Journal of Magnetism and Magnetic Materials 597 (2024) 172001

Contents lists available at ScienceDirect

Journal of Magnetism and Magnetic Materials

journal homepage: www.elsevier.com/locate/jmmm

Research article

GPTArticleExtractor: An automated workflow for magnetic material database construction

Yibo Zhang [a,b,*], Suman Itani [a], Kamal Khanal [a], Emmanuel Okyere [a], Gavin Smith [a], Koichiro Takahashi [a], Jiadong Zang [a,*]

[a] *Department of Physics and Astronomy, University of New Hampshire, 9 Library Way, Durham, 03824, NH, USA*
[b] *Department of Chemistry, University of New Hampshire, 23 Academic Way, Durham, 03824, NH, USA*

ARTICLE INFO

Keywords:
Magnetic materials
Database
Large language models

ABSTRACT

A comprehensive database of magnetic materials is valuable for researching the properties of magnetic materials and discovering new ones. This article introduces a novel workflow that leverages large language models for extracting key information from scientific literature. From 22,120 articles in the Journal of Magnetism and Magnetic Materials, a database containing 2,035 magnetic materials was automatically generated, with ferromagnetic materials constituting 76% of the total. Each entry in the database includes the material's chemical compounds, as well as related structures (space group, crystal structure) and magnetic temperatures (Curie, Néel, and other transitional temperatures). To ensure data accuracy, we meticulously compared each entry in the database against the original literature, verifying the precision and reliability of each entry.

1. Introduction

Magnetic materials play an indispensable role in modern science and engineering domains. Their applications span a wide range, from data storage devices like hard disks and tapes to electrical power conversion and transmission systems, and extend to medical and consumer electronics. The unique magnetic properties of these materials enable various technological applications to function efficiently and precisely [1,2].

While discovering antiferromagnets is relatively easy due to the superexchange nature therein [3], predicting new ferromagnets poses significant challenges. The interplay of the correlation and itinerancy is believed to be the origin of most ferromagnetism, but a complete understanding is yet to be available [4]. The first-principles calculations, such as density functional theory (DFT), are generally accurate in determining the energy of non-magnetic materials. Unfortunately, they often result in incorrect predictions of magnetic properties, especially for magnetic materials where the electronic correlation is relevant, particularly ferromagnetism. Many magnetic materials are strongly correlated, and one often has to go beyond DFT or DFT + U approaches [5]. Moreover, the limitations of the DFT+U method in predicting magnetic structures are demonstrated by the disagreement between the observed experimental and the calculated magnetic configurations [6]. Another drawback of the DFT calculations is that to get a successful description of exchange interactions in some materials (like antiferromagnetic Mn_3O_4), one has to use hybrid exchange–correlation functionals, which have a considerable computational cost [7]. Another study [8] also suggests that none of the functionals work in all conditions, indicating the need for further development of exchange–correlation functionals. The paramount impact is clearly doomed if a reliable way of predicting new ferromagnets cannot be achieved.

Data-driven materials discovery is an effective solution to this challenge. For example, the high-throughput (HT) methods efficiently pinpoint materials with distinct properties by conducting broad screenings within databases [9–11]. However, HT outcomes, rooted in the first-principles calculations, are still constrained by these algorithms' drawbacks above. On the other hand, Data mining techniques in materials science are significantly enhanced by various analytical methods. Key among these are machine learning [12,13], deep learning [14], linear regression [15], and trend analysis [16], which all garnered success in identifying materials and developing screening models. However, the effectiveness of these methods relies heavily on the availability of a comprehensive dataset, which has yet to be ready for magnetic materials.

Databases based on experimentally verified information do exist [17–20], yet they all lack magnetic property information, such as Curie or Néel temperatures. The Bilbao Magnetic Materials Database

[*] Corresponding authors.
E-mail addresses: yibo.zhang@unh.edu (Y. Zhang), jiadong.zang@unh.edu (J. Zang).

https://doi.org/10.1016/j.jmmm.2024.172001
Received 22 December 2023; Received in revised form 24 February 2024; Accepted 26 March 2024
Available online 28 March 2024

[21] comes closest to meeting these criteria. It offers a complete lattice and magnetic structure for each entry and leads to success in discovering magnetic topological insulators [22]. However, its dataset is limited to 1890 entries, a scale insufficient to support more intricate deep-learning models. Some efforts to create a magnetic material database have been noticed in articles [12,13,23] and books [24–26]. Notably, Matthew et al. [27] developed a rule-based syntax toolkit for the automatic extraction of chemical information. It requires substantial programming efforts to use regular expressions, which may fail to adapt to different articles' writing styles and layouts. Callum et al. [28] automated the analysis of 68,078 articles in physics and chemistry, extracting information related to chemical composition and magnetic phase transitions (e.g., Curie and Néel temperatures) from 39,822 of them. Luke et al. [29] implemented literature scraping without rules by fine-tuning a BERT model. These works do not address the extraction of structural information, which is essential for the descriptor [30] in data-based materials discovery. However, we admit the difficulty in extracting structural information since they are often embedded within extensive text. Furthermore, these works do not distinguish experimental and theoretical papers, risking contamination of experimental data with theoretical data in the database. Data quality needs to be further evaluated.

The recent development of large language models (LLM) in natural language processing (NLP) brings new insight into this challenge. It can be traced back to the advent of the Transformer model [31]. This model introduced the Self-Attention mechanism, laying the groundwork for NLP. Further advances came with the pre-trained BERT model [32], solidifying the field's foundation. OpenAI subsequently followed suit by releasing their version of a pre-trained model [33]. This technology's widespread application and significant advancement are attributed to OpenAI's theory of fine-tuning [34]. This theory empowers LLMs to understand human intent and engage in meaningful dialogue and interaction based on these intents. It makes machine reading of scientific literature possible.

The process of training LLMs generally consists of two main steps. The first step, the pre-training [32,33,35], involves embedding human knowledge and linguistic structures into the model's internal parameters, similar to lossy compression [36] of this knowledge in the model's memory. The second step, Fine-Tuning, particularly through the Instruction Tuning [35], transforms a pre-trained large language model into a chat model capable of effective human interaction.

However, the model's effectiveness is limited by its training methods. It performs well in areas with ample data but struggles in less-explored fields. Additionally, as it is not trained specifically for literature extraction, using it to convert magnetic material texts directly into a database could be ineffective. These limitations highlight the importance of precise instructions and prompt engineering [37].

Prompt Engineering enhances interactions with LLMs by designing effective prompts to guide more accurate outputs, even when the model has limited knowledge in areas such as magnetic materials or when instruction tuning is not specifically optimized for certain tasks.

In this work, through prompt engineering of LLM, we have developed a novel method named GPTArticleExtractor for extracting bibliographic data to create a database. To validate the feasibility of our approach, we have constructed a database for magnetic materials based on published papers in the Journal of Magnetism and Magnetic Materials.

2. Material and methods

This section presents a comprehensive overview of the workflow employed in GPTArticleExtractor. Our principal aim is to extract critical information pertaining to each experimentally reported material, including the material's chemical formula, various magnetic temperatures (such as Curie, Curie–Weiss, Néel, and other transitional temperatures), as well as structural details like the space group and crystal structure.

Table 1
Prompts.

Task	Prompt
① Study type filter	Based on the provided abstract and title of an article, determine if this article is a research study about materials. Please return 'True' if it is and 'False' if it is not.
② Temperature	Please determine the Curie temperature, Néel temperature, transitional temperature or critical temperature discussed in this article. Please include the specific name of these temperature. Consider the following documents: {doc_tem}
③ Materials	Please identify the chemical compounds being studied in this article. Consider the following documents: {doc_material}
④ Structure	Please help me find and summarize the paragraphs related to material structure, crystal structure, lattice structure, and other relevant concepts in the given text: {doc_structure}
⑤ Space group	Please help me find and summarize the paragraphs discussing space groups, their properties, and their relevance to crystal structures in the given text: {doc_space_group}
⑥ Summary	As a research assistant, your task is to identify the material being studied in this article, its associated Curie temperature, transition temperatures and critical temperatures, and structural information like space group, lattice Constant, crystal structure, lattice structure etc., based on the following details provided to you: the title, abstract, and summary details from GPT. The title and abstract should be your primary sources, while GPT's summaries should serve as secondary references. Title: {title}, Abstract: {abstract}, GPT Materials Summary: {materials}, GPT Temperature Summary: {temperature}, GPT Structural Information: {structure}. The output should be in a JSON readable format.

2.1. Data and tools

DATA: This study utilizes Elsevier's Text Mining Application Programming Interface (API) for text retrieval through DOI queries. It focuses on articles from the Journal of Magnetism and Magnetic Materials (JMMM) from 2000 to 2023. Each article is provided in Extensible Markup Language (XML) format. A substantial database comprising 22,120 corpora was downloaded.

Model: In this work, we employ two major language models, GPT-3.5 and GPT-4, for the automated extraction of structured data from scientific literature. Each model has unique strengths: GPT-3.5 ($0.002 per 1K tokens) is effective for initial filtering and categorization, while GPT-4 ($0.06 per 1K tokens) excels at deeper analysis and summarization. To ensure both efficiency and cost-effectiveness, the majority of the tasks are handled by GPT-3.5. For the final summarization phase, we switch to GPT-4 to ensure quality.

Vector Database: Our method is initiated by tokenizing articles segmenting the text into chunks of 500 tokens each. We then analyze these segments in a vector space, employing metrics such as Euclidean distance to measure word similarity. This analysis involves comparing the vectorized text segments with a question to determine similarity and identify the most relevant article segments for the question. The five most closely related segments to a query are identified using Facebook AI Similarity Search (FAISS) [38]. This approach allows us to concentrate on the most pertinent segments, significantly enhancing the quality of our answers by focusing on relevance and minimizing input size.

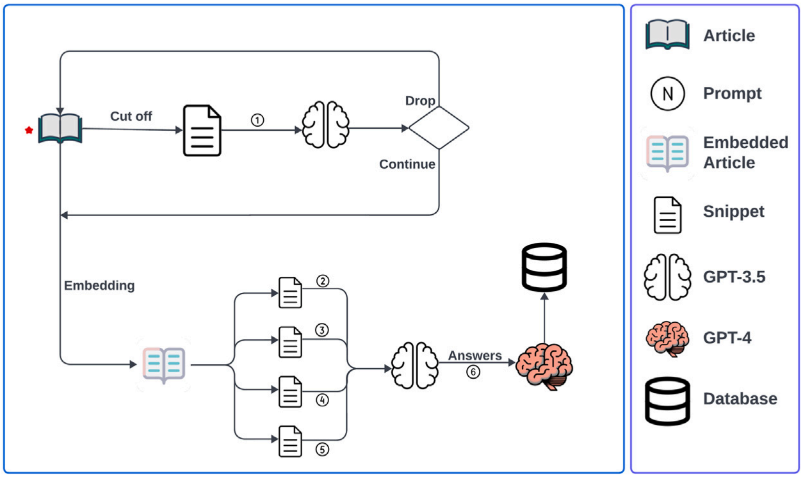

Fig. 1. Flowchart depicting the transformation from original articles to structured data through various prompts and processing stages, utilizing different GPT versions.

2.2. *GPTArticleExtractor structure*

Our platform offers a streamlined, automated text analysis process that accepts plain text and produces structured data. Fig. 1 illustrates the complete workflow. The corresponding prompts for each step are listed in Table 1.

First, we extract the title and abstract from an article and use GPT-3.5, guided by question ①, for an initial screening to identify articles focused on new material research. If the model deems it relevant (judged as True), we proceed to the next step: vectorizing the article using a vector database. This vectorization allows us to find the most relevant snippets based on specific sub-questions (②, ③, ④, ⑤). These snippets and the questions are fed back into GPT-3.5, which generates corresponding answers. This step can also be regarded as a form of text summary. Finally, these answers, along with question ⑥, are fed into GPT-4 for answer aggregation and structuring, which are then inserted into the database.

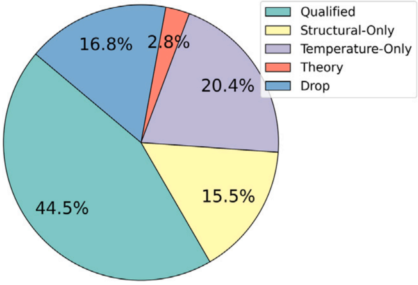

Fig. 2. Data Completeness Pie Chart - 44.5% entries ready, 15.5% contain structure only, 20.4% contain temperature only, 2.8% theory, 16.8% irrelevant. Finalized database: 2035 entries out of 4639 articles, yielding an accuracy rate of 83.2%.

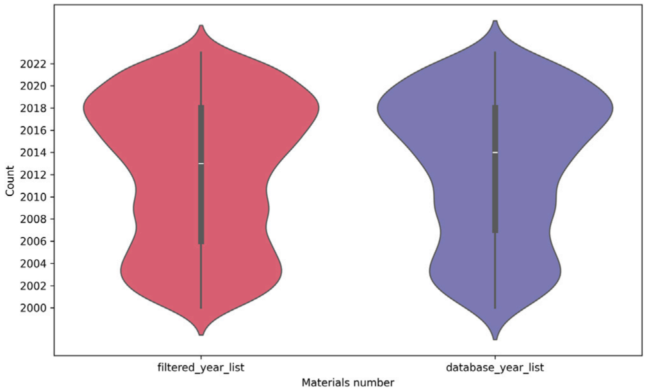

Fig. 3. Violin plots comparing the Curie temperature distribution over the years for two datasets: filtered and complete database records.

3. Results

Using GPTExtractor, 4639 articles were identified to report magnetic materials. To evaluate the algorithm's effectiveness, we manually checked the GPT-generated entry with the original article. As shown in Fig. 2, 44.5% entries have complete structural and transition temperature information. These entries are ready to be included in the database. 15.5% and 20.4% are attributed to structure and temperature only, respectively. These articles may be used as backup options and could be incorporated into the final database once the missing information is supplied. Noticeably, 2.8% entries are extracted from theory papers only. Most transition temperatures therein are computed by Monte Carlo simulations or combined with the first-principles calculations. Since only experimentally verified data are included in this database, they have to be discarded. Finally, the remaining 16.8% entries are completely irrelevant. Some of them are non-magnetic materials. Our finalized database comprises 2035 entries, with duplications removed. Entries sharing identical chemical formulas are distinguished by their different lattice structures, ensuring the database's integrity. The overall 83.2% yield rate shows the accuracy of the algorithm. This suggests that constructing a larger database focused on magnetic materials is viable once the pool of journal articles is expanded.

Violin plots in Fig. 3 show the distribution of publication years of the articles identified. The plot on the left represents the distribution of publication years for articles in the database screened by workflow. In contrast, the plot on the right shows the distribution of articles eventually listed in the database. It is evident that a significant number of articles published before approximately 2012 were excluded during the manual review process. This observation suggests that the model is more proficient at comprehending articles from the last decade and is less susceptible to errors. Articles published before 2000 are notably absent since they are not analyzed in this work. Those articles are less commonly preserved in text form but often exist in PDF format only. Incorporating Optical Character Recognition (OCR) [39] in future work could remedy this limitation.

Concerning Curie temperatures, as illustrated in Fig. 4, the distribution of this database spans from 0 to 1500 K. Notably, a significant concentration of data points resides in the low-temperature range between 0 and 400 K, whereas the high-temperature region is relatively underrepresented. As for Néel temperatures, Fig. 5 shows that the

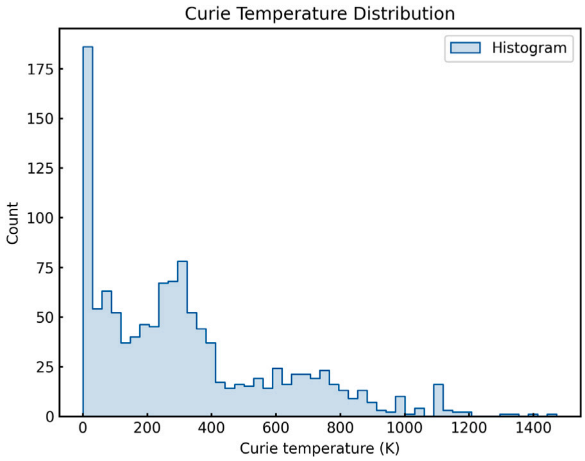

Fig. 4. Histogram depicting the distribution of Curie temperatures within our comprehensive database of magnetic materials.

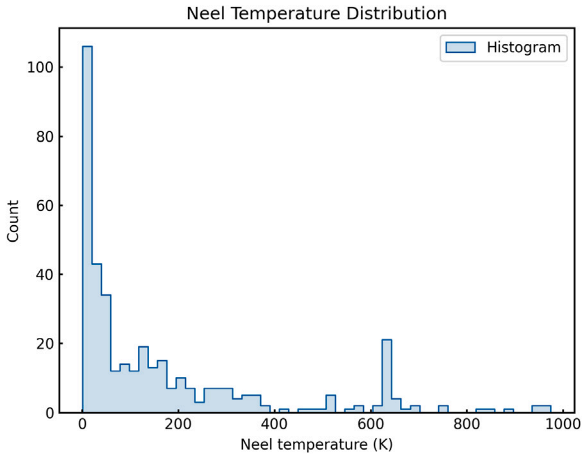

Fig. 5. Histogram depicting the distribution of Néel temperatures within our comprehensive database of magnetic materials.

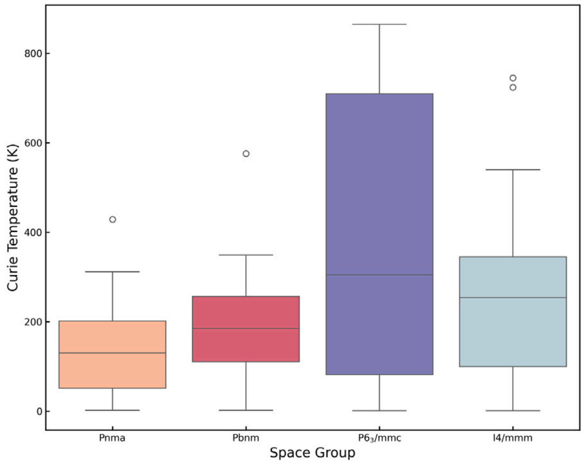

Fig. 6. Box plot displaying the distribution of Curie temperatures across different space groups in our database. Each box represents the interquartile range (IQR) and median, and outliers are indicated by individual points.

principal distribution ranges from 0 to 1000 K, revealing a greater prevalence of data in the low-temperature area. Additionally, we investigated the influence of different space groups on the distribution of Curie temperatures. Fig. 6 reveals the significant impact of diverse space groups on Curie temperatures. The *P63/mmc* space group notably exhibits a broader distribution, especially at higher temperatures. Across various structures, notable contributions from Perovskite include 18% in *Pbnm*, 10% in *Pnma*, and 5.2% in *I4/mmm*. Wurtzite constitutes 5.8% in *P63/mmc*, and Spinel structures account for 1.3% in *I4/mmm*. Calculating ratios for the entire database further highlights these correlations: Perovskite at 10.9%, Spinel at 8.5%, and Wurtzite at 0.64%. This data underscores a discernible link between crystal structure and space group, providing valuable insights into understanding Curie temperatures.

After constructing the database, we randomly selected 200 articles that the GPT model for manual check had not initially identified. We discovered that 15.5% journals were missed by the algorithm. Among them 5% articles present relevant information in terms of figures and tables. It is thus not surprising that they were not captured by the current text-based algorithm and should be excluded in the count. Relevant sentences in these articles are usually long and complex, posing challenges to the language model's ability to extract answers. The 500-token limit during the embedding process could truncate such sentences, leading to incomplete source data for the GPT model. Additionally, information regarding the temperature of magnetic materials is often presented in images or tables, which the GPT model could not interpret due to formatting constraints. Based on the error rate, about 2000 additional articles are potentially eligible to be included in the database.

Several attempts could be explored in future work to address these limitations and include all eligible entries to the database. First, given the recent upgrades to the GPT model with the increasing token limit, the length of truncated embeddings could be extended to ensure answers are not prematurely cut off. The prompts could be refined, and additional few-shot learning samples could be used to accommodate articles that use diverse language styles. Future efforts could employ OCR to transcribe text from visual content to address issues associated with images, PDFs, and tables. Multimodal language models may also be leveraged to recognize and convert image-based content into text, which can then be processed further by the GPT model.

4. Conclusions

GPTArticleExtractor provides an efficient way to automatically extract information about chemical compounds, magnetic temperature (Curie and Néel temperatures), and material structures. This tool is highly scalable and can be easily adapted for multiple database domains by simply changing the input queries. Its utility in magnetic materials is particularly notable, offering a user-friendly alternative to more complex extraction methods. A database covering various magnetic materials is available on our website (https://MagneticMaterials.org). The database will be constantly updated and expanded.

CRediT authorship contribution statement

Yibo Zhang: Writing – original draft, Visualization, Validation, Software, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization, Writing – review & editing. **Suman Itani:** Data curation, Writing – review & editing. **Kamal Khanal:** Data curation. **Emmanuel Okyere:** Data curation. **Gavin Smith:** Data curation. **Koichiro Takahashi:** Data curation. **Jiadong Zang:** Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgment

This work was supported in part by a Collaborative Research Excellence (CoRE), USA grant from the University of New Hampshire.

References

[1] J.M.D. Coey, Magnetism and Magnetic Materials, Cambridge University Press, 2012.

[2] N.A. Spaldin, Magnetic Materials: Fundamentals and Applications, Cambridge University Press, 2010.

[3] P.W. Anderson, Antiferromagnetism. Theory of superexchange interaction, Phys. Rev. 79 (2) (1950) 350.

[4] H. Tasaki, The origin of ferromagnetism, in: Physics and Mathematics of Quantum Many-Body Systems, Springer International Publishing, Cham, 2020, pp. 371–455.

[5] E. Pavarini, Solving the strong-correlation problem in materials, Riv. Nuovo Cimento 44 (11) (2021) 597–640.

[6] J.A. Blanco, P.J. Brown, Comment on DFT+U search for the energy minimum among eight collinear and noncollinear magnetic structures of GdB_4, Phys. Rev. B 79 (2009) 216401.

[7] R. Ribeiro, S. de Lazaro, S. Pianaro, Density functional theory applied to magnetic materials: Mn3O4 at different hybrid functionals, J. Magn. Magn. Mater. 391 (2015) 166–171.

[8] A.H. Romero, M.J. Verstraete, From one to three, exploring the rungs of Jacob's ladder in magnetic alloys, Eur. Phys. J. B 91 (8) (2018) 193.

[9] S. Curtarolo, G.L. Hart, M.B. Nardelli, N. Mingo, S. Sanvito, O. Levy, The high-throughput highway to computational materials design, Nat. Mater. 12 (3) (2013) 191–201.

[10] A. Jain, S.P. Ong, G. Hautier, W. Chen, W.D. Richards, S. Dacek, S. Cholia, D. Gunter, D. Skinner, G. Ceder, et al., Commentary: The materials project: A materials genome approach to accelerating materials innovation, APL Mater. 1 (1) (2013).

[11] M. Liu, S. Meng, Atomly. net materials database and its application in inorganic chemistry, Sci. Sin-Chim. 53 (2023) 19–25.

[12] J. Nelson, S. Sanvito, Predicting the Curie temperature of ferromagnets using machine learning, Phys. Rev. Mater. 3 (10) (2019) 104405.

[13] P. Singh, T. Del Rose, A. Palasyuk, Y. Mudryk, Physics-informed machine-learning prediction of Curie temperatures and its promise for guiding the discovery of functional magnetic materials, Chem. Mater. 35 (16) (2023) 6304–6312.

[14] M. Alverson, S. Baird, R. Murdock, J. Johnson, T. Sparks, et al., Generative adversarial networks and diffusion models in material discovery, 2023.

[15] D.-N. Nguyen, T.-L. Pham, V.-C. Nguyen, A.-T. Nguyen, H. Kino, T. Miyake, H.-C. Dam, A regression-based model evaluation of the Curie temperature of transition-metal rare-earth compounds, in: Journal of Physics: Conference Series, Vol. 1290, IOP Publishing, 2019, 012009.

[16] J.K. Byland, Y. Shi, D.S. Parker, J. Zhao, S. Ding, R. Mata, H.E. Magliari, A. Palasyuk, S.L. Bud'ko, P.C. Canfield, et al., Statistics on magnetic properties of Co compounds: A database-driven method for discovering co-based ferromagnets, Phys. Rev. Mater. 6 (6) (2022) 063803.

[17] A. Vaitkus, A. Merkys, S. Gražulis, Validation of the crystallography open database using the crystallographic information framework, J. Appl. Crystallogr. 54 (2) (2021) 661–672.

[18] D. Zagorac, H. Müller, S. Ruehl, J. Zagorac, S. Rehme, Recent developments in the inorganic crystal structure database: theoretical crystal structure data and related features, J. Appl. Crystallogr. 52 (5) (2019) 918–925.

[19] G. Bergerhoff, I. Brown, F. Allen, et al., Crystallographic databases, Int. Union Crystallogr. Chester 360 (1987) 77–95.

[20] P. Villars, K. Cenzual, et al., Pearson's crystal data: crystal structure database for inorganic compounds, 2007.

[21] S.V. Gallego, J.M. Perez-Mato, L. Elcoro, E.S. Tasci, R.M. Hanson, K. Momma, M.I. Aroyo, G. Madariaga, MAGNDATA: towards a database of magnetic structures. I. The commensurate case, J. Appl. Crystallogr. 49 (5) (2016) 1750–1776.

[22] Y. Xu, L. Elcoro, Z.-D. Song, B.J. Wieder, M. Vergniory, N. Regnault, Y. Chen, C. Felser, B.A. Bernevig, High-throughput calculations of magnetic topological materials, Nature 586 (7831) (2020) 702–707.

[23] Y. Xu, M. Yamazaki, P. Villars, Inorganic materials database for exploring the nature of material, Japan. J. Appl. Phys. 50 (11S) (2011) 11RH02.

[24] T.F. Connolly, Bibliography of Magnetic Materials and Tabulation of Magnetic Transition Temperatures, Springer Science & Business Media, 2012.

[25] K.J. Buschow, Handbook of Magnetic Materials, Elsevier, 2003.

[26] P. Villars, K. Cenzual, J. Daams, Y. Chen, S. Iwata, Data-driven atomic environment prediction for binaries using the mendeleev number: Part 1. Composition AB, J. Alloys Compd. 367 (1–2) (2004) 167–175.

[27] M.C. Swain, J.M. Cole, ChemDataExtractor: a toolkit for automated extraction of chemical information from the scientific literature, J. Chem. Inf. Model. 56 (10) (2016) 1894–1904.

[28] C.J. Court, J.M. Cole, Auto-generated materials database of Curie and Néel temperatures via semi-supervised relationship extraction, Sci. Data 5 (1) (2018) 1–12.

[29] L.P. Gilligan, M. Cobelli, V. Taufour, S. Sanvito, A rule-free workflow for the automated generation of databases from scientific literature, 2023, arXiv preprint arXiv:2301.11689.

[30] L. Himanen, M.O. Jäger, E.V. Morooka, F.F. Canova, Y.S. Ranawat, D.Z. Gao, P. Rinke, A.S. Foster, DScribe: Library of descriptors for machine learning in materials science, Comput. Phys. Comm. 247 (2020) 106949.

[31] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A.N. Gomez, Ł. Kaiser, I. Polosukhin, Attention is all you need, Adv. Neural Inf. Process. Syst. 30 (2017).

[32] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, Bert: Pre-training of deep bidirectional transformers for language understanding, 2018, arXiv preprint arXiv:1810.04805.

[33] T. Brown, B. Mann, N. Ryder, M. Subbiah, J.D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, et al., Language models are few-shot learners, Adv. Neural Inf. Process. Syst. 33 (2020) 1877–1901.

[34] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray, et al., Training language models to follow instructions with human feedback, Adv. Neural Inf. Process. Syst. 35 (2022) 27730–27744.

[35] H. Touvron, L. Martin, K. Stone, P. Albert, A. Almahairi, Y. Babaei, N. Bashlykov, S. Batra, P. Bhargava, S. Bhosale, et al., Llama 2: Open foundation and fine-tuned chat models, 2023, arXiv preprint arXiv:2307.09288.

[36] G. Delétang, A. Ruoss, P.-A. Duquenne, E. Catt, T. Genewein, C. Mattern, J. Grau-Moya, L.K. Wenliang, M. Aitchison, L. Orseau, et al., Language modeling is compression, 2023, arXiv preprint arXiv:2309.10668.

[37] Y. Zhou, A.I. Muresanu, Z. Han, K. Paster, S. Pitis, H. Chan, J. Ba, Large language models are human-level prompt engineers, 2022, arXiv preprint arXiv:2211.01910.

[38] J. Johnson, M. Douze, H. Jégou, Billion-scale similarity search with GPUs, IEEE Trans. Big Data 7 (3) (2019) 535–547.

[39] C. Wick, C. Reul, F. Puppe, Calamari-a high-performance tensorflow-based deep learning package for optical character recognition, 2018, arXiv preprint arXiv:1807.02004.