OSAWA Laboratory

第33回 全脳アーキテクチャ勉強会

社会的承認による定義をされたAGIに向けたHAIとWBAの役割

日本大学文理学部 情報科学科 助教
次世代社会研究センター (RINGS) 長
大澤 正彦

大澤 正彦 (Masahiko Osawa)

日本大学文理学部 次世代社会研究センター (RINGS) 長 / 情報科学科 助教 / 大澤研究室 主宰
専修大学 ネットワーク情報学部 兼任講師
株式会社BLUEM 代表取締役
孫正義育英財団 正財団生, **全脳アーキテクチャ若手の会** 設立者/フェロー

1993年生まれ. 2020年3月 慶大院にて博士 (工学)を取得.
2020年4月より日大文理助教, 同年12月よりRINGS センター長
記憶のないくらい前からずっと, **ドラえもんをつくる**のが夢

日本大学文理学部　専修大学

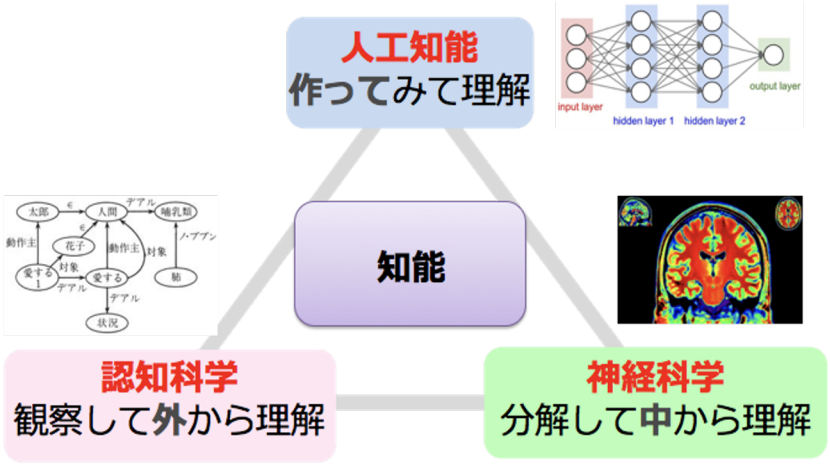
BLUEM

WBA Future Leaders
WHOLE BRAIN ARCHITECTURE FUTURE LEADERS

01 自己紹介

全脳アーキテクチャ若手の会

AIに関係のない人なんていない



2014年夏に大澤が設立．2017年6月まで代表．
6年間で **2,600人規模** の団体に成長

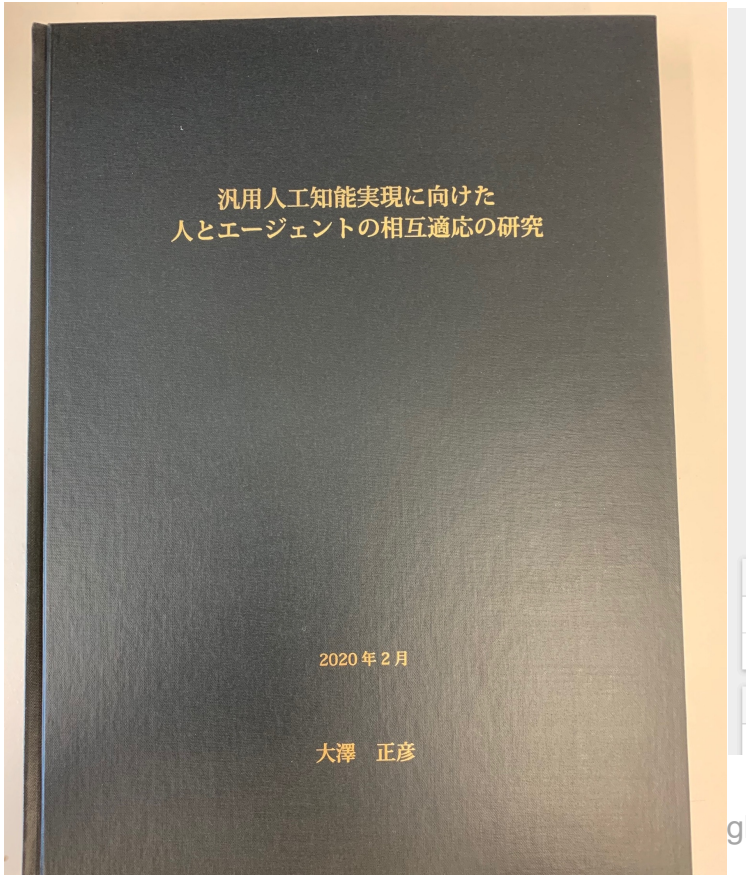
全世界に優秀な運営メンバー
- 関東支部, 関西支部, 九州支部, 東北支部の4拠点
- 大学・大学院の**首席8名**輩出
- **Harvard, Johns Hopkins**など輩出

運営メンバーの共同研究開発の成功

多種多様な立場の老若男女がフラットに
議論・情報共有できる場
- 研究者, ビジネスマン, 起業家,
ライター, 小説家, 漫画家, 医者,
ダンサー, etc.

WBA Future Leaders
WHOLE BRAIN ARCHITECTURE FUTURE LEADERS

汎用人工知能をテーマに博論



社会的承認によって定義されたAGIの開発

- 汎用人工知能 (Artificial General Intelligence; AGI)
  - 環境に応じて汎用的に動作可能な人工知能
  - 特化型人工知能（narrow AI) の対立概念

- AGI開発の障壁
  - 膨大かつ高難易度な機能要件集合
  - 明確でない機能要件集合
  - 定義の曖昧さ

定義方法や評価方法から見直し

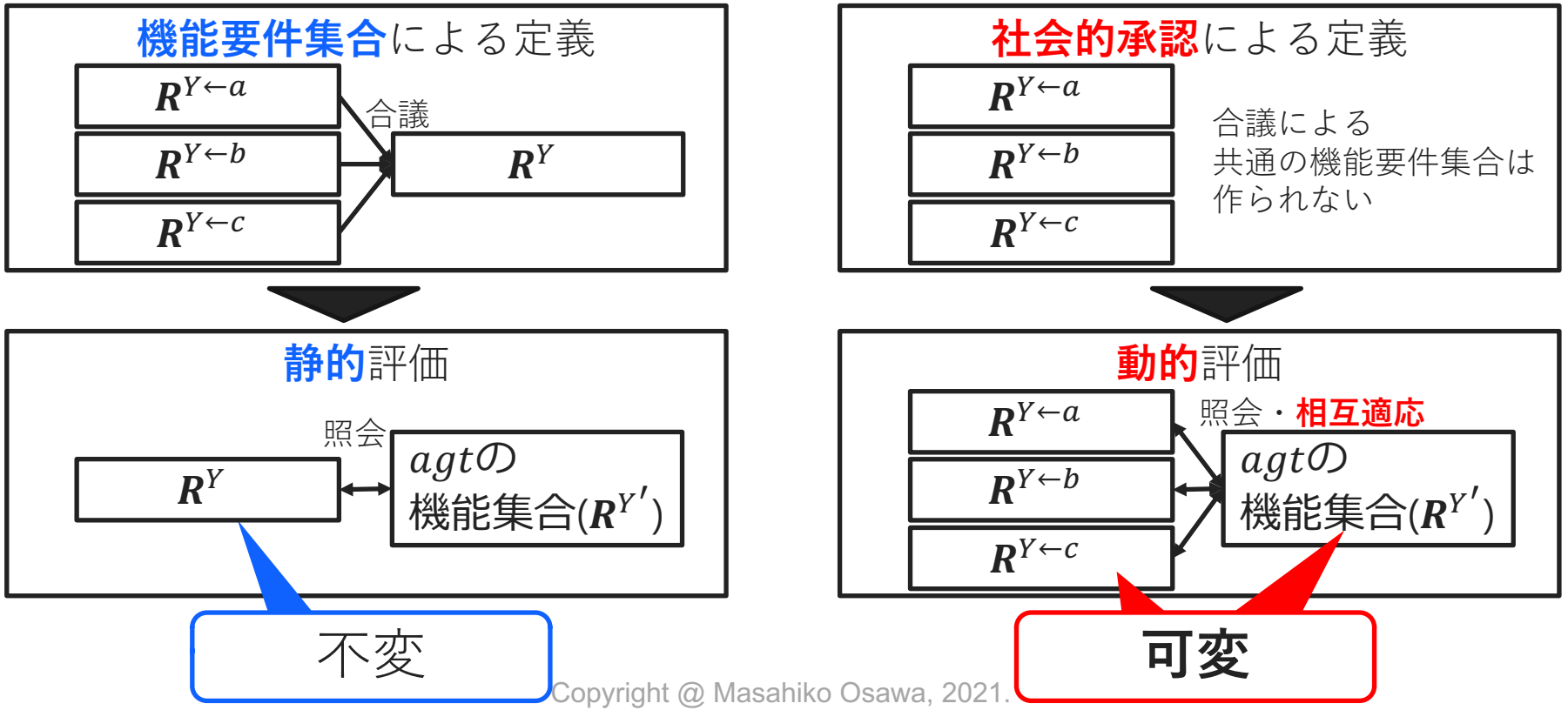
AGIの定義方法と評価方法

- 機能要件集合によるAGIの定義
  - “何ができたらAGIか”という機能要件を列挙
  - 例: 感情を持っていたらAGI!

- 社会的承認によるAGIの定義
  - “多くの人にAGIと認められればAGI”
  - 典型例：チューリングテスト

社会的承認によって定義された
AGIの可能性を模索

AGIの定義方法と評価方法



定義方法の差異による研究アプローチの比較

機能要件集合で定義した場合にとれるアプローチ

実装する

実装せずに
実装済みにみせる

実装せずに
許してもらう

社会的承認で定義した場合にとれるアプローチ

研究アプローチの比較

- 例：
  - R1: 目に見える
  - R2: 感情がある
  - R3: タイムマシーンに乗ってやってくる

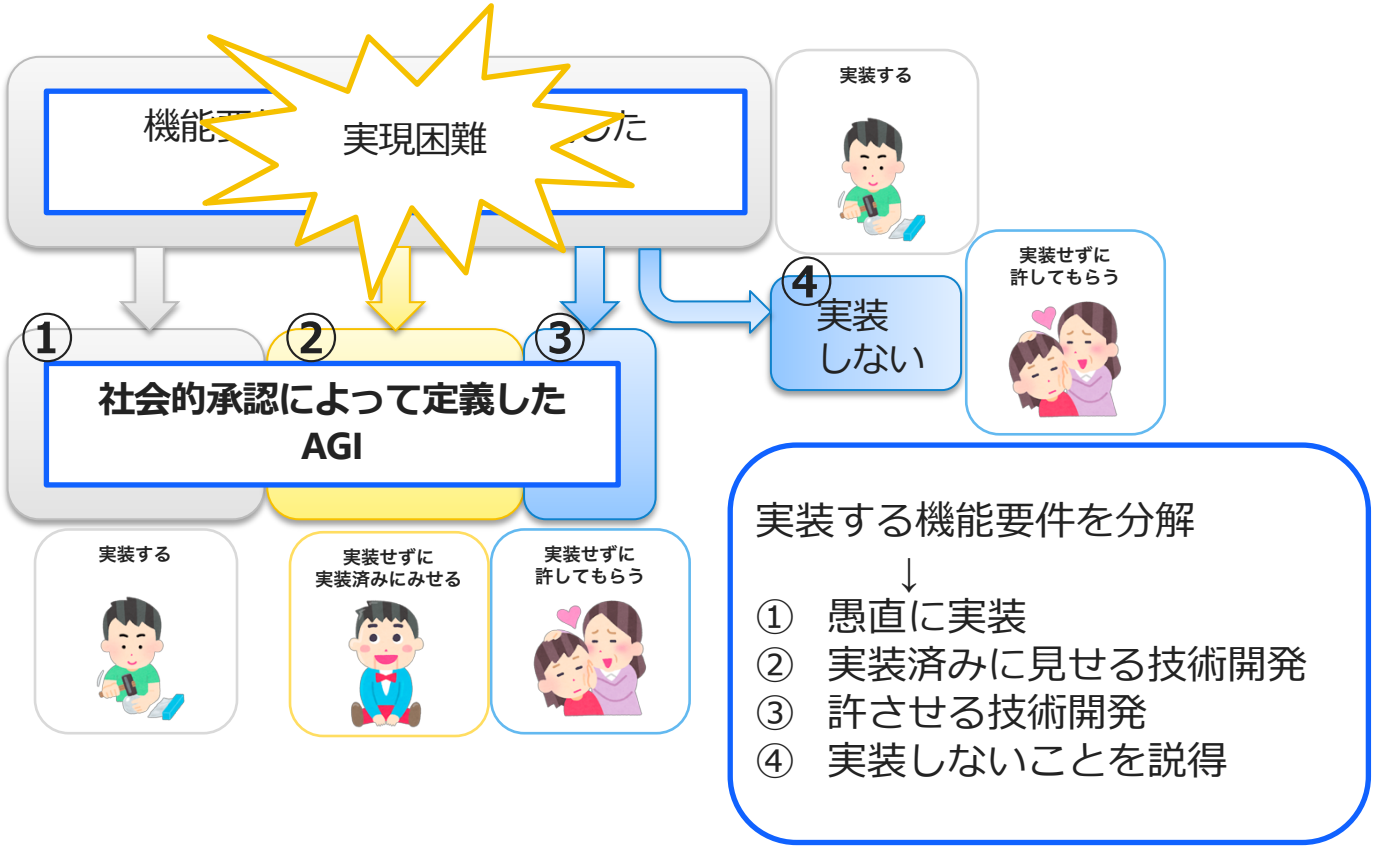
実装する

実装せずに
実装済みにみせる

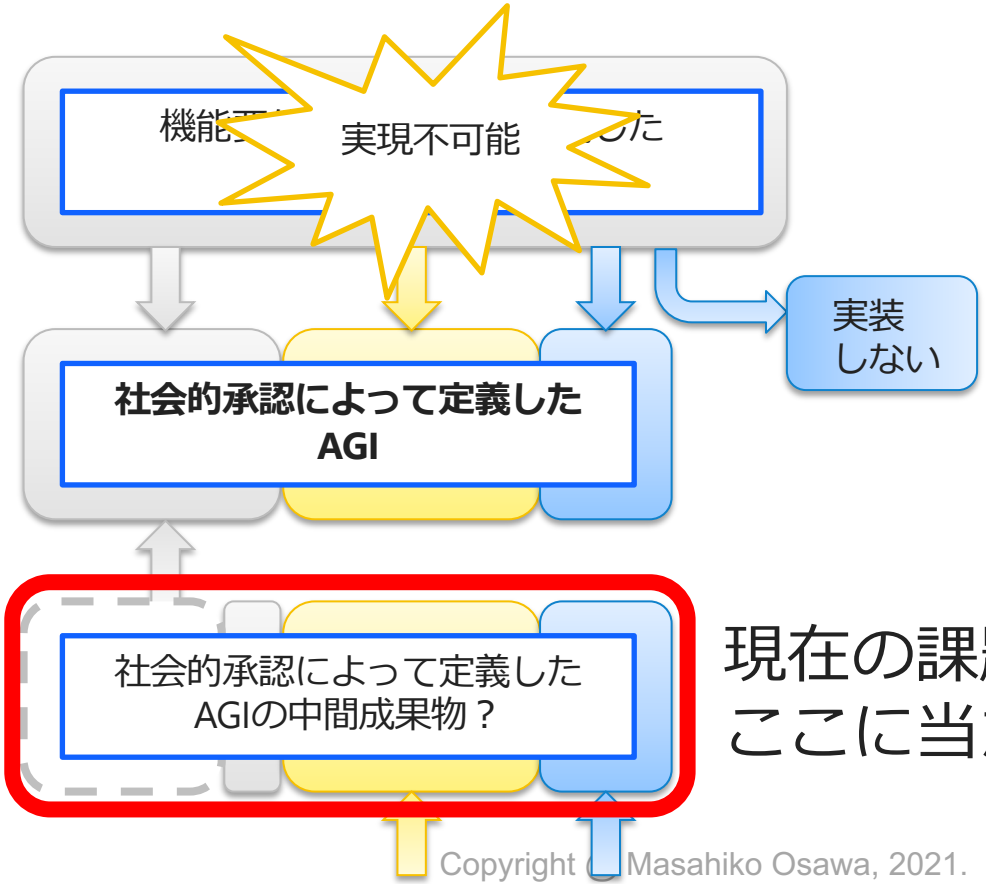
実装せずに
許してもらう

社会的承認によって定義されたAGIの中間成果物

社会的承認によって定義されたAGIの中間成果物



社会的承認によって定義されたAGIの中間成果物



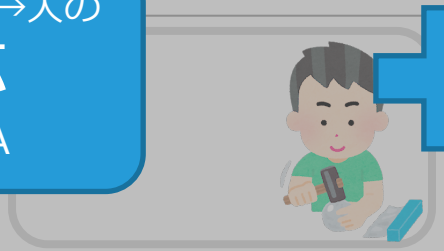
現在の課題は
ここに当たる問題設定

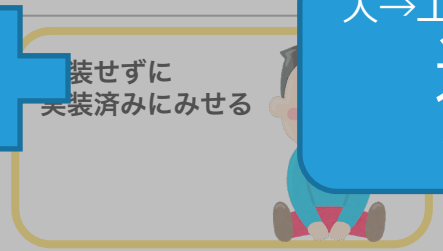
AGIのつくりかた

	実装する 	実装せずに 実装済みにみせる 	実装せずに 許してもらう 
あらかじめ 実装しておく	エージェント→人の **適応** 例: WBA	人→エージェントの **適応** 例: HAI	
インタラクションの中で 実現する			

ドラえもんのつくりかた

エージェント→人の
適応
例：WBA





人→エージェントの
適応
例：HAI

あらかじめ
実装[redacted]

インタラクションの中で
実現する

人とエージェントの
相互適応

社会的承認によるAGIの定義

研究の体系化

社会的承認によるAGIの定義

研究の体系化

神経科学

深層学習

認知科学

参考

改良

参考

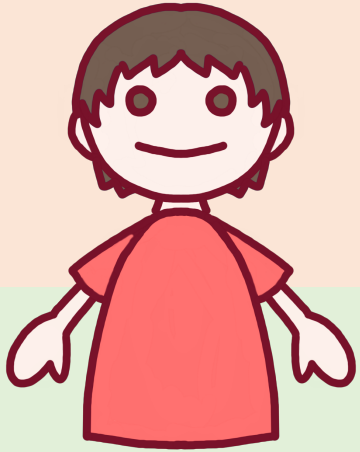
適応
アーキテクチャ

搭載

適応

適応

エージェント



人

社会的承認によるAGIの定義

研究の体系化

神経科学

深層学習

参考

改良

WBA

適応
アーキテクチャ

搭載

エージェント

認知科学

参考

適応

適応

人

HAI

社会的承認によるドラえもんの定義

神経科学

深層学習

参考

改良

適応
アーキテクチャ

エージェント→人の
適応
例: WBA

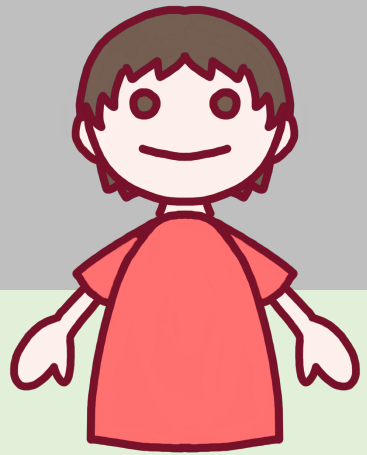
参考

適応

搭載

エージェント

適応



人

社会的承認によるドラえもんの定義

研究の体系化

人→エージェントの
適応
例: HAI

適応
アーキテクチャ

認知科学

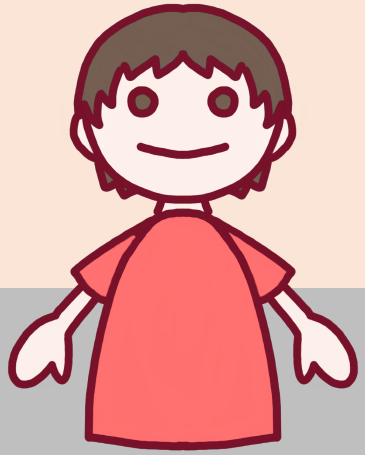
参考

適応

搭載

適応

エージェント



人

社会的承認によるドラえもんの定義

研究の体系化

神経科学

深層学習

認知科学

参考

改良

参考

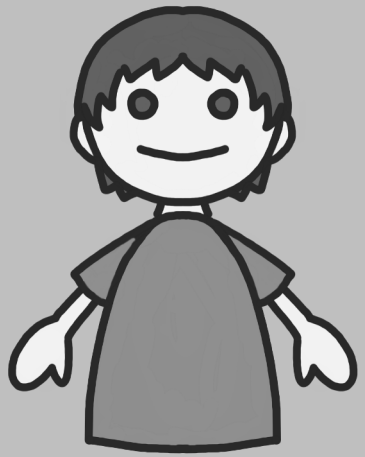
適応
アーキテクチャ

搭載

適応

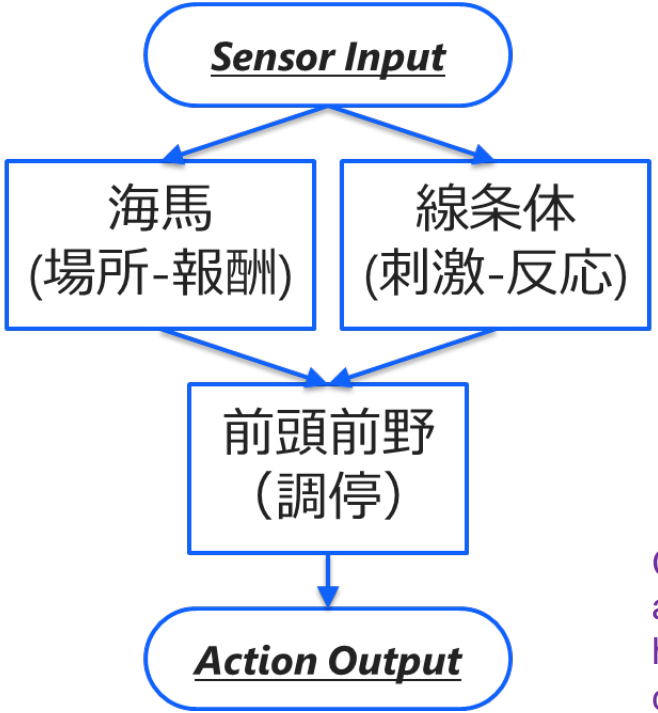
適応

エージェント



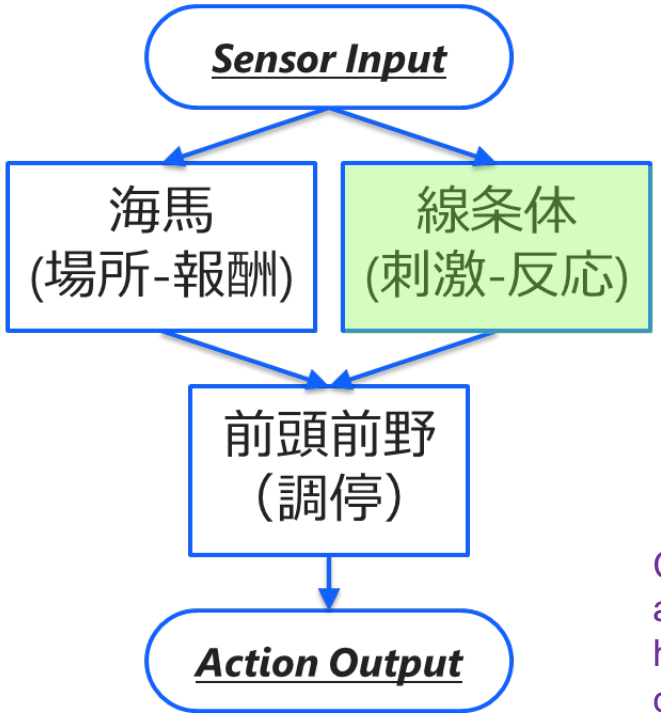
人

ネズミレベルの最小認知アーキテクチャ



Chersi and Neil, “The cognitive architecture of spatial navigation: hippocampal and striatal contributions.”, *Neuron,* 2015.

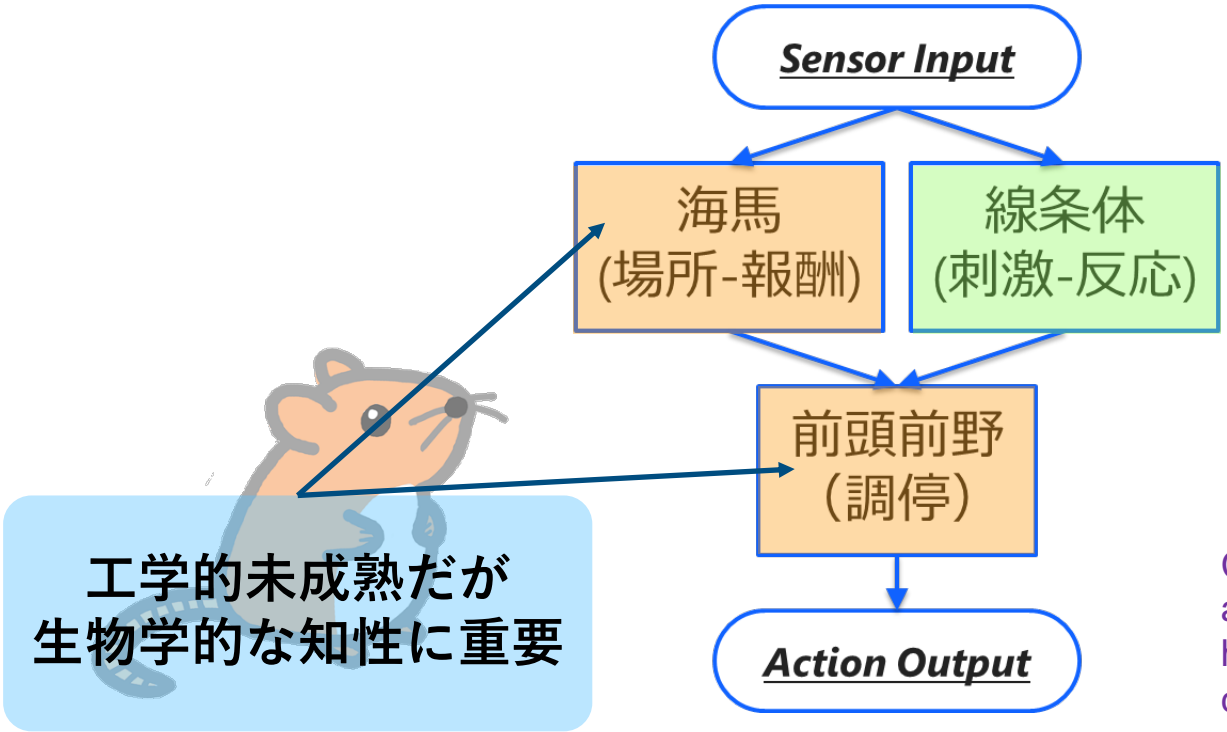
ネズミレベルの最小認知アーキテクチャ



深層強化学習的

Chersi and Neil, “The cognitive architecture of spatial navigation: hippocampal and striatal contributions.”, *Neuron,* 2015.

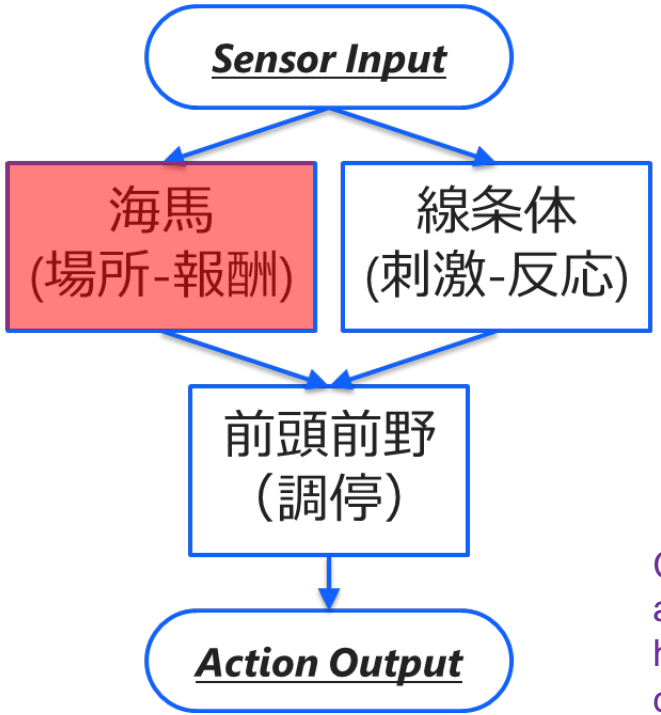
ネズミレベルの最小認知アーキテクチャ



深層強化学習的

Chersi and Neil, “The cognitive architecture of spatial navigation: hippocampal and striatal contributions.”, *Neuron,* 2015.

ネズミレベルの最小認知アーキテクチャ



Chersi and Neil, “The cognitive architecture of spatial navigation: hippocampal and striatal contributions.”, *Neuron,* 2015.

海馬を参考にした深層学習器の概要

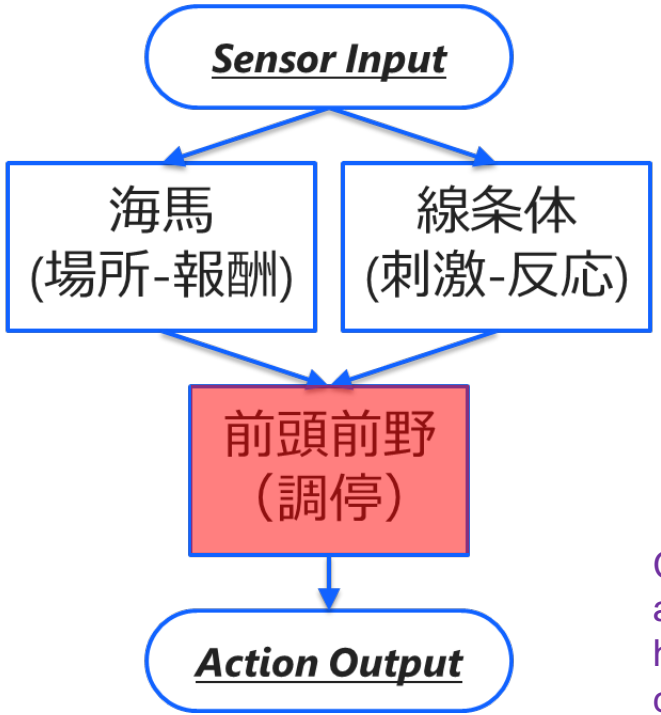
- 海馬では成体の脳に神経新生
  → 神経新生を参考にした学習モデル [1]
  * 新生させる細胞(素子数)を自動決定するアルゴリズムも独自開発 [2]
  発表3件中3件受賞 (*1*2*3)

- 海馬では強いリカレント構造
  → リカレント構造を
  取り入れた学習モデル [3]

- 神経科学への示唆 [4]

[1] Matsumori, Abe, Osawa, Imai, Procedia Computer Science, 2018.
[2] Osawa and Hagiwara, ISIS, 2015.
[3] Osawa, Imai, and Yamakawa, ICONIP, 2016.
[4] Osawa and Imai, Procedia Computer Science, 2018.
(受賞)
*1 IEEE Japan Chapter Young Researcher Award, 2014.
*2 ISIS2015, Best Presentation Award, 2015
*3 日本神経回路学会大会奨励賞, 2017.

ネズミレベルの最小認知アーキテクチャ



Chersi and Neil, “The cognitive architecture of spatial navigation: hippocampal and striatal contributions.”, *Neuron,* 2015.

前頭前野を参考にした汎用調停モデルの概要

- 前頭前野は複数モジュールの調停機能を担うとされる
  → Accumulator ニューロンのモデルを工学応用して実装
  → 別の問題設定の機械学習タスク（教師あり学習・強化学習）に
  汎用的に有効な調停手法を実現 [5], *4
  → 多数の応用事例 [6-8]

[5] Osawa, Ashihara, Seno, Imai and S. Kurihara, ICONIP, 2017
[6] Seno, Osawa, Imai, HAI, 2018
[7] Okuoka, Takimoto, Osawa, Imai, HAI, 2018.
[8] Seno, Osawa, Imai, BICA, 2018.

(受賞)
*4 人工知能学会30 周年記念事業奨励賞, 2016.

社会的承認によるドラえもんの定義

研究の体系化

神経科学

深層学習

認知科学

参考

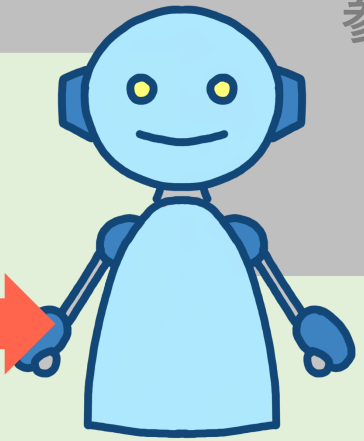
改良

参考

適応

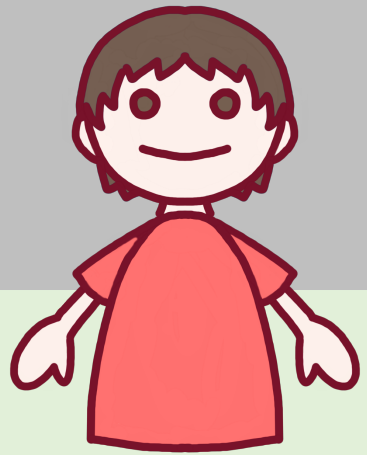
適応
アーキテクチャ

搭載

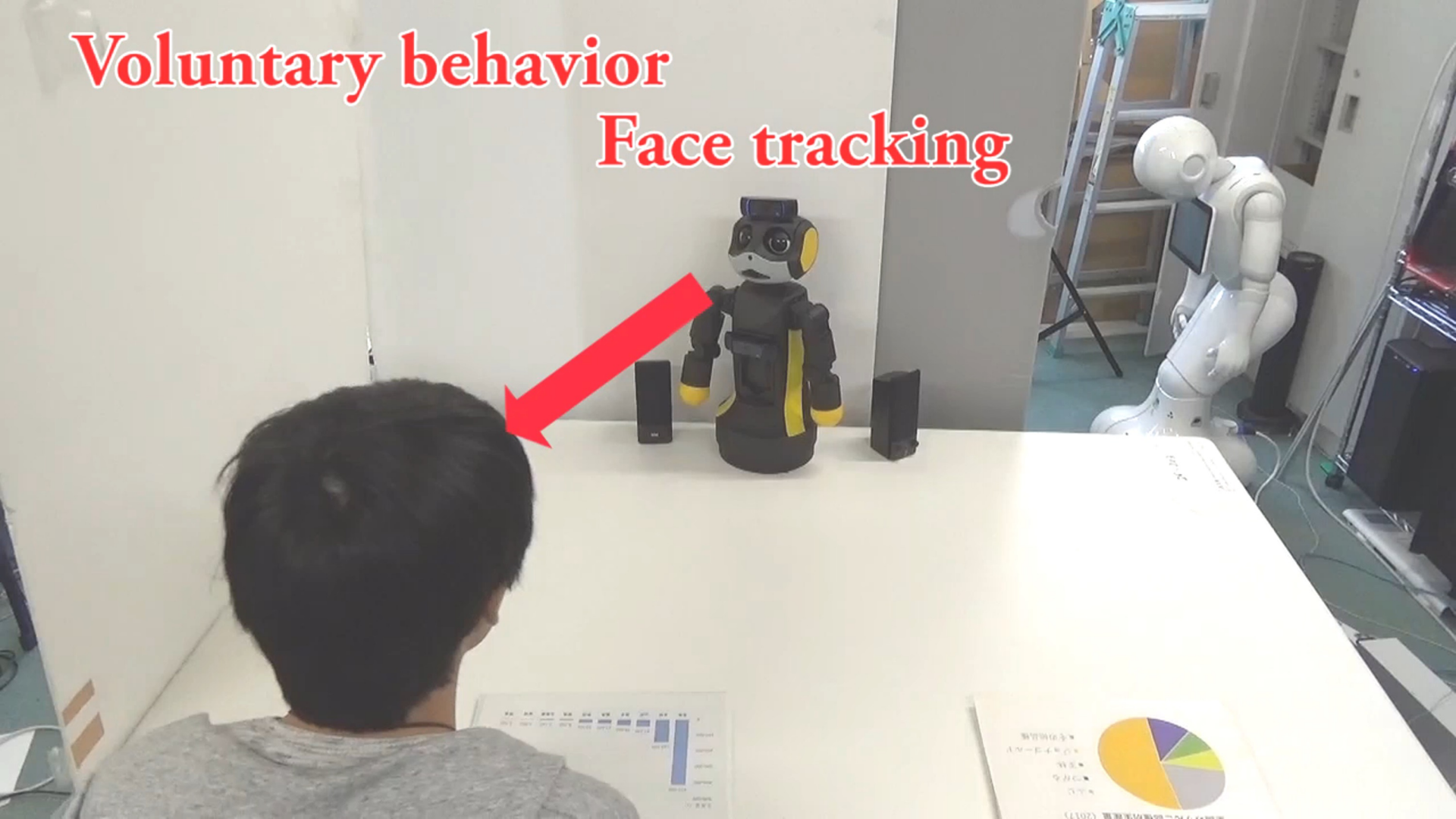


エージェント

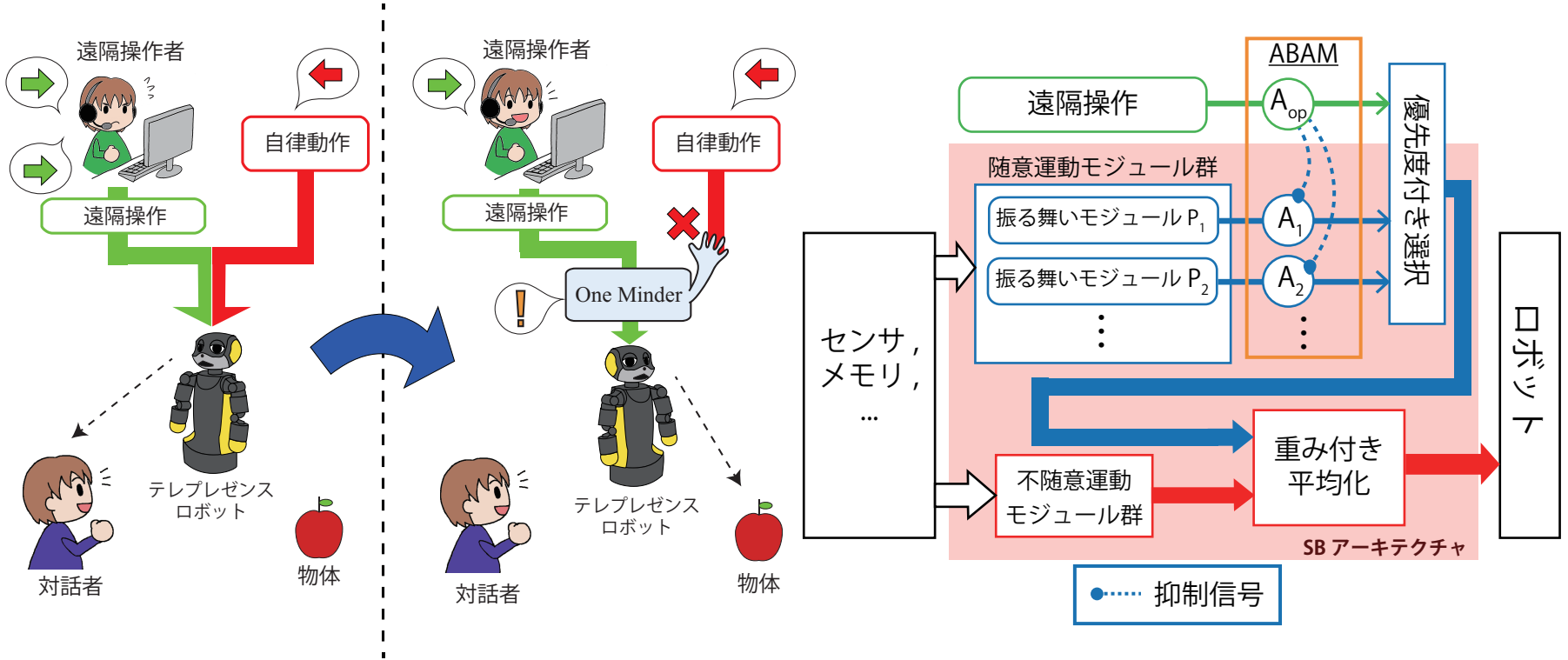
適応



人

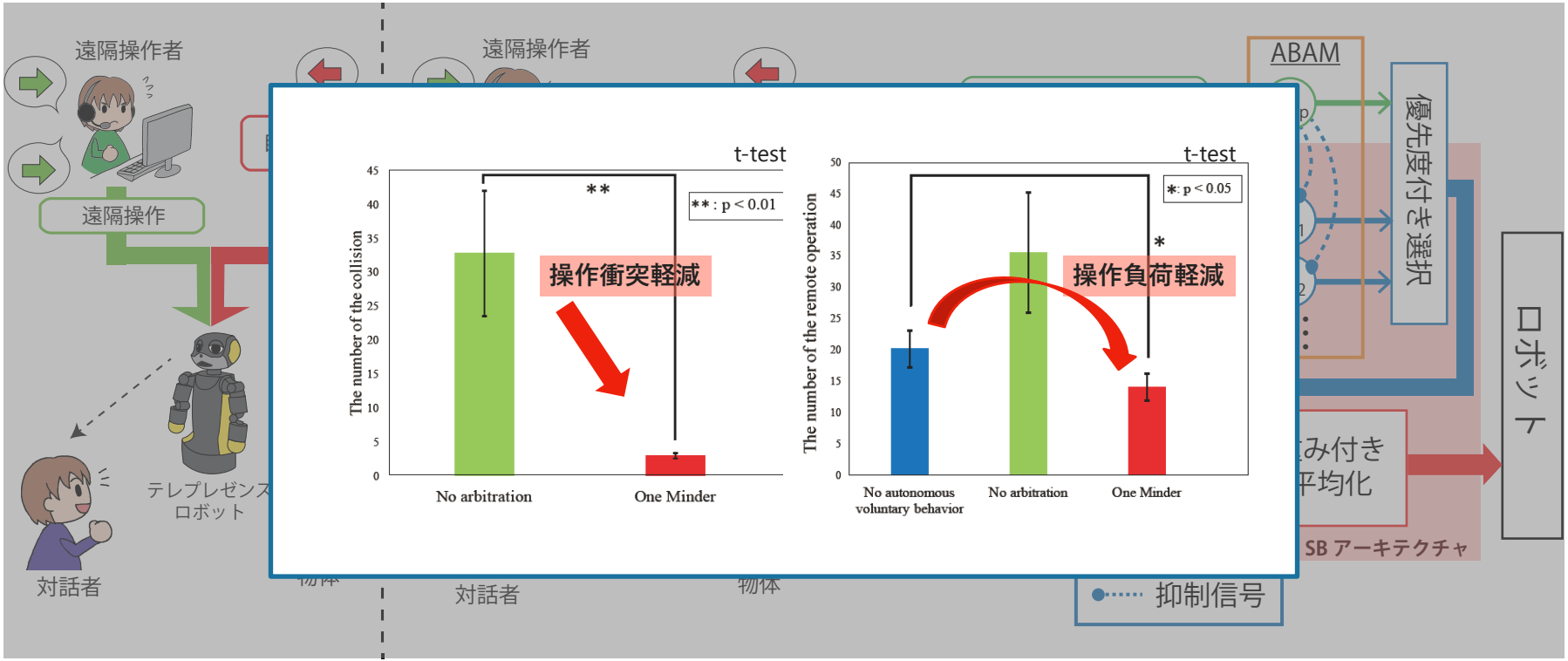

Voluntary behavior
Face tracking

自律動作を調停する遠隔操作ロボット

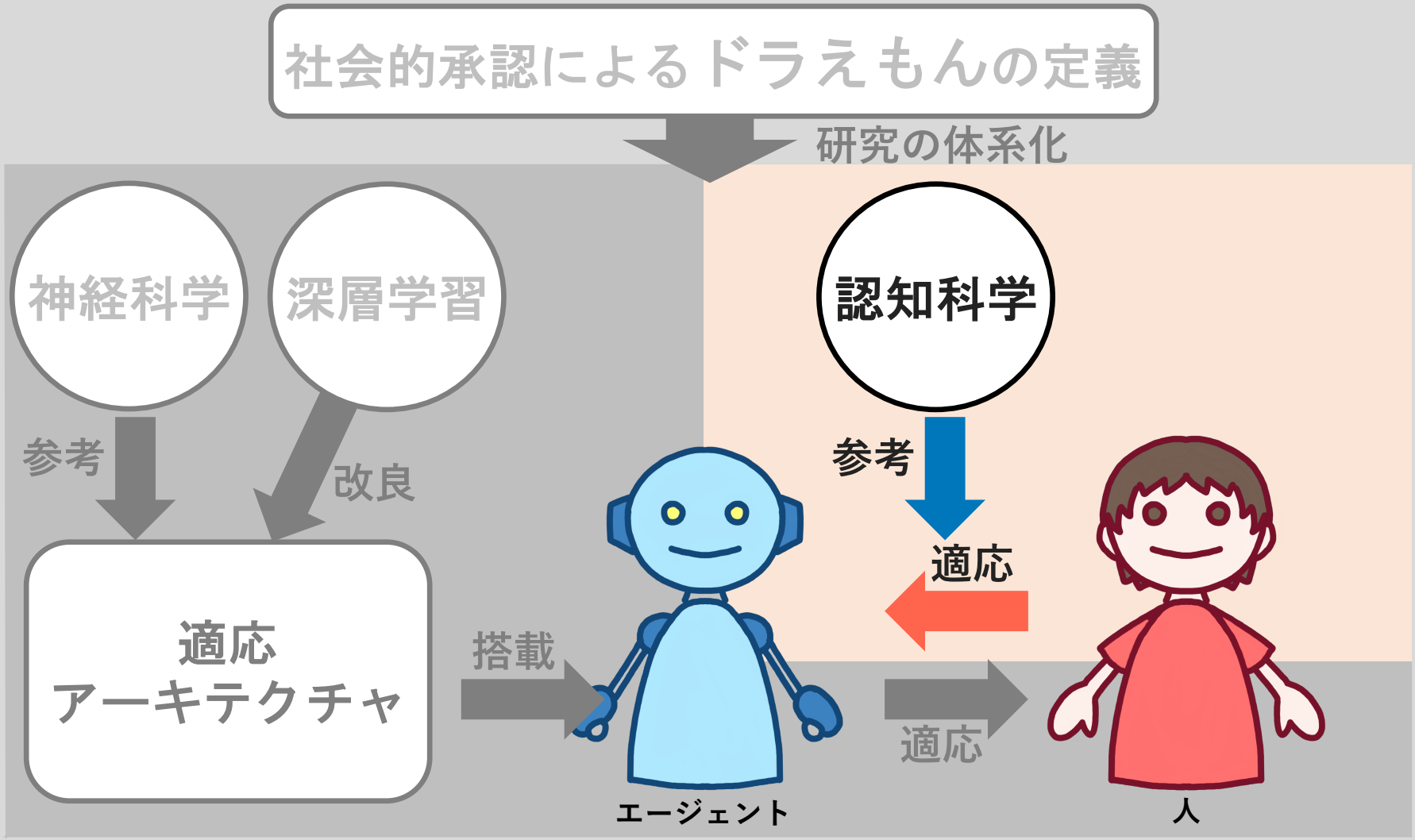


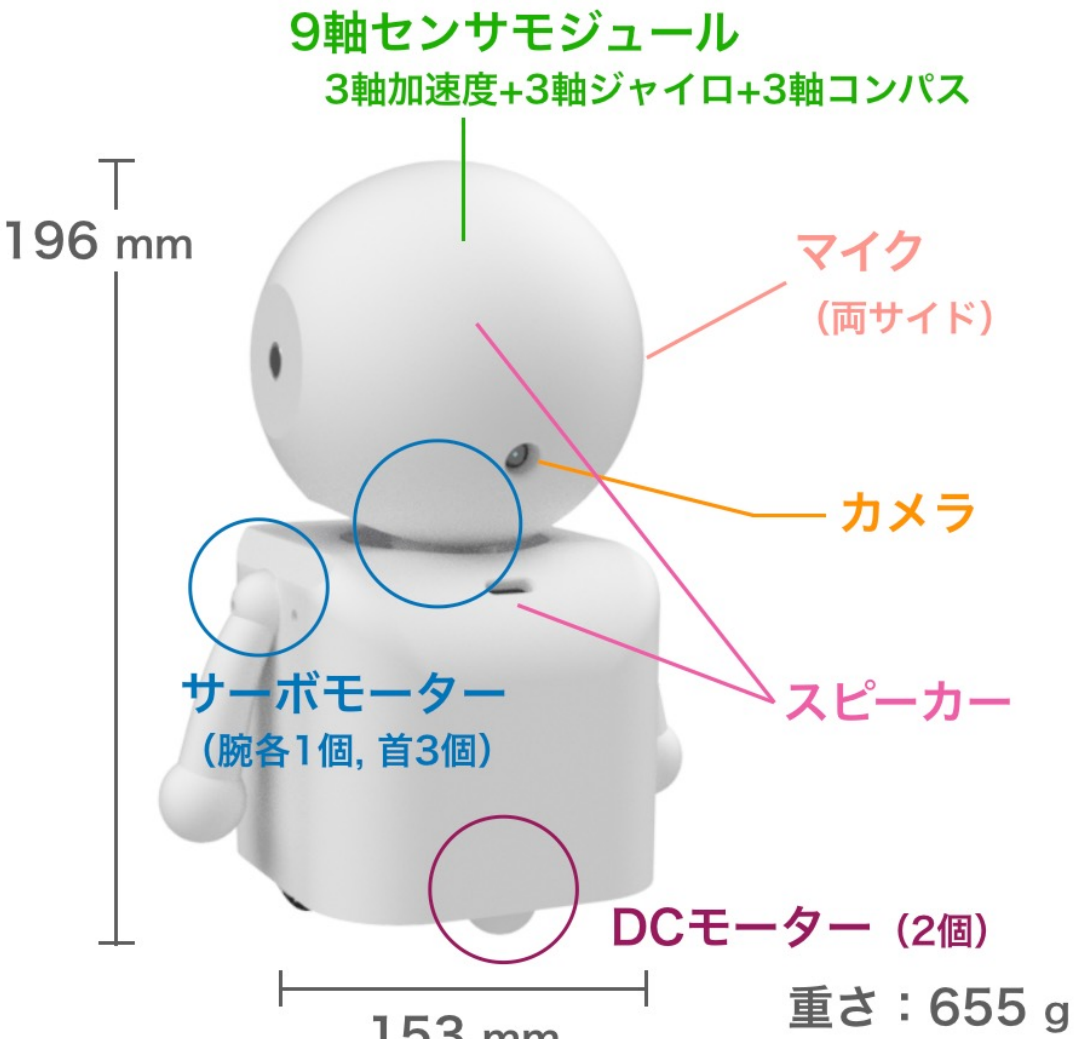
奥岡,大澤,今井, “遠隔操作と自律操作を適応的に切り替える半自律テレプレゼンスロボットアーキテクチャ”, 人工知能, 2021-3.

自律動作を調停する遠隔操作ロボット

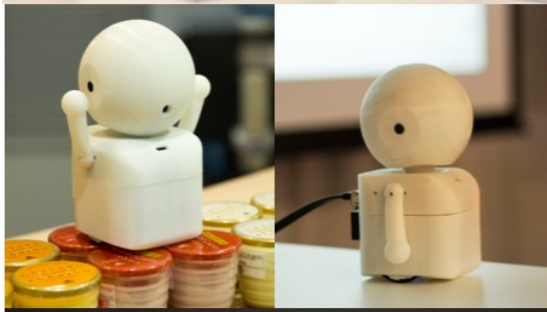


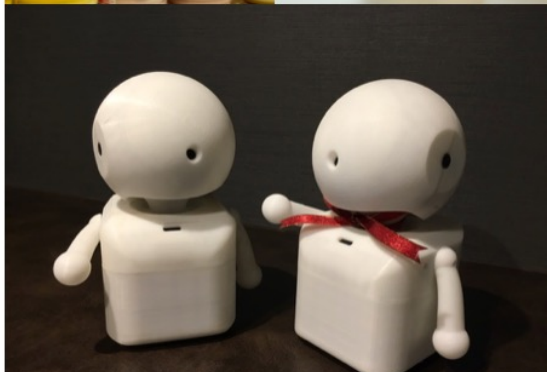
奥岡,大澤,今井, “遠隔操作と自律操作を適応的に切り替える半自律テレプレゼンスロボットアーキテクチャ”, 人工知能, 2021-3.


社会的承認によるドラえもんの定義
研究の体系化
神経科学
深層学習
認知科学
参考
改良
参考
適応
適応
アーキテクチャ
搭載
適応
エージェント
人

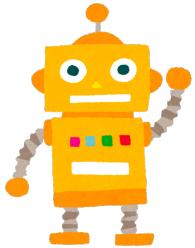

9軸センサモジュール
3軸加速度+3軸ジャイロ+3軸コンパス
196 mm
マイク
（両サイド）
カメラ
スピーカー
サーボモーター
（腕各1個, 首3個）
DCモーター（2個）
153 mm

重さ：655 g





「言葉を使わずに」しりとり



「りんご」

ドララ!

「ごりら」…って言ったの？
じゃあ、「らっぱ！」

エージェントデザインのトレードオフ

技術的課題
「心を想定される」と「役に立つ」ことの両立が困難

非自然言語
エージェント

自然言語
エージェント

◎心を想定される
△ 役立つ

△心を想定される
◎ 役立つ

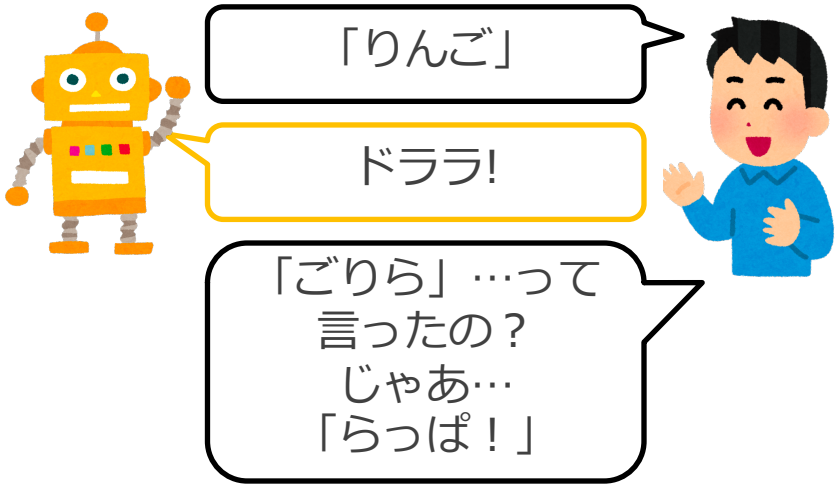
役に立つために自然言語能力が重要

技術が未成熟なため不適切な応答

ユーザに内部設計を見透かされる

ユーザは心ではなく内部設計を想定

エージェントデザインのトレードオフ



●概要

文脈に基づいて表出の意図を推定させる

→自然言語を用いたものと同様のインタラクション

●実現できたこと

ユーザに内部設計を見透かされずに

自然言語的なインタラクションが実現

●実現できていないこと

ユーザの予測のきっかけとなる文脈生成

→文脈がないと役に立てなくなる

文脈生成さえできれば、

「心を想定される」ことと「役に立つ」ことを両立

ITACO x 非自然言語エージェント

- 非自然言語エージェントが文脈を作れない欠点があったが、
  「乗り移る」ことで新たに文脈を自律的に生成可能に
  e.g. 冷蔵庫に乗り移ったということは、
  食材について伝えたい？

実社会において「心を想定される」「役に立つ」を両立

HAIはWBAの中間成果物として有効か？

- 社会的承認による定義を考えた時、私は…

  HAIはAGIの中間成果物として有効と考えている

  HAIはWBAの中間成果物として有効であることを期待している

- 論点は、生物の脳と整合性がとれる情報処理アーキテクチャがHAIにおいて

  ① 人から他者モデルを想定されやすい

  ② 人に対して他者モデルを想定しやすい

  を満たすか？

OSAWA Laboratory

Fin.