

## 脳を参考としてヒトのような汎用人工知能を開発する方法論を標準化

1. 発表者： 山川 宏（東京大学大学院医学系研究科脳神経医学専攻客員研究員、東京大学工学系研究科技術経営戦略学専攻特任研究員、特定非営利活動法人全脳アーキテクチャ・イニシアティブ代表、理化学研究所生命機能科学研究センター客員主管研究員）

### 2. 発表のポイント：

- ◆脳のように振る舞う汎用人工知能などのソフトウェアを実装する際の仕様情報となる、脳参照アーキテクチャ(BRA, 注1)データ形式と、それをを用いた開発方法論を標準化しました。
- ◆上記方法論で扱う解剖学的構造の標準的な記述粒度の決定指針を示し、その構造記述に沿った形で計算機能を付与する方法論を定式化しました。
- ◆BRA形式のデータが蓄積および共有が進み、脳の計算論的理解や、認知モデル/脳型人工知能の開発、それら知見の精神医学やヒューマン・エージェント・インタラクションへの応用などの波及効果が期待されます。

### 3. 発表概要：

人間の認知機能を脳全体の神経回路を参照しながら再現するソフトウェアの構築は、認知科学や神経科学などの人間科学や、人工知能やロボット工学などの工学的応用において高い価値があると考えられています。しかし、そうしたソフトウェアの設計基盤となる神経科学の知識は膨大かつ複雑であるため、特定の個人の能力によって設計することは困難です。さらに、脳科学の知見を認知機能に適切に反映させるためには、それに必要な解剖学的記述粒度を適切に揃える必要があります。東京大学大学院医学系研究科脳神経医学専攻において、山川宏客員研究員は、そうしたソフトウェアを実装する際の仕様情報となる脳参照アーキテクチャ(BRA, 注1)データ形式と、それをを用いた開発方法論を標準化しました。この方法論では、解剖学的構造を記述する粒度のガイドラインを示し、その構造記述に従って設計された計算機能の仮説をデータとして蓄積・共有するための方法論を提示しました。この方法論によってBRA形式のデータを蓄積・共有が進展すれば、脳型ソフトウェアの開発と活用が促進されると期待できます。

### 4. 発表内容：

#### 【研究の背景】

人工知能の分野では、その黎明期から、ヒトのような汎用性を備えた知能の実現を夢見てきましたが、未だにその目標は達成されていません。深層学習が発展した2010年代は、主に機械学習装置を組み合わせることに期待が高まりました。しかし、多様な計算装置を組み上げる設計空間は膨大であり、試行錯誤を通じての構築は容易ではありません。2014年から「脳全体のアーキテクチャに学び人間のような汎用人工知能を創る(工学)」ことで設計空間を制約する全脳アーキテクチャ・アプローチが、特定非営利活動法人全脳アーキテクチャ・イニシアティブによって推進されてきました。しかし当初は、現状の神経科学の知見に基づいて、脳のようなソフトウェアを構築する方法論は明らかではありませんでした。しかし、数年を経て3つの主要な課題が明らかとなりました。1つ目は、脳科学とソフトウェア開発の両方に精

通した人材が少なく、その育成が困難であること。2つ目は、脳全体をカバーする神経科学知見は膨大であり、個人の認知能力に頼って脳全体の機能を統合したソフトウェアを設計することには無理があること。3つ目には、脳の認知機能をソフトウェアに適切に反映させるためには、参照する脳の記述粒度を適切に選択する必要があるということです。

### 【研究内容】

本研究では、以上の課題を踏まえ、脳型ソフトウェア開発を、脳参照アーキテクチャ（BRA、注1）を共同で描く設計作業と、BRAに基づいてソフトウェアの実装と統合を行う開発作業に分けることによるBRA駆動開発方法論を提案しました。ここで、BRAとは、脳型ソフトウェアの仕様情報として標準化されたデータ形式です。主に、メゾスコピックレベルの解剖学的構造を記述した脳情報フロー（BIF、注2）形式のデータと、それに付随する1つまたは複数の仮説的コンポーネント図（HCD、注4）で構成されています（図2参照）。この方法論では、まず神経科学の論文やデータを収集・整理してBIFデータを作成します。BIFは、図3に示されるように、脳内における様々な粒度の「サーキット」をノードとし、それらの間の軸素投射にあたる「コネクション」をリンクとする有向グラフです。なお、コネクションは終点として任意のサーキットを選べますが、始点としてはサーキットの最小記述単位であるユニフォームサーキットのみを選べるものとしています。設計を通じて定義されるユニフォームサーキットは、原則としてある脳領域内の特定細胞タイプで構成されるニューロン群です。また、BIF内の任意のサーキットにおける神経活動がNBP（注3）として記述されます。BIFに割り当てられた計算機能は、仮説であるため、仮説的コンポーネント図（HCD、注4）と呼ばれます。そのため、いずれのサーキットに対しても複数のHCDを割り当てうるものです。現在の神経科学分野では、解剖学的構造に関する知識は比較的網羅的に蓄積されつつあり、これを利用してHCDを設計するSCID法（注5）は、脳の比較的広い領域に適用できるという利点があります。

必ずしも脳を深く理解していない開発者であっても、HCDを仕様として参照することで、脳型ソフトウェアの実装に携われるようになります。こうして作成されたソフトウェアの生物学的妥当性は、その構造と動作がBRAと一致しているかどうかによって評価されます。さらにBRA自体の信憑性は、神経科学の知見の裏付けによって評価されます。

### 【社会的意義・今後の予定】

今回の成果として、脳型ソフトウェアの設計空間を効果的に制約させる方法論を提示できました。この方法論に賛同する協力者が増えれば、BRA形式のデータの蓄積と共有が速やかに進むでしょう。蓄積されたデータは、脳の計算論的理解、認知モデルの開発、脳を利用した人工知能やヒューマノイドロボットの開発、精神医学やヒューマン・エージェント・インタラクション等への応用などに波及することが期待されます。

全脳アーキテクチャ・アプローチの目標である汎用人工知能の実現に関しては、BRA駆動開発をベースにしたロードマップが提案されました（図4参照）。この目標の達成には、脳のほぼ全体をカバーするシステムを構築する必要があります。ロードマップの、初期段階では、脳全体のBIFとそれに対応するHCDを構築し、その後、脳全体を統合した初期ソフトウェアを実装することが中間目標となります。その後の開発工程においては、異なるプロジェクトで設計・実装されたソフトウェアモジュールを統合しつつ、チューニングを行う作業が中心となってゆきます。このとき、ソフトウェアモジュールがBIFに準拠して作成されていれば、統合すべき対象がお互いに明確であるため、統合作業の効率化が期待できます。

## 5. 発表雑誌：

雑誌名： Neural Networks オンライン版（2021年9月14日）

論文タイトル： The whole brain architecture approach: Accelerating the development of artificial general intelligence by referring to the brain

著者： Hiroshi Yamakawa (東京大学)

DOI 番号： 10.1016/j.neunet.2021.09.004

アブストラクト URL： <https://www.sciencedirect.com/science/article/pii/S0893608021003543>

## 6. 問い合わせ先：

東京大学大学院医学系研究科 細胞分子生理学 新学術領域研究「脳情報動態」事務局付き

山川 宏 (やまかわ ひろし) 宛て

Tel: 03-5841-3470 (内線：23470)

Fax: 03-5841-3471,

E-mail: [brainfodyn@m.u-tokyo.ac.jp](mailto:brainfodyn@m.u-tokyo.ac.jp)

## 7. 用語解説：

注1) 脳参照アーキテクチャ：脳型ソフトウェアの外部設計仕様書。構成要素として BIF、NBP、HCD を含む。

注2) 脳情報フロー：脳神経回路のメゾスコピックレベルの解剖学的構造の情報フロー。Brain Information Flow (BIF) と呼ばれる。

注3) ニューラル・ビヘイビア&プロセス：脳神経回路上における局所的な神経活動の振る舞いやプロセス。記述は自然言語でなされる場合と、データとして示される場合がある。Neural Behavior and Process (NBP) と呼ばれる。

注4) 仮説的コンポーネント図：ROI が担う最上位機能を解剖学的構造と一致するように機能を体系化した仮説の図。Hypothetical Component Diagram (HCD) と呼ばれる。

注5) SCID 法：特定の神経回路(ROI) 内の解剖学的構造に整合するように、その ROI が担う最上位機能を達成できる機能体系を仮説として構築するソフトウェア設計手法である。設計においては適宜に神経活動現象も参照する。Structure-Constrained Interface Decomposition method の略称である。

8. 添付資料：

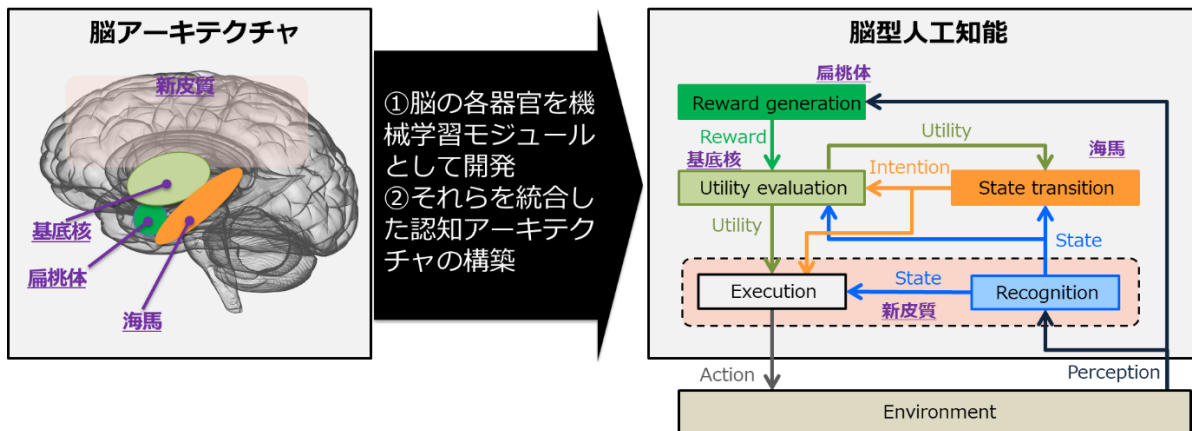


図1：全脳アーキテクチャ・アプローチの基本概念図：

2014年から徐々に形成されてきたWBAアプローチの基本コンセプトを簡略化した図を改訂したものです。図の左側には、大脳新皮質、大脳基底核、海馬、扁桃体など、脳内の大きな器官の主な例が示されています。また、これらの器官の間の結合によって、さらなる脳の構造が形成されています(図示せず)。図の右側には、機械学習を利用するものを含む計算モジュールが、脳の構造を参考にして配置・接続されています。このようにして、身体を介して環境と相互作用しながら動作するAIソフトウェアシステムを構築していきます。

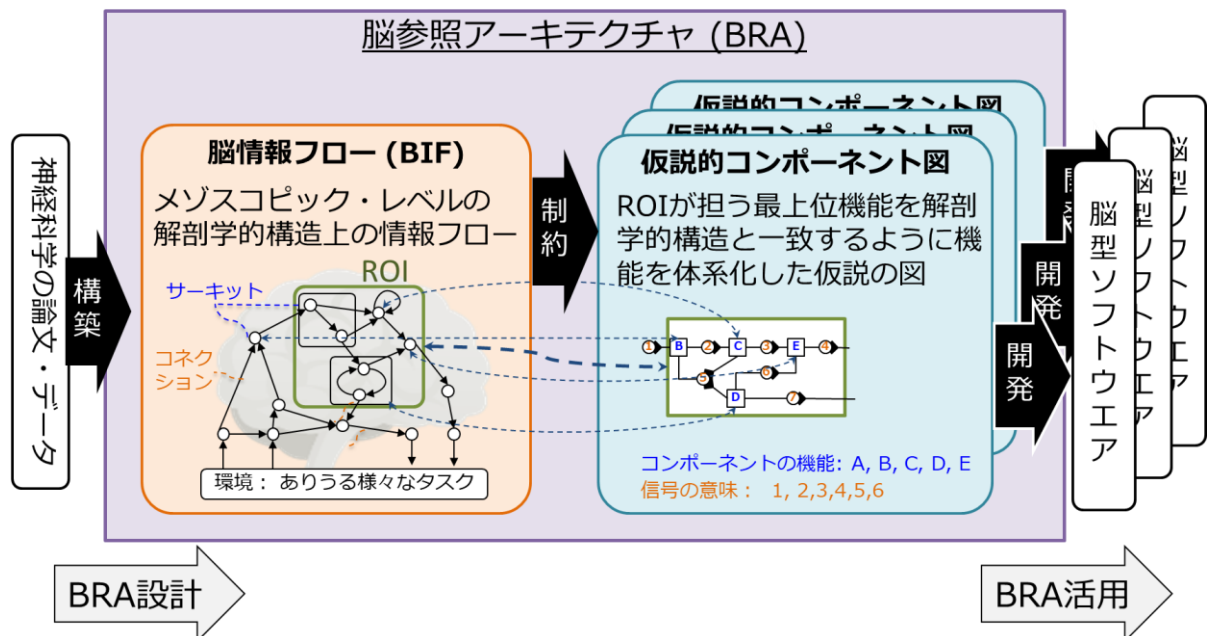


図2：BRAは、BIFとHCDで構成されています。

BIFは、解剖学的構造におけるメソスコピックレベルの情報の流れを提供するものです。HCDは、ある回路の解剖学的構造と一致するように機能を整理した図である。BIFでは、1つの回路に対して複数のHCDを使用することができます。

す。すべてのソフトウェア開発プロジェクトは、基本的に特定のHCDに基づいている。

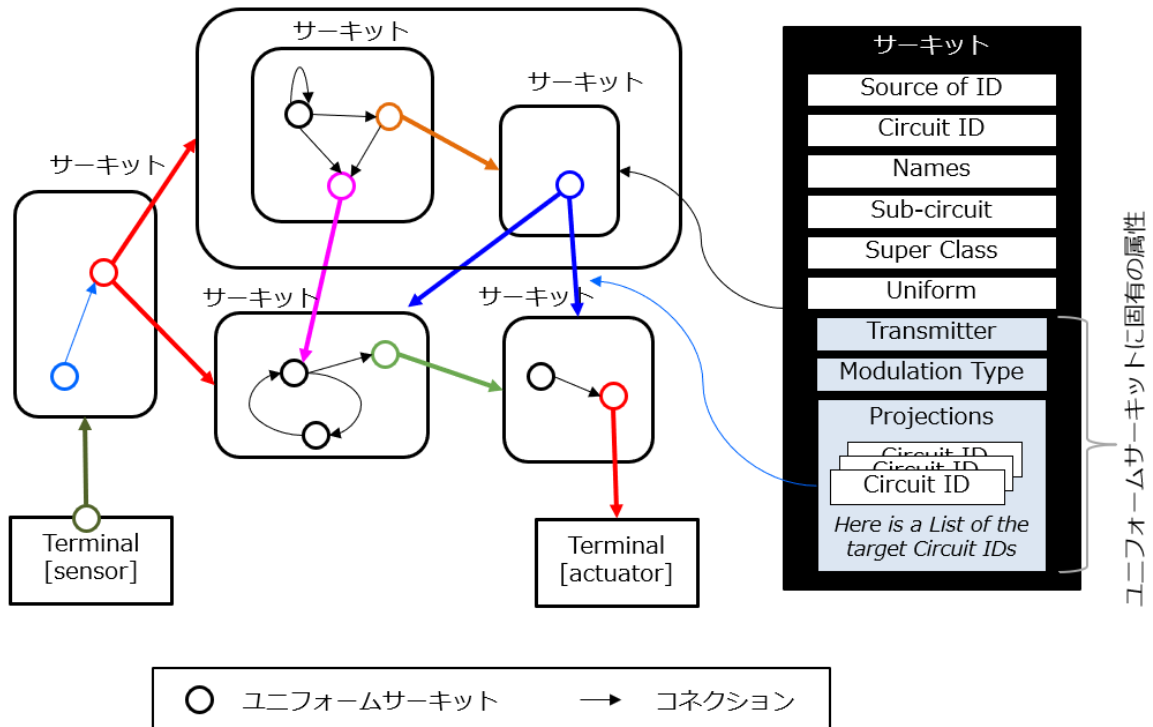


図3 脳情報フロー (Brain Information Flow: BIF)形式

BIFは、脳のメゾスコピックレベルの解剖学的構造における情報の流れを記述するものです。左図：BIFは、回路とその回路間の接続からなるグラフです。回路とは、脳内の任意の器官や領域と、それらが接続されたグラフである。接続の起点となるのは、機能的に同じ種類の意味をコード化すると考えられるユニフォームサーキットである。右図：BIFデータの各回路を記述する属性。同様回路に特有の属性がいくつかあるが、代表的なものはProjectionsプロパティで、投影対象のCircuit IDのリストである。

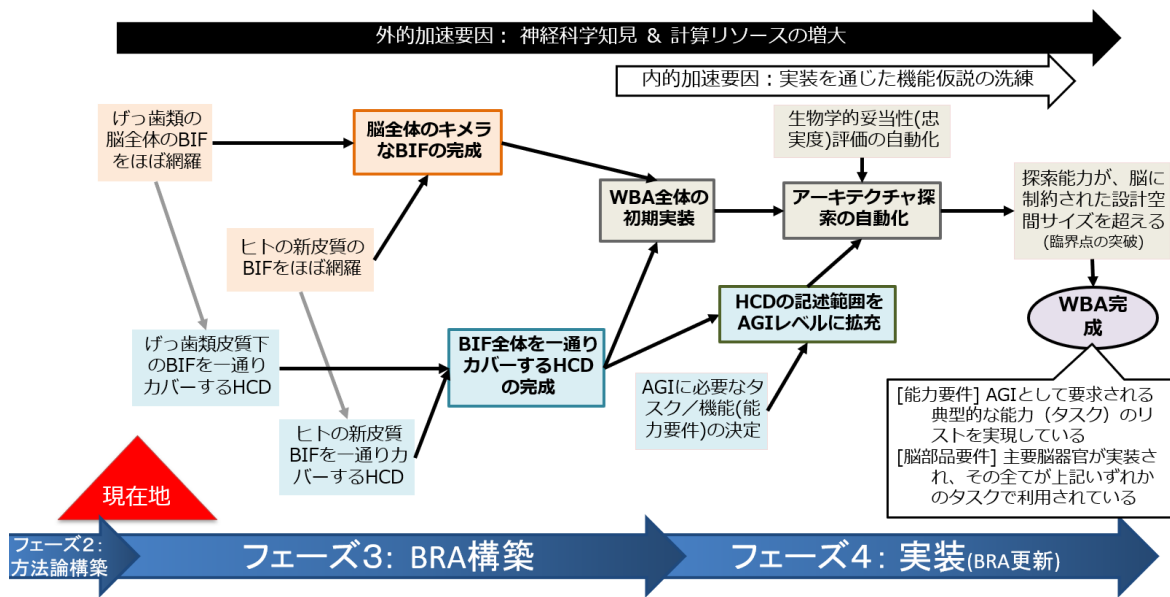


図4： 全脳アーキテクチャ開発のロードマップ

BRA 駆動開発に基づいて検討された、全脳アーキテクチャの完成に至るまでのロードマップ。BRA 構築段階では脳全体の BIF と HCD を一通り作成して、初期的に動作する WBA システムを実装する。実装段階では AGI として妥当な能力を実現できるように評価指標を改善しながら、それを実現するアーキテクチャを見つけ出しながら開発を行う。