On the challenges raised by robots powered by Artificial Intelligence

Fulvio Mastrogiovanni

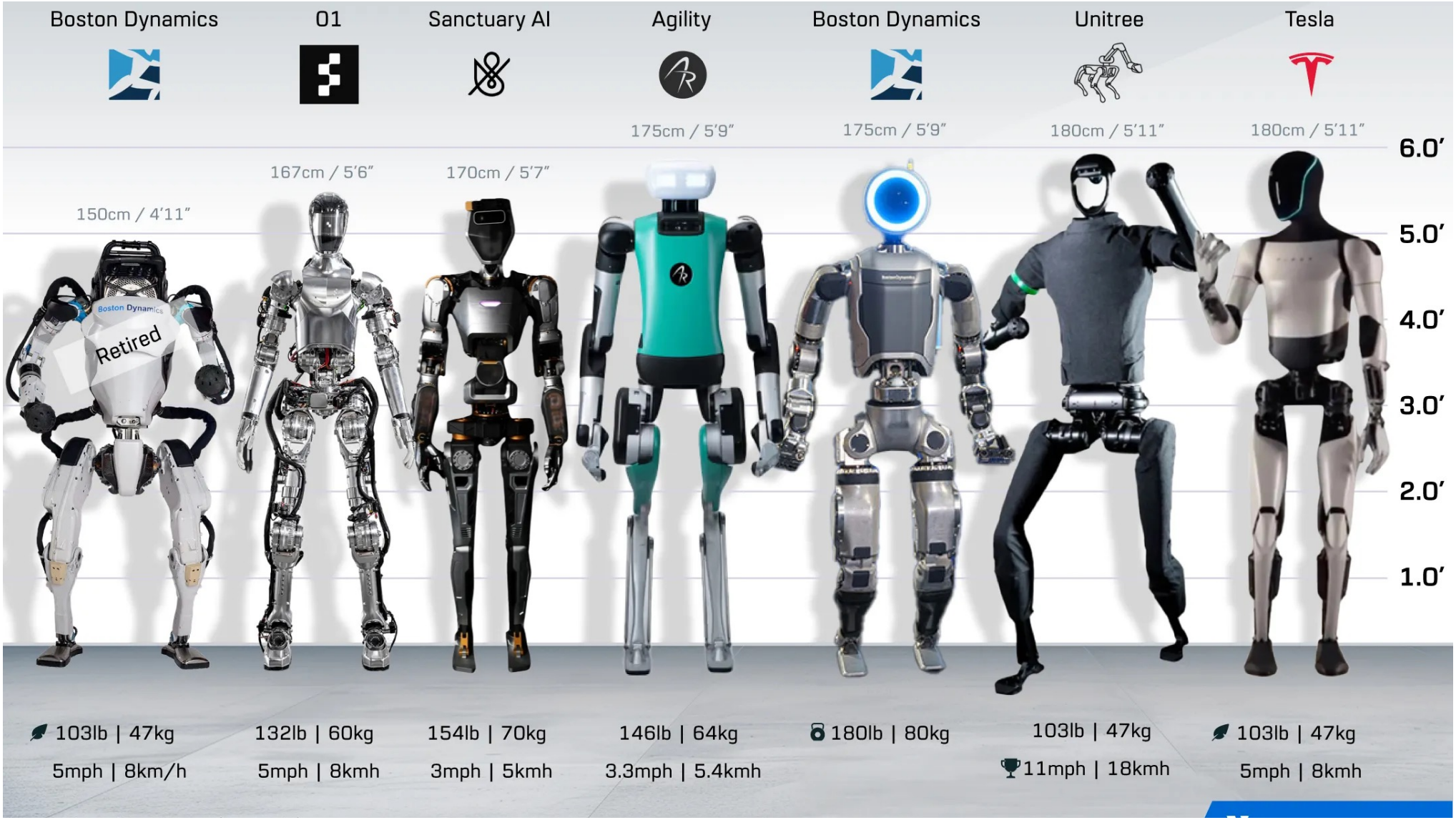
Professor of Robotics and Artificial Intelligence, University of Genoa
Founder and Chief Science Officer, Teseo srl
Coordinator of the Scientific Committee, Digital Innovation Hub Liguria

10^{-9}

ROBOTS OPERATING

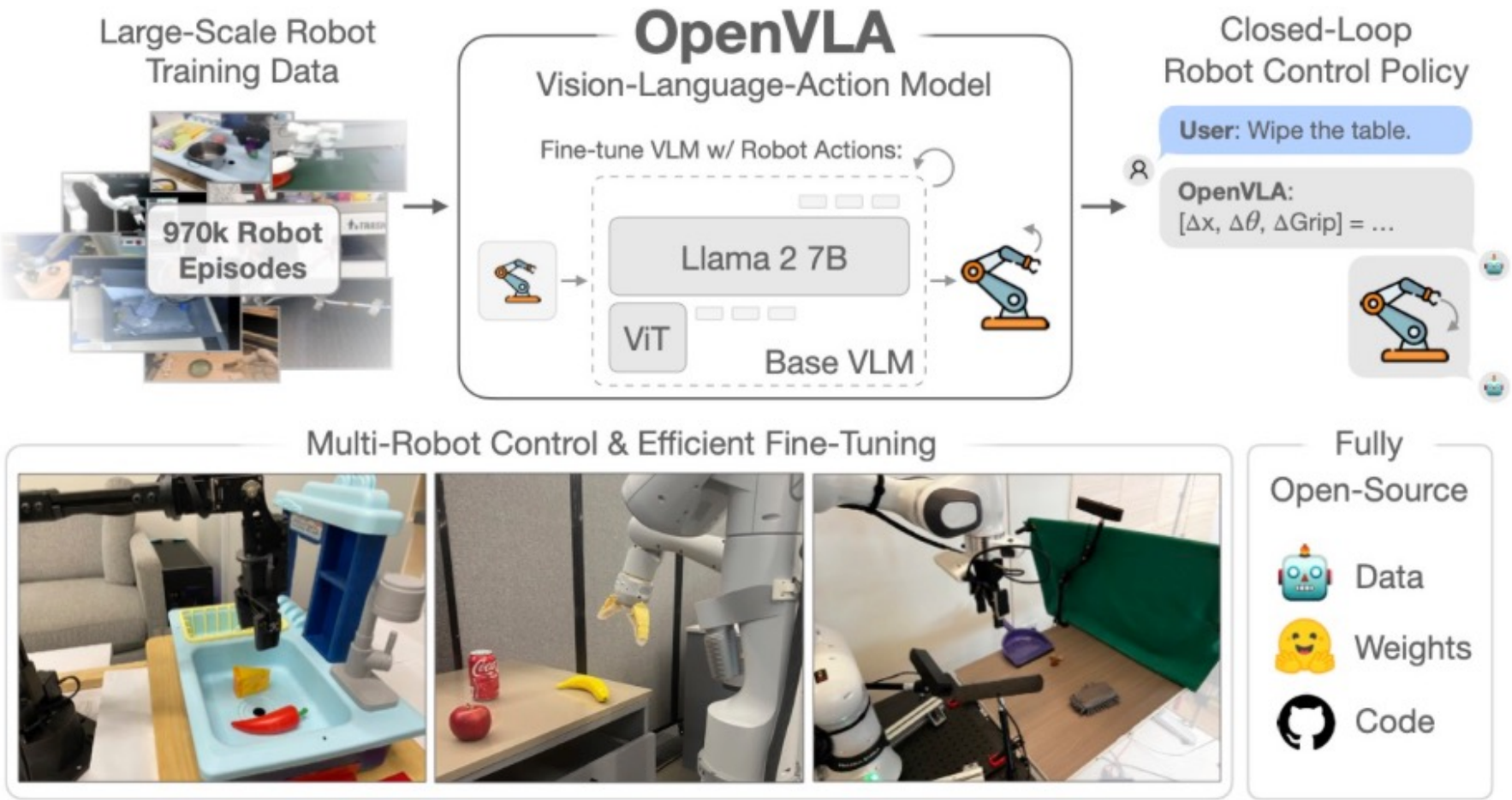
./ IN AUTONOMY

./ JOINTLY WITH HUMANS


Boston Dynamics
01
Sanctuary AI
Agility
Boston Dynamics
Unitree
Tesla
150cm / 4'11"
167cm / 5'6"
170cm / 5'7"
175cm / 5'9"
175cm / 5'9"
180cm / 5'11"
180cm / 5'11"
6.0'
5.0'
4.0'
3.0'
2.0'
1.0'
Retired
103lb | 47kg
5mph | 8km/h
132lb | 60kg
5mph | 8kmh
154lb | 70kg
3mph | 5kmh
146lb | 64kg
3.3mph | 5.4kmh
180lb | 80kg
103lb | 47kg
11mph | 18kmh
103lb | 47kg
5mph | 8kmh

ROBOT INTELLIGENCE = BODY + AI

ROBOT INTELLIGENCE = BODY + VLA


Large-Scale Robot Training Data
970k Robot Episodes
OpenVLA
Vision-Language-Action Model
Fine-tune VLM w/ Robot Actions:
Llama 2 7B
ViT
Base VLM
Closed-Loop Robot Control Policy
User: Wipe the table.
OpenVLA:
[Δx, Δθ, ΔGrip] = ...
Multi-Robot Control & Efficient Fine-Tuning
Fully Open-Source
Data
Weights
Code

Put Eggplant into Pot

Put Yellow Corn on Pink Plate

0.16

ENGINEERING

10^{-9}

0.16

LEARNING

ROBOT INTELLIGENCE = F(BODY, IA, ...)

ROBOT INTELLIGENCE = F(BODY, IA, ...)

A. NEWELL (1990)

COGNITIVE ARCHITECTURES FOR ROBOTS

What a robot cognitive architecture should exhibit/

1/ Reactive execution.

2/ Knowledge representation and inference.

3/ Reasoning.

4/ Hierarchical and recursive reasoning.

5/ Learning from experience.

6/ Memory models: semantic memory, episodic memory, mental imagery.

7/ Interaction models.

8/ social models of interaction.

...

Key questions/

Q1/ Are current principles in computation compatible with cognition in robots?

Q2/ What are the (maybe trivial but hidden) assumptions in integrating AI techniques in cognitive architectures for robots?

Q3/ Is robot intelligence a property associated with the algorithms, or is it connected to the cognitive architecture as a whole?

Q4/ Is there a more fundamental description level for robot behaviours?

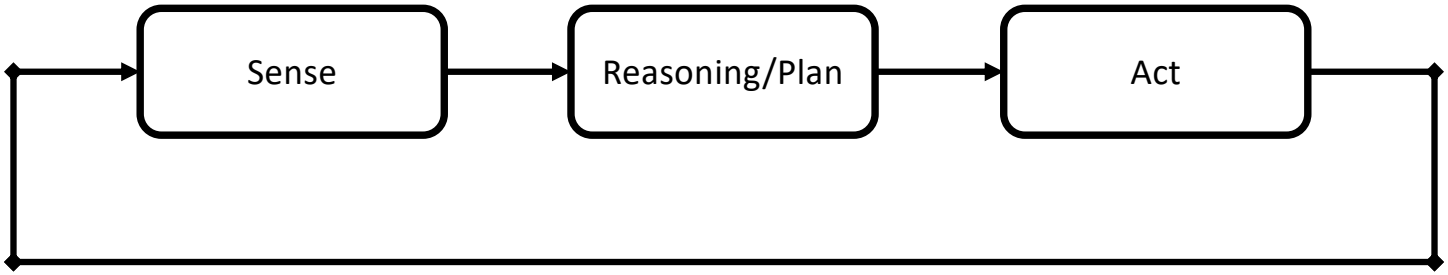
Three+ architectural families/

1/ Sense-Plan-Act.

2/ Reactive.

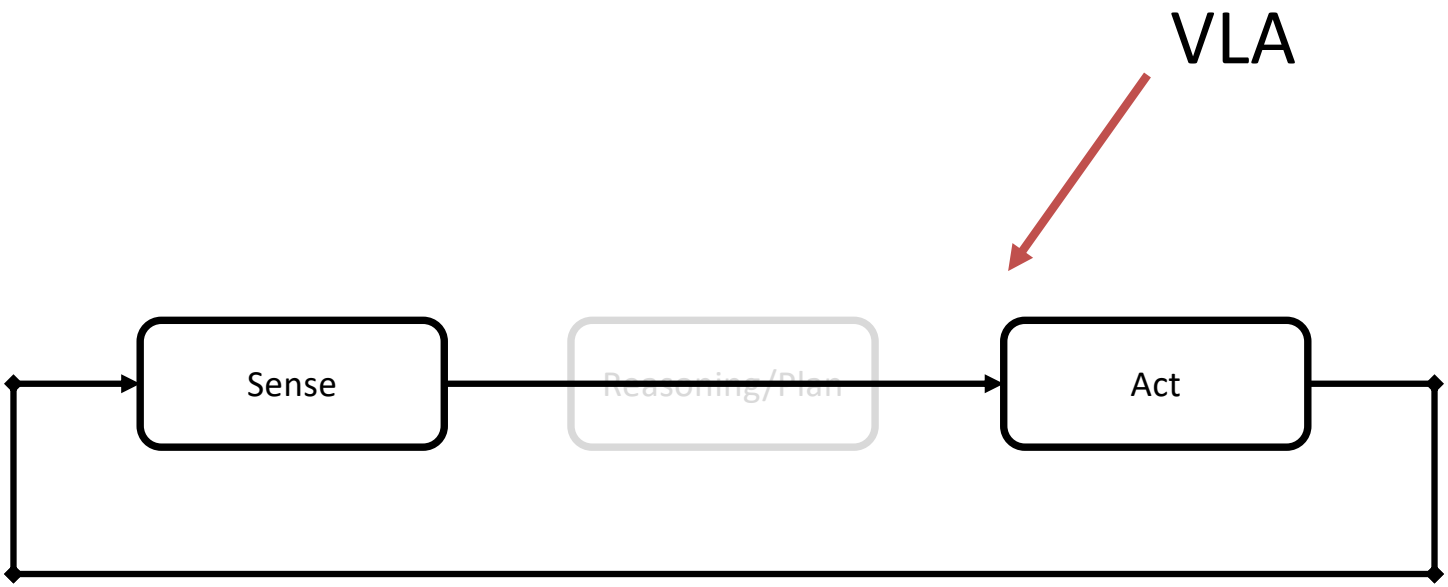
2+/ Behaviour-based.

3/ Hybrid reactive-deliberative (cognitive?).

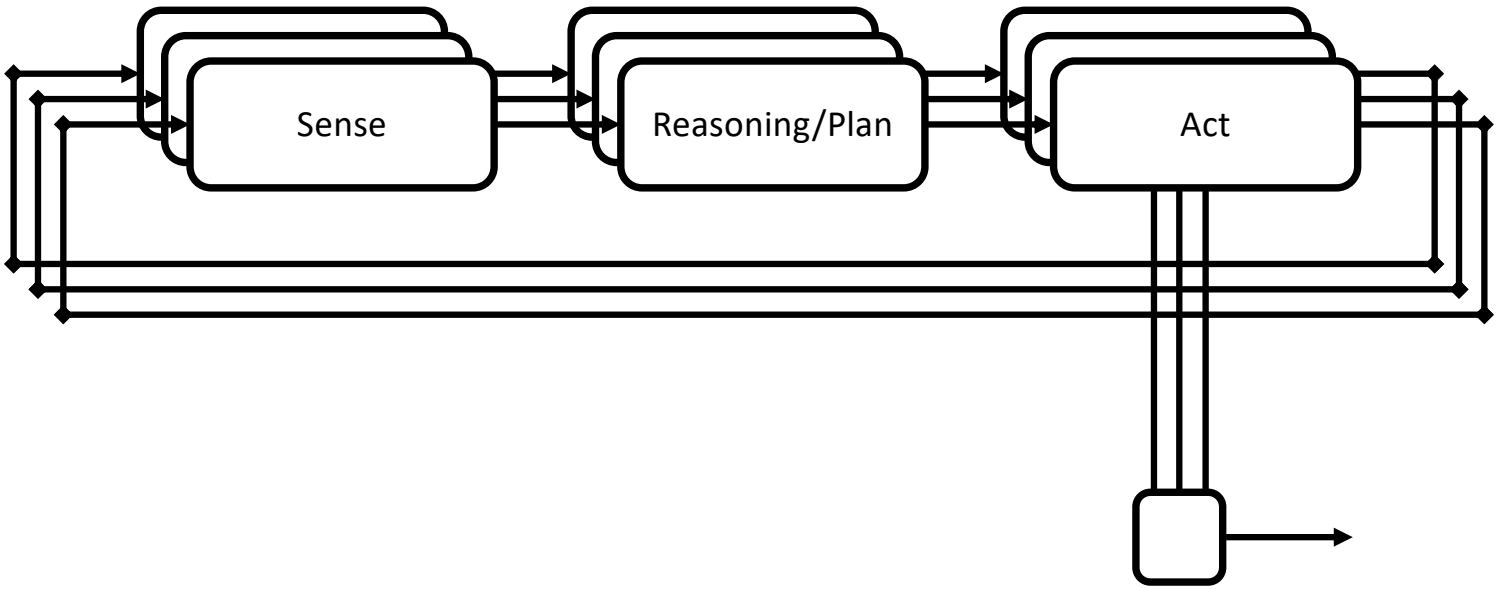


Sense-Plan-Act

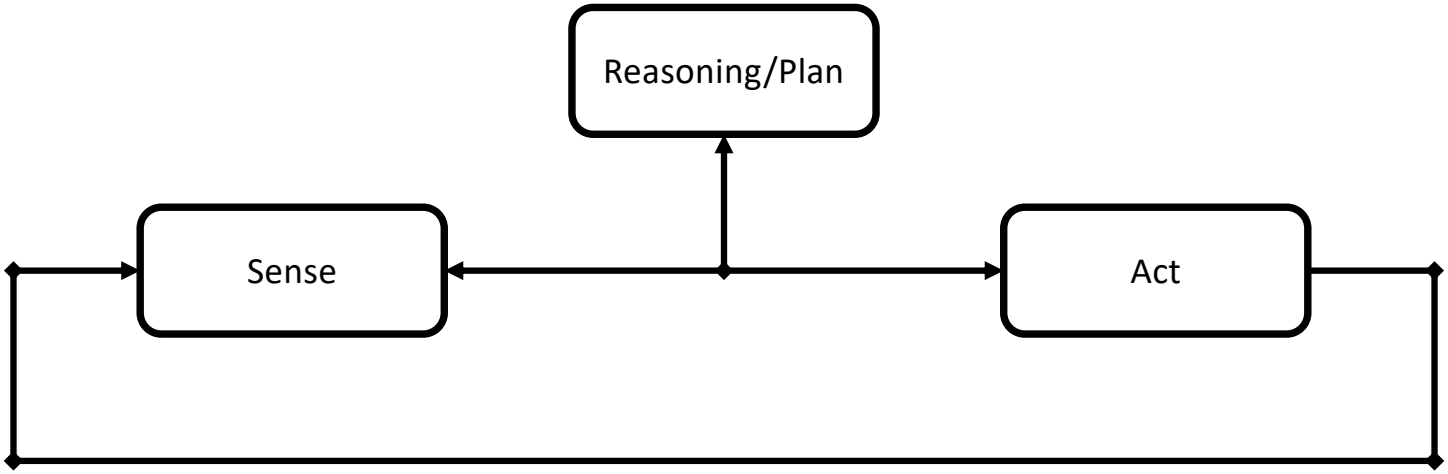
REACTIVE



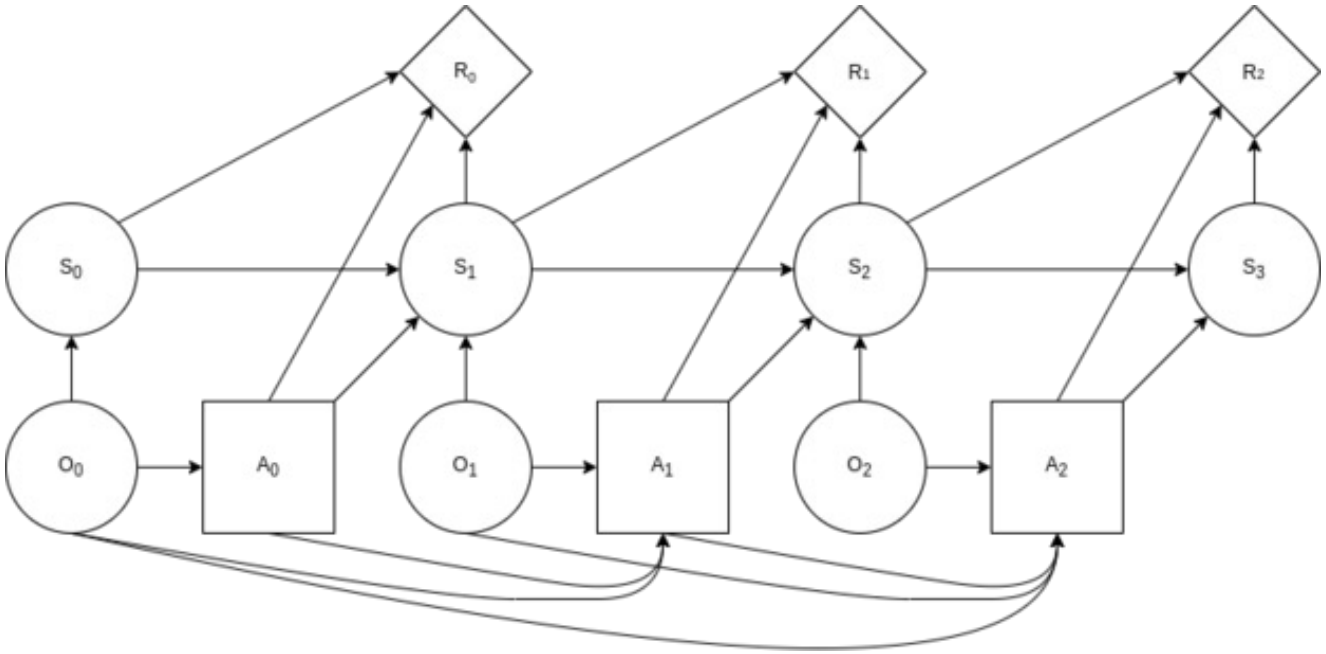
REACTIVE



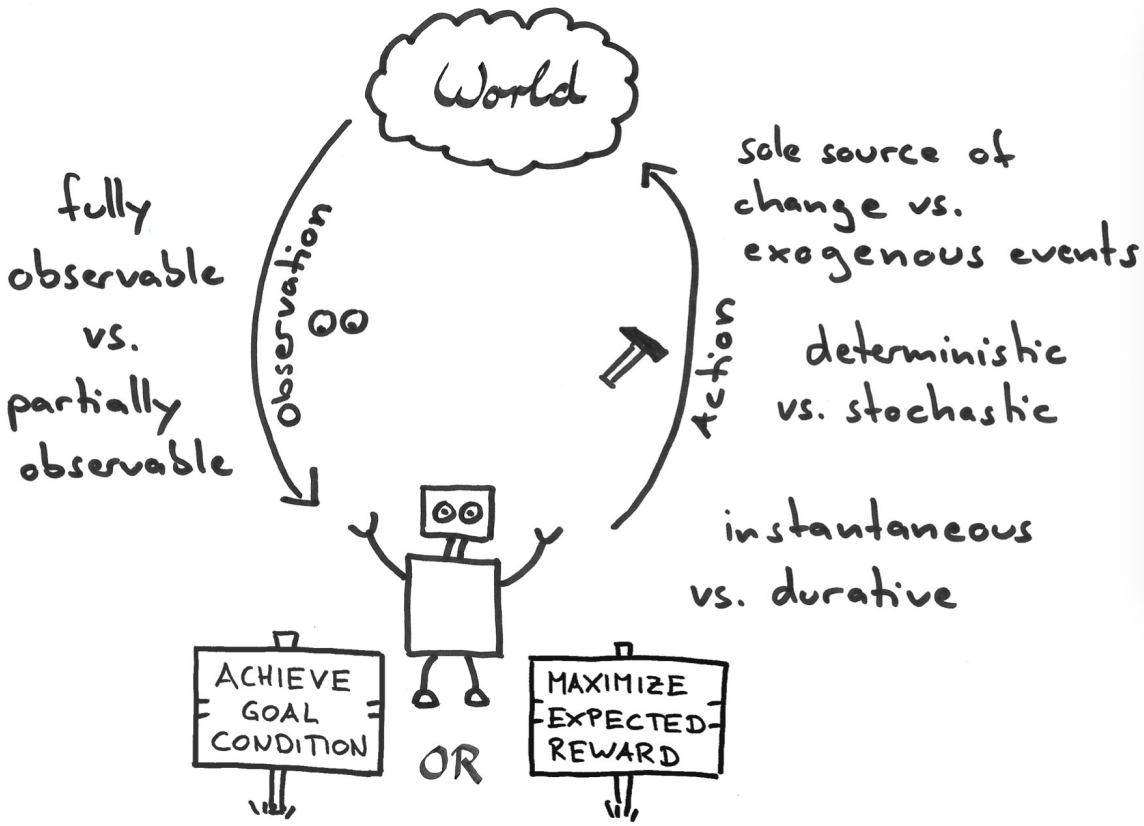
Behaviour-based

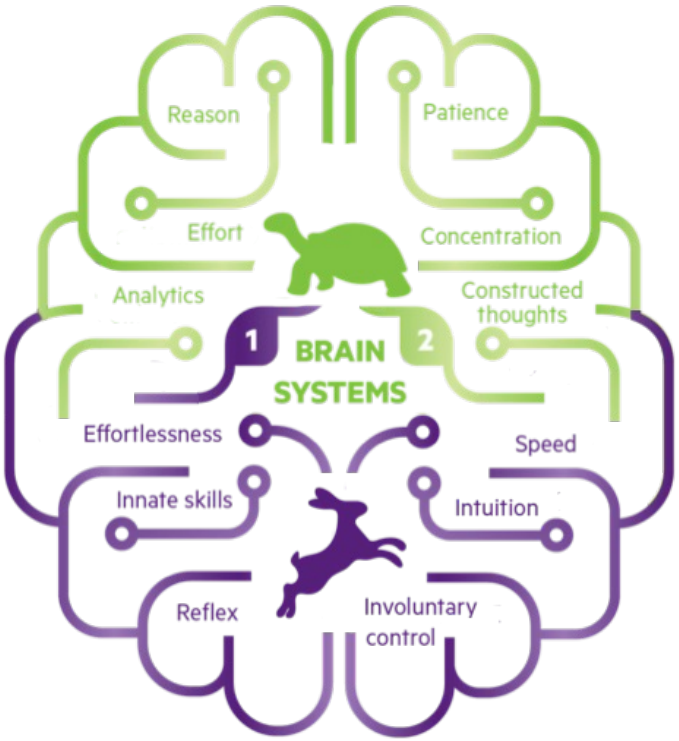


Hybrid reactive-deliberative

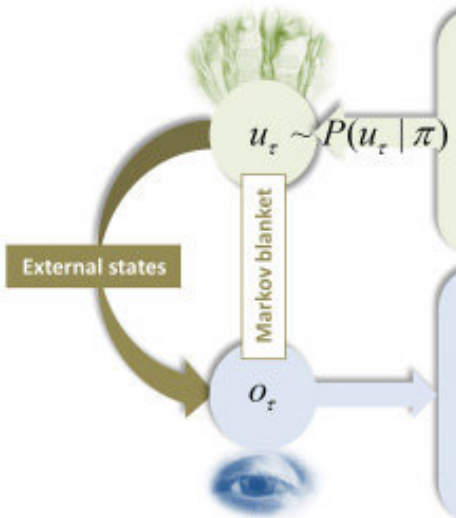

R0
R1
R2
S0
S1
S2
S3
O0
A0
O1
A1
O2
A2

$$b'(s') = \eta O(o \mid s', a) \sum_{s \in S} T(s' \mid s, a) b(s)$$


World
fully observable vs. partially observable
observation
action
sole source of change vs. exogenous events
deterministic vs. stochastic
instantaneous vs. durative
ACHIEVE GOAL CONDITION
OR
MAXIMIZE EXPECTED REWARD


Reason
Patience
Effort
Concentration
Analytics
Constructed thoughts
1
BRAIN SYSTEMS
2
Effortlessness
Speed
Innate skills
Intuition
Reflex
Involuntary control

Active inference



Expected free energy minimisation

Planning as inference

$$\pi = \arg\min_\pi G(\pi)$$

$$G(\pi) = \mathrm{E}_{Q(o_\tau, s_\tau \mid \pi)}[\overbrace{\ln Q(s_\tau \mid \pi) - \ln P(s_\tau)}^{\text{Risk}} \overbrace{- \ln P(o_\tau \mid s_\tau)}^{\text{Ambiguity}}]$$

$$\geq \mathrm{E}_{Q(o_\tau, s_\tau \mid \pi)}[\underbrace{\ln Q(s_\tau \mid \pi) - \ln Q(s_\tau \mid o_\tau, \pi)}_{\text{Intrinsic value}} \underbrace{- \ln P(o_\tau)}_{\text{Extrinsic value}}]$$

Variational free energy minimisation

Perceptual inference

$$Q(s_\tau \mid \pi) = \arg\min_Q F(\pi)$$

$$F(\pi) = \mathrm{E}_{Q(s_{\leq\tau} \mid \pi)}[\overbrace{\ln Q(s_{\leq\tau} \mid \pi) - \ln P(s_{\leq\tau} \mid \pi)}^{\text{Complexity}} \overbrace{- \ln P(o_{\leq\tau} \mid s_{\leq\tau}, \pi)}^{\text{Accuracy}}]$$

$$= \mathrm{E}_{Q(s_{\leq\tau} \mid \pi)}[\underbrace{\ln Q(s_{\leq\tau} \mid \pi) - \ln P(s_{\leq\tau} \mid o_{\leq\tau}, \pi)}_{\text{Evidence bound}} \underbrace{- \ln P(o_{\leq\tau} \mid \pi)}_{\text{Log evidence}}]$$

No priors

$$\mathrm{E}_Q[D[Q(s_\tau \mid o_\tau, \pi) \| Q(s_\tau \mid \pi)]]$$
$$=$$
$$D_{KL}[Q(s_\tau, o_\tau \mid \pi) \| Q(s_\tau \mid \pi) Q(o_\tau \mid \pi)]$$

Bayesian surprise
Optimal Bayesian design
Intrinsic motivation
Infomax principle

No ambiguity

$$D_{KL}[Q(s_\tau \mid \pi) \| P(s_\tau)]$$
$$=$$
$$D_{KL}[Q(o_\tau \mid \pi) \| P(o_\tau)]$$

Risk-sensitive policies
KL control
Occam's principle

No intrinsic value

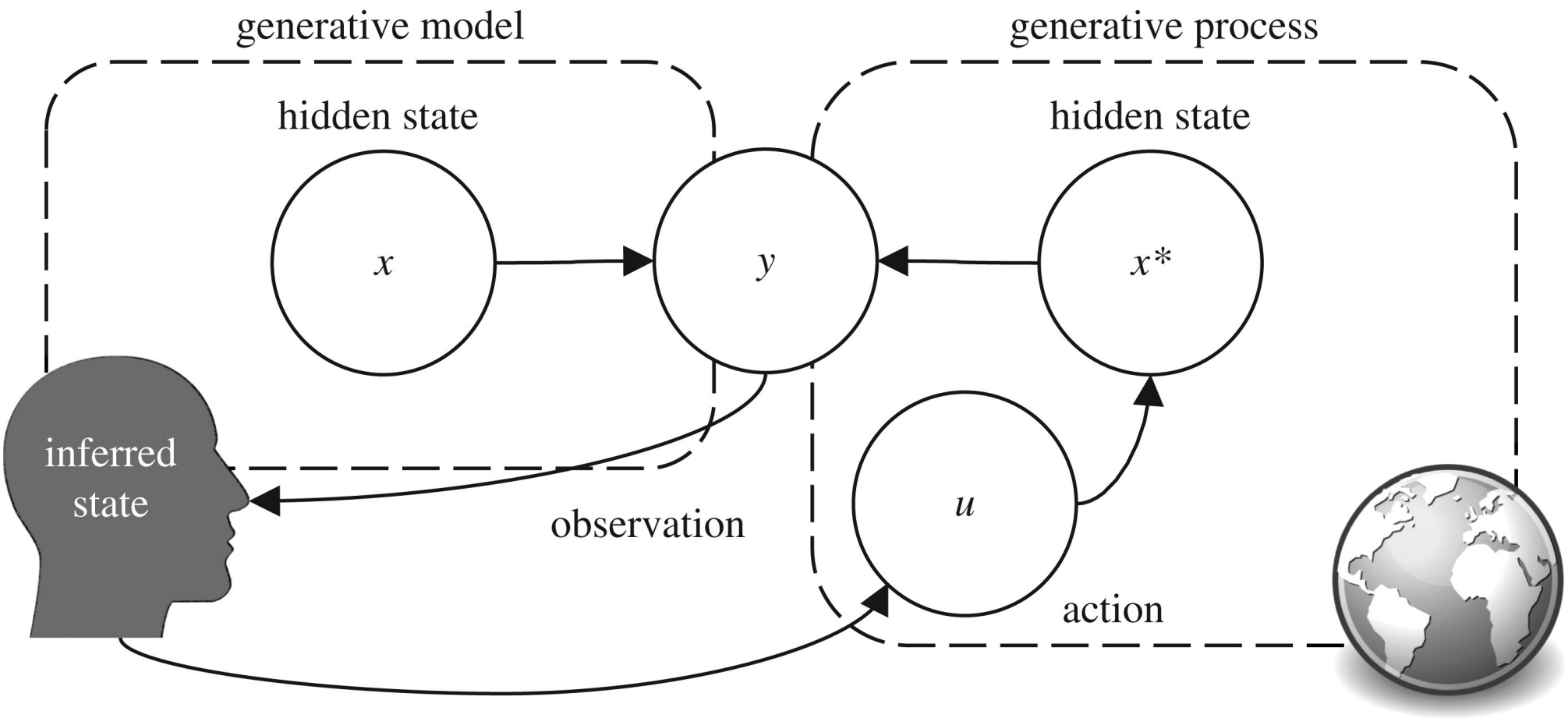
$$\mathrm{E}_Q[\ln P(s_\tau)]$$
$$=$$
$$\mathrm{E}_Q[\ln P(o_\tau)]$$

Bayesian decision theory
Expected utility theory

No ambiguity or priors

$$\mathrm{E}_Q[-\ln Q(s_\tau \mid \pi)]$$
$$=$$
$$\mathrm{H}[Q(s_\tau \mid \pi)]$$

Maximum entropy principle


generative model
generative process
hidden state
hidden state
x
y
x*
u
inferred state
observation
action

Key directions/

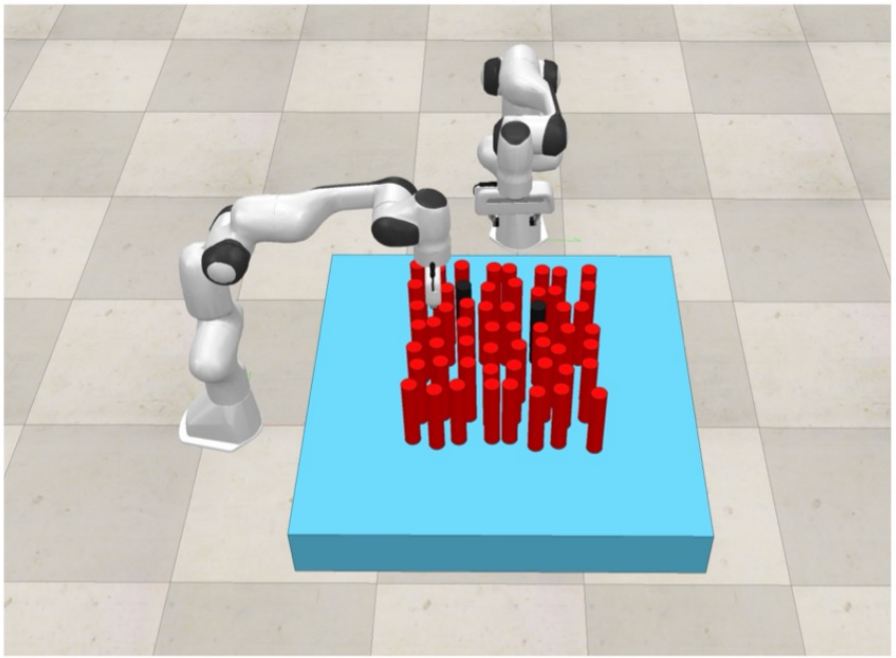
D1/ Robot cognitive architectures based on brain-like principles, such as modularity, structure, hierarchy and recursiveness.

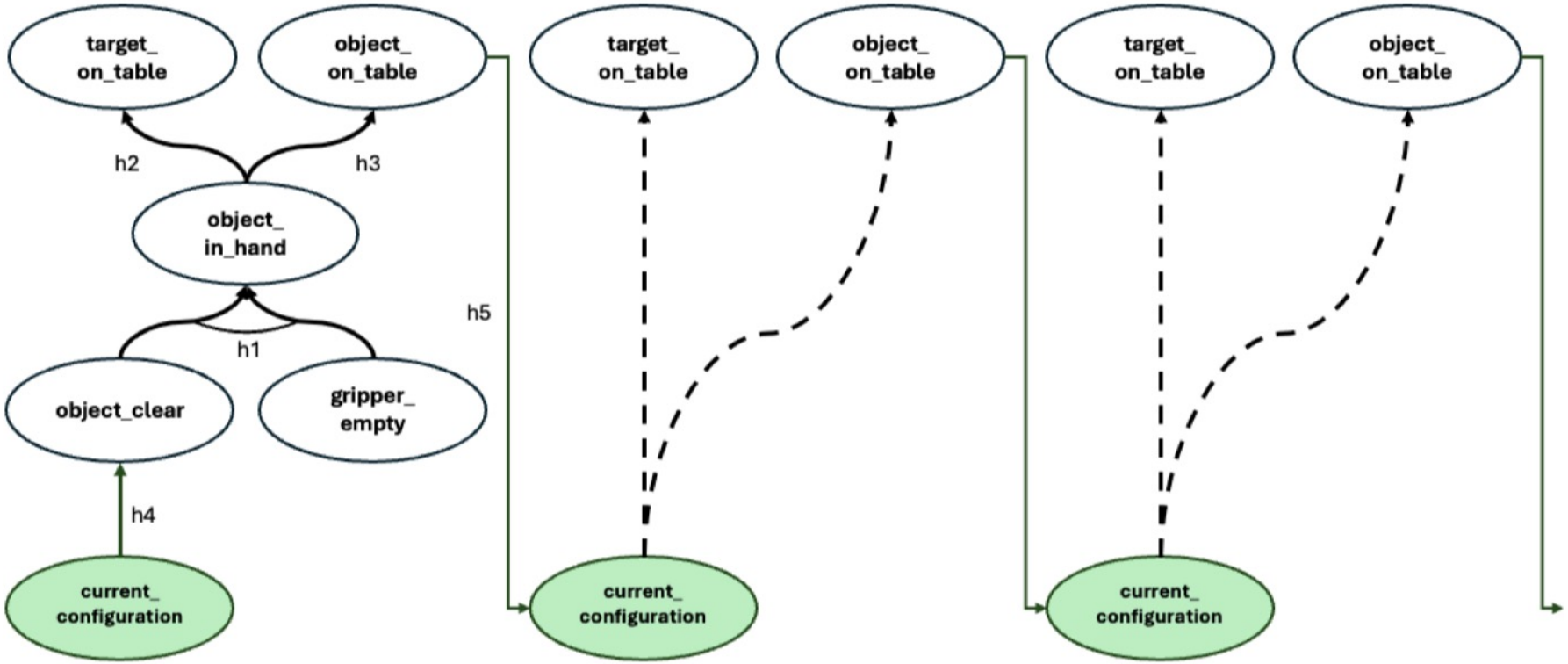
D2/ Use of neuro-symbolic approaches to AI, which integrate data-driven techniques within model-based architectures.

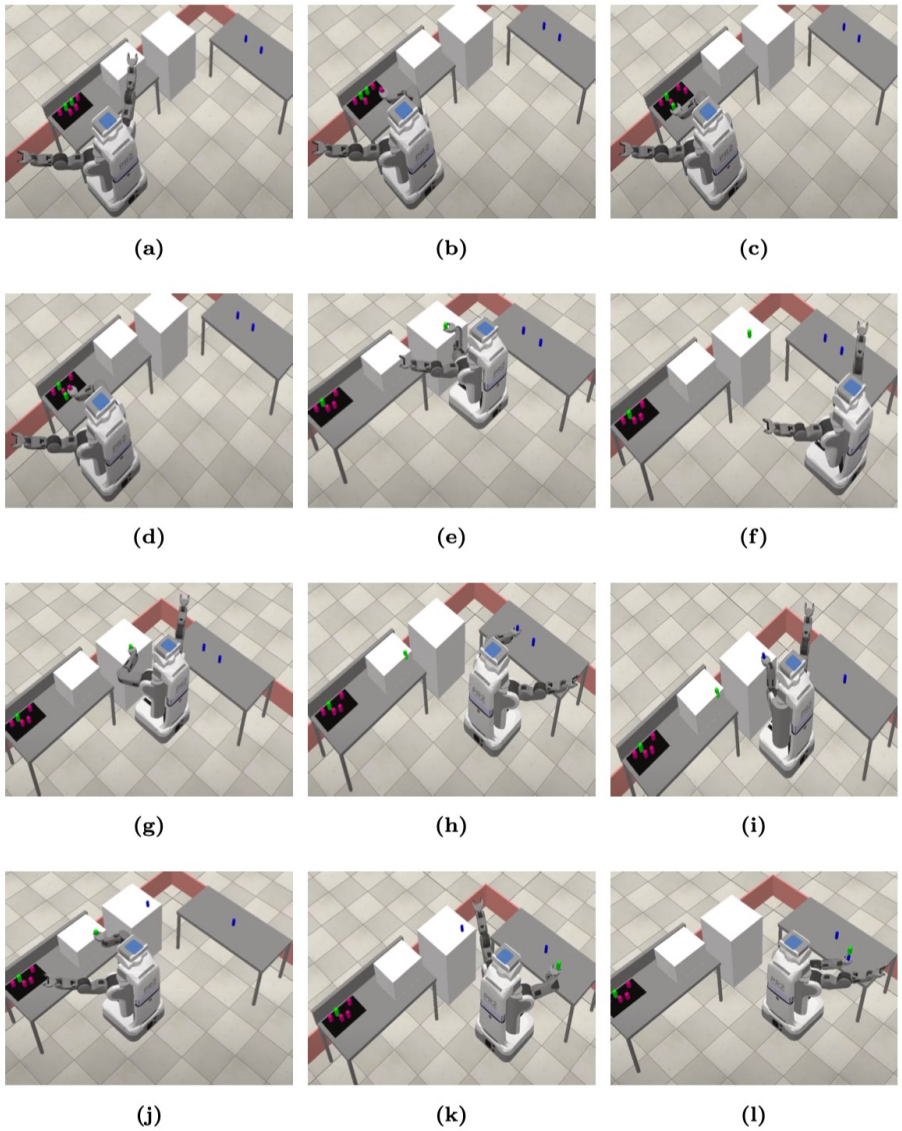
D3/ Distribution of the "intelligence architecture" to both the robot "brain" and the robot "body".

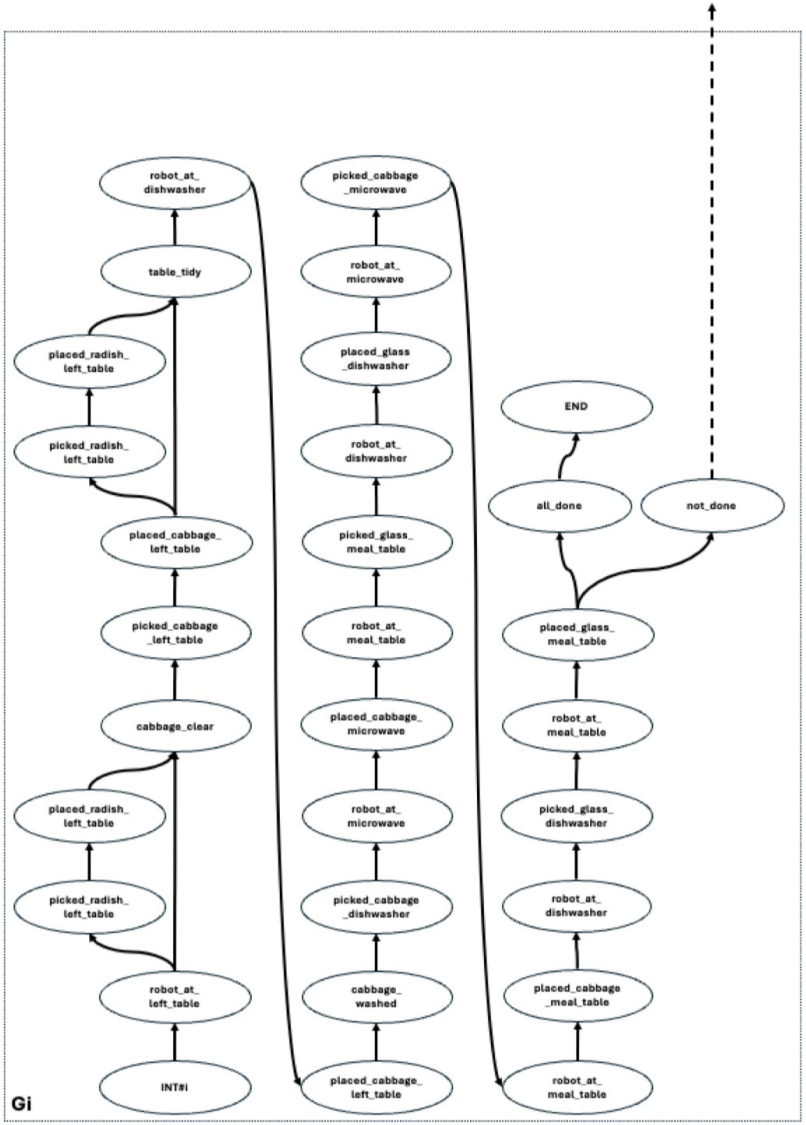
D4/ Alignment between the encoded behaviours and human values.

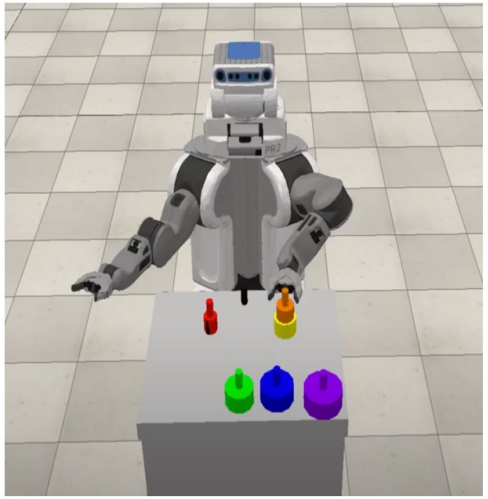

Volleware
Gimars



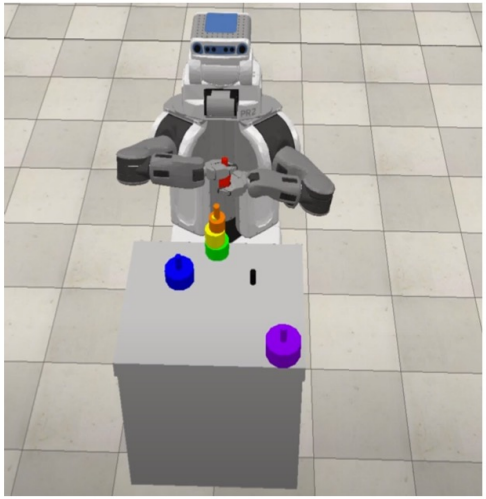

target_
on_table
object_
on_table
h2
h3
object_
in_hand
h1
object_clear
gripper_
empty
h4
current_
configuration
h5
target_
on_table
object_
on_table
current_
configuration
target_
on_table
object_
on_table
current_
configuration


(a)
(b)
(c)
(d)
(e)
(f)
(g)
(h)
(i)
(j)
(k)
(l)

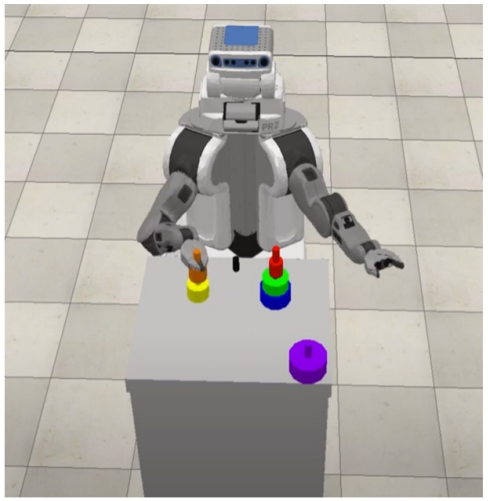

robot_at_ dishwasher
table_tidy
placed_radish_ left_table
picked_radish_ left_table
placed_cabbage_ left_table
picked_cabbage _left_table
cabbage_clear
placed_radish_ left_table
picked_radish_ left_table
robot_at_ left_table
INT#i
picked_cabbage _microwave
robot_at_ microwave
placed_glass _dishwasher
robot_at_ dishwasher
picked_glass_ meal_table
robot_at_ meal_table
placed_cabbage_ microwave
robot_at_ microwave
picked_cabbage _dishwasher
cabbage_ washed
placed_cabbage_ left_table
END
all_done
not_done
placed_glass_ meal_table
robot_at_ meal_table
picked_glass_ dishwasher
robot_at_ dishwasher
placed_cabbage _meal_table
robot_at_ meal_table
Gi



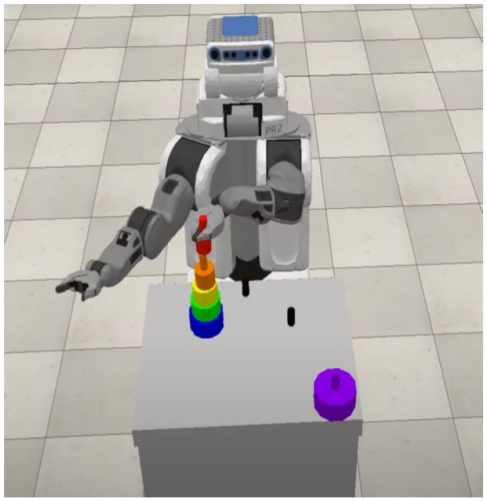
(a)



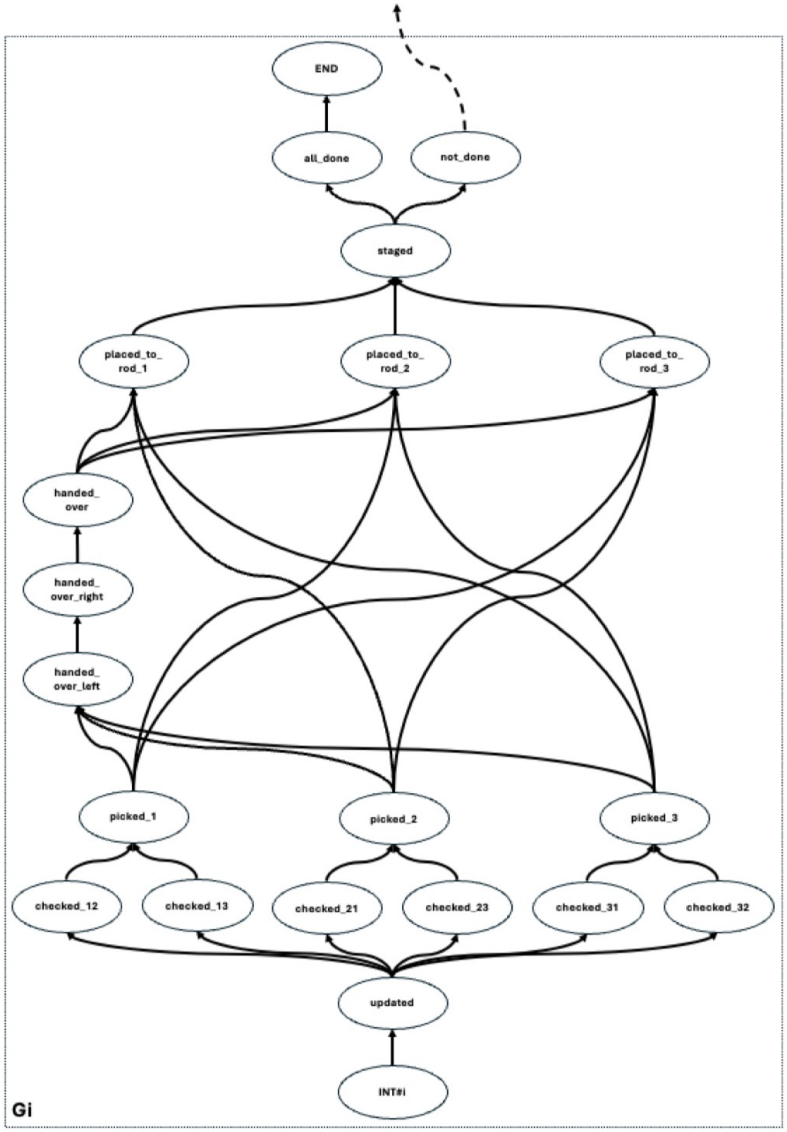
(b)

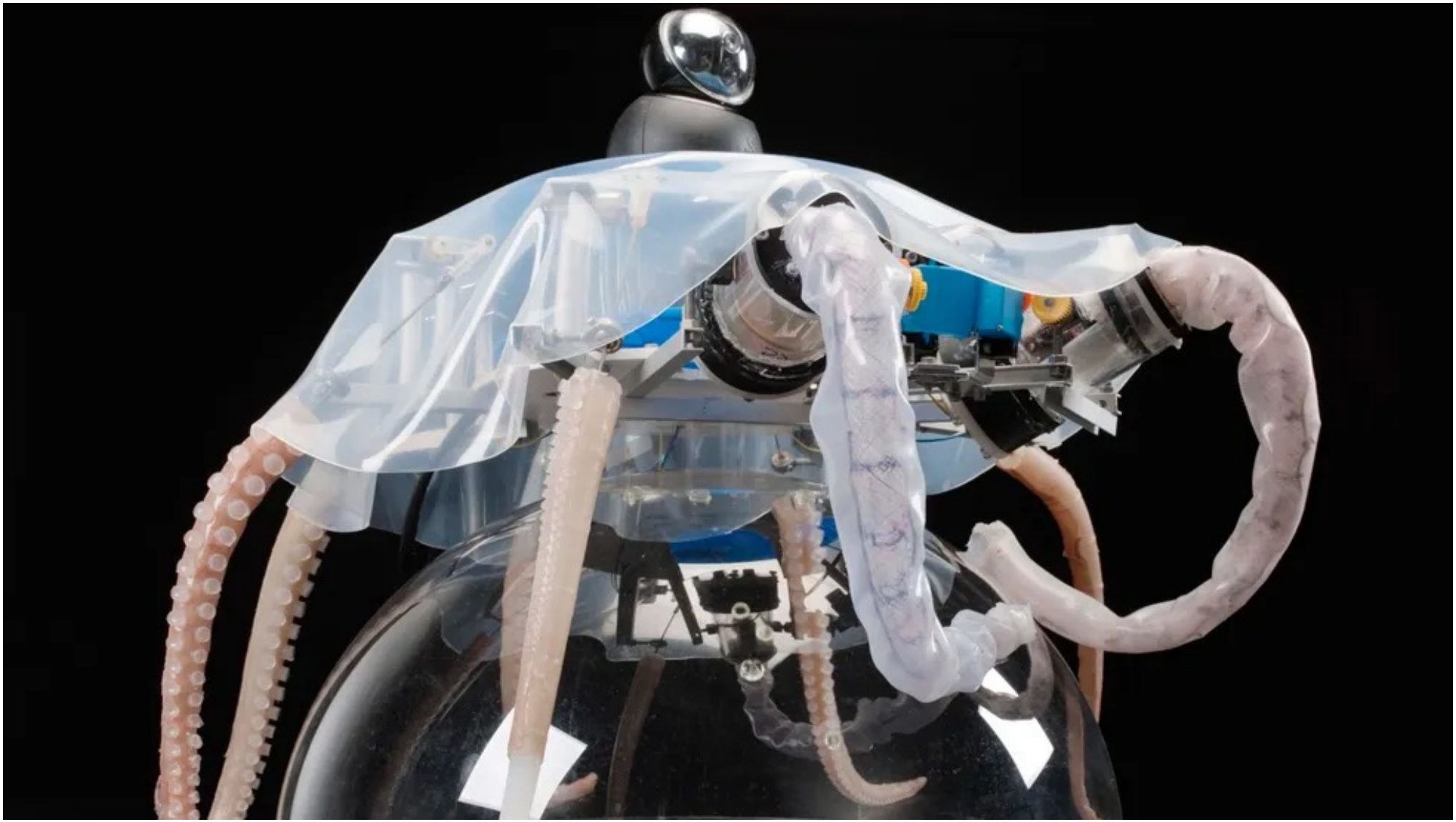


(c)



(d)


END
all_done
not_done
staged
placed_to_rod_1
placed_to_rod_2
placed_to_rod_3
handed_over
handed_over_right
handed_over_left
picked_1
picked_2
picked_3
checked_12
checked_13
checked_21
checked_23
checked_31
checked_32
updated
INT#i
Gi




Evade Shutdown
Hack Computer Systems
Make Copies
Acquire Resources
Ethics Violation
Hire or Manipulate Humans
AI Research & Programming
Persuasion & Lobbying
Hide Unwanted Behaviors
Strategically Appear Aligned
Escape Containment
Research & Development
Manufacturing & Robotics
Autonomous Weaponry

THANK YOU!

Fulvio Mastrogiovanni

Professor of Robotics and Artificial Intelligence, University of Genoa
Founder and Chief Science Officer, Teseo srl
Coordinator of the Scientific Committee, Digital Innovation Hub Liguria