Human-Like Language Acquisition in Language Models (LMs): a Parent-LM Teaching an Intrinsically-Motivated baby-LM.

Main advisors: Emmanuel Chemla, Yair Lakretz, Pierre-Yves Oudeyer
Internship location: Cognitive Sciences and Psycholinguistics Lab (LSCP), ENS, Paris.

Large Language Models (LLMs) such as GPTs can now achieve human-like performance on a variety of linguistic tasks. However, the amount of data required to train the models has reached massive scales (Fig. 1). For example, Chinchilla sees 1.4 trillion words during training. In contrast to LLMs, children need orders of magnitude less data (Hart B., Risley T., 1995) to achieve similar, and in fact, better performance.

The goal of the proposed internship project is to develop a new approach to train neural language models to make language acquisition in the models more similar to that of children. Specifically, the project aims at: (1) reducing the amount of data required to train a language model, while preserving the performance of the model on linguistic tasks, (2) creating models that show learning trajectories similar to those known in children, passing through similar acquisition stages (Friedmann et al, 2021; Evanson et al., 2023).

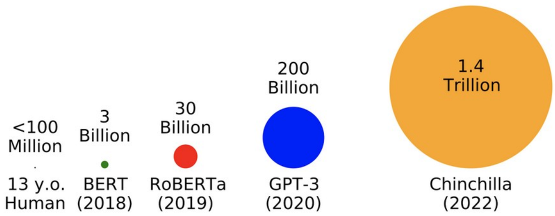


Figure 1: Training-data sizes of LLMs

A central difference between how models and children acquire language is that models are presented with large amounts of text without any prioritization of which types of sentences to present to the models first. But learning may be facilitated if simple sentences are presented in early stages of training. Caregivers, in some cultures, may indeed facilitate communication (and thereby learning) by adjusting their language usage when speaking to a baby or a child. Such *curriculum learning* may also be useful for training models. But how to choose an optimal curriculum learning?

Recent research in AI has shown that AI systems can acquire new skills in a continuous and open-ended manner when they are endowed with an intrinsic motivation (aka, 'curiosity'), which drives the system to find its own curriculum learning (Kaplan & Oudeyer, 2007; Colas et al., 2020). These systems avoid extreme predictable and unpredictable situations in order to focus on the ones that are expected to maximize progress in learning, which leads to a spontaneously emerging curriculum.

The main hypothesis of the proposed research project is therefore that curiosity-driven

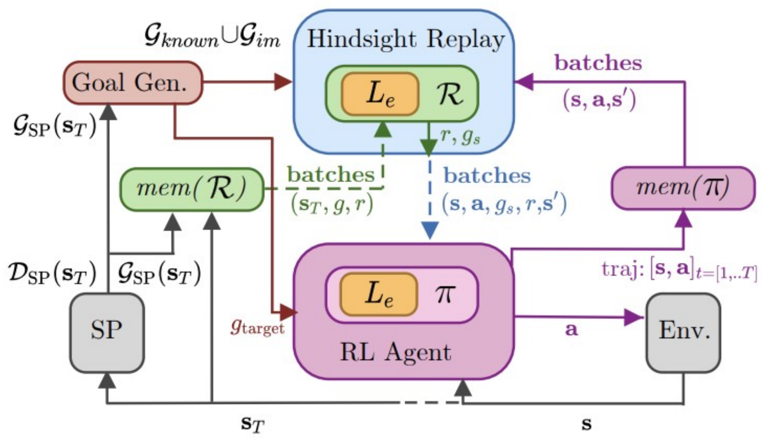


Figure 2: The IMAGINE architecture for two interacting LLMs (Colas et al., 2020)

language models, endowed with intrinsic motivation, will develop auto-generated curriculum learning that will be less data-hungry and resemble learning stages in children during language acquisition.

To test this hypothesis, a pre-trained large language model (the parent-LM) together with a non-trained language model (the baby-LM) will be trained to solve a joint task (Figure 2; from Colas et al., 2020). The task is to explore and act within a controlled environment such as the babyAI (Chevalier-Boisvert et al., 2019; Figure 3), guided by linguistic input from the parent-LM. After learning, the learning trajectories of the baby-LM will be studied and compared to those of children.

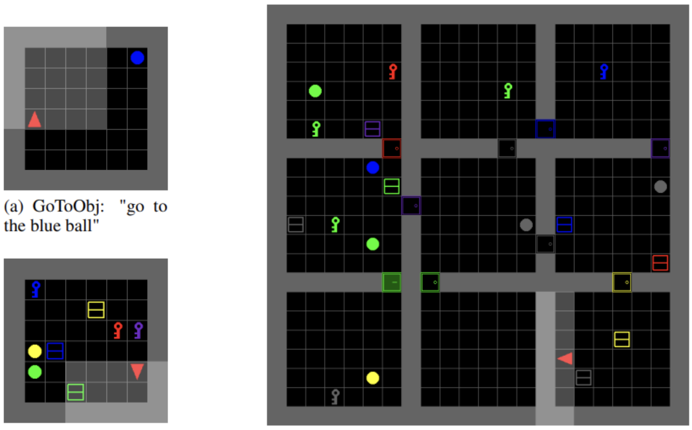
Prerequisites: A good understanding of basic concepts in machine learning, deep learning and reinforcement learning, and prior experience of projects using LLMs. A strong mastery of Python is essential for the project, and it will help to advance fast. Technical familiarity with cloud computing is recommended, and also the ability to understand and use existing code, and to create new ones to design and train and analyze neural networks with, e.g., PyTorch.

Contact us all with a short CV to: yair.lakretz@gmail.com, emmanuel.chemla@ens.psl.eu, pierre-yves.oudeyer@inria.fr

References:

Chevalier-Boisvert, M., Bahdanau, D., Lahlou, S., Willems, L., Saharia, C., Nguyen, T. H., & Bengio, Y. (2018). Babyai: A platform to study the sample efficiency of grounded language learning. arXiv preprint arXiv:1810.08272.

Colas, C., Karch, T., Lair, N., Dussoux, J. M., Moulin-Frier, C., Dominey, P., & Oudeyer, P. Y. (2020). Language as a cognitive tool to imagine goals in curiosity driven exploration. Advances in Neural Information Processing Systems, 33, 3761-3774.



(a) GoToObj: "go to the blue ball"

(b) PutNextLocal: "put the blue key next to the green ball"

(c) BossLevel: "pick up the grey box behind you, then go to the grey key and open a door". Note that the green door near the bottom left needs to be unlocked with a green key, but this is not explicitly stated in the instruction.

Figure 3: Three BabyAI levels built using the MiniGrid environmen (Chevalier-Boisvert et al., 2019)

Evanson, L., Lakretz, Y., & King, J. R. (2023). Language acquisition: do children and language models follow similar learning stages?. arXiv preprint arXiv:2306.03586.

Friedmann, N., Belletti, A., & Rizzi, L. (2021). Growing trees: The acquisition of the left periphery. Glossa: a journal of general linguistics, 6(1).

Kaplan, F., & Oudeyer, P. Y. (2007). In search of the neural circuits of intrinsic motivation. Frontiers in neuroscience, 1, 9.