Proceedings of Machine Learning Research vol 284:1–3, 2025 19th Conference on Neurosymbolic Learning and Reasoning

Neuro-Symbolic Data Collection Automata for Training Language Models on Edge Devices

Jake Ryland Williams JW3477@DREXEL.EDU
Drexel University, Philadelphia, PA 19104

August Lilley AUGUST@THEAUGUSTHAUS.COM
August Haus, Philadelphia, PA 19103

Ankur Mali ANKURARJUNMALI@USF.EDU
University of South Florida, Tampa, FL 33620

Editors: Leilani H. Gilpin, Eleonora Giunchiglia, Pascal Hitzler, and Emile van Krieken

Abstract

Language models (LMs) have achieved significant success in centralized settings, but their utility in localized, real-time applications on edge devices remains constrained. These environments—where direct interaction between users and devices occurs—lack the vast training resources available to general-purpose cloud-based models. The typical development pipeline for LMs involves (1) large-scale unsupervised pretraining to develop generalist behaviors before (2) supervised fine-tuning on small, task-specific datasets. The second step remains a bottleneck for edge deployment, as it requires labeled data, which is rarely available or easily collected *in situ*. We address this challenge by introducing a neuro-symbolic framework for data collection and learning on edge devices. At the core of our approach is a finite-state machine (FSM), called a Data Collection Automaton (DCA), that supervises an LM through interaction with the environment. This FSM enables automatic labeling of user inputs by tracking conversational and physical interactions, transforming them into usable training data. Our implementation focuses on a voice-controlled smart lamp that learns from its user without external data—only through spoken commands and switch toggles. The DCA operates as an FSM: $M = (S, I, \delta, s_0)$ defined by the states $S = \{\text{Empty}, \text{Full}, \text{Lit}, \text{Dark}\}$, input alphabet $I = \{x, +, -, \emptyset, \tau\}$, and transition function $\delta : S \times I \to S$. Text ($x$) is labeled with `<lit>` or `<dar>` depending on whether it is followed by a lamp-On ($+$) or lamp-Off ($-$) interaction. The FSM is a minimal supervision interface that transforms natural user behavior into a training signal.

Training proceeds asynchronously, driven by user interaction. Whenever sufficient data are collected, a new model is trained using a four-stage pipeline: (0) build a byte-pair encoding tokenizer that learns character-merge rules and defines a vocabulary Sennrich et al. (2016), (1) assign vocabulary embeddings via a bit-cipher extending one- to multi-hot encodings for low-dimensional vectors Zhao and Williams (2023), (2) apply a data-dependent weight initialization based on differential analysis of feed-forward layers Williams and Zhao (2023b), (3) optimize with Adam while freezing embeddings, and (4) fine-tune with unfrozen embeddings and a reduced learning rate. This procedure allows for continual model updates with increasing intervals, determined by the rate at which user interaction generates data. Two sizes of precision LM (PLM) were developed around small vocabulary sizes N and embedding and hidden dimensions (d_E and d_h), and based on a decoder-only architecture using self-attention and three distinct token-context types: fixed-length blocks of size b, local neighborhoods of radius r, and the distribution across D documents Williams and Zhao (2023a); whose hidden states (h_b, h_r, and h_D) are combined into a unified representation $h = [h_b, h_r, h_D]$ with hidden dimension is set as: $d_h = 1.25 d_E$, and connected to

an output matrix O that projects to a given vocabulary size: N. Specific model hyperparameters used in this abstract's experiments are: $(b, r, N, d_E, d_h) = (64, 8, 2048, 128, 160)$, and these define a PLM with fewer than one million parameters.

Wav2Vec 2.0 is used for audio transcription Baevski et al. (2020), followed by byte-pair encoding, ensuring robustness to linguistic variability and phonetic ambiguity. Extensive pilot experiments were conducted to evaluate whether PLMs can be trained *from scratch* in real-time through natural lamp usage. Each trial began with a fully untrained system. Users, given only a microphone and button interface, trained the lamp by speaking commands and pressing the lamp switch. No technical expertise was assumed. Over multiple sessions, users successfully trained the lamp to respond to custom commands. In **Fig. 1**, re-labeled and -trained models for six distinct from-scratch training sessions demonstrated average performance improvements in precision and robustness (gray/pink curves), indicating that an optimized configuration for the user interface—with better aligned timing between speeches and switches—would greatly increase the speed at which a command is learned. A live demo was also conducted by an advanced user, and highlights the generalizability and consistency of the presented learning process. This demo is also presented as the red/black curves in **Fig. 1**, and a video the demo's training session is available at https://youtu.be/IxBu7VbeIbI.

In summary, we demonstrate the feasibility of an entirely user-driven, zero-shot learning framework for language models deployed on low-resource devices. By integrating symbolic control, phonetic transcription, and lightweight neural architectures, our approach opens new frontiers in adaptive, private, and context-aware AI. This neuro-symbolic model proves capable of building semantic mappings from unstructured interaction alone—providing a template for autonomous edge learning systems beyond static deployment models.

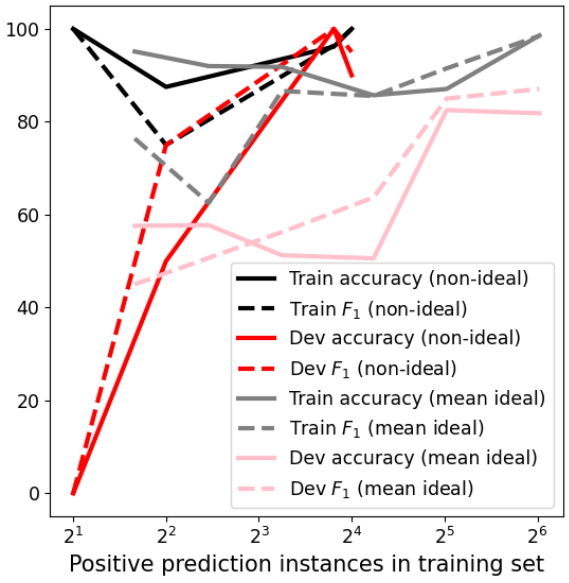

Figure 1: Performance of a live demo (red/black) and the average of 6 sessions (grey/pink).

References

Alexei Baevski, Henry Zhou, Abdelrahman Mohamed, and Michael Auli. wav2vec 2.0: a framework for self-supervised learning of speech representations. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20, Red Hook, NY, USA, 2020. Curran Associates Inc. ISBN 9781713829546.

Rico Sennrich, Barry Haddow, and Alexandra Birch. Neural machine translation of rare words with subword units. In Katrin Erk and Noah A. Smith, editors, *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1715–1725, Berlin, Germany, August 2016. Association for Computational Linguistics. doi: 10.18653/v1/P16-1162. URL https://aclanthology.org/P16-1162/.

Jake Ryland Williams and Haoran Zhao. Explicit foundation model optimization with self-attentive feed-forward neural units, 2023a.

Jake Ryland Williams and Haoran Zhao. Reducing the need for backpropagation and discovering better optima with explicit optimizations of neural networks, 2023b.

Haoran Zhao and Jake Ryland Williams. Bit cipher – a simple yet powerful word representation system that integrates efficiently with language models, 2023.