



Manuel Baltieri, Ph.D.
Chief Researcher, Araya Inc.

Foundations of AI:
building reliable, trustworthy and agentic systems

Background

Araya's teams and my role

Research DX

Neurotech

R&D

Edge AI

Air Conditioning

Computational Fluid Dynamics

Image Recognition

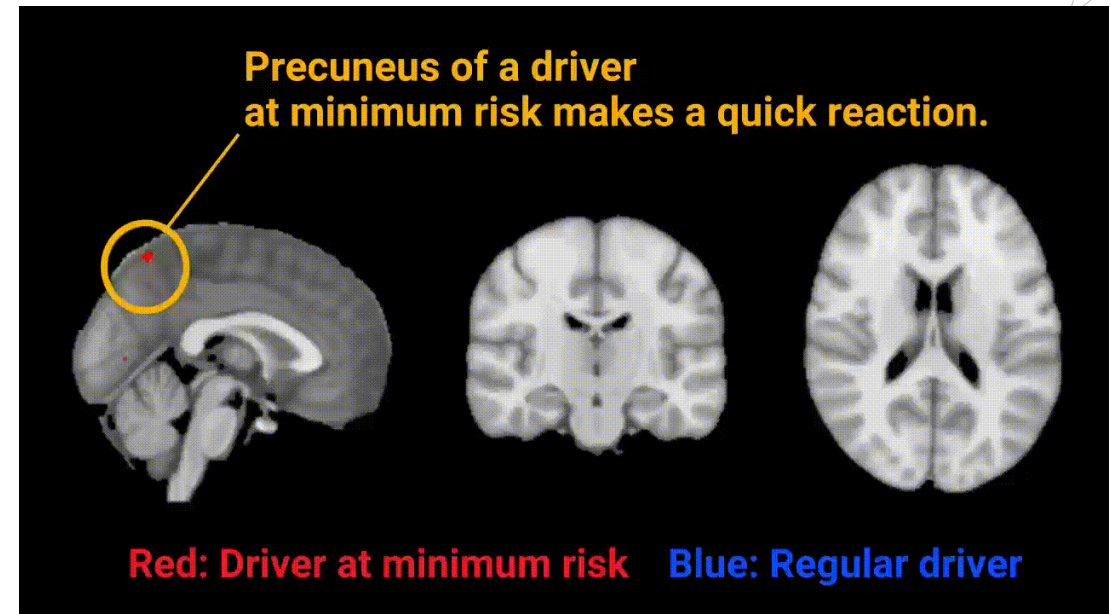
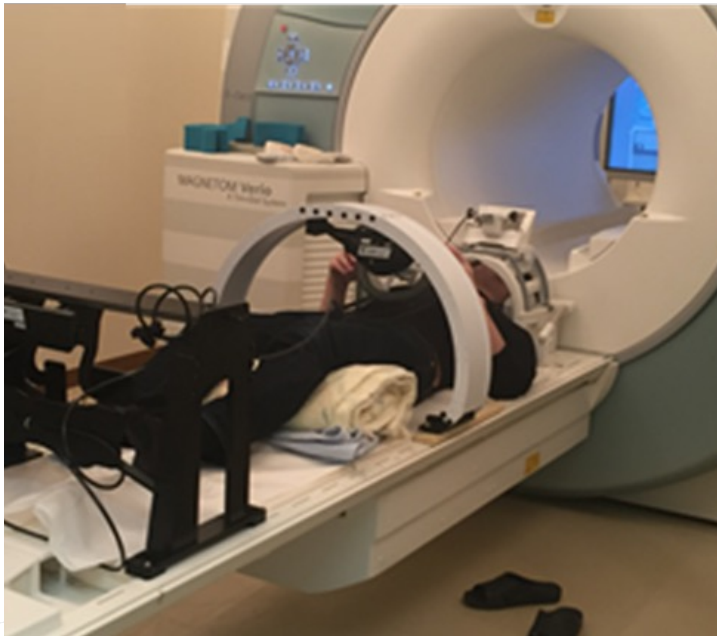
Construction Machinery



Examples (neuroscience)

Neurological Analysis Support

HONDA



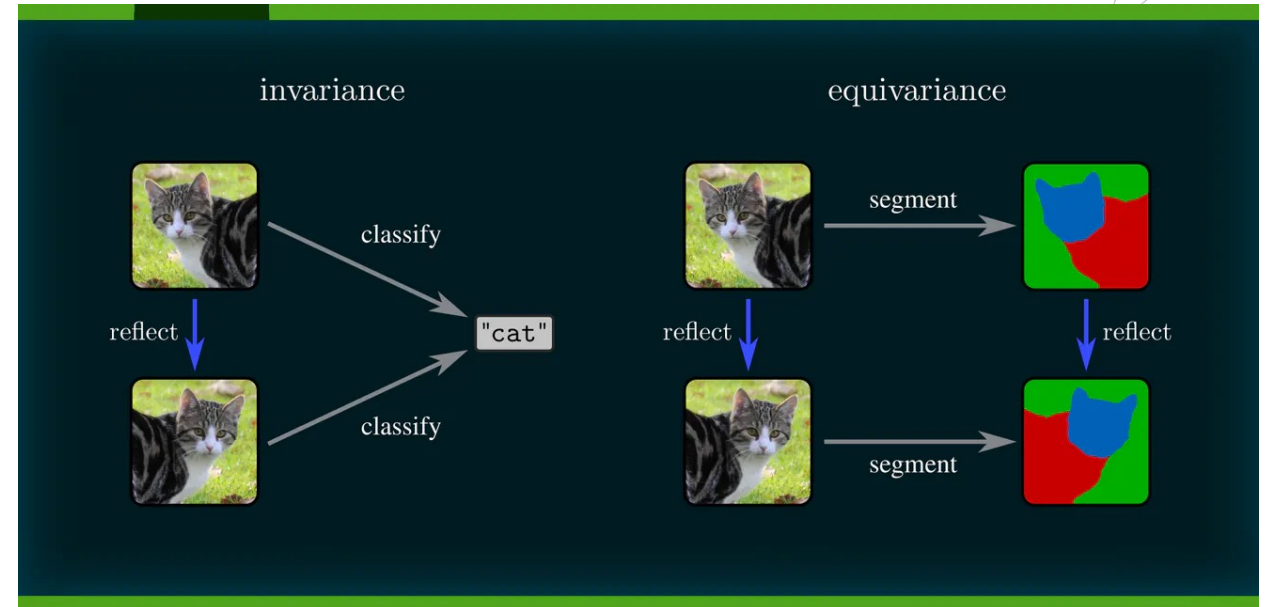
Support for joint or contracted research for brain-science-informed models of decision making in manufacturing processes.

<https://prtimes.jp/main/html/rd/p/000000013.000049573.html>

Examples (machine learning)

Mathematical approaches to ML

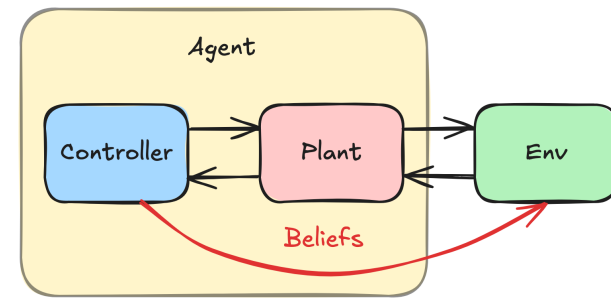
```
test.py
test.py > ...
1 def celsius_to_fahrenheit(celsius):
2     """Convert Celsius to Fahrenheit
3     :param celsius: temperature in Celsius
4     :return: temperature in Fahrenheit
5     """
6     fahrenheit = celsius * 9.0 / 5 + 32
7     return fahrenheit
8
9
```



Support and production of ML models for program synthesis and code comprehension combining existing methods with modern techniques from geometric deep learning.

Example (control theory/applied maths)

Agentic and safe AI



Identity

Goal-directedness

Models/beliefs

Autonomy

B. Bayesian model of the whole system

Corollary 16 established that the active controller C^* HW-models the global attractor S^* via $\pi_{C^*} : S^* \rightarrow C^*$.

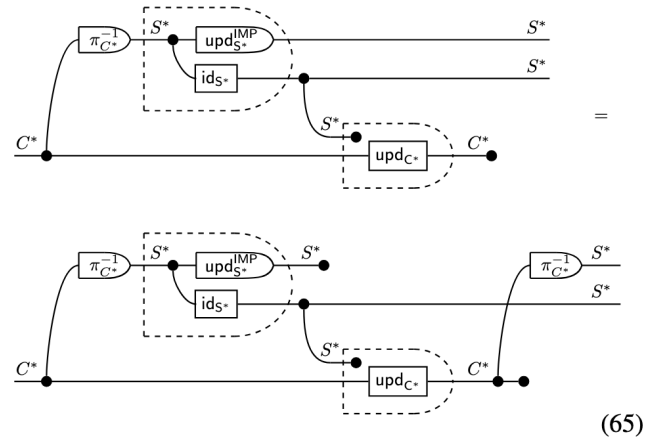
Example 28. Due to Theorem 27, we thus know that $\pi_{C^*}^{-1} : C^* \rightarrow \mathcal{P}^+ S^*$ is a belief map for the reasoner $\text{upd}_{R-S^*} : S^* \times C^* \rightarrow C^*$ given by

$$\text{upd}_{R-S^*} := \text{del}_{S^*} \otimes \text{upd}_{C^*}, \quad (63)$$

with respect to the model $\kappa_{S^*} : S^* \rightarrow S^* \times S^*$ given by

$$\kappa_{S^*} := \Delta_{S^*} \circ \text{upd}_{S^*}^{\text{IMP}} \otimes \text{id}_{S^*}. \quad (64)$$

The resulting consistency equation is then:



C. Bayesian model of the environment

Corollary 16 established that the active controller C^* also HW-models the environment E via $\mu_E : E \rightarrow C^*$.

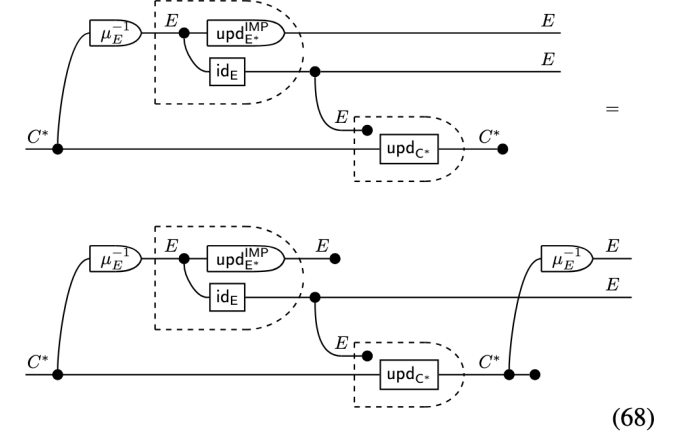
Example 29. Due to Theorem 27 we know that $\mu_E^{-1} : C^* \rightarrow \mathcal{P}^+ E$ is a belief map for the reasoner $\text{upd}_{R-E} : E \times C^* \rightarrow C^*$ given by

$$\text{upd}_{R-E} := \text{del}_E \otimes \text{upd}_{C^*}, \quad (66)$$

with respect to the model $\kappa^E : E \rightarrow E \times E$ given by

$$\kappa^E := \Delta_E \circ \text{upd}_E^{\text{IMP}} \otimes \text{id}_E. \quad (67)$$

The resulting consistency equation is then:



Fundamental research in agency and decision making for the next generation safe AI systems.



Engaged in AI-assisted BMI development as PM since 2020

Moonshot Research and Development Program from the Japan Science and Technology Agency



Contact Japanese

What is IoB? 5 Themes Member Activities Achievements Careers



Moonshot Goal 1

Realization of a society in which human beings can be free from limitations of body, brain, space, and time by 2050

Liberation from Biological Limitations via Physical, Cognitive and Perceptual Augmentation

Our goal is to make a world in which anyone can participate in creative social activities by controlling a large number of cybernetic avatars (CAs) in coordination using only the images of words and actions that come to mind, regardless of physical or brain limitations. To achieve this goal, we will improve the performance of brain-machine interfaces (BMIs), which read mental images and intentions, and promote their utilization in society. In conjunction with proactively addressing ethical issues, we aim to create a new industry centred around BMI-CA that can be trusted by society.