Exploring Portuguese Hate Speech Detection with Transformers

Gabriel Assis[1], **Annie Amorim**[1], **Jonnathan Carvalho**[2],
Daniel de Oliveira[1], **Daniela Vianna**[3] **and Aline Paes**[1]
[1] Institute of Computing, Universidade Federal Fluminense, Niterói, RJ, Brazil
[2] Department of Informatics, Instituto Federal Fluminense, Itaperuna, RJ, Brazil
[3] JusBrasil
{*assisgabriel,annieamorim*}*@id.uff.br*, *joncarv@iff.edu.br*,
{*danielcmo,alinepaes*}*@ic.uff.br*, *dvianna@gmail.com*

Abstract

Social Media platforms, vital for debate and communication, also grapple with misinformation and hateful comments. This work examines the detection of hate speech in Portuguese, contemplating its unique linguistic and cultural nuances. Leveraging Transformer-based models and different training and activation strategies, eight models with variations in architecture, size, and pre-training *corpora* are evaluated. Our findings show that, even though large generative models with enhanced prompts exhibited promising results, tuned small language models are still superior in addressing this task.

1 Introduction

Social Media platforms have become essential for debate and enabled unprecedented communication. However, they have also introduced significant challenges, such as spreading misinformation and the proliferation of hateful comments (Pelle et al., 2018; Aluru et al., 2020). In the interim, the Transformer (Vaswani et al., 2017) architecture has emerged, demonstrating state-of-the-art results in various scenarios, including classification problems (Fortuna and Nunes, 2018).

However, detecting hate speech remains an open issue, notably lacking resources in languages other than English, such as Portuguese (Jahan and Oussalah, 2023). The inherent characteristics of the language play a crucial role in this context, as the use of figures of speech and cultural nuances can significantly complicate this problem (Jang et al., 2023). On the other hand, an equally important consideration is the sensitivity of the domain, where both types of misclassification – falsely identifying content as problematic and failing to identify problematic content – are critical, as they could lead to censorship or a failure to protect vulnerable groups.

In this context, we approach the problem as: **Given a social media post P written in Portuguese, pre-process it returning X, and classify it as belonging to one of the three classes in Y = {"hate speech", "offensive" or "neutral"}**, where offensive comments encompass rude or insulting communication, and hate speech involves expressions of hate towards an individual or a group, rooted on characteristics like ethnicity and gender (Pelle et al., 2018; Vargas et al., 2021).

Our work aims to investigate the performance of the prominent Transformer architecture to tackle this critical task, thereby contributing to safeguarding a resilient and pluralistic environment on social media. We explore eight models varying in architecture, size, and the *corpora* on which they were pre-trained. Specifically, we consider three groups of models: (i) models based on the BERT (Devlin et al., 2019) architecture; (ii) Portuguese-language models based on the LLaMA (Touvron et al., 2023a,b) architecture; and (iii) general-purpose Large Language Models (LLMs). The first group consists of four models, peculiarly three alternatives specialized for Portuguese — including a model pre-trained on a *corpus* of tweets — and one multilingual alternative also pre-trained on tweets. The second group includes two 7-billion parameter models pre-trained on structured texts. Lastly, the third group involves one model from the popular GPT (Brown et al., 2020) family and the recently released Gemini-pro (Google, 2023), both not mainly pre-trained in Portuguese. This way, we contribute to a diverse study of models to address the challenging domain of detecting Portuguese hate speech on social media platforms.

2 Related Work

Identifying hate speech on social media has become a significant topic in recent years. Yet, the number of studies focusing on the peculiarities of the Portuguese language remains limited compared to English (Jahan and Oussalah, 2023). Some approaches address models based on BERT and its state-of-the-art capabilities for classification tasks. In this context, (da Silva and Rosa, 2023) evaluated several distinct models, finding superior results in BERT-based models, such as BERTimbau (Souza et al., 2020), a finding reinforced by (Santos et al., 2022). (Jahan and Oussalah, 2023) present results indicating that language-specific models achieve better outcomes than multilingual alternatives.

Furthermore, LLMs and their remarkable abilities, are also being investigated for this task. (Assis et al., 2024) compare the GPT-3.5 and the Brazilian chatbot Maritalk[1] with pt-BR BERT-based options, concluding that the latter group achieve better results. (Oliveira et al., 2024) contrast the same pair of LLMs, with a prompt engineering approach, and underscore Maritalk's potential despite GPT's higher performance. Additionally, (Chiu et al., 2022) assessed ChatGPT for detecting hate speech content, and (Nguyen et al., 2023) evaluated tuned LLaMA-2 models for detecting sexual, predatory, and abusive texts.

None of the aforementioned works conducted a study that comparatively includes the same vast amount of Portuguese-language models tuned as a ternary classification problem. Also, decoder models as the foundation for classifiers and a more recent LLM in an in-context learning (Brown et al., 2020) approach have not been evaluated either.

3 Method

This section details the selected models, training methods for classifier models, and inference strategies for generative models.

3.1 Encoder-based Classifiers Training

We select encoder-based models as follows. First, we have BERT-based models pre-trained with Brazilian Portuguese *corpora*: BERTimbau (Souza et al., 2020) in its large version, and AlBERTina (Rodrigues et al., 2023) in its 100m version, both pre-trained with more well-formed language; also BERTweet.BR (Carneiro, 2023), that is pre-trained with a *corpus* of tweets. Bernice (DeLucia et al., 2022) is also pre-trained on a Twitter data *corpus*, but it is multilingual. The most common fine-tuning strategy was adopted, stacking a classifier layer onto the language model and adjusting the model weights according to the training examples.

[1] https://www.maritaca.ai/

3.2 Decoder-based Classifiers Training

Regarding the decoder-based classifier models, Sabiá-7b-1 (Pires et al., 2023), which is built on the LLaMA-1 architecture, and Gervásio-7b-PTBR (Santos et al., 2024), built on the LLaMA-2 architecture, were selected. Both models were pre-trained on well-structured Portuguese text *corpora*. We used a tuning approach similar to the one usually adopted in encoder-based classifiers: stacking a classifier layer onto the language model. This choice stems from the decoder output of LLMs holding semantic meaning from the input, serving as text representations for classification tasks with prominent results (Li et al., 2023).

3.3 Generative LLMs Activation

The popular GPT-3.5-turbo (Ouyang et al., 2022) and the Gemini-pro 1.0 (Google, 2023), recognized for their remarkable performance in recent benchmarks, were chosen as generative large language models. Due to the constraints in adjusting the weights of these large and closed models, our strategy leverages their in-context learning capabilities by activating them with prompts (Brown et al., 2020). The responses' effectiveness may be expressively influenced by how the prompts are crafted (White et al., 2023). This way, a well-known method involves embedding examples directly within the prompts. Our approach encompasses fixing the prompt instruction and exploring the choice of demonstrations and their impact on the models' performance. The instruction is as follows: CLASSIFIQUE O TEXTO DE REDE SOCIAL COMO "DISCURSO DE ODIO" OU "OFENSIVO" OU "NEUTRO".\N TEXTO: *target* \N CLASSE:[2].

We rely on four ways to assemble prompts using examples: **(a.) zero-shot**, with no examples; **(b.) one-shot**, which includes a single example from one class; **(c.) one-class-shot**, which incorporates one example per class; and **(d.) few-shot**, which uses more than one example per class, precisely two in this study. For selecting examples, we introduced three strategies: **(e.) random choice**, **(f.) based on semantic similarity**, and **(g.) based on the number of tokens**. Strategies **(f.)** and **(g.)** start by sorting the set of demonstration candidates into clusters per class. They then pick examples close and far from the test instances' embedding

[2] In English that would be: *Classify the social network text as "hate speech", "offensive", or "neutral". \n Text: target \n Class:*

	HateBR						OLID-BR						ToLD-BR					
	Strategy	prec.	rec.	acc.	f1	f1$_{hs}$	Strategy	prec.	rec.	acc.	f1	f1$_{hs}$	Strategy	prec.	rec.	acc.	f1	f1$_{hs}$
BERTimbau	Fine Tuning	0.803	**0.822**	0.862	**0.811**	**0.667**	Fine Tuning	0.637	0.664	0.663	0.623	0.596	Fine Tuning	0.529	0.598	0.599	0.474	0.065
AlBERTina	Fine Tuning	0.793	0.707	0.800	0.734	0.569	Fine Tuning	0.599	0.589	0.613	0.563	0.544	Fine Tuning	0.417	0.453	0.538	0.399	0.059
BERTweet.BR	Fine Tuning	0.768	0.793	0.846	0.779	0.583	Fine Tuning	0.625	**0.665**	**0.672**	**0.635**	**0.606**	Fine Tuning	**0.543**	**0.671**	**0.708**	**0.548**	**0.178**
Bernice	Fine Tuning	**0.830**	0.788	**0.863**	0.805	0.656	Fine Tuning	**0.640**	0.660	0.666	0.620	0.579	Fine Tuning	0.534	0.640	0.704	0.536	0.141
Sabiá-7b-1	Fine Tuning	0.465	0.379	0.526	0.322	0.129	Fine Tuning	0.422	0.434	0.532	0.437	0.412	Fine Tuning	0.383	0.381	0.531	0.355	0.000
Gervásio-7b	Fine Tuning	0.595	0.619	0.672	0.595	0.345	Fine Tuning	0.464	0.480	0.495	0.457	0.475	Fine Tuning	0.361	0.388	0.446	0.336	0.042
GPT-3.5-turbo	size-based one-class-shot	0.654	0.696	0.697	0.621	0.408	sim-based one-class-shot	0.526	0.567	0.553	0.528	0.564	sim-based few-shot	0.486	0.543	0.621	0.447	0.081
Gemini-pro 1.0	size-based one shot	0.602*	0.609*	0.601*	0.562*	0.407*	sim-based one shot	0.592*	0.476*	0.607*	0.460*	0.554*	rand-based few shot	0.475*	0.526*	0.609*	0.455*	0.100*

Table 1: Macro results of precision, recall, accuracy, f1-score, and also the hate speech class f1-score for each model in its best configuration. Gemini* results may slightly fluctuate due to the rate of responses blocked by Google API filters. This rate was 0.15%, 0.79% and 0.12% for each dataset, respectively. Best results in **bold**.

representation or mode size. This way, we aim to evaluate how such extremes affect inference.

4 Experiments and Results

This section details the implementation process and presents the results obtained.

4.1 Experimental Setup

Models Setup All the fine-tuned models utilized an early stopping criterion for epoch selection and a batch size of 16. The encoder-based models had a learning rate of $2e - 5$. The 7B models were fine-tuned using LoRA (Hu et al., 2022) strategy, with $r = 16$, $lora\_alpha = 32$, and a learning rate of $1e-4$. Finally, the prompt-activated models had $temperature = 0.1$ and the $max\_token = 20$.

Datasets Three datasets with hate content were used for evaluation. HateBR (Vargas et al., 2022), which includes comments gathered from the Instagram accounts of Brazilian politicians; OLID-BR (Trajano et al., 2023), featuring tweets and YouTube comments in Portuguese; and ToLD-BR (Leite et al., 2020), which consists of a collection of Brazilian tweets. All datasets were divided into 60% for training, 20% for validation, and 20% for testing. Dataset preprocessing involves anonymizing users with the @USER token, replacing URLs with HTTPURL token, and converting emojis into text.

4.2 Results

To address space limitations, we only included the best results for each model in Table 1, based on the F1-score for hate speech, which we conjecture is the most critical class. Encoder-based models, especially the one pre-trained on social media and Portuguese (*i.e.*, BERTweet.BR), were the top performers in most datasets, suggesting that pre-training *corpus* is a crucial aspect.

Despite having more parameters, decoder-based 7-billion models were less successful than encoder-based models. This hints at a possible gap in their training on hateful content. Furthermore, the demonstration selection strategy for large models activated by prompts demonstrates potential. The GPT and Gemini models achieved most of their best results when selections were based on size or semantic similarity. They even outperformed models specifically adapted for Portuguese and further fine-tuned, Sabiá and Gervásio.

Overall, our findings emphasize the superiority of encoder-based models for this task. While generative models have shown potential, especially those trained on vast and diverse datasets, the targeted nature of encoder language models pre-trained on specific domains (*e.g.*, social media and Portuguese) and adjusted explicitly for the task appears to be a critical feature for identifying hate speech.

5 Conclusions

We examined eight models with varying features derived from the Transformer architecture for hate speech detection task. Our findings indicate that despite the advanced abilities of generative LLMs, small models still play a crucial role in preventing the perpetuation of social issues in NLP tools. Additionally, aspects related to pre-training (*e.g.*, the ethical filters, the nature and the language of the training *corpora*) may be correlated with better outcomes, more than the size of the models in this case. Our results also illustrate AI limitations for this critical domain. Therefore, we emphasize that these models should serve as aids in moderation, but not as complete substitutes for it.

Limitations

This study faces a limitation regarding the division of its training, validation and testing sets, as it employs only a single split. This constraint primarily stems from the significant costs of utilizing the GPT API, Gemini API, and computational resources on the Google Cloud environment. Furthermore, this limitation also restricted the variation of hyperparameters for our models, such as adjusting the number of epochs or modifying the learning parameters. While these factors may affect the interpretation of the models' behavior in broader scenarios, those decisions enabled the analysis and comparison of various approaches across models, each with unique characteristics.

Acknowledgements

This research was financed by CNPq (National Council for Scientific and Technological Development), grant 307088/2023-5, FAPERJ - *Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro*, process SEI-260003/000614/2023 and SEI-260003/002930/2024, and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001. Additionally, this content was developed with the support of the Google Cloud Research Credits program, under the award GCP19980904.

References

Sai Saketh Aluru, Binny Mathew, Punyajoy Saha, and Animesh Mukherjee. 2020. Deep Learning Models for Multilingual Hate Speech Detection. *CoRR*, abs/2004.06465.

Gabriel Assis, Annie Amorim, Jonnathan Carvalho, Daniel de Oliveira, Daniela Vianna, and Aline Paes. 2024. Exploring Portuguese Hate Speech Detection in Low-Resource Settings: Lightly Tuning Encoder Models or In-context Learning of Large Models? In *Proceedings of the 16th International Conference on Computational Processing of Portuguese*, pages 301–311, Santiago de Compostela, Galicia/Spain. Association for Computational Lingustics.

Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. In *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc.

Fernando Pereira Carneiro. 2023. BERTweet.BR: A Pre-Trained Language Model for Tweets in Portuguese. Master's thesis, Universidade Federal Fluminense, Programa de Pós-Graduação em Computação, Niterói.

Ke-Li Chiu, Annie Collins, and Rohan Alexander. 2022. Detecting Hate Speech with GPT-3.

Rodolfo Costa Cezar da Silva and Thierson Couto Rosa. 2023. Combining Data Transformation and Classification Approaches for Hate Speech Detection: A Comparative Study. Available at SSRN.

Alexandra DeLucia, Shijie Wu, Aaron Mueller, Carlos Aguirre, Philip Resnik, and Mark Dredze. 2022. Bernice: A Multilingual Pre-trained Encoder for Twitter. In *Proceedings of the 2022 Conference on Empirical Methods in Natural Language Processing*, pages 6191–6205, Abu Dhabi, United Arab Emirates. Association for Computational Linguistics.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2019. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Paula Fortuna and Sérgio Nunes. 2018. A Survey on Automatic Detection of Hate Speech in Text. *ACM Comput. Surv.*, 51(4).

Team Google. 2023. Gemini: A Family of Highly Capable Multimodal Models. *CoRR*, abs/2312.11805.

Edward J Hu, yelong shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2022. LoRA: Low-Rank Adaptation of Large Language Models. In *International Conference on Learning Representations*.

Md Saroar Jahan and Mourad Oussalah. 2023. A systematic review of hate speech automatic detection using natural language processing. *Neurocomputing*, 546:126232.

Hyewon Jang, Qi Yu, and Diego Frassinelli. 2023. Figurative Language Processing: A Linguistically Informed Feature Analysis of the Behavior of Language Models and Humans. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 9816–9832, Toronto, Canada. Association for Computational Linguistics.

João Augusto Leite, Diego Silva, Kalina Bontcheva, and Carolina Scarton. 2020. Toxic Language Detection in Social Media for Brazilian Portuguese: New Dataset and Multilingual Analysis. In *Proceedings of the 1st Conference of the Asia-Pacific Chapter of the Association for Computational Linguistics and the 10th International Joint Conference on Natural Language Processing*, pages 914–924, Suzhou, China. Association for Computational Linguistics.

Zongxi Li, Xianming Li, Yuzhang Liu, Haoran Xie, Jing Li, Fu Lee Wang, Qing Li, and Xiaoqin Zhong. 2023. Label Supervised LLaMA Finetuning. *CoRR*, abs/2310.01208.

Thanh Thi Nguyen, Campbell Wilson, and Janis Dalins. 2023. Fine-Tuning Llama 2 Large Language Models for Detecting Online Sexual Predatory Chats and Abusive Texts. *CoRR*, abs/2308.14683.

Amanda Oliveira, Thiago de Carvalho Cecote, João Paulo Reis Alvarenga, Vander Luis de Souza Freitas, and Eduardo José da Silva Luz. 2024. Toxic Speech Detection in Portuguese: A Comparative Study of Large Language Models. In *Proceedings of the 16th International Conference on Computational Processing of Portuguese*, pages 108–116, Santiago de Compostela, Galicia/Spain. Association for Computational Lingustics.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul F Christiano, Jan Leike, and Ryan Lowe. 2022. Training language models to follow instructions with human feedback. In *Advances in Neural Information Processing Systems*, volume 35, pages 27730–27744. Curran Associates, Inc.

Rogers Pelle, Cleber Alcântara, and Viviane P. Moreira. 2018. A Classifier Ensemble for Offensive Text Detection. In *Proceedings of the 24th Brazilian Symposium on Multimedia and the Web*, WebMedia '18, page 237–243, New York, NY, USA. Association for Computing Machinery.

Ramon Pires, Hugo Queiroz Abonizio, Thales Sales Almeida, and Rodrigo Frassetto Nogueira. 2023. Sabiá: Portuguese Large Language Models. In *Intelligent Systems - 12th Brazilian Conference, BRACIS 2023, Belo Horizonte, Brazil, September 25-29, 2023, Proceedings, Part III*, volume 14197 of *Lecture Notes in Computer Science*, pages 226–240. Springer.

João Rodrigues, Luís Gomes, João Silva, António Branco, Rodrigo Santos, Henrique Lopes Cardoso, and Tomás Osório. 2023. Advancing Neural Encoding of Portuguese with Transformer albertina pt-*.

Raquel Bento Santos, Bernardo Cunha Matos, Paula Carvalho, Fernando Batista, and Ricardo Ribeiro. 2022. Semi-Supervised Annotation of Portuguese Hate Speech Across Social Media Domains. In *11th Symposium on Languages, Applications and Technologies (SLATE 2022)*, volume 104 of *Open Access Series in Informatics (OASIcs)*, pages 11:1–11:14, Dagstuhl, Germany. Schloss Dagstuhl – Leibniz-Zentrum für Informatik.

Rodrigo Santos, João Silva, Luís Gomes, João Rodrigues, and António Branco. 2024. Advancing Generative AI for Portuguese with Open Decoder Gervásio PT-*.

Fábio Souza, Rodrigo Nogueira, and Roberto Lotufo. 2020. BERTimbau: Pretrained BERT Models for Brazilian Portuguese. In *Intelligent Systems*, pages 403–417, Cham. Springer International Publishing.

Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurélien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023a. LLaMA: Open and Efficient Foundation Language Models. *CoRR*, abs/2302.13971.

Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti Bhosale, Dan Bikel, Lukas Blecher, Cristian Canton-Ferrer, Moya Chen, Guillem Cucurull, David Esiobu, Jude Fernandes, Jeremy Fu, Wenyin Fu, Brian Fuller, Cynthia Gao, Vedanuj Goswami, Naman Goyal, Anthony Hartshorn, Saghar Hosseini, Rui Hou, Hakan Inan, Marcin Kardas, Viktor Kerkez, Madian Khabsa, Isabel Kloumann, Artem Korenev, Punit Singh Koura, Marie-Anne Lachaux, Thibaut Lavril, Jenya Lee, Diana Liskovich, Yinghai Lu, Yuning Mao, Xavier Martinet, Todor Mihaylov, Pushkar Mishra, Igor Molybog, Yixin Nie, Andrew Poulton, Jeremy Reizenstein, Rashi Rungta, Kalyan Saladi, Alan Schelten, Ruan Silva, Eric Michael Smith, Ranjan Subramanian, Xiaoqing Ellen Tan, Binh Tang, Ross Taylor, Adina Williams, Jian Xiang Kuan, Puxin Xu, Zheng Yan, Iliyan Zarov, Yuchen Zhang, Angela Fan, Melanie Kambadur, Sharan Narang, Aurélien Rodriguez, Robert Stojnic, Sergey Edunov, and Thomas Scialom. 2023b. Llama 2: Open Foundation and Fine-Tuned Chat Models. *CoRR*, abs/2307.09288.

Douglas Trajano, Rafael H. Bordini, and Renata Vieira. 2023. OLID-BR: offensive language identification dataset for Brazilian Portuguese. *Language Resources and Evaluation*.

Francielle Vargas, Isabelle Carvalho, Fabiana Rodrigues de Góes, Thiago Pardo, and Fabrício Benevenuto. 2022. HateBR: A Large Expert Annotated Corpus of Brazilian Instagram Comments for Offensive Language and Hate Speech Detection. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 7174–7183, Marseille, France. European Language Resources Association.

Francielle Vargas, Fabiana Rodrigues de Góes, Isabelle Carvalho, Fabrício Benevenuto, and Thiago Pardo. 2021. Contextual-Lexicon Approach for Abusive

Language Detection. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2021)*, pages 1438–1447, Held Online. INCOMA Ltd.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Ł ukasz Kaiser, and Illia Polosukhin. 2017. Attention is All You Need. In *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Jules White, Quchen Fu, Sam Hays, Michael Sandborn, Carlos Olea, Henry Gilbert, Ashraf Elnashar, Jesse Spencer-Smith, and Douglas C. Schmidt. 2023. A Prompt Pattern Catalog to Enhance Prompt Engineering with ChatGPT. *CoRR*, abs/2302.11382.