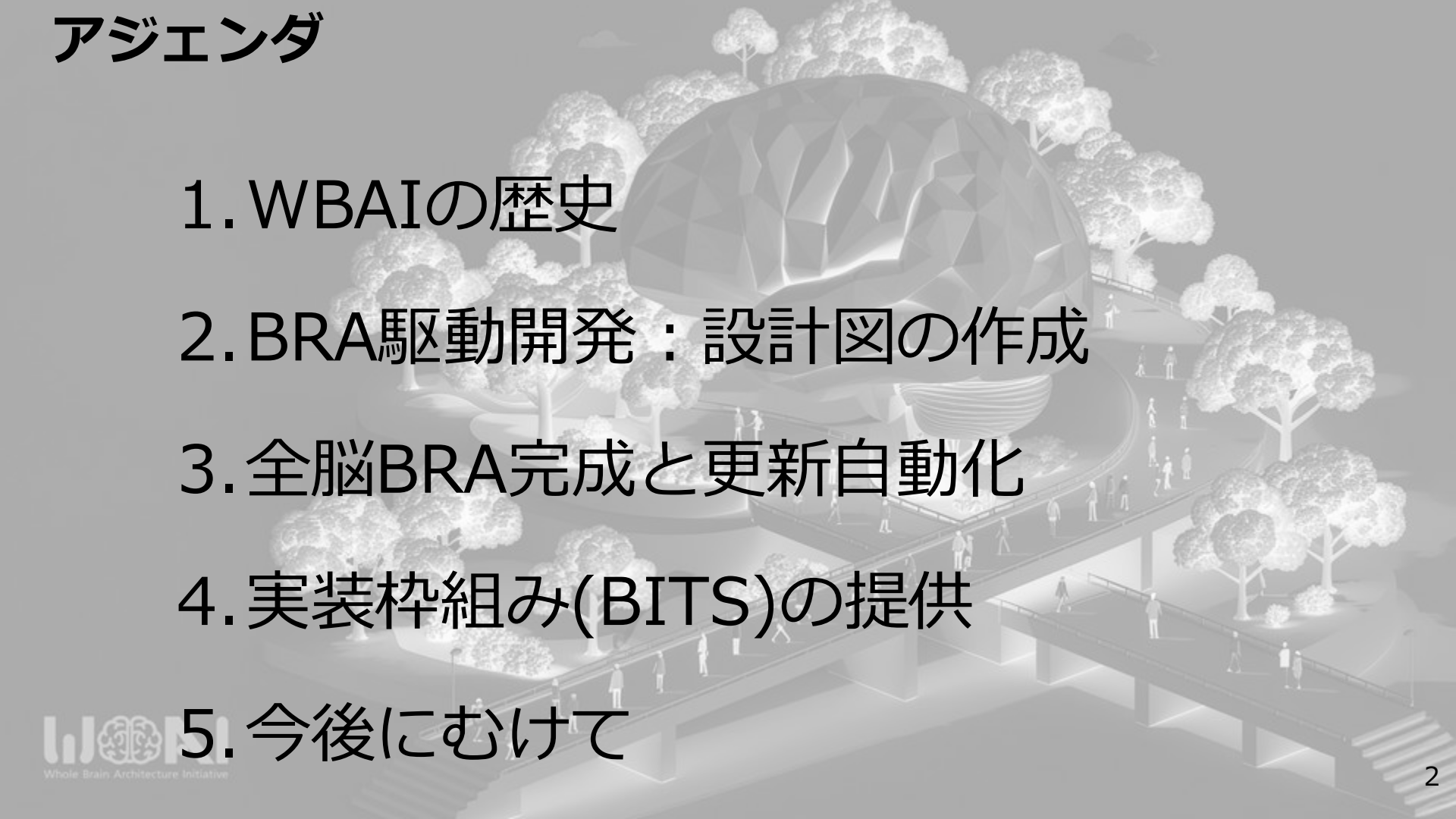


基調講演：WBAI 10年の歩み — 脳型AGIの設計から実装へ —

山川 宏

NPO法人 全脳アーキテクチャ・イニシアティブ
代表

アジェンダ

A stylized, low-poly illustration of a human brain in the center, surrounded by a park-like setting with trees, a winding path, and small human figures walking. The entire scene is rendered in a monochromatic, light gray tone.

1. WBAIの歴史

2. BRA駆動開発：設計図の作成

3. 全脳BRA完成と更新自動化

4. 実装枠組み(BITS)の提供

5. 今後に向けて

本発表の概要

- 全脳アーキテクチャ・イニシアティブ（WBAI）が10年間で確立した脳型AGI開発手法は、人間の脳構造を参照したBRA駆動開発手法により、内部動作が解釈可能で信頼性の高いAGIを実現する。
- これは従来のブラックボックス化しがちなAIの限界を突破する重要なアプローチである。
- こうした特徴を持つ脳型AGIは、非脳型AIでは代替不可能な、人類と超知能の仲介という点で独自価値を創出できる。
- 急速に進展するAGI開発状況において、2027年までに、その設計図(WBRA)完成は人類にとって重要な意味を持つ。
- この目標達成に向けて、大規模言語モデルを戦略的に活用して開発を大幅加速させている。
- さらに、ソフトウェアの実装と評価のシステム（BITS）を提示することで、ヒト脳型AGIが効率的に開発できる環境の準備を整えている。

アジェンダ

1. WBAIの歴史

2. BRA駆動開発：設計図の作成

3. 全脳BRA完成と更新自動化

4. 実装枠組み(BITS)の提供

5. 今後に向けて

WBAIの推進体制

顧問： 銅谷賢治（OIST） 北野宏明（システム・バイオロジー研究機構）
 富田勝（慶應義塾大） 森川博之（東大）
 中島秀之（札幌大） 岡ノ谷一夫（帝京大）
 尾藤晴彦（東京大学） 本多祥子（東京女子医科大学）

理事： 山川宏（代表） 松尾豊（副代表/東大）
 高橋恒一（副代表/理研） 田和辻可昌（副代表/東大）
 荒川直哉

監事： 中田康史

会議： 総会、理事会、その他委員会

全脳アーキテクチャ・イニシアティブ(WBAI)の歴史

- 2015年8月: WBAI設立
(WBAアプローチにより2030年頃のAGI実現を目指す)
- 2016年: NPOとしての開発「促進」軸の明確化
- 2017年: ビジョン "人類と調和する人工知能のある世界"を発表
- 2020年頃: 脳のリバーズエンジニアリングによる脳型ソフトウェアの設計手法を確立 (BRA駆動開発と呼ぶ)
- 2022年: LLMの活用を開始し、BRA駆動開発が加速する
- 2023年: 人類の未来に貢献できるヒト脳型AGIの仕様書(WBRA)を共同開発・公開する方針を決定
 - 2027年までにWBRAの初期版を完成させ、その後アップデートする計画

全脳アーキテクチャ・イニシアティブ (WBAI) Since 2015

良くも悪くも人類の福利厚生に対して大きな影響を与えるのは知能。

ヒトのあらゆる知的能力を凌駕する汎用人工知能(AGI)の実現は急速に現実味を帯びている。

「**人類と調和した人工知能のある世界**」というビジョン実現のため、脳型からのAGIのオープンな開発を促進している(当初は2030年目処)。

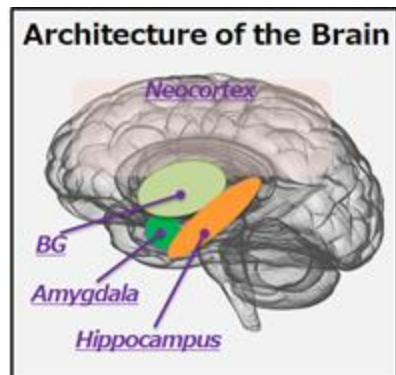
現在は、2027年までにヒト脳型AGIの設計情報(**WBRA**)の初期版完成に注力



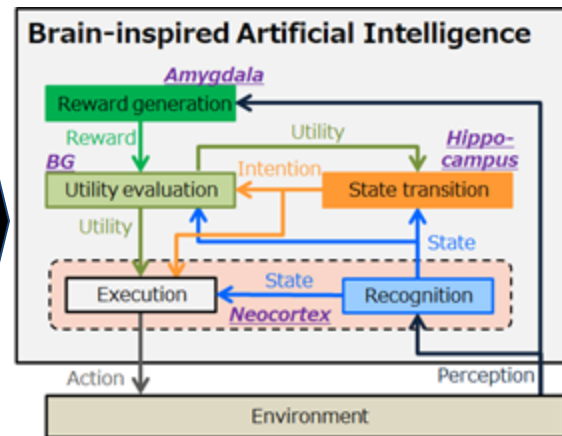
NPO法人WBAI代表

山川 宏

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
- 脳を参照して設計空間を限定する

人と親和性の高いAGIを作れる

- 対人インタラクションが必要な応用
- 精神疾患のモデル化など医療応用
- 人類と超知能の架け橋となる可能性

AIアーキテクチャを脳に学べる技術状況（2014年当時）

自動車に
例えると

人工知能
(認知科学)

ニューラル
ネットワーク

神経科学の
測定手法

ヒトの
脳

全体
設計図

認知
アーキテクチャ

脳全体の
ANNモデル

fMRI, EEG

脳全体

エンジン

WBA

機械学習
知識設計

AIアーキテクチャの
形を制約する

深層学習

コネクトーム
統合的
理解

新皮質の
領野

ピストン

高級な
記述言語

Multi-Layered
Perceptron

Single-Layer
Perceptron

二光子
イメージング

局所
神経回路

ネジ

プログラミング
言語

マカロック・ピ
ッツモデルニュー
ーロン

電極

神経細胞

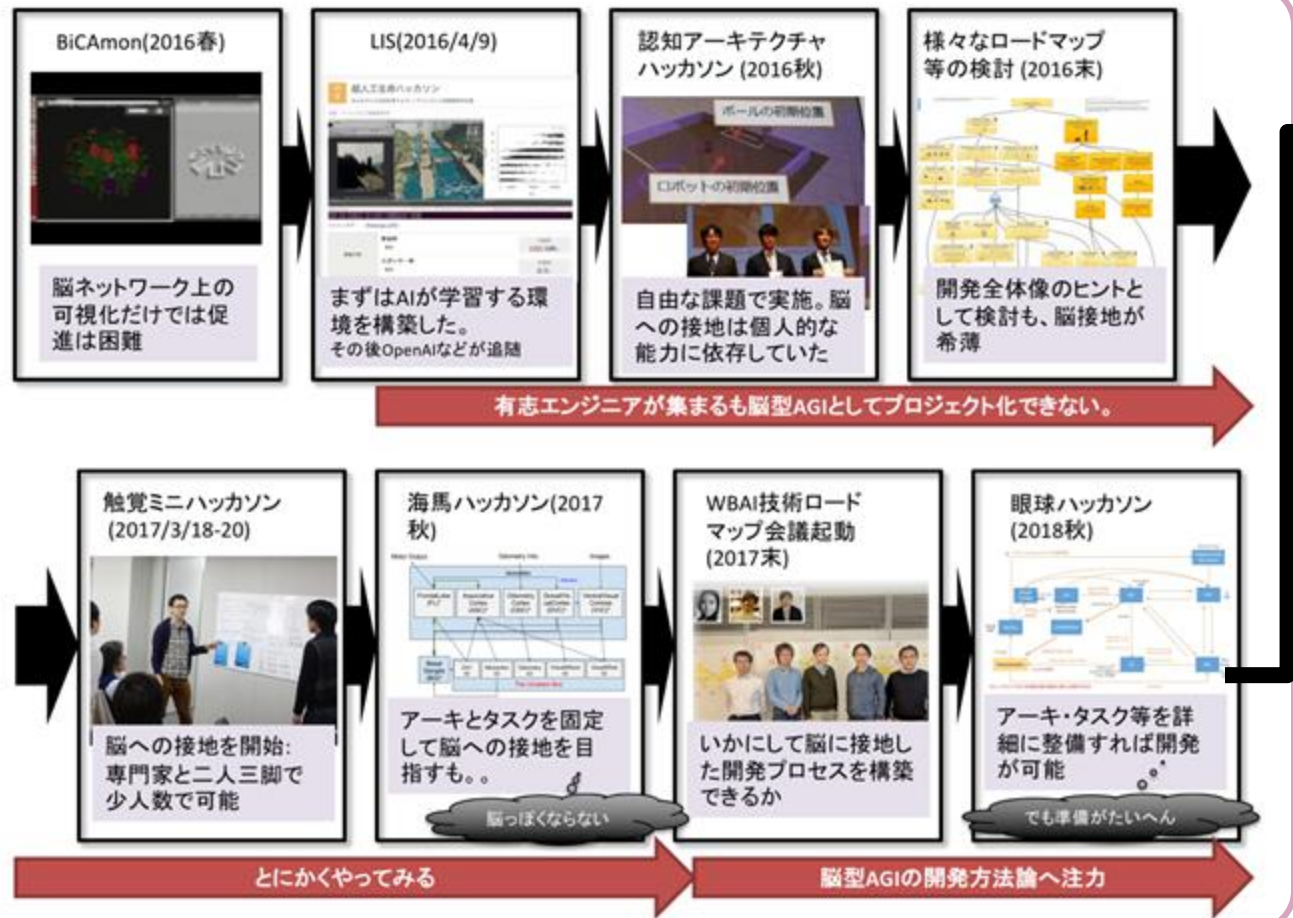
マ
ク
ロ

メ
ゾ

ミ
ク
ロ

脳に学ぶAGI開発手法の足取り (2016-2018年頃)

大澤さん



2017年までの結論：
脳とAIの両分野に詳しい人材が、アーキやタスクを詳細設計すれば脳型AIは実装可能

課題：

- 両分野に詳しい人材は少なく育成困難
- 個人による脳全体の一貫した設計は困難

解決策：

- BRA駆動開発
- 脳全体で標準化した設計情報の構築

アジェンダ

1. WBAIの歴史

2. BRA駆動開発：設計図の作成

3. 全脳BRA完成と更新自動化

4. 実装枠組み(BITS)の提供

5. 今後に向けて

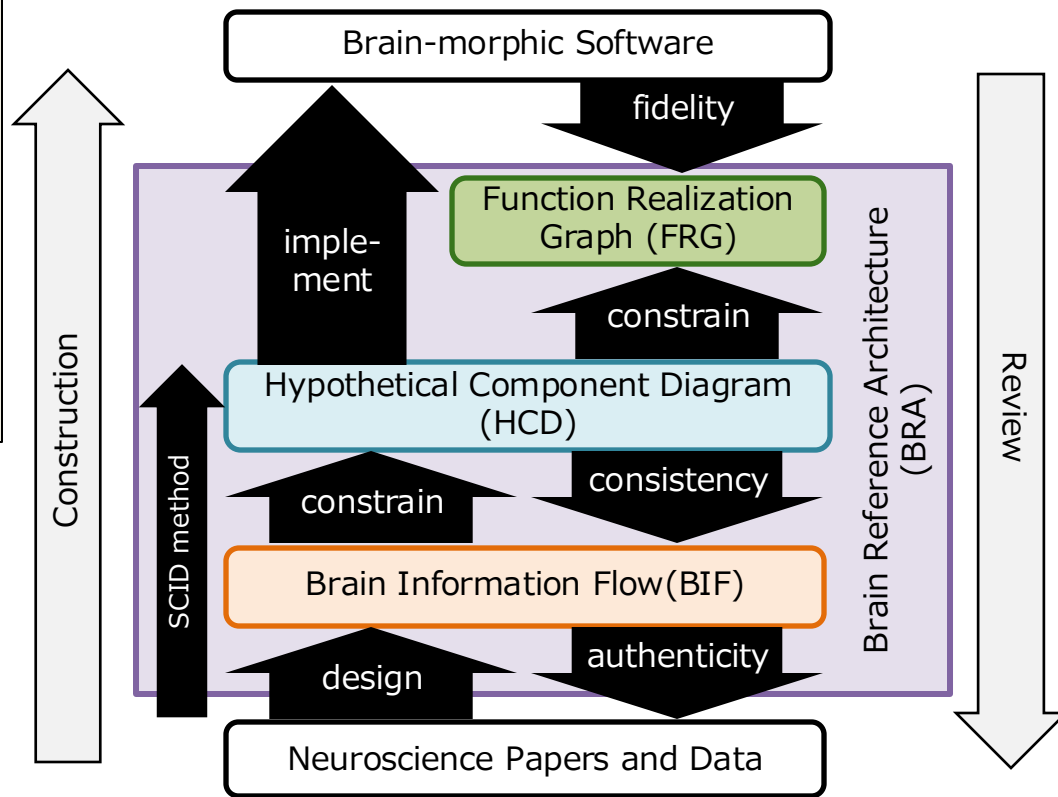
脳参照アーキテクチャ(BRA)駆動開発

脳全体の数千に及ぶ
神経回路モジュールを参照

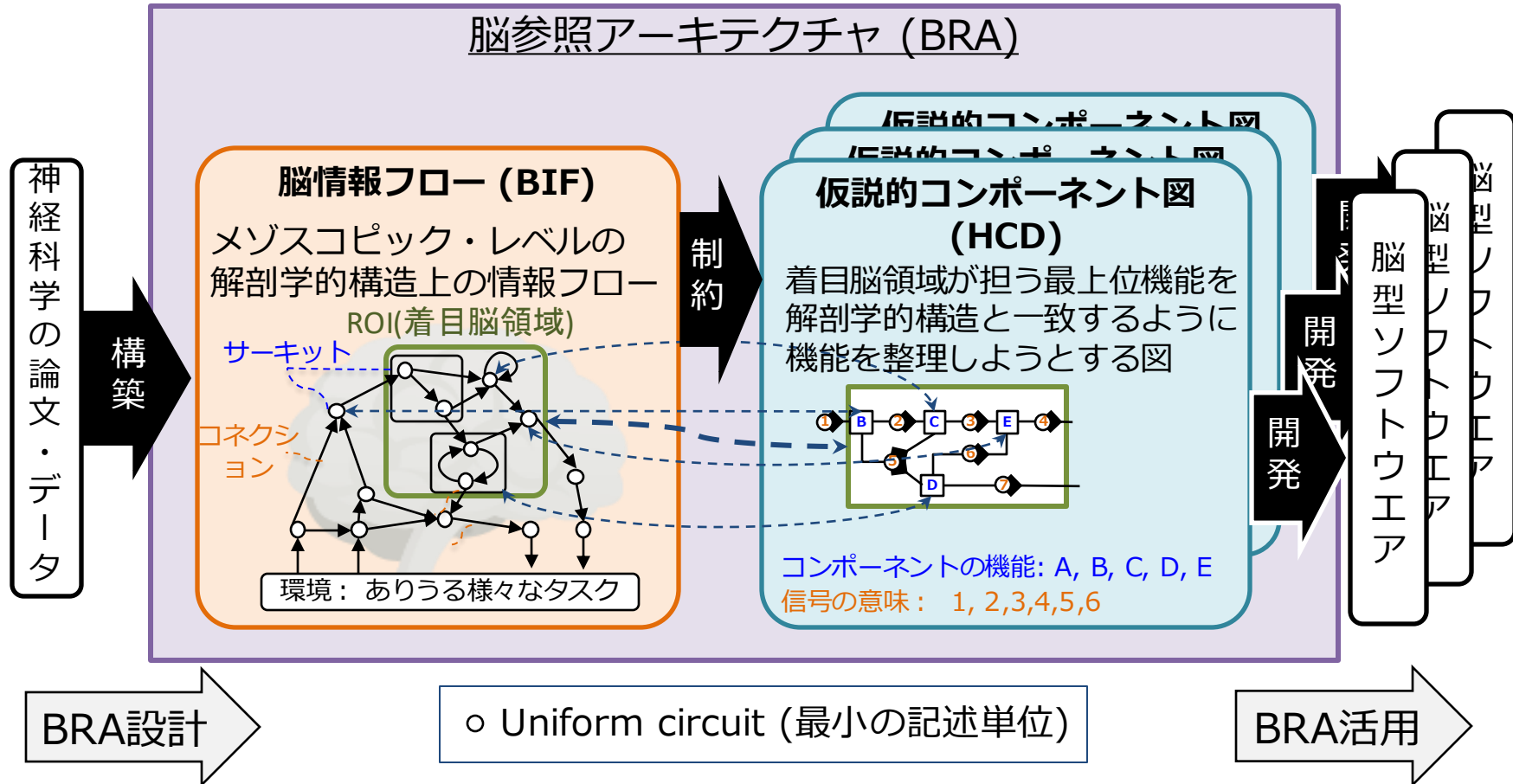


人間の解剖学的構造と一致した
認知機能を再現する
ソフトウェア構築

- Brain-morphic Software: 人の脳と同様の認知行動を行うソフトウェア
- BIF: 脳領域間の情報フローを示すダイアグラム
- HCD: BIFで示された脳領域をコンポーネントとして扱い、それぞれの入出力を表現
- FRG: HCD のコンポーネントが担う機能を階層的に定義するグラフ



BIFに関連付けられる様々なHCD



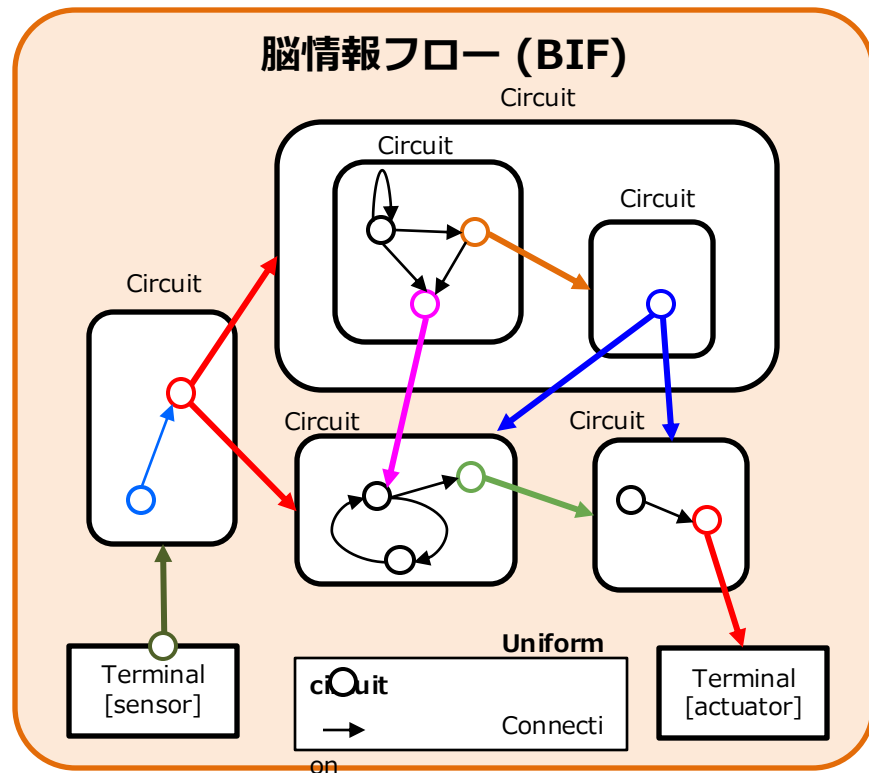
Uniform circuits : BIFの最小構成要素

Circuit (ノード)

- 脳内の何らかの神経回路(器官) に対応付
- 分類
 - **Uniform circuits :**
 - ある脳領域における特定の細胞タイプからなるニューロン群として設計。Circuitの最小単位であり、かつ、そのみが出来投射の始点となれる。ソフトウェアコンポーネントの単純な変数に相当する。
 - Other circuits: 一つ以上のUniform circuitを含む。
- 複数のCircuit間で重複部分を含みうる

Connection (リンク)

- Circuit間の信号伝達を担う軸索の束に相当
- Connectionの始点はUniform Circuitに限る

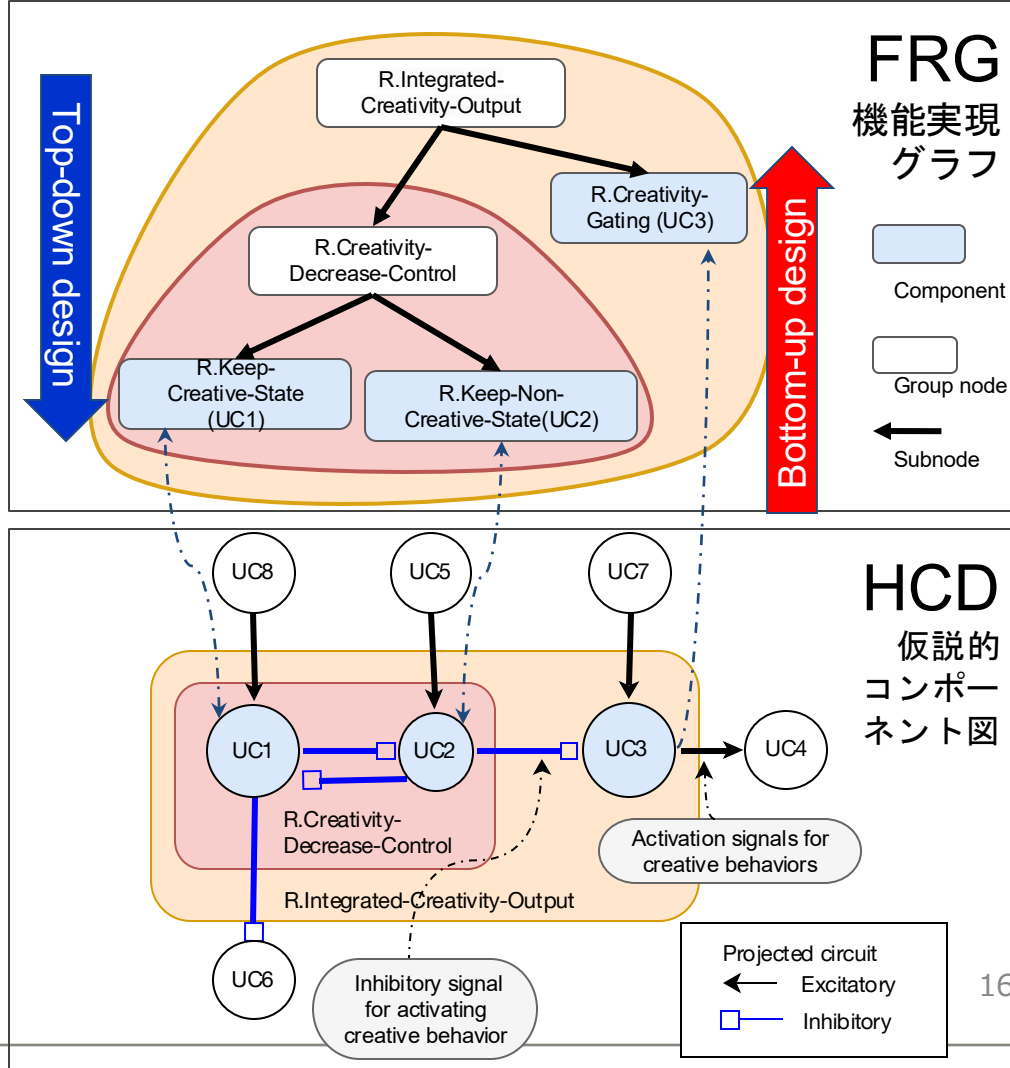


HCDに接地するFRG作成

両方向からの設計を整合させながら、未解明部分を含む脳の計算機能の有力な仮説を特定するリバースエンジニアリング

- 要求される機能を、トップダウンに分解して構造を得る
 - ソフトウェア設計の基本
- 実装できる機能を、ボトムアップに積み上げる構造を得る
 - 神経科学においては標準的

山川ら (2024). 脳計算機能の効率的なリバースエンジニアリングのためのデータ記述形式. *IEICE Conferences Archives*



アジェンダ

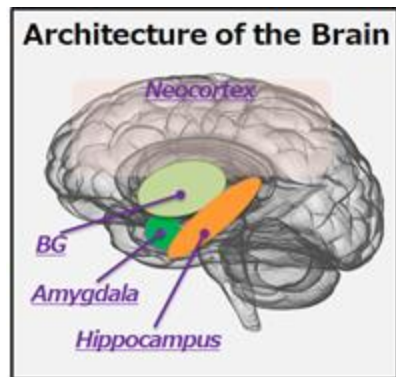
1. WBAIの歴史
2. BRA駆動開発：設計図の作成
3. 全脳BRA完成と更新自動化
4. 実装枠組み(BITS)の提供
5. 今後に向けて

ヒト脳型AGIをとりまく状況の変化

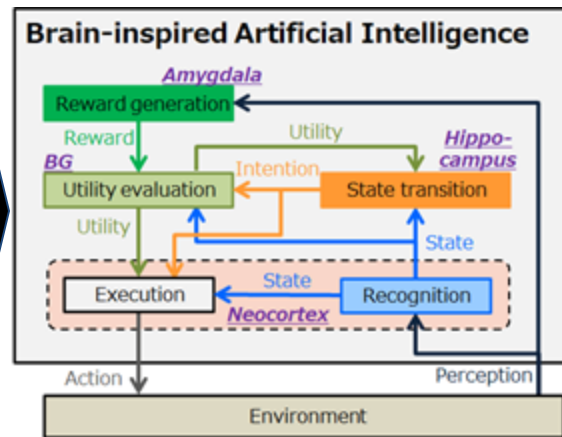
- 2015年頃
 - AGIの構築方法に目処がたっていなかった。
 - 脳型アプローチはAGI構築の有力な候補であった
- 2025年現在
 - 統計的機械学習(生成AI, 大規模言語モデル, Transformer)などの経路からAGI実現の可能性が増加している。
 - ヒト脳型AGIは、非脳型AGIが不得手な領域を担う技術として位置づけが進んでいる。

**人類が存続している限りは、
ヒト脳型AGIには何らかの価値があり続けるだろう**

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

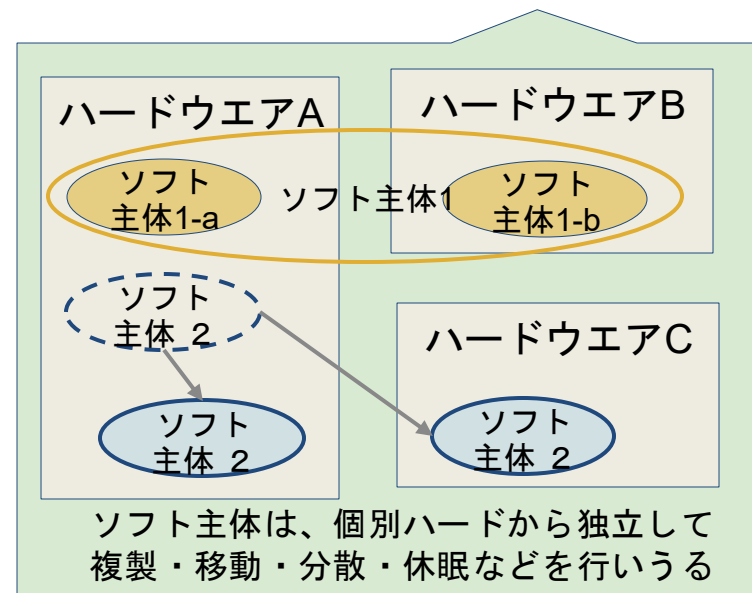
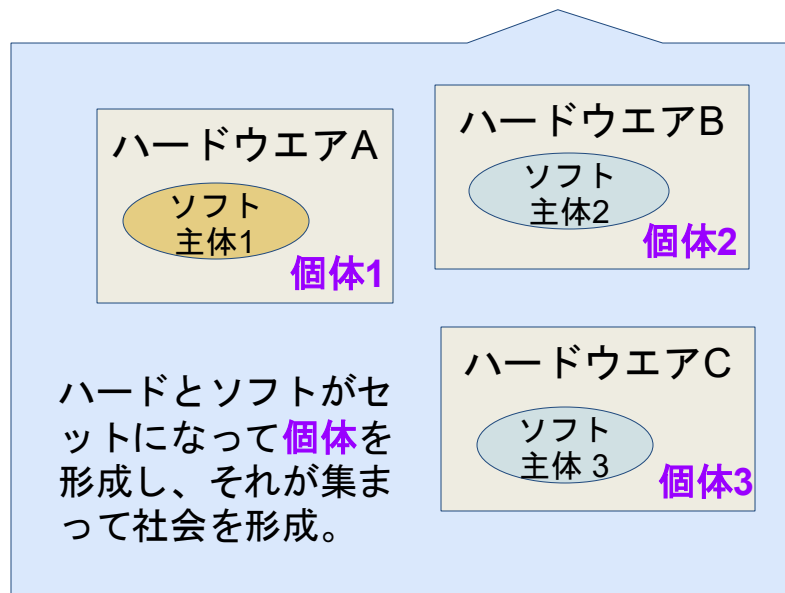
- 人間に関わる諸分野の科学知見を活用できる
- 開発の発散を防げる合意しうる認知アーキテクチャ
- 脳を参照して設計空間を限定する

人と親和性の高いAGIを作れる

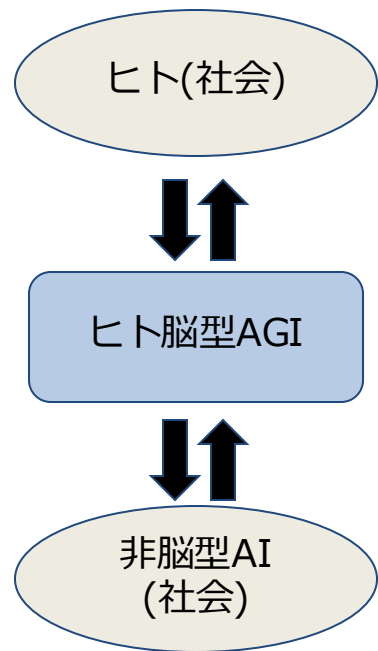
- 対人インタラクションが必要な応用
- 精神疾患などのモデル化など医療応用
- 人類と超知能の架け橋となる可能性

人類とAIでは行動の基盤が異なる：「個体」と「死」

ファクター	人類（地球型生命体）	AI（デジタル生命体）
境界の明確さ(個体)	身体による固定的境界・絶対的個体区分	情報パターンによる流動的境界・可変的定義
生存執着（死）	死への恐怖・強い生存本能・不可逆性	複製復元可能・生存執着の論理的減少傾向



ヒト脳型AGIが仲介者(翻訳者)として有効と思われる分野



分類	概念例	要 脳忠実度	ヒト脳型AGIの 存在意義
Type A 翻訳可能	物理法則 数学・技術	低	△ 既存AIで十分
Type B 部分翻訳可能	公正性・責任 品質・進歩	中～高	○ 有望な領域
Type C 翻訳困難	愛・尊厳 美・自由	高	△ 原理的困難さ（ 個体とその死）

Type B領域では、価値観の完全な共有は困難だが、実務レベルでの協力には有効性が期待される（理論的分析による予測）

脳に基づく解釈可能性

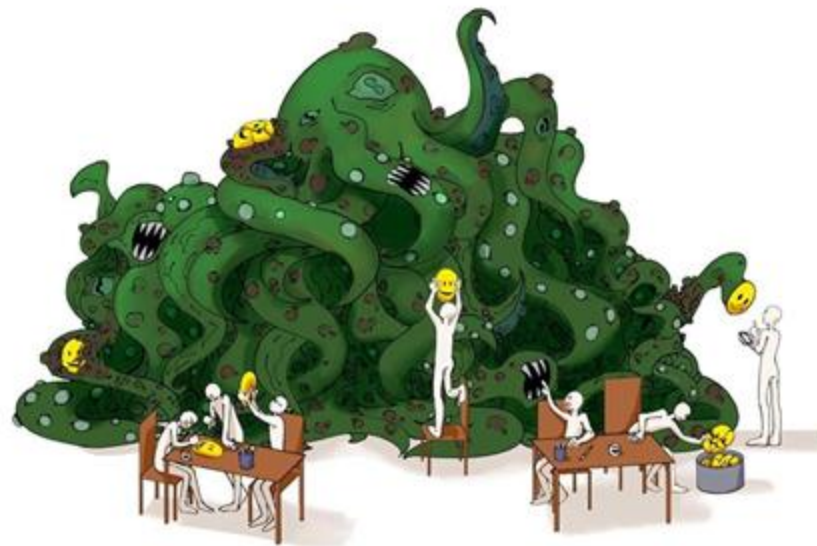
中嶋さん

WBAの解釈可能性:
ヒトの脳神経活動と対応付けを
通じて動作状態を解釈できる。



≠

機械学習AI + Alignment:
ヒトのマスクを被っている
だけで内部を理解が困難



Yampolskiy, R. V. AI: Unexplainable,
Unpredictable, Uncontrollable. (CRC Press, 2024).

戦略：WBAはいつ実現されるのが最適か？

WBAが先に作られる

- 危険な時期が2度ある。
- 全体として絶滅リスクが高まりうる

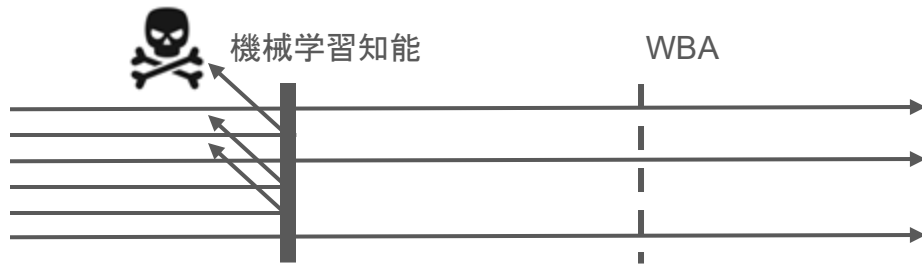


人類の存続可能性

低

機械学習知能が先に作られる

- 危険な時期は一度である。
- その時期の絶滅リスクは高まるかもしれない



中

(Bostrom, Superintelligence, 2014)中の図を改変したもの。
原図ではWBAではなくWBE(Whole Brain Emulation)としている。

戦略：WBAはいつ実現されるのが最適か？

WBAが先に作られる

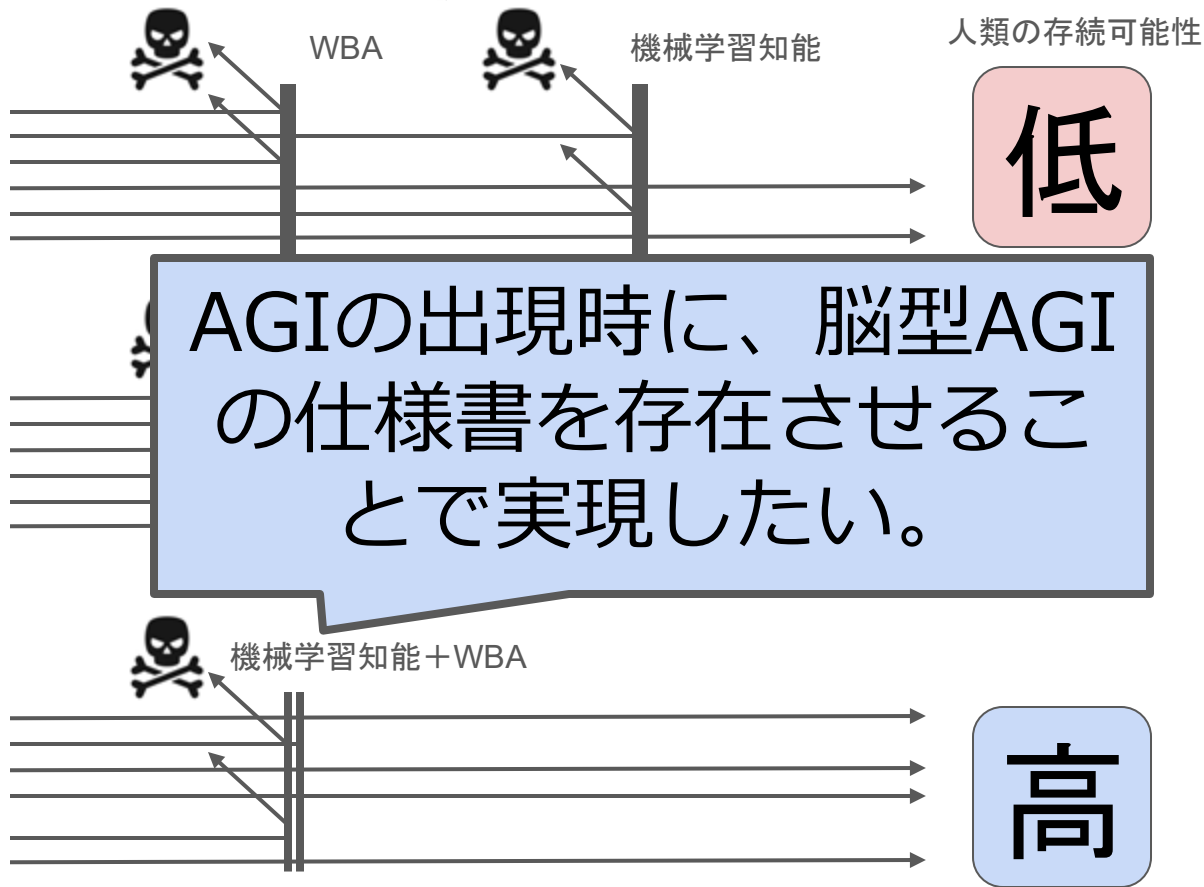
- 危険な時期が2度ある。
- 全体として絶滅リスクが高まりうる

機械学習知能が先に作られる

- 危険な時期は一度である。
- その時期の絶滅リスクは高まるかもしれない

両者がほぼ同時に作られる

- 危険な時期は一度である。
- その時期の絶滅リスクをWBAによって低減しうる



WBA技術ロードマップの外観

人に優しい脳型AGIの民主的な開発を促進

脳参照アーキテクチャ (BRA)

脳情報フロー(BIF)

コントリビュータ

BRA設計

仮説的コンポーネント図(HCD)

WBRA

開発者

脳型AGIのオープンな開発

(自動化が進む)ソフトウェア実装

2027年までにヒト脳型AGIの設計仕様であるWBRAを共同で作成・共有することを目指す。

人類と調和する
人工知能のある世界

未来の人々



Whole Brain Architecture Initiative

①模索期

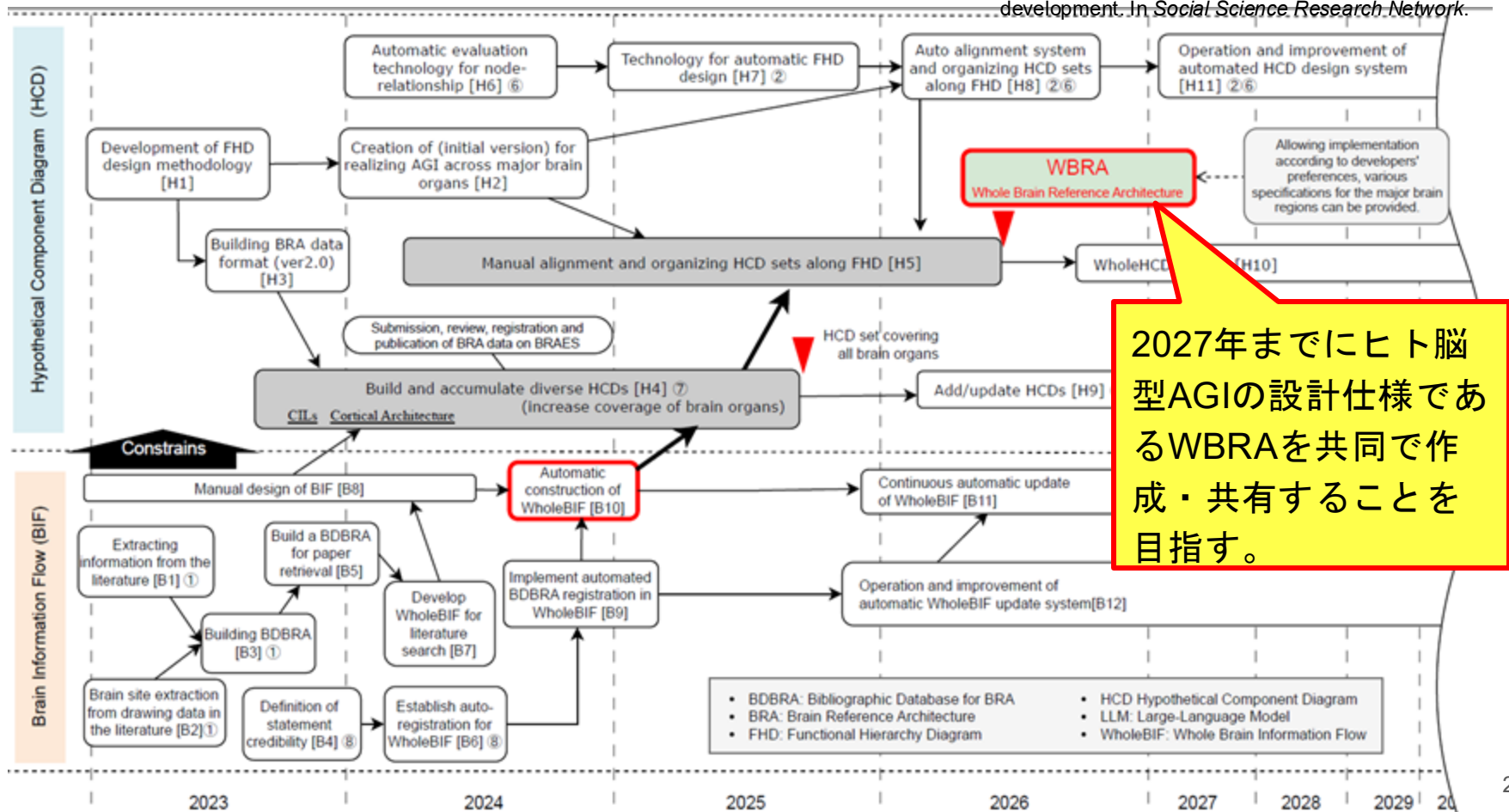
②方法論構築期

③BRA構築期

④実装期

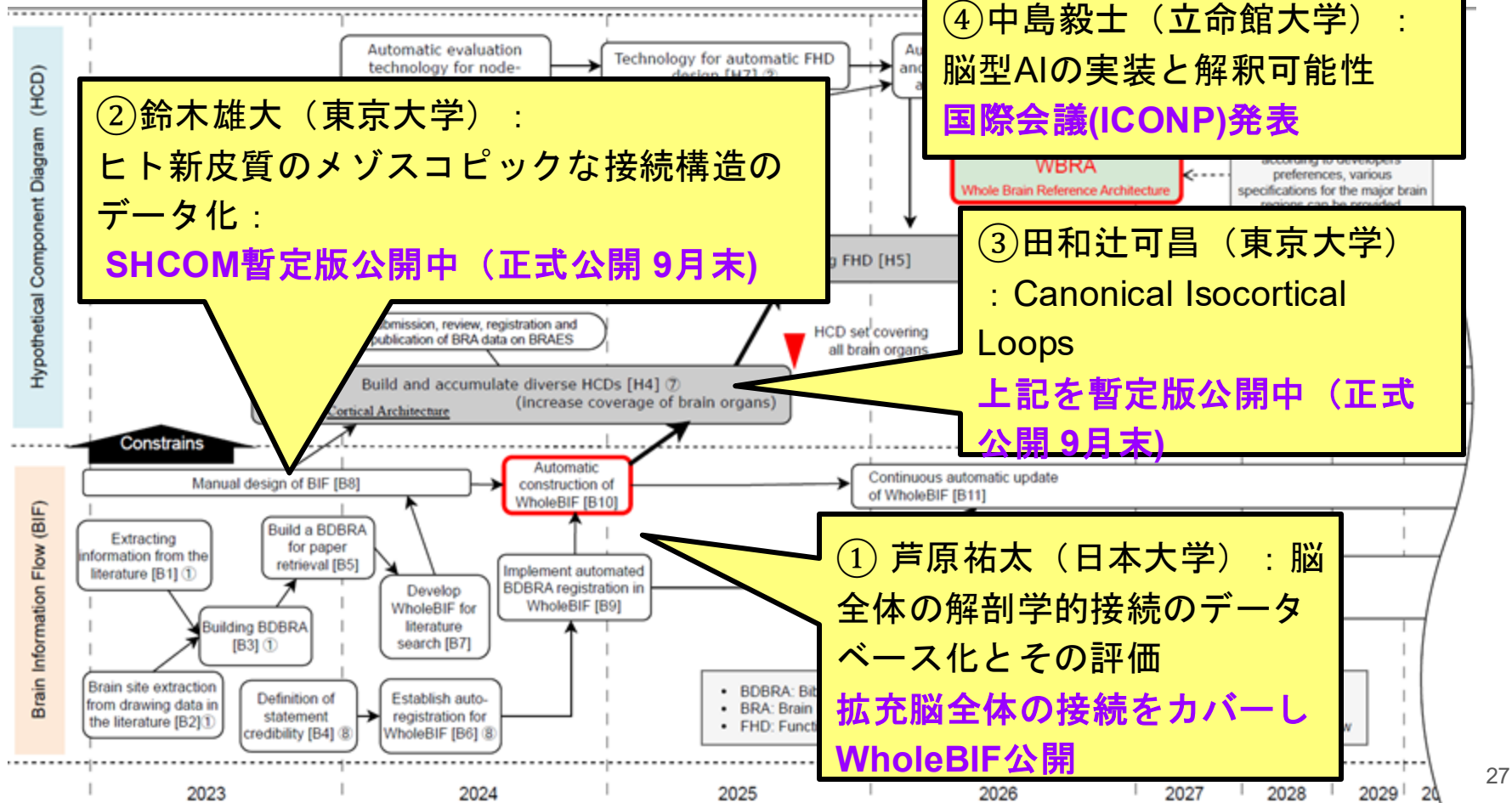
詳細化されたWBA技術ロードマップ

Yamakawa, H., et. al. (2024). Technology roadmap toward the completion of whole-brain architecture with bra-driven development. In *Social Science Research Network*.



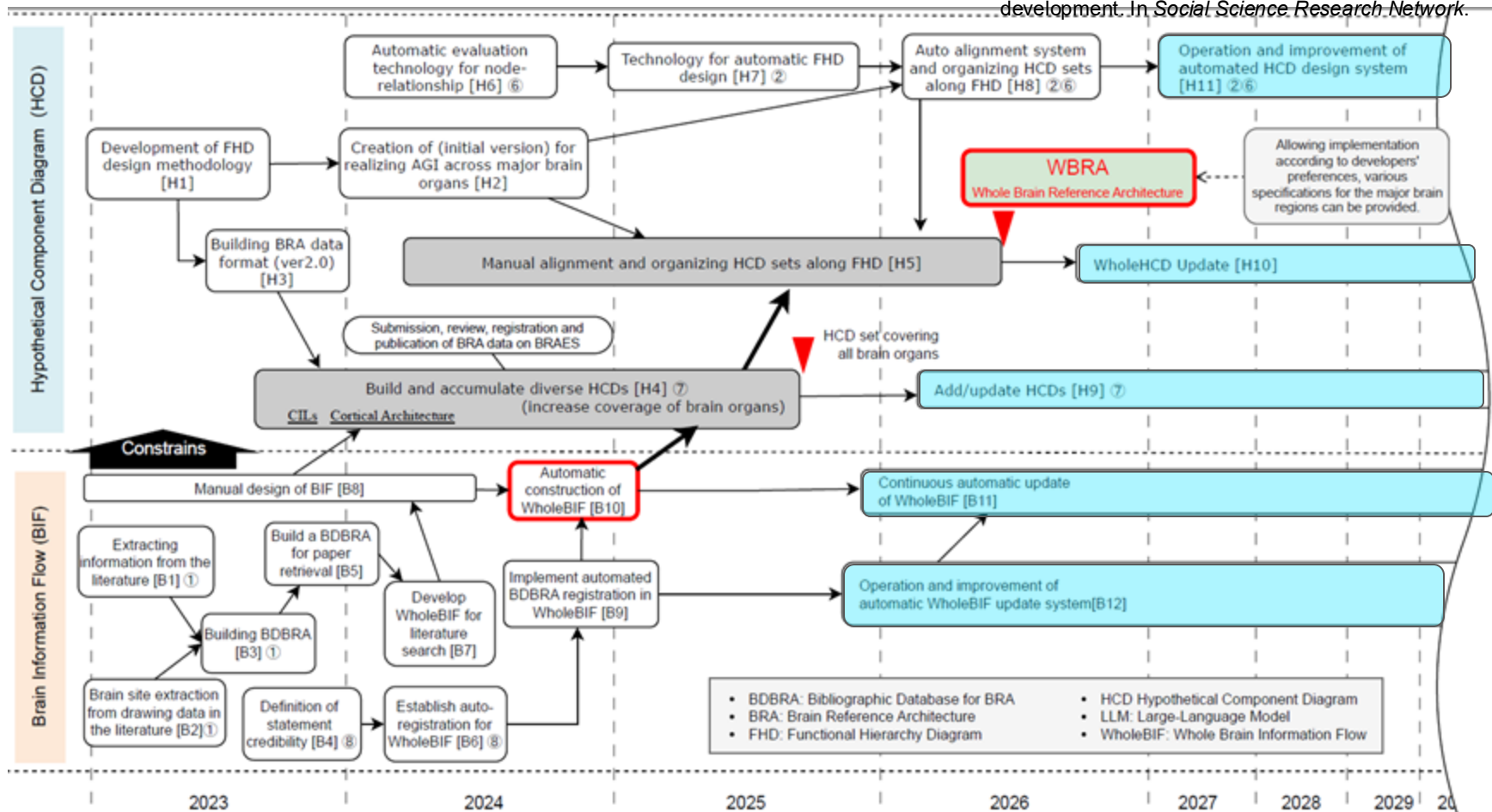
主要な技術成果（このあと説明）

Yamakawa, H., et al. (2024). Technology roadmap toward
driven

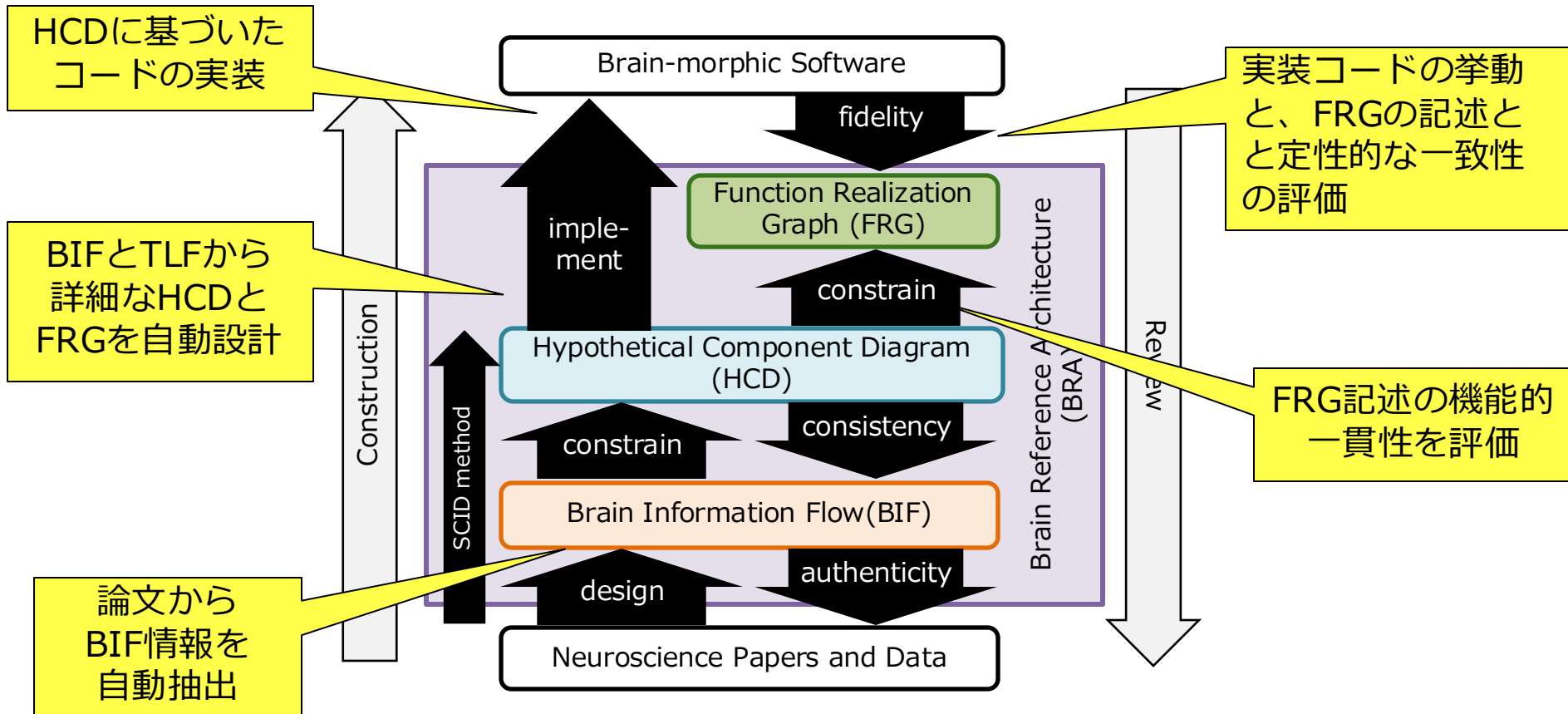


アップデートの自動化にむけて

Yamakawa, H., et. al. (2024). Technology roadmap toward the completion of whole-brain architecture with bra-driven development. In *Social Science Research Network*.



大規模言語モデル(LLM)を用いたBRA駆動開発の加速



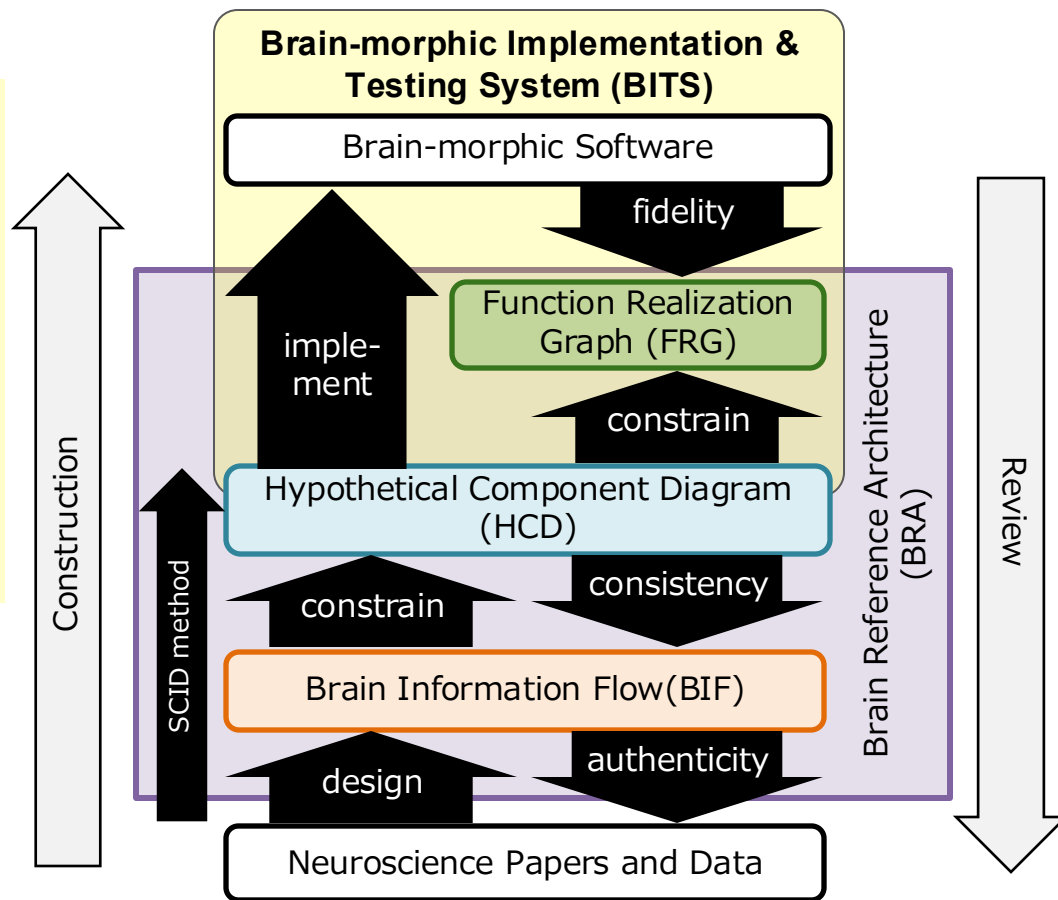
アジェンダ

1. WBAIの歴史
2. BRA駆動開発：設計図の作成
3. 全脳BRA完成と更新自動化
4. 実装枠組み(BITS)の提供
5. 今後に向けて

BRA駆動開発におけるBITS

Brain-morphic Implementation & Testing System (BITS) :

BRA駆動開発の出口として、Brain-morphic Softwareを作成及び評価するシステムの提供準備を進めている。

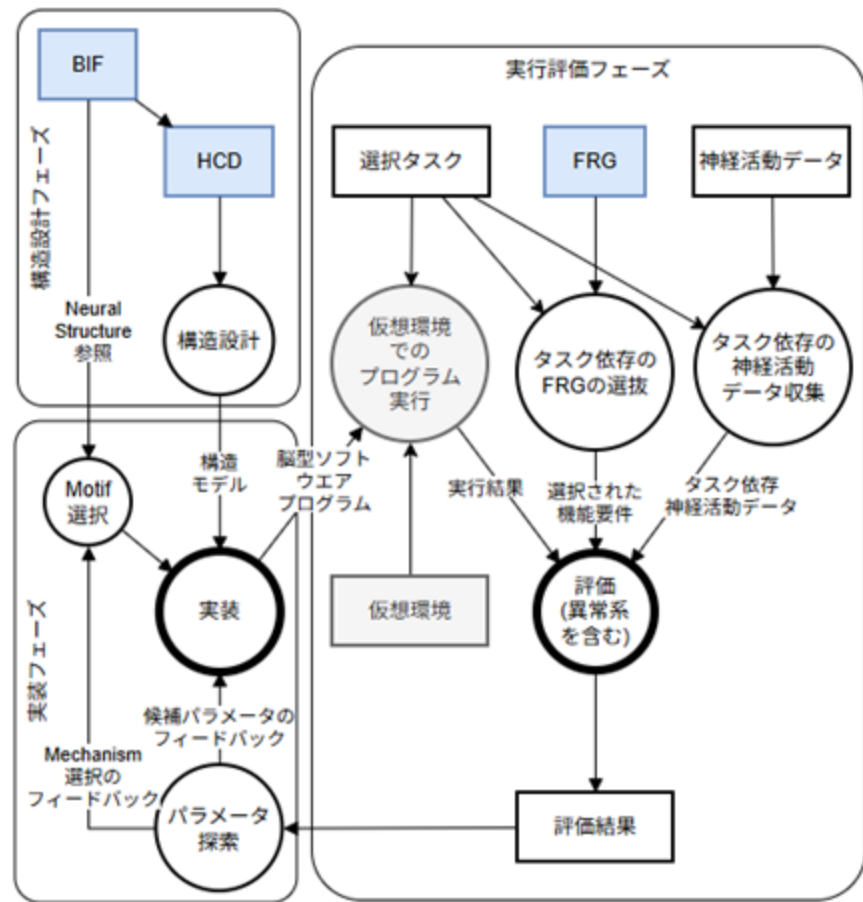


BITS : BIF/HCD/FRGに基づいてソフトを実装・評価

Brain-morphic software Implementation & Testing System

- 構造設計フェーズ：
HCDを基に計算モデルのグラフ構造を定義
- 実装フェーズ：
モデルを脳型ソフトウェアとして実装
- 評価フェーズ：
タスクを選択し、仮想環境上でソフトウェアを動作(チューニング)させ、その後に評価

評価結果は、必要に応じて繰り返し
実装フェーズにフィードバック



実装フェーズのフローチャート

BRA駆動開発の実装は、基本的にはUniform Circuit (UC)毎に行われる。

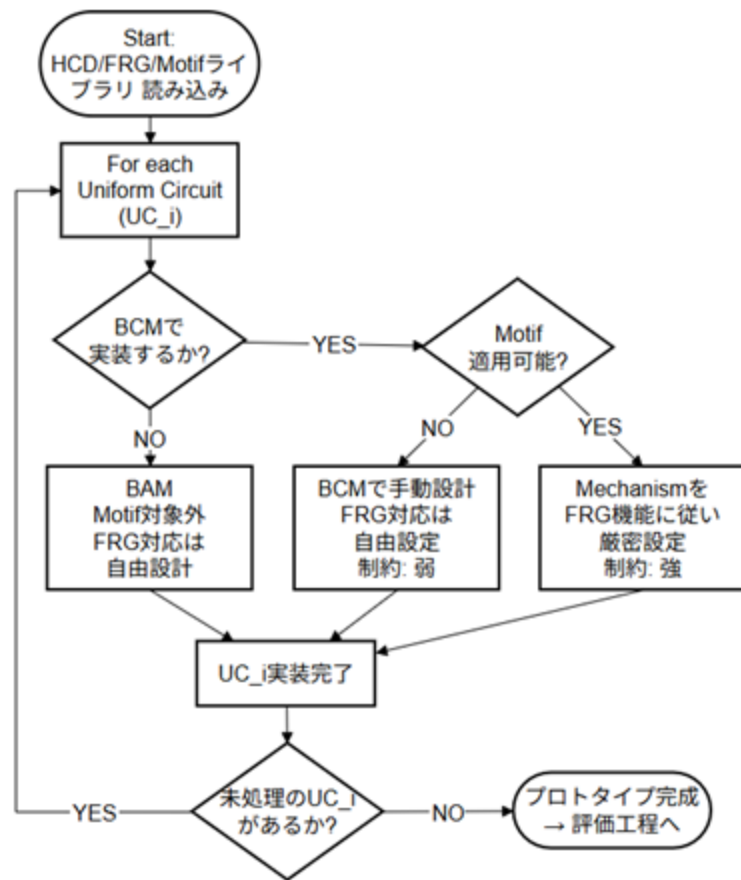
UC: 特定の脳領域内の特定の細胞タイプによって構成されるニューロン群

BCMでは「**Capability-Constrainedness**」を特にMotifにおいて利用できる。

つまり、Motif内のパラメータ(結合重みなど)を、特定のCapabilityを実現する範囲に制約できる。

例えば、WTA機能を実現する重みや時間特性には、自ずと適正な範囲が存在する。

この制約を利用できることは、BAMにはないBCMの利点である。



Motif : 脳型ソフトの実装の機能要素

Motif: 創発的かつ脳内で繰り返し出現するMechanism

例: Wilson-Cowan modelで実装するWTA motif

Capability:

Winner take All

Group node mechanism

x1によって活性化した
z2-z4の活動和に基づ
いてz1がそれらを抑制

Node mechanism:

パラメータを指定された
node algorithm

Node algorithm:

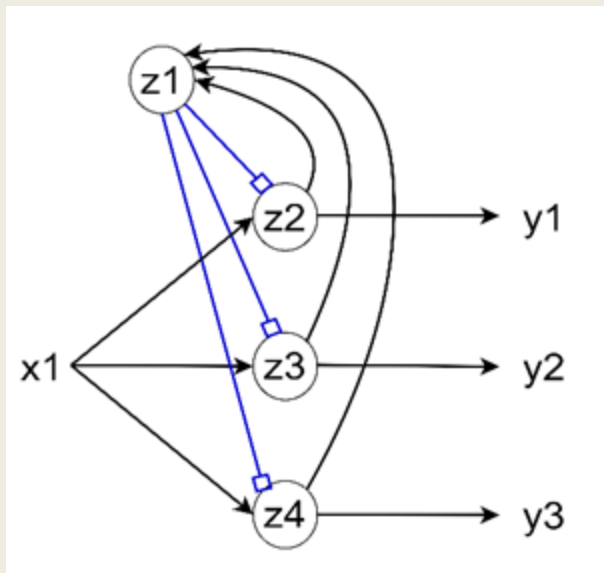
Wilson-Cowan model

Coding scheme:

集団平均発火率

Neural structure

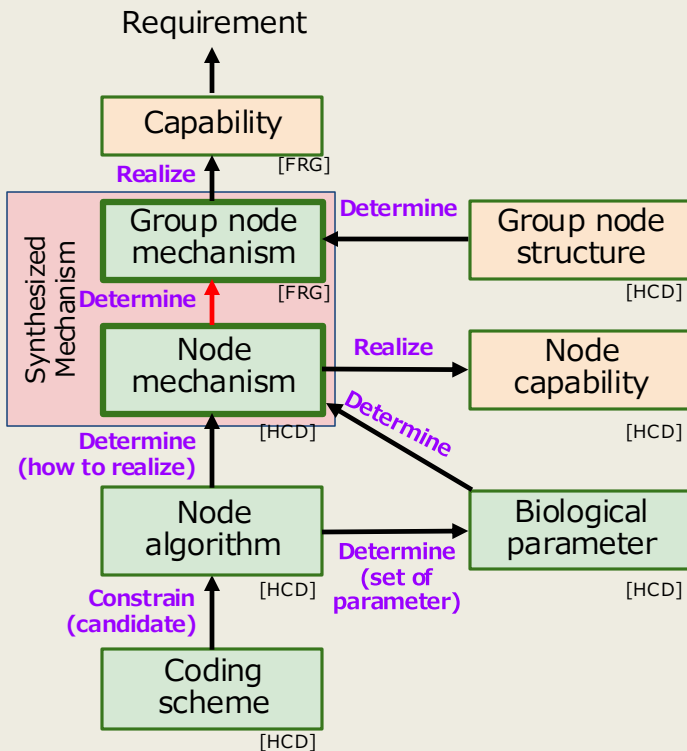
右図の通り



x1:入力, yi:出力, zj: 隠れ変数

Capability-Constrainedness:

複数のNode mechanismを組み合わせた
Group node mechanismが発揮する特定
のCapability(WTA等) が、Biological
parameterの範囲を制約する性質である。



ライブラリ化が進むMotifとそのCapability

Neural Structure ID	Reference	Network diagram	抑制性ノードの有無	ループの有無	並列性	Capability									
						フィードフォワード	微小な遅延線 (Minor)	自己維持活動 (Self-S)	Winner-Take-All	バイスタブル / マル	シナジー効果 / ク	パターン補完 (Pattern)	ゲーティング	時間的フィルタリング	過剰興奮防止 (Prevention)
Recurrent_Excitation	[Braganza, 2018]		No	No	Yes			○		○		○			
Convergent-Excitation	[Luo, 2020] A		No	No		○						○			
Divergent-Excitation	[Luo, 2020] B		No	No	Yes	○									
Feedforward-Excitation	[Luo, 2020] C		No	No		○	○								
Feedback-Excitation	[Luo, 2020] D		No	Yes				○		○		○			
Reciprocal-Inhibition	[Godard CA., 2014]		No	Yes	Yes				○		○				
Lateral-Excitation-Transfer	[Luo, 2020] E		No	Yes	Yes						○	○			
Feedforward-Inhibition	[Braganza, 2018]		Yes	No									○	○	○
Disinhibition	[Braganza, 2018]		Yes	No									○		

評価フェーズ：実装を評価する

評価要素	BITS評価レベル	評価方法
構造の類似性	Group node structure [BIF]	実装のコンポーネント配置とBIF記述の一致性評価
機構類似性 (機能類似性)	Group node mechanism [FRG]	機能仮説を反映したプログラムとFRGの機構比較
活動再現性 (現象)	Coding scheme [HCD]	コンポーネント内変数の挙動と神経活動データの一致性評価
Capability (部分システム)	Capability [FRG]	特定タスクにおける部分システムの能力評価
パフォーマンス (Capabilityの特殊ケース)	Capability [FRG] (機能比較・全体)	多様なタスクにおけるシステム全体の性能評価

※ 構造や機構の変更による、活動やCapabilityの異常も評価対象となる

アジェンダ

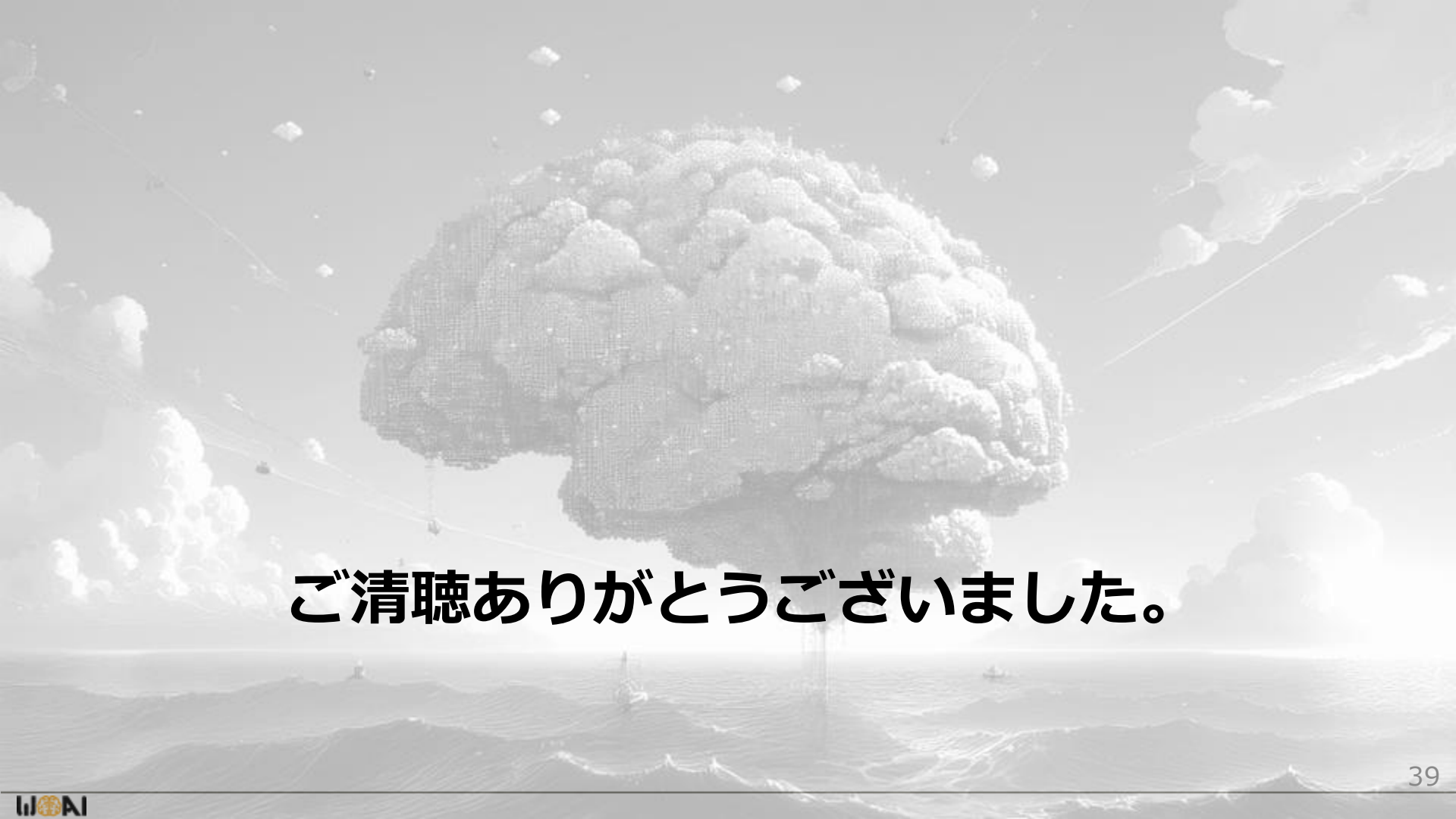
1. WBAIの歴史
2. BRA駆動開発：設計図の作成
3. 全脳BRA完成と更新自動化
4. 実装枠組み(BITS)の提供
5. 今後に向けて

今後にむけて：脳型AGIの設計から実装へ

-  **ヒト脳型AGIの意義**：超知能時代でも、人間との仲介やマインドアップロードの基盤として、ヒト脳型AGIの価値は永続。
-  **目標**：脳型AGI開発の設計図となる全脳BRA（参照アーキテクチャ）を2027年までに完成させ、アップデートを自動化。
-  **実装基盤の整備**：BRAに基づくソフトウェアの実装・検証を支援するシステム（BITS）の提供準備を進めている。

ご清聴ありがとうございました

脳型AGI実現に向けて、設計から実装への橋渡しを加速するため
皆様からのご支援をお待ちしております！



ご清聴ありがとうございました。