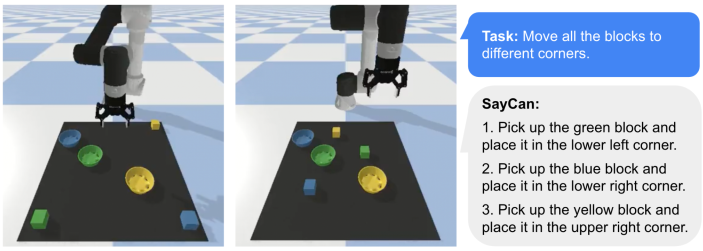
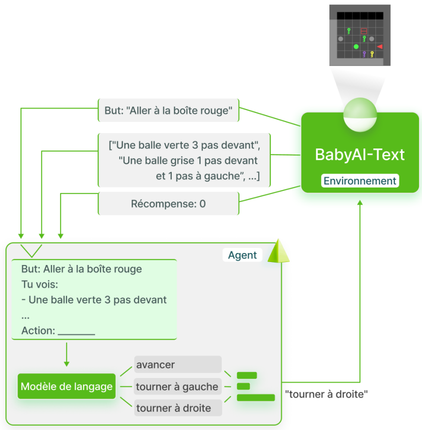
# Flowers team internship proposal, master 2, 2024
# Studying the impact of RL-based grounding on LLMs




**Title:** Studying the impact of RL-based grounding methods on LLMs
**Supervision:** Clément Romac (Inria, Hugging Face), Thomas Carta (Inria), Pierre-Yves Oudeyer (Inria)
**Host team:** Flowers team, Inria Bordeaux
**Duration:** 6 months (march - august 2024)
**How to apply:** contact pierre-yves.oudeyer@inria.fr and clement.romac@inria.fr and thomas.carta@inria.fr with CV and letter of motivation, and **using the [application] tag in the email object.**
**Keywords:** Large Language Models, Deep Learning, Transformers, Reinforcement Learning, Grounding, Natural Language Processing

**TL;DR**
A recent work from the Flowers team proposed to tackle the lack of grounding of LLMs by using them as policies to solve tasks in textual environments and leverage Reinforcement Learning for finetuning. This internship aims to study and lead to a better understanding of what is the broader impact of such an RL-based finetuning on the LLM's knowledge and structure.

**Background:**
The recent rise of Transformer-based Large Language Models (LLMs) trained on massive text datasets has led to models exhibiting impressive capabilities (e.g. natural language generation, question answering, reasoning, translation...) [1][2][3][4][5]. Recently, LLMs were shown to also capture aspects of the physical rules in our world, e.g. about space [6], colors [7] or even affordances between bodies and objects [8].

However, LLMs are known to suffer from a lack of grounding (i.e. connecting their inner representation of language to a world) which prevents them from properly dealing with the meaning of concepts and their direct use to solve tasks in interactive environments [9]. Indeed, alignment between statistical structures in such LLMs and environments can be very limited, or even sometimes entirely wrong (e.g. up to a recent update, ChatGPT would still propose a plan on how to cut a bowl with a knife). This is partly due to 1) a training process (predicting next words) that is not directly incentivized to solve problems in an environment, 2) lack of abilities to intervene in the environment to identify causal structures; 3) lack in abilities to learn based on data collected as a result of interacting with the environment [10].

Focusing on such a functional competence, a recent work from the Flowers team [11] proposed to use an LLM as the policy interacting with a textual environment (i.e. a textual description of the scene is provided by the environment and possible actions are text commands) for decision-making problems. Using Reinforcement Learning (RL) to finetune the LLM to make it solve various tasks in this environment, the proposed GLAM method "functionally grounds"  the LLM on the environment. That is, grounding the dynamics and physical rules of an environment to solve problems and obtain, in the end, an operational LLM able to use natural language to solve tasks in this interactive environment.

Applying GLAM to Flan-T5 780M showed how such an LLM can be functionally grounded and be able to solve the tasks in an interactive textual environment. However, more general questions related to the impact of this RL-based functional grounding on the LLM's knowledge arises. For instance, were the grounded physical rules of the environment deeply encoded in the LLM (e.g. are they exhibited when one asks the LLM about its knowledge)? Moreover, did GLAM lead to side effects also affecting the LLM's ability to answer questions unrelated to the environment the grounding was performed on? Complementary to these empirical tests, there is the question of how such an RL-based finetuning changes the weights and inner structure of the LLM.For this, studying drifts in weights or patterns in gradient updates may help to grasp how GLAM changes the LLM to functionally ground it.

**Project:**
In this project, the intern will attempt to answer the above questions. This will imply:

- Designing a set of tests and experiments assessing GLAM's impact on LLMs. This implies modifying or creating new textual environments specifically designed for this analysis and applying GLAM with an LLM on them.
- Carrying out these experiments with LLMs of various sizes (60M-10B) functionally grounded with GLAM using clusters of GPUs.
- Designing and carrying out an in-depth analysis of how GLAM changes the LLM's weights and inner structure (e.g. computing distances between weights or analysing gradients). This may also imply studying the impact of lightweight recent finetuning methods such as LoRA [12].

**Requirements:**
We are looking for motivated Master 2 (MSc) students with a background in applied math and/or computer science and good command of python and pytorch. Experience with transformers (e.g. through huggingface's transformers library), NLP or (deep) RL/control is a plus.
Additionally the student should have good oral/written communication and reporting skills.

**References**

[1] Brown, Tom B., Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, et al. Language Models are Few-Shot Learners. ArXiv, abs/2005.14165, 2020

[2] Chowdhery, Aakanksha, Sharan Narang, Jacob Devlin, Maarten Bosma, Gaurav Mishra, Adam Roberts, Paul Barham, et al. PaLM: Scaling Language Modeling with Pathways. ArXiv, abs/2204.02311, 2022.

[3] Devlin, Jacob, Ming-Wei Chang, Kenton Lee, et Kristina Toutanova. BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. ArXiv, abs/1810.04805, 2019.

[4] Rae, Jack W., Sebastian Borgeaud, Trevor Cai, Katie Millican, Jordan Hoffmann, Francis Song, John Aslanides, et al. Scaling Language Models: Methods, Analysis & Insights from Training Gopher. ArXiv, abs/2112.11446, 2021.

[5] Scao, Teven, Angela Fan, Christopher Akiki, Ellie Pavlick, Suzana Ilić, Daniel Hesslow, Roman Castagné, et al. BLOOM: A 176B-Parameter Open-Access Multilingual Language Model, ArXiv, abs/2211.05100, 2022.

[6] Patel, Roma, et Ellie Pavlick. Mapping Language Models to Grounded Conceptual Spaces. In International Conference on Learning Representations, 2022.

[7] Mostafa Abdou, Artur Kulmizev, Daniel Hershcovich, Stella Frank, Ellie Pavlick, and Anders Søgaard. Can language models encode perceptual structure without grounding? a case study in color. In Proceedings of the 25th Conference on Computational Natural Language Learning (CoNLL), 2021.

[8] Michael Ahn, Anthony Brohan, Noah Brown, Yevgen Chebotar, Omar Cortes, Byron David, Chelsea Finn, et al. Do as i can, not as i say: Grounding language in robotic affordances. ArXiv, abs/2204.01691, 2022.

[9] Kyle Mahowald, Anna A. Ivanova, Idan Asher Blank, Nancy G. Kanwisher, Joshua B. Tenenbaum, and Evelina Fedorenko. Dissociating language and thought in large language models: a cognitive perspective. ArXiv, abs/2301.06627, 2023.

[10] Emily M. Bender and Alexander Koller. Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, pages 5185–5198, Online, July 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.acl-main.463.

[11] Carta, Thomas, Clément Romac, Thomas Wolf, Sylvain Lamprier, Olivier Sigaud, et Pierre-Yves Oudeyer. Grounding Large Language Models in Interactive Environments with Online Reinforcement Learning. In Proceedings of the 40th International Conference on Machine Learning, 3676‑3713. PMLR, 2023. https://proceedings.mlr.press/v202/carta23a.html.

[12] Hu, Edward J., Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, et Weizhu Chen. LoRA: Low-Rank Adaptation of Large Language Models, 2021. ArXiv, abs/2106.09685.