PhD in Communication Efficient Large Scale Distributed Training

Location: Mila / Concordia University (Montreal, Canada) with visits and stays at Sorbonne University (Paris, France)
Supervisors: Prof. Eugene Belilovsky and Dr. Edouard Oyallon

Training large-scale foundational models—including large language models (LLMs) and image/video diffusion models—demands highly distributed systems and efficient use of computational resources. However, most current distributed training algorithms are designed for homogeneous data centers with high-cost, low-latency interconnects, and suffer from significant communication inefficiencies. This PhD project will focus on designing novel communication-efficient algorithms and training paradigms that enable scalable and robust model training across heterogeneous and lower-cost compute environments.

MS students and postdoc with relevant experience may also be considered

References:

Joseph, C.-É., Thérien, B., Moudgil, A., Knyazev, B., & Belilovsky, E. (2025). Meta-learning Optimizers for Communication-Efficient Learning. *Transactions on Machine Learning Research*. https://openreview.net/forum?id=uRbf9ANAns

Nabli, A., Fournier, L., Erbacher, P., Serrano, L., Belilovsky, E., & Oyallon, E. (2024). ACCO: Accumulate While You Communicate, Hiding Communications in Distributed LLM Training. *arXiv preprint arXiv:2406.02613*. https://arxiv.org/abs/2406.02613

Rivaud, S., Fournier, L., Pumir, T., Belilovsky, E., Eickenberg, M., & Oyallon, E. (2024). PETRA: Parallel End-to-End Training with Reversible Architectures. *International Conference on Learning Representations (ICLR) 2025*. https://openreview.net/forum?id=0fhzSFsGUTarXiv+6OpenReview+6OpenReview+6

Douillard, A., Feng, Q., Rusu, A. A., Chhaparia, R., Donchev, Y., Kuncoro, A., Ranzato, M., Szlam, A., & Shen, J. (2024). DiLoCo: Distributed Low-Communication Training of Language Models. *arXiv preprint arXiv:2311.08105*. https://arxiv.org/abs/2311.08105

Peng, B., Quesnelle, J., & Kingma, D. P. (2024). DeMo: Decoupled Momentum Optimization. *arXiv preprint arXiv:2411.19870*. https://arxiv.org/abs/2411.19870

Nabli, A., Belilovsky, E., & Oyallon, E. (2023). A²CiD²: Accelerating Asynchronous Communication in Decentralized Deep Learning. *Advances in Neural Information Processing Systems*, 36.

Contact: Email CV and informal requests to
eugene.belilovsky@concordia.ca and edouard.oyallon@cnrs.fr