Gabriele Prato

gabriele.prato@mila.quebec
gabprato.github.io
linkedin.com/in/gabprato

SUMMARY

I am a **PhD Candidate** at **Mila**, Université de Montréal, working on **knowledge consolidation** in Large Language Models (**LLMs**). My long-term research goal is to advance current state-of-the-art models towards Artificial General Intelligence (**AGI**). I am looking to collaborate with some of the most talented researchers in the world on this challenge.

EDUCATION

Université de Montréal, Montréal, QC - *PhD in Artificial Intelligence*

September 2019 - September 2024

- Thesis Topic: Knowledge Consolidation in Large Language Models
- GPA: 4.15/4.3
- Advisors: Sarath Chandar, Alain Tapp

PUBLICATIONS

EpiK-Eval: Evaluation for Language Models as Epistemic Models - *First Author*

EMNLP 2023 (Oral)

First study to investigate LMs' capability to combine information seen in different training documents during inference (knowledge consolidation).

PatchBlender: A Motion Prior for Video Transformers - *First Author*

NeurIPS 2022 Workshop

Introduced a learnable blending function that operates over patch embeddings across the temporal dimension of the latent space of Vision Transformers.

Scaling Laws for the Few-Shot Adaptation of Pretrained Image Classifiers - *First Author*

ICML 2021 Workshop

Showed that the few-shot generalization performance of image classifiers is well approximated by power laws as the pre-training set size increases.

Fully Quantized Transformer for Machine Translation - *First Author*

Findings of EMNLP 2020

First paper to show that the entire Transformer neural network could be quantized to 8-bit without impairing performance.

Towards Lossless Encoding of Sentences - *First Author*

ACL 2019

Proposed a near lossless method for encoding long sequences of texts into feature rich representations.

EXPERIENCE

Huawei, Montréal, QC - *Associate Researcher*

January 2019 - December 2019

I was assigned the task of quantizing the Transformer to 8 bits without compromising performance, and I successfully completed it, publishing a paper on the subject.

AWARDS

I received an **Excellence Scholarship** for my Bachelor's Degree in Computer Science at Université de Montréal.

TECHNICAL SKILLS

Programming Languages

- Python
- C
- C++
- Java

AI and Machine Learning Frameworks

- PyTorch
- Huggingface Transformers & Accelerate
- DeepSpeed
- Numpy

Large Scale Training

- I've trained models up to 60 billion parameters on multi-node compute clusters.

Other

- Git/Github
- Docker

LANGUAGES SPOKEN

- English
- French