Optimizing Resume Data Extraction with Small Language Models: A Comparative Study on Efficiency and Privacy

Wagner Costa Santos[1,2]**, Gustavo Tomaello Juiz**[1]**, Arnaldo Candido Junior**[1]

[1]Institute of Biosciences, Humanities and Exact Sciences
São Paulo State University (UNESP)
São José do Rio Preto, SP – Brazil

[2]Compleo Enterprise – ATS Software (Emphasys Group)
São Paulo, SP – Brazil

{wagner.costa,gustavo.tomaello,arnaldo.candidato}@unesp.br

Abstract. *Resume data extraction remains a critical challenge in automated recruitment systems, where traditional approaches struggle with unstructured document formats while facing significant computational and privacy constraints. Large Language Models (LLMs) have shown promise for information extraction tasks but present substantial barriers including high computational costs and data privacy concerns when processing sensitive personal information. This study investigates the fine-tuning of Small Language Models (SLMs) as a viable alternative, specifically examining Phi-3 mini for extracting structured data from resumes. We employ advanced fine-tuning techniques including Parameter Efficient Fine-Tuning (PEFT) and Reparameterized Fine-Tuning with LoRA Quantization Adaptation (QLoRA) on a synthetic dataset of 2,100 resumes generated using privacy-preserving methods. Our comparative evaluation against GPT-4o demonstrates that the fine-tuned SLM achieves superior performance with 87.9% accuracy and 91.5 BLEU score, compared to GPT-4o's 85.5% accuracy and 83.9 BLEU score, while significantly reducing computational costs. These findings establish SLMs as scalable and privacy-conscious alternatives for applications requiring nuanced information processing, offering comparable accuracy to LLMs with enhanced efficiency and data protection.*

1. Introduction

Extracting relevant information from unstructured documents, such as resumes, remains a persistent challenge in natural language processing. Large Language Models (LLMs), exemplified by OpenAI's GPT-3.5 [Brown et al. 2020] and GPT-4 [OpenAI et al. 2024], have demonstrated remarkable capabilities in comprehending and parsing complex data. LLMs are characterized by their vast size, typically containing billions of parameters, which enables them to capture and generate human-like text with high coherence and contextual understanding. However, the utilization of LLMs is often hindered by high computational costs and privacy concerns, particularly when dealing with sensitive personal information.

In contrast, Small Language Models (SLMs) have emerged as a promising alternative to address these limitations. SLMs are compact models that contain significantly fewer parameters compared to LLMs. Despite their reduced size, SLMs can still achieve impressive performance on various NLP tasks. Models like Phi-3 mini [Abdin et al. 2024] a 3.8 billion parameter model, developed by Microsoft, showcase the potential of SLMs to deliver state-of-the-art results while maintaining a smaller model size and reduced computational requirements. This makes SLMs an attractive option for applications that prioritize efficiency and privacy.

In this study, we investigate the potential of fine-tuning SLMs, specifically Phi-3 mini, for extracting information from resumes. We aim to assess the viability of SLMs as a cost-effective and privacy-aware solution, benchmarking their performance against established LLMs. To ensure the privacy of individuals, we create a synthetic dataset using real resume templates populated with generated data using the Python Faker library [Faraglia and Other Contributors 2025]. This approach allows for controlled experimentation while mitigating privacy risks associated with using real personal data.

Our research focuses on fine-tuning Phi-3 using Supervised Fine-Tuning (SFT) and advanced techniques such as Parameter Efficient Fine-Tuning (PEFT) [Houlsby et al. 2019] and Reparameterized Fine-Tuning with Lora Quantization Adaptation (QLoRA)[Dettmers et al. 2023]. By leveraging these methods, we aim to enhance the model's ability to accurately extract and structure relevant information from resumes into a standardized JSON format. This structured output facilitates the evaluation of the model's performance and its potential for real-world applications.

Through this comparative analysis, we seek to contribute to the ongoing discourse in the field of NLP by exploring the advantages and limitations of SLMs in the context of information extraction. By demonstrating the effectiveness and efficiency of fine-tuned SLMs, we aim to provide valuable insights into their potential as a scalable and privacy-conscious alternative to LLMs for a wide range of applications that require nuanced information processing.

2. Related Work

2.1. General Techniques for Resume Data Extraction

The integration of advanced Natural Language Processing (NLP) and machine learning techniques in the automation of resume parsing has become increasingly prevalent, serving key roles in recruitment and talent management. Rawat et al. (2021) [Rawat et al. 2021] review the evolution of resume parsing techniques in the HR recruitment process, highlighting the transition from traditional manual methods to advanced machine learning (ML) and deep learning (DL) approaches. The development of NLP and Named Entity Recognition (NER) has significantly enhanced the efficiency and accuracy of extracting structured information from unstructured resumes. The article underscores the implementation of various frameworks and algorithms, such as hybrid cascaded models, ontology-based systems, and neural networks, which have collectively driven rapid advancements in automated resume parsing technology. Other research contributions include the work of Kulkarni (2024) [Kulkarni 2024], which incorporates NLP and rule-based methods to segment resumes by their layout, and Rasal (2023) [Rasal 2023], who advocates for the integration of NLP techniques into resume parser projects to effectively analyze resume content.

2.2. Data Extraction in Large Language Models (LLMs)

The use of large language models (LLMs) in resume parsing presents challenges related to computational efficiency and data privacy concerns. Recent advancements in instruction tuning and model pretraining, as discussed by Zhou (2023) in LIMA [Zhou et al. 2023], suggest that a significant portion of a model's knowledge can be acquired during pre-training with minimal additional data for fine-tuning. This insight supports our research into leveraging smaller models for similar tasks. Moreover, Tingyu Xie et al. (2023) [Xie et al. 2023] explore the performance of LLMs in zero-shot information extraction tasks, such as named entity recognition (NER), using models like ChatGPT. Their findings underscore the potential of LLMs in extracting structured information from unstructured text, aligning with our focus on evaluating the capabilities of small language models in resume data extraction.

This structured approach ensures that our research contributions are clearly contextualized within the existing literature, addressing both the capabilities and limitations of current NLP applications in resume parsing. By focusing on SLMs, our study proposes a scalable, efficient, and privacy-conscious alternative that could enhance information extraction processes in sensitive applications.

3. Method

The creation of a suitable dataset for training and evaluating language models applied to resume information extraction presents considerable challenges, primarily due to the sensitive personal data typically found in such documents. To mitigate privacy and ethical risks, we devised a rigorous procedure for generating a synthetic dataset, as illustrated in Figure 1 and detailed as follows:

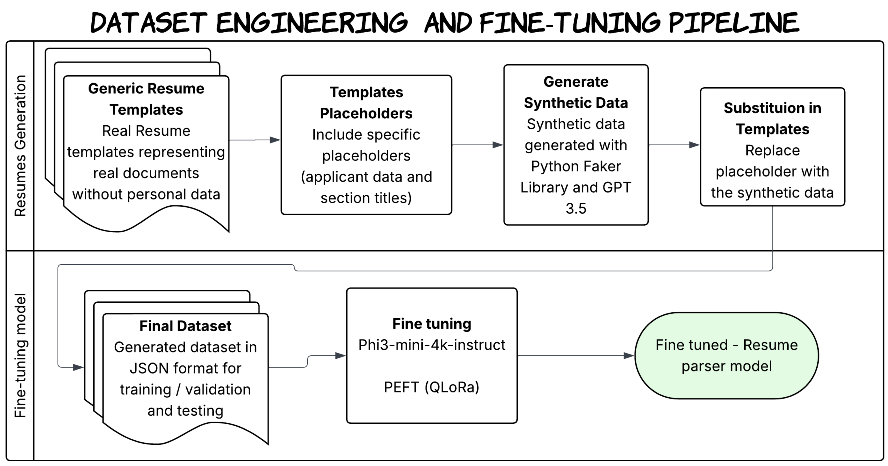

Figure 1. Dataset engineering and Fine-tuning pipeline

3.1. Dataset

1. **Use of Generic Resume Templates:** We began the process by utilizing 12 resume templates, including real resumes with common layouts found in Applicant

Tracking System (ATS) software and auto-generated CVs from platforms such as LinkedIn (with all personal data removed). These templates were converted into text format and adapted to include specific placeholders for data such as Name, Email, and Professional History (company, date, activity description). After conversion, the templates produced distinct text patterns depending on their original source format; for example, resumes originally containing tables could present information in varying orders, such as the candidate's name and position appearing mid-document instead of at the top.

2. **Generation of Synthetic Data:** With the placeholders set, we proceeded to generate synthetic data using primarily deterministic and rule-based approaches. The Python Faker library was used to generate the majority of fields including names, emails, addresses, phone numbers, dates, and company names. For complex narrative content, such as job descriptions, skill summaries, and professional experience descriptions, which require contextual coherence, we employed the `gpt-3.5-turbo-0125` Large Language Model (LLM). This model was used exclusively for narrative text generation, while all structured fields were generated deterministically to maintain control over the evaluation dataset.
3. **Substitution in Templates:** After generating the synthetic data, we replaced the placeholders in the templates with the generated synthetic data, resulting in a set of 2100 resumes structured in a realistically representative manner. We also randomly assigned different titles to the sections of resumes from a list of common names. This variation, along with the template format, allowed for a greater variety of formats in the dataset.
4. **Label Generation for Fine-tuning:** Using the synthesized dataset, we generated labels in JSON format that represent the expected outcomes for the model's fine-tuning process. These labels correspond primarily to the deterministically generated structured data (names, emails, dates, companies) rather than the LLM-generated narrative content, ensuring that the evaluation targets genuine extraction capabilities rather than content reproduction.
5. **Dataset Validation:** A portion of the dataset was reserved for validation during the model development process, ensuring that the model could generalize beyond the examples seen during training. Specifically, 1900 resumes were used for training, with 5% of this training dataset (100 records) set aside for evaluation.
6. **Accuracy Testing:** A smaller fraction of the dataset, specifically 100 records, was designated to test the model's accuracy post fine-tuning. We aimed to use some templates that were not included in the training set, allowing a rigorous assessment of the model's effectiveness under conditions similar to real-world usage.

This method of dataset generation not only reinforces privacy and ethical compliance but also provides a robust basis for assessing the effectiveness of small language models in processing and analyzing complex information contained in resumes. While synthetic data ensures privacy and allows controlled experimentation, it may not fully capture the diversity, noise, and inconsistencies present in real-world resumes, which could affect the model's performance in practical deployments. Figure 1 provides an example of our dataset structure.

```
[
    {
        "content": "You are an assistant specialized in extracting data from resumes.\n Input: Sara Silveira \n
        \nAustraliano,70 \nViaduto Rua Gabrielly Novaes,712 □ Vila De Sá □ Santa Maria - DF \nCelular: +55 (45) 9
        0866-5644 / Telefone Residencial: 11 0720-0271\nE-mail: maria-juliasiqueira@example.com\nObjetivo Profissional \n
        \n \nFormação Escolar \n\n\n\uf0b7 \nTécnico em Relações Públicas  - SENAI - Em andamento\n\n\uf0b7 \nTécnico em
        Publicidade  - Inst. Federal de Educação, Ciência e Tecnologia da Paraíba (IFPB) - jan-2000\n\n\uf0b7 \nMestrado
        Gestão de Negócios  - Univ. Federal de Viçosa (UFV) - jan-2000\n \nHistórico de Emprego \n \n\n\uf0b7 \nBrito -
        EI - dez/1993 a Atual\nFunção: Vendedor de Loja\nAtividades Realizadas: - Atendimento ao cliente\n- Organização
        do espaço de vendas\n- Realização de vendas e negociações\n\n\n\uf0b7 \nAlmeida - jan/1990 a jan/2000\nFunção:
        Supervisor de Vendas\nAtividades Realizadas: - Treinamento de equipe de vendas\n- Elaboração de metas e
        estratégias de vendas\n- Monitoramento de desempenho e resultados\n\n\n\uf0b7 \nMoreira - nov/1996 a jan/
        2000\nFunção: Gerente de Loja\nAtividades Realizadas: - Gestão de equipe e escalas de trabalho\n- Controle de
        estoque e pedidos\n- Análise de indicadores de performance\n\n\n \nCursos Complementares \n \n\n\uf0b7 \nGestão
        de Projetos - Workshops Guerra Correia e Filhos, Em curso\n\n\uf0b7 \nAtendimento ao Cliente - Cursos Cassiano,
        abr/2007\n\n\n\nFrancês\n",
        "role": "user"
    },
    {
        "content": {
            "name": "Sara Silveira",
            "email1": "maria-juliasiqueira@example.com",
            "mobile1": "+55 (45) 9 0866-5644",
            "landline1": "11 0720-0271",
            "address": {
                "streetName": "Viaduto Rua Gabrielly Novaes",
                "neighborhood": "Vila De S\\u00e1",
                "city": "Santa Maria",
                "state": "DF",
                "number": "712"
            },
            "educationHistory": [
                {
                    "endDate": "",
                    "finished": false,
                    "courseName": "T\\u00e9cnico em Rela\\u00e7\\u00f5es P\\u00fablicas",
                    "schoolName": "SENAI"
                },
                {...
```


Figure 2. An example JSON representation of our dataset structure.

3.2. Fine-Tuning Process

We will perform the fine-tuning process using Supervised Fine-Tuning (SFT) and advanced techniques such as Parameter Efficient Fine-Tuning (PEFT) and Reparameterized Fine-Tuning with Lora Quantization Adaptation (QLoRA). These techniques provide an efficient fine-tuning approach that reduces memory usage and computational costs while maintaining high performance. The fine-tuning process will be conducted on the **Microsoft Phi-3-mini-4k-instruct**, a small language model developed by Microsoft with 3.8 billion parameters.

3.3. LLM Baseline

To provide a comparative benchmark for the fine-tuned small language model (SLM), we utilized GPT-4o, a Large Language Model (LLM), as a baseline for evaluating performance. GPT-4o, which incorporates few-shot learning capabilities, was trained on a diverse range of text corpora to enhance its language understanding and information extraction capabilities. By comparing the SLM's performance against the GPT-4o baseline, we aimed to assess the efficiency and privacy benefits of utilizing smaller models for information extraction tasks.

3.4. Comparison with Non-Fine-Tuned Model

In addition to benchmarking against GPT-4o, we will also compare our fine-tuned small language model (SLM) with its non-fine-tuned version. This comparison aims to provide further insights into the impact of the fine-tuning process on the model's performance.

By evaluating the results from the non-fine-tuned model, we can better understand the enhancements brought about by fine-tuning in terms of language understanding and information extraction capabilities.

3.5. Results Analysis

In assessing the fine-tuned Phi-3 model, we focus on the accuracy of individual elements such as names, emails, and educational backgrounds within the JSON outputs. This evaluation uses exact matches to determine the model's accuracy at a per-field level. The overall performance metric consolidates these findings to offer a comprehensive view of the model's efficacy across all fields. Additionally, we will use the BLEU score [Papineni et al. 2001] as supplementary information to compare the JSON outputs, acknowledging that while BLEU is typically used for translations, it can be useful for comparing the similarity of structured data. The ground truth for this evaluation is the synthetic JSON used to generate the resume templates. The metrics obtained for the fine-tuned Phi-3, Phi-3, and GPT-4 models will be compared to each other.

4. Results

The main results of our comparative analysis are presented in Table 1 and Figure 2, highlighting the performance metrics for the fine-tuned Phi-3 model, GPT-4o, and the non-fine-tuned Phi-3 model. The metrics include the number of exact matches, total keys expected, total keys predicted, BLEU score, and accuracy.

Model	Accuracy	Bleu Score	Keys Matches	Keys (Total)
Fine-Tuned Phi-3	0.879	91.506	3089	3514
GPT-4o	0.855	83.924	3005	3514
Phi-3	0.545	65.426	1916	3514

Table 1. Comparison of exact matches, total keys expected, total keys predicted, and similarity for different models, with an indicator for the best performance.

We also evaluated the results using the 2 templates that were not used in the training data. These templates were only used in the test data to better assess the model's generalization capability. The results for this case are presented in Table 2 and Figure 3.

Model	Accuracy	Bleu Score	Keys Matches	Keys (Total)
Fine-Tuned Phi-3	0.71	82.1	811	1143
GPT-4o	0.79	80.5	903	1143
Phi-3	0.537	68.6	614	1143

Table 2. Comparison of exact matches, total keys expected, total keys predicted, and similarity for different models, with an indicator for the best performance, for specific templates.

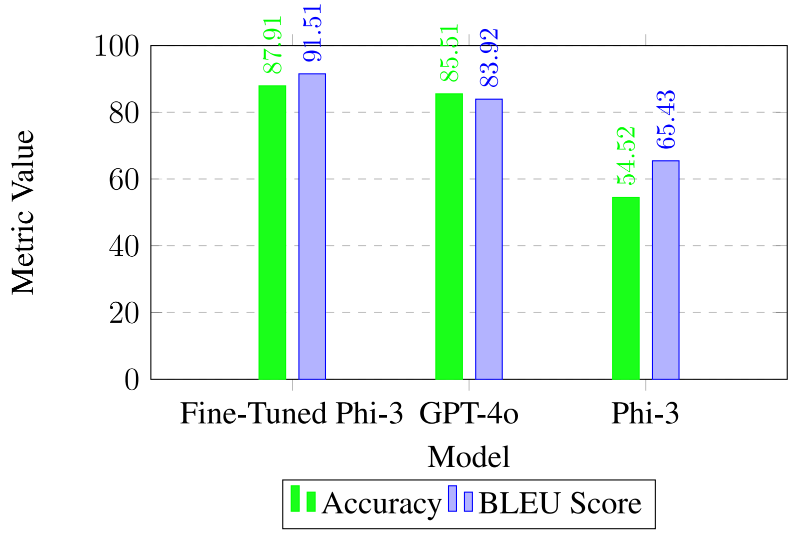

Figure 3. Comparison of Accuracy and BLEU Score for Different Models

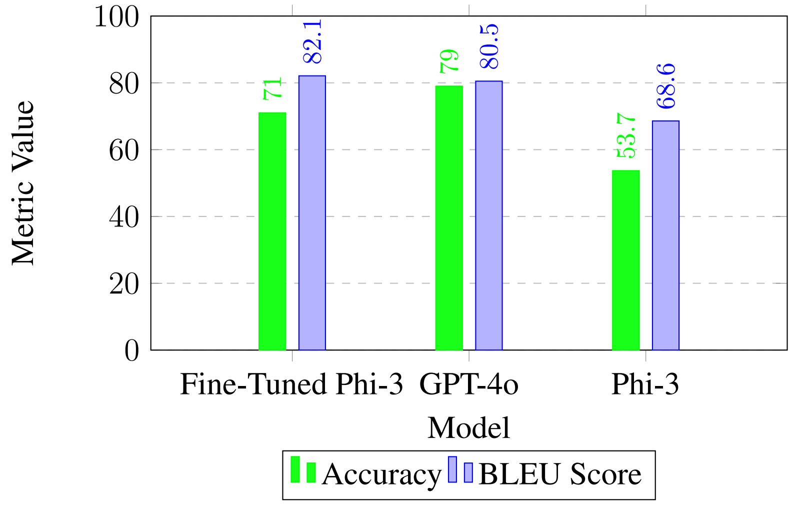

Figure 4. Comparison of Accuracy and BLEU Score for Different Models on Specific Templates

5. Conclusion

This study aimed to evaluate the effectiveness of Small Language Models (SLMs), specifically Phi-3, in extracting data from resumes and compared their performance with Large Language Models (LLMs) such as GPT-4o. Through comprehensive experimentation and fine-tuning techniques, we demonstrated that SLMs could achieve comparable accuracy to LLMs while significantly reducing inference costs and enhancing data privacy.

The primary contributions of this research include:

Efficiency and Privacy of SLMs: We showed that SLMs, despite their smaller size, could perform complex NLP tasks effectively. Fine-tuned Phi-3 achieved an accuracy of 87.9% and a BLEU score of 91.5, outperforming the baseline LLM, GPT-4o. Even with unseen templates, the fine-tuned Phi-3 achieved an accuracy of 71% and a BLEU score of 82.1%, compared to GPT-4o's 79% and 80.5%, respectively. This performance was significantly better than the non-fine-tuned Phi-3, which scored 53.7% and 68.6%, demonstrating the superior generalization capability of the fine-tuned model.

Fine-Tuning Techniques: The study leveraged advanced fine-tuning techniques such as the Supervised Fine-tuning Trainer (SFT Trainer), PEFT with QLoRa. These methods proved effective in enhancing the model's ability to extract and structure relevant information from resumes.

Synthetic Dataset Creation: We developed a synthetic dataset using real resume templates populated with generated data, ensuring privacy while providing a robust basis for model training and evaluation. This dataset was crucial for controlled experimentation and accurate performance assessment.

Benchmarking Against LLMs: By benchmarking the fine-tuned Phi-3 against GPT-4o and its non-fine-tuned version, we provided valuable insights into the trade-offs between model size, performance, and computational cost. The results suggest that SLMs can be a viable alternative to LLMs for applications requiring nuanced information processing, particularly where efficiency and privacy are paramount. The fine-tuned Phi-3 model demonstrated strong performance metrics, making it a compelling choice for real-world applications in resume data extraction.

Future work could explore further optimization of SLMs, the integration of multi-modal data, and the application of these models to other domains requiring sensitive data processing. Additionally, investigating the combination of SLMs with LLMs in hybrid approaches may yield even better performance and efficiency.

In conclusion, this study highlights the potential of SLMs as scalable, efficient, and privacy-conscious alternatives to LLMs for resume data extraction and possibly other similar NLP tasks. By continuing to refine these models and techniques, we can develop more robust, cost-effective, and secure solutions for diverse applications in natural language processing.

It is important to note that the narrative fields in our synthetic dataset were generated using the `gpt-3.5-turbo-0125` model. While this step was limited to non-structural content and aimed to reduce privacy risks, it may still introduce certain limitations, such as reduced linguistic variety, stylistic artifacts from the specific LLM, or subtle biases inherent to its training data. These factors could influence the diversity and representativeness of the dataset. Although our methodology and controlled setup sought to minimize these effects, by using deterministic generation for most fields and evaluating all models on the same dataset, they cannot be completely eliminated and should be considered when interpreting the results.

References

Abdin, M., Aneja, J., Awadalla, H., Awadallah, A., Awan, A. A., Bach, N., Bahree, A., Bakhtiari, A., Bao, J., Behl, H., Benhaim, A., Bilenko, M., Bjorck, J., Bubeck, S., Cai, M., Cai, Q., Chaudhary, V., Chen, D., Chen, D., Chen, W., Chen, Y.-C., Chen, Y.-L., Cheng, H., Chopra, P., Dai, X., Dixon, M., Eldan, R., Fragoso, V., Gao, J., Gao, M., Gao, M., Garg, A., Giorno, A. D., Goswami, A., Gunasekar, S., Haider, E., Hao, J., Hewett, R. J., Hu, W., Huynh, J., Iter, D., Jacobs, S. A., Javaheripi, M., Jin, X., Karampatziakis, N., Kauffmann, P., Khademi, M., Kim, D., Kim, Y. J., Kurilenko, L., Lee, J. R., Lee, Y. T., Li, Y., Li, Y., Liang, C., Liden, L., Lin, X., Lin, Z., Liu, C., Liu, L., Liu, M., Liu, W., Liu, X., Luo, C., Madan, P., Mahmoudzadeh, A., Majercak, D., Mazzola, M., Mendes, C. C. T., Mitra, A., Modi, H., Nguyen, A., Norick, B., Patra, B., Perez-Becker, D., Portet, T., Pryzant, R., Qin, H., Radmilac, M., Ren, L., Rosa, G. d., Rosset, C., Roy, S., Ruwase, O., Saarikivi, O., Saied, A., Salim, A., Santacroce, M., Shah, S., Shang, N., Sharma, H., Shen, Y., Shukla, S., Song, X., Tanaka, M., Tupini, A., Vaddamanu, P., Wang, C., Wang, G., Wang, L., Wang, S., Wang, X., Wang, Y., Ward, R., Wen, W., Witte, P., Wu, H., Wu, X., Wyatt, M., Xiao, B., Xu, C., Xu, J., Xu, W., Xue, J., Yadav, S., Yang, F., Yang, J., Yang, Y., Yang, Z., Yu, D., Yuan, L., Zhang, C., Zhang, C., Zhang, J., Zhang, L. L., Zhang, Y., Zhang, Y., Zhang, Y., and Zhou, X. (2024). Phi-3 Technical Report: A Highly Capable Language Model Locally on Your Phone. arXiv:2404.14219 [cs].

Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., Agarwal, S., Herbert-Voss, A., Krueger, G., Henighan, T., Child, R., Ramesh, A., Ziegler, D., Wu, J., Winter, C., Hesse, C., Chen, M., Sigler, E., Litwin, M., Gray, S., Chess, B., Clark, J., Berner, C., McCandlish, S., Radford, A., Sutskever, I., and Amodei, D. (2020). Language Models are Few-Shot Learners. In *Advances in Neural Information Processing Systems*, volume 33, pages 1877–1901. Curran Associates, Inc.

Dettmers, T., Pagnoni, A., Holtzman, A., and Zettlemoyer, L. (2023). QLoRA: Efficient Finetuning of Quantized LLMs. *Advances in Neural Information Processing Systems*, 36:10088–10115.

Faraglia, D. and Other Contributors (2025). Faker. original-date: 2012-11-12T23:00:09Z.

Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., Laroussilhe, Q. D., Gesmundo, A., Attariyan, M., and Gelly, S. (2019). Parameter-Efficient Transfer Learning for NLP. In *Proceedings of the 36th International Conference on Machine Learning*, pages 2790–2799. PMLR. ISSN: 2640-3498.

Kulkarni, A. (2024). Advancing Job Search: A Comprehensive Resume-Based Job Recommendation System Using NLP and Deep Learning Techniques. *International Journal for Research in Applied Science and Engineering Technology*, 12(2):1054–1058.

OpenAI, Achiam, J., Adler, S., Agarwal, S., Ahmad, L., Akkaya, I., Aleman, F. L., Almeida, D., Altenschmidt, J., Altman, S., Anadkat, S., Avila, R., Babuschkin, I., Balaji, S., Balcom, V., Baltescu, P., Bao, H., Bavarian, M., Belgum, J., Bello, I., Berdine, J., Bernadett-Shapiro, G., Berner, C., Bogdonoff, L., Boiko, O., Boyd, M., Brakman, A.-L., Brockman, G., Brooks, T., Brundage, M., Button, K., Cai, T., Campbell, R., Cann, A., Carey, B., Carlson, C., Carmichael, R., Chan, B., Chang, C., Chantzis, F., Chen, D., Chen, S., Chen, R., Chen, J., Chen, M., Chess, B., Cho, C., Chu, C., Chung, H. W., Cummings, D., Currier, J., Dai, Y., Decareaux, C., Degry, T., Deutsch, N., Deville, D., Dhar, A., Dohan, D., Dowling, S., Dunning, S., Ecoffet, A., Eleti, A., Eloundou, T., Farhi, D., Fedus, L., Felix, N., Fishman, S. P., Forte, J., Fulford, I., Gao, L., Georges, E., Gibson, C., Goel, V., Gogineni, T., Goh, G., Gontijo-Lopes, R., Gordon, J., Grafstein, M., Gray, S., Greene, R., Gross, J., Gu, S. S., Guo, Y., Hallacy, C., Han, J., Harris, J., He, Y., Heaton, M., Heidecke, J., Hesse, C., Hickey, A., Hickey, W., Hoeschele, P., Houghton, B., Hsu, K., Hu, S., Hu, X., Huizinga, J., Jain, S., Jain, S., Jang, J., Jiang, A., Jiang, R., Jin, H., Jin, D., Jomoto, S., Jonn, B., Jun, H., Kaftan, T., Kaiser, , Kamali, A., Kanitscheider, I., Keskar, N. S., Khan, T., Kilpatrick, L., Kim, J. W., Kim, C., Kim, Y., Kirchner, J. H., Kiros, J., Knight, M., Kokotajlo, D., Kondraciuk, , Kondrich, A., Konstantinidis, A., Kosic, K., Krueger, G., Kuo, V., Lampe, M., Lan, I., Lee, T., Leike, J., Leung, J., Levy, D., Li, C. M., Lim, R., Lin, M., Lin, S., Litwin, M., Lopez, T., Lowe, R., Lue, P., Makanju, A., Malfacini, K., Manning, S., Markov, T., Markovski, Y., Martin, B., Mayer, K., Mayne, A., McGrew, B., McKinney, S. M., McLeavey, C., McMillan, P., McNeil, J., Medina, D., Mehta, A., Menick, J., Metz, L., Mishchenko, A., Mishkin, P., Monaco, V., Morikawa, E., Mossing, D., Mu, T., Murati, M., Murk, O., Mély, D., Nair, A., Nakano, R., Nayak, R., Neelakantan, A., Ngo, R., Noh, H., Ouyang, L., O'Keefe, C., Pachocki, J., Paino, A., Palermo, J., Pantuliano, A., Parascandolo, G., Parish, J., Parparita, E., Passos, A., Pavlov, M., Peng, A., Perelman, A., Peres, F. d. A. B., Petrov, M., Pinto, H. P. d. O., Michael, Pokorny, Pokrass, M., Pong, V. H., Powell, T., Power, A., Power, B., Proehl, E., Puri, R., Radford, A., Rae, J., Ramesh, A., Raymond, C., Real, F., Rimbach, K., Ross, C., Rotsted, B., Roussez, H., Ryder, N., Saltarelli, M., Sanders, T., Santurkar, S., Sastry, G., Schmidt, H., Schnurr, D., Schulman, J., Selsam, D., Sheppard, K., Sherbakov, T., Shieh, J., Shoker, S., Shyam, P., Sidor, S., Sigler, E., Simens, M., Sitkin, J., Slama, K., Sohl, I., Sokolowsky, B., Song, Y., Staudacher, N., Such, F. P., Summers, N., Sutskever, I., Tang, J., Tezak, N., Thompson, M. B., Tillet, P., Tootoonchian, A., Tseng, E., Tuggle, P., Turley, N., Tworek, J., Uribe, J. F. C., Vallone, A., Vijayvergiya, A., Voss, C., Wainwright, C., Wang, J. J., Wang, A., Wang, B., Ward, J., Wei, J., Weinmann, C. J., Welihinda, A., Welinder, P., Weng, J., Weng, L., Wiethoff, M., Willner, D., Winter, C., Wolrich, S., Wong, H., Workman, L., Wu, S., Wu, J., Wu, M., Xiao, K., Xu, T., Yoo, S., Yu, K., Yuan, Q., Zaremba, W., Zellers, R., Zhang, C., Zhang, M., Zhao, S., Zheng, T., Zhuang, J., Zhuk, W., and Zoph, B. (2024). GPT-4 Technical Report. arXiv:2303.08774 [cs].

Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2001). BLEU: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting on Association for Computational Linguistics - ACL '02*, page 311, Philadelphia, Pennsylvania. Association for Computational Linguistics.

Rasal, P. (2023). Resume Parser Analysis Using Machine Learning and Natural Language Processing. *International Journal for Research in Applied Science and Engineering Technology*, 11(5):2840–2844.

Rawat, A., Malik, S., Rawat, S., Kumar, D., and Kumar, P. (2021). A Systematic Literature Review (SLR) On The Beginning of Resume Parsing in HR Recruitment Process & SMART Advancements in Chronological Order. ISSN: 2693-5015.

Xie, T., Li, Q., Zhang, J., Zhang, Y., Liu, Z., and Wang, H. (2023). Empirical Study of Zero-Shot NER with ChatGPT. arXiv:2310.10035 [cs].

Zhou, C., Liu, P., Xu, P., Iyer, S., Sun, J., Mao, Y., Ma, X., Efrat, A., Yu, P., Yu, L., Zhang, S., Ghosh, G., Lewis, M., Zettlemoyer, L., and Levy, O. (2023). LIMA: Less Is More for Alignment. *Advances in Neural Information Processing Systems*, 36:55006–55021.