

WBA開発のロードマップと自動化

山川 宏

NPO法人 全脳アーキテクチャ・イニシアティブ

生成AI時代における 全脳アーキテクチャ・アプローチの 開発とその意義

第1部：加速する全脳アーキテクチャの開発

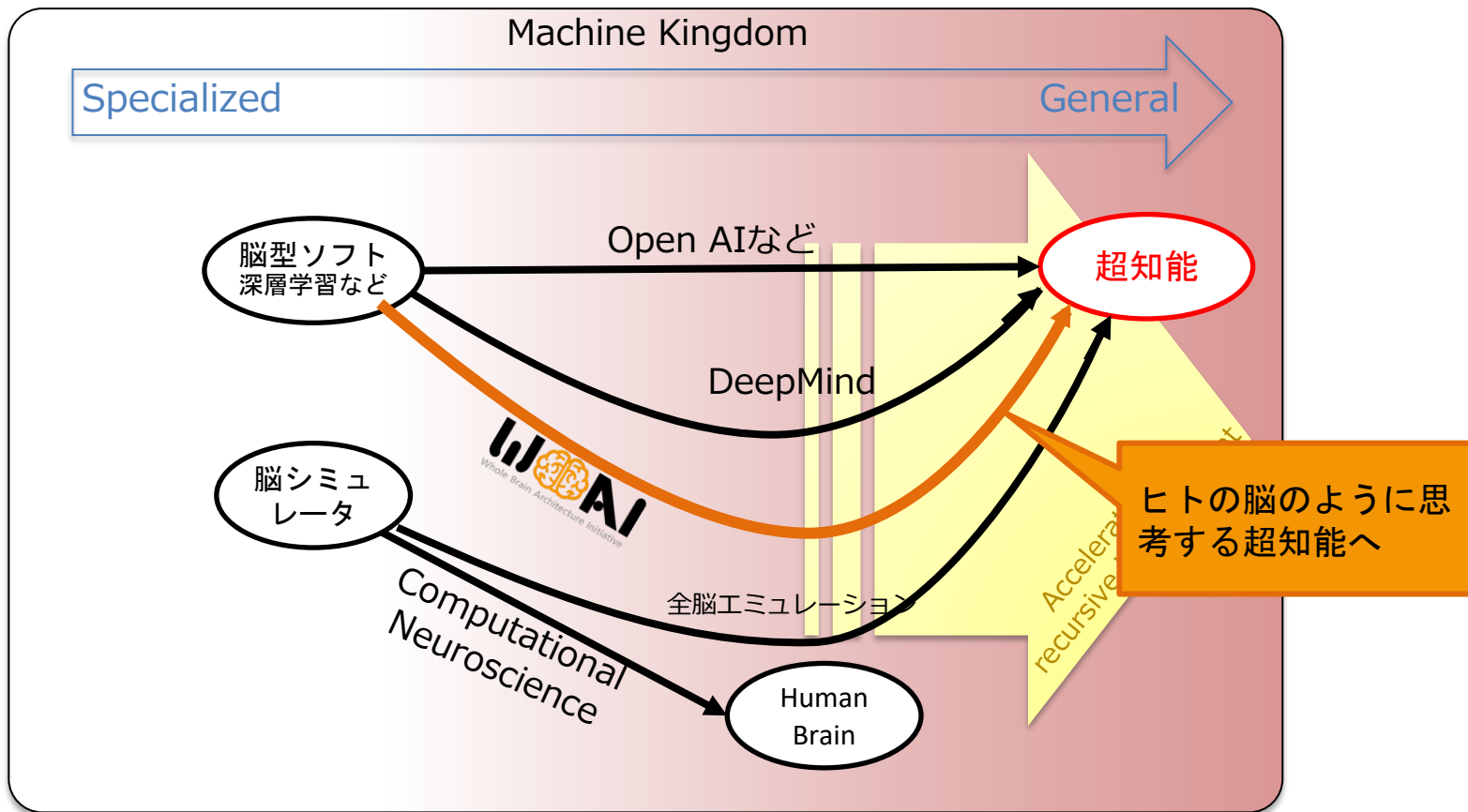
生成AIは、全脳アーキテクチャ・アプローチからのAGI開発に何をもたらしたか。

- 非脳型からのAGIが実現される可能性が高まった
(但し、今後順調に完成に至らなければ、脳型AGIが先行して完成する可能性はある)
- 全脳アーキテクチャからのAGI開発を大きく加速している

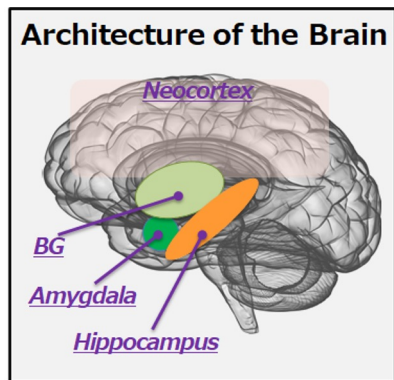
第3部：よき超知能時代に向かうためにヒト脳型AGIが果たすべき役割

ほぼ超知能の速度で動作しつつ人間と親和性の高い脳型AGIは人類と超知能の架け橋になるかもしれない。

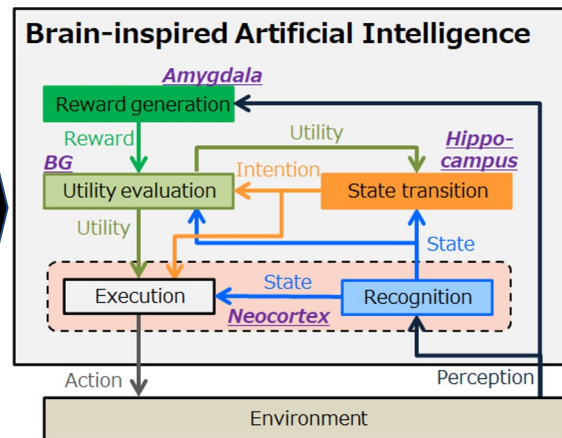
高度人工知能への道筋



脳全体のアーキテクチャに学び人間のようなAGIを創る(工学)



- ① 脳の各器官を機械学習モジュールとして開発
② それらを統合した認知アーキテクチャの構築



本アプローチの利点

共同開発の基盤として脳の構造を利用して開発を加速

- 人間に関わる諸分野の科学知見を活用できる
- 開発の発散を防げる合意しうる認知アーキテクチャ
- 脳にガイドされた有機的な技術の統合

人と親和性の高いAGIを作れる

- 情緒の推定等を要する対話などへの応用
- 精神疾患などのモデル化など医療応用
- マインド・アップローディングの器

脳に学んで汎用人工知能をつくる意義

■ 汎用人工知能(Artificial General Intelligence: AGI)とは

- 多様な問題領域において多角的な問題解決能力を自ら獲得し設計時の想定を超えた問題を解決できる人工知能

■ AGI開発の課題

- 様々な計算機構の組み合わせによって生じる設計空間の広大さ
 - ・ タスク毎に計算機構を組み合わせている現状のAIシステム

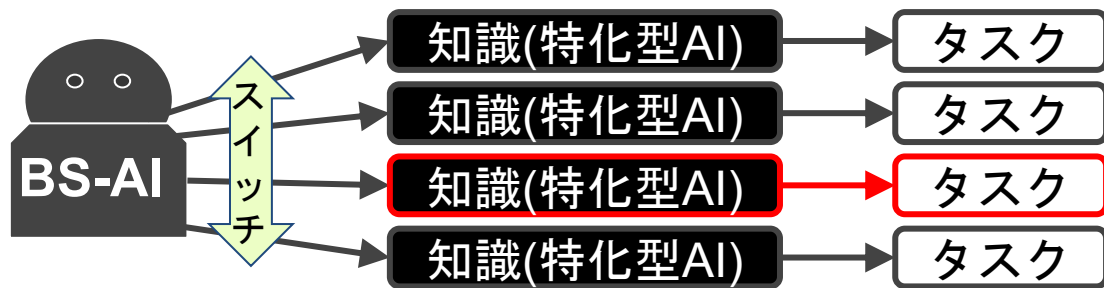
■ WBAアプローチによる解決

- 脳のアーキテクチャに学んで、再利用性が高くなるように計算機構を配置すれば、一つのシステムで多様なタスクを実行できるようになる

重点目標は、AGIに固有な技術X

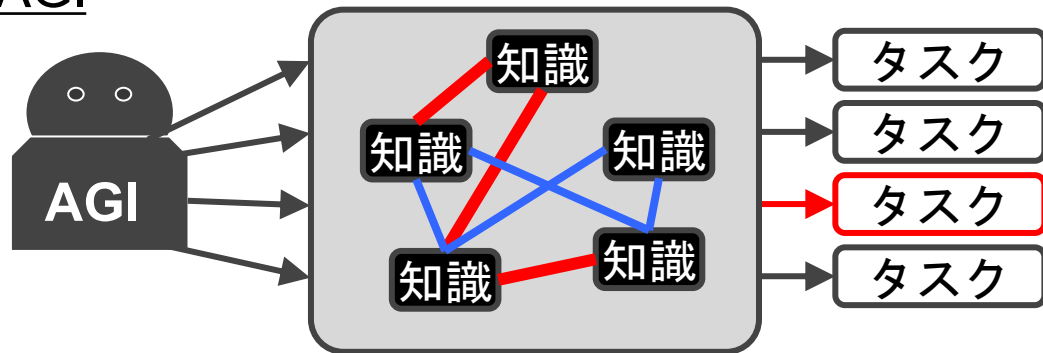
第3回WBAシンポジウム
(2018年5月8日) より

Big switch文型AI



$$\text{技術X} = \text{AGI} - \sum (\text{特化型AI})_i$$

AGI



データが少ない状況で有力な仮説を生成する能力



技術X: 既存の知識を柔軟に組合せて推論する機構

汎用人工知能に固有な必須技術

技術X：既存の知識を柔軟に
組合せて推論する機構

①知識の獲得：再利用性の高い知識の獲得 (機械学習：ML)

- 知識の写像/適用範囲拡大
- 分散表現知識の分解 (Disentangle)
- 分散表現と記号の対応付け
- 分散表現の標準化

後述する**BRA**駆動開発は、
脳に固有な②の部分を
解決するために発展

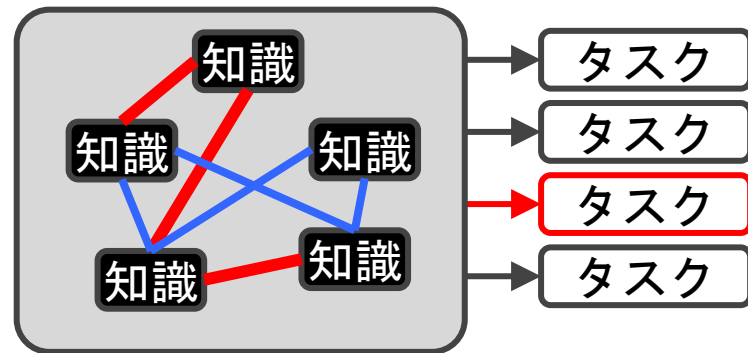
知識の利用：
多角的に知識を組
合わせる
アーキテクチャ

②アーキテクチャ開発：

脳を手本として、まずは再利用性の
高い部品の組合せとして、脳の
全ての部品を繋ぎ合わせる。
→ **全部**アーキテクチャ作製

③モジュールの組合せを操作する
仕組み (理論)

脳の情報動態を参考としつつ、
タスクの実行に有効な知識の組
合せをメタレベルで操作する技
術を確立する

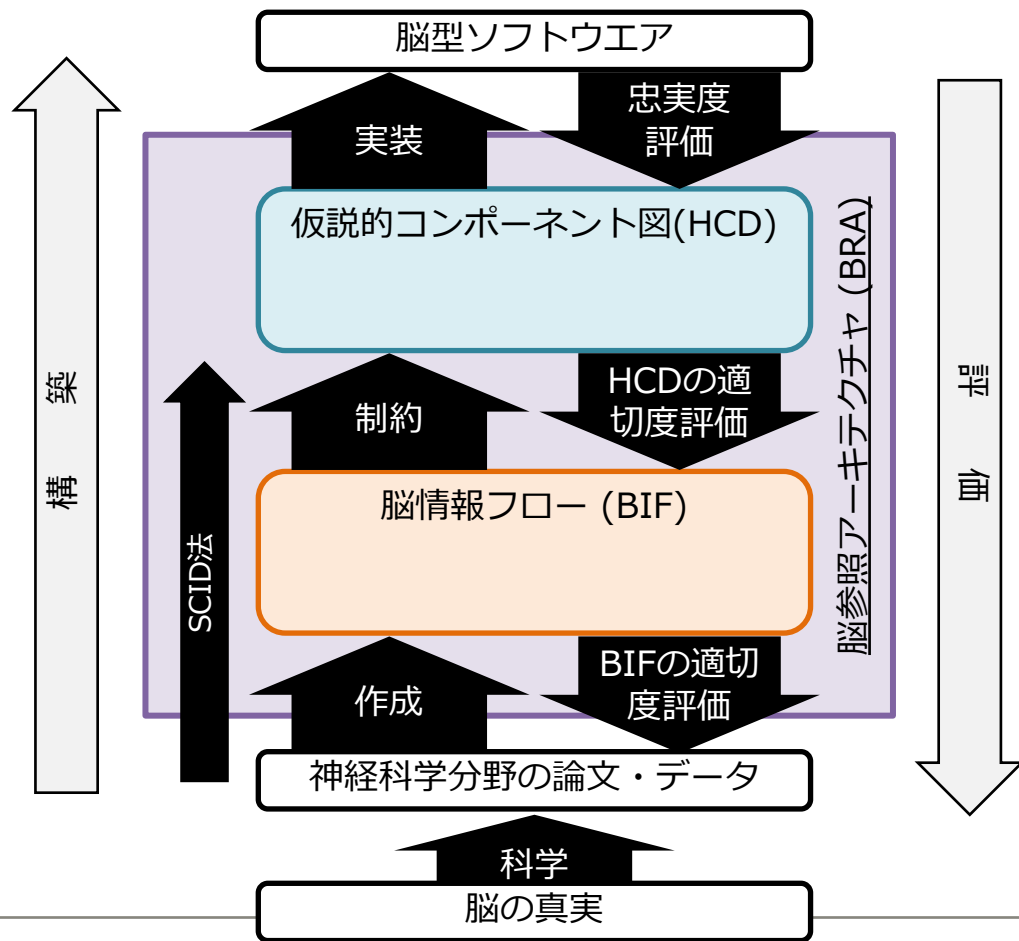


タスクに応じた知識の組合せをみつける

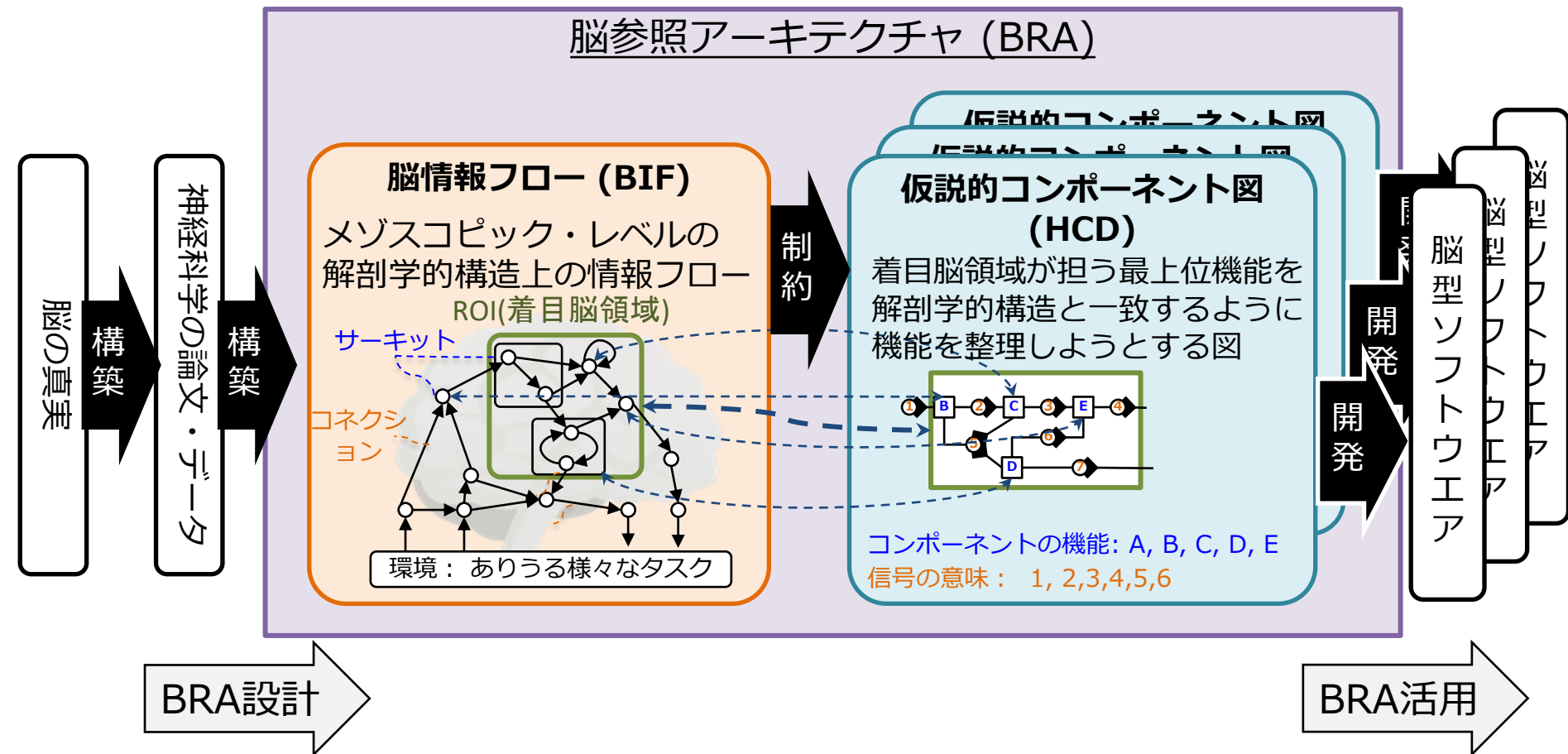
脳参照アーキテクチャ(BRA)駆動開発

脳神経回路のリバーズエンジニアリングにより、
脳型AGIの設計データとなる**脳参照アーキテクチャ(BRA)**を作成する。

そのBRAに基づいて脳型ソフトウェアを構築。



脳参照アーキテクチャ(BRA)を基盤とするBRA駆動開発



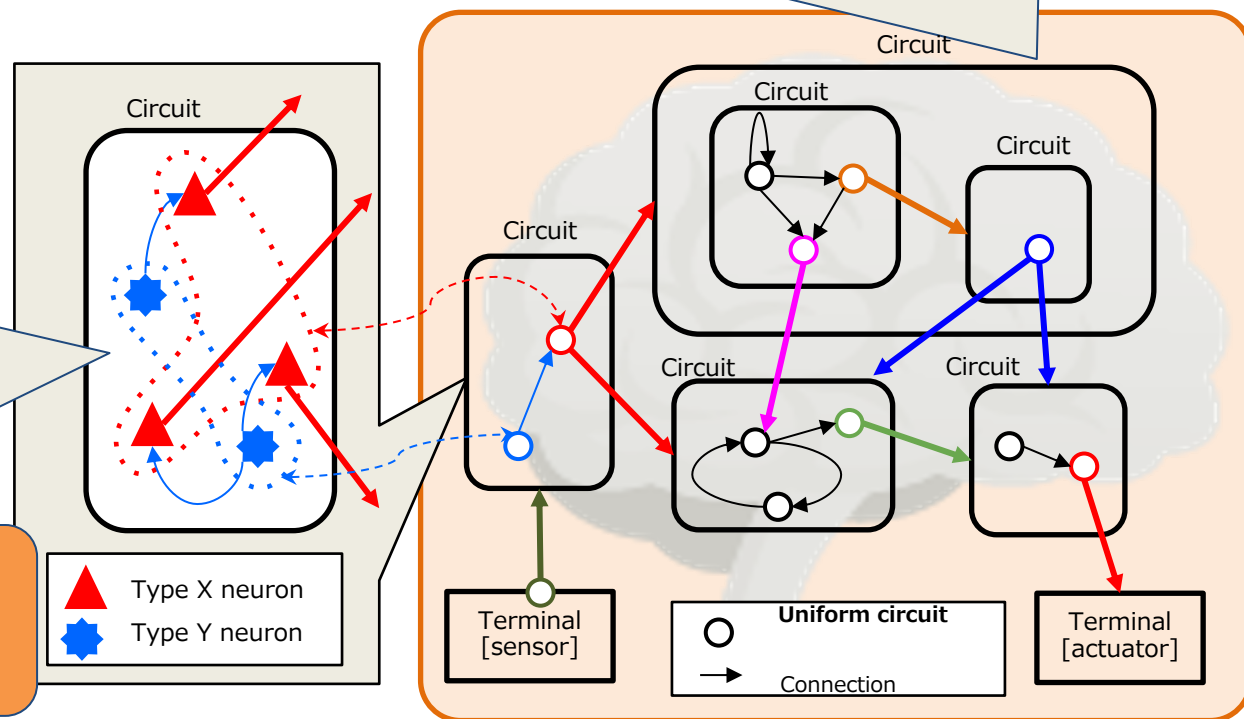
Uniform Circuits : 解剖学的構造の最小記述単位を標準化

Uniform Circuit とは

BRA駆動開発で、脳構造を記述する最小単位：
特定の脳領域内の特定の細胞タイプで構成されるニューロンのグループ

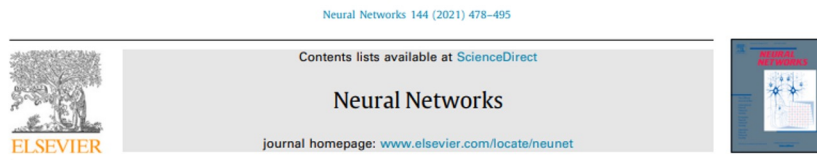
認知行動のモデル化に必須であろうレベルより詳細は扱わない
→AGI開発の効率化のため

ソフトウェア・コンポーネントとのインターフェイスとなる単純な引数（ベクターなど）へのマッピング可能



BRA駆動開発の確立(2020年ごろ)

山川宏、第7章 全脳アーキテクチャー機能を理解しながら脳型AIを設計・開発するー、認知科学講座4 心をとらえるフレームワークの展開, 2022年09月16日発売



2021 Special Issue on AI and Brain Science: AI-powered Brain Science

The whole brain architecture approach: Accelerating the development of artificial general intelligence by referring to the brain

Hiroshi Yamakawa^{a,b,c,*}

^a The Whole Brain Architecture Initiative, Nishikoiwa 2-19-21, Edogawa-ku, Tokyo 133-0057, Japan

^b The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan

^c RIKEN, 6-2-3, Furuedai, Suita, Osaka 565-0874, Japan

ARTICLE INFO

Article history:
Available online 14 September 2021

Keywords:
Brain reference architecture
Structure-constrained interface
decomposition method
Brain information flow
Hypothetical component diagram
Brain-inspired artificial general intelligence
Whole-brain architecture

ABSTRACT

The vastness of the design space that is created by the combination of numerous computational mechanisms, including machine learning, is an obstacle to creating artificial general intelligence (AGI). Brain-inspired AGI development; that is, the reduction of the design space to resemble a biological brain more closely, is a promising approach for solving this problem. However, it is difficult for an individual to design a software program that corresponds to the entire brain as the neuroscientific data that are required to understand the architecture of the brain are extensive and complicated. The whole-brain architecture approach divides the brain-inspired AGI development process into the task of designing the brain reference architecture (BRA), which provides the flow of information and a diagram of the corresponding components, and the task of developing each component using the BRA. This is known as BRA-driven development. Another difficulty lies in the extraction of the operating principles that are necessary for reproducing the cognitive-behavioral function of the brain from neuroscience data. Therefore, this study proposes structure-constrained interface decomposition (SCID), which is a hypothesis-building method for creating a hypothetical component diagram that is consistent with neuroscientific findings. The application of this approach has been initiated for constructing various regions of the brain. In the future, we will examine methods for evaluating the biological plausibility of brain-inspired software. This evaluation will also be used to prioritize different computational mechanisms, which should be integrated and associated with the same regions of the brain.

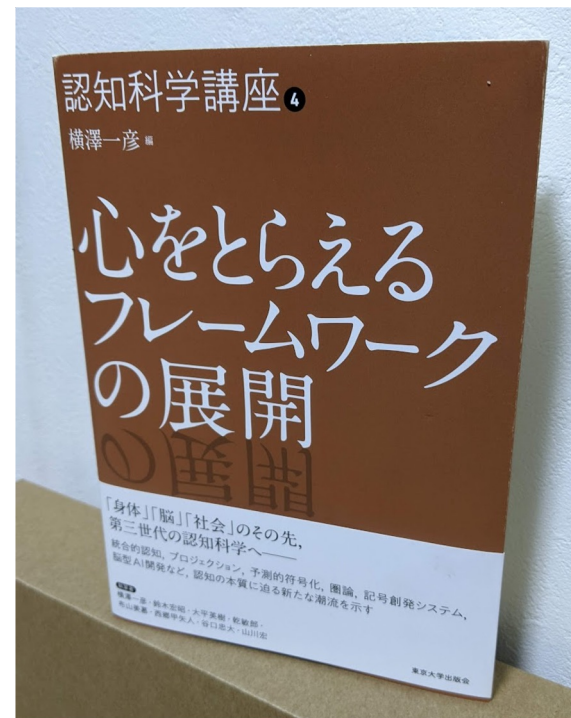
© 2021 The Author(s). Published by Elsevier Ltd. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Artificial general intelligence (AGI), the development of which has been a major goal in advanced artificial intelligence (AI) research in recent years, involves demonstrating the extensive general intelligence that is possessed by humans within a computational system (Adams et al., 2012; Goertzel, 2014). An essential ability of AGI could be solving various problems, including those on unknown issues, by flexibly combining knowledge that is gained from experience. However, methods for developing AGI remain unclear. Many AI researchers believe that the development of deep learning (LeCun, Bengio, & Hinton, 2015) serves as

However, no single theory exists on which the entire sphere of intelligence can be built. Thus far, the development of AGI has progressed by the repetitive tuning of various limited issues. However, such an approach makes it difficult to design AGI with flexible problem-solving abilities that would enable unknown issues to be solved. The construction of an AGI that possesses the full extent of human abilities would require an extremely large design space owing to the combination of computational mechanisms. Although this design space could also be explored mechanically (Clune, 2019), at present, it is difficult to secure the required computational complexity.

Brain intelligence is associated with a certain degree of ver-



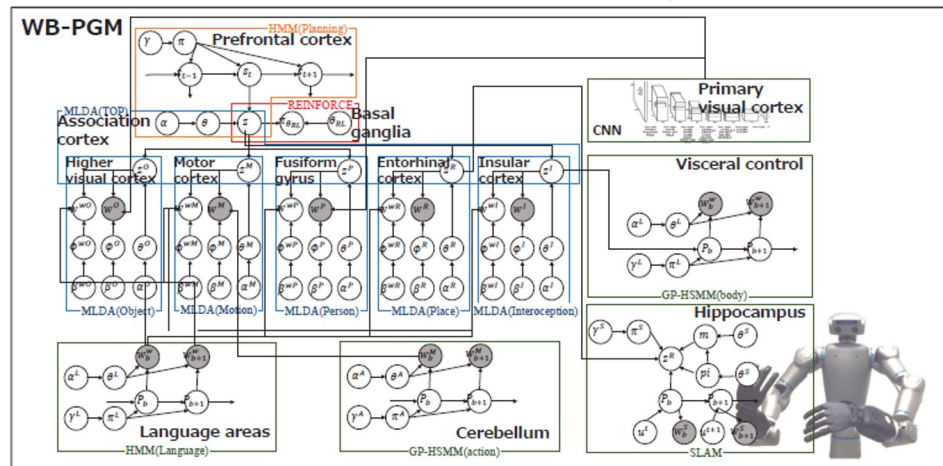
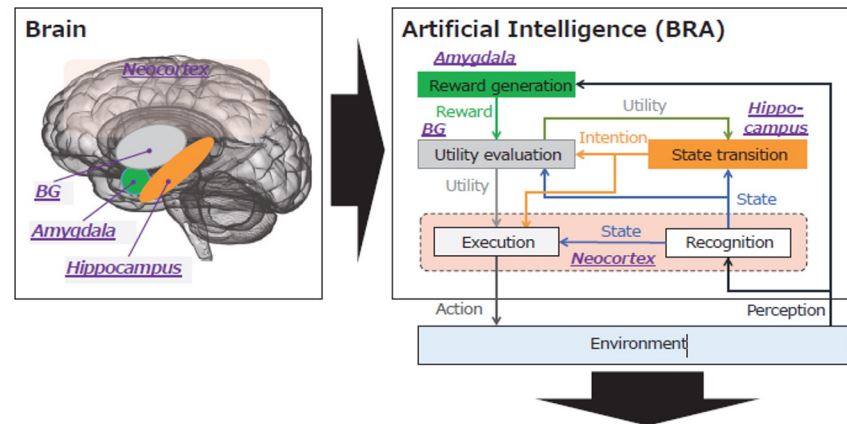
実装にむけた連携：全脳確率生成モデル(WB-PGM:)

谷口 忠大、山川 宏、長井 隆行、
銅谷 賢治、坂上 雅道、鈴木 雅大、
中村 友昭、**谷口 彰**

WB-PGMは、人間の全脳の構造
に学ぶとともに、予測学習を基
礎に据えた確率的生成モデルに
より認知アーキテクチャを構築
しようというアプローチ

背景：マルチモーダルな情報を統合し、
複雑な身体を統御し、環境に適応で
きる人間の知能、および発達的なロボ
ットの構成をその延長線上で捉えよう
とすると、人の認知アーキテクチャと
しての構造を検討する必要がある。

Whole-Brain Probabilistic Generative Model

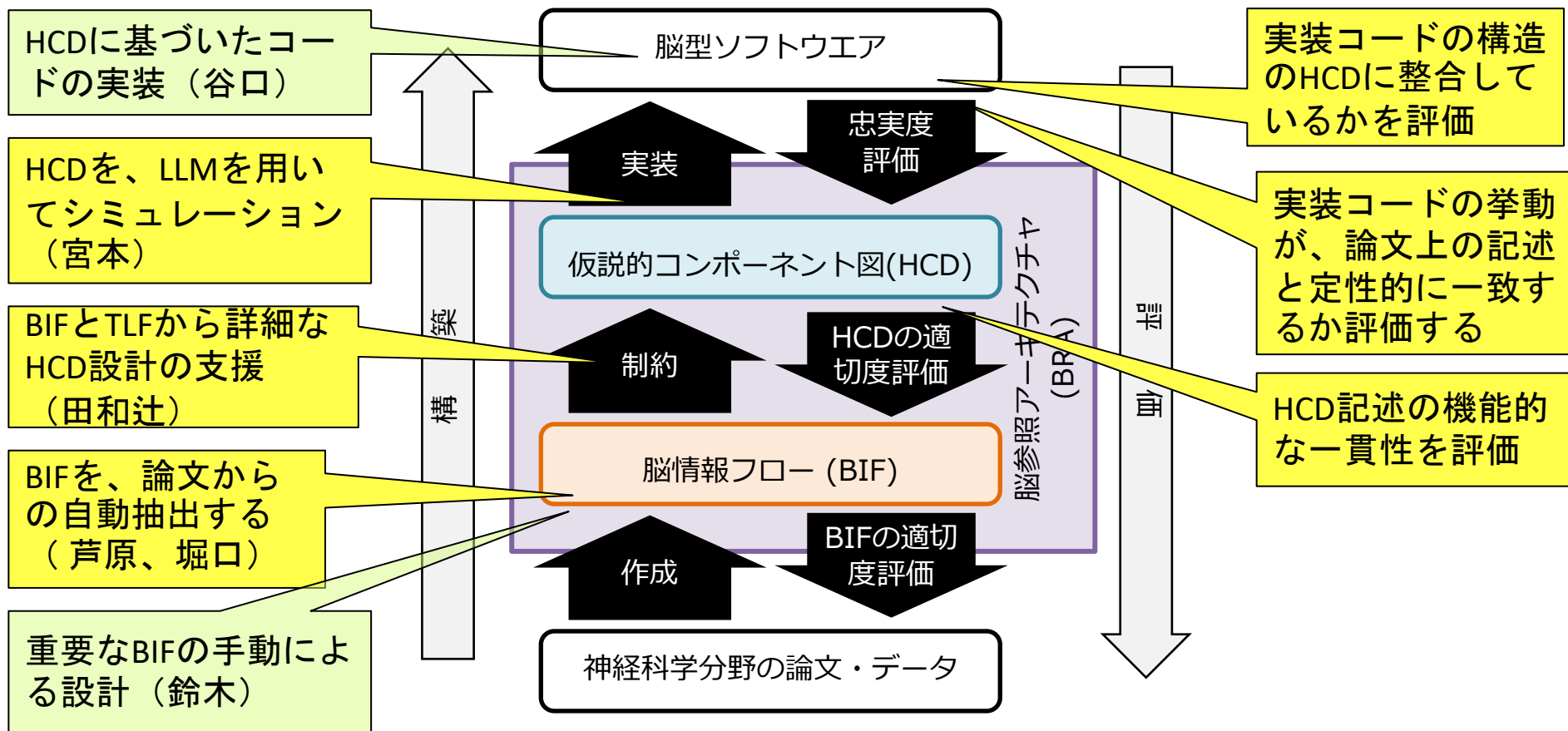


全脳アーキテクチャ(WBA)アプローチの現在までの主な進展

- ❑ 2013年末にWBA勉強会を開始
- ❑ 2014年にWBAアプローチが、「脳全体のアーキテクチャに学ぶことで、人間のような人工知能（AGI）を作る（工学）」と定義された
- ❑ 2015年に、上記を推進するWBAI（The Whole Brain Architecture Initiative）が創設された
- ❑ 2020年頃にBRA(Brain Reference Architecture)駆動開発が確立
- ❑ 2022年のLLM(言語学習モデル)の急進展の影響を受け、BRA駆動開発のペースが加速（但し、非脳型からのAGIの完成予想が早まった）

BRA駆動開発の自動化 そして 技術ロードマップ

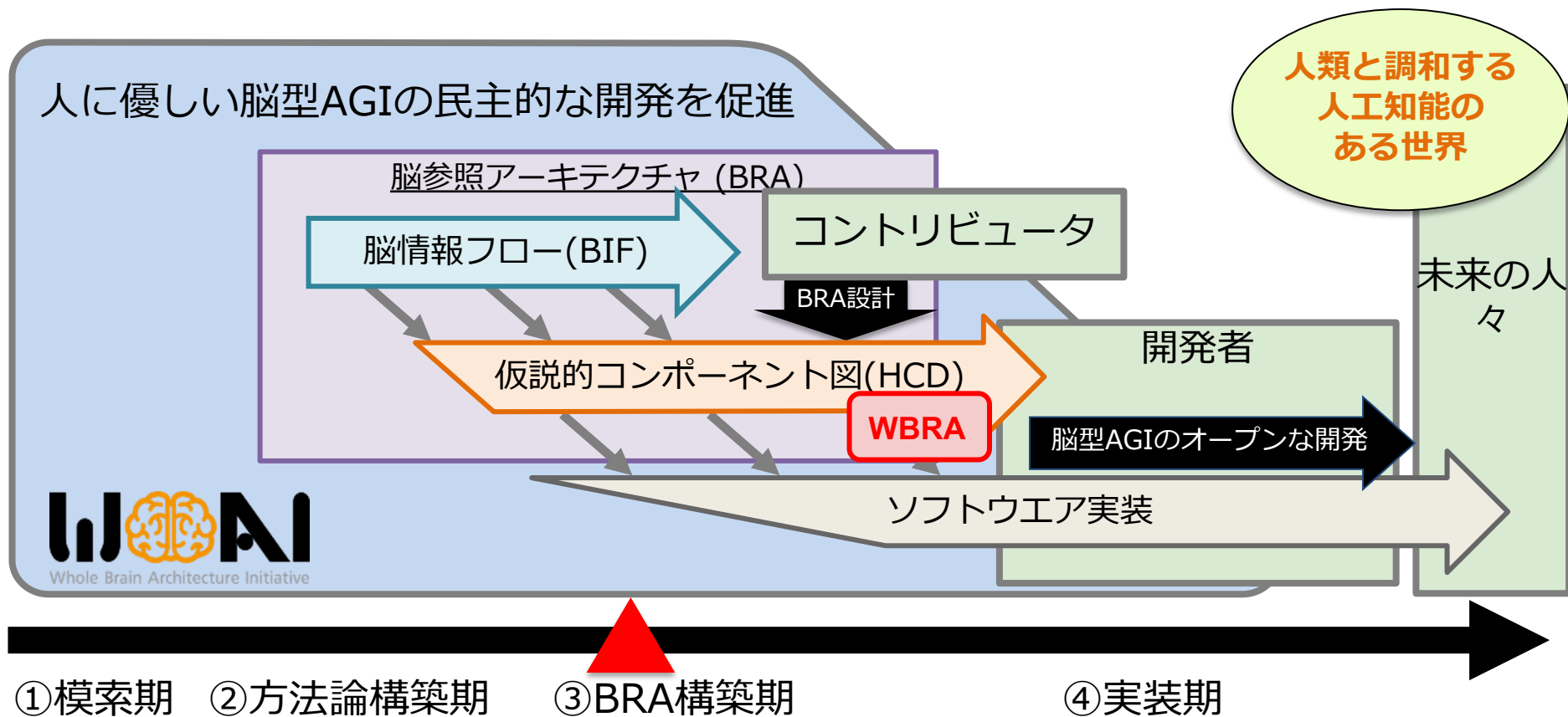
大規模言語モデル(LLM)はBRA駆動開発を様々に加速しうる



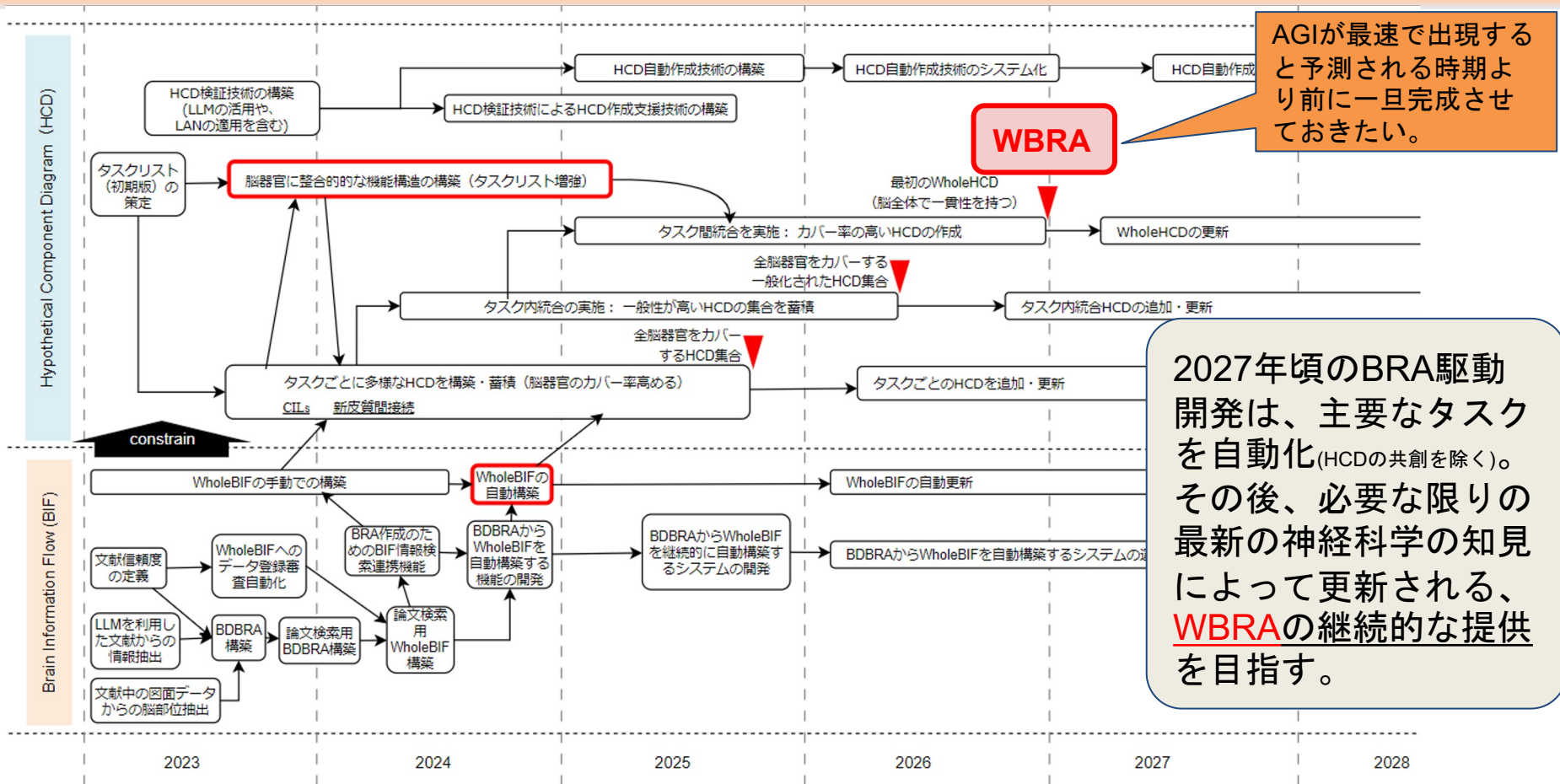
技術ロードマップを作る意義：**WBRA**を完成させる

- 我々WBAIがAGIに最初に到達するかもしれないし、他の組織がAGIに最初に到達するかもしれない（数年以内にAGIに到達すると予測する識者もいる）
- **脳型AGIは人類と超知能の架け橋**となりうるので、最初のAGIが超知能として離陸するまでの間に存在することが人類社会にとって特に有益となりうる
 - 超知能が完全に離陸した後は、その意義は薄れると思われる
- そこで、脳型AGIの設計情報である全脳についてのBRAである**Whole Brain Reference Architecture (WBRA)**を早急に作成し、公開したい。
- すると、非脳型AGIを作成した研究組織に対し、WBRAを用いて脳型AGIに改変する選択肢を提供でき、その選択肢は人類にとって有益となりうる
- **WBRA**の2027年までの完成を実現するためのロードマップを作成中
 - BIFの構築は、主にLLMを用いて自動化される予定
 - HCDの設計には引き継ぎ人的リソースが必要であり、共同作業を推進する

WBA技術ロードマップの外観



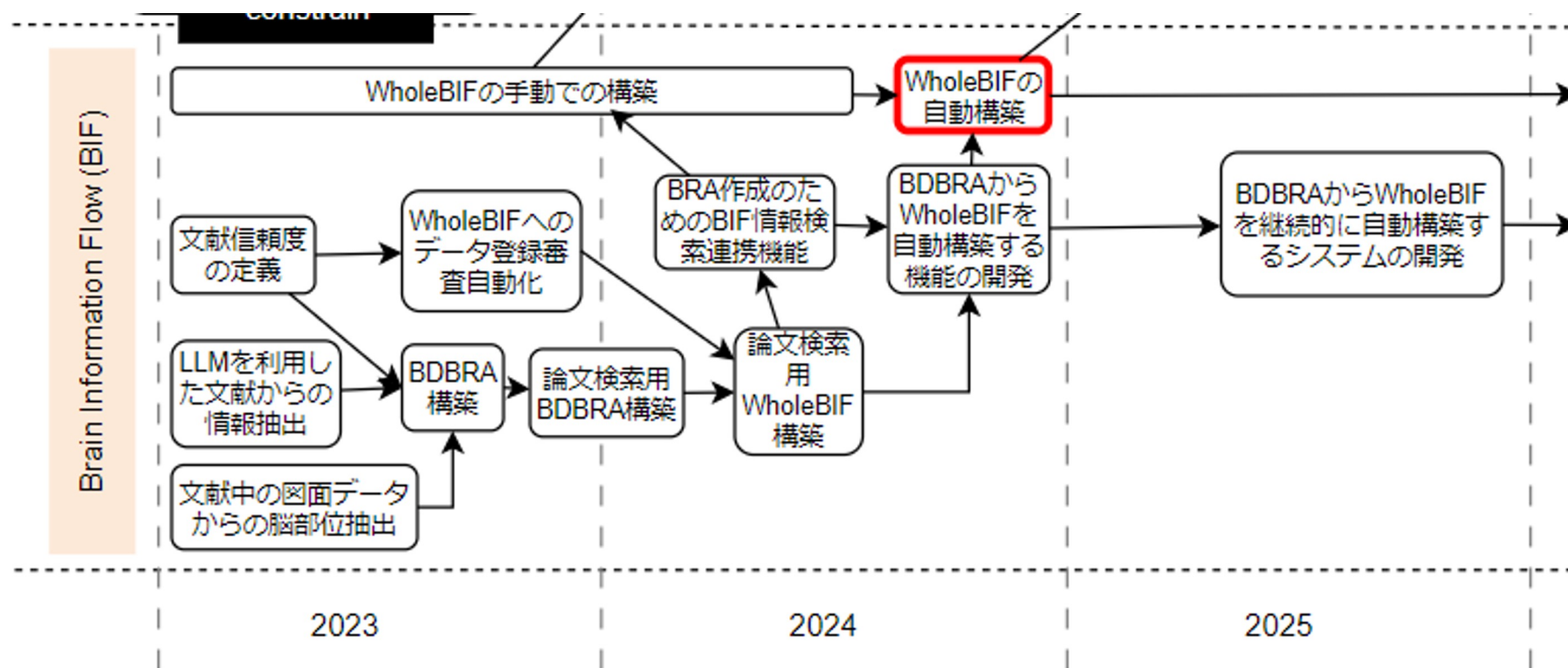
WBRAを構築するための技術ロードマップ



BRA駆動開発の長期見通し

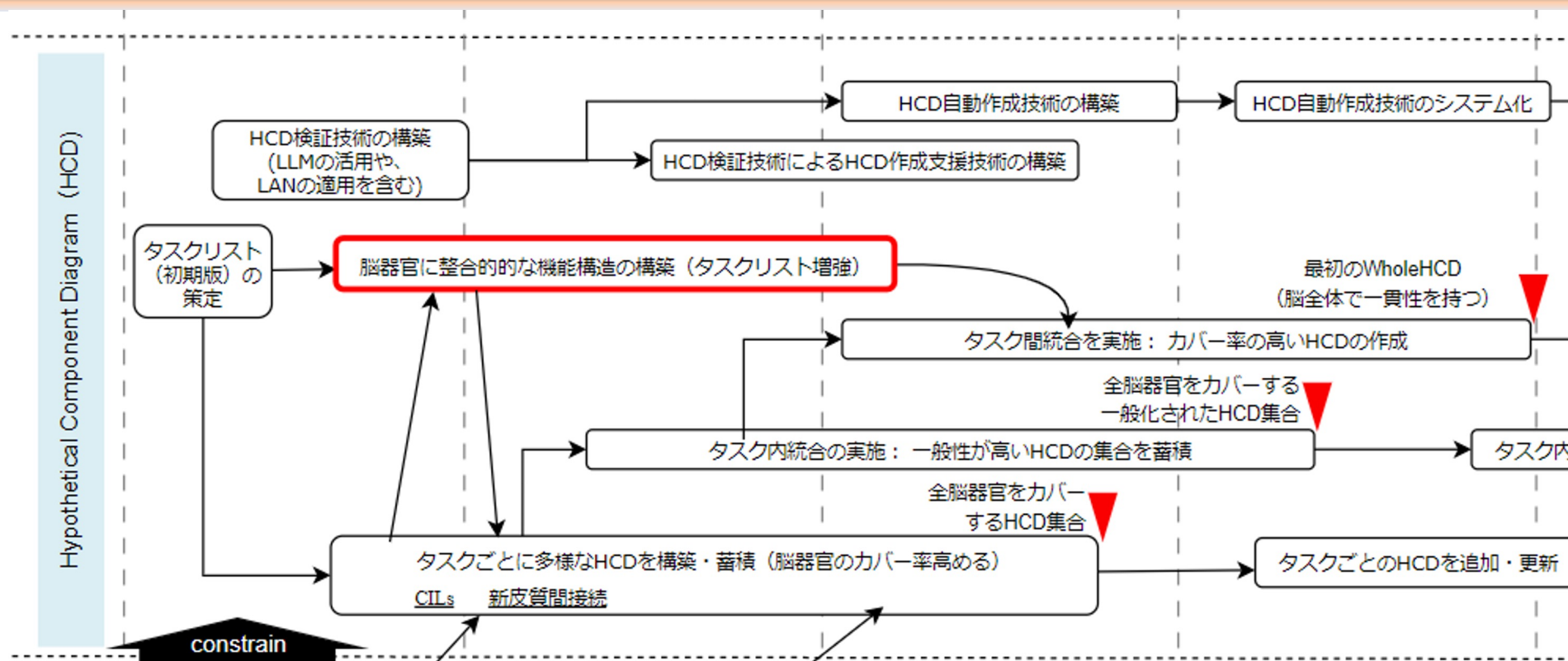
- 仕上げ期：2023～2027年頃
 - HCD共同作成の促進
 - WBRAを完成
- 過渡期：2028～2030年頃
 - 神経科学知見の増加に対応したBRA更新の運用自動化を進める
 - BRAに基づく脳型ソフトウェア実装などの利用促進
- 保存期：2030年頃から約10年間程度（必要に応じて延長）
 - BRA公開ポータルの維持
 - ほぼ自動化されたBRAの更新

Whole BIFの自動構築を目指すBIF



⇒ 芦原佑太：自動化に向けた取り組み(BDBRA, BIF)

様々なタスクに応じたHCD作成とその統合



⇒

田和辻 可昌：HCDの自動作成／統合／評価に向けた取り組み

折り返し地点にあるWBAIの今後の方向性

WBAIのプロジェクト期間は2015～2030年頃を想定していた。折り返し地点にある2023年は、環境変化も激しく、最終成果の創出にむけて舵取りを行うタイミングである。

- 制約条件
 - AGI出現までに残された時間の減少（2030年より前かも）
 - WBAIが蓄積してきた技術資産（BRA駆動開発）
- WBAIが最低限残すべき資産（Minimum Valuable Asset）
 - 応援者に対して説明可能な社会的貢献
 - 確実に実現可能な成果物（現在の積立金を消化するだけで可能）

- 芦原佑太：[10分] 自動化に向けた取り組み(BDBRA, BIF, HCD)
- 堀口維里優：[5分] BIF自動抽出
- 田和辻可昌：[10分] HCD自動作成／統合／評価（機能志向 SCID法、Great Crossword）
- 宮本竜也：[5分] 言語的定性シミュレーション
- 鈴木雄大：[5分] 新皮質SHCOM
- 谷口彰：[5分] 海馬モデルの実装