



# 第5回 WBAレクチャー趣旨説明 全脳計算モデルの実現に向けた 認知モデルの役割

静岡大学 情報学部 行動情報学科

森田純哉

# WBAIのゴール



協力者として参加するには

[ホーム](#) [イベント・勉強会](#) [研究](#) [活動に参加する](#) [ブログ](#) 言語: 

## WBAIについて

Home / WBAIについて

全脳アーキテクチャ・イニシアティブ（WBAI）は、脳全体のアーキテクチャに学ぶことにより人間のような知的能力を持つ汎用人工知能の実現を目指す研究開発活動にたいして支援と促進を行うことを目的とする NPO（特定非営利活動法人）です。

WBAIは、特定非営利活動に係る事業として教育事業と研究開発事業を行っています。教育事業としては、勉強会やハッカソンの開催などを行っています。研究開発事業としては、汎用人工知能研究を側面から支援するためのソフトウェアプラットフォームの開発を行う他、関連する研究者や研究コミュニティとの交流を行っています。

# 人間のような知的能力を持つ汎用人工知能の実現

# 汎用人工知能 (AGI) に対する 一般的イメージ

ChatGPT 3.5 ▾



**You**

ChatGPTは汎用人工知能と呼べるでしょうか。



**ChatGPT**

はい、ChatGPTは一般的な人工知能 (AI) と呼ばれることができます。ChatGPTは自然言語処理や対話生成などのタスクにおいて広範な応用が可能なAIモデルです。そのため、汎用的な人工知能と見なすことができます。ただし、ChatGPTは特定のタスクや領域に特化しているわけではなく、多岐にわたる問題に対応できる柔軟性を持っています。



# OpenAIのスタンス

<https://openai.com/research/overview>



[Menu](#)

Pioneering research  
on the path to AGI

We believe our research will eventually lead to artificial general intelligence, a system that can solve human-level problems. Building safe and beneficial AGI is our mission.

[View research index](#)

[Learn about safety](#)

# OpenAIのスタンス

<https://openai.com/research/overview>

“Safely aligning powerful AI systems is one of the most important unsolved problems for our mission. Techniques like learning from human feedback are helping us get closer, and we are actively researching new techniques to help us fill the gaps.”


???

Josh Achiam  
Researcher at OpenAI

# AGIに向けた別のアプローチ



協力者として参加するには

[ホーム](#) [イベント・勉強会](#) [研究](#) [活動に参加する](#) [ブログ](#) 言語: 

## WBAIについて

Home / WBAIについて

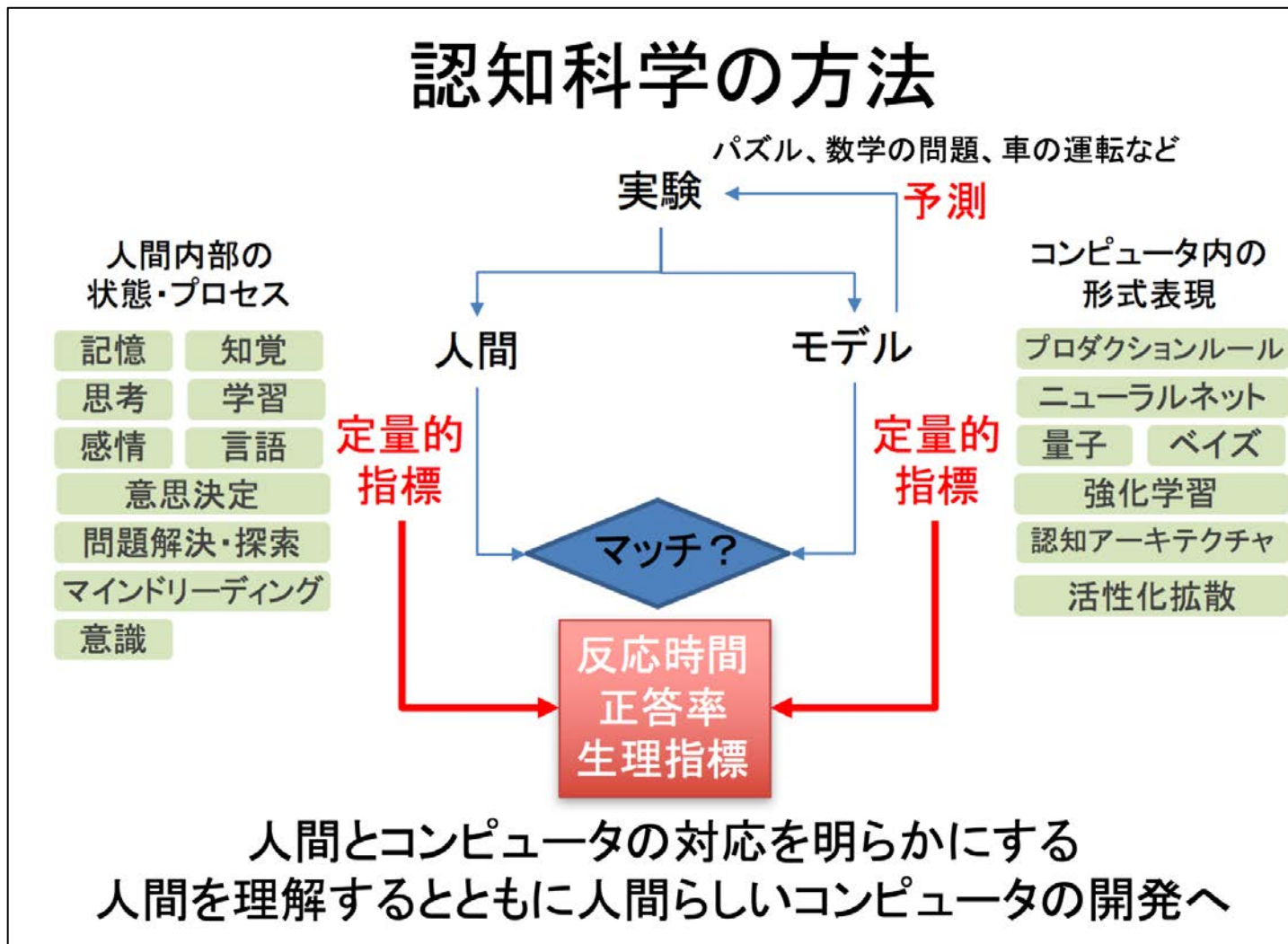
全脳アーキテクチャ・イニシアティブ（WBAI）は、脳全体のアーキテクチャに学ぶことにより人間のような知的能力を持つ汎用人工知能の実現を目指す研究開発活動にたいして支援と促進を行うことを目的とする NPO（特定非営利活動法人）です。

WBAIは、特定非営利活動に係る事業として教育事業と研究開発事業を行っています。教育事業としては、勉強会やハッカソンの開催などを行っています。研究開発事業としては、汎用人工知能研究を側面から支援するためのソフトウェアプラットフォームの開発を行う他、関連する研究者や研究コミュニティとの交流を行っています。

<https://wba-initiative.org/about/>

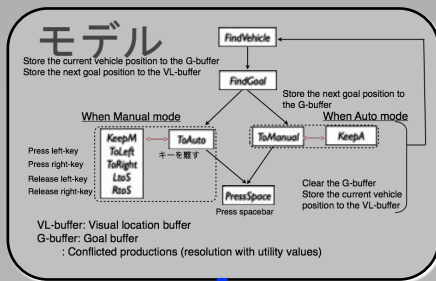
# アラインメントの方法

- 静岡大学情報学部  
「認知科学と行動情報」  
(1年・必修)授業スライド



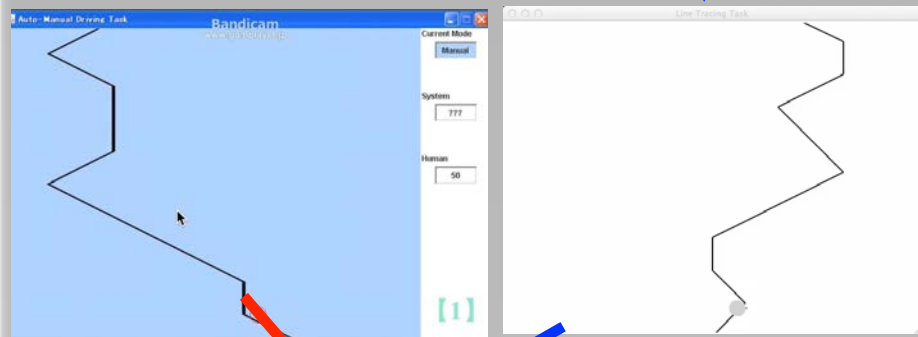
# 認知モデルの研究事例

自動操縦の利用 (Morita et al. CogSci2011)

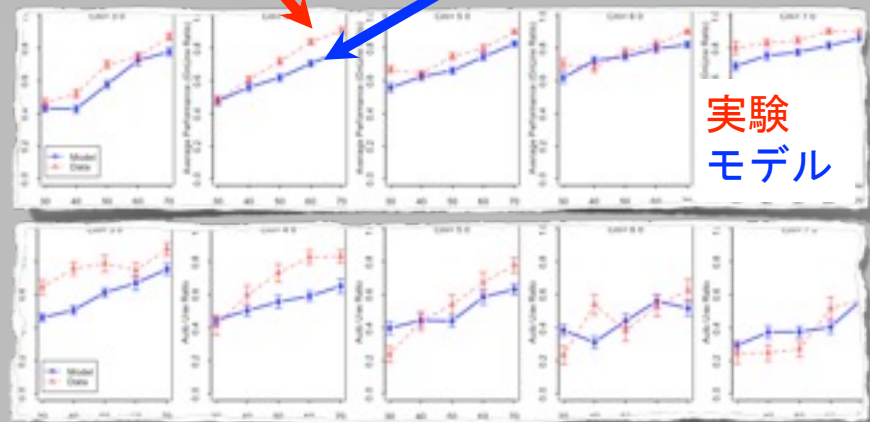


実験

シミュレーション



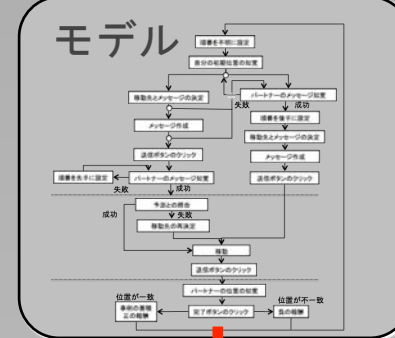
結果の対応づけ



実験  
モデル

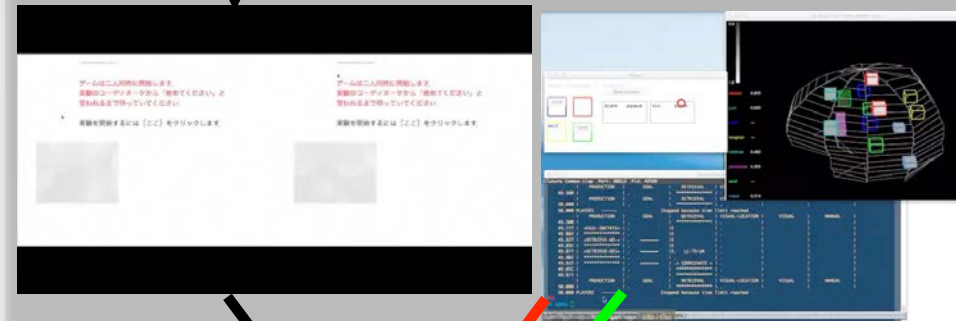
階層的強化学習の必要

コミュニケーションの発生 (Morita et al. CogSci2012)

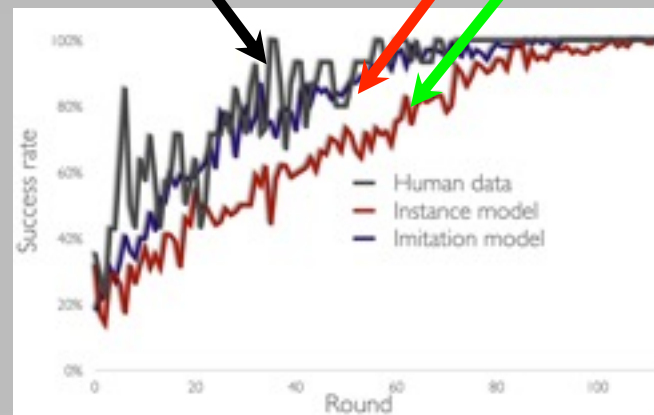


実験

シミュレーション



結果の対応づけ



相互模倣の必要



特集 | 編集委員 今年の抱負 2024 |

認知モデラーとしてのサステナブルなライフワークの  
下地づくり

森田 純哉 静岡大学

## だが、しかし

ふと周りを見渡すと、このアプローチは、今や人工知能研究にとっても、認知科学にとっても、**絶滅危惧種となりつつある**。論文や研究費のリジェクト率、査読者から受けるコメント、授業での受講生の反応、いずれの方面からも心を折るフィードバックが増えている。もちろん、この背景には、単純な作業の質の低下など、いろいろな要因があり得る。しかし、自身のアプローチが時代に合わなくなっているのでは？という疑念も否定できない。

年代によらず、現代に生きる多くの人々が、時代の流れの速さを感じている。特に「人工知能」をめぐる状況は、まさに日進月歩である。この分野を本職とする研究者でさえ、その進歩を追うことが難しくなっている。今や専門的な文献からだけでなく、マスメディアや大学の運営会議など、多方面から分野の最新情報が入ってくる。目まぐるしく変化する State of the Art (SOTA) を尻目に、一時のトレンドに惑わされない長期的な仕事【持続可能なライフワーク】を追求したいと多くの研究者が願っているのではないだろうか。

筆者自身、上記の願望を抱く一人である。筆者のこれまでのライフワークは、認知モデル・認知アーキテクチャの応用であった。認知モデルの研究は、情報処理として人間を理解し、その理解を計算機上で確認することによって進められる。この作業は、認知アーキテクチャと呼ばれるツールを使って行われることがある。認知アーキテクチャとは、複数の認知モデルに共通する構造であり、認知モデルを記述するプログラミング言語の形態をとる。人間の情報処理に対する研究者間の理解を統合するコモングラウンドと言い換えてもよい。そして、このコモングラウンドは、認知アーキテクチャを利用した認知モデルの開発を通して更新される。つまり、この分野の研究者は、コミュニティとして人間の情報処理の理解を追求していく実践者となる。

こういった理解をベースとしたアプローチは人工知能研究の源流の一つである。さらに、人工知能研究の近接領域である認知科学においても、心理学実験をシミュレーションする認知モデルの開発が、かつて盛んに行われていた。筆者自身、学部生であった頃にこのアプローチに魅せられ、この世界にはまり込んでいった。これまでの研究者人生を通し、さまざまなモデルに触れ、その度に目を見張るような驚きを感じ、思考の枠組みの転換を繰り返してきた。当時から28年を経た現在も、この分野への知的好奇心は減衰していない。この仕事を生涯にわたり続けたとしても、飽和することはないという自信がある。さらに、認知モデルの構築によって得られる理解を世に広めることこそ自身の使命であるという揺るがない信念（思い込み）も有している。

しかし、ふと周りを見渡すと、このアプローチは、今や人工知能研究にとっても、認知科学にとっても、絶滅危惧種となりつつある。論文や研究費のリジェクト率、査読者から受けるコメント、授業での受講生の反応、い

ずれの方面からも心を折るフィードバックが増えている。もちろん、この背景には、単純な作業の質の低下など、いろいろな要因があり得る。しかし、自身のアプローチが時代に合わなくなっているのでは？という疑念も否定できない。

2016年頃の深層学習ブームの開始期は、主流の方向ではないにせよ、認知モデル・認知アーキテクチャの研究への期待を感じていた。しかし、最近では、人々の注意が GPT-n などに代表される「生成 AI」に惹き付けられ過ぎている。おそらく、この危惧は、認知モデルの研究者に限らず、人工知能の多くの分野の研究者が抱えているものだろう。理解を目指す人工知能「研究」にとって、厳冬の時代を感じさせる空気が漂っている。

2024年とは、この流れに抗い、認知モデラーとしてのライフワークを続ける下地づくりの一年としたい。そして、幸いなことに本稿を執筆している2023年10月上旬の時点で、いくつかの好材料が得られている。

材料1：生成 AI と認知アーキテクチャの統合に向けた国際コミュニティ

一つの材料は、国際的なコミュニティにおける認知アーキテクチャと生成 AI の統合に関する議論である。2023年10月末、AAAI 秋のシンポジウムにて、認知モデルの分野をリードする研究者により「Integration of Cognitive Architectures and Generative Models」と題するワークショップが開催される。事前公開されているプロシーディングスによれば、主に三つの方向が議論されそうである。

1. 生成 AI を制御する構造としての認知アーキテクチャ。生成 AI の学習におけるモダリティ間のインタラクションを認知アーキテクチャにより、明示的に設計する。これにより、ブラックボックス化された AI から信頼される AI への転換が目指される。
2. 生成 AI による認知モデルの動作・内部状態の可視化。旧来の抽象化された認知モデルに生成モデルを組み入れる。これにより、人間と同様にイメージを生成し、言語を生成するタスクのプロセスがデモンストレーションされる。このようにデモンストレーションされたモデルは、研究者以外の人間によっても容易に理解でき、かつ人間との共同作業を可能にする。つまり、社会にグラウンディングした認知モデルが誕生する。
3. 生成モデルの利用による認知モデル構築のコスト

# 生き延びるために

## 材料2：生成AI時代に**認知モデルの意義をアピールする特集**

国際的な場での活動を持続的なものとするためには、自身が生活をする地域での活動の下地（地域からの理解）も必要である。つまり、筆者のライフワーク継続の下地の一部は、日本のコミュニティに対して、この分野の素晴らしさを伝えていくことでつくられる。

の低下、プログラミングコードをChatGPTによって生成するのと同様、生成AIによって認知アーキテクチャに搭載する知識を構築する。人間の思考の背後にあるルールセットを抽出する、あるいは記号的な知識の背後に存在する分散表現を抽出する。生成AIを活用した認知モデルの構築は、研究者の負荷を下げるとともに、認知モデリングの普及につながる。高度なプログラミングスキルを有する研究者のみが構築できた認知モデリングが一般化される未来が描かれる。

筆者自身、現在、上記の二つ目のアプローチの研究に取り組んでおり、現地での発表を予定している。また、このワークショップは、生成モデルを組み入れた次世代認知アーキテクチャに向けたディスカッションが大半を占める。このような国際的な場へのコミットを通して、自身の研究を持続可能なものとし、コミュニティへの貢献を目指していく。

### 材料2：生成AI時代に認知モデルの意義をアピールする特集

国際的な場での活動を持続的なものとするためには、自身が生活をする地域での活動の下地（地域からの理解）も必要である。つまり、筆者のライフワーク継続の下地の一部は、日本のコミュニティに対して、この分野の素晴らしさを伝えていくことでつくられる。

先述したように、認知モデルは絶滅の危機に瀕している。とはいえ、筆者自身は、このアプローチの意義が失われたとは全く考えていない。その存在意義は、人工知能研究の開始時に語られたような人間らしい人工物をつくるというわかりやすいものではない。深層学習の爆発的な発展以降、人間を理解することが、知的な人工物をつくる必要条件でないことを、多くの人は理解している。

筆者が考える認知モデリングの存在意義は、一つには人間に対する理解への欲求の充足である。筆者自身の知的好奇心が尽きないことに示唆されるように、人間には理解への根源的な欲求が存在すると考えられる。筆者は、その欲求（報酬）は、知識を利用して環境からパターンを発見することで生じると考える[長島21]。逆に言えば、環境から新規なパターンが発見されなくなることで、飽きが生じる。この点に関して、人間を対象とした認知モデリングほど知的好奇心を駆動するものはない。モデルとはパターンであり、変化する人間の行動に対して無限に生成される。

しかし、上記のような抽象的な説明では、この時代に生きる人々へのアピール材料とはならないだろう。人を惹き付けるには、やはり、この活動により駆動される知的好奇心（楽しさ）の具体例を示す必要がある。この点

に関して、ありがたいことに会誌にて特集の企画を組む話をいただいた。自身が認知モデルの道にはまり込むきっかけをつくっていただいた先生方、同じ時代のなかで同期をしてきた方々、そして次代の担い手となる方々に、生成AI時代の認知モデリングの面白さ、およびその行く先を描いていただく。本特集はタイムリーなことに次号（Vol. 39, No. 2）にて掲載の予定である。

### 材料3：認知モデルをベースとした社会デザインプロジェクト

先ほどは、最近の研究費や論文投稿に関するネガティブな状況を記載した。しかし、折られた心を継ぎ接ぎしつつ、多方面へのアピール（もがき）を続け、認知モデルをテーマとする研究の機会をいただくことができた。2023年11月より、JST RISTEX（科学技術振興機構社会技術研究開発センター）「SDGsの達成に向けた共創的研究開発プログラム（情報社会における社会的側面からのトラスト形成）」のもとで、「ローカルエコチェーンをステアリングするトラスト調和メカニズムに関する認知科学的検討」をスタートさせる。

このプロジェクトは、筆者が所属する静岡大学情報学部にも所属する人工知能、認知科学、社会科学の7人の研究者が関わる。地域社会を活性化させるSNSを構築するという、ある種わかりやすいアウトプットを掲げつつ、その背後に存在する集団ダイナミクスを認知モデルによって解明することを目指す。

プロジェクトの採択から、現代における認知モデルの意義は、知的好奇心の充足以外でも主張可能であることが示される。情報社会において生じている問題の多くは、人間の認知の仕組みから始まる。人間は進化の過程のなかで、生存のために、種々の認知的傾向（バイアス）を獲得してきた。しかし、急速に変化する環境に対し、旧来のバイアスは不適応を起こしている。その具体的な例が、エコチェーンバや行動嗜癖である。

こういった問題を解決していくためには、現状に流されるデータ駆動のアプローチではなく、あり得る社会の姿を描いたうえで理解に基づくアプローチが必要になる。筆者は認知モデルこそが現代社会をステアリングする鍵となると考えており、この考えを本プロジェクトにおいて実証し、自身のライフワークの下地を盤石なものとする所存である。

### ◇ 参考文献 ◇

[長島 21] 長島一真, 森田純哉, 竹内勇剛: ACT-R による内発的動機づけのモデル化. 人工知能学会論文誌, Vol. 36, No. 5, pp. AG21-E\_1-13 (2021)

## 【会誌発行】人工知能学会誌 Vol.39 No.2 (2024/3)

巻頭言  
[学会の魅力向上への取組み：多様性と変化への対応に向けて](#) ..... 小野 智弘 91

特集：「人の学びとAI」  
[特集「人の学びとAI」にあたって](#) ..... 山元 翔・平嶋 宗 92  
 哲学から見た時間と学習 ..... 三宅 陽一郎 94  
 「分からない」状態のモデル化 ..... 溝口 理一郎 103  
 AI・ビッグデータによるアダプティブラーニング ..... 植野 真臣 111  
 対話型LLMとのインタラクションに起因する学習の機相の探究に向けて ..... 中小路 久美代・岡 隆之介 118  
 学習科学に基づく人の学びとAI  
 — 未来のクラスルームアナリティクス — ..... 白水 始・飯窪 真也・齊藤 萌木・益川 弘如 125  
 ロボットによる学習支援における非言語コミュニケーションの効果 ..... 松居 辰則 134  
 ユビキタスな学習環境とリアルワールド教育データが実現する学びの変革 ..... 緒方 広明・堀越 泉 143

特集：「生成AI時代における認知モデリング」  
[特集「生成AI時代における認知モデリング」にあたって](#) ..... 森田 純哉 153  
 認知モデリングにおける記号処理アーキテクチャ ..... 三輪 和久 155  
 認知モデリング研究は大規模言語モデル時代をどう生き抜くか？ ..... 内海 彰 163  
 生成AIと認知モデリングの比較  
 — 活用例と倫理的問題 — ..... 松室 美紀 171  
 認知アーキテクチャと生成AIが織りなす人と機械の共通基盤 ..... 森田 純哉・東中 竜一郎・竹内 勇剛 178  
 手続き合理性に基づく学習とインタラクションのモデル ..... 安西 祐一郎 187  
 解釈可能性を高めて信頼し得るエージェントを実現するための脳型認知モデル  
 ..... 田和辻 可昌・大森 隆司・太田 博三・宮本 竜也・芦原 佑太・荒川 直哉・山川 宏 197  
 脳科学とロボティクスを結ぶ統合認知モデルの実践と展望 ..... 谷口 彰・山川 宏・谷口 忠大 205  
 大規模言語モデルはドラえもんになれるか ..... 大澤 正彦 214  
 ChatGPTの知らない「意図推定の世界」 ..... 植田 一博 222  
 人は人の理解をやめるか  
 — AIを用いた詳細価値観モデリングと精密行動予測の可能性 — ..... 寺田 和憲 230

レクチャーシリーズ：「AIと社会と人間～ぶつかる・なじむ・とけこむ～」(第8回)  
 AI化社会における人の在り方に関する一考察 ..... 武田 英明 240

研究会報告  
[合同研究会2023 開催報告](#) ..... 大原 剛三・村田 博士・松井 藤五郎・馬場 雪乃・砂山 渡・服部 宏充 248

私のブックマーク  
[人の移動](#) ..... 落合 桂一 256

グローバルアイ (第69回)  
[PhD 学生のアメリカンライフ～NY 州イサカへ](#) ..... 久壽米木 啓悟 260

学生フォーラム (第121回)  
[植原絵里奈先生インタビュー「プログラミング教育の魅力と個性の尊重」](#) ..... 大西 舞子・村上 綾菜 263

会議報告  
[2023 IEEE International Smart Cities Conference](#) ..... 三宅 陽一郎 267

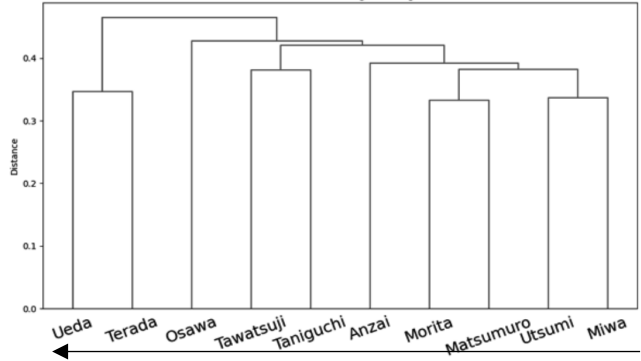
書 評  
[一般社団法人情報科学技術協会 監修 南山泰之 編 池内有為、尾城孝一、佐藤翔、林和弘 林豊著：オープンサイエンスにまつわる論点 変革する学術コミュニケーション pp.168 榎村房 \(2023\)](#) ..... 池田 圭佑 270

アーティクル  
[表紙企画 人工知能をめぐる旅：JSAI 研究会の窓② — 人工知能基本問題研究会 \(FPAI\) —](#)  
 ..... 企画：榎 剛史・山野 泰子・井上 昂治・山本 祐輔, イラスト：岩澤 駿 271

会 告 ..... 1～5  
 表紙イラスト・デザイン：©岩澤 駿

# 人工知能学会誌特集 (3月)

GPTによる埋込み表現を用いた記事間のクラスタリング  
 Hierarchical Clustering Dendrogram



1. 生成 AI 時代における認知モデリングの現状と展望
2. 生成 AI と認知アーキテクチャ
3. 生成 AI と認知モデルの自覚的な橋渡し
4. 次のステップとしての解釈可能性
5. 生成 AI と認知モデルの日本的融合
6. 生成 AI と認知モデルが導く社会

# 人工知能学会誌特集（3月）

認知モデリングは人間の認知機能を計算機上で模倣する学術的な営みである。その目的は、人間あるいは生物が有する認知機能の理論的な理解である。

認知モデリングの研究は長い歴史をもつものの、現在、**Big Tech**により構築された生成AIの影響を強く受けている。

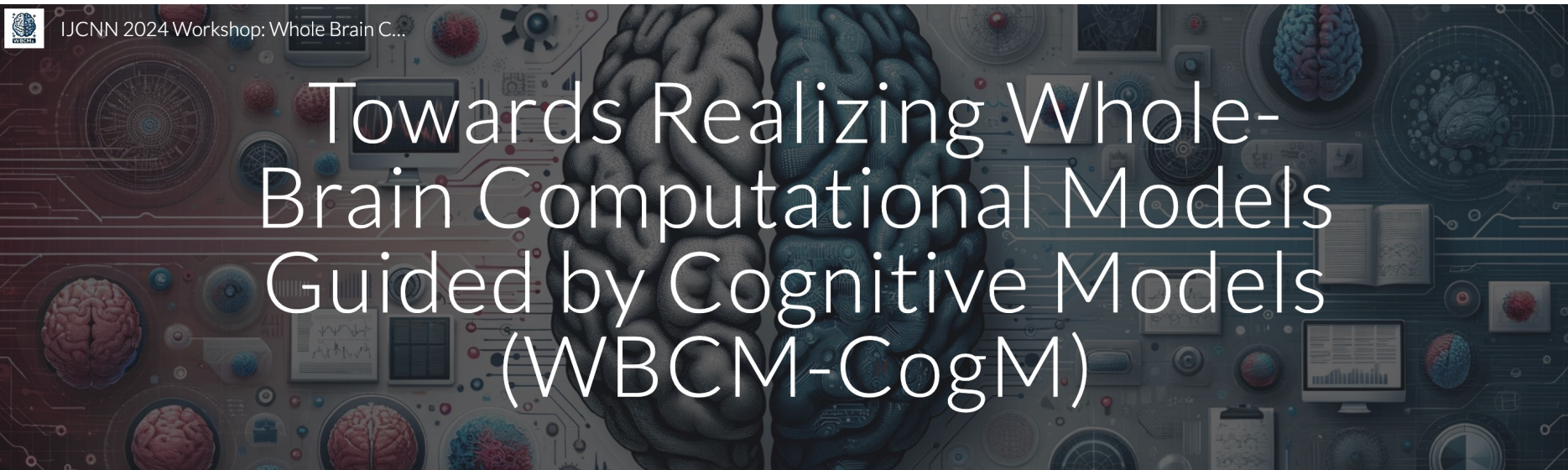
生成モデルの出現は、人間に対する理解の仕方を大きく変化させつつ、同時にAI研究における認知モデリングの意義を揺らがせている。以前は、人間あるいは生物を理解することは知的な人工物の構築に役立つというストーリーが成り立った。しかし、近年の生成AIの構築において、心理学や言語学の理論が利用されたという話は聞かない。

人間を理論的に理解することの応用的価値が問われる現代において、認知モデルの研究はどのような方向に向かうべきか。あるいは、どのようにこの時代の中で研究を持続可能とし、社会にインパクトを与えていくか。

# Future

認知モデルによりガイドされる全脳AI

*How can we develop socially agreeable interpretable (explainable) brain-inspired cognitive architectures?*



## Summary:

In recent years, there has been a focused effort to develop Whole Brain Computational Models (WBCMs), aiming to represent the entire brain's functions and contribute to creating artificial intelligence with human-level capabilities. WBCMs involve not only neuroscientific but also cognitive models, especially in constructing a cognitive architecture for consistency. Cognitive models enhance interpretability in implementing WBCMs into AI agents, providing insight into thought processes. This approach, resembling human cognition, offers potential psychological reassurance to users. The discussion about the relationship between cognitive models and WBCMs is linked to AI alignment debates, crucial as powerful AI systems develop. The workshop aims to discuss methodologies to realize WBCMs, emphasizing the role of cognitive models.

## The topic of interests:

- Cognitive model / architecture
- Whole-brain computational model / Neural network
- AI alignment / Interpretability
- Artificial general intelligence
- General purpose robotics
- Predictive coding
- Free energy principle
- Emotion and memory
- Generative AI / Large-Language Model / GPT Utilization
- Experiment-based model construction

6/30-7/5のうち1日  
投稿期日: 5/1



# 認知アーキテクチャと生成AIが織りなす 人と機械の共通基盤

静岡大学 情報学部 行動情報学科

森田純哉

# 森田純哉

静岡大学情報学部行動情報学科 教授 (2023.4～)

- 経歴

- 静岡大学情報学部行動情報学科 准教授 (2016 -)
- 名古屋大学未来社会創造機構特任助教 (2014 -2016)
- 北陸先端科学技術大学院大学助教 (2006-2014)
- 名古屋大学人間情報学研究科博士課程修了 (2006.3)

- 所属学会

- 日本認知科学会: 常任運営委員 (国際担当)
- 人工知能学会: 編集委員

# 研究分野

- 応用認知モデリング研究室
  - 認知モデル・認知アーキテクチャの社会応用を目指す研究室

# Applied Cognitive Modeling Laboratory

静岡大学情報学部 森田研究室 認知モデルを利用したミラーリングエージェント Applied Cognitive Modeling Lab(ACML)

研究コンセプト 研究アプローチ 研究内容 研究実施者 論文リスト

- 行動を生み出す心の情報処理モデル**  
人間の行動を再現するコンピュータプログラムの開発に取り組んでいます。  
[詳細△](#)
- 情報技術を利用した人間行動の定量的分析**  
マルチモーダルなデータをモデルベースで分析することで、人間の知性や感情の状態を推定し、その発生メカニズムを理解することを目指しています。  
[詳細△](#)
- 情報技術による行動変容への働きかけ**
  - ・デザイン教育
  - ・大学生/大学院生による認知科学の学習
  - ・高齢者の心の問題への支援[詳細△](#)

2016年4月に設立された研究室です。行動情報学の領域でメタ認知強化装置としての認知モデルの研究開発を進めています。自分の認知を客観化し、よりよい自分へ向けた自己決定へ。

Activity [一覧を見る](#)

- 2024/3/2 [人工知能学会・特集「生成AI時代における認知のモデリング」](#)
- 2023/11/25 [人工知能学会 第99回 先進的学習科学と工学研究会\(SIG-ALST\)](#)
- 2023/11/12 [静岡大学浜松キャンパス・第28回テクノフェスタでの研究者展示](#)

Pickup 研究コンセプト [詳細を見る](#)

Pickup 研究アプローチ

関連リンク

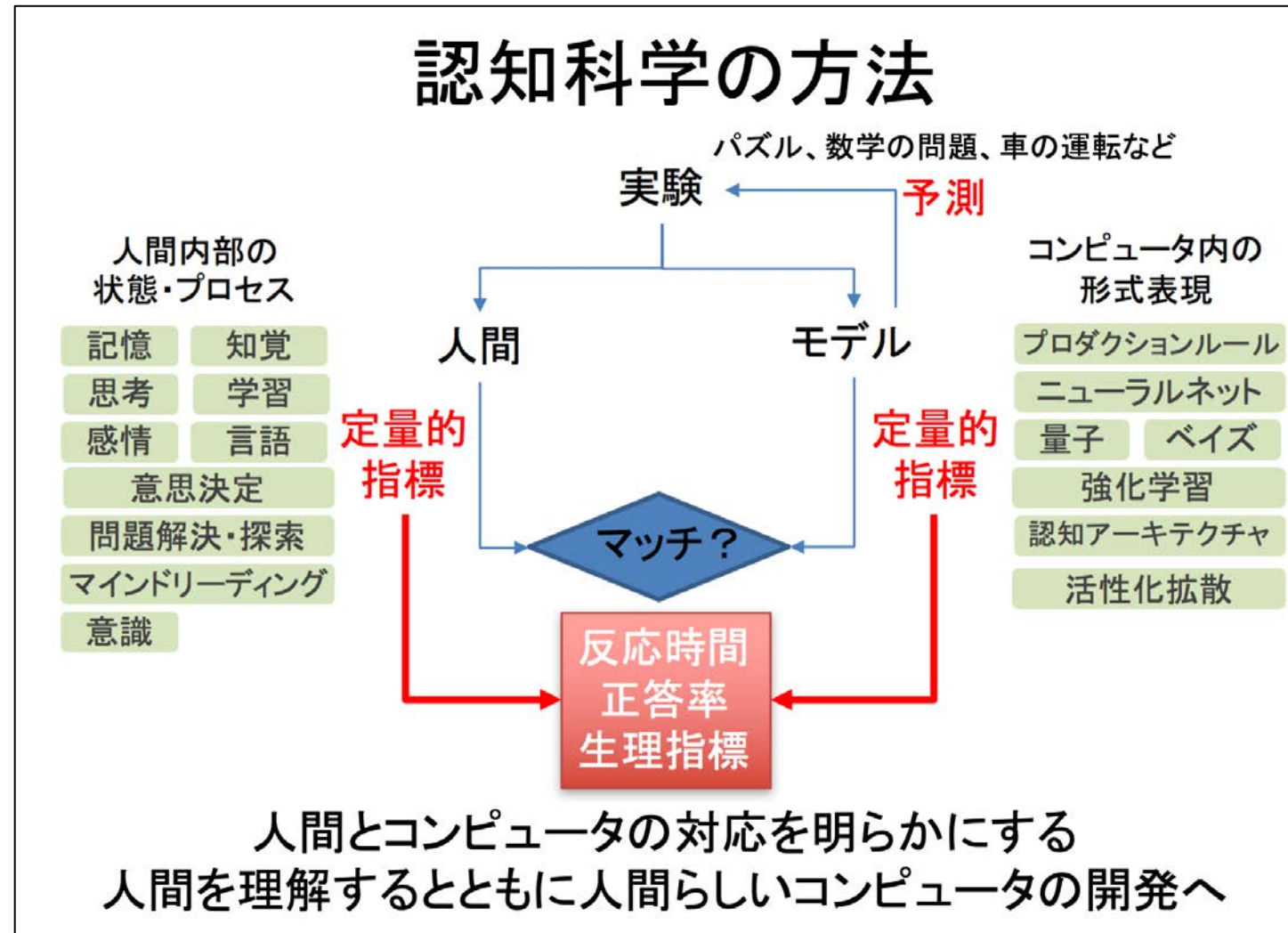
- ・RISTEXデジタルトラスト (~2026)
- ・脳研新学術：対話知能学 (~2023)
- ・脳研新学術：共創言語進化 (~2021)
- ・脳研新学術：認知的インタラク ション (~2018)
- ・静岡大学

<http://acml-shizuppi.net/>



# 認知モデルとは 一人と機械のアライメントー

- 静岡大学情報学部  
「認知科学と行動情報」  
講義資料

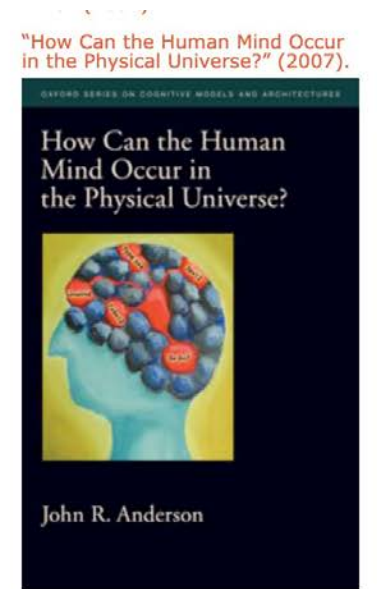
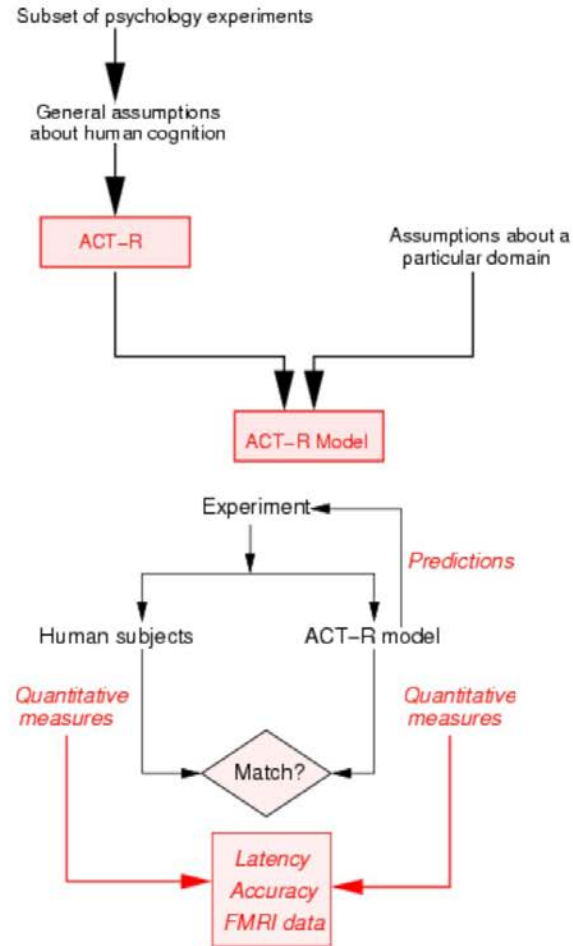


# 図の出典

**models** (aka programs) that are written in ACT-R and that, beside incorporating the ACT-R's view of cognition, add their own assumptions about the particular task. These assumptions can be tested by comparing the results of the model with the results of people doing the same tasks. By "results" we mean the traditional measures of cognitive psychology:

- time to perform the task,
- accuracy in the task, and,
- (more recently) neurological data such as those obtained from fMRI.

One important feature of ACT-R that distinguishes it from other theories in the field is that it allows researchers to collect quantitative measures that can be directly compared with the quantitative measures obtained from human participants.



Oxford University Press website for "How Can the Human Mind Occur in the Physical Universe?"

## VIDEOS



Allen Newell describes cognitive architectures as the way to answer one of the ultimate scientific questions: "How can the human mind occur in the physical universe?"

<http://act-r.psy.cmu.edu/about/>

# もう一つの汎用人工知能？

<http://act-r.psy.cmu.edu/>

Carnegie Mellon University

Contact | Sign up | Submit Publication | Mailing List |

Search

HOME

ABOUT

PEOPLE

PUBLICATIONS & MODELS

SOFTWARE

WORKSHOPS

LINKS

ACT-R \akt-ahr\ , noun;

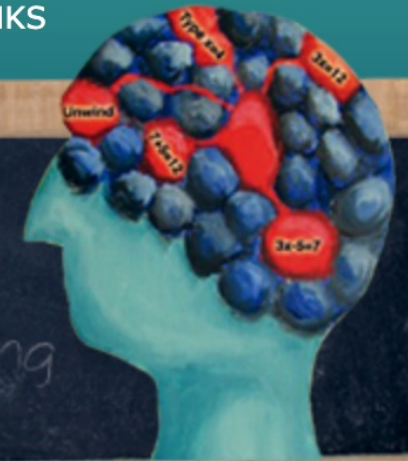
1. cognitive architecture
2. a theory for simulating and understanding human cognition

$$7+5=12$$

$$8+2=10$$

$$3x-5=7$$

cognitive modeling



## Home

ACT-R is a cognitive architecture: a theory for simulating and understanding human cognition. Researchers working on ACT-R strive to understand how people organize knowledge and produce intelligent behavior. As the research continues, ACT-R evolves ever closer into a system which can perform the full range of human cognitive tasks: capturing in great detail the way we perceive, think about, and act on the world.

### RECENTLY UPLOADED

An Integrated Computational Framework for the Neurobiology of Memory Based on the ACT-R Declarative Memory System



# 認知アーキテクチャとは？



WIKIPEDIA  
The Free Encyclopedia

Search

## Cognitive model

🌐 12 languages

Contents hide

Article [Talk](#)

[Read](#) [Edit](#) [View history](#) [Tools](#)

(Top)

[Relationship to cognitive architectures](#)

[History](#)

[Box-and-arrow models](#)

▼ [Computational models](#)

[Symbolic](#)

[Subsymbolic](#)

[Hybrid](#)

▼ [Dynamical systems](#)

[Early dynamical systems](#)

[Associative memory](#)

[Language acquisition](#)

[Cognitive development](#)

From Wikipedia, the free encyclopedia

*This article is about models of cognition's operation. For a model of external reality as modeled in the mind, see [mental model](#).*

A **cognitive model** is an approximation of one or more [cognitive processes](#) in humans or other animals for the purposes of comprehension and prediction. There are many types of cognitive models, and they can range from box-and-arrow diagrams to a set of equations to software programs that interact with the same tools that humans use to complete tasks (e.g., computer mouse and keyboard).<sup>[1]</sup><sup>[page needed]</sup> In terms of [information processing](#), **cognitive modeling** is modeling of human perception, reasoning, memory and action.<sup>[2]</sup><sup>[3]</sup>

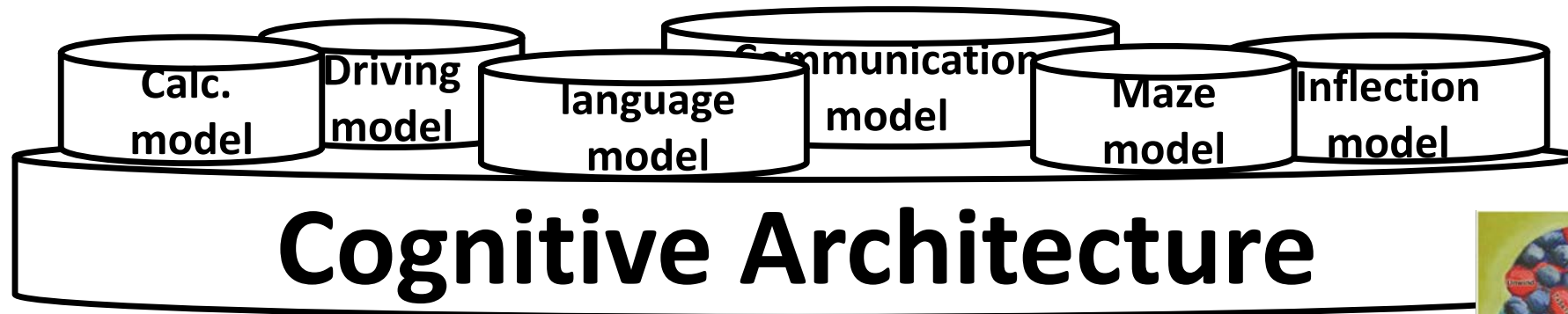
### Relationship to cognitive architectures [\[ edit \]](#)

Cognitive models can be developed within or without a [cognitive architecture](#), though the two are not always easily distinguishable. In contrast to cognitive architectures, cognitive models tend to be focused on a single cognitive phenomenon or process (e.g., list learning), how two or more processes interact (e.g., visual search bsc1780 decision making), or making behavioral predictions for a specific task or tool (e.g., how instituting a new software package will affect productivity). Cognitive architectures tend to be focused on the structural properties of the modeled system, and help constrain the development of cognitive models within the architecture.<sup>[4]</sup> Likewise, model development helps to inform limitations and shortcomings of the architecture. Some of the most popular architectures for cognitive modeling include [ACT-R](#), [Clarion](#), [LIDA](#), and [Soar](#).

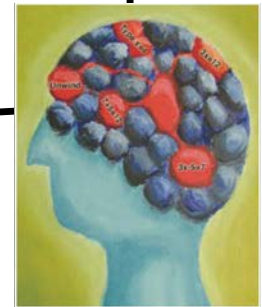
# 認知モデルと認知アーキテクチャ

- 認知モデル：人間の認知プロセスをシミュレートするプログラム
- 認知アーキテクチャ：認知モデルを作る土台・構造・制約

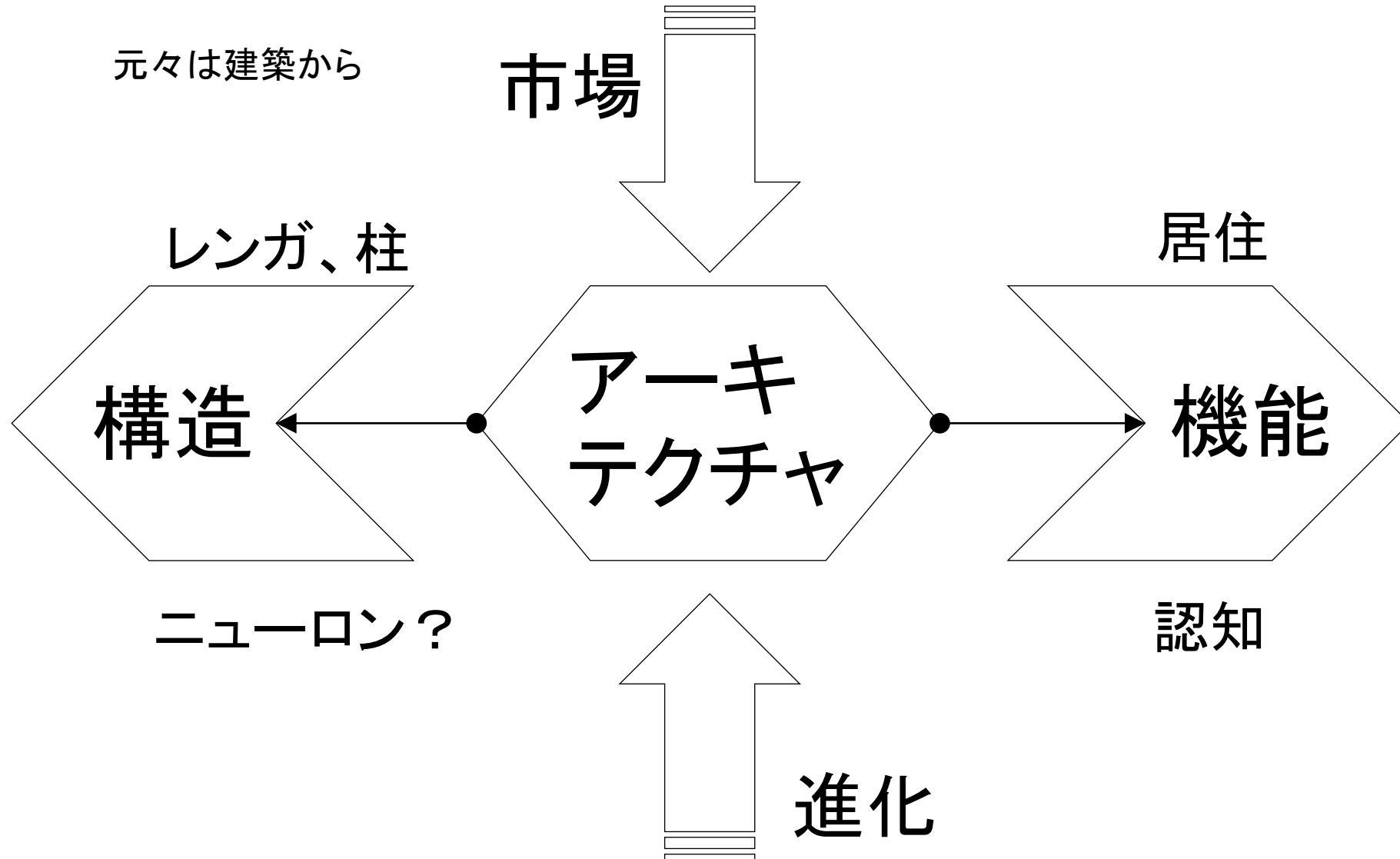
## Models for different tasks



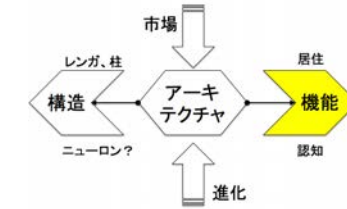
共通の認知アーキテクチャをベースに  
多様なタスクの認知モデルが作られる



# アーキテクチャ: 構造と機能の対応



# ACT-Rがカバーする認知機能



## Browse by Category

### ACT-R Theory

Architecture

### Language Processing

Analogy and Metaphor

Language Learning

Lexical and General Language Processing

Parsing

Sentence Memory

### Perception and Attention

Attention

Driving and Flying Behavior

Eye Movements

Graphical User Interfaces

Multi-Tasking

Psychophysical Judgements

Situational Awareness and Embedded Cognition

Stroop

Subitizing

Task Switching

Time Perception

Visual Search

### Problem Solving and Decision Making

Choice and Strategy Selection

Dynamic Systems

Errors

Game Playing

Insight and Scientific Discovery

Mathematical Problem Solving

Programming

Reasoning

Spatial Reasoning and Navigation

Tower of Hanoi

Use and Design of Artifacts

言語処理

知覚と注意

問題解決と  
意思決定

### Learning and Memory

Category Learning

Causal Learning

Cognitive Arithmetic

Declarative Memory

Implicit Learning

Interference

Learning by Exploration and Demonstration

List Memory

Practice and Retention

Reinforcement Learning

Representation

Skill Acquisition

Updating Memory and Prospective Memory

Working Memory

### Other

Cognitive Development

Cognitive Workload

Communication, Negotiation, and Group Decision Making

Comparative (Architectures)

Comparative (Inter-species)

Computer Generated Forces, Video Games, and Agents

fMRI

Individual Differences

Information Search

Instructional Materials

Intelligent Tutoring Systems

Motivation, Emotion, Cognitive Moderators, & Performance

Neuropsychology

Tools

Unrelated to ACT-R

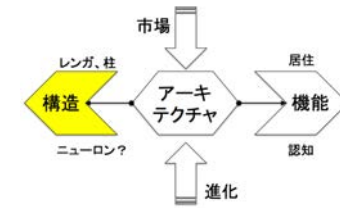
User Modeling

### Uncategorized

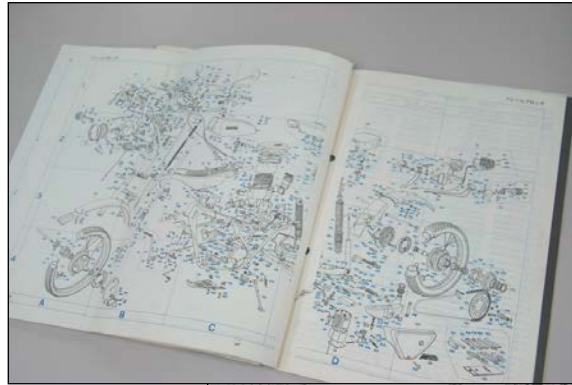
学習と記憶

ACT-Rの論文を整理したIndex

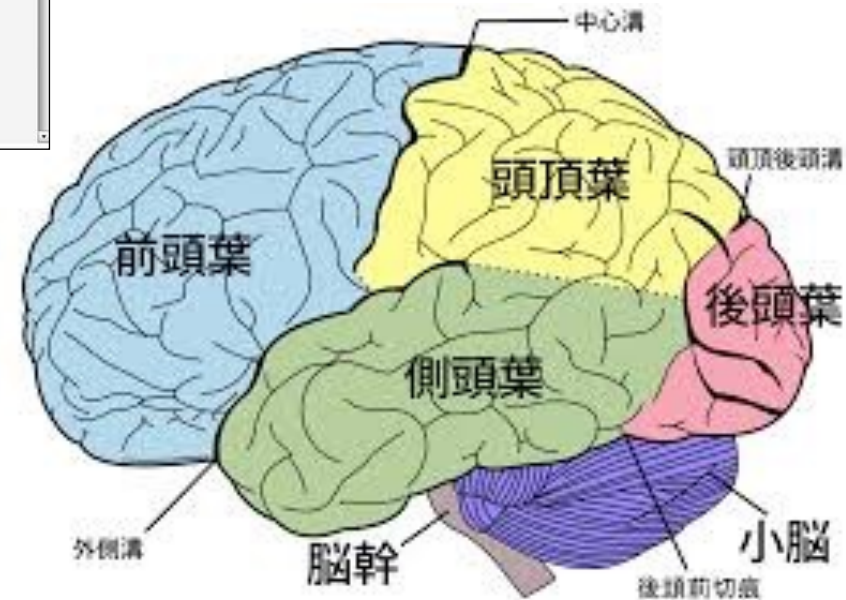
# モジュール単位で構造を実装



エンジン、フォーク、フレーム  
などのパーツ単位で掲載



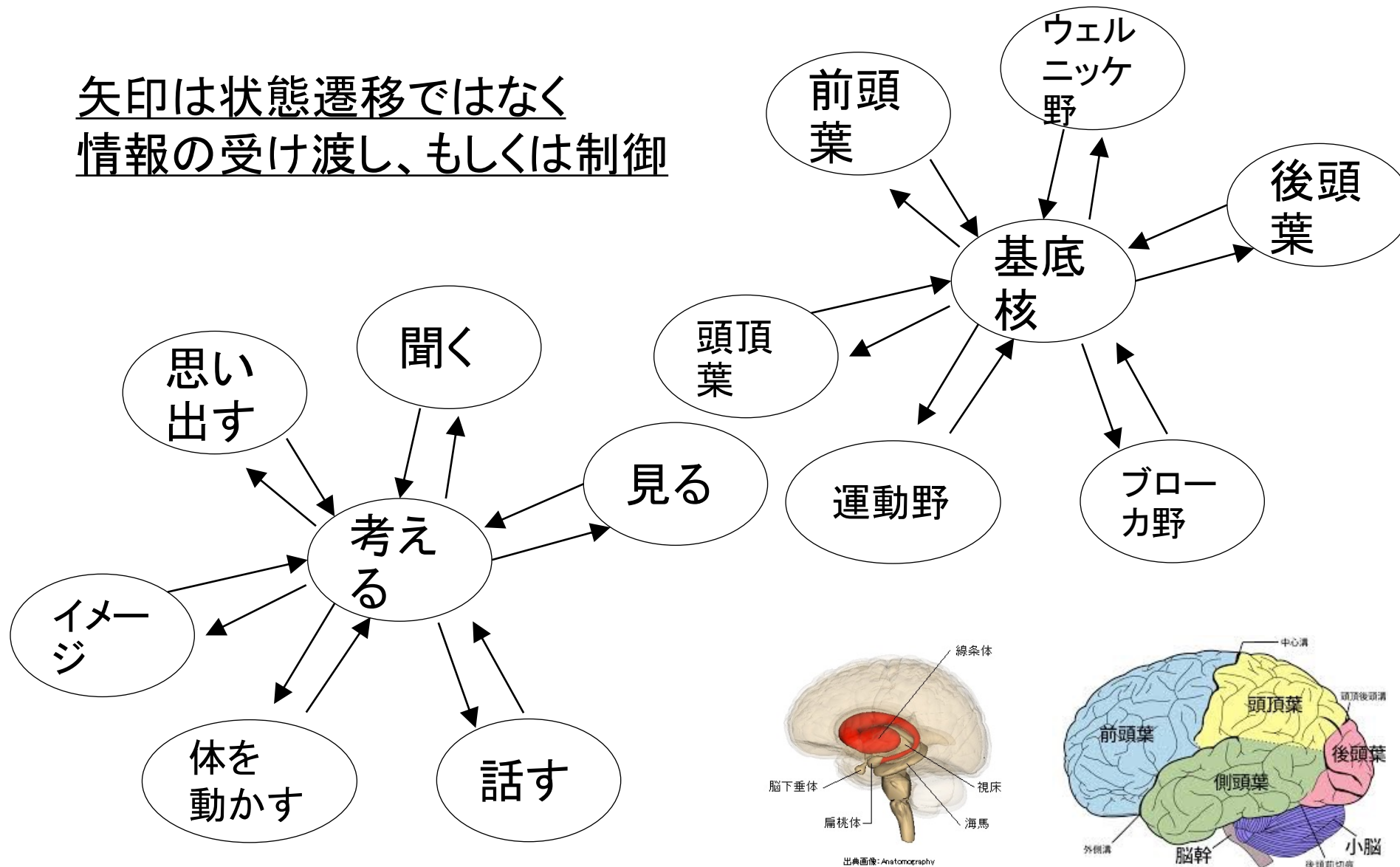
脳の部位と機能が対応  
例: 側頭葉の一部を欠損  
すると言葉が出なくなる



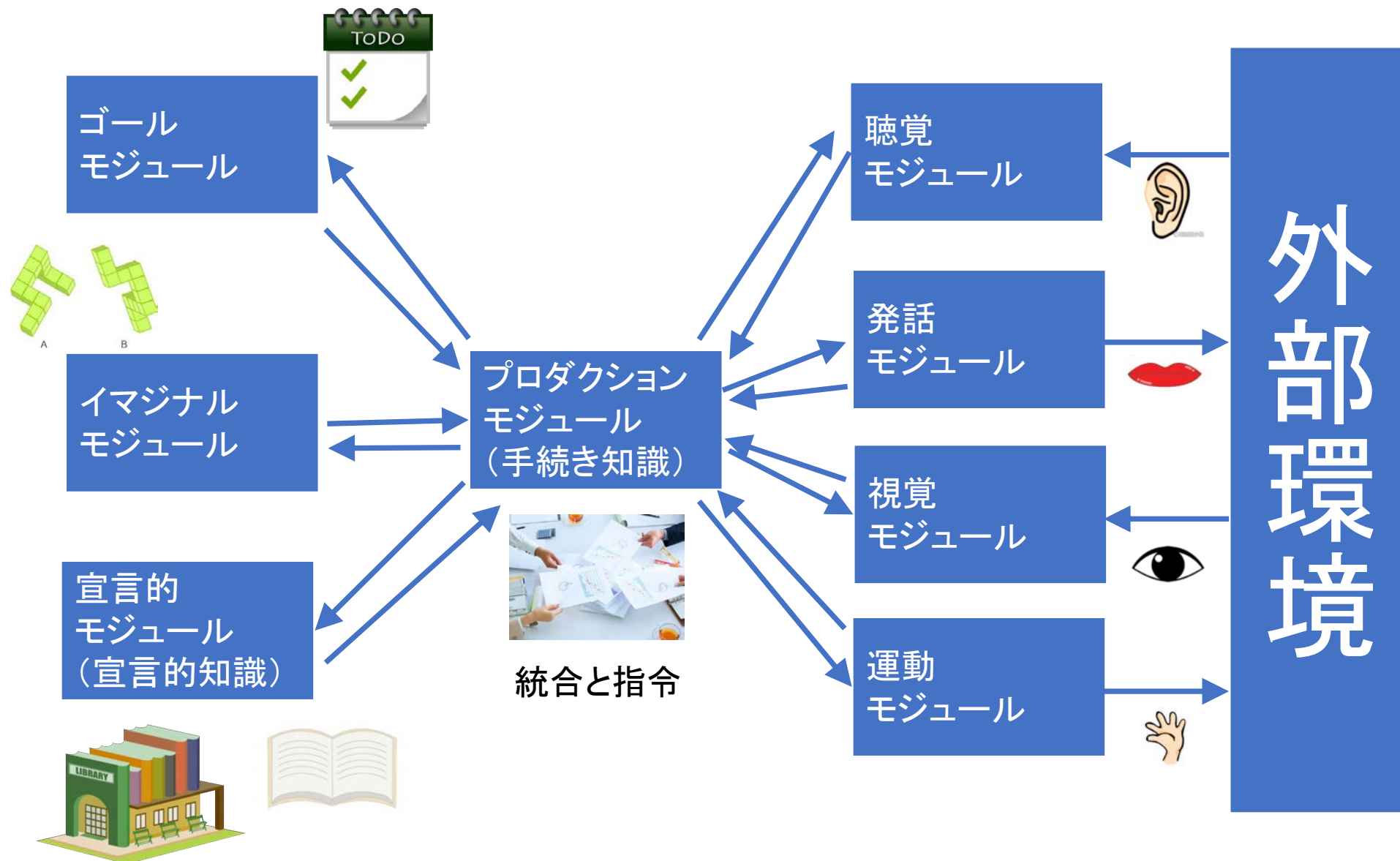


# 脳の部位と認知のモジュールの対応

矢印は状態遷移ではなく  
情報の受け渡し、もしくは制御



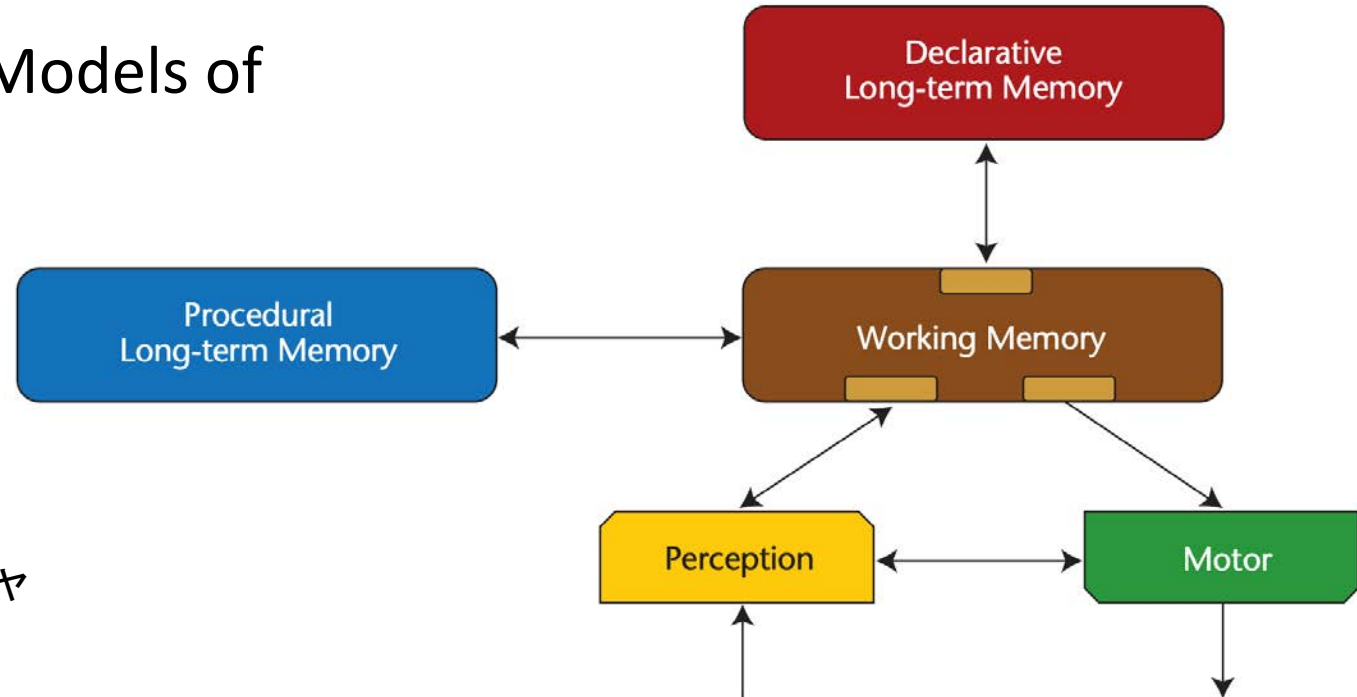
# ACT-Rのモジュール構造



# 認知アーキテクチャ界隈の最近の動向

## 共通認知モデル (Common Models of Cognition) による統合

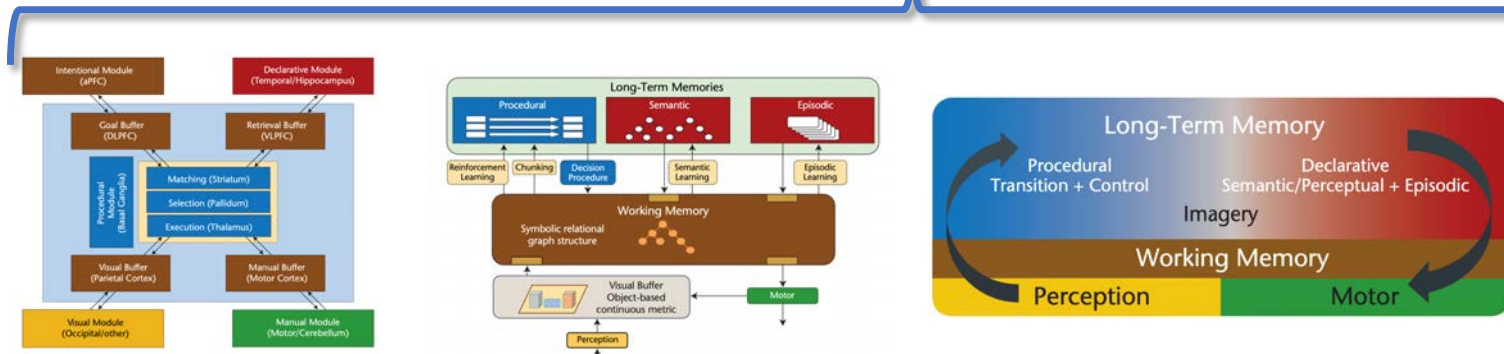
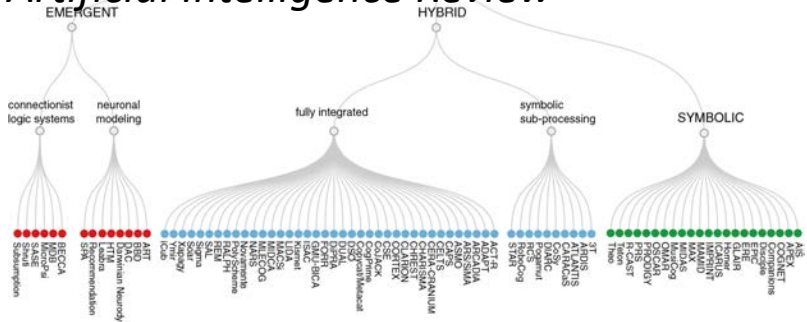
Laird, Lebiere & Rosenbloom (2017).  
A Standard Model of the Mind,  
*AI Magazine*



共通構造の抽出

過去に構築された様々な認知アーキテクチャ

Kotseruba & Tsorsos (2020).  
40 years of cognitive architecture,  
*Artificial Intelligence Review*



# 認知アーキテクチャ界隈の さらに最近の動向



About Us ▾ Research ▾ News/Publications ▾ Work With CIC ▾

## FSS2023: Integration of Cognitive Architectures and Generative Models

### Description of the symposium:

The purpose of this symposium is to explore possible integrations of cognitive architectures and generative models, two very different approaches for developing general embodied intelligent agents with possibly complementary strengths and weaknesses. An important goal is to understand the strengths, weaknesses, tradeoffs, and challenges among alternative approaches to integration, as well as potential directions for future research. We hope to include a broad range of researchers whose interests span cognitive architectures and generative models.

### Organizing Committee:

- John Laird, Center for Integrated Cognition/IQMRI
- Christian Lebiere, Carnegie Mellon University
- David Reitter, Google Deepmind
- Paul S. Rosenbloom, University of Southern California
- Andrea Stocco, University of Washington

### Meeting Schedule:

- The FSS is Oct 25-27 at the Westin Arlington Gateway, Arlington, Virginia, USA
- See sections below for details on the schedule of talks, posters, and discussions
- [Papers and printable schedule](#) (contact moderators for password)

# 認知アーキテク チャと生成AIの 統合

<https://acml-shizuppi.net/aaai-fss2023/>

Menu
<a href="#">Activity</a>
<a href="#">研究室へのアクセス</a>
<a href="#">サイトマップ</a>

## AAAI秋のシンポジウム (FSS2023: Integration of Cognitive Architectures and Generative Models)

アメリカのワシントンDC (正確にはその隣のアーリントン) にて開催されたAAAI (アメリカ人工知能学会) のシンポジウムに森田が参加し、発表し、議論してきました。テーマは生成AIと認知アーキテクチャの統合についてです。

### FSS2023: Integration of Cognitive Architectures and Generative Models

生成モデルは現在大変な社会現象になっています。この流れのなかにもどのように認知モデル・認知アーキテクチャを組み入れられるか (プレゼンスを発揮できるか) という議論がされました。テーマの重要性からか、認知モデル・認知アーキテクチャを古くから研究されている大御所の先生方が多くいらっしゃっていました。またシンポジウムは全体を通してディスカッションの時間が多く、ポトムアップにコミュニティを形成していこうとする流れができていました。認知アーキテクチャを使うことで、生成AIに記号的に複雑な推論を付加できること、生成AIにおける身体性の不在をカバーでき、ロボット執事のような心の理論をもった人間の意図を察するエージェントを構築できることなどが議論されました。一方で認知アーキテクチャの分野におけるベンチマークの不在、資金調達などの問題などについては乗り越えるべき弱点として議論されました。

ACMLからの発表は以下のものです (リンクは先行公開されているarxivのもの)。

[Cognitive Architecture Toward Common Ground Sharing Among Humans and Generative AIs: Trial Modeling on Model-Model Interaction in Tangram Naming Task](#)  
Junya Morita, Tatsuya Yui, Takeru Amaya, Ryuichiro Higashinaka and Yugo Takeuchi

生成AIと認知アーキテクチャの統合により、人間と人工知能のコモングラウンド (コミュニケーションを行う際の前提) を形成するというもの。今回のシンポジウムのなかでは異色の発表で、会場からの反応はそれなりによかったのではないかと思います。



<https://arxiv.org/abs/2311.05851>

Cognitive Architecture Toward Common Ground Sharing Among Humans and Generative AIs:

# **Trial Modeling on Model-Model Interaction in Tangram Naming Task**

Junya Morita, Tatsuya Yui, Takeru Amaya,  
Ryuichiro Higashinaka, Yugo Takeuchi

Shizuoka University, Nippon Telegraph and Telephone Corporation

# My opinions on the theme

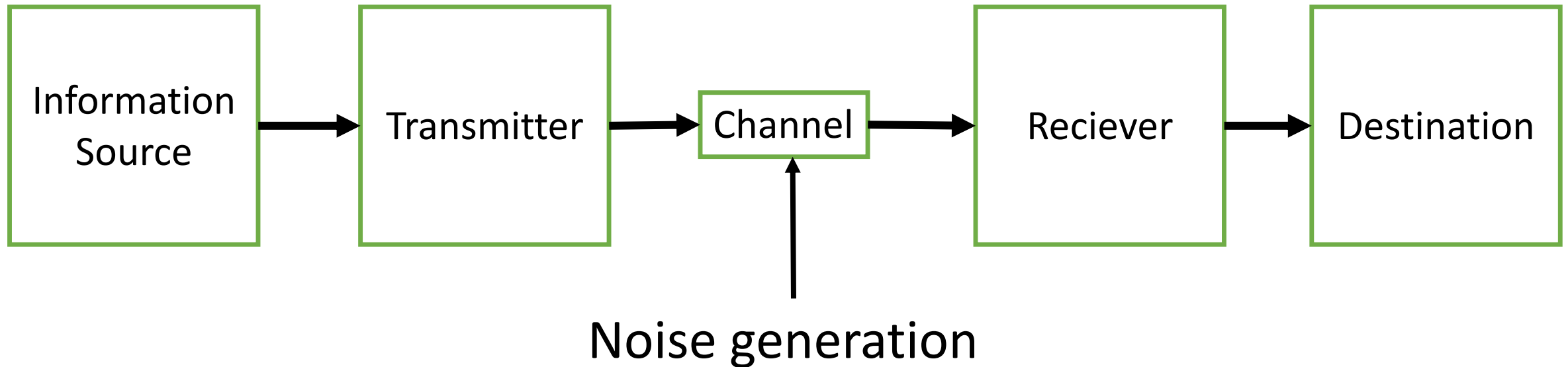
**1. Controlling** generative AIs by cognitive architectures:  
Toward transparent and trustworthy AIs

**2. Constructing** cognitive models by generative AIs:  
Toward richer and friendly cognitive modeling

**3. Concretizing** cognitive models by generative AIs:  
Toward a common grounds between humans and AIs

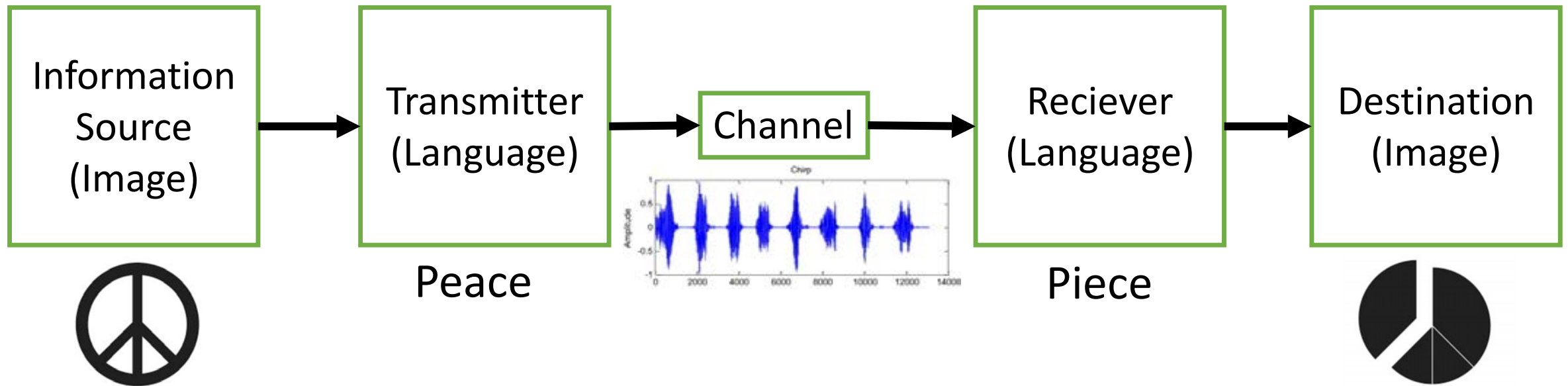
**Our research**

# Communication (Shannon model)



Miscommunication is explained by noise generation

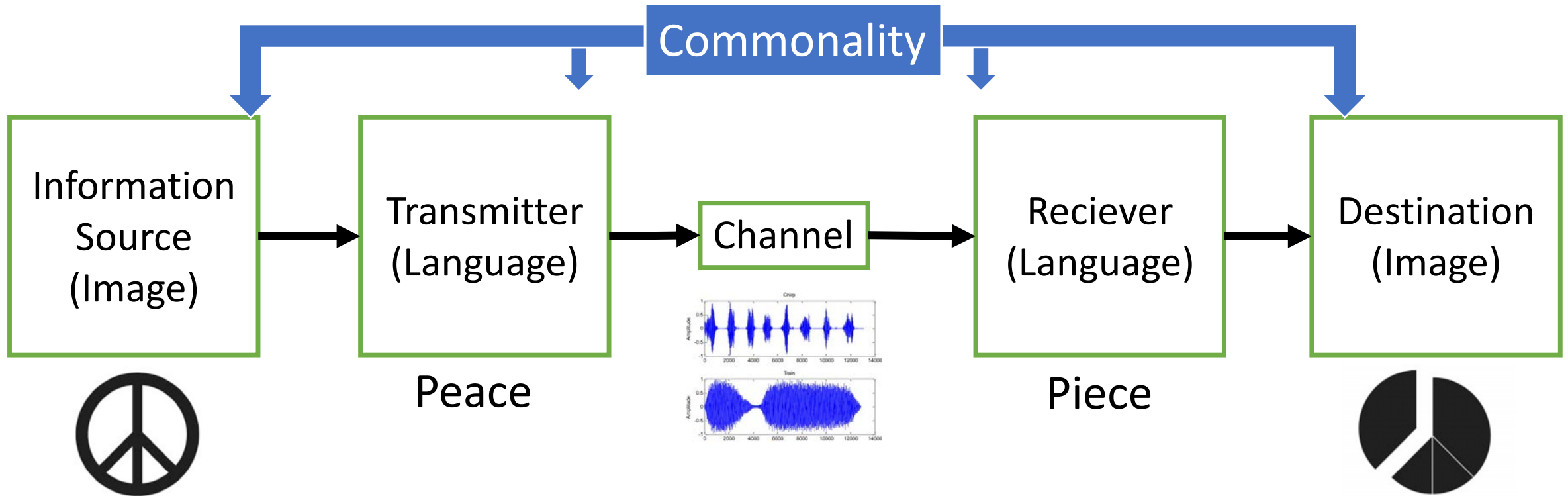
# Communication (Cognitive view)



Miscommunication is explained by different interpretations on ambiguous language expressions



# Communication (Cognitive view)



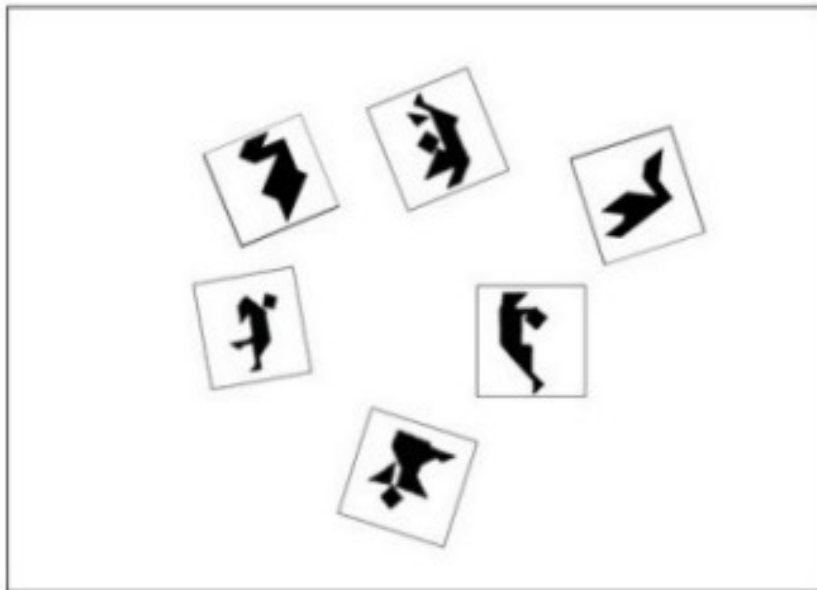
Common ground is needed  
Commonality of attention and imagination mechanisms

# Tangram naming task (Clark, 1986)

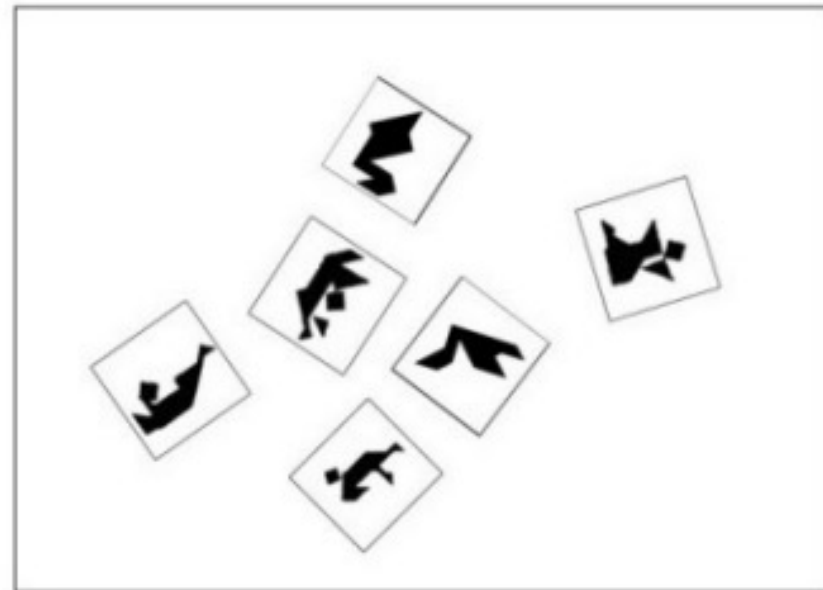
Try to share names for each tangram between participants

- They cannot see the other's arrangement
- Only verbal communication can be used

Participant A



Participant B



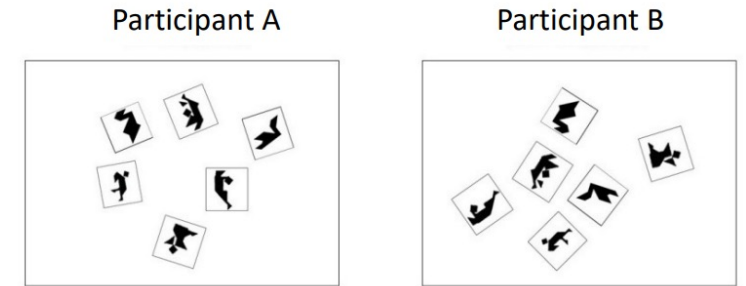
# Typical communication

Two types of expressions

- **Holistic utterance**  
Metaphorical expression
- **Analytic utterance**  
Describing arrangement of geometric shapes

## Holistic utterance is dominant

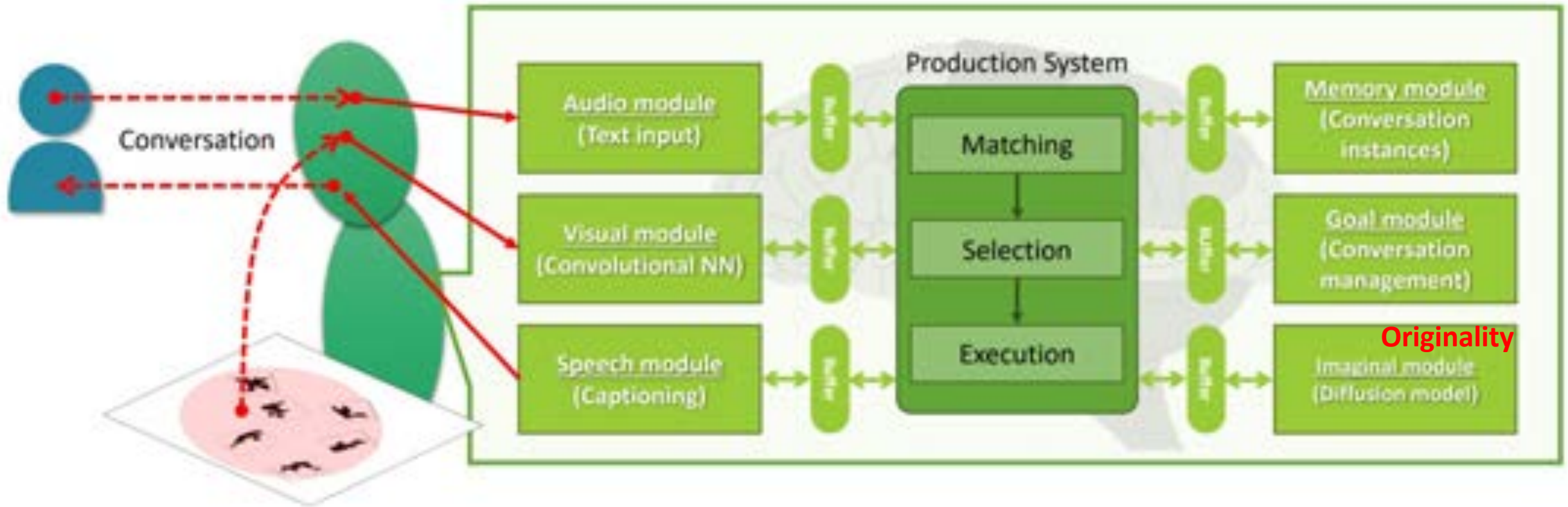
- People have a rich imagination
- *Imagination is the foundation of our society* (by Harari's Sapiens)



- A And, you know, kicking, like, kicking a ball or something.
- B I can't see.
- A There is some kind of ball behind the head, and the feet are shaped like the guy is kicking a ball.
- B You know, the one with the separate squares?
- A Aha, yes, yes, yes.
- B Like Hokkaido?
- A Hokkaido
- B Like a map of Japan
- A Oh, no, no, no, no, not that.
- A It's kind of a 90-degree kink.
- B Ah
- A Foot-like, ball-kicking kind of thing.
- B Yes, yes, yes, like a little cross-legged thing?
- A Oh, yes, and that one with the little square behind it.
- B Yeah, I kind of get it.

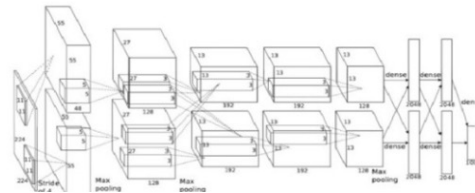
# Assumed architecture

- Modules made by NN are controlled by the central rule engine
- *Imaginative process* is realized by a generative AI

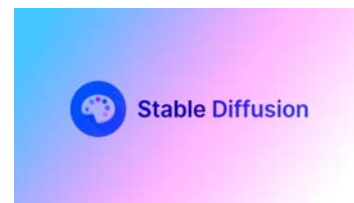


**Too expensive for a single cognitive process**

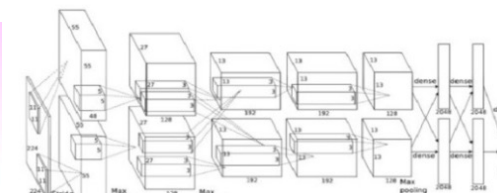
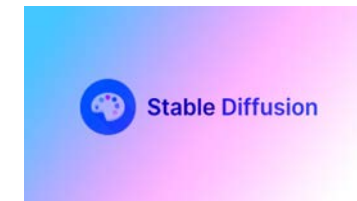
# Pipeline behind the holistic utterances



Originally trained CNN



GPT-2

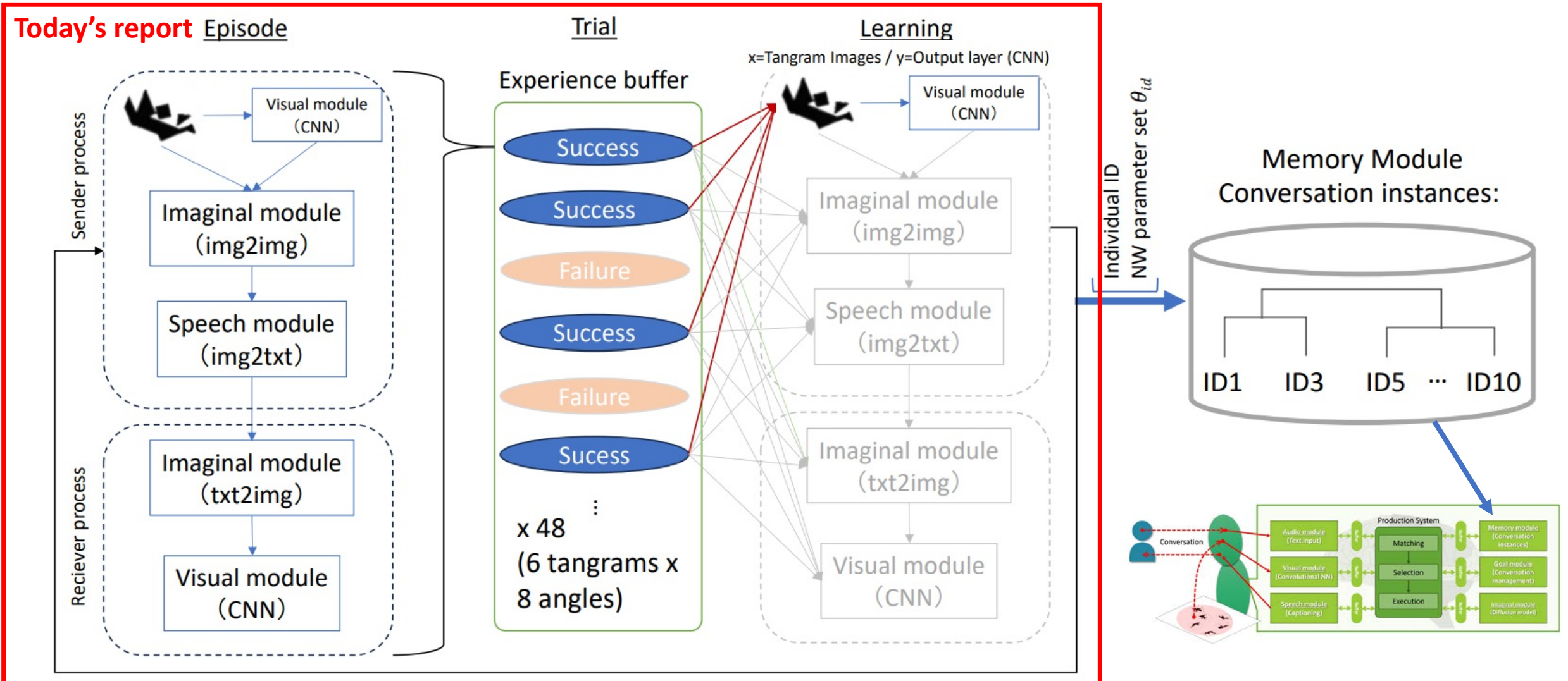


Originally trained CNN



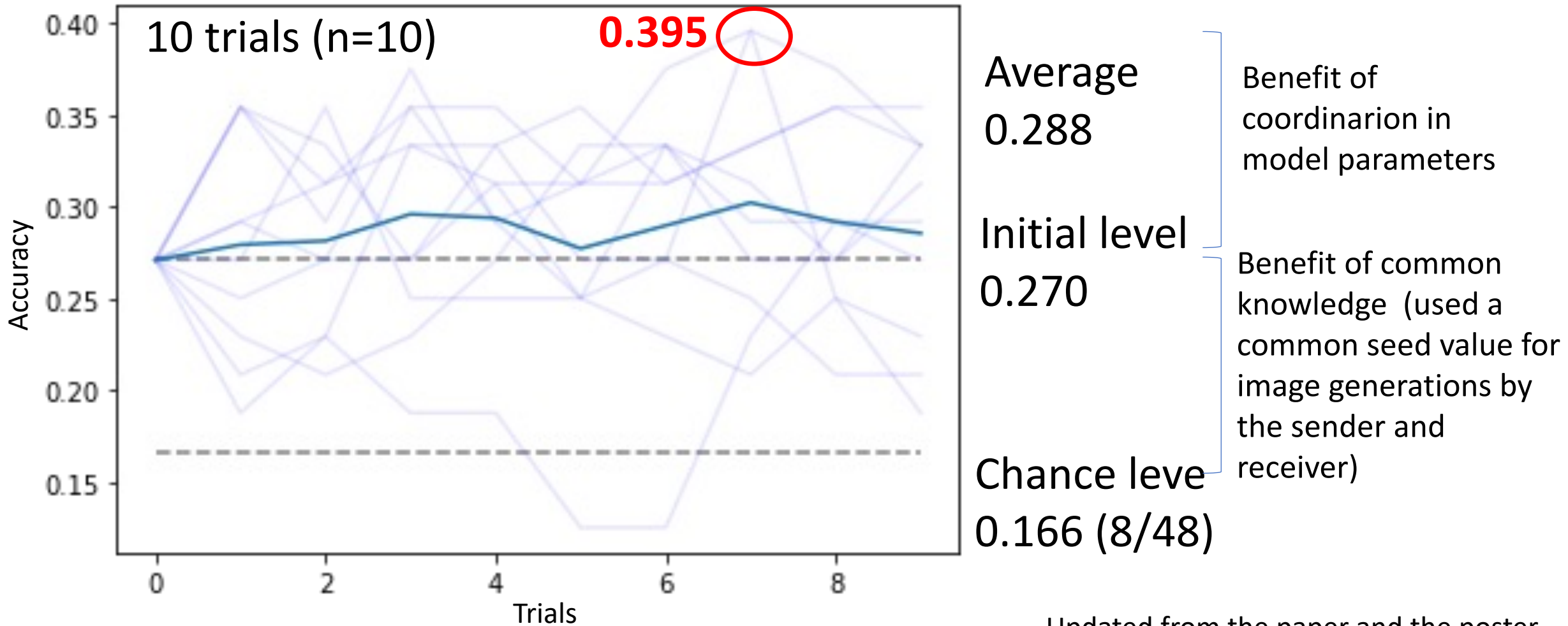
# Learning procedure

Pick up success trials for backprop on component networks



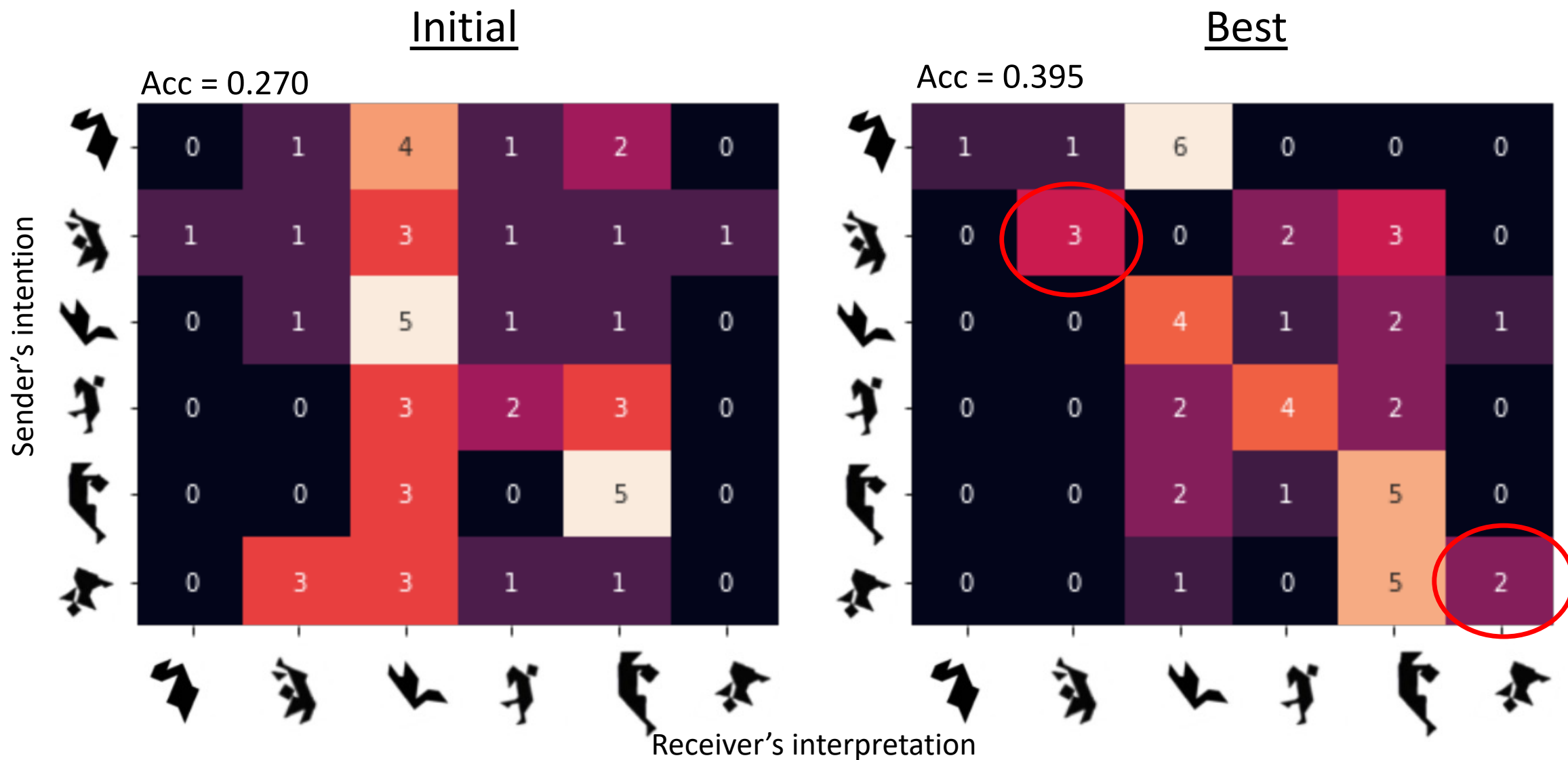
# Tentative results

- Large improvement from the chance
- (Slightly but) Statistically significant improvement by the learning ( $p < .01$ )



Updated from the paper and the poster

# Initial and best confusion matrix





# Implication and Significance 1

**Generative AI concretizes** cognitive process to be easily interpretable by humans →

Leads a common grounding between humans and AIs

## Implication and Significance 2

**Cognitive Architecture provides structure** of organizing implicit cognitive process generated by genAIs →  
Leads another common grounding between humans and AIs

# 人とAIの共通基盤形成に向けて

## 認知アーキテクチャ

- 人間に対する科学的な理解に基づく汎用人工知能
- 文明的な認知

## 生成AI

- 人間からのフィードバックにより学習する汎用人工知能
- 自然な認知

両者の補完的な関係により  
理解の平面（グラウンド）が完成