



Applying the **PICO** Framework for Human-Centered AI: Towards Inclusive and Adaptive AI Through Small Language Models

Suchir Salhan

Department of Computer Science & Technology, University of Cambridge, U.K. sas245@cam.ac.uk

Small Language Models (SLMs): Goals and Rationale

Small Language Models (SLMs), typically under 1B parameters, provide valuable, interpretable, and efficient alternatives to large models. They are especially suitable for proprietary, task-specialised applications such as query routing in chatbot systems or edge/on-device learning.

The **PICO** Framework

A Modular Framework for Hypothesis-Driven Small Language Model Research

The Buttery Group has developed **Pico**, an open-source framework with two libraries for hypothesis-driven pre-training of small language models from scratch [1].

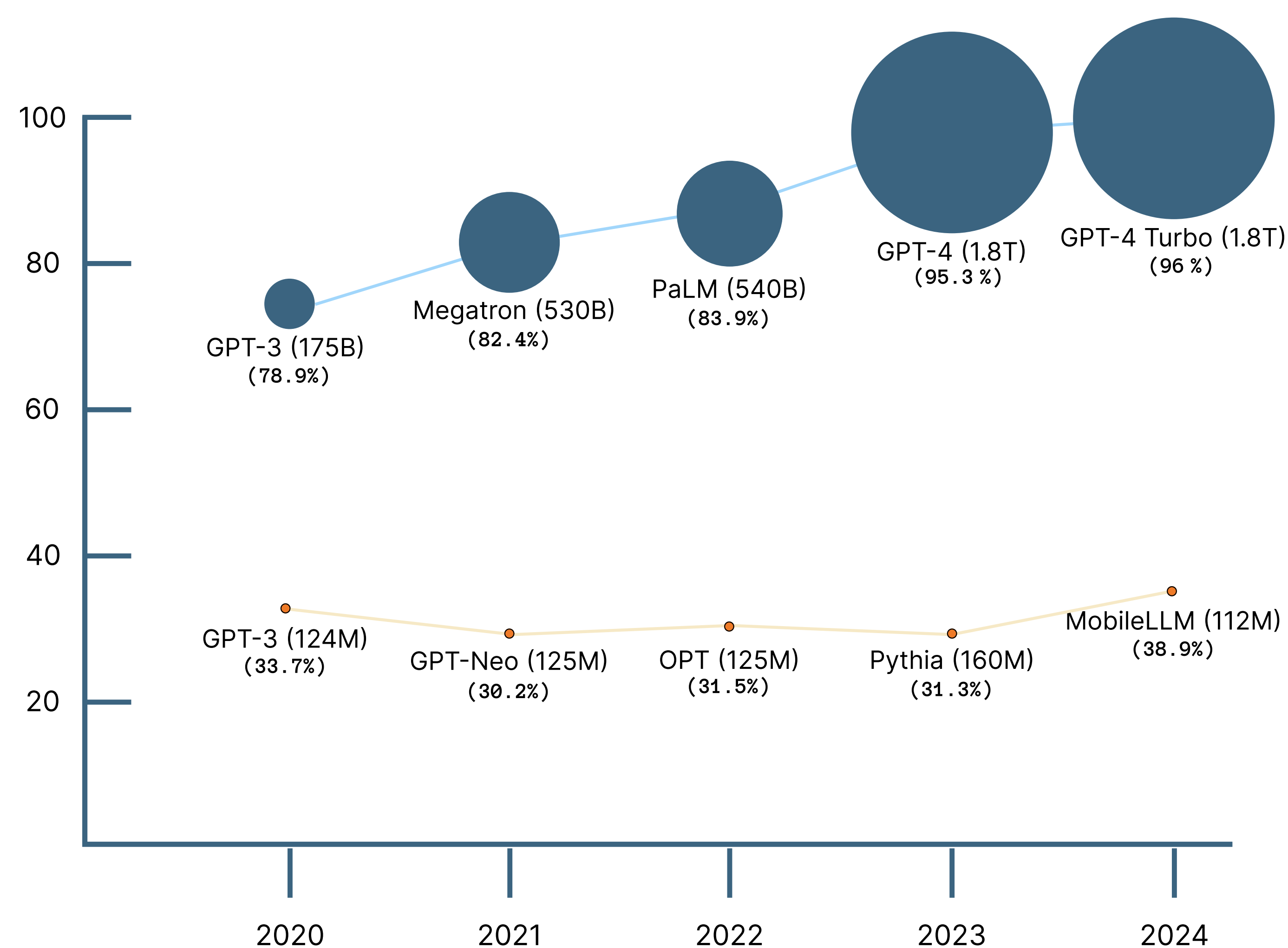


Figure 1: Best performance of fixed-size SLMs and LLMs on MMLU per year, size of the circles represents model size. Note that the size of GPT-4 models is speculative 1 and has not been confirmed by OpenAI. Figure by Richard Diehl Martinez

Pico computes quantitative metrics— such as sparsity, rank, similarity, and norm measures— on different model components throughout small model pretraining, where components refer to specific elements like weight matrices, gradients, or higher-level structures that combine multiple parts (e.g., attention circuits), enabling fine-grained and customizable interpretability analyses. To support reproducible experimentation, we also release a suite of baseline models, pico-decoder, trained under standardized conditions and open-sourced for the community. Case studies highlight how **Pico** can support iterative small LM design and analysis.

My PhD Work: SLMs as *Learner Models*

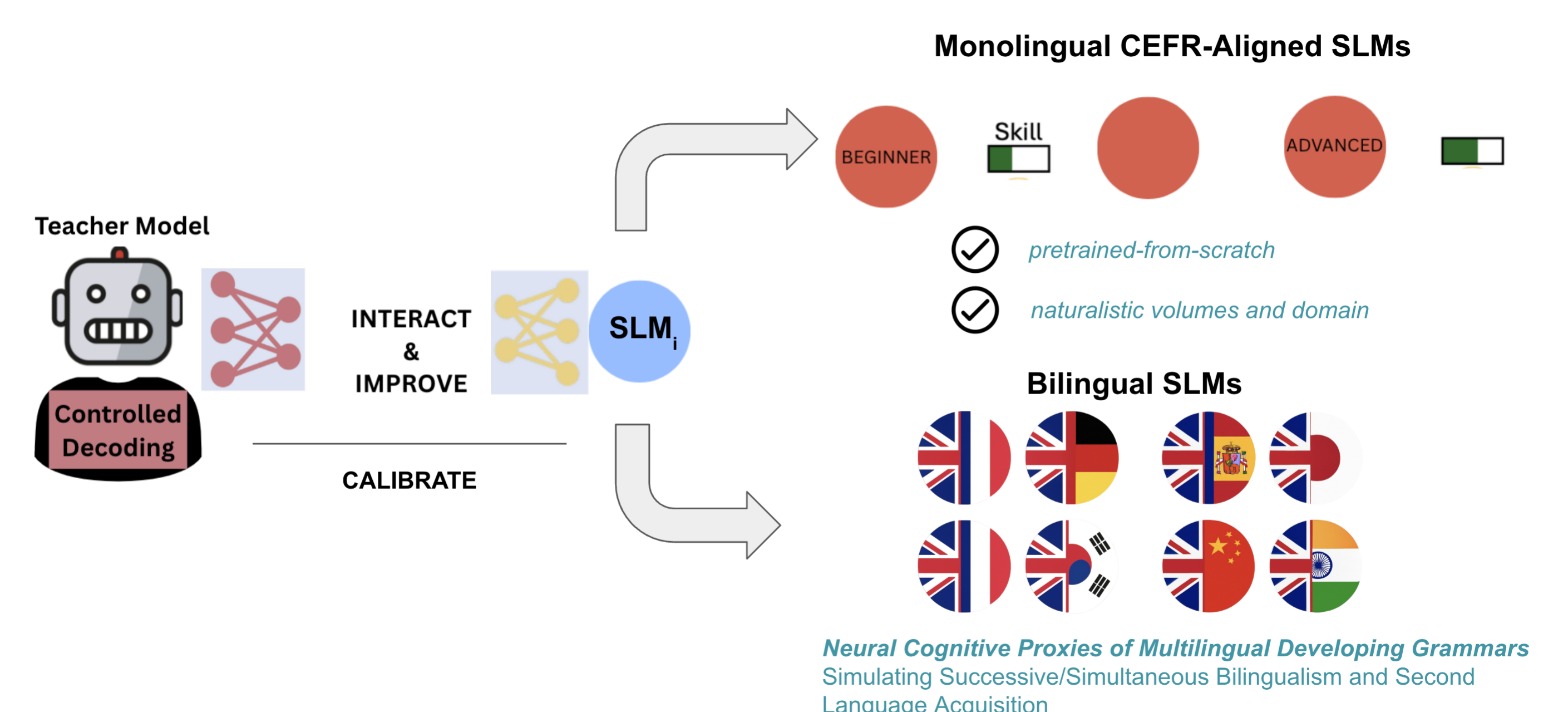
Problem: High-precision learner representations are essential for **aligning LLMs to be more human-centered and inclusive** across different learner capabilities and attributes (e.g., in personalised and adaptive automated learning and assessment).

Solution: I propose bilingual SLMs for second language adaptation — or **L2LMs** — which simulate the developmental trajectories of second-language learners. Model design reflects the *diversity of backgrounds of language learners* (over 55 distinct L1 language families in learner corpora) [2,3].

Aims: Rich Learning Dynamics and Checkpointing of L2LMs for enhanced interpretability, as a **multilingual extension** to the **Pico** Framework [3]. Design prioritises **developmental interpretability** to see exactly where models begin to represent certain linguistic features, how representations stabilise (or oscillate) over training, and which parts of the network converge faster than others to develop more cognitively-aligned pretraining strategies.

Multi-Agent LLM-SLM Ecosystem for Multi-Capability Alignment

Figure 2: Proposed Framework for **LLM-L2LM Alignment** and **LLM Interaction with CEFR-Aligned LLMs**



Selected References

- [1] Diehl Martinez, R., Demitri Africa, D., Weiss, Y., **Salhan, S.A.**, Daniels, R., & Buttery. P.J. (2025) **Pico: A Lightweight Framework for Studying Language Model Learning Dynamics** (*EMNLP Systems Demonstration*) [2] **Salhan, S.A.**, Diehl Martinez, R., Goriely, Z., & Buttery. P.J. (2024) **Less is More: Pre-Training Cross-Lingual Small-Scale Language Models with Cognitively-Plausible Curriculum Learning Strategies** (Accepted Paper @ BabyLM Workshop in EMNLP 2024). [3] 'BLISS: Evaluating Bilingual Learner Competence in Second Language Small Language Models' by Yuan Gao, Suchir Salhan, Andrew Caines, Paula Buttery & Weiwei Sun

Suchir Salhan is supported by Cambridge University Press & Assessment. Pico was supported by a grant from the Accelerate Programme for Scientific Discovery, made possible by a donation from Schmidt Futures.