PENDEKATAN BERKELANJUTAN BERBASIS *PARAMETER-EFFICIENT FINE-TUNING* UNTUK MEMBANGUN SISTEM PERCAKAPAN BERBASIS AI DENGAN DUKUNGAN BAHASA DAERAH PADA *LARGE VISION LANGUAGE MODEL*

SKRIPSI

Diajukan untuk memenuhi salah satu syarat untuk memperoleh gelar Sarjana Teknik pada Program Studi Mekatronika dan Kecerdasan Buatan

Oleh:

Ramadhirra Azzahra Putri

NIM 2100188

PROGRAM STUDI MEKATRONIKA DAN KECERDASAN BUATAN

KAMPUS UPI DI PURWAKARTA

UNIVERSITAS PENDIDIKAN INDONESIA

2025

LEMBAR HAK CIPTA

PENDEKATAN BERKELANJUTAN BERBASIS *PARAMETER-EFFICIENT FINE-TUNING* UNTUK MEMBANGUN SISTEM PERCAKAPAN BERBASIS AI DENGAN DUKUNGAN BAHASA DAERAH PADA *LARGE VISION LANGUAGE MODEL*

Oleh

Ramadhirra Azzahra Putri

Sebuah skripsi yang diajukan untuk memenuhi salah satu syarat memperoleh gelar Sarjana Teknik pada Program Studi Mekatronika dan Kecerdasan Buatan

Universitas Pendidikan Indonesia

Agustus 2025

LEMBAR PENGESAHAN

Ramadhirra Azzahra Putri

2100188

PENDEKATAN BERKELANJUTAN BERBASIS *PARAMETER-EFFICIENT FINE-TUNING* UNTUK MEMBANGUN SISTEM PERCAKAPAN BERBASIS AI DENGAN DUKUNGAN BAHASA DAERAH PADA *LARGE VISION LANGUAGE MODEL*

Disetujui dan disahkan oleh pembimbing,

Pembimbing I

Liptia Venica, S.T., M.T.

NIP. 920210919941203201

Pembimbing II

Dewi Indriati Hadi Putri, S.Pd., M.T.

NIP. 920190219900126201

Mengetahui,

Ketua Program Studi

Mekatronika dan Kecerdasan Buatan,

Dewi Indriati Hadi Putri, S.Pd., M.T.

NIP. 920190219900126201

ABSTRAK

PENDEKATAN BERKELANJUTAN BERBASIS *PARAMETER-EFFICIENT FINE-TUNING* UNTUK MEMBANGUN SISTEM PERCAKAPAN BERBASIS AI DENGAN DUKUNGAN BAHASA DAERAH PADA *LARGE VISION LANGUAGE MODEL*

Ramadhirra Azzahra Putri

2100188

Indonesia memiliki >700 bahasa daerah yang kurang terreprentasi secara digital, sehingga adaptasi *Large Vision Language Models (LVLMs)* perlu pendekatan yang kontekstual dan hemat sumber daya. Studi ini mengadaptasi dua *backbone* berbasis *LLaMa 3.2 11B Vision* dan *Gemma 3 12B PT* untuk bahasa Sunda dan Jawa melalui *Parameter-Efficient Fine-Tuning (PEFT)* skema *QSBoRA-FA* (matriks *A* dibekukan/di-*one-hot*, adaptasi pada *B*), *assistant-only loss*, kuantisasi *4-bit* via *Unsloth*, serta pelacakan emisi dengan *CodeCarbon*. Rangkaian meliputi *continued pretraining* pada *Cendol Collection*, *instruction tuning* pada *Alpaca Sundanese/Javanese (cleaned)* dan korpus *distillation* (≈*7k* per bahasa), lalu *Direct Preference Optimization (DPO)* pada pasangan *chosen/rejected* (≈*1k+1k*). Pada *NusaX-MT*, *LLaMa 3.2 11B Vision* meraih *chrF++* 60,50 (rata-rata 6 arah terjemahan: Sun↔Ind, Jav↔Ind, Sun↔Jav) dan *Gemma 3 12B PT* meraih skor 45,70; catatan bahwa pada skenario *low-resource chrF++* ≥50 lazimnya sudah "sangat baik", sehingga capaian 60,50 menandai kualitas terjemahan yang matang. Pada *NusaX-Senti*, *LLaMa 3.2 11B Vision* mencapai *Weighted-F1-score* 95,42% (akurasi 94,00%), sementara *Gemma 3 12B PT* mencapai *Weighted-F1-score* 89,76% (akurasi 95,00%); *Weighted-F1-score* ditekankan karena ketahanannya terhadap ketidakseimbangan label daripada akurasi yang cepat jenuh. Jejak operasional tercatat ≈21,854 kg CO_2e (energi 48,260 kWh, waktu 88,23 jam); tahap *CP* mendominasi konsumsi, sedangkan *DPO* berskala menit. Angka tersebut sekitar 0,096% dari emisi BLOOM-176B (24,7 ton) dan 0,0047% dari GPT-3 (502 ton), sehingga skema *QSBoRA-FA* (dengan *Unsloth*) menunjukkan adaptasi *LVLM low-carbon* tanpa *trade-off* kualitas yang berarti pada tugas target, sekaligus menawarkan *baseline Green AI* untuk konteks Indonesia.

Kata kunci: *LVLM*, *PEFT, QSBoRA-FA, Green AI*, *CodeCarbon, NusaX-MT, NusaX-Senti, DPO,* bahasa Sunda, bahasa Jawa.

ABSTRACT

A SUSTAINABLE APPROACH BASED ON PARAMETER-EFFICIENT FINE-TUNING FOR BUILDING AN AI-BASED CONVERSATIONAL SYSTEM WITH LOCAL-LANGUAGE SUPPORT ON A LARGE VISION LANGUAGE MODEL

Ramadhirra Azzahra Putri

2100188

Indonesia hosts 700+ local languages that are digitally underrepresented, making Large Vision Language Model (LVLM) adaptation a problem of linguistic context and resource efficiency. We adapt two backbones, LLaMa 3.2 11B Vision and Gemma 3 12B PT, using QSBoRA-FA Parameter-Efficient Fine-Tuning (PEFT) (freezing/one-hotting matrix A, adapting B), assistant-only loss, 4-bit quantization via Unsloth, and CodeCarbon for emissions tracking. The pipeline comprises continued pretraining on Cendol Collection, instruction tuning on cleaned Alpaca Sundanese/Javanese and 7k/language distilled corpora, followed by Direct Preference Optimization (≈1k+1k chosen/rejected pairs). On NusaX-MT, LLaMa 3.2 11B Vision attains chrF++ 60.50 averaged over 6 translation directions (Sun↔Ind, Jav↔Ind, Sun↔Jav) while Gemma 3 12B PT scores 45.70; notably, in low-resource settings chrF++ ≥50 is commonly considered "very good," so 60.50 indicates strong translation quality. On NusaX-Senti, LLaMa 3.2 11B Vision reaches 95.42% Weighted-F1-score (94.00% accuracy) and Gemma 3 12B PT 89.76% Weighted-F1-score (95.00% accuracy); we highlight Weighted-F1-score as IT is more robust to class imbalance than accuracy. The total operational footprint is ≈21.854 kg CO_2e (48.260 kWh, 88.23 h), with CP dominating and DPO being minute-scale; this is ≈0.096% of BLOOM-176B (24.7 t) and ≈0.0047% of GPT-3 (~502 t). Overall, QSBoRA-FA (with Unsloth) delivers low-carbon LVLM adaptation without meaningful quality trade-offs on the target tasks, providing a practical Green AI baseline for the Indonesian context.

Keywords: *LVLM, PEFT, QSBoRA-FA, Green AI, CodeCarbon, NusaX-MT, NusaX-Senti, DPO, Sundanese, Javanese.*

DAFTAR ISI

DAFTAR TABEL

DAFTAR GAMBAR

DAFTAR LAMPIRAN

DAFTAR PUSTAKA

[1]. A. F. Aji, G. I. Winata, F. Koto, S. Cahyawijaya, A. Romadhony, R. Mahendra, K. Kurniawan, D. Moeljadi, R. E. Prasojo, T. Baldwin, J. H. Lau, dan S. Ruder, "One Country, 700+ Languages: NLP Challenges for Underrepresented Languages and Dialects in Indonesia," *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics*, vol. 1: Long Papers, hlm. 7226–7249, 22–27 Mei 2022.

[2]. B. Wilie, K. Vincentio, G. I. Winata, S. Cahyawijaya, X. Li, Z. Y. Lim, S. Soleman, R. Mahendra, P. Fung, S. Bahar, dan A. Purwarianti, "IndoNLU: Benchmark and Resources for Evaluating Indonesian Natural Language Understanding," *arXiv preprint* arXiv:2009.05387v3 [cs.CL], 38 hlm., 8 Okt. 2020.

[3]. S. Cahyawijaya, H. Lovenia, F. Koto, R. Putri, W. Cenggoro, J. Lee, S. Akbar, E. Dave, N. Nuurshadieq, M. Mahendra, Rr Putri, B. Wilie, G. I. Winata, A. F. Aji, A. Purwarianti, dan P. Fung, "Cendol: Open Instruction-tuned Generative Large Language Models for Indonesian Languages," dalam *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, hlm. 14899–14914, Bangkok, Thailand, Ags. 2024. [Daring]. Tersedia: https://aclanthology.org/2024.acl-long.796/ doi: 10.18653/v1/2024.acl-long.796

[4]. L. Owen, V. Tripathi, A. Kumar, dan B. Ahmed, "Komodo-7B: A Linguistic Expedition into Indonesia's Regional Languages," *arXiv preprint* arXiv:2403.09362v2 [cs.CL], 30 hlm., 19 Mar. 2024. [Daring]. Tersedia: https://doi.org/10.48550/arXiv.2403.09362

[5]. T. Dettmers, A. Pagnoni, A. Holtzman, dan L. Zettlemoyer, "QLoRA: Efficient fine-tuning of Quantized LLMs," *arXiv preprint* arXiv:2305.14314, 2023.

[6]. L. Po, Y. Liu, H. Wu, T. Zhang, W. Yu, Z. Wang, Z. Jiang, dan K. Li,

"SBoRA: Low-*rank* adaptation with regional weight updates," *arXiv preprint* arXiv:2407.05413v3, 2024.

[7]. N. Islam dan S. Moushi, "GPT-4o: The cutting-edge advancement in multimodal LLM," *arXiv preprint* arXiv:2401.09210, 2024.

[8]. A. Younesi, M. Ansari, M. A. Fazli, A. Ejlali, M. Shafique, dan J. Henkel, "A Comprehensive Survey of Convolutions in Deep Learning: Applications, Challenges, and Future Trends," *arXiv preprint* arXiv:2402.15490v2 [cs.LG], 28 Feb. 2024.

[9]. Z. Ji, Z. Liu, dan H. Chen, "EMMA-500: Enhancing massively multilingual adaptation of Large Language Models," *Journal of Multilingual Multimodal AI*, 2024.

[10]. A. Myrzakhan, S. M. Bsharat, dan Z. Shen, "Open-LLM-Leaderboard: From Multi-choice to Open-style Questions for LLMs Evaluation, Benchmark, and Arena," *arXiv preprint* arXiv:2406.07545, Jun. 2024. doi: 10.48550/arXiv.2406.07545

[11]. J. Parmar, S. Satheesh, M. Patwary, M. Shoeybi, dan B. Catanzaro, "Reuse, Don't Retrain: A Recipe for Continued Pretraining of Language Models," *arXiv preprint* arXiv:2407.07263, 2024. [Daring]. Tersedia: https://arxiv.org/abs/2407.07263

[12]. S. Chandra, T. Ho, dan C. Wang, "Parameter efficient fine-tuning: A comprehensive analysis across applications," dalam *Proceedings of the IEEE International Conference on Pattern Recognition (ICPR)*, 2024.

[13]. S. Balavadhani, R. Wang, Y. Lin, dan J. Stotzka, "The Ultimate Guide to Fine-tuning LLMs from Basics to Breakthroughs: An Exhaustive Review of Technologies, Research, Best Practices, Applied Research Challenges and Opportunities," *arXiv preprint* arXiv:2408.13296v3, 2024.

[14]. O. Kaynak, "The golden age of Artificial Intelligence," *Discover Artificial Intelligence*, vol. 1, no. 1, hlm. 1–7, 2021.

[15]. E. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, dan W. Chen, “LoRA: Low-*rank* Adaptation of Large Language Models,” *arXiv preprint* arXiv:2106.09685v2, 2021.

[16]. L. Zhang, L. Zhang, S. Shi, X. Chu, dan B. Li, “LoRA-FA: Memory-efficient Low-*rank* Adaptation for Large Language Models Fine-tuning,” *arXiv preprint* arXiv:2308.03303, 2023.

[17]. S.-Y. Liu, C.-Y. Wang, H. Yin, P. Molchanov, Y.-C. F. Wang, K.-T. Cheng, dan M.-H. Chen, “DoRA: Weight-Decomposed Low-*rank* Adaptation,” *arXiv preprint* arXiv:2402.09353v6, 2024.

[18]. T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, dan D. Amodei, “Language Models are Few-shot Learners,” dalam *Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, hlm. 1877–1901, 2020.

[19]. J. Kaplan, S. McCandlish, T. Henighan, T. B. Brown, B. Chess, R. Child, S. Gray, A. Radford, J. Wu, dan D. Amodei, “Scaling Laws for Neural Language Models,” *arXiv preprint* arXiv:2001.08361 [cs.LG], 23 Jan. 2020. doi: 10.48550/arXiv.2001.08361

[20]. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, dan I. Polosukhin, “Attention Is All You Need,” *arXiv preprint* arXiv:1706.03762v7 [cs.CL], 2 Agt. 2023. doi: 10.48550/arXiv.1706.03762

[21]. S. Minaee, T. Mikolov, N. Nikzad, M. Chenaghlu, R. Socher, X. Amatriain, dan J. Gao, “Large Language Models: A Survey,” *arXiv preprint* arXiv:2402.06196v3 [cs.CL], 23 Mar. 2025.

[22]. Llama Team, “The Llama 3 Herd of Models,” 2024.

[23]. J. Chen, Z. Chen, J. Wang, K. Zhou, Y. Zhu, J. Jiang, Y. Min, W. X. Zhao, Z. Dou, J. Mao, Y. Lin, R. Song, J. Xu, X. Chen, R. Yan, Z. Wei, D. Hu, W. Huang, dan J.-R. Wen, “Towards Effective and Efficient Continual

Pre-training of Large Language Models," *arXiv preprint* arXiv:2407.18743 [cs.CL], 2024. doi: 10.48550/arXiv.2407.18743

[24]. R. Rafailov, A. Sharma, E. Mitchell, S. Ermon, C. D. Manning, dan C. Finn, "Direct Preference Optimization: Your Language Model Is Secretly a Reward Model," *arXiv preprint* arXiv:2305.18290, 2024.

[25]. A. Manevich dan R. Tsarfaty, "Mitigating hallucinations in large vision-language models (LVLMs) via language-contrastive decoding (LCD)," *arXiv* 2408.04664 [cs.CL], Ags. 2024. doi: 10.48550/arXiv.2408.04664.

[26]. Y. Chen, K. Sikka, M. Cogswell, H. Ji, dan A. Divakaran, "DRESS: Instructing large vision-language models to align and interact with humans via natural language feedback," dalam *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 17–21 Jun. 2024, hlm. 14239–14250. doi: 10.1109/CVPR52733.2024.01404

[27]. S. O'Brien dan M. Lewis, "Contrastive decoding improves reasoning in Large Language Models," *arXiv preprint* arXiv:2309.09117, 2023.

[28]. Y. Guan, J. Liu, Z. Dou, Y. Zhang, S. Bai, W. Wei, H. Wu, dan H. Wang, "HallusionBench: An advanced diagnostic suite for entangled language hallucination and visual illusion in large vision-language models," *arXiv preprint* arXiv:2312.11513, 2023.

[29]. Meta AI, "Llama 3.2 *11B* Vision Architecture Diagram," *GitHub*, 2024. [Daring]. Tersedia: https://github.com/meta-Llama/Llama-models/blob/main/models/Llama3_2/mm-model.png [Diakses: 11-Oktober-2024].

[30]. Z. Han, C. Gao, J. Liu, J. Zhang, dan S. G. Zhang, "Parameter-efficient fine-tuning for Large Models: A Comprehensive Survey," *arXiv preprint* arXiv:2403.14608v7, 2024.

[31]. S. Bhaduri, C. C. S. Balne, A. Mehrotra, A. Dasgupta, dan S. Agrawal,

"Parameter efficient fine tuning: A comprehensive analysis across applications," dalam *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, hlm. 20477–20488, 2024.

[32]. S.-C. Huang, S.-H. Wang, M.-H. Shih, S. Sahay, dan H.-y. Lee, "Systematic Analysis for Pretrained Language Model Priming for Parameter-efficient fine-tuning," *arXiv preprint* arXiv:2212.01032, Mei 2024.

[33]. E. Frantar, S. Ashkboos, T. Hoefler, dan D. Alistarh, "GPTQ: Accurate Post-Training Quantization for Generative Pre-trained Transformers," *arXiv preprint* arXiv:2210.17323v2 [cs.LG], 22 Mar. 2023.

[34]. R. Gong, Y. Ding, Z. Wang, C. Lv, X. Zheng, J. Du, H. Qin, J. Guo, M. Magno, dan X. Liu, "A Survey of Low-*bit* Large Language Models: Basics, Systems, and Algorithms," *arXiv preprint* arXiv:2409.16694v2 [cs.AI], 30 Sep. 2024.

[35]. C. Wang, Z. Wang, X. Xu, Y. Tang, J. Zhou, dan J. Lu, "Q-VLM: Post-training Quantization for Large Vision-Language Models," *arXiv preprint* arXiv:2410.08119v3 [cs.CV], 24 Feb. 2025.

[36]. A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, dan N. Houlsby, "An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale," *arXiv preprint* arXiv:2010.11929v2 [cs.CV], 3 Jun. 2021.

[37]. Gemma Team, *Gemma 2: Improving Open Language Models at a Practical Size*, Google DeepMind, *arXiv preprint* arXiv:2408.00118v3, Okt. 2024. [Daring]. Tersedia: https://arxiv.org/abs/2408.00118

[38]. Unsloth AI, *Dokumentasi Resmi Unsloth*. [Daring]. Tersedia: https://docs.unsloth.ai. [Diakses: 18-April-2025].

[39]. Ç. Yıldız, N. K. Ravichandran, N. Sharma, M. Bethge, dan B. Ermis, “Investigating Continual Pretraining in Large Language Models: Insights and Implications,” *arXiv preprint* arXiv:2402.17400, 2025. [Daring].

[40]. S. Zhang, L. Dong, X. Li, S. Zhang, X. Sun, S. Wang, J. Li, R. Hu, T. Zhang, F. Wu, dan G. Wang, “Instruction Tuning for Large Language Models: A Survey,” *arXiv preprint* arXiv:2308.10792v8, 1 Des. 2024.

[41]. Q. Liu, F. Zhou, Z. Jiang, L. Dou, dan M. Lin, “From Zero to Hero: Examining the Power of Symbolic Tasks in Instruction Tuning,” *arXiv preprint* arXiv:2304.07995v1, 17 Apr. 2023.

[42]. P. Chen, S. Ji, N. Bogoychev, A. Kutuzov, B. Haddow, dan K. Heafield, “Monolingual or Multilingual Instruction Tuning: Which Makes a Better Alpaca,” *arXiv preprint* arXiv:2309.08958v2, 31 Jan. 2024.

[43]. J. Gou, B. Yu, S. J. Maybank, dan D. Tao, “Knowledge Distillation: A Survey,” *International Journal of Computer Vision*, vol. 129, hlm. 1789–1819, 2021, doi: 10.1007/s11263-021-01453-z. [Daring].

[44]. X. Xu, M. Li, C. Tao, T. Shen, R. Cheng, J. Li, C. Xu, D. Tao, dan T. Zhou, “A Survey on Knowledge Distillation of Large Language Models,” *arXiv preprint* arXiv:2402.13116v4 [cs.CL], 21 Okt. 2024.

[45]. Secure and Assured Intelligent Learning (SAIL) Lab, “Javanese Alpaca Cleaned,” *Hugging Face Datasets*, 2024. [Daring]. Tersedia: https://huggingface.co/datasets/saillab/alpaca-javanese-cleaned. [Diakses: 4-April-2025].

[46]. Secure and Assured Intelligent Learning (SAIL) Lab, “Sundanese Alpaca Cleaned,” *Hugging Face Datasets*, 2024. [Daring]. Tersedia: https://huggingface.co/datasets/saillab/alpaca-sundanese-cleaned. [Diakses: 4-April-2025].

[47]. M. Wortsman, G. Ilharco, S. Y. Gadre, R. Roelofs, R. Gontijo-Lopes, A. S. Morcos, H. Namkoong, A. Farhadi, Y. Carmon, S. Kornblith, dan L.

Schmidt, “Model soups: Averaging weights of multiple fine-tuned models improves Accuracy without increasing inference time,” *arXiv preprint* arXiv:2203.05482, versi ke-3, 1 Jul. 2022. [Daring]. Tersedia: https://doi.org/10.48550/arXiv.2203.05482

[48]. A. Chronopoulou, M. E. Peters, A. Fraser, dan J. Dodge, “AdapterSoup: Weight Averaging to Improve Generalization of Pretrained Language Models,” *arXiv preprint* arXiv:2302.07027, Mar. 2023. [Daring]. Tersedia: https://doi.org/10.48550/arXiv.2302.07027

[49]. M. Bagane, “Google Colab: The Free Cloud Platform Powering Machine Learning,” *International Journal of Science, Engineering and Technology*, vol. 12, no. 1, 2024. [Daring].

[50]. W. Vallejo, C. Díaz-Uribe, dan C. Fajardo, “Google Colab and Virtual Simulations: Practical e-Learning Tools to Support the Teaching of Thermodynamics and to Introduce Coding to Students,” *ACS Omega*, vol. 7, no. 8, hlm. 7421–7429, Mar. 2022, doi: 10.1021/acsomega.2c00362

[51]. J. Jones, W. Jiang, N. Synovic, G. K. Thiruvathukal, dan J. C. Davis, “What do we know about Hugging Face? A systematic literature review and quantitative validation of qualitative claims,” *Proceedings of the 18th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement (ESEM)*, 2024. [Daring].

[52]. C. Osborne, J. Ding, dan H. R. Kirk, “The AI Community Building the Future? A Quantitative Analysis of Development Activity on Hugging Face Hub,” *arXiv preprint* arXiv:2405.13058, 2024. [Daring].

[53]. J. Castaño, S. Martínez-Fernández, X. Franch, dan J. Bogner, “Analyzing the Evolution and Maintenance of ML Models on Hugging Face,” *Proceedings of the 21st IEEE/ACM International Conference on Mining Software Repositories (MSR)*, 2024. [Daring].

[54]. *CodeCarbon Documentation, Methodology*, 2020. [Daring]. Tersedia:

https://mlco2.github.io/CodeCarbon/methodology.html [Diakses: 05-Januari-2024].

[55]. L. B. Heguerte, A. Bugeau, dan L. Lannelongue, "How to estimate carbon footprint when training deep learning models? A guide and review," *Environmental Research Communications*, vol. 5, no. 11, 2023. [Daring]. Tersedia: https://doi.org/10.48550/arXiv.2306.08323

[56]. T. Zhong, Z. Yang, Z. Liu, R. Zhang, Y. Liu, H. Sun, Y. Pan, Y. Li, Y. Zhou, H. Jiang, J. Chen, dan T. Liu, "Opportunities and Challenges of Large Language Models for Low-resource Languages in Humanities Research," *arXiv preprint* arXiv:2412.04497v2 [cs.CL], 9 Des. 2024, doi: 10.48550/arXiv.2412.04497.

[57]. A. Singh, N. P. Patel, A. Ehtesham, S. Kumar, dan T. T. Khoei, "A Survey of Sustainability in Large Language Models: Applications, Economics, and Challenges," *arXiv* 2412.04782v2 [cs.AI], revisi terakhir 18 Jan. 2025. [Online].

[58]. L. Wang, S. Chen, L. Jiang, S. Pan, R. Cai, S. Yang, dan F. Yang, "Parameter-efficient fine-tuning in Large Language Models: a survey of methodologies," *Artificial Intelligence Review*, vol. 58, art. no. 227, 3 Mei 2025. [Online].

[59]. G. I. Winata, A. F. Aji, S. Cahyawijaya, R. Mahendra, F. Koto, A. Romadhony, K. Kurniawan, D. Moeljadi, R. E. Prasojo, P. Fung, T. Baldwin, J. H. Lau, R. Sennrich, dan S. Ruder, "NusaX: Multilingual Parallel Sentiment *Dataset* for 10 Indonesian Local Languages," dalam *Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*, hlm. 815–834, 2–6 Mei 2023.

[60]. S. Joshi, "Evaluation of Large Language Models: Review of Metrics, Applications, and Methodologies," *Preprints*, 2025. doi:10.20944/preprints202504.0369.v1.

[61]. NLLB Team, “Scaling neural machine translation to 200 languages,” *Nature*, vol. 630, hlm. 841–846, Jun. 2024, doi:10.1038/s41586-024-07335-x.

[62]. A. Mukherjee dan M. Shrivastava, “chrF-S: Semantics Is All You Need,” dalam *Proceedings of the Ninth Conference on Machine Translation*, Miami, FL, USA, 15–16 Nov. 2024, hlm. 470–474.

[63]. M. C. Hinojosa Lee, J. Braet, dan J. Springael, “Performance Metrics for Multilabel Emotion Classification: Comparing Micro, Macro, and Weighted F1-Scores,” *Applied Sciences*, vol. 14, no. 21, p. 9863, Nov. 2024, doi:10.3390/app14219863.

[64]. T. Hu dan X.-H. Zhou, “Unveiling LLM Evaluation Focused on Metrics: Challenges and Solutions,” *arXiv preprint* arXiv:2404.09135 [cs.CL], Apr. 2024, doi:10.48550/arXiv.2404.09135.

[65]. P. Vickers, L. Barrault, E. Monti, dan N. Aletras, “We Need to Talk About Classification Evaluation Metrics in NLP,” *arXiv preprint* arXiv:2401.03831 [cs.CL], Jan. 2024, doi:10.48550/arXiv.2401.03831.

[66]. I. Bojic, J. Chen, S. Y. Chang, Q. C. Ong, S. Joty, dan J. Car, “Hierarchical Evaluation Framework: Best Practices for Human Evaluation,” *arXiv preprint* arXiv:2310.01917v2 [cs.CL], 2023, doi:10.48550/arXiv.2310.01917.

[67]. R. Likert, “A Technique for the Measurement of Attitudes,” *Archives of Psychology*, no. 140, 1932.

[68]. L. Chen, J. Li, X. Dong, P. Zhang, Y. Zang, Z. Chen, H. Duan, J. Wang, Y. Qiao, D. Lin, dan F. Zhao, “Are We on the Right Way for Evaluating Large Vision-Language Models?,” *arXiv preprint* arXiv:2403.20330v2 [cs.CV], 2024.

[69]. H. Qiu, W. Hu, Z.-Y. Dou, dan N. Peng, “VALOR-Eval: Holistic Coverage and Faithfulness Evaluation of Large Vision-Language

Models," *arXiv preprint* arXiv:2404.13874v4 [cs.CL], 2024.

[70]. M. Ohi, M. Kaneko, N. Okazaki, dan N. Inoue, "Multi-modal, Multi-task, Multi-criteria Automatic Evaluation with Vision Language Models," *arXiv preprint* arXiv:2412.14613v2 [cs.CL], 2024–2025.

[71]. John W. Creswell dan J. David Creswell, *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, edisi ke-6, Thousand Oaks, CA: SAGE, 18 Okt. 2022.

[72]. Meredith D. Gall, Joyce P. Gall, dan Walter R. Borg, *Educational Research: An Introduction*, edisi ke-8, Boston: Allyn dan Bacon, 2007.

[73]. S. Wu, O. Irsoy, S. Lu, V. Dabravolski, M. Dredze, S. Gehrmann, P. Kambadur, D. Rosenberg, dan G. Mann, "BloombergGPT: A Large Language Model for Finance," *arXiv preprint* arXiv:2303.17564v3 [cs.LG], 21 Des. 2023. doi:10.48550/arXiv.2303.17564.

[74]. Z. Jiang, H. Lin, Y. Zhong, Q. Huang, Y. Chen, Z. Zhang, Y. Peng, X. Li, C. Xie, S. Nong, Y. Jia, S. He, H. Chen, Z. Bai, Q. Hou, S. Yan, D. Zhou, Y. Sheng, Z. Jiang, H. Xu, H. Wei, Z. Zhang, P. Nie, L. Zou, S. Zhao, L. Xiang, Z. Liu, Z. Li, X. Jia, J. Ye, X. Jin, dan X. Liu, "MegaScale: Scaling Large Language Model Training to More Than 10,000 GPUs," *arXiv preprint* arXiv:2402.15627v1 [cs.LG], 23 Feb. 2024.

[75]. *ISO/IEC/IEEE 15288:2023, Systems and Software Engineering—System Life Cycle Processes*, ISO/IEC/IEEE, 2023.

[76]. *ISO/IEC/IEEE 24748-1:2024, Systems and Software Engineering—Life Cycle Management—Part 1: Guidelines for Life Cycle Management*, ISO/IEC/IEEE, 2024.

[77]. H. van Cuylenburg dan T. Ginige, "Emotion Guru: A smart emotion tracking application with AI conversational agent for exploring and preventing depression," dalam *Proc. 2021 2nd Int. Conf. Ubiquitous Comput. Emerging Technol. (UCET)*, Dhaka, Bangladesh, Nov. 2021,

hlm. 1–6. doi: 10.1109/UCET54125.2021.9674993.

[78]. ISO/IEC/IEEE, *ISO/IEC/IEEE 15288:2023, Systems and Software Engineering, System Life Cycle Processes*, 2023.

[79]. ISO/IEC/IEEE, *ISO/IEC/IEEE 24748-1:2024, Systems and Software Engineering, Life Cycle Management, Part 1: Guidelines for Life Cycle Management*, 2024.

[80]. N. W. Kit, "AI Mood Tracking Application," *Applied Information Technology and Computer Science (AITCS)*, 2024. [Daring]. Tersedia: https://publisher.uthm.edu.my/periodicals/index.php/aitcs/article/view/16162 [Diakses: 11-Mei-2025].

[81]. H. Schuff, L. Vanderlyn, H. Adel, dan N. T. Vu, "How to do human evaluation: A brief introduction to user studies in NLP," *Natural Language Engineering*, vol. 29, no. 5, hlm. 1199–1222, 2023. doi:10.1017/S1351324922000535

[82]. M. Freitag, G. Foster, D. Grangier, V. Ratnakar, Q. Tan, dan W. Macherey, "Experts, Errors, and Context: A Large-Scale Study of Human Evaluation for Machine Translation," *Transactions of the Association for Computational Linguistics*, vol. 9, hlm. 1460–1474, 2021. doi:10.1162/tacl_a_00437

[83]. M. Freitag, R. Rei, N. Mathur, C.-k. Lo, C. Stewart, G. Foster, A. Lavie, O. Bojar, (eds. L. Barrault, O. Bojar, F. Bougares, R. Chatterjee, M. R. Costa-jussà, C. Federmann, M. Fishel, A. Fraser, M. Freitag, Y. Graham, R. Grundkiewicz, P. Guzman, B. Haddow, M. Huck, A. J. Yepes, P. Koehn, T. Kocmi, A. Martins, M. Morishita, dan C. Monz), "Results of the WMT21 Metrics Shared Task: Evaluating Metrics with Expert-based Human Evaluations on TED and News Domain," dalam *Proceedings of the Sixth Conference on Machine Translation*, Online, Nov. 2021, hlm. 733–774. Tersedia: https://aclanthology.org/2021.wmt-1.73/

[84]. A. Jaelani, “The Interchange of Language and Philosophy in Sundanese Speech Level Systems,” *Dialectical Literature and English Journal*, 2024.

[85]. D. Atmawati, “Language Politeness in the Javanese Verb Speech Level,” *Lingua Cultura*, 2021.

[86]. Gemma Team, “Gemma 3 Technical Report,” *arXiv* 2503.19786, 2025.

[87]. T. Hui, dkk., “HFT: Half Fine-tuning for Large Language Models,” dalam *Proceedings of ACL 2025 (Long)*, 2025.

[88]. D. Kalajdzievski, “A *Rank* Stabilization Scaling Factor for Fine-tuning with LoRA,” *arXiv* 2312.03732, 2023.

[89]. M. F. Adilazuarda, M. I. Wijanarko, L. Susanto, K. Nur'aini, D. T. Wijaya, dan A. F. Aji, “NusaAksara: A Multimodal and Multilingual Benchmark for Preserving Indonesian Indigenous Scripts,” dalam *Proceedings of the 63rd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Vienna, Austria, Jul. 2025, hlm. 28371–28401. doi:10.18653/v1/2025.acl-long.1377

[90]. *Google Developers Blog*, “Gemma explained: What’s new in Gemma 3,” 2025.

[91]. A. Köpf, Y. Kilcher, D. von Rütte, S. Anagnostidis, Z.-R. Tam, K. Stevens, A. Barhoum, N. M. Duc, O. Stanley, R. Nagyfi, S. ES, S. Suri, D. Glushkov, A. Dantuluri, A. Maguire, C. Schuhmann, H. Nguyen, dan A. Mattick, “OpenAssistant conversations -- democratizing large language model alignment,” dalam *Proc. Thirty-seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track (NeurIPS)*, 2023. doi: 10.48550/arXiv.2304.07327.

[92]. M. Post, O. Caglayan, M. Popel, dkk., “*chrf.py*—implementasi metrik ChrF/ChrF++,” *SacreBLEU (GitHub repo)*, rilis v2.5.0, 3 Jan. 2025. [Daring]. Tersedia: https://github.com/mjpost/sacrebleu/blob/master/sacrebleu/metrics/chrf.p

y. Diakses: 23 Juli 2025.

[93]. M. Post, O. Caglayan, M. Popel, dkk., "*SacreBLEU 2.x*—dokumentasi & catatan skala 0–100," *GitHub/PyPI*, 2024–2025. [Daring]. Tersedia: https://github.com/mjpost/sacrebleu; https://pypi.org/project/sacrebleu/. Diakses: 25 Juli 2025.

[94]. P. Pakray, S. Pal, A. Vetagiri, R. Krishna, A. K. Maji, S. Dash, L. Laitonjam, L. Sarah, dan R. Manna, "Findings of WMT 2024 Shared Task on Low-Resource Indic Languages Translation," *Proceedings of the Ninth Conference on Machine Translation (WMT 2024)*, Miami, Florida, USA, hlm. 654–668, Nov. 2024. Association for Computational Linguistics.

[95]. HPLT Project (konsorsium), *Translation Models for Select Language Pairs (Deliverable 5.1)*, 29 Feb. 2024. [Daring]. Tersedia: https://hplt-project.org/HPLT_D5_1___Translation_models_for_select_language_pairs.pdf. Diakses: 02 Agustus 2025.

[96]. W. Wongso, D. S. Setiawan, S. Limcorn, dan A. Joyoadikusumo, "NusaBERT: Teaching IndoBERT to be Multilingual and Multicultural," *Proceedings of the Second Workshop in South East Asian Language Processing (SEALP 2025)*, hlm. 10–26, Jan. 2025. Association for Computational Linguistics.

[97]. A. S. Luccioni, S. Viguier, dan A.-L. Ligozat, "Estimating the Carbon Footprint of BLOOM, a 176B Parameter Language Model," *arXiv preprint* arXiv:2211.02001 [cs.LG], Nov. 2022. doi: 10.48550/arXiv.2211.02001.