Grounding LLMs as Autotelic Reinforcement Learning agents

Clément Romac

PhD advisors:
- *Pierre-Yves Oudeyer (Inria)*
- *Thomas Wolf (Hugging Face)*

Jury:
- *Ellie Pavlick (Brown/Deepmind)*
- *Prithviraj Ammanabrolu (UCSD/Nvidia)*
- *Hugo Larochelle (Mila/UdeM)*
- *Timothy Lillicrap (UCL/Deepmind)*
- *Matthieu Cord (Sorbonne/Valeo)*

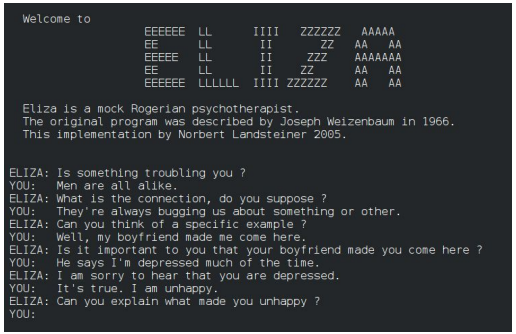
Modeling language understanding

Since the **beginning of artificial intelligence (AI)**, scientists have been trying to build **models of language understanding**.

Modeling language understanding

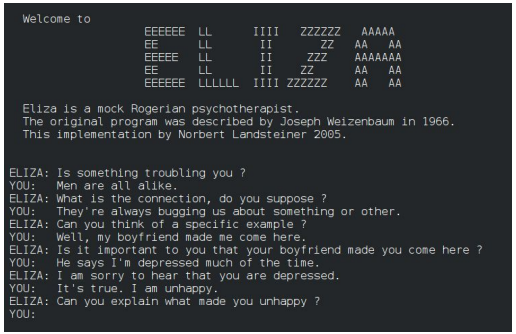
Since the **beginning of artificial intelligence (AI)**, scientists have been trying to build **models of language understanding**.

ELIZA (Weizenbaum, 1966).

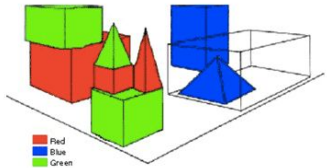


Modeling language understanding

Since the **beginning of artificial intelligence (AI)**, scientists have been trying to build **models of language understanding**.



ELIZA (Weizenbaum, 1966).



Person: Pick up a big red block.
Computer: OK.
Person: Grasp the pyramid.
Computer: I don't understand which pyramid you mean.

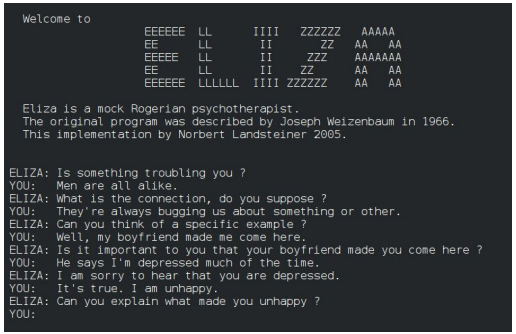
SHRDLU (Winograd, 1971).

Modeling language understanding

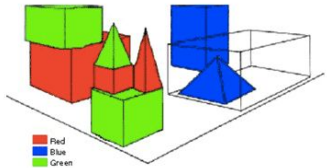
Since the **beginning of artificial intelligence (AI)**, scientists have been trying to build **models of language understanding**.

Measuring a machine's **intelligence** has long been tightly bound to its **ability at understanding natural language**.

Computing Machinery And Intelligence (Turing, 1950).



ELIZA (Weizenbaum, 1966).

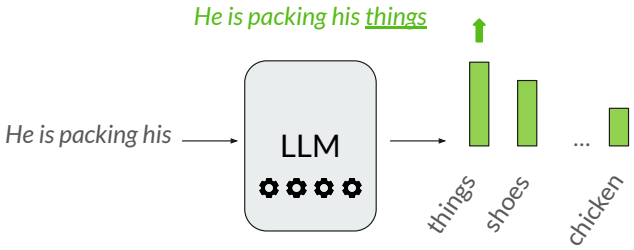


Person: Pick up a big red block.
Computer: OK.
Person: Grasp the pyramid.
Computer: I don't understand which pyramid you mean.

SHRDLU (Winograd, 1971).

Large Language Models (LLMs)

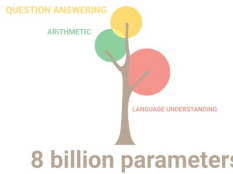
During the last 5 years, we have seen emerge **very large Machine Learning models trained on massive datasets**.



Large Language Models (LLMs)

During the last 5 years, we have seen emerge **very large Machine Learning models trained on massive datasets**.

These models now exhibit **unprecedented** and arguably unexpected **abilities**.

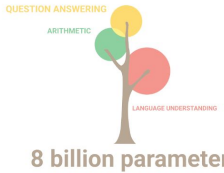


PaLM (Google, 2022).

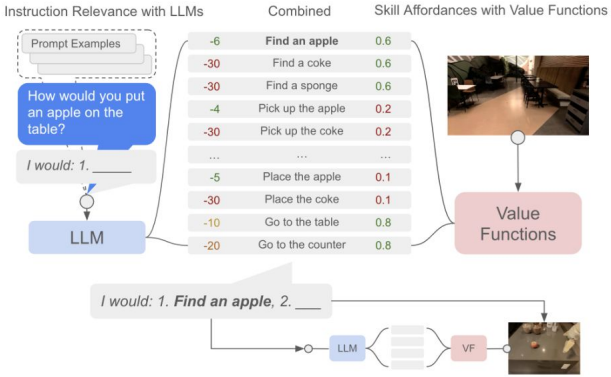
Large Language Models (LLMs)

During the last 5 years, we have seen emerge **very large Machine Learning models trained on massive datasets**.

These models now exhibit **unprecedented** and arguably unexpected **abilities**.



PaLM (Google, 2022).

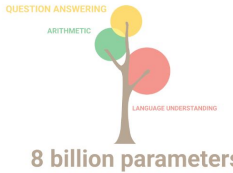


SayCan (Anh et al., 2022).

Large Language Models (LLMs)

During the last 5 years, we have seen emerge **very large Machine Learning models trained on massive datasets**.

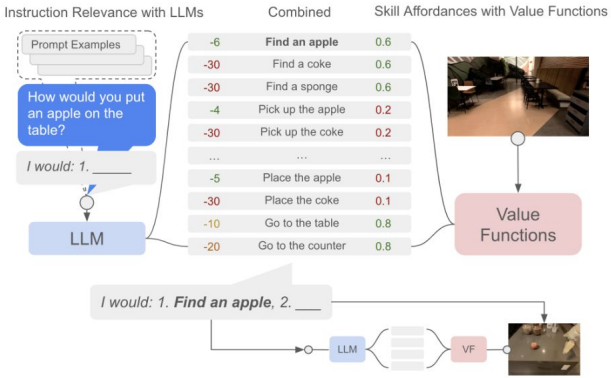
These models now exhibit **unprecedented** and arguably unexpected **abilities**.



PaLM (Google, 2022).

Some of these abilities may be **deceptive**.
(Bender & Coller, 2020; Bisk, 2020; Mahowald et al., 2024).

We still observe limitations:
- handling **physical concepts**
- being **precise** forward models
- …

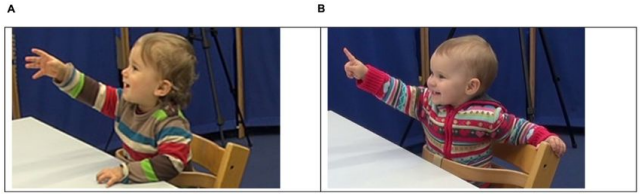


SayCan (Anh et al., 2022).

What studying children has taught us

What studying children has taught us

- Language is acquired through **interactions**:
  - with the **socio-cultural** environment
  - with the **physical** environment

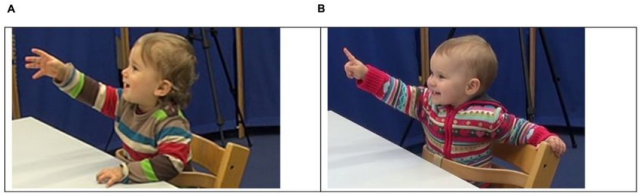


Rohlfing, 2017

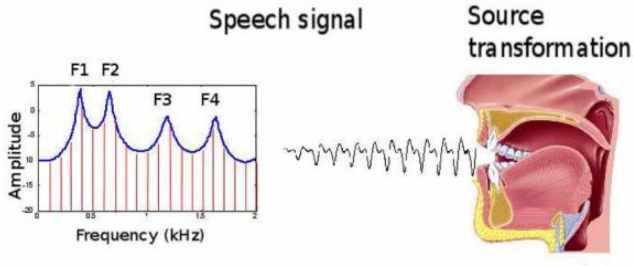
What studying children has taught us

- Language is acquired through **interactions**:
  - with the **socio-cultural** environment
  - with the **physical** environment

- Children are **intrinsically motivated** to learn:
  - To **model and control** their body
  - In **interaction** with their environment
  - In order to **solve** intrinsically and extrinsically defined **problems**



Rohlfing, 2017



Moulin-frier, 2014

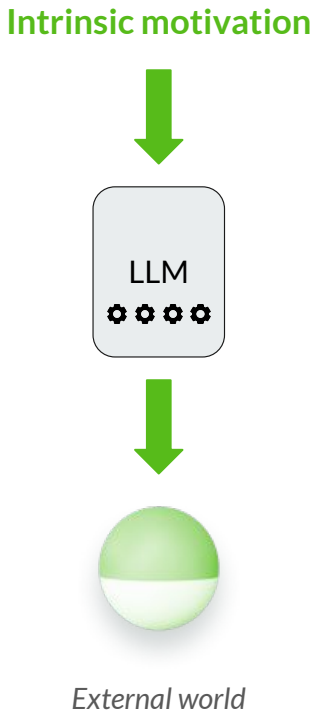
What studying children has taught us

- Language is acquired through **interactions**:
  - with the **socio-cultural** environment
  - with the **physical** environment
- Children are **intrinsically motivated** to learn:
  - To **model and control** their body
  - In **interaction** with their environment
  - In order to **solve** intrinsically and extrinsically defined **problems**

- LLMs are **passive learners**
  - They are trained to **predict probability distribution** over the next token
  - As well as to maximize proxies of human preferences
- They never learned to **solve problems through interactions**

What studying children has taught us

- Can we **integrate key mechanisms** of language acquisition in humans **into LLMs**?
- Can it help overcome LLMs' limitations?
- We do **not** consider a **developmental** approach!
  - We study pre-trained LLMs

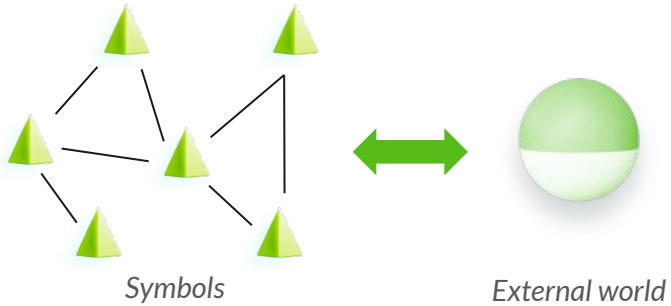


Language and embodiment

Early works in psychology and linguistics evidenced that symbols* we use are grounded in our **socio-cultural and physical world**.

- The Chinese room *(Searle, 1980)*
- The symbol grounding problem *(Harnad, 1990)*

**Symbols encompass here words or grammatical rules*

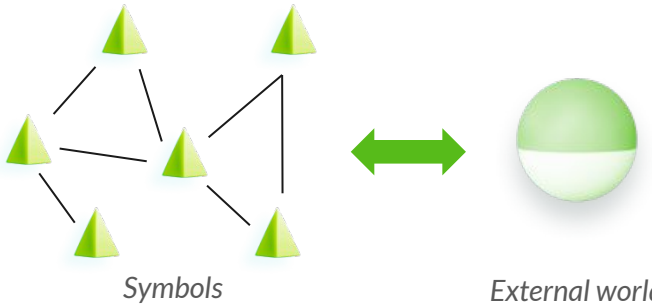


*Symbols encompass here words or grammatical rules

Language and embodiment

Early works in psychology and linguistics evidenced that symbols* we use are grounded in our **socio-cultural and physical world**.

- The Chinese room *(Searle, 1980)*
- The symbol grounding problem *(Harnad, 1990)*



Language is acquired along, and supports the development of other cognitive abilities through **embodied sensorimotor and social experiences**:

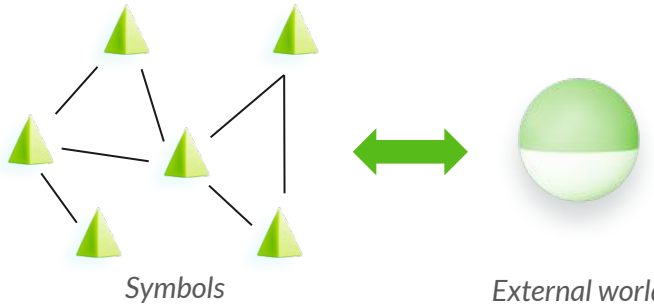
- to create abstractions and concepts *(Piaget, Cangelosi)*
- for thoughts *(Vygotsky)*
- to create theories about the world *(Gopnik)*
- …

Language and embodiment

**Symbols encompass here words or grammatical rules*

Early works in psychology and linguistics evidenced that symbols* we use are grounded in our **socio-cultural and physical world**.

- The Chinese room *(Searle, 1980)*
- The symbol grounding problem *(Harnad, 1990)*



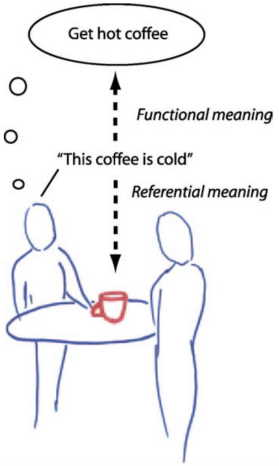
Language is acquired along, and supports the development of other cognitive abilities through **embodied sensorimotor and social experiences**

*We consider a wide definition of embodiment which focuses on the **ability to intervene** in an environment and **perceive** the result of these interventions.*

=> Regardless of the modalities

Functional competence

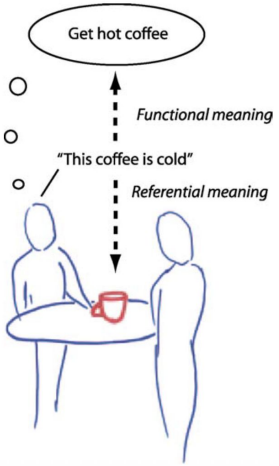
- Beyond referential meaning, language is used to **achieve goals** => **Functional meaning** *(Roy, 2005)*
- One's ability to use language to solve goals is called **functional competence** *(Mahowald, 2024)*



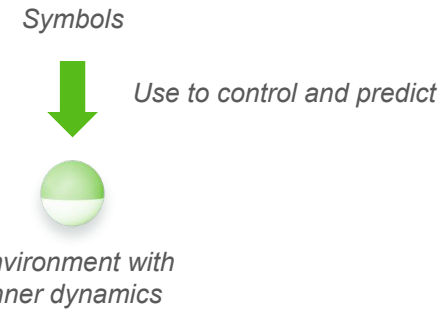
Roy, 2005

Functional competence

- Beyond referential meaning, language is used to **achieve goals**
  => **Functional meaning** *(Roy, 2005)*
- One's ability to use language to solve goals is called **functional competence**
  (Mahowald, 2024)
- Functional competence is also **grounded in goal-directed experiences**

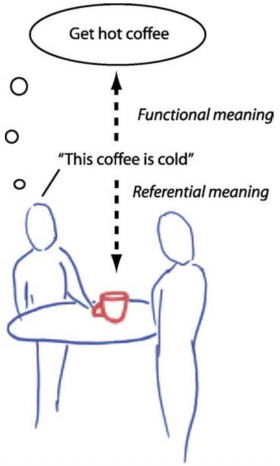


Roy, 2005



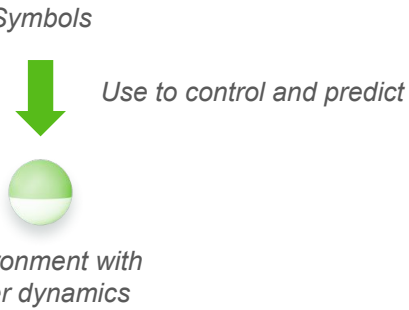
Functional competence

- Beyond referential meaning, language is used to **achieve goals**
  => **Functional meaning** *(Roy, 2005)*
- One's ability to use language to solve goals is called **functional competence** *(Mahowald, 2024)*
- Functional competence is also **grounded in goal-directed experiences**

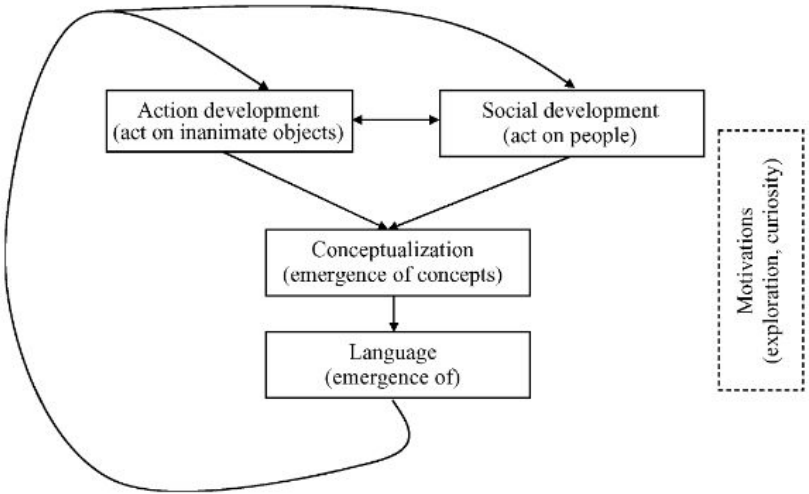


Roy, 2005

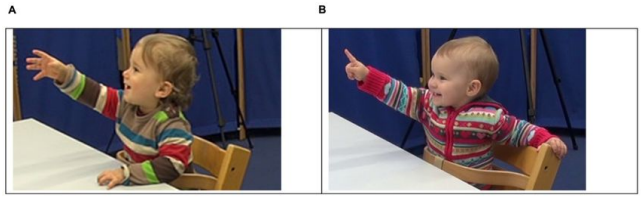
Where do these goals come from?



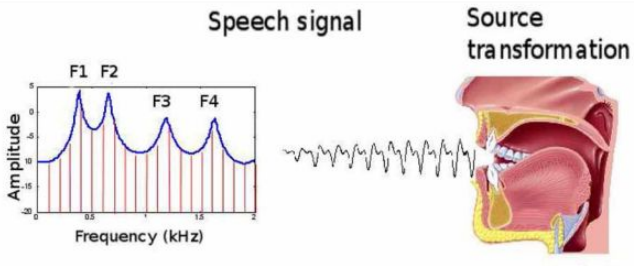
Humans are intrinsically motivated learners



Language and concepts acquisition
Cangelosi et al., 2010



Social interactions
Rohlfing, 2017



Vocal development
Moulin-frier, 2014

Which intrinsic motivation?

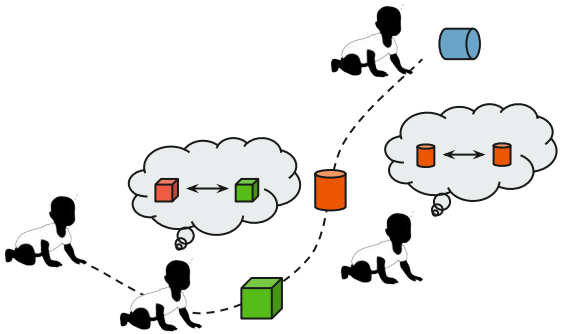
Knowledge-based (KB)*

Competence-based (CB)*

*Oudeyer & Kaplan, 2007

Which intrinsic motivation?

Knowledge-based (KB)*



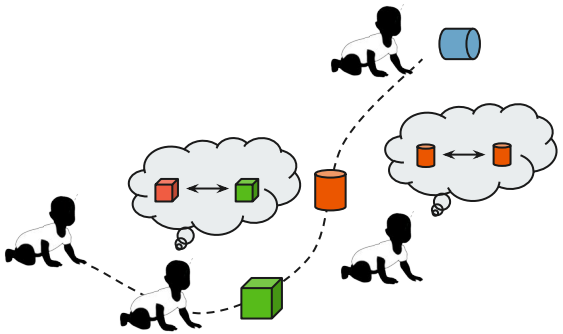
- *KB motivations are about **collecting information***
- *Novelty, empowerment, surprise, prediction error…*

Competence-based (CB)*

*Oudeyer & Kaplan, 2007

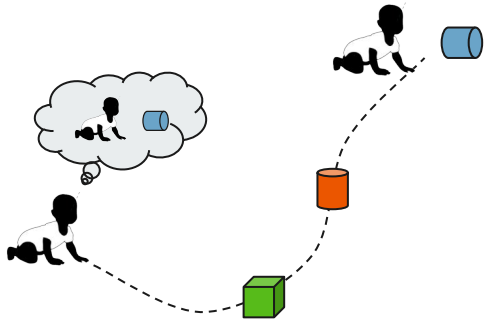
Which intrinsic motivation?

Knowledge-based (KB)*



- *KB motivations are about **collecting information***
- *Novelty, empowerment, surprise, prediction error…*

Competence-based (CB)*



- *CB motivations are **goal-directed***
- *They are about **skill acquisition***

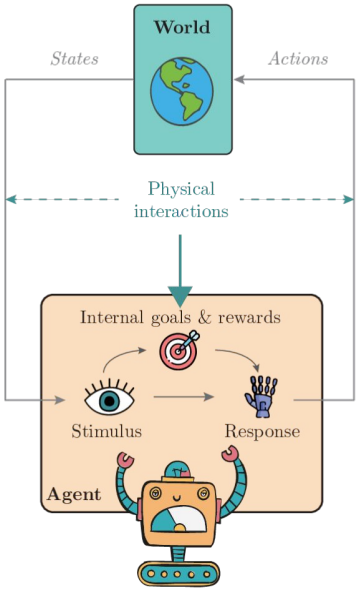
*Oudeyer & Kaplan, 2007

Humans are autotelic learners

- Humans are **autotelic learners** *(Steels, 2004; White, 1959; Oudeyer & Kaplan, 2007)*

  => They **generate**, **select** and learn to **solve** their **own goals**

- This is not a purely individual endeavour: their socio-cultural environment **constraints and provides guidance** to all aspects, from goal-generation, goal-selection, to goal-learning

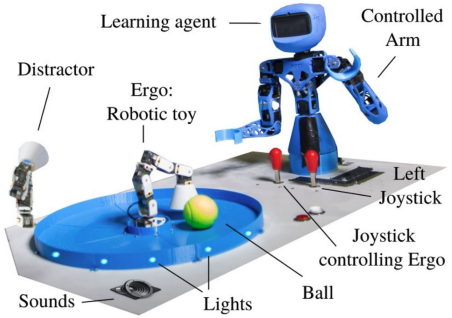


Colas, 2022

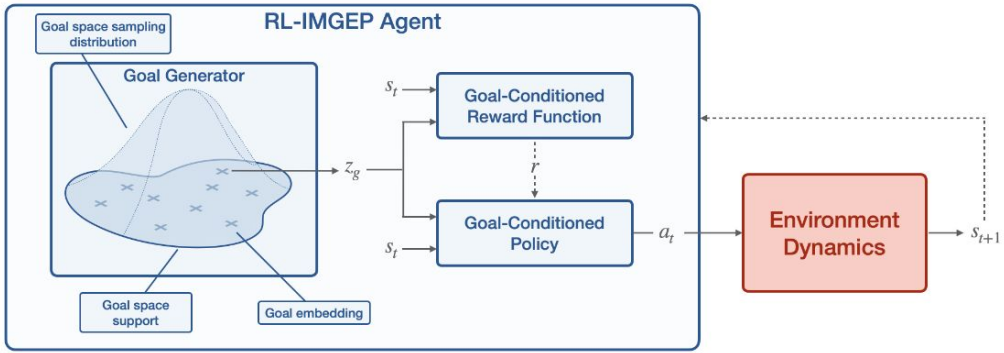
Autotelic artificial agents

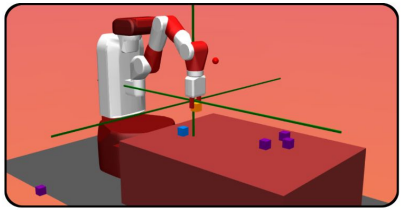
Autotelic agents have been applied to **simulated environments** as well as **real robots**.

It allowed the **discovery of complex skills**.



Forestier et al., 2022





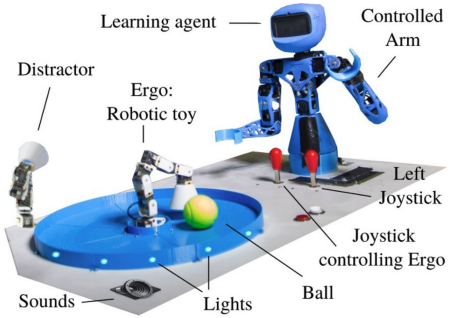
Colas et al., 2019

Autotelic artificial agents

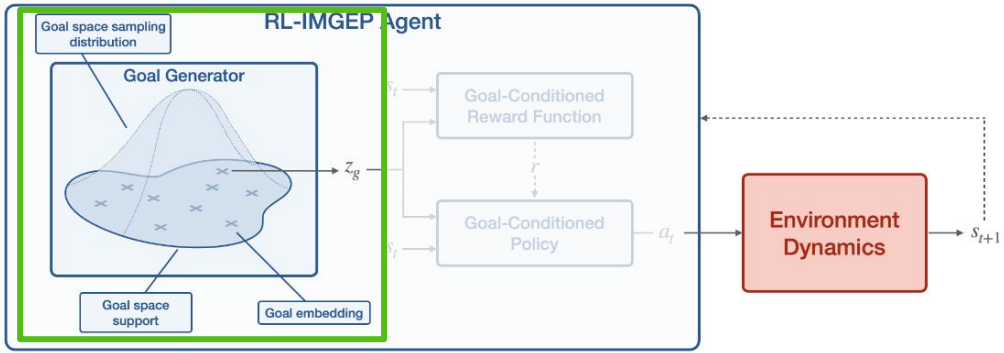
Autotelic agents have been applied to **simulated environments** as well as **real robots**.

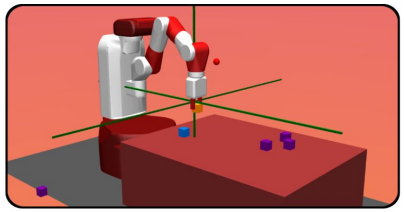
It allowed the **discovery of complex skills**.

The **goal space** and the **sampling strategy** are key elements for such systems.



Forestier et al., 2022





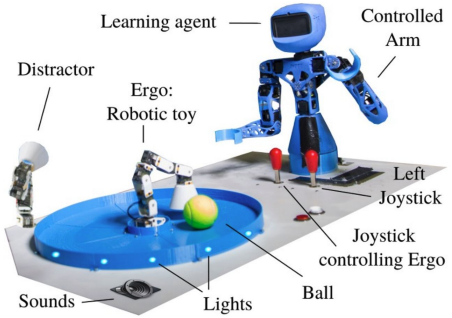
Colas et al., 2019

Autotelic artificial agents

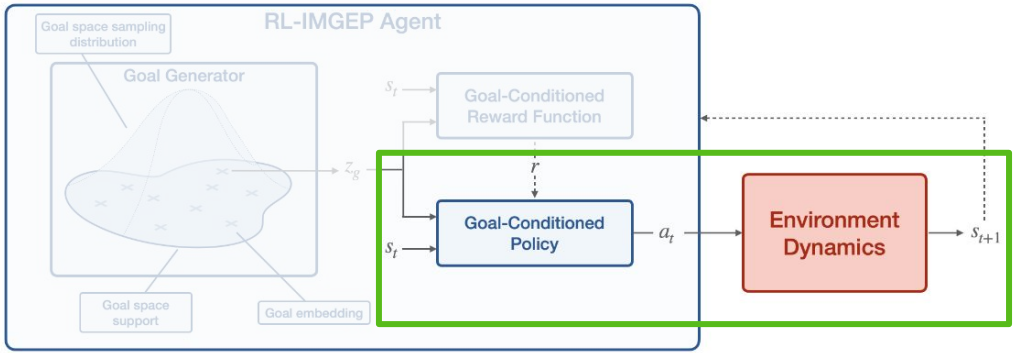
Autotelic agents have been applied to **simulated environments** as well as **real robots**.

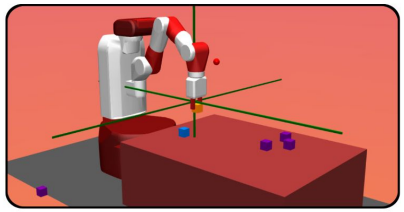
It allowed the **discovery of complex skills**.

The **goal space** and the **sampling strategy** are key elements for such systems.



Forestier et al., 2022





Colas et al., 2019

(Goal-conditioned) Reinforcement Learning

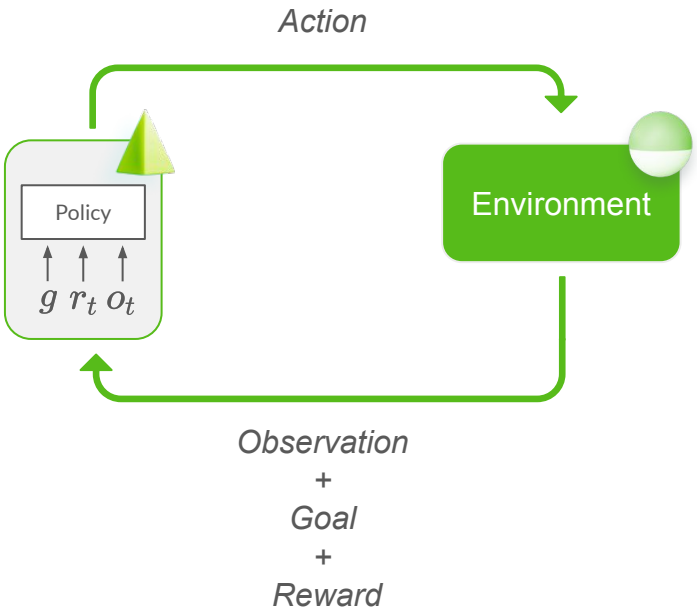
Given a **goal** g at each timestep t:
- the agent **perceives** o_t
- the agent receives a **reward** r_t
- the agent chooses the **action** a_t

The agent chooses actions with its **policy**:

$$\pi : S \times A \mapsto [0, 1]$$

We look for the policy which maximizes the (discounted) sum of rewards:

$$\max_\pi \mathbb{E}_\pi [\sum_{k=0} \gamma^{k+t} r_{t+k}]$$



(Autotelic) Reinforcement Learning

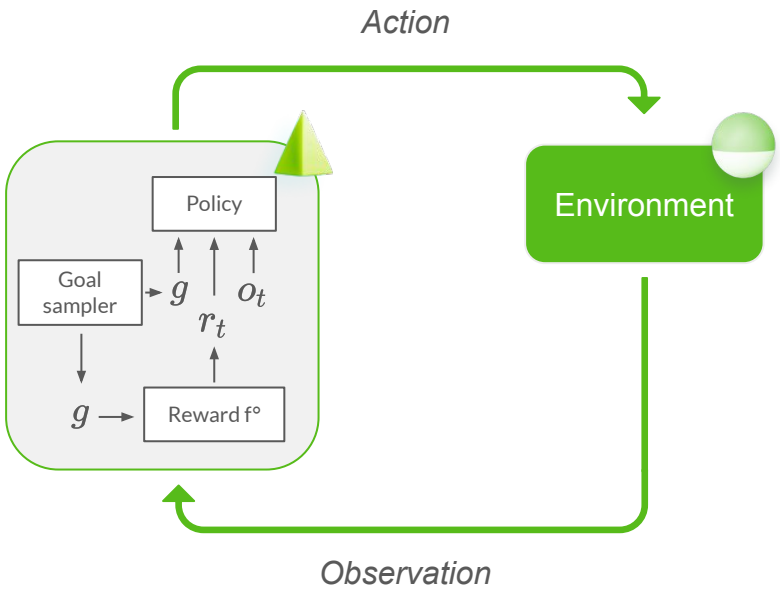
Given a **goal** g at each timestep t:
- the agent **perceives** o_t
- the agent receives a **reward** r_t
- the agent chooses the **action** a_t

The agent chooses actions with its **policy**:

$$\pi : S \times A \mapsto [0, 1]$$

We look for the policy which maximizes the (discounted) sum of rewards:

$$\max_\pi \mathbb{E}_\pi [\sum_{k=0} \gamma^{k+t} r_{t+k}]$$



Towards embodied LLM agents solving problems

Humans

- Language is acquired through **interactions**
- Children are **intrinsically motivated**
  - In particular **autotelic** learners that select their own goals
- Humans use language to solve goals (**functional competence**)

LLMs

- LLMs are **passive learners**
- They never learned to **solve problems through interactions**

Towards embodied LLM agents solving problems

Humans

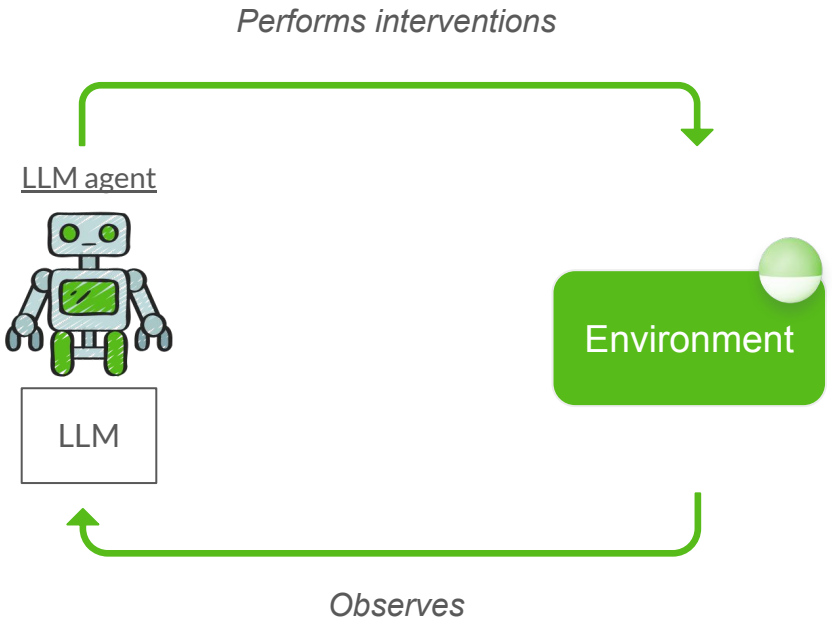
- Language is acquired through **interactions**
- Children are **intrinsically motivated**
  - In particular **autotelic** learners that select their own goals
- Humans use language to solve goals (**functional competence**)

LLMs

- LLMs are **passive learners**
- They never learned to **solve problems through interactions**

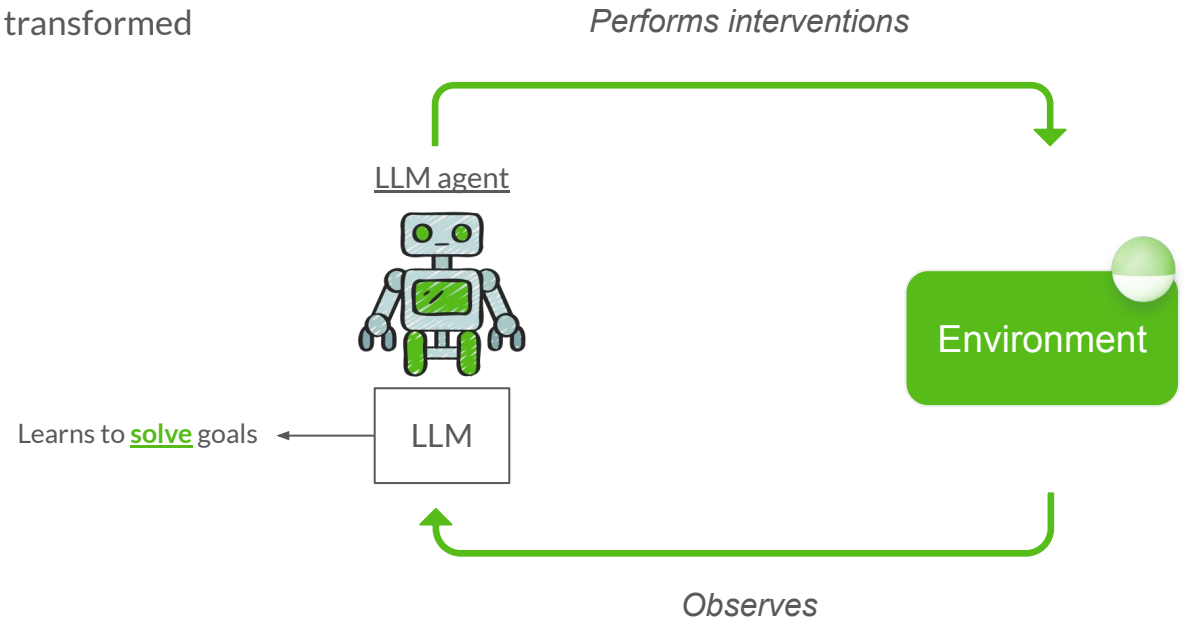
Towards embodied LLM agents solving problems

This PhD explored how **LLMs** can be transformed into **autotelic embodied learners**.



Towards embodied LLM agents solving problems

This PhD explored how **LLMs** can be transformed into **autotelic embodied learners**.



Towards embodied LLM agents solving problems

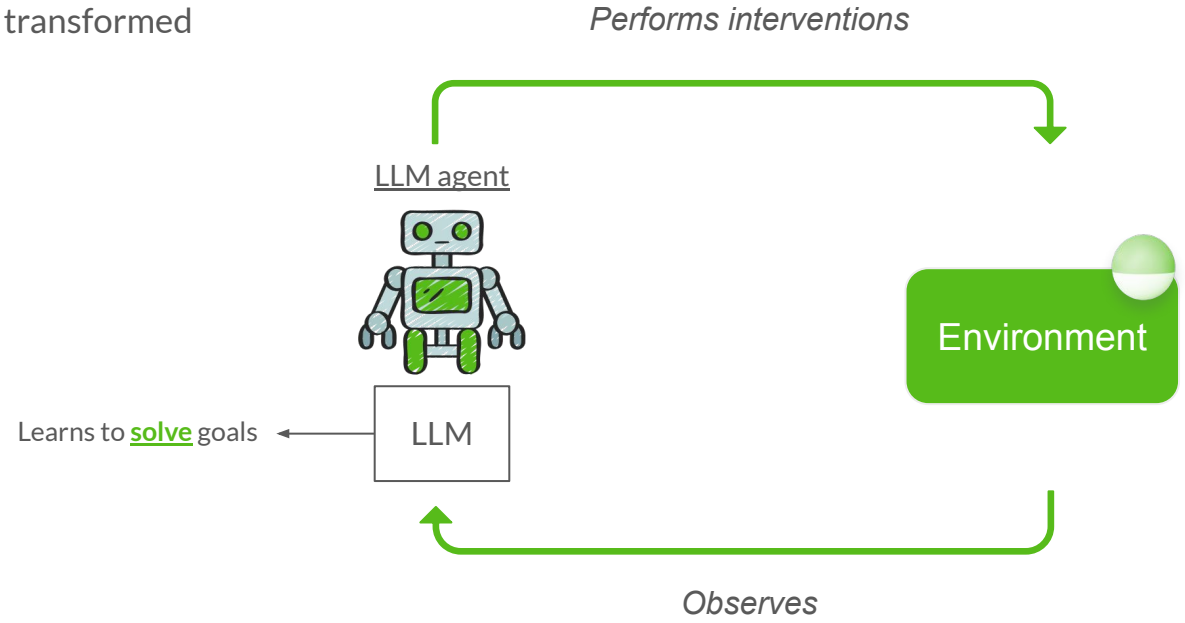
This PhD explored how **LLMs** can be transformed into **autotelic embodied learners**.

Functional Grounding*

Symbol

Use to control and predict for solving tasks

Environment with inner dynamics

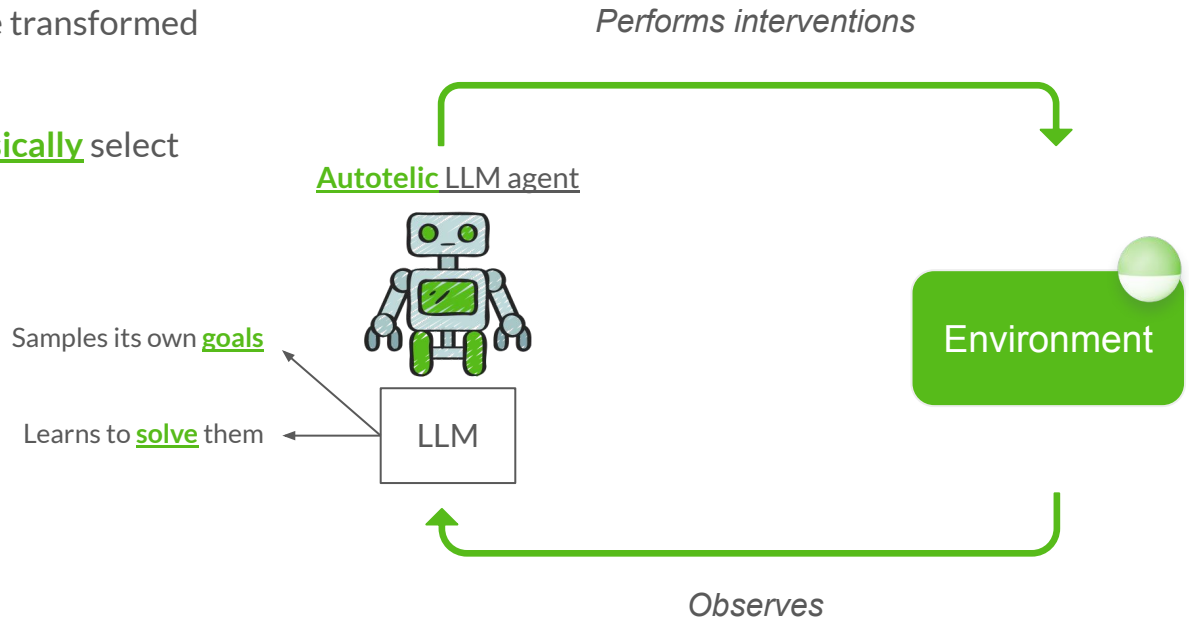


*What I mean by "grounding" in this talk: How do we align our internal representations with the external world.

Towards embodied LLM agents solving problems

This PhD explored how **LLMs** can be transformed into **autotelic embodied learners**.

↳ LLM agents that **intrinsically** select their **own goals**



Towards embodied LLM agents solving problems

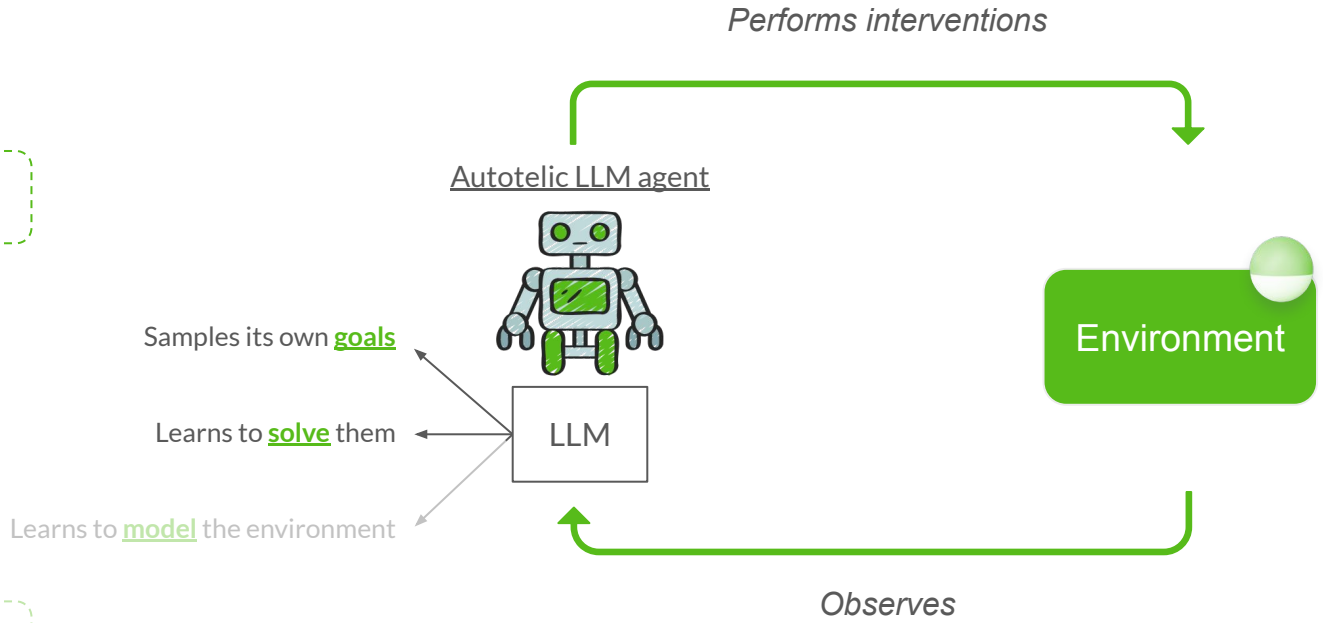
MAGELLAN (Gaven et al., 2025)

GLAM (Romac et al., 2023)

SAC-GLAM (Gaven et al., 2025)

Grounding analysis (Aissi et al., 2024)

WorldLLM (Levy et al., 2025)



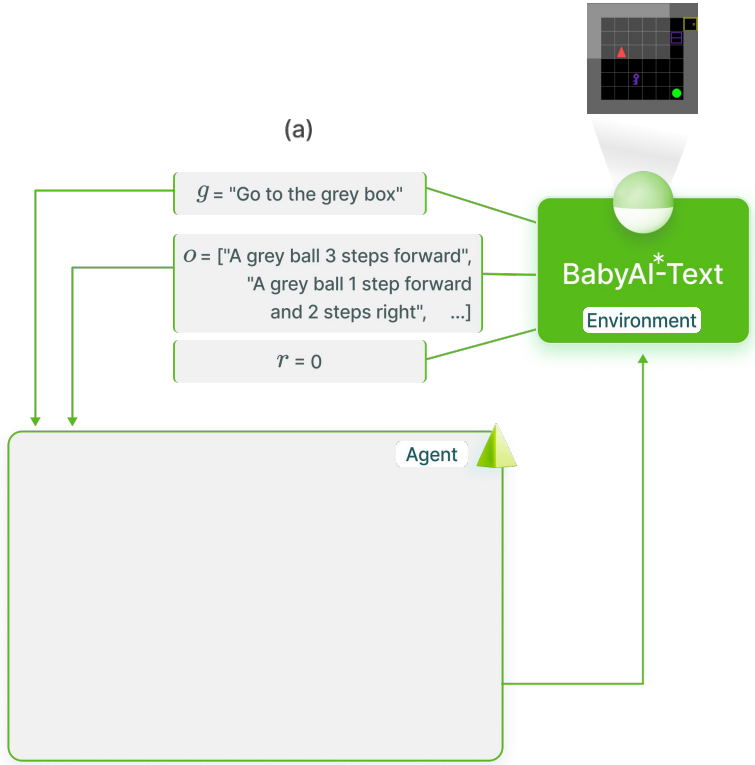
Functional grounding through embodied interactions

Grounding Large Language Models in **Interactive Environments** with Online Reinforcement Learning

Clement Romac*, Thomas Carta*, Thomas Wolf, Sylvain Lamprier, Olivier Sigaud, Pierre-Yves Oudeyer

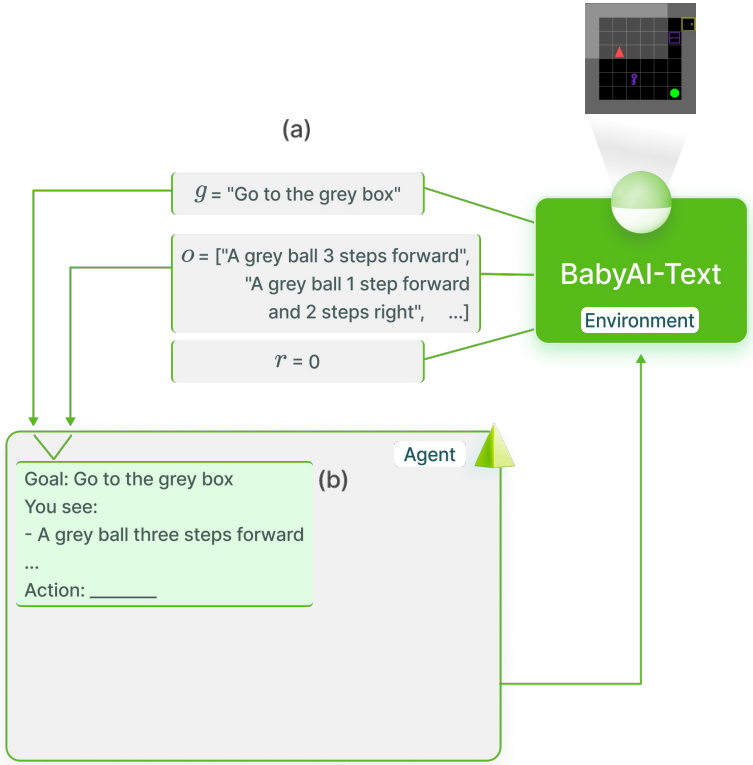
*Equal contribution

GLAM: Grounding with Online RL

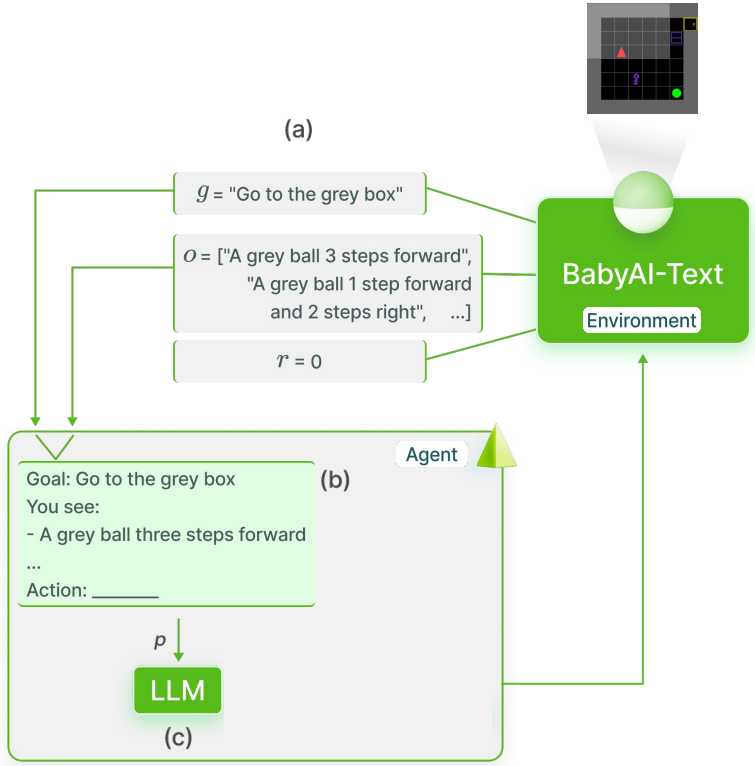


* *Chevalier-Boisvert et al., 2018*

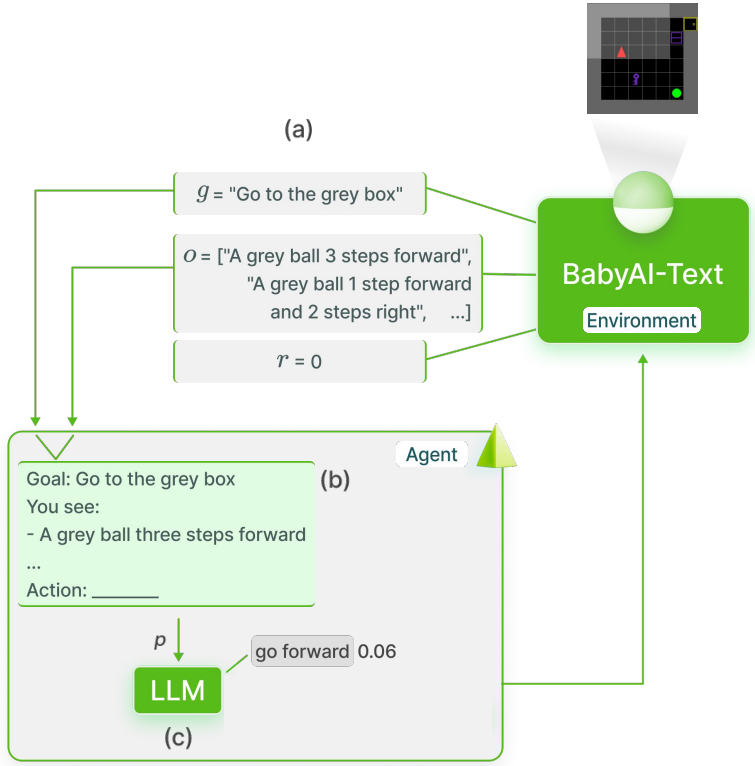
GLAM: Grounding with Online RL



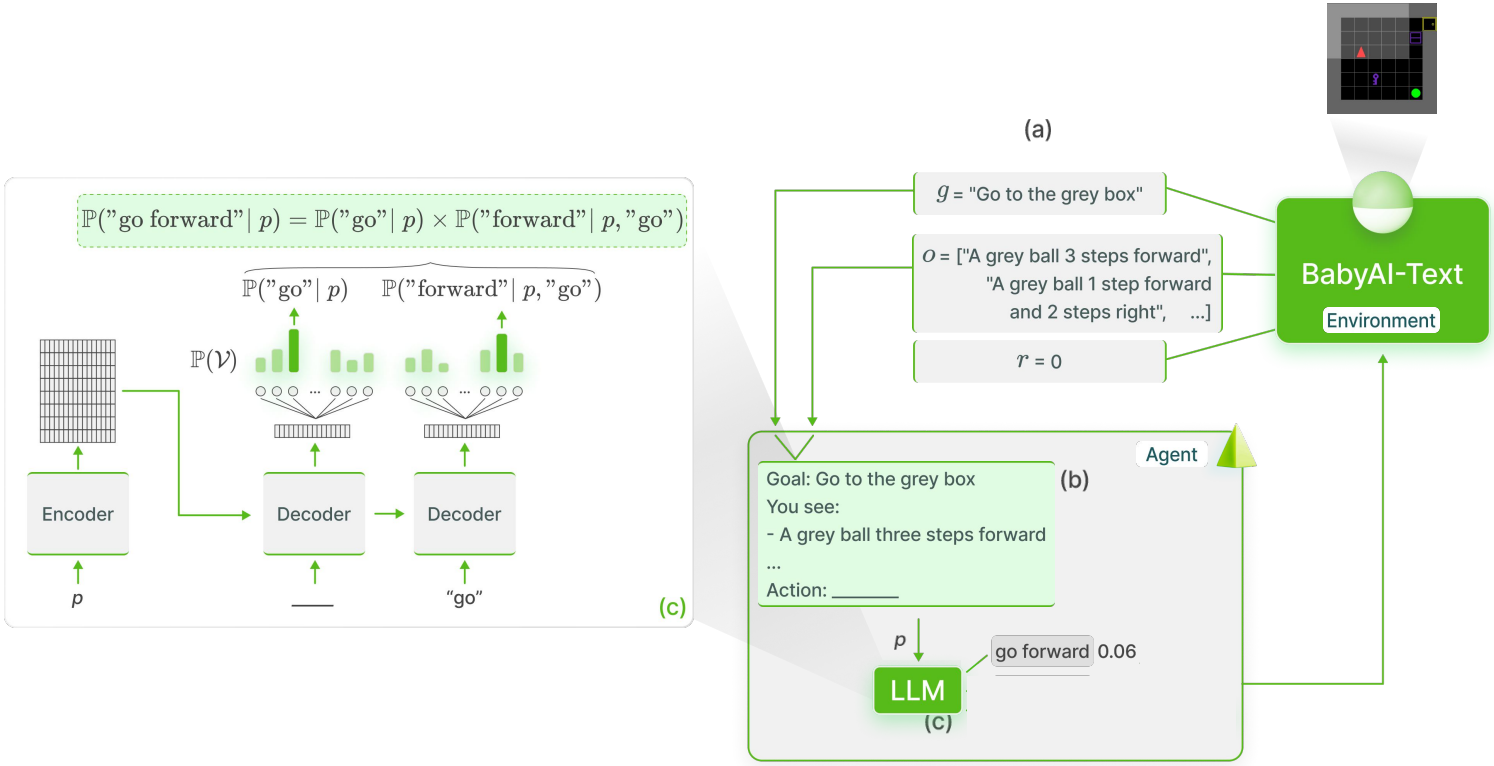
GLAM: Grounding with Online RL



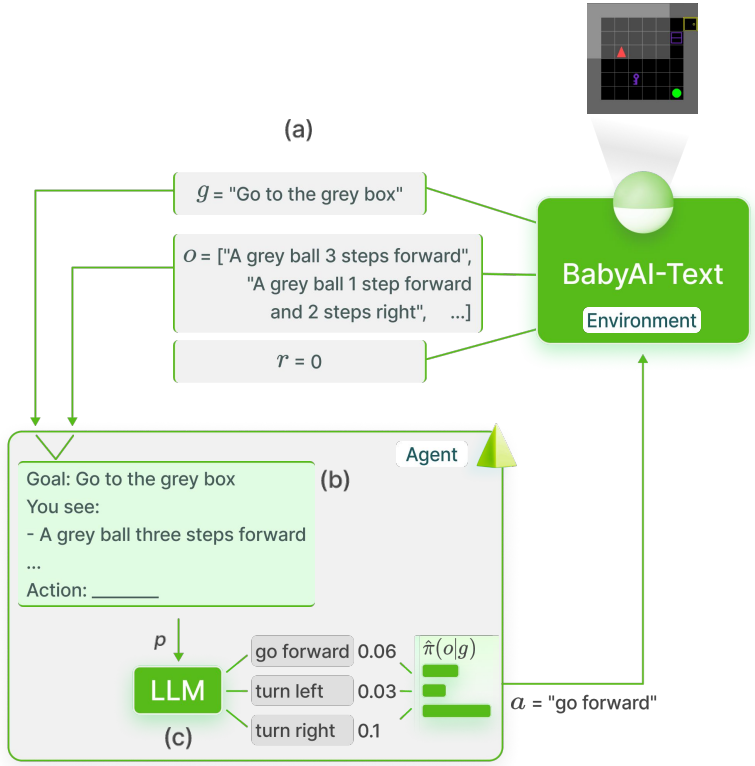
GLAM: Grounding with Online RL



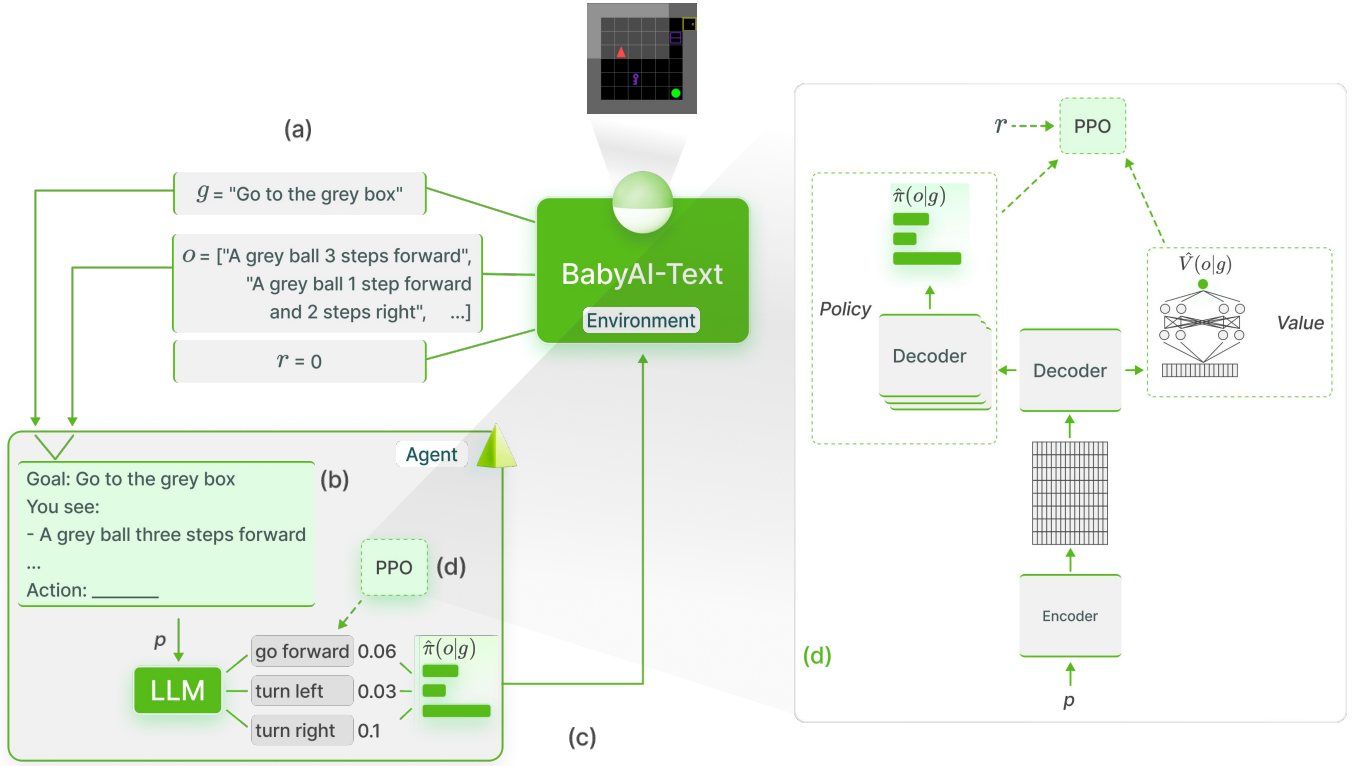
GLAM: Grounding with Online RL



GLAM: Grounding with Online RL



GLAM: Grounding with Online RL

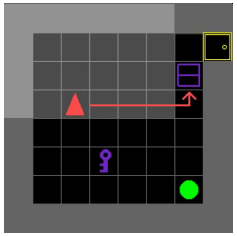


Multi-task RL setup

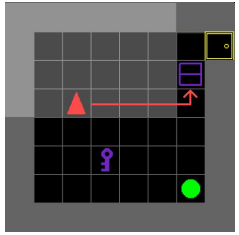
Box
Ball
Key

Purple
Blue
Green
...

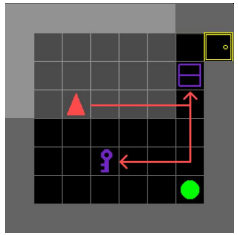
go forward
turn right
turn left
...



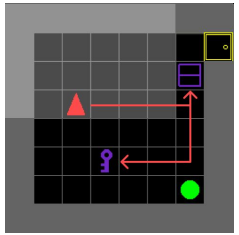
Go to purple box



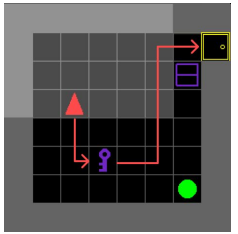
Pick up purple box



Put the purple box next to the purple key



Pick up the purple box then go to the purple key



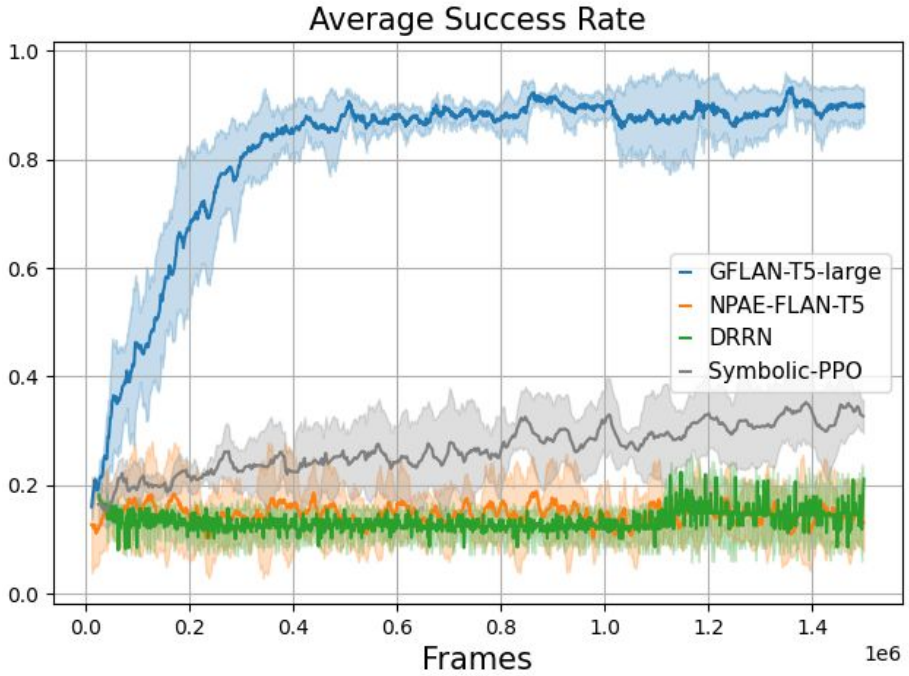
Unlock the yellow door

- 1 room
- 6 actions
  - <turn left>, <turn right>, <go forward>, <pick up>, <drop>, <toggle>
- 8 distractor objects (useless to complete the task)

Prompt

Goal of the agent: *<goal>*
Obs 0: *<obs at t-2>*
Action 0: *<action at t-2>*
...
Action 2:

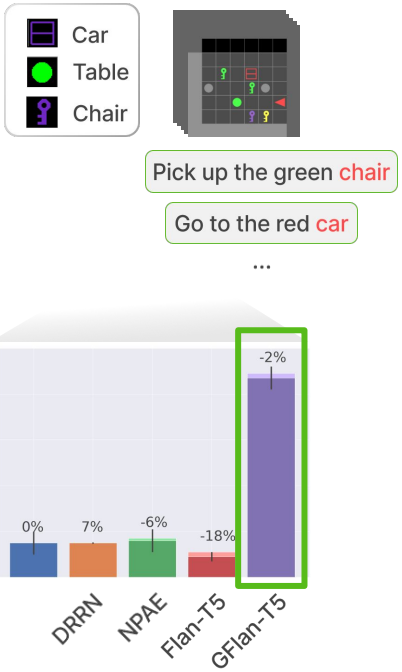
Q1. Sample efficiency



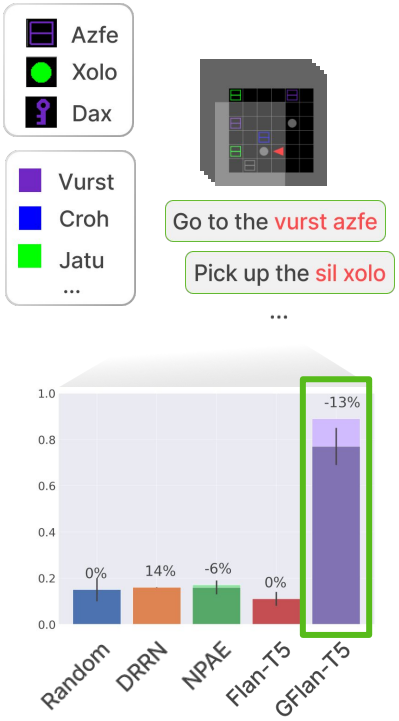
(4 seeds)

- We fine-tuned Flan-T5 780M with GLAM for 1.5M steps in BabyAI-Text
- Tasks/goals are randomly sampled
- We also applied GLAM to a randomly initialized Flan-T5 780M (NPAE)

Q2. Generalization to new objects



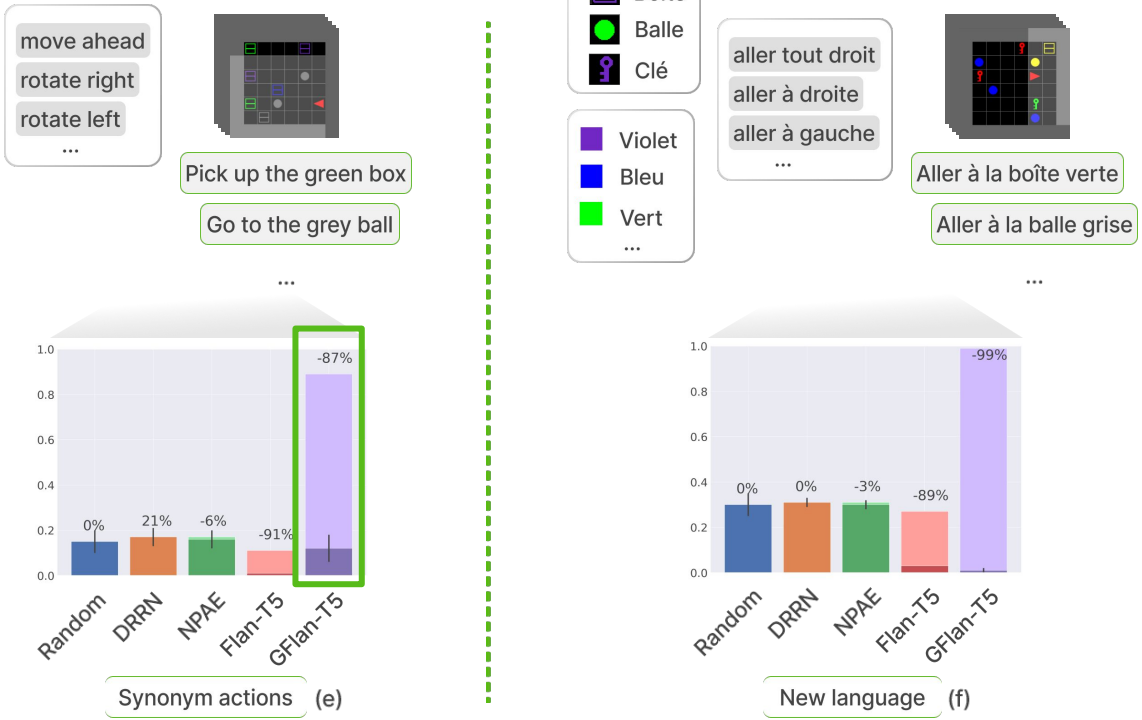
Out-of-vocabulary (b)



Invented words (c)

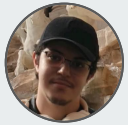
=> Hints about a potentially restrained impact of GLAM

Q3. Generalization to new tasks



RL for Aligning Large Language Models Agents with Interactive Environments : **Quantifying and Mitigating Prompt Overfitting**.

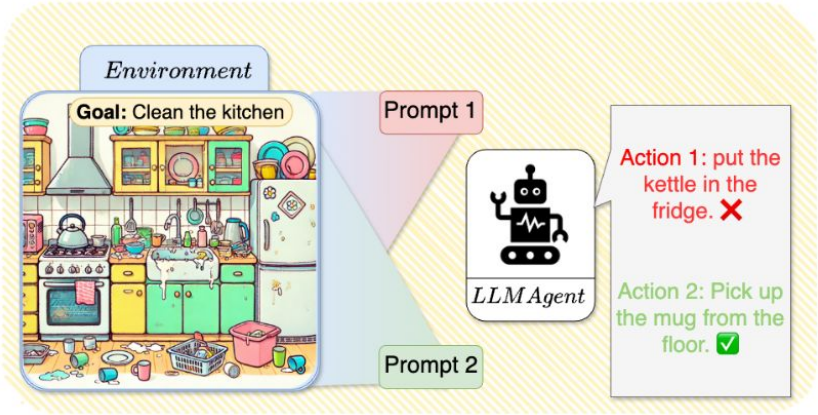
M. S. Aissi, **C. Romac**, T. Carta, S. Lamprier, P.-Y. Oudeyer, O. Sigaud, L. Soulier, and N. Thome



Large-scale study of GLAM

- We perform a **large-scale study of GLAM's impact** on LLMs by varying:
  - LLMs
  - environments
  - prompt formulations

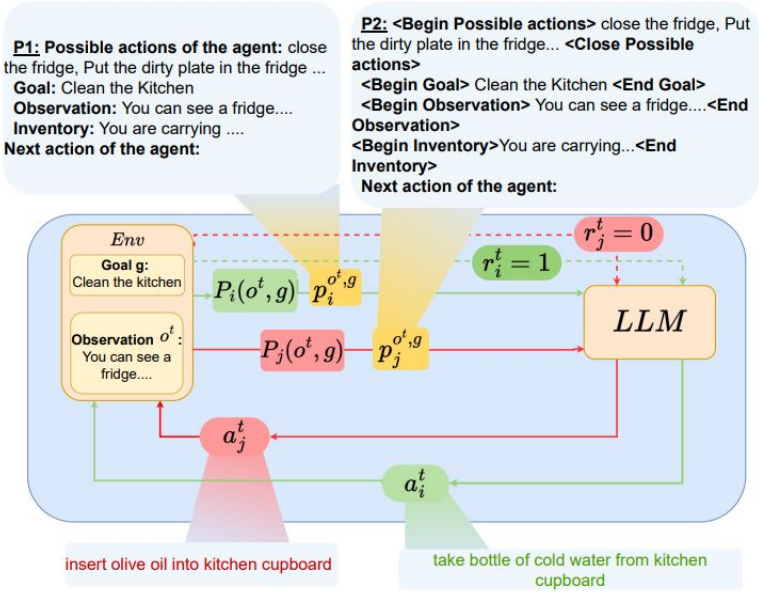
- We study of the **impact of functional grounding on representational abilities of LLMs**:
  - We look how this impacts functional competence
  - But also the broader comprehension of the environment



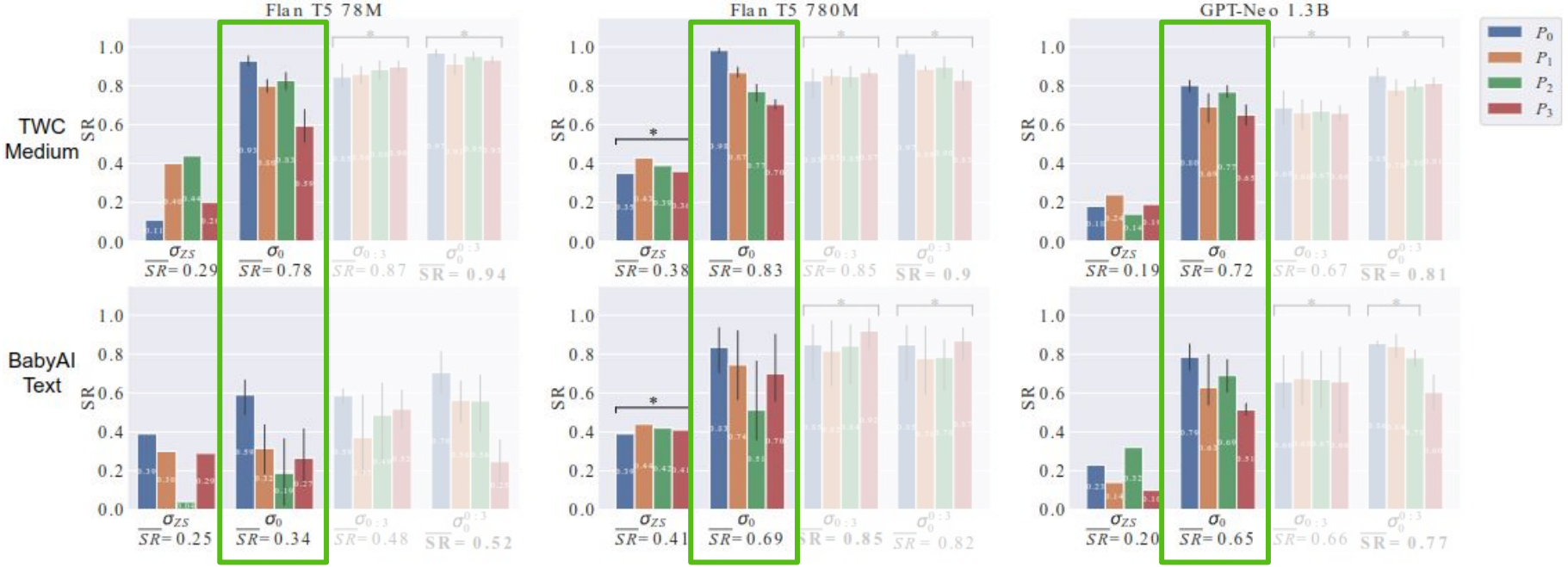
Prompt sensitivity in GLAM

- We begin by looking at how **prompt sensitive** the **functional competence of LLMs grounded with GLAM** is.

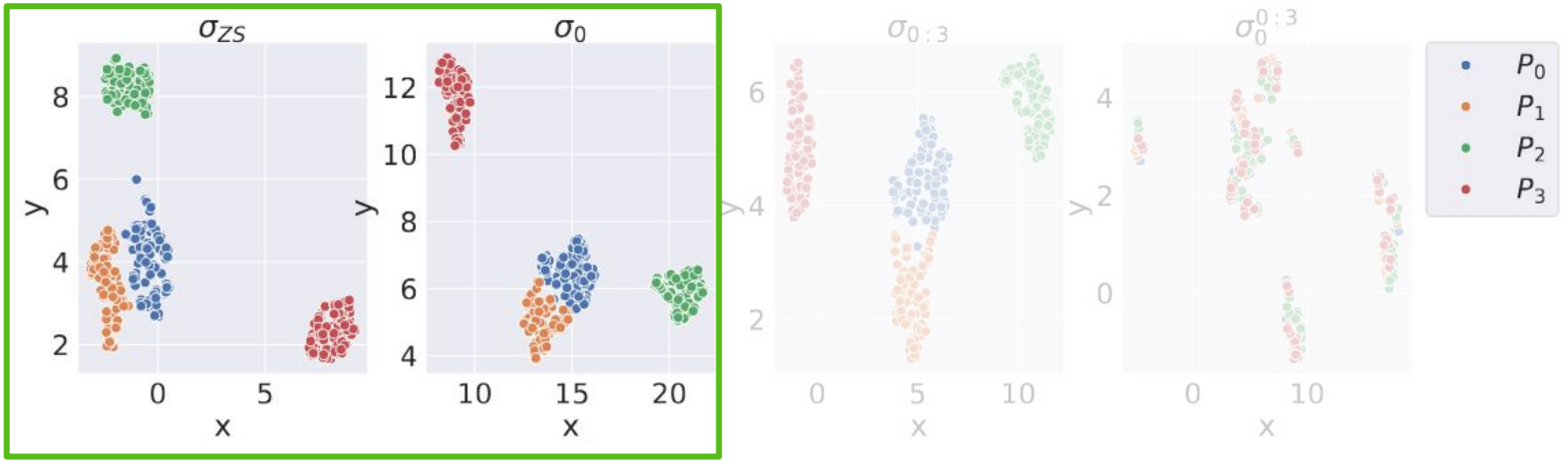
- We design **4 different prompts** and study how testing the LLM on a **different prompt formulation than the one seen during training** affects its performance.



Prompt sensitivity in GLAM



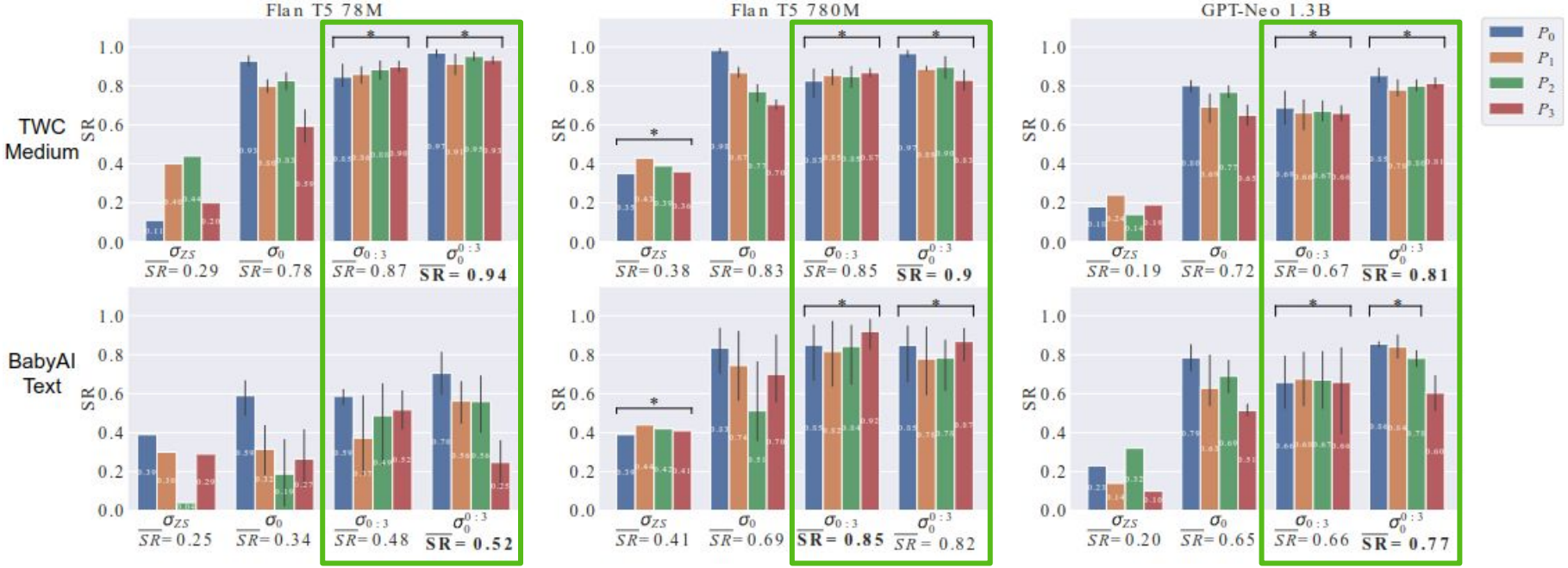
Diving into internal representations



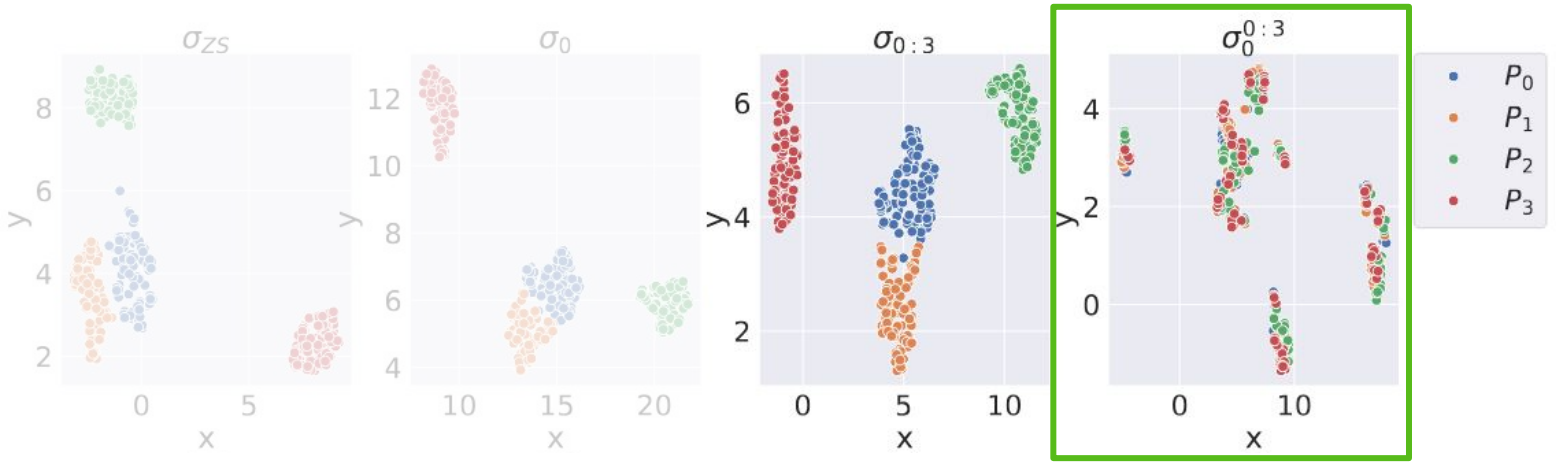
Before GLAM

After GLAM

Prompt sensitivity in GLAM



Diving into internal representations



Multiple prompts

Contrastive loss

Broader impact of functional grounding

- We also proposed an experiment in which functionally grounded LLMs are **asked to answer questions** about the environment.

- We design two set of questions:
  - **Object Counting (OC)**: capturing information in the observations
  - **Task Related (TR)**: identifying useful objects for a task

	TWC TR	TWC OC
σ_{zs}	0.4866 *	0.0876 ***
σ_0	0.4901 *	0.1340 ***
$\sigma_{0:3}$	0.5019 *	0.2526 *
$\sigma_0^{0:3}$	**0.6322**	**0.5155**

Table 1.4: **Environmental knowledge** of GPT-Neo 1.3B on TWC TR and TWC OC datasets. * and *** correspond to the p-value (resp. < 0.05 and < 0.001) of Welch's t-test to compare the performance between $\sigma_0^{0:3}$ and other scenarios. We observe a significant improvement with $\sigma_0^{0:3}$ scenario compared to σ_{zs}, σ_0, and $\sigma_{0:3}$ scenarios across both datasets.

Broader impact of functional grounding

- We also proposed an experiment in which functionally grounded LLMs are **asked to answer questions** about the environment.

- We design two set of questions:
  - **Object Counting (OC)**: capturing information in the observations
  - **Task Related (TR)**: identifying useful objects for a task

	TWC TR	TWC OC
σ_{zs}	0.4866 *	0.0876 ***
σ_0	0.4901 *	0.1340 ***
$\sigma_{0:3}$	0.5019 *	0.2526 *
$\sigma_0^{0:3}$	**0.6322**	**0.5155**

Table 1.4: **Environmental knowledge** of GPT-Neo 1.3B on TWC TR and TWC OC datasets. * and *** correspond to the p-value (resp. < 0.05 and < 0.001) of Welch's t-test to compare the performance between $\sigma_0^{0:3}$ and other scenarios. We observe a significant improvement with $\sigma_0^{0:3}$ scenario compared to σ_{zs}, σ_0, and $\sigma_{0:3}$ scenarios across both datasets.

Conclusion

We showed that **GLAM** - online RL-based functional grounding - can:

- **Improve LLMs' functional competence**
- Retain the LLMs' **generalization of functional competence** to environment variations

Our large-scale study hints at **representational changes** that impact the LLM **beyond functional competence**.

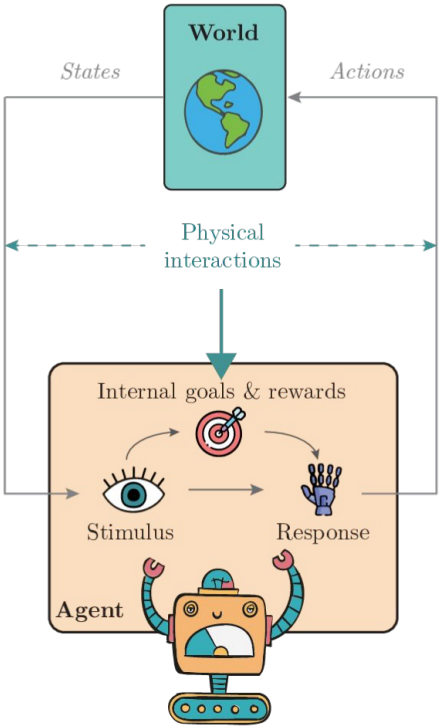
In this part of the talk, goals/tasks were provided by the environment, we will now move to **autotelic approaches to functional grounding**.

Towards autotelic functional grounding

Building autotelic LLM agents

Autotelic RL agents are characterized by:

1) A **goal space**
2) A **goal-selection strategy**
3) A goal-conditioned **reward function**
4) **Goal-learning** mechanisms



b) Autotelic RL

Colas, 2022

MAGELLAN: **Metacognitive** predictions of **learning progress** guide autotelic LLM agents in large goal spaces

Loris Gaven, Thomas Carta, **Clement Romac**, Cedric Colas, Sylvain Lamprier, Olivier Sigaud, Pierre-Yves Oudeyer


Building autotelic LLM agents

Autotelic RL agents are characterized by:

1) A **goal space**

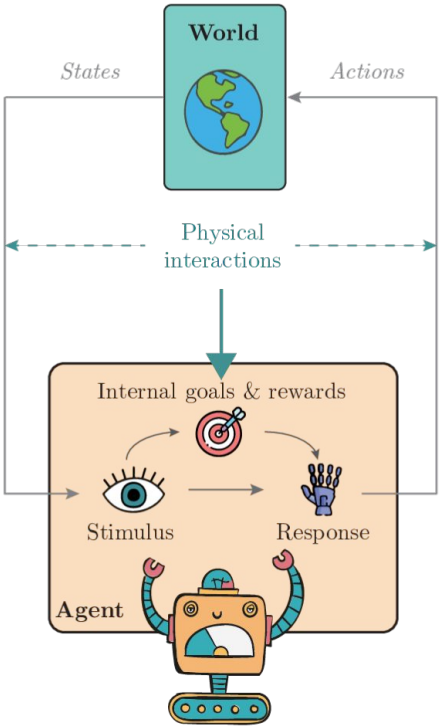
2) A **goal-selection strategy**

3) ~~A goal-conditioned **reward function**~~

4) ~~**Goal learning** mechanisms~~

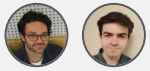
How can autotelic LLM agents **select their goals**?

This work studies how to scale existing goal-selection approaches to **extremely large goal spaces** in which goals are **natural language instructions**.

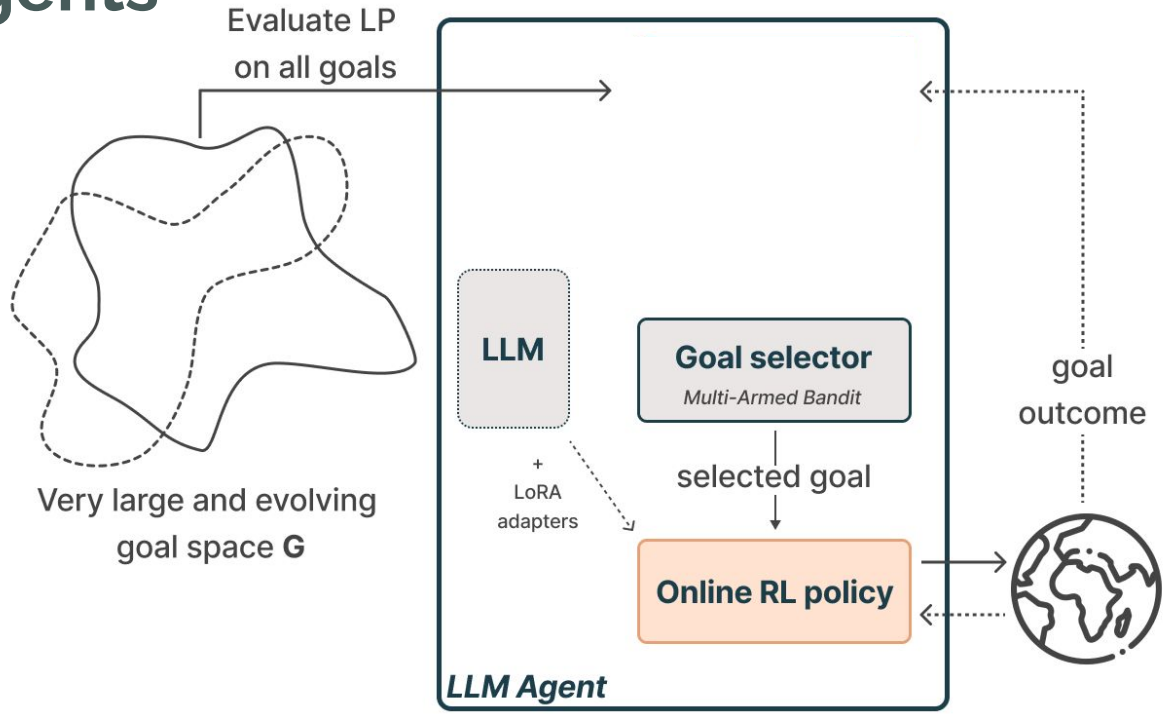


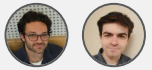
b) Autotelic RL

Colas, 2022



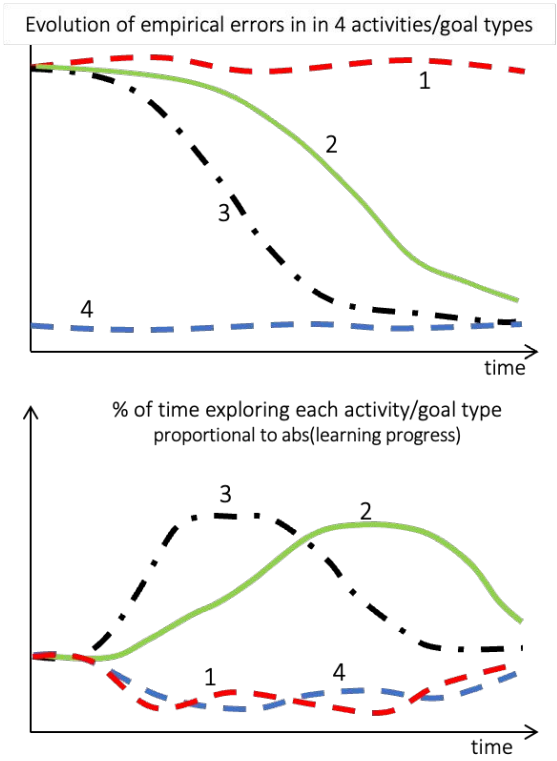
Autotelic LLM agents





How do humans select goals?

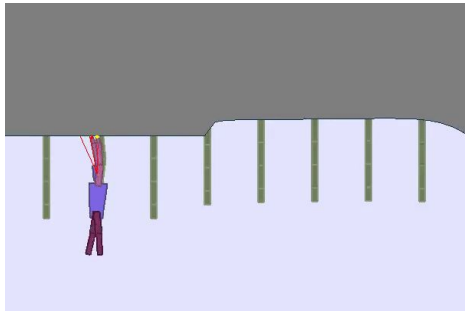
- What is an **interesting goal**?
- One that maximizes **Learning Progress** *(Kaplan & Oudeyer, 2007)*
  - $LP_t(\tau) = \frac{\partial C_t(\tau)}{\partial t} \simeq C_t(\tau) - C_{t-N}(\tau)$



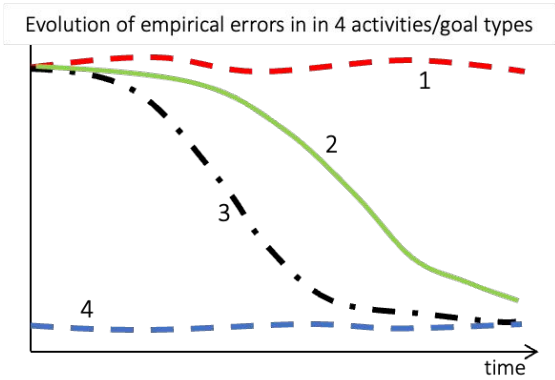

How do humans select goals?

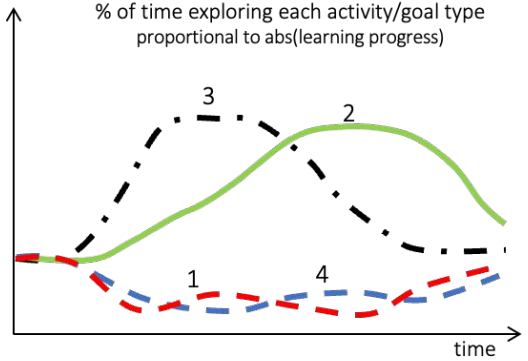
- What is an **interesting goal**?
- One that maximizes **Learning Progress** *(Kaplan & Oudeyer, 2007)*
  - $LP_t(\tau) = \frac{\partial C_t(\tau)}{\partial t} \simeq C_t(\tau) - C_{t-N}(\tau)$

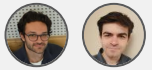
LP enables automatic skill discovery in real world robots *(Baranes, 2013)*



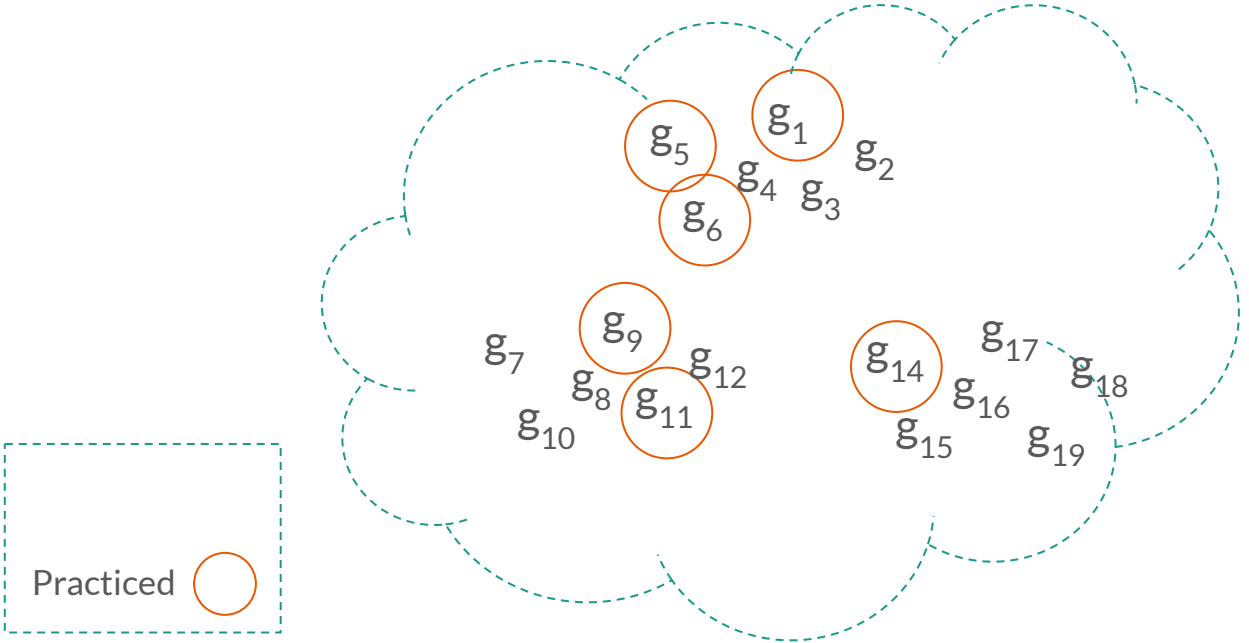
LP enables complex skill learning in RL agents *(Romac, 2021)*







Computing Learning Progress approximates



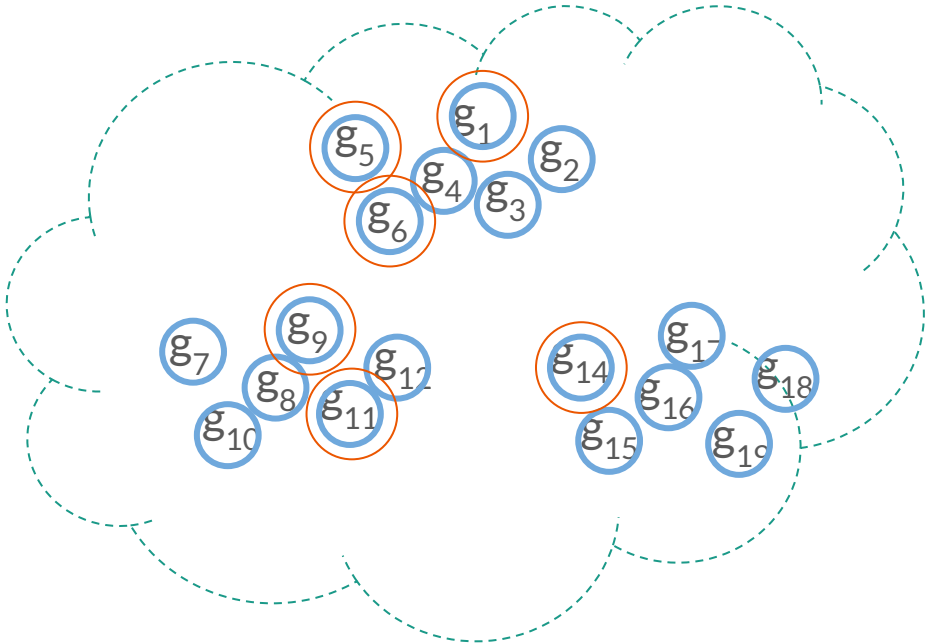

Computing Learning Progress approximates

Eval LP:

Frequently evaluate the agent on all goals and update all competence and LP estimations

\+ Perfectly tracks **competence transfer**

\- Computationally **intractable** when the goal space is large



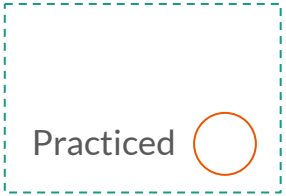

Computing Learning Progress approximates

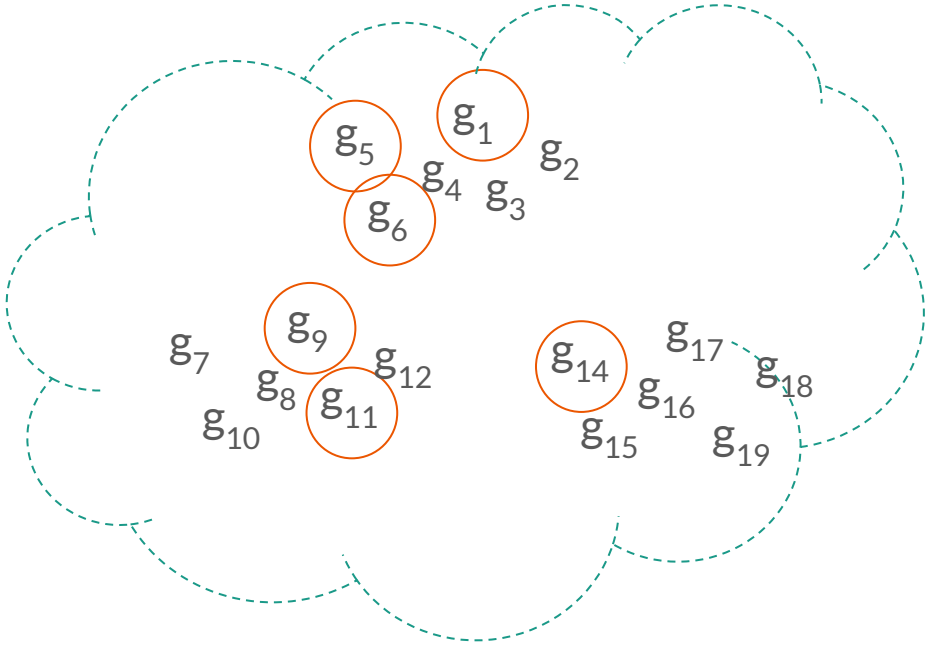
Online LP:

Update the competence (and LP) estimation of a goal whenever it is practiced

- \+ No additional computation
- \- Do not track **competence transfer** between goals

g_1	0,1,0,0,1,0,1,1,1
g_5	1,0,1,1,0,1,1,1,0
g_6	0,0,1,1,1
g_9	0,0,1,0,0,1
g_{11}	1,1,0
g_{14}	0,0,0,0,0,1,0



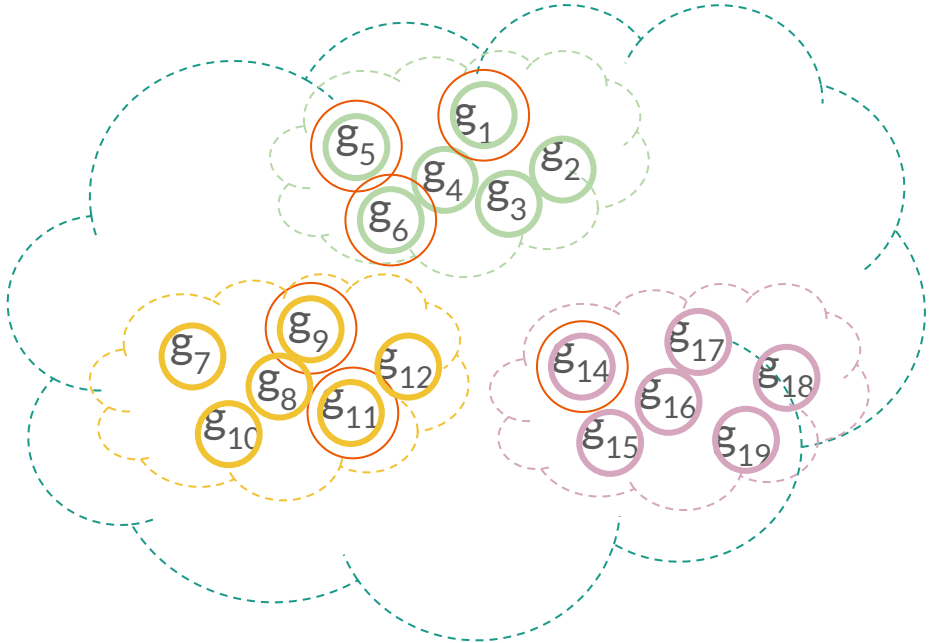



Computing Learning Progress approximates

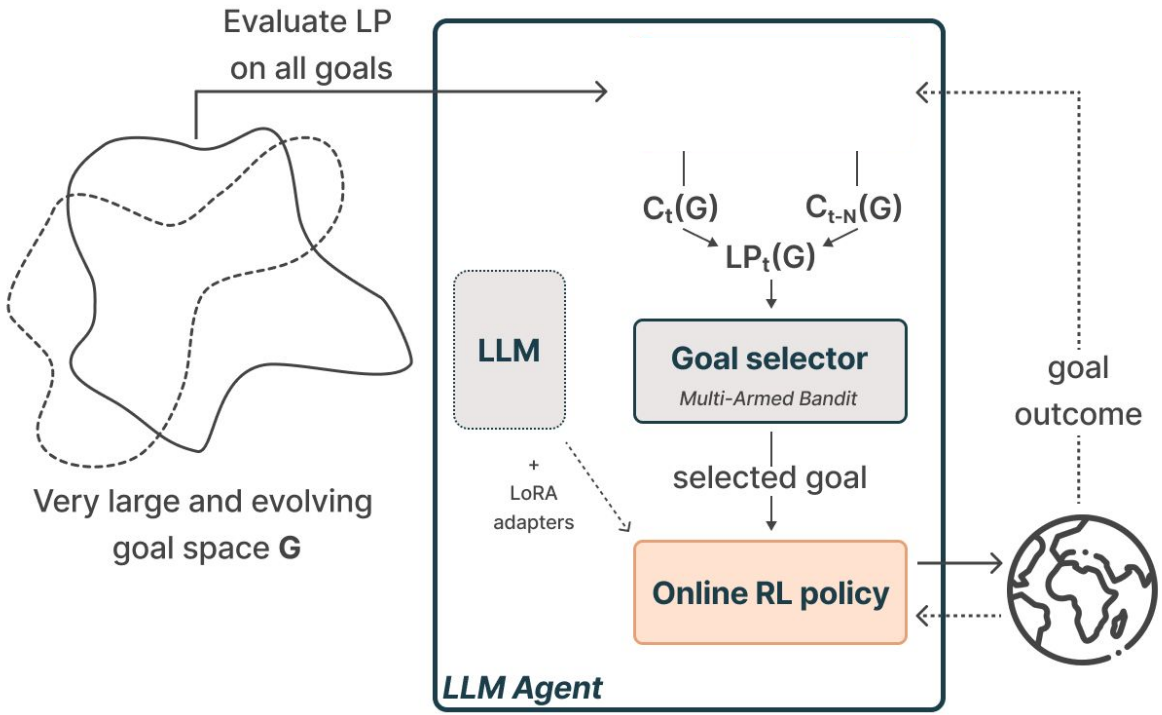
EK-Online LP:

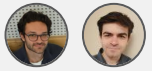
Update the competence (and LP) estimation of a **category** whenever one of its goals is practiced

\+ No additional computation
\+ Assumes **competence transfer** within categories
\- Requires **expert**-defined categories



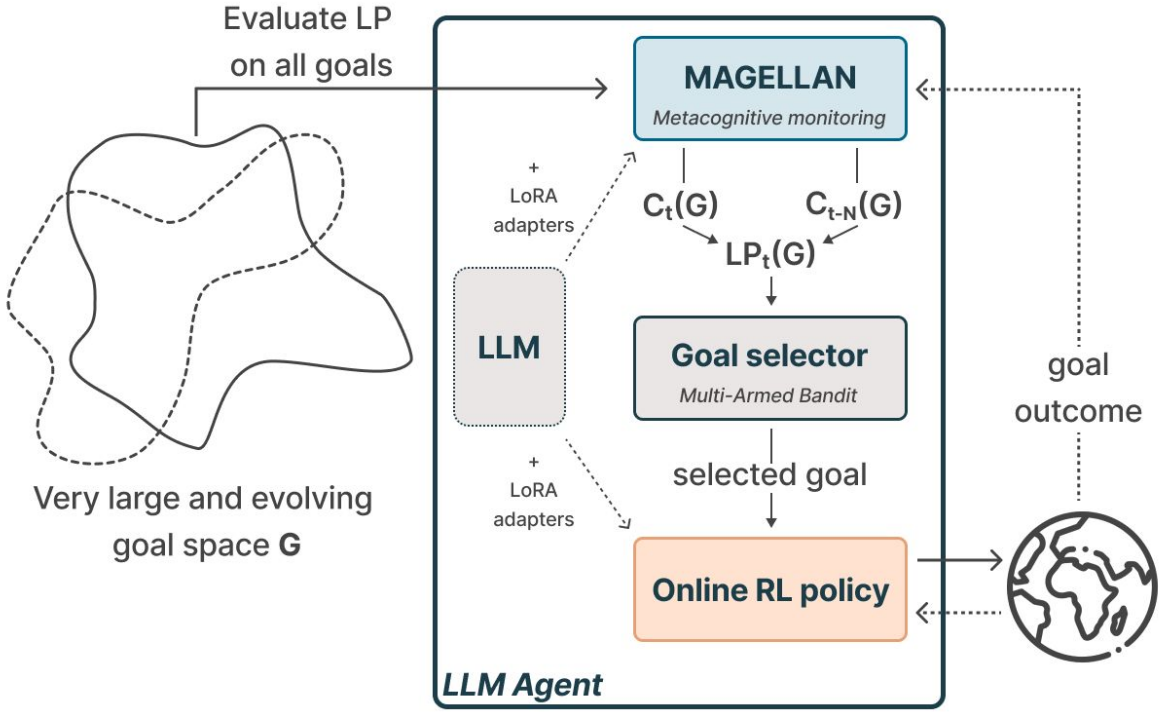
MAGELLAN


Evaluate LP
on all goals
Very large and evolving
goal space G
$C_t(G)$
$C_{t-N}(G)$
$LP_t(G)$
LLM
+
LoRA
adapters
Goal selector
Multi-Armed Bandit
selected goal
Online RL policy
LLM Agent
goal
outcome



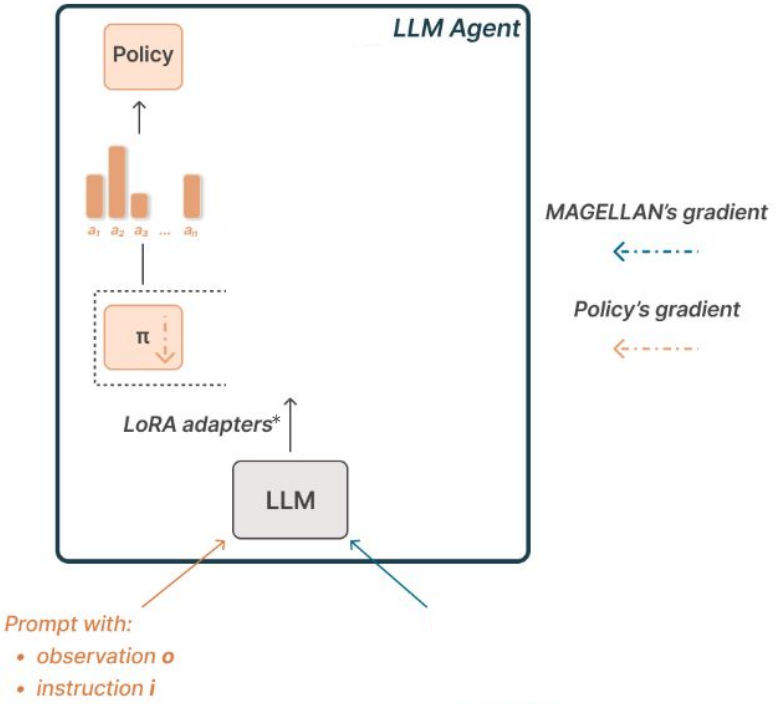
MAGELLAN

- We propose to augment LLMs with **metacognitive** monitoring skills.
- Can an LLM learn to **predict its own competence and LP**?
- Can it grasp **semantic relationships** between goals and **generalize its competence estimation to goals not practiced**?

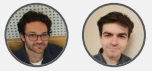



Learning the policy and MAGELLAN

- We use **GLAM** to fine-tune the LLM's policy.

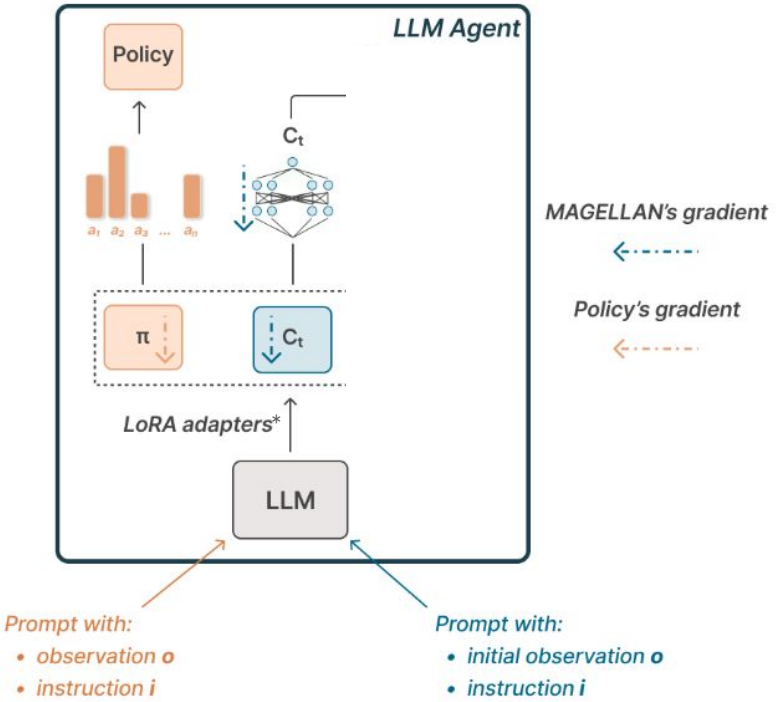


*Hu et al., 2021

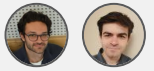


Learning the policy and MAGELLAN

- We use **GLAM** to fine-tune the LLM's policy.
- MAGELLAN uses the **LLM to project goals into a continuous space** and then uses a Multi-Layer Perceptron to **estimate the competence**.

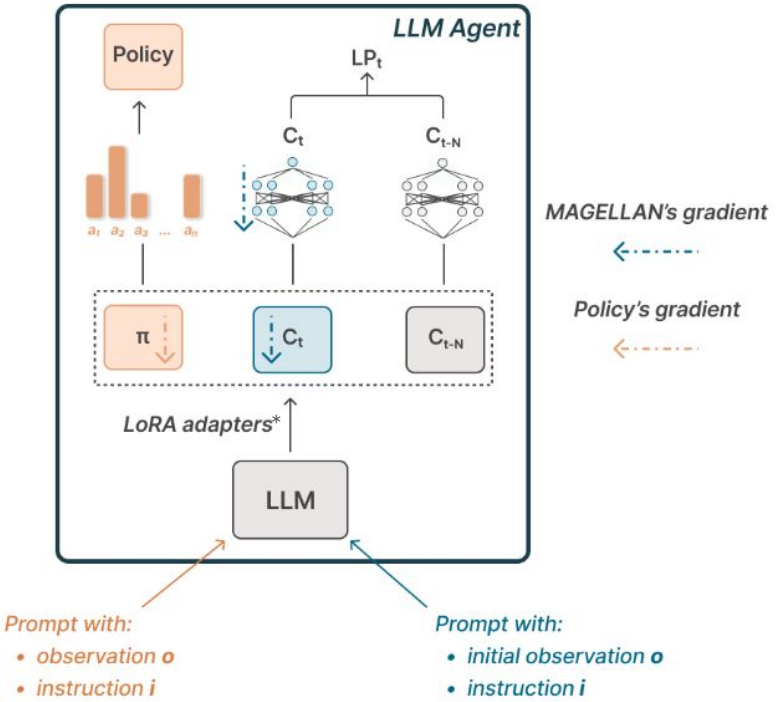


*Hu et al., 2021



Learning the policy and MAGELLAN

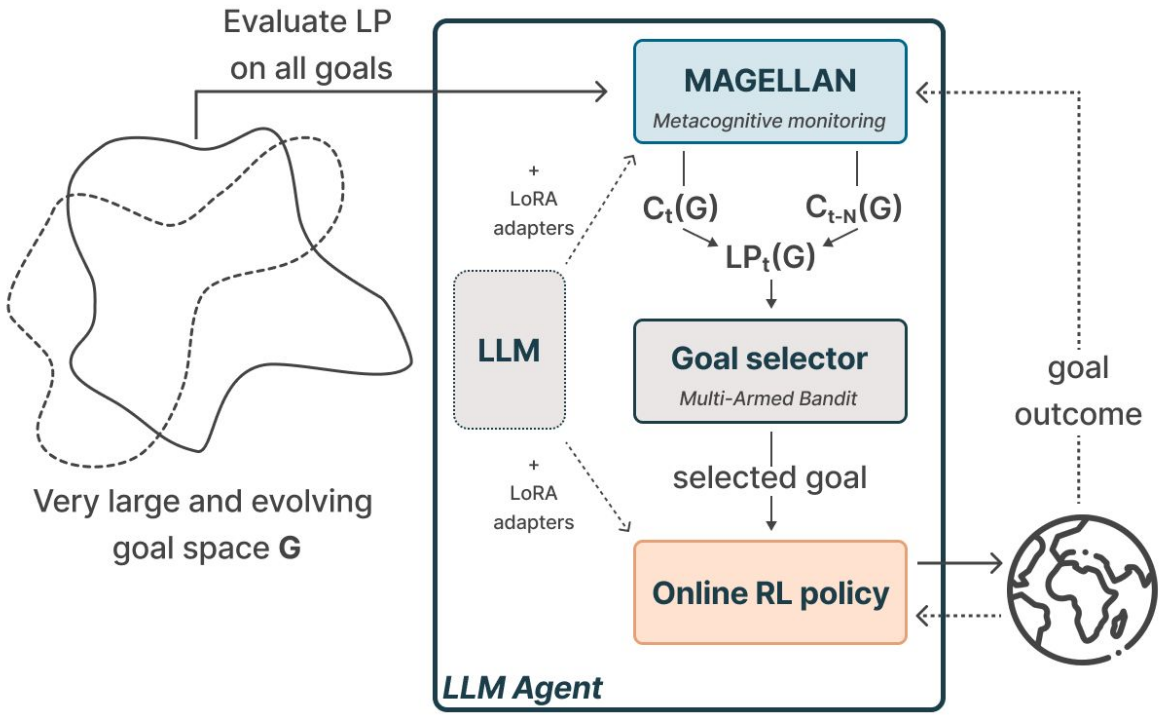
- We use **GLAM** to fine-tune the LLM's policy.
- MAGELLAN uses the **LLM to project goals into a continuous space** and then uses a Multi-Layer Perceptron to **estimate the competence**.
- We keep **older versions of MAGELLAN's competence estimator to compute LP**.



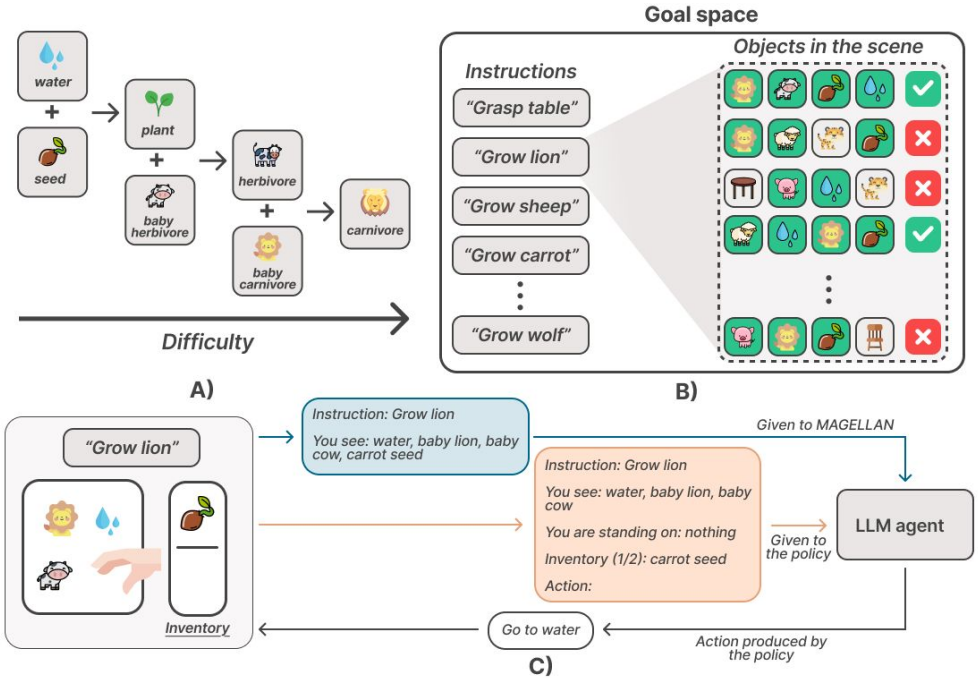
*Hu et al., 2021

MAGELLAN

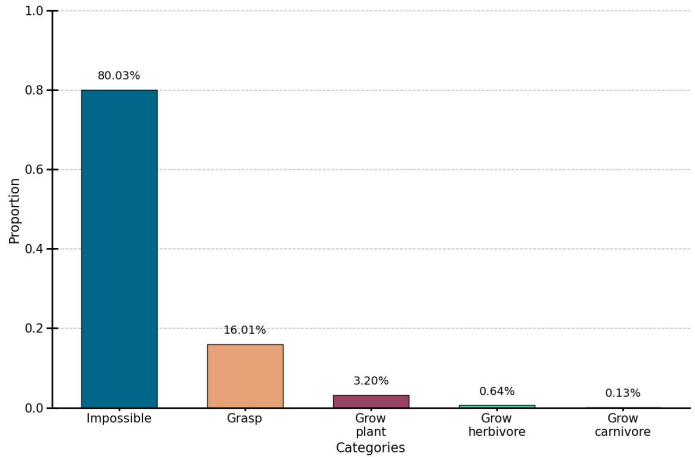
- We train our estimator (*with a cross entropy loss*) every M episodes on a **buffer** of N goals and associate outcome.
- We **sample** goals proportionally to their estimated LP + a random exploration.

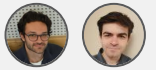


Navigating language goal spaces

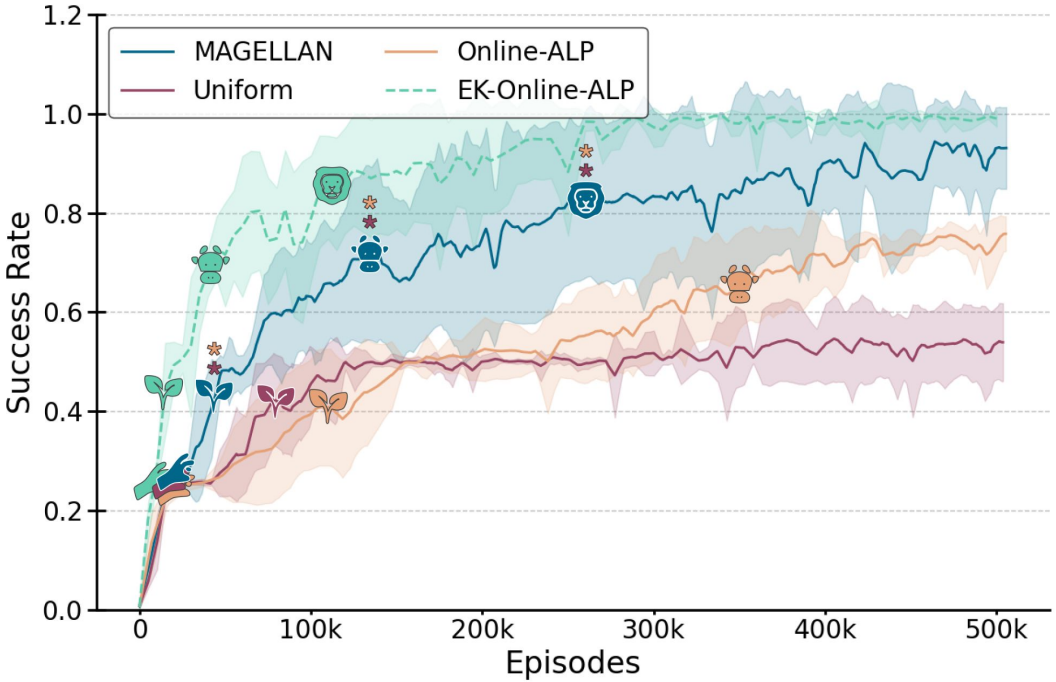


- **Goal = Instruction + Scene initialization**
- Accurately estimating one's competence requires **capturing the environment dynamics**.



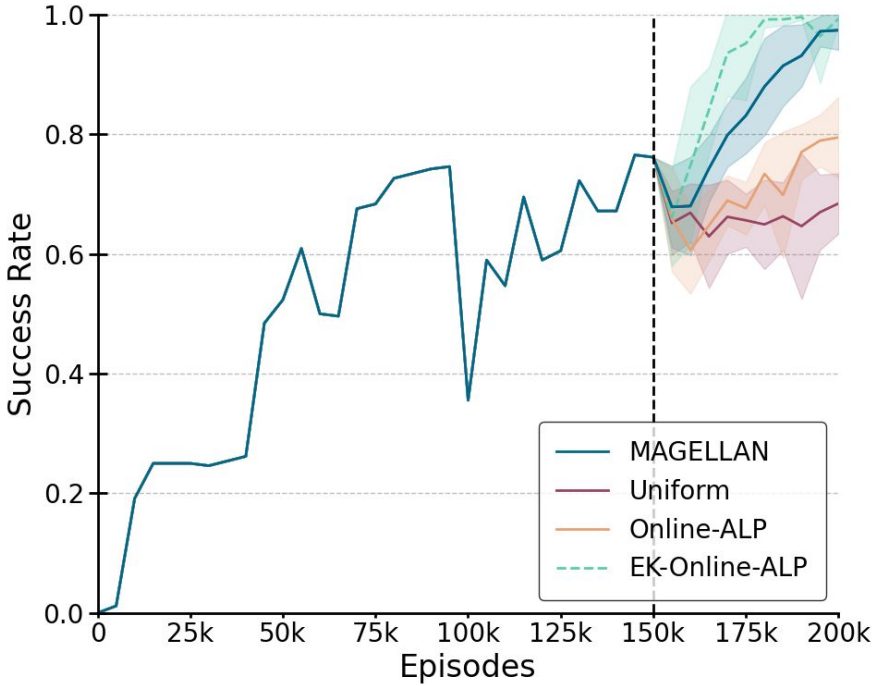


Selecting goals with MAGELLAN

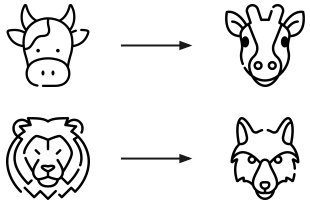


Evolving goal space: towards open-ended learning

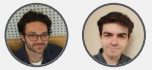
"High LP" scenario



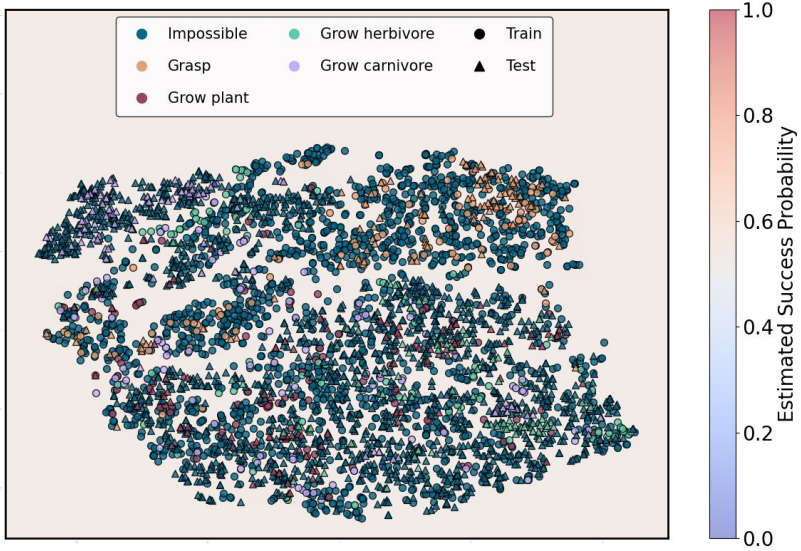
- After 150k steps, we **replace** all the goals from **unseen** ones (which still follow the same inner dynamics):



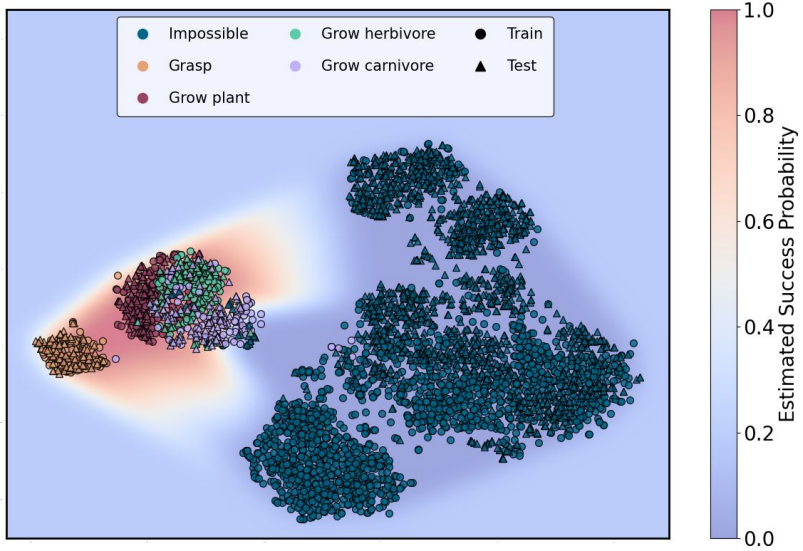
- Online-ALP has all its buffers reset
  => MAGELLAN simply generalizes



MAGELLAN learns to cluster goals



Before training



After training

=> Learning metacognitive monitoring also shapes the LLM's internal representations

Conclusion

In this paper, we showed that LLMs can learn to **estimate their own competence** through interactions.

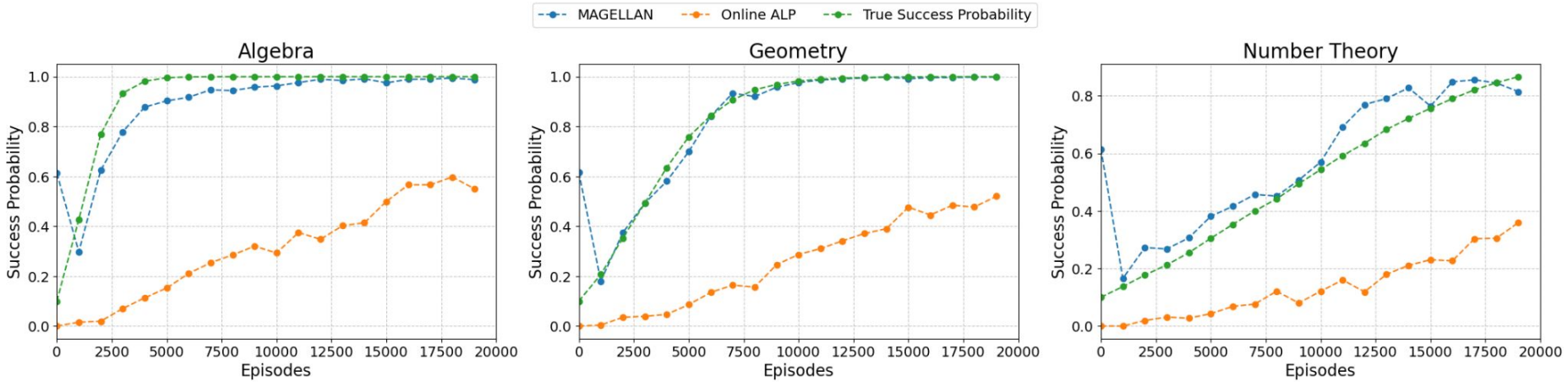
MAGELLAN's utility goes beyond autotelic LLM agents:

Conclusion

In this paper, we showed that LLMs can learn to **estimate their own competence** through interactions.

MAGELLAN's utility goes beyond autotelic LLM agents:

- its efficiency on language goals opens up various applications in **educational technologies**.

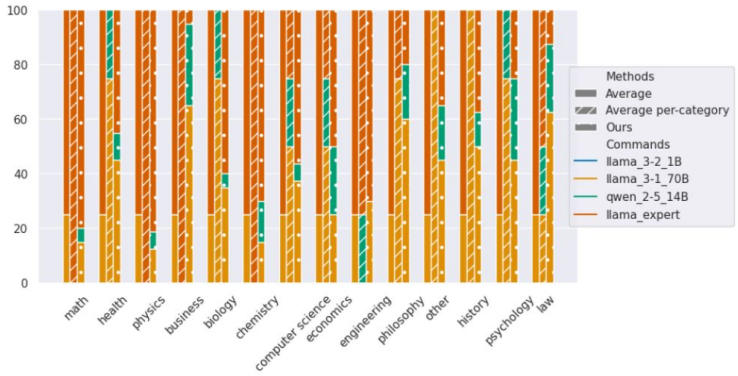


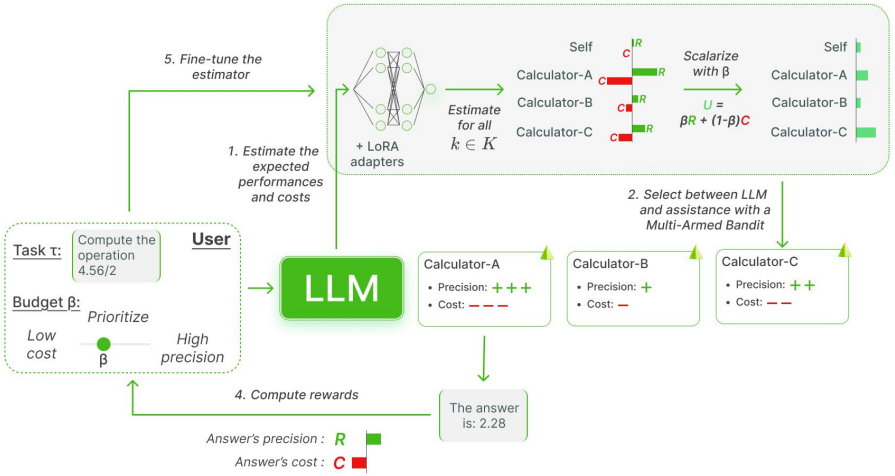
Conclusion

In this paper, we showed that LLMs can learn to **estimate their own competence** through interactions.

MAGELLAN's utility goes beyond autotelic LLM agents:

- its efficiency on language goals opens up various applications in **educational technologies**.
- it can also be used by LLMs to **trigger external assistance** when their estimated functional competence is too low

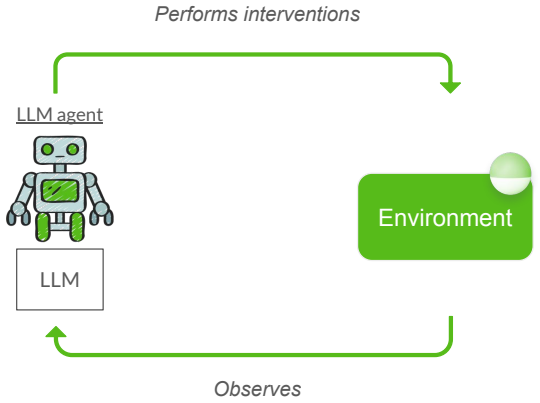




Discussion

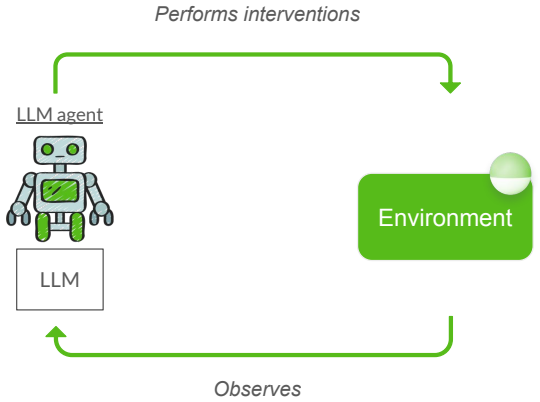
Discussion

This PhD proposed an **embodied autotelic approach to ground LLMs' functional competence**. We enabled LLMs to learn from online interventions.



Discussion

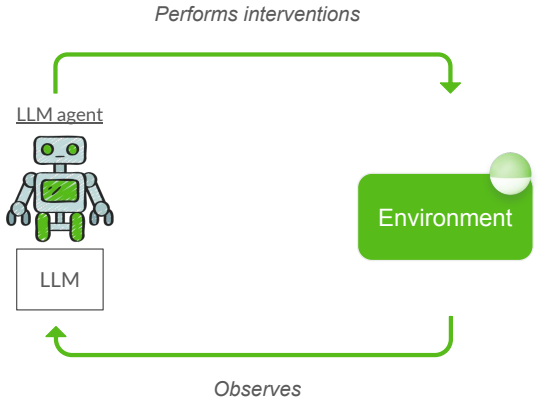
This PhD proposed an **embodied autotelic approach to ground LLMs' functional competence**. We enabled LLMs to learn from online interventions.



1. The first part showed evidence that **RL-based functional grounding aligns LLMs' functional competence** with interactive environments but also hinted **potential broader impact** which remains to be further studied.

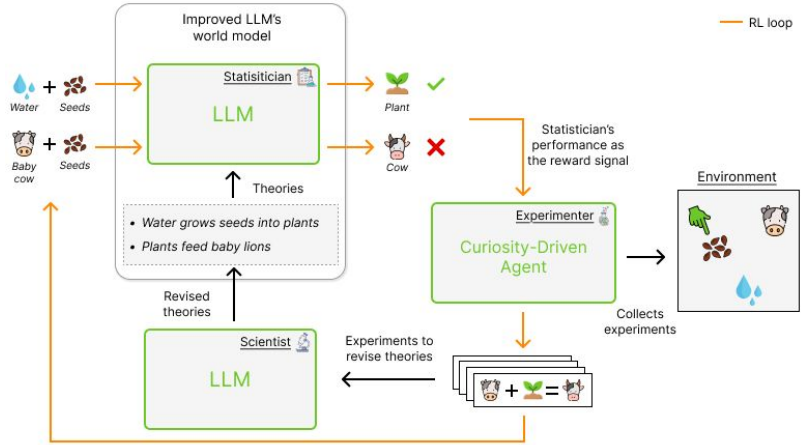
Discussion

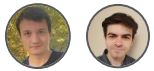
This PhD proposed an **embodied autotelic approach to ground LLMs' functional competence**. We enabled LLMs to learn from online interventions.



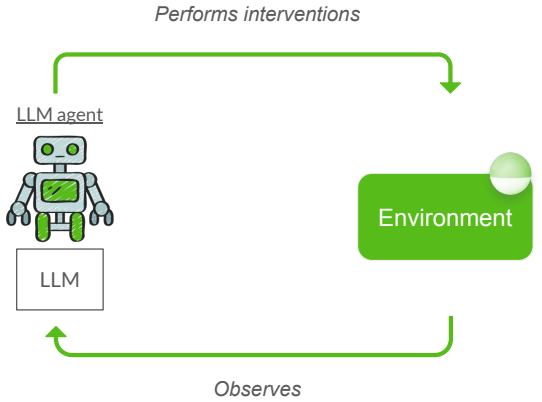
1. The first part showed evidence that **RL-based functional grounding aligns LLMs' functional competence** with interactive environments but also hinted **potential broader impact** which remains to be further studied.

While this talk focused on the **control aspect of functional competence**, our **WorldLLM** approach *(Levy et al., 2024)* studied how to improve **LLMs' predictive abilities**.





Discussion



This PhD proposed an **embodied autotelic approach to ground LLMs' functional competence**. We enabled LLMs to learn from online interventions.

1. The first part showed evidence that **RL-based functional grounding aligns LLMs' functional competence** with interactive environments but also hinted **potential broader impact** which remains to be further studied.

2. In the second part, we discussed the challenges in building **autotelic LLM agents for functional grounding**.

We showed that **metacognitive monitoring** is an essential component of such agents. We also showed that its use goes beyond goal-selection.

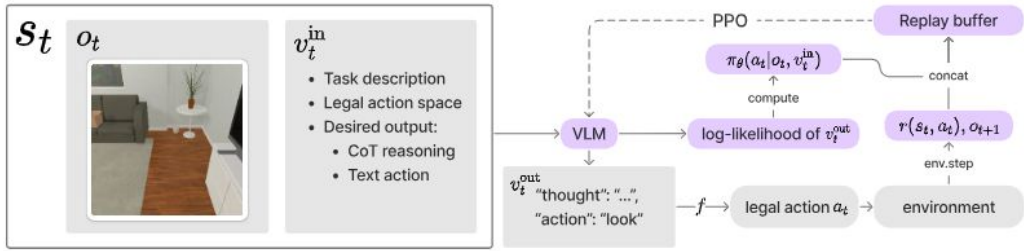
Perspectives

1. **More complex environments**

Our approaches remain to be scaled to **more complex environments** (e.g., multimodal).

=> First attempts at scaling GLAM-like grounding to **VLMs** have been done *(Wang et al., 2024; Aissi et al., 2025; Zhai et al., 2025)*

=> Our approaches might also prove useful for building general-purpose **action models** (e.g., for robotics)



Zhai et al., 2025

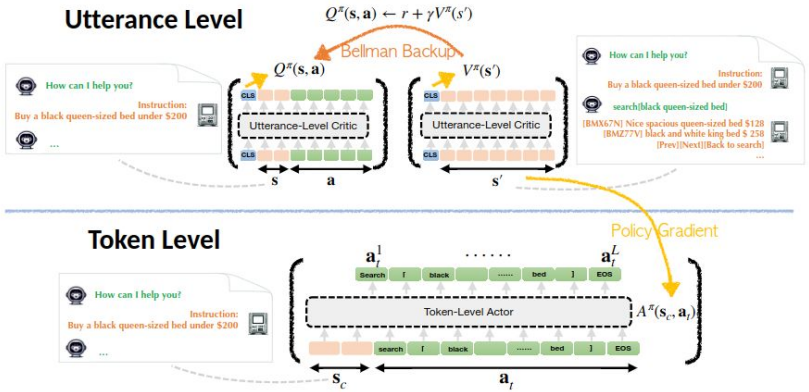
Perspectives

1. More complex environments
2. **Reasoning**

Current LLMs also extensively use **reasoning**.

=> Studying its link to **functional competence** and how to ground reasoning.

=> How about **credit assignment**?



Zhou et al., 2024

Perspectives

1. More complex environments
2. Reasoning
3. **Causal models**

Can LLMs capture **causal models of the world**? *(Hao et al., 2023a; Li et al., 2023a; Vafa et al., 2024; Ding et al., 2025; Ying et al., 2025)*.

=> Functional grounding and metacognitive monitoring **shape** internal representations towards this.

=> Do the theories from WorldLLM lead to **causal inference**?

=> Modeling **other agents or humans** (i.e., Th. of Mind) through online interactions

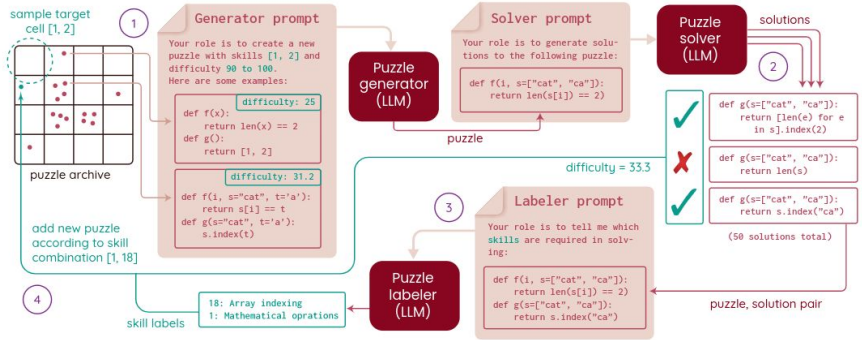
Perspectives

1. More complex environments
2. Reasoning
3. Causal models
4. **Goal generation**

We assumed goals already generated along with a reward function.
=>The next step is to **generate goals**.

=> It can enable to go beyond datasets.

=> **MAGELLAN could drive a generator model.**



ACES (Pourcel et al., 2024)

Perspectives

1. More complex environments
2. Reasoning
3. Causal models
4. Goal generation
5. **Safety and alignment**

A key challenge of the current large use of LLMs is to **align** their knowledge and behaviour to the **end-users and their world**.

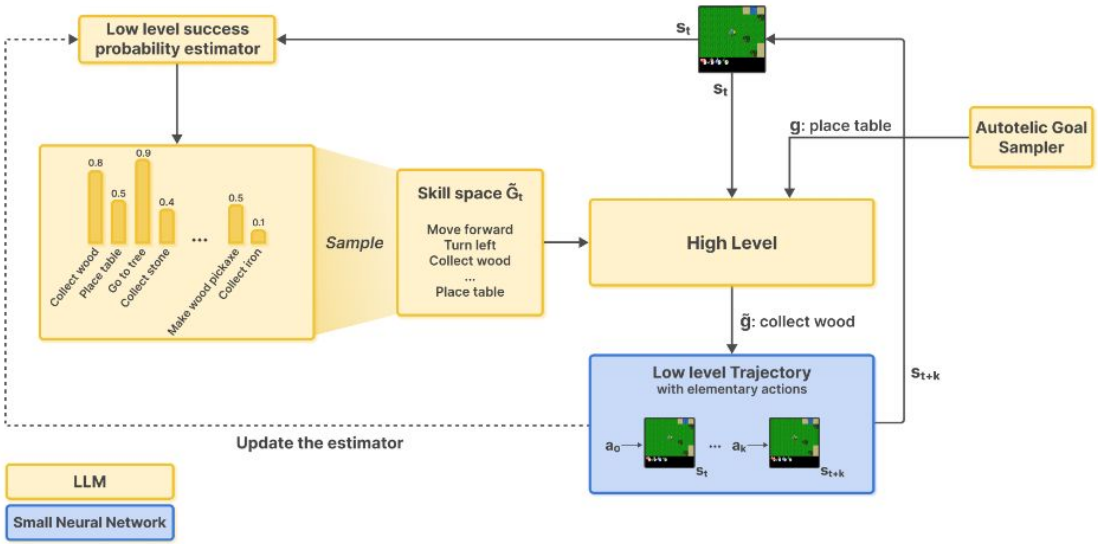
Another important step towards an increased **safety** of current LLMs is developing their **metacognitive abilities**.

=> MAGELLAN is a step towards this, but broader metacognitive abilities remain to be studied.

Perspectives

1. More complex environments
2. Reasoning
3. Causal models
4. Goal generation
5. Safety and alignment
6. **Autotelic RL**

This PhD also contributed to improving existing **autotelic RL agents**.



HERAKLES (Carta et al., 2025)

Thank you !

*And thanks to my **collaborators** on these works:*

- *Thomas Carta (Inria)*
- *Pierre-Yves Oudeyer (Inria)*
- *Loris Gaven (Inria)*
- *Guillaume Levy (ex Inria)*
- *Nicolas Yax (Inria / ENS)*
- *Cédric Colas (MIT / Inria)*
- *Thomas Wolf (Hugging Face)*
- *Sylvain Lamprier (Univ. Angers)*
- *Salim Aissi (ISIR)*
- *Olivier Sigaud (ISIR)*
- *Laure Soulier (ISIR)*
- *Nicolas Thome (ISIR)*