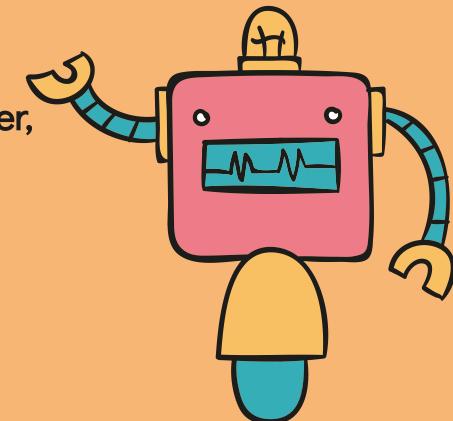
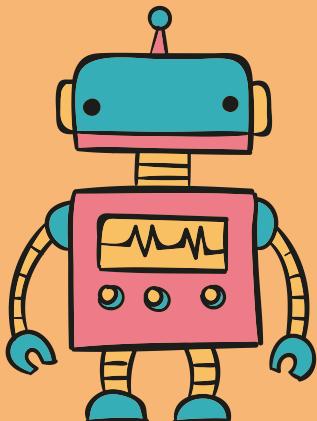


Towards a Vygotskian Autotelic Artificial Intelligence

The Internalization of Cognitive Tools from
Rich Socio-Cultural Worlds

Cédric Colas*, Tristan Karch*, Thomas Carta, Clément Moulin-Frier,
Pierre-Yves Oudeyer (Inria, Univ. de Bordeaux)

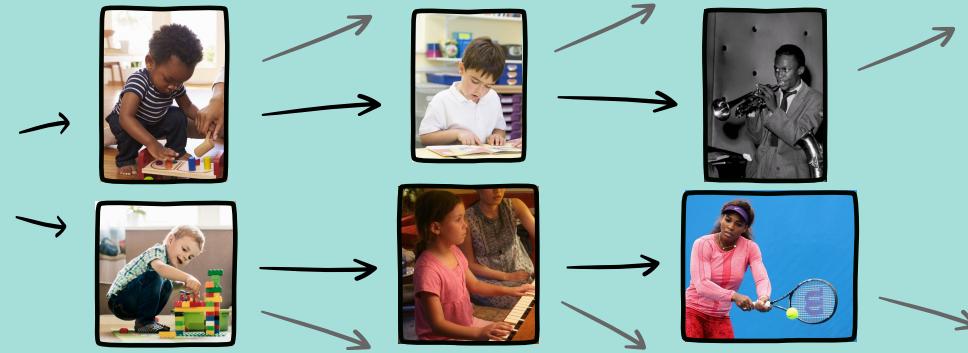
* equal contribution - cedric.colas@inria.fr



Why Am I Here?

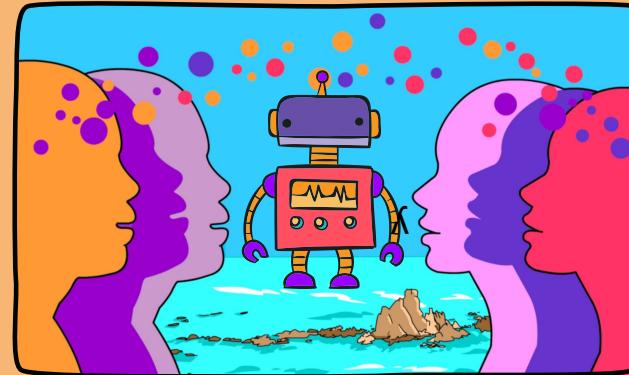
Research objective:

building an artificial open-ended skill discovery process.



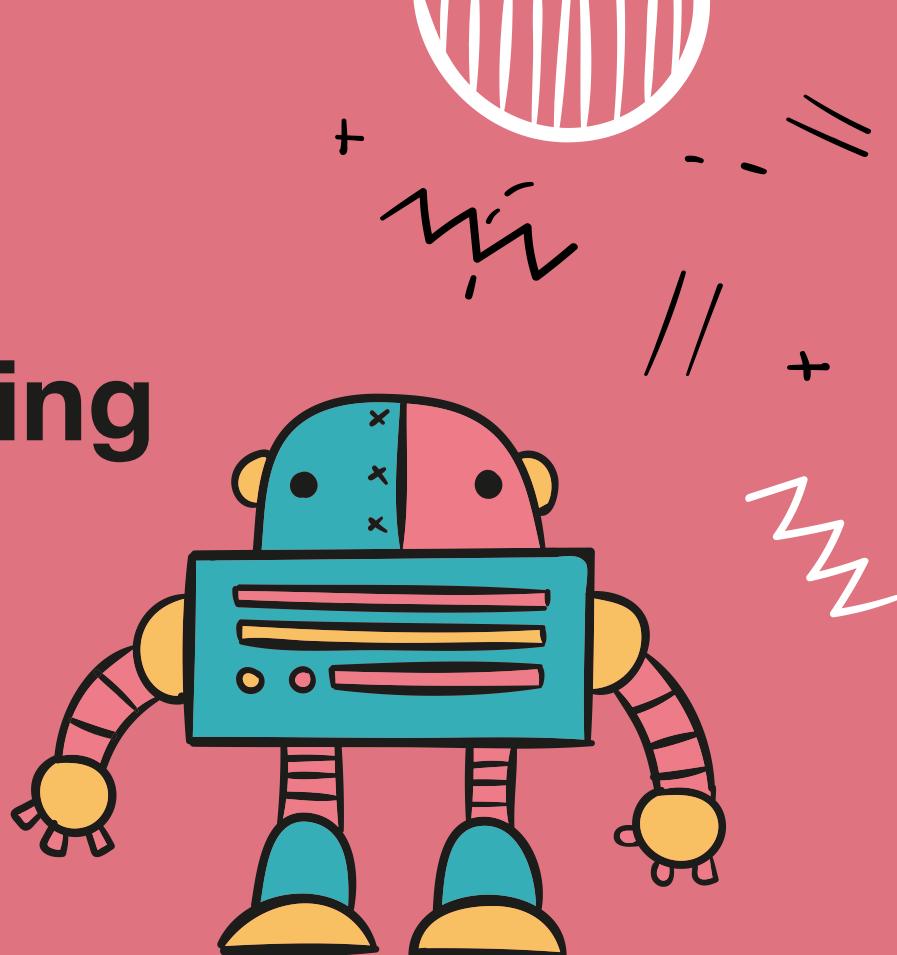
To move forward in AI, we need to immerse agents into rich socio-cultural worlds, to let them turn social interactions into cognitive tools.

We need LCM!

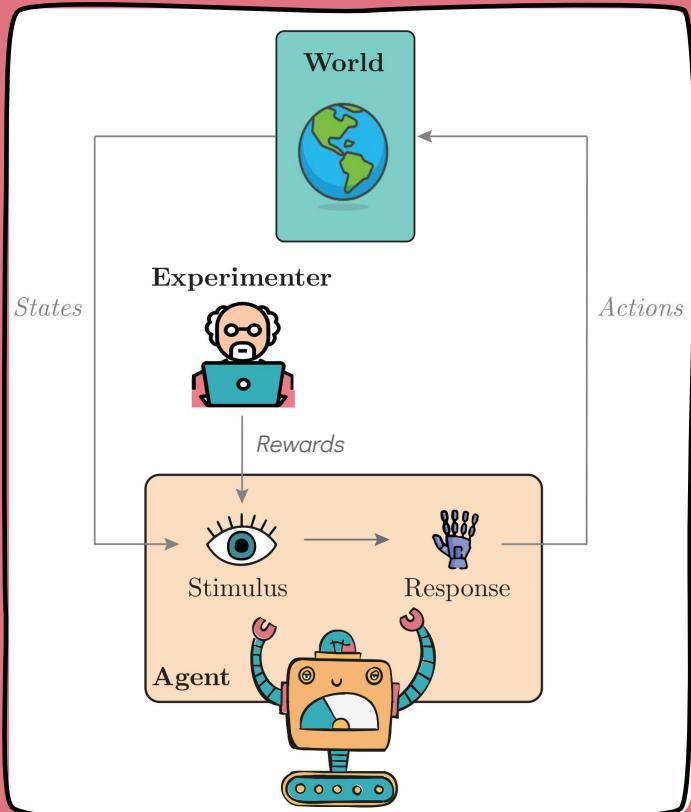


Piagetian Learning

Intrinsically motivated goal directed learners.



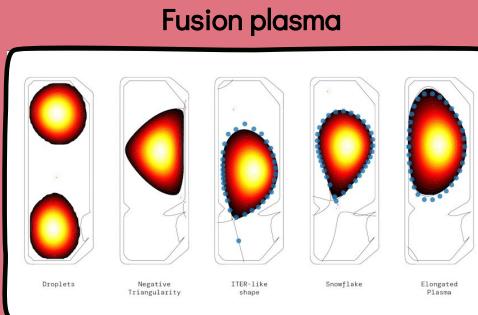
Reinforcement Learning (RL)



The RL framework



Go



Fusion plasma

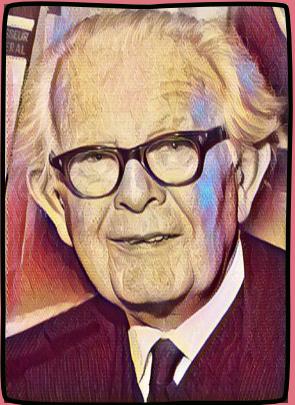


Atari Games

... but they can only solve
a specific predefined task at a time.



Intrinsically Motivated Learners



Jean Piaget
(1896-1980)



Credit: Francis Vachon

Intrinsic motivations

Defined by psychologists
(Berlyne, 1950/1966; Czikszentmihalyi, 1990;
Ryan & Deci, 2000; Kidd, 2012).

Implemented by RL and developmental
robotics researchers (Schmidhuber, 1991;
Oudeyer & Kaplan, 2004, 2007).

Scaled with deep learning
(Bellemare, 2016; Pathak, 2018; Burda, 2019).

Autotelic Learners

Autotelic Learning

(Steels, 2004; Colas, 2021)

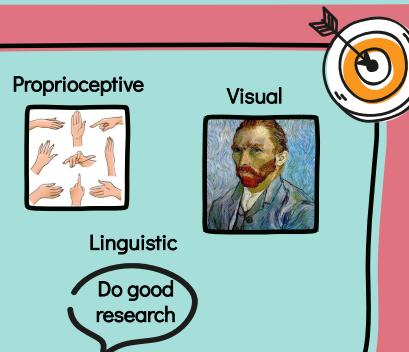
Autotelic agents are intrinsically motivated to learn to represent, generate, pursue and master their own goals.

Goals

“A goal is a cognitive representation of a future object that the organism is committed to approach or avoid.”

(Elliott & Fryer, 2008)

AI: goal representation + reward function

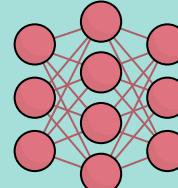


Repertoire of Skills

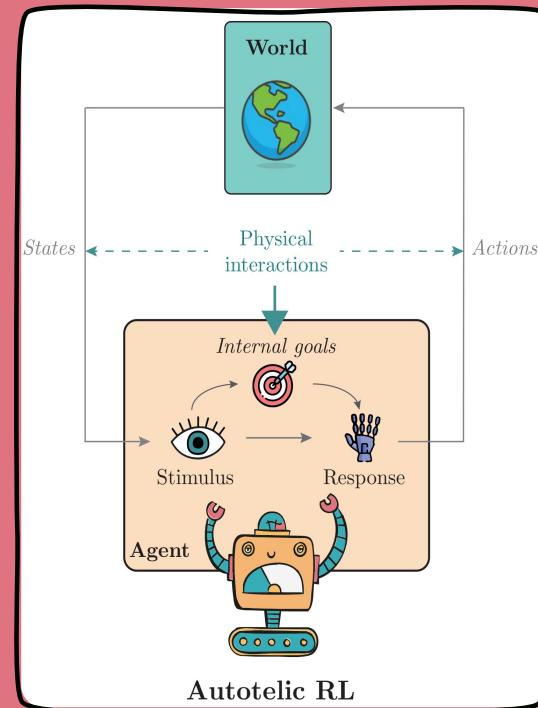


set of goal representations

+

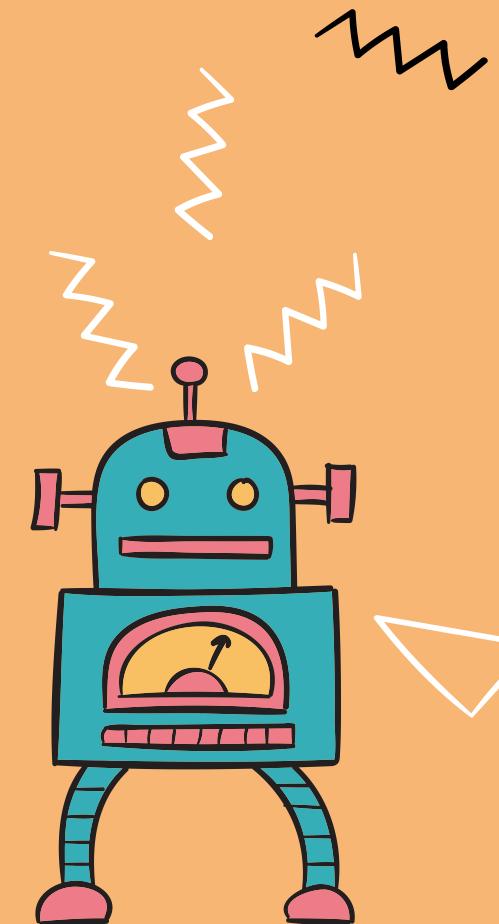


goal-directed behaviors



Vygotskian Learning

Turning socio-cultural interactions into cognitive tools.



Social Autotelic Learners

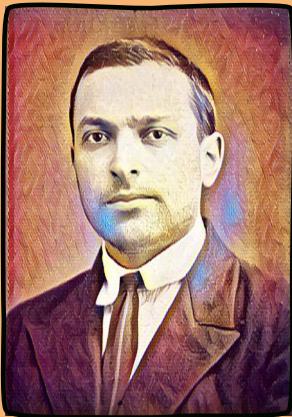


Open-ended
repertoire of skills



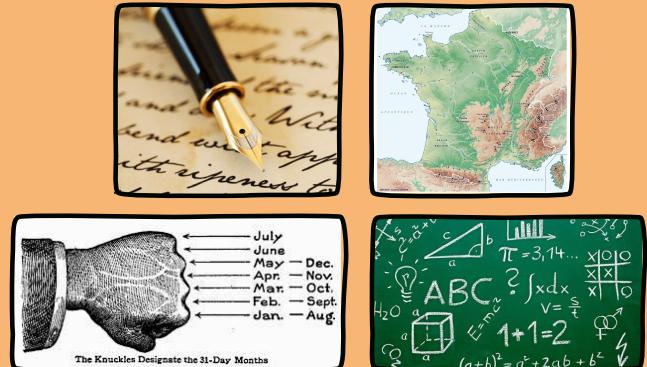
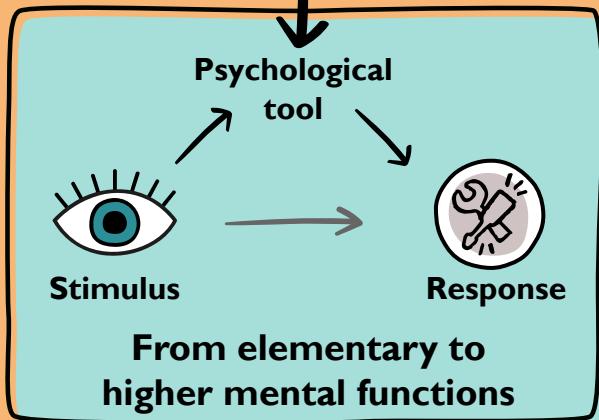
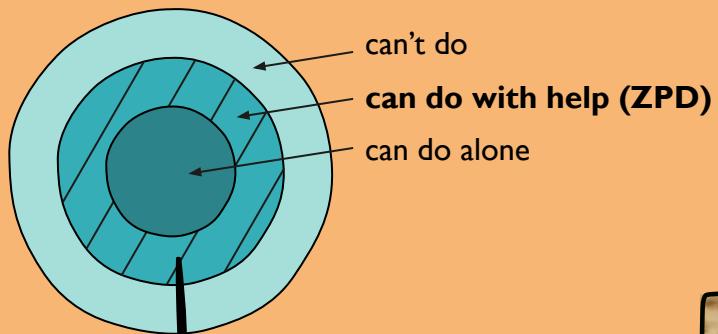
Socio-Cultural Situatedness
Humans learn from others in a rich socio-cultural world.

Vygotskian View on Human Development



Lev Vygotsky
(1896-1934)

Zone of Proximal Development (ZPD)



Examples of psychological tool

Language as a Cognitive Tool



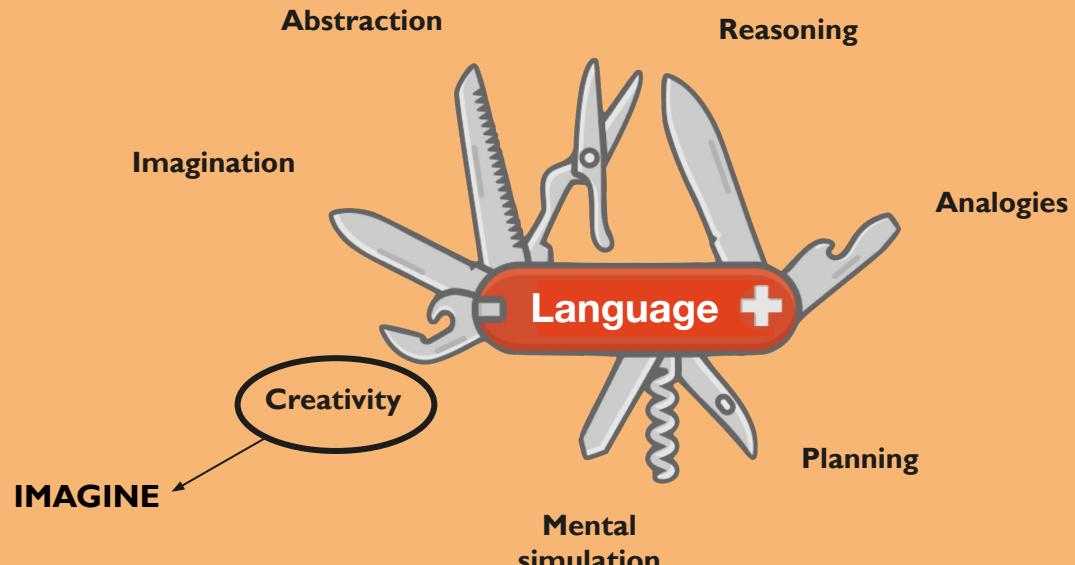
Jean Piaget

Egocentric speech is a sign of cognitive immaturity.
inside → outside

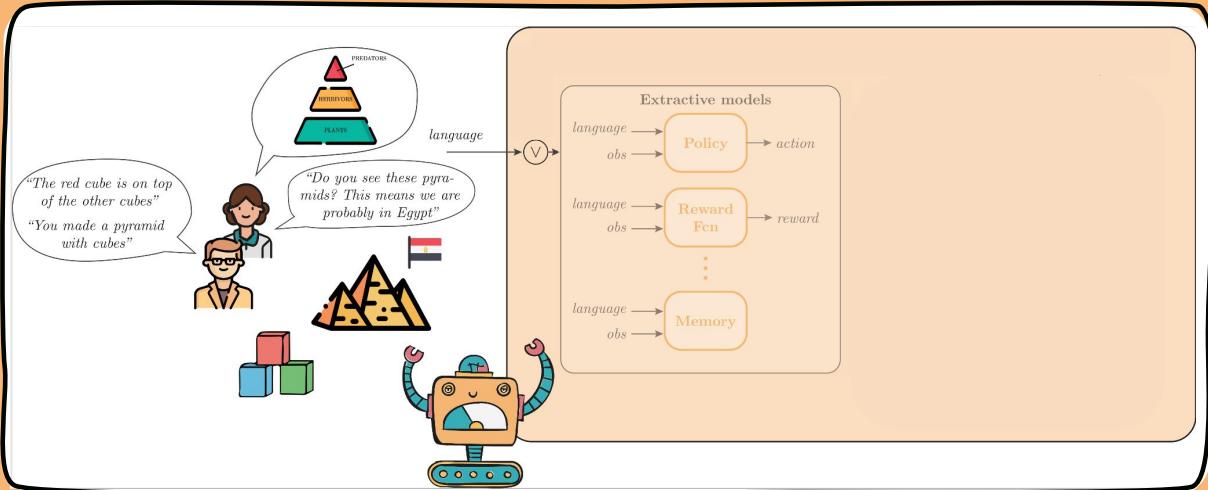
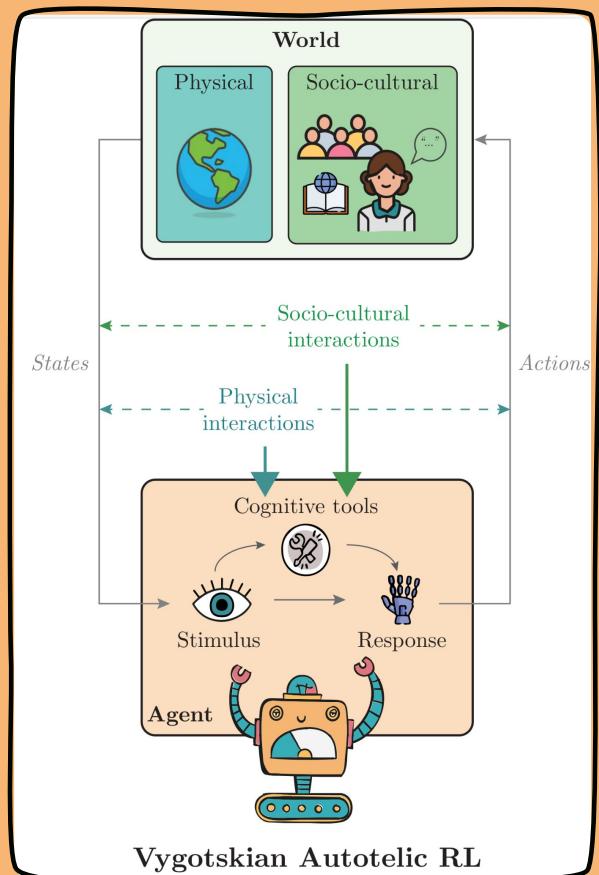


Lev Vygotsky

Egocentric speech is the internalization of social speech.
outside → inside



Vygotskian Autotelic AI

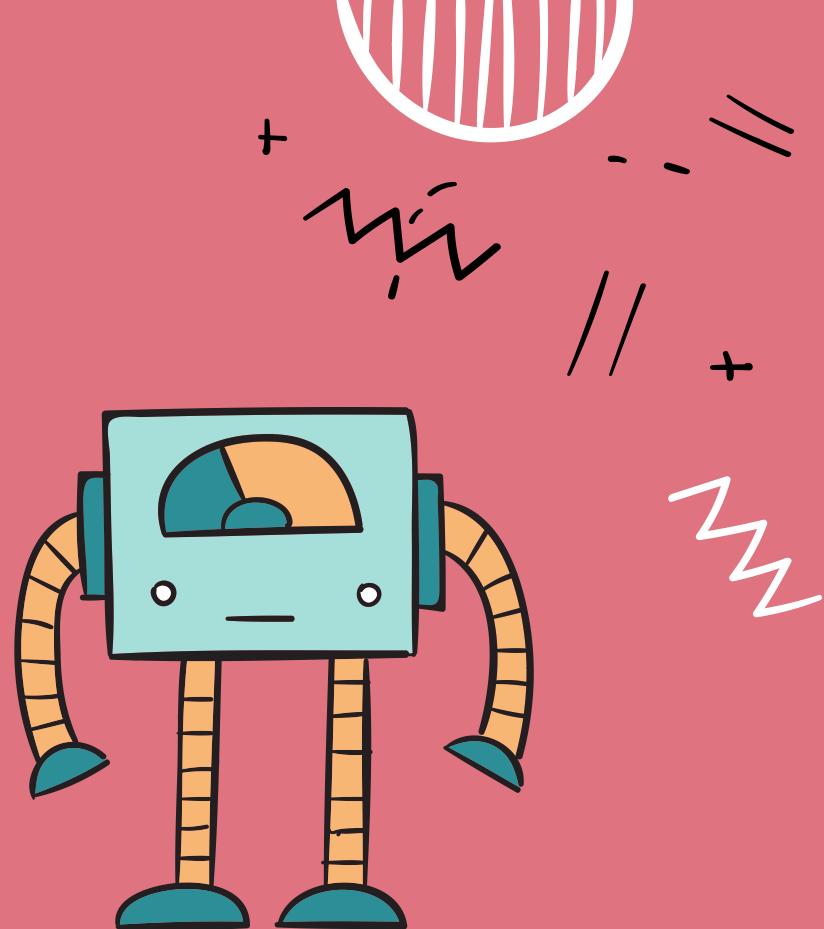


Extractive models exploit linguistic structure and content to reshape inner computations.

Productive models internalize models of social partners to generate feedback through a form of inner speech.

IMAGINE

*Linguistic creativity for exploration
and generalization.*



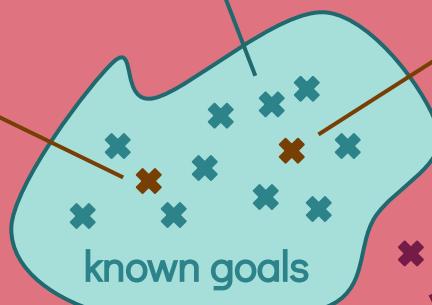
Towards Out-of-Distribution Goal Generation

+

-



in-distribution



out-of-distribution



Linguistic Creativity

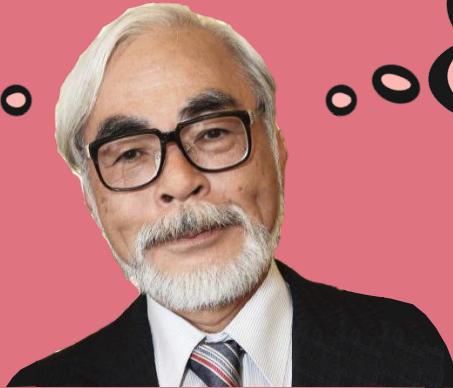
Creativity = novelty x appropriateness.

(Simonton, 2012)

Linguistic creativity: generate new utterances (novelty) from a known grammar/known constructions (appropriateness).

(Chomsky, 1957; Hoffmann, 2020)

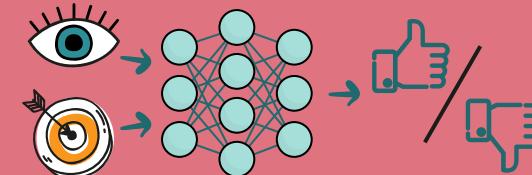
cat + bus
= cat-bus!



Internalization of Linguistic Goals

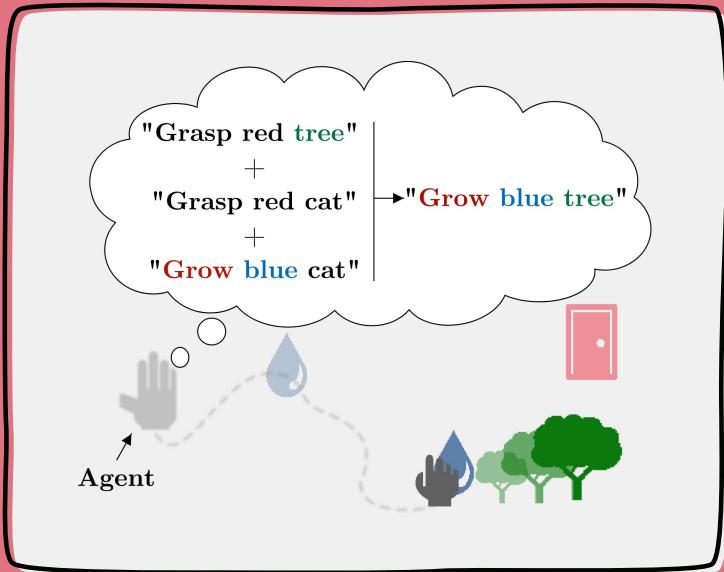


Autotelic exploration with
social descriptions



Learn a goal-conditioned reward function
(extractive model)
(Bahdanau, 2019)

Language as a Cognitive Tool to Imagine Goals + -



Creative autotelic exploration

Idea

Use language compositionality to systematically compose novel, out-of-distribution goals.

Internalized goal generation and reward functions let the agent train autonomously.

New Cognitive Functions

Extractive
models

Abstraction

Agents now represent cultural concepts such as colors (red), objects (bonsai) and categories (animals).

Systematic Generalization

Agents can represent and achieve new goals by systematic generalization.

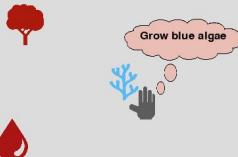
Productive
models

Creative Exploration

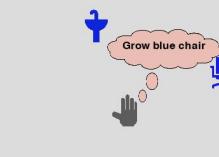
Agents can generate creative goals that drive their exploration.

Cultural Attention

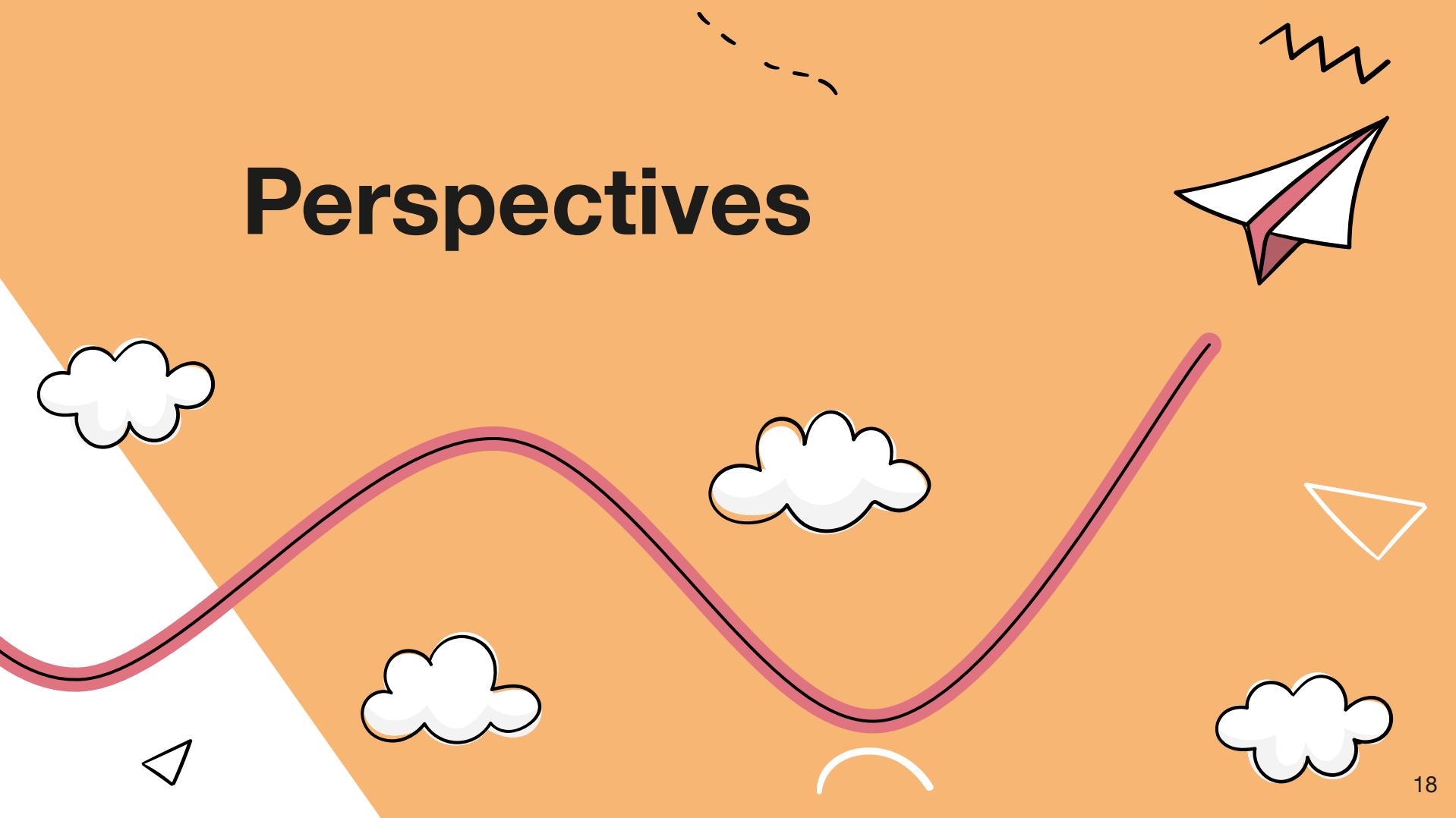
Agents invent goals based on culturally-transmitted concepts worth of attention (object, attributes, object interactions).



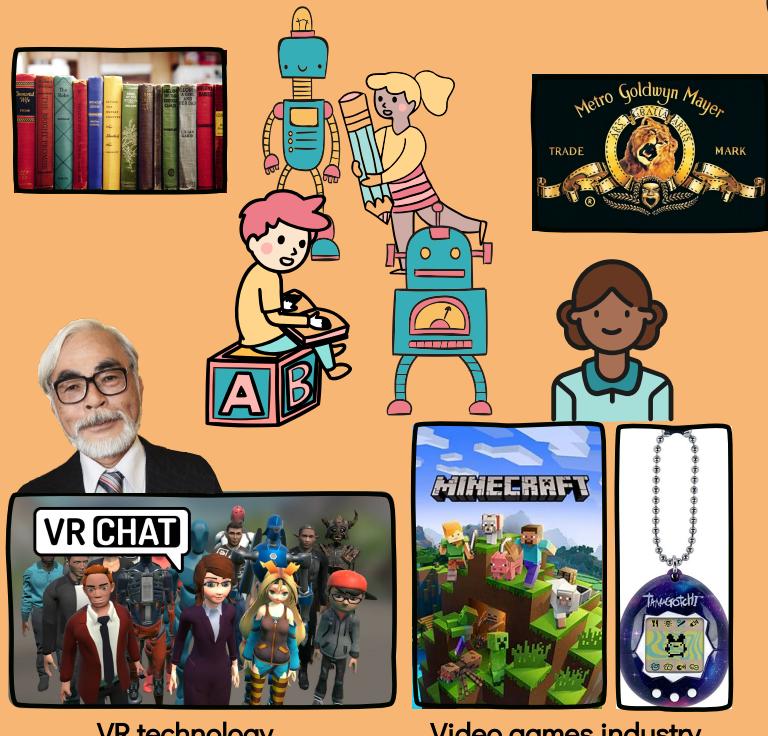
The agent imagines it could feed plants and furniture.



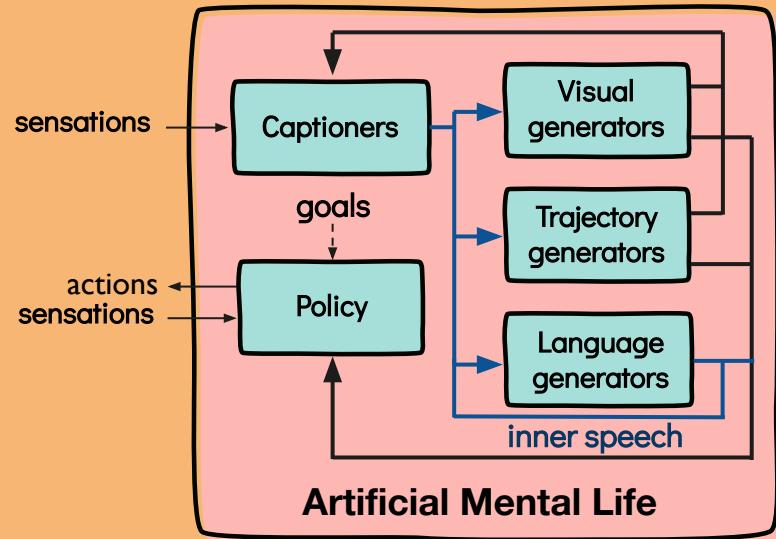
Perspectives



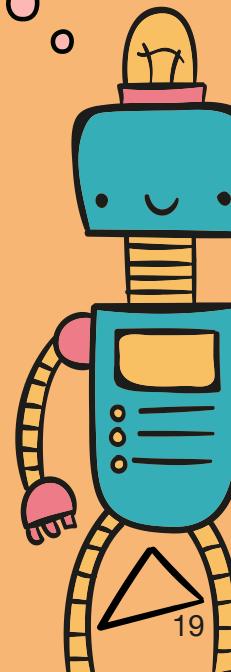
Challenges



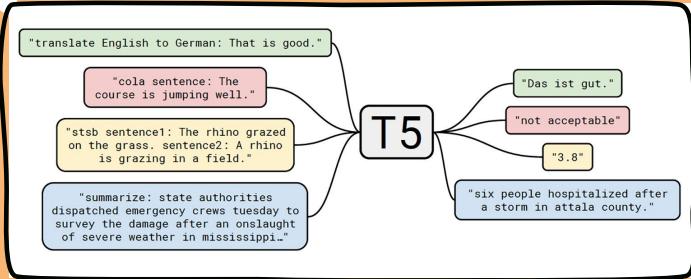
Challenge #1
Immersion in rich socio-cultural worlds



Challenge #2
Artificial mental life with extractive and productive models



Challenges



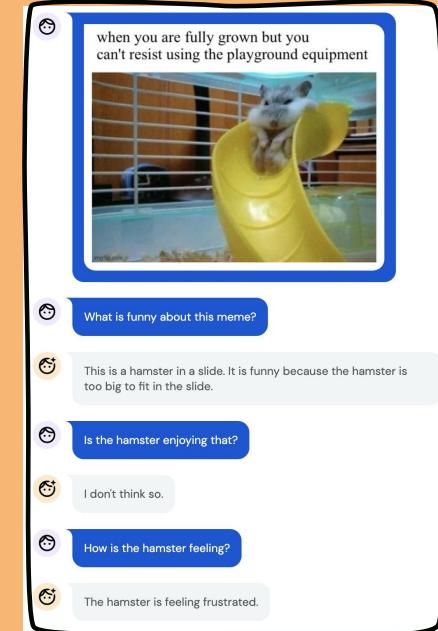
T5 model (Raffel et al., 2019)

But also:

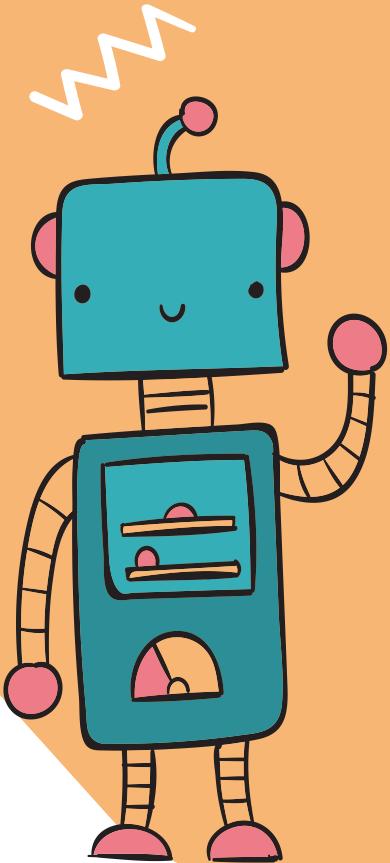
- Planning (Huang et al., 2022, Ahn et al., 2022)
- Common sense (West et al., 2022)
- Cultural differences (Hershcovich et al., 2022)
- Moral values (Schramowski et al., 2022)
- ...

Joke understanding by the Flamingo model (grey)
(Alayrac et al., 2022)

From @antoine77340



Challenge #3
Leveraging large language models as cultural models



Thanks!

Collaborators:



Tristan Karch



Thomas Carta



Clément
Moulin-Frier



Pierre-Yves Oudeyer

Contact:

cedric.colas@inria.fr

<https://ccolas.github.io/>

<https://github.com/flowersteam>