilização das informações, potencializando, assim, o impacto das pesquisas relacionadas à temática das “violências”.

2 PROCEDIMENTOS METODOLÓGICOS

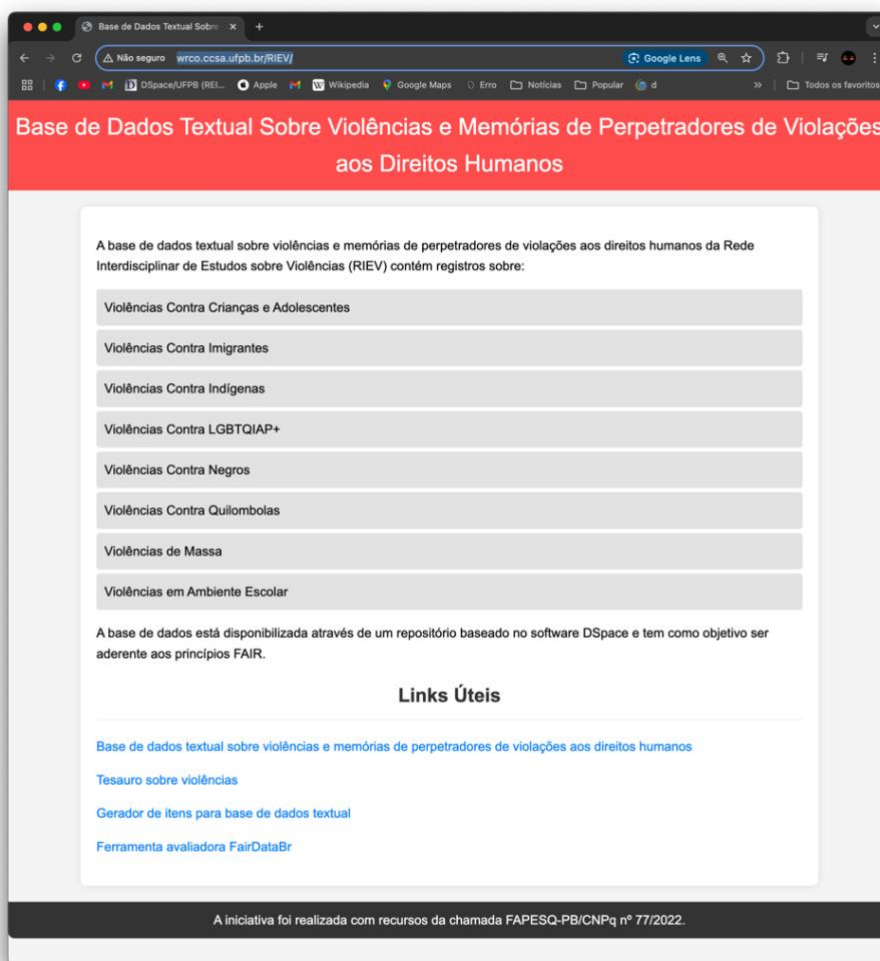
A metodologia adotada para o desenvolvimento da base de dados textual de referência sobre violências caracteriza-se como pesquisa aplicada, pois tem como objetivo a criação de um sistema para apoiar pesquisadores na recuperação e organização de informações sobre o tema. A abordagem utilizada é predominantemente qualitativa, centrada na estruturação e depósito dos dados. O trabalho integra tecnologias emergentes, automação de processos e princípios da Ciência da Informação, proporcionando maior eficiência na gestão e disseminação do conhecimento sobre violências.

O *ChatGPT* foi utilizado como uma ferramenta de suporte para a criação de instrumentos necessários à operacionalização da base de dados. A seguir, indicamos quais foram: 1) Página na *Web* para acesso a base de dados e demais ferramentas de suporte; 2) Tesouro sobre violências; 3) Gerador de itens para a base de dados e 4) Imagens ilustrativas das coleções e itens da base de dados. Utilizamos a versão com assinatura do *ChatGPT*, que emprega o modelo *GPT-4*.

A página *Web* para acesso à base de dados e demais ferramentas de suporte foi criada com auxílio do *ChatGPT*, por meio de um *prompt* contendo instruções textuais detalhadas sobre como gostaríamos que ela fosse estruturada e apresentada. O resultado foi obtido na forma de um arquivo contendo código HTML, posteriormente testado localmente em um navegador. Durante o desenvolvimento foram realizadas diversas

interações com o *ChatGPT* até que chegássemos a uma versão plenamente às nossas expectativas iniciais. Essa abordagem mostrou-se bastante eficiente, uma vez que possibilitou ajustes rápidos e pontuais, garantindo, assim, maior controle sobre o resultado final. A Figura 1 ilustra o código gerado, exibido em um navegador *Web* sob a forma de página, destacando a clareza e a funcionalidade obtidas por meio dessa abordagem interativa.

Figura 1 – Página de acesso aos instrumentos necessários à operacionalização da base de dados



Fonte: Elaboração própria.

Para realizar o depósito dos registros sobre violência na base de dados, verificamos a necessidade de desenvolver um vocabulário controlado, especificamente um tesouro, para garantir um controle terminológico eficiente das palavras-chave associadas a cada item inserido. Essa estratégia contribui bastante para aumentar a

precisão na recuperação dos dados, reduzindo ambiguidades e facilitando o acesso pelos usuários.

O uso do tesauro permite estabelecer uma padronização consistente na descrição dos conteúdos, ampliando as possibilidades de integração com outras bases de dados e favorecendo futuras ações de compartilhamento e colaboração científica.

A extração de termos a partir de um tesauro é uma etapa fundamental para a indexação de documentos na base de dados. Com a organização criteriosa dos termos, consolidou-se um vocabulário controlado que facilita a recuperação da informação. Esse processo assegurou uma indexação inicial padronizada e reduziu significativamente o esforço humano na categorização dos conteúdos, otimizando o tempo de ingestão de dados na base.

Os termos iniciais para a criação do vocabulário controlado foram extraídos a partir de títulos e resumos associados de 105 documentos selecionados previamente com temas relacionados a diversos tipos de “violências”. Os títulos e resumos foram submetidos ao *ChatGPT* por meio de um arquivo de texto. Solicitamos através do *prompt* a geração de um arquivo em formato DOCX para *download* que reunisse em uma lista os termos mais representativos de “violências” contidos nos respectivos títulos e resumos submetidos. A lista de termos representativos das “violências” foi utilizada como um base norteadora inicial para a construção de um tesauro por pesquisadores na área da Ciência de Informação. Para a construção do tesauro foi realizada com auxílio da ferramenta *Thesa*⁵ (Tesauro Semântico Aplicado) (Gabriel Junior; Laipelt, 2017), desenvolvida por pesquisadores da Universidade Federal do Rio Grande do Sul.

O desenvolvimento de um formulário para a geração dos elementos descritivos dos recursos informacionais sobre “violências” (itens da base de dados) mostrou-se necessário, visto que o que é efetivamente depositado no *DSpace* não é o recurso informacional fonte propriamente dito (artigo, dissertação, teses, notícia ou outro tipo de documento), mas sim uma descrição desse recurso. Essa característica configura a base como uma de dados textual de referência.

⁵ Disponível em: <https://www.ufrgs.br/tesauros/index.php/thesa>. Acesso em: 19 maio 2025.

Assim como na construção da página para acesso à base de dados, nesse caso também utilizamos o *ChatGPT* em diversas interações até alcançar uma configuração adequada do formulário de entrada de dados. Cabe mencionar que alguns ajustes no código HTML do formulário precisaram ser realizados manualmente, até obtermos uma versão que atendesse plenamente aos nossos propósitos. O uso do formulário possibilitou gerar os itens a serem depositados na base de dados com maior agilidade e melhor controle de qualidade.

Após a captura dos dados pelo formulário, estes são exportados para dois arquivos antes de serem depositados no *DSpace*: um arquivo em formato de texto contendo os principais dados do recurso e um segundo, também em formato texto, porém estruturado na Linguagem XML. A Figura 2 exemplifica o conteúdo de um dos arquivos gerados com os descritores do recurso informacional, no caso a versão em XML.

Figura 2 – Item da base de dados (versão XML)

```
<?xml version="1.0" encoding="UTF-8" ?>
<rdf:RDF xmlns:rdf="http://www.w3.org/1999/02/22-rdf-syntax-ns#" xmlns:dc="http://purl.org/dc/elements/1.1/" ?>
  <dc:Description rdf:about="http://example.org/resource/1/" ?>
    <dc:title>Desconstruindo os arquivos da violência escolar: reflexões sobre a relação entre a docência e a Filosofia da Educação</dc:title>
    <dc:type>Artigo</dc:type>
    <dc:creators>Amarildo Luiz Trevisan</dc:creators>
    <dc:keywords>Violência simbólica</dc:keywords>
    <dc:keywords>Preconceito</dc:keywords>
    <dc:keywords>Exclusão social</dc:keywords>
    <dc:keywords>Violência psicológica</dc:keywords>
    <dc:coverage>Ponta Grossa</dc:coverage>
    <dc:date>20/07/2023</dc:date>
    <dc:format>application/pdf</dc:format>
    <dc:language>pt</dc:language>
    <dc:relation>n/a</dc:relation>
    <dc:source> Práxis Educativa</dc:source>
    <dc:doi>https://doi.org/10.5212/PraxEduc.v.18.22099.068</dc:doi>
    <dc:issn>1809-4031</dc:issn>
    <dc:orcid> https://orcid.org/0000-0002-3575-4369</dc:orcid>
    <dc:link>https://revistas.uepg.br/index.php/praxiseducativa/article/view/22099</dc:link>
    <dc:reference>TREVISAN, A. L. Desconstruindo os arquivos da violência escolar: reflexões sobre a relação entre a docência e a Filosofia da Educação. Práxis Educativa, [S. l.], v. 18, p. 1-19, 2023. DOI: 10.5212/PraxEduc.v.18.22099.068. Disponível em: https://revistas.uepg.br/index.php/praxiseducativa/article/view/22099.</dc:reference>
    <dc:description>Este texto apresenta uma reflexão crítica sobre os arquivos da violência escolar como um problema filosófico e psicanalítico. Utilizando a expressão "mal de arquivo", de Derrida, como ponto de partida, discutem-se os movimentos da docência diante do problema da violência escolar. A partir de uma experiência de pesquisa, são apresentadas as opiniões de professores sobre as dificuldades que enfrentam no contexto escolar em relação à violência. Com base na hermenêutica desconstrutiva, o texto compreende a escola como um espaço de enfrentamento de estereótipos, preconceitos e fantasmas sobre a violência, facilitando a convivência com as diferenças. Conclui-se pela necessidade de pensar o arquivo escolar como uma "tela desconstrucionista", capaz de romper com o individualismo e a indiferença em relação ao outro. Palavras-chave: Mal de arquivo. Filosofia da Educação. Violência escolar.</dc:description>
    <dc:abstract>This text presents a critical reflection on the archives of school violence as a philosophical and psychoanalytical problem. Using Derrida's expression "archive fever" as a starting point, it discusses teaching movements against the school violence issue. From a research experience, teachers' opinions about the difficulties they face in the school context in relation to violence are presented. Based on the deconstructive hermeneutics, the text understands the school as a space for confronting stereotypes, prejudices and ghosts about violence, facilitating coexistence with differences. It is concluded that there is a need to think of the school archive as a "deconstructionist canvas", capable of breaking with individualism and indifference towards the other.</dc:abstract>
    <dc:extent>19 p.</dc:extent>
    <dc:quality>n/a</dc:quality>
    <dc:rights>CC BY 4.0</dc:rights>
  </dc:Description>
</rdf:RDF>
```

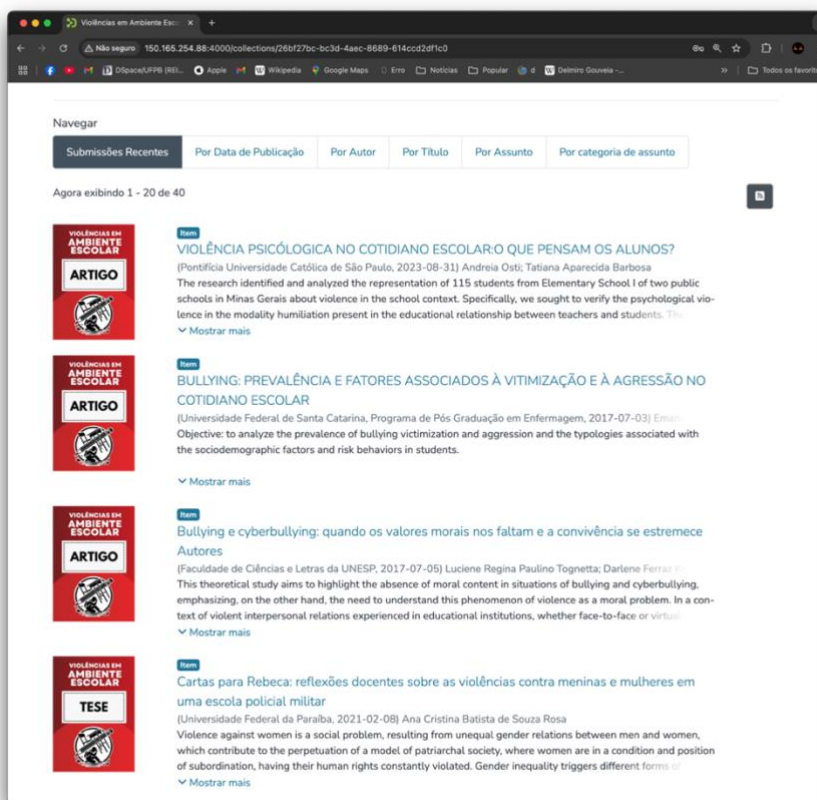
Fonte: Elaboração própria.

As imagens ilustrativas das coleções e itens da base de dados (*thumbnails*) também foram geradas através do *ChatGPT*. Cada coleção e cada item cadastrado na base possuem uma imagem específica que os representa visualmente, facilitando a identificação rápida pelos usuários. Essa abordagem tornou possível obter uma padronização estética e funcional nas representações gráficas, contribuindo para uma interface mais intuitiva.

O produto de *software* selecionado para armazenar e disponibilizar os registros da base de dados foi o *DSpace 7*. Essa versão foi escolhida porque, no momento da instalação, era a versão de produção recomendada (*current release*). Embora o *DSpace 8* já estivesse disponível, ainda não era considerada a opção mais estável. O *DSpace 7* foi instalado de forma containerizada em um servidor *Linux Ubuntu* virtualizado nas dependências da Superintendência de Tecnologia da Informação da Universidade Federal da Paraíba (STI/UFPB).

Para estruturar logicamente a base de dados, criamos uma comunidade *DSpace* intitulada *Violências e Memórias - Rede Interdisciplinar de Estudos sobre Violências (RIEV)*. Dentro desta comunidade criamos oito coleções, a saber: 1-*Violências Contra Crianças e Adolescentes*, 2-*Violências Contra Imigrantes*, 3-*Violências Contra Indígenas*, 4-*Violências Contra LGBTQIAPN+*, 5-*Violências Contra Negros*, 6-*Violências Contra Quilombolas*, 7-*Violências de Massa* e 8-*Violências em Ambiente Escolar* (Figura 3).

Figura 3 – Comunidade e Coleções da base de dados



Fonte: Elaboração própria.

Para ampliar a aderência dos dados disponibilizados ao padrão FAIR, utilizamos a ferramenta de avaliação de conjuntos de dados *FairDataBR* (*FairDataBR*, 2025) a fim de verificar seu grau de conformidade. Indicamos também que os elementos de metadados de cada item depositado na base de dados são compatíveis com o padrão *Dublin Core*.

3 RESULTADOS

A base de dados textual de referência sobre violências está operacional em fase de testes de povoamento, encontra-se disponível para consulta pela comunidade no endereço <http://150.165.254.88:4000/home> (Figura 4).

Figura 4 – Exemplos de itens depositados na base de dados



Fonte: Elaboração própria

Uma das possibilidades de aprimoramento no processo de construção da base de dados refere-se à demora na coleta de artigos, dissertações, teses, notícias e outros documentos para a extração de seus descritores pelos pesquisadores. Esse desafio poderia ser mitigado por meio da implantação de um *harvester* compatível com OAI-PMH,

permitindo a coleta automatizada de metadados diretamente dos portais de periódicos, o que é particularmente relevante, considerando que os artigos representam a maior parte dos recursos informacionais depositados na base.

De forma análoga, o processo de ingestão (*ingestion*) dos itens na base de dados, atualmente realizado de maneira manual, é bastante oneroso em termos de tempo e poderia ser otimizado por meio da automação dessa etapa. Além de reduzir o esforço humano, a automação contribuiria para minimizar erros de inserção de dados, garantindo maior confiabilidade no repositório. Uma possibilidade seria o recurso de ingestão por lotes (*batch import*) disponível no *DSpace 7*, que permite a inserção massiva de registros e padroniza o fluxo de ingestão, reduzindo a necessidade de intervenção manual.

4 CONSIDERAÇÕES FINAIS

A aplicação da Inteligência Artificial na organização de bases de dados pode representar um avanço significativo para a otimização de processos na Ciência da Informação, permitindo maior agilidade na estruturação e gestão de repositórios. O uso de ferramentas como o *ChatGPT* possibilita a automação de tarefas operacionais, como a geração de metadados e a padronização de descritores, além de reduzir o tempo necessário para a realização dessas atividades, tornando a administração das informações mais eficiente e acessível. No entanto, é fundamental considerar que, apesar das vantagens proporcionadas pela Inteligência Artificial, a validação e a supervisão humanas continuam indispensáveis para garantir a qualidade, a consistência e a adequação dos dados armazenados.

No caso deste estudo, a base de dados desenvolvida atende às necessidades dos pesquisadores vinculados à Rede Interdisciplinar de Estudos sobre Violências, oferecendo um repositório temático estruturado e de fácil acesso, além de se apresentar como uma ferramenta estratégica para o fortalecimento das pesquisas sobre o tema. Ao reunir e organizar informações de forma padronizada, a base contribui para a disseminação do conhecimento sobre os diferentes tipos de violência, proporcionando uma abordagem mais sistemática na recuperação de dados e na formulação de novas investigações acadêmicas.

Mencionamos ainda que a adoção dos princípios FAIR expande o potencial de colaboração científica, promovendo a interoperabilidade com outras bases e repositórios e incentivando o compartilhamento responsável de informações.

Um desafio encontrado na condução do projeto foi a dificuldade em localizar profissionais na área de Tecnologia da Informação com domínio no processo de configuração do *DSpace 7*. Por conta disso, muitas de nossas demandas referentes à customização desse *software*, no que diz respeito à configuração de *layouts* e à estrutura de apresentação de metadados, foram afetadas.

Este estudo mostra a importância da incorporação de novas tecnologias na gestão de dados, ao mesmo tempo em que evidencia a necessidade de reflexões contínuas sobre as possibilidades e desafios do uso da Inteligência Artificial na Ciência da Informação.

REFERÊNCIAS

FAIRDATABR. **FairDataBR**: uma ferramenta para a avaliação de conjuntos de dados. Disponível em: <https://wrco.ufpb.br/fair/index.html>. Acesso em: 24 fev. 2025.

GABRIEL JUNIOR, R. F.; LAIPELT, R. C. Thesa: ferramenta para construção de tesauro semântico aplicado interoperável. **Revista P2P & Inovação**, Rio de Janeiro, v. 3, n. 2, p. 124-145, mar./set. 2017.

RUSSEL, S.; NORVIG, P. **Inteligência artificial**. Rio de Janeiro: Elsevier, 2004.

UNIVERSIDADE FEDERAL DA PARAÍBA. **Rede interdisciplinar de estudos sobre violências**. Disponível em: <https://www.ufpb.br/riev/index.php>. Acesso em: 20 fev. 2025.

WILKINSON, M. D. *et al.* The FAIR guiding principles for scientific data management and stewardship. **Scientific Data**, [S. l.], v. 3, art. 160018, 2016. Disponível em: <https://www.nature.com/articles/sdata201618>. Acesso em: 20 fev. 2025.

AGRADECIMENTOS

Trabalho apoiado pelas da Chamadas FAPESQ-PB/CNPq nº 77/2022 e CNPq nº 09/2018.



[4.0 Internacional](https://creativecommons.org/licenses/by/4.0/)

Esta licença permite compartilhamento, remixe, adaptação e criação a partir do trabalho, mesmo para fins comerciais, desde que sejam atribuídos créditos ao(s) autor(es). Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.

Como citar este trabalho:

DIAS, Guilherme Ataíde; OLIVEIRA, Tassyara Onofre; CAMPELO, Arthur Antônio Martins Sales; CARVALHO, Vivian Roberta Lima de. Utilizando o ChatGPT como ferramenta de suporte para a construção de uma base de dados textual sobre violências: apresentação e análise do processo de desenvolvimento. *In: WORKSHOP DE INFORMAÇÃO DADOS E TECNOLOGIA*, 8., 2025, Marília, SP. **Anais [...]**. Marília, SP: Universidade de Marília, 2025. DOI: <https://doi.org/10.22477/viii.widat.224>.