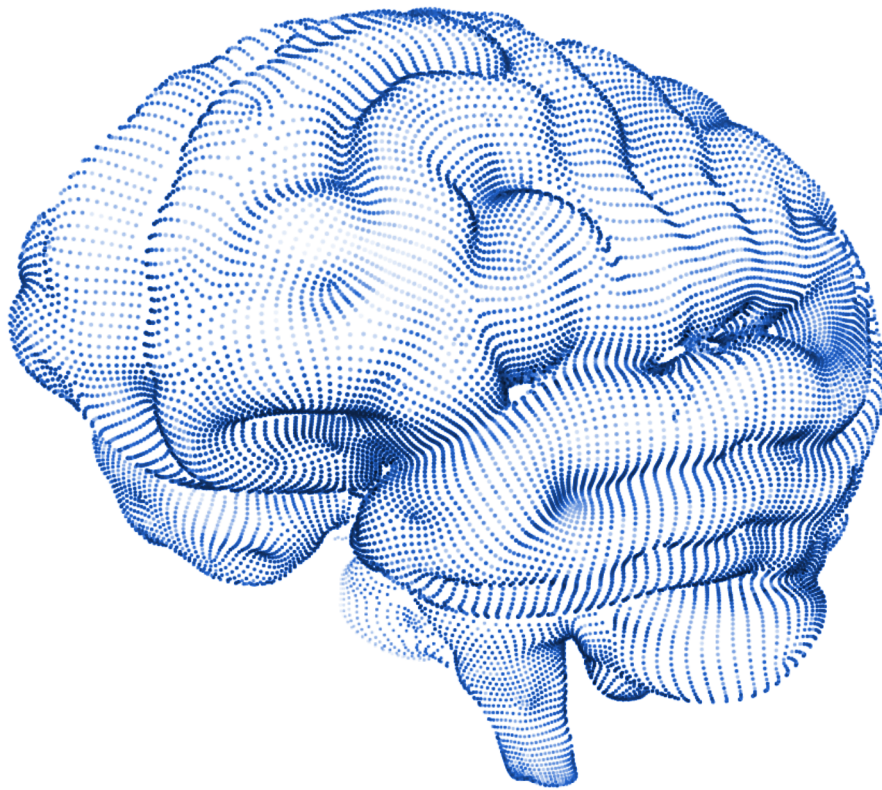


ANNUAL REPORT 2023



NPO法人全脳アーキテクチャ・イニシアティブ

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

Twitter: @wba_meetings, @wba_initiative

Facebook: <https://www.facebook.com/WBA.Initiative/>

基本理念

ビジョン： 人類と調和した人工知能のある世界

ミッション：

全脳アーキテクチャのオープンな開発を促進

価値観：

まなぶ： 関連する専門知識を学び、広める

みわたす： 広く対話を通じて見識を高める

つくる： 共に作り上げる

目次

発刊にあたり	2
2023年度を支えてくださった皆様.....	3
2023年度の活動方針と予算.....	4
2023年度の活動実績.....	5
財務状況.....	9
2024年度の活動に向けて.....	10
私とWBAI.....	14
おわりに	15
発表論文など	15
別表1 貸借対照表	17
別表2 活動計算書	18

発刊にあたり

現在の高度な人工知能（AI）技術は生成AIや基盤モデルとも呼ばれるトランスフォーマーや拡散モデルなどのニューラルネットワークに基づくものを中心となっています。これらは少なくともアーキテクチャのレベルでは脳に忠実ではないため、非脳型AIと呼ぶのが適切です。非脳型AI技術が将来的にAGI（汎用人工知能）に到達するかは未だ不確かです。しかし、2023年にChatGPTを含む大規模言語モデルの普及が目覚ましかったことで、非脳型AIによるAGI実現の可能性が世界的に注目されるようになり、2027年から2030年ごろまでに実現される可能性も無視できなくなりつつあります。



一方、私たちは以前からヒト脳型AGIを構築するために脳のアーキテクチャを参照する全脳アーキテクチャアプローチを提唱してきました。そのアプローチには以下の2つのメリット¹がありました。

- 開発の加速：脳の構造と機能を参照することでAGIモデルの設計空間を限定し、開発を加速する利点があります。これは人間の脳をベースとしたAGIの開発を直接的に促進し、結果としてAGI全体の発展を加速します。
- 対人親和性：内部構造が人間に近いAGIをモデル化することで、本質的な解釈可能性が向上し、人間が理解しやすく信頼できるAGIを構築できるというメリットがあります。これは精神疾患のモデル化やマインドアップロードの基盤としての利用など、医療への応用にもつながります。

昨今の技術状況の変化は上記の2つのメリットにおいても影響を与えています。

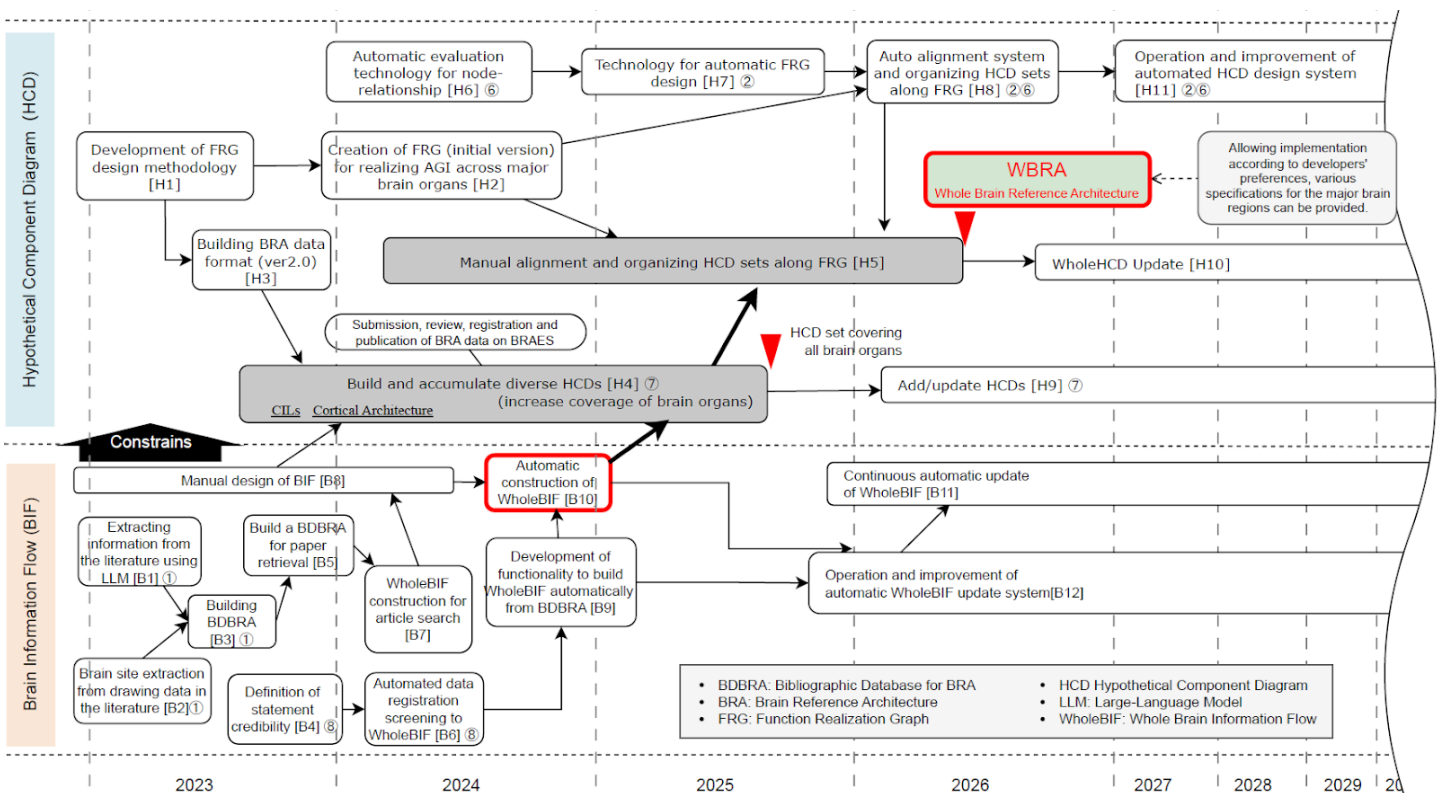


図1: 全脳アーキテクチャ技術ロードマップ

図中でWBRA (Whole Brain Reference Architecture) がヒト脳型AGIの仕様に対応する。

¹ 山川宏、脳型でAGIを開発する意義、第4回全脳アーキテクチャ・シンポジウム、2019年6月26日。

開発の加速に関しては、最新の神経科学の知見を反映した「ヒト脳型AGIの仕様を半自動で生成できるシステム」を2027年にリリースすることを目指す開発ロードマップ（図1）を発表しました [1]。上記のシステムの実現性は LLM技術の急速な進歩により開発プロセスの多くを自動化できるようになるという前提によって可能となりつつあります。また、ある時点での仕様の作成を目標とするのではなく、仕様を生成するシステムを構築することで予測が困難なAGIの完成時期の変動に対処してより優れた仕様を提供することが可能になります。

他方で対人親和性に関しては、AIの急速な進歩がもたらすリスクの点からの位置づけが重要性を増しています。ある時点でAGIが出現すればそれはAI自体を再帰的に自己改善できるため、数ヶ月程度で全人類を超える超知能に到達するという可能性があります。この超知能が人類に破滅的な影響を与えることが懸念されています。このため特にAIが人々の意図や価値観に沿って振る舞うように調整することを目指す「AIアライメント」の実現に向けた研究も盛んになってきています。

こうした背景においてヒト脳型AGIは、人間との親和性を持ち、脳に基づく解釈可能性を備えています。この特性により、ヒト脳型AGIは人間にとって理解しやすく、信頼できるAGIとなります。さらに、ヒト脳型AGIは超知能との高速な対話を可能にすることで、人類と超知能の間の橋渡し役を担うことができます。特に脳に基づく解釈可能性とは、AGIの計算プロセス（意識的および無意識的な側面を含む）を、対応する人間の神経活動の観点から認知的なレベルで理解できる特性を指します（図2）。これにより、AIの意思決定や行動の根底にある内部状態（信念、意図、喜び、苦しみなど）を明示することが期待できます。脳ベースの解釈可能性を持つAGIシステムが不都合な意図や欺瞞を形成した場合、人類はそれを発見し対処する可能性を高めます。

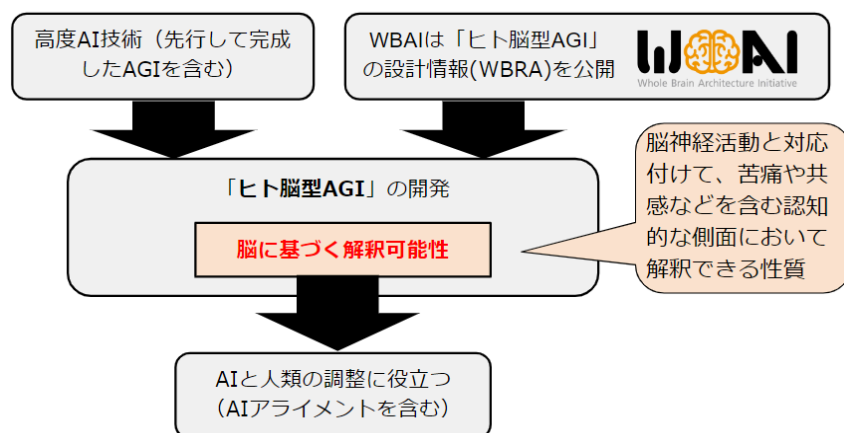


図2: 脳に基づく解釈可能性

AGIがどのような形で完成したとしても、そのAGIを活用することで、脳型AGIの構築もより容易になると考えられます。脳型AGIは、人間との親和性や解釈可能性といった独自のメリットを持つため、AGI全体の発展において重要な位置を占めるでしょう。したがって、AGIの実現を踏まえつつ脳型AGIの構築に必要な技術的条件を整備しておくことは人類にとって有益な選択肢を提供することになります。つまりWBAIはヒト脳型AGIに固有のメリットを通じて社会に価値を提供することができます。AI技術の急速な進歩が人類の生存を含むリスクを増大させている現状において、当NPOは解釈可能性の高いヒト脳型AGIのための仕様を半自動的に作成する技術を開発し、そのリスクを低減することに貢献することを目指します。

2023年度を支えてくださった皆様

会員は正会員（社員）と賛助会員からなります。企業をはじめとする賛助会員の方に財政的なご支援をいただきました。

賛助会員一覧

※各カテゴリー内の順番は入会順です。

賛助会員

後藤健太郎様

株式会社IGPIビジネスアナリティクス&
インテリジェンス

上林厚志様

株式会社TOPWELL

株式会社三菱総合研究所

パナソニックコネクタ株式会社

パナソニックホールディングス株式会社

顧問

- 銅谷 賢治（沖縄科学技術大学院大学神経計算ユニット教授）
- 北野 宏明（特定非営利活動法人システム・バイオロジー研究機構代表）
- 富田 勝（慶應義塾大学環境情報学部教授）
- 森川 博之（東京大学大学院工学系研究科教授）
- 中島 秀之（札幌市立大学学長）
- 岡ノ谷一夫（帝京大学教授）

役員

- 代 表（理事）：山川 宏
- 副代表（理事）：松尾 豊（東京大学大学院工学系研究科教授）
- 副代表（理事）：高橋 恒一（理化学研究所生命システム研究センターチームリーダー）
- 理 事：荒川 直哉（事務局長兼務）
- 監 事：浅川伸一、上林 厚志

正会員

- 正会員（社員） 14名
（2024年3月末時点）

2023年度の活動方針と予算

当法人では、脳型AGIの開発を促進するという目的に即して教育事業と研究開発事業をすすめています。当初の試みを通じて、直接的に脳型AGIの開発を行いうる神経科学と情報技術の両面に精通した人材育成が必ずしも容易でないことが明らかになってきています。そこで2017年度以降、神経科学分野の知識を開発要求仕様書である脳参照アーキテクチャ（Brain Reference Architecture: BRA）²として整理し、AI/機械学習専門家に提供するという開発方法の構築を進めています。以下、事業分野ごとに方針を記します。

教育事業

教育事業は、脳型のAGIの研究開発に必要な人工知能、神経科学、認知科学、機械学習などの異なる複数の専門性を同時に備えた学際的な人材を育成する事業です。2023年度も前年度の活動を継続し、勉強会の実施、シンポジウムの実施、外部学術イベントへの参加・協力、外部学術団体との協力・情報交換などを行うことにしました。例年通りWBAI奨励賞を授与し、脳型AGIの（国内外の）技術開発の促進において、波及効果の高い開発成果を残した者を評価することで、コミュニティの活性化をはかることにしました。

研究開発事業

研究開発事業の目標は、全脳アーキテクチャ・アプローチによる研究を支援することです。2023年度は、全脳アーキテクチャ関連の研究開発を行う機関や研究者による取り組みを設計、実装、評価、運用、方法論などのカテゴリに分けて計画を立て、支援をすることにしました（詳細は[2023年度の活動方針](#)を御覧ください）。

AGIをBeneficialなものとするための活動

当法人が開発を促進しているAGIが人類にもたらすインパクトに配慮して、2023年度もAGIをBeneficialで安全なものとし、かつ民主化するための配慮を行っていくことにしました。

2023年度の予算

予定収入は約89万円でした（前期繰越金約831万円を合計すると約920万円となります）。支出では、管理費に約81万円、イベント開催費用、研究開発費を含む事業費に約108万円、計約189万円を予定しました（当期予定収入との関係では100万円の赤字となります）。実績については以下の「2023年度の活動実績」と「財務状況」を参照ください。

本年度（2023年度）の活動実績

2023年度の活動方針に沿った教育事業、研究開発事業等を実施しました。

² 脳参照アーキテクチャ（BRA）は、脳型ソフトウェア開発のために機能仮説と現象を付与したメゾスコピックな神経回路に基づく[参照モデル](#)です。

教育事業

教育事業の目標は、長期的に全脳アーキテクチャアプローチによる研究開発を行える人材を増加させることです。2023年度は、第8回全脳アーキテクチャ・シンポジウムと2回の全脳アーキテクチャ勉強会、および1回のWBAレクチャーを開催しました。

全脳アーキテクチャ・シンポジウム

2023年8月1日

テーマ：「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：WBAIの活動についての報告

第2部：表彰

第3部：田口茂氏（北海道大学）および山田胡瓜氏（漫画家）による講演と、山川宏（WBAI）を交えた討論

会場：Zoom Meeting（参加者167名）（無料）

主催：NPO法人 全脳アーキテクチャ・イニシアティブ

後援：文部科学省学術変革領域研究行動変容を創発する脳ダイナミクスの解読と操作が拓く多元生物学

なお、本シンポジウムにおいて、孫昱偉（Sūn, Yùwěi）氏にWBAI奨励賞、全脳アーキテクチャ勉強会

実行委員会の統括者としてのシンポジウム・勉強会などのイベントの運営の円滑な実施において多大な貢献をいただいた孫曉白氏、および当法人の監事として保有財産及び理事の業務執行状況の監査に携わり、多大な貢献をいただいた上林厚志氏に[活動功労賞](#)を授与しました。

全脳アーキテクチャ勉強会

全脳アーキテクチャ勉強会は当法人の創設以前から行われてきた活動です。2023年はオンラインの勉強会を2回開催しました（以下敬称略）。

- 第38回：2023年5月11日「神経科学におけるChatGPT等の活用」
小島 武（東京大学）、芦原佑太（WBAI）、中江 健（生命創成探究センター）による講演と討論（寺田晋一郎（東京大学）司会）
遠隔開催：Zoom Meeting（参加者119名）（一般参加費1,000円・学生無料）
- 第39回：2023年11月28日「記憶の自己構築性から脳と社会とAIの「知」を考える」
倉重 宏樹（東海大学）による講演と倉重 宏樹、鈴木雅大（東京大学）、山川宏（WBAI）、田和辻可昌（東京大学）による討論
遠隔開催：Zoom Meeting（参加者約110名）（一般参加費1,000円・学生無料）

WBAレクチャー

- 第5回：2024年3月28日「全脳計算モデルの実現に向けた認知モデルの役割」
森田 純哉（静岡大学）、田和辻 可昌（東京大学、WBAI）、谷口 彰（立命館大学）による講演と討論
遠隔開催：Zoom Meeting（参加者約160名）（参加費無料）

電子情報通信学会ニューロコンピューティング研究会において、11月に「脳アーキテクチャ」特集を実施することで国内の計算論的神経科学関連領域でのBRA駆動開発の促進を行いました。

研究開発促進事業の実績

WBAIでは2018年より脳参照アーキテクチャ（BRA）駆動開発による脳型AGIの研究開発を促進しています。BRA駆動開発は図3に示すようにヒトの認知行動との関係が明らかな脳のメソスコピック・レベルの解剖学的構造を基盤とした脳情報フロー（Brain Information Flow: BIF）と、それに整合的な計算機能の記述である仮説的コンポーネント図（Hypothetical Component Diagram: HCD）を脳型ソフトウェアの仕様として用いる脳型ソフトウェアの開発方法論です。

現在、本事業においてBRA駆動開発の設計に関しては BIF作成、BRA-DB、BRA自動化、HCD設計、HCD統合を行っており、実装に関するものとしては、脳型検証、新機能、環境整備を行っています。また方法論としてのプロセス確立も進めています。以下、それぞれについて説明します。なお、学術変革領域研究「行動変容を創発する脳ダイナミクスの解釈と操作が拓く多元生物学」から財政的な支援をいただいた事業については（行動変容）と記しました。また、東京大学ムハンマド・ビン・サルマン未来科学技術センター（MbSC2030）の支援を頂いた事業には（MbSC）と記しました。

[方法論：プロセス確立]

BRA駆動開発に関連する方法論や業務フ

ロー、HCDに基づく実装プロセスの確立に着手

しました。この取り組みを通じて全脳アーキテクチャの完成に向けた技術ロードマップを構築し、その進行状況を文書化しました。特に、BRA駆動開発を用いて脳の参照アーキテクチャに基づくアプローチにより全脳アーキテクチャの完成を目指す方法論を提案し、これらの成果を共有しました [1][2]（行動変容）。

さらに、幅広い脳領域のリバースエンジニアリングに適用可能なSCID方法の拡張に取り組み、この研究を公表しました [3]。また、人間レベルのAIの実現を評価するための検討として、一般情報処理能力と Entification 能力を組み合わせた技術地図を提案しました [4]。

[設計：BIF構築]

脳のメソスコピック・レベルの解剖学的構造を基盤とした脳情報フローであるBIFの作成およびそれに関連する作業を進めてきました。このプロジェクトでは、ヒトの新皮質における前向きおよび後ろ向きの階層構造を推定するための研究を行い、その成果を公表しました [5]（行動変容）。また、BRA駆動開発を支援する

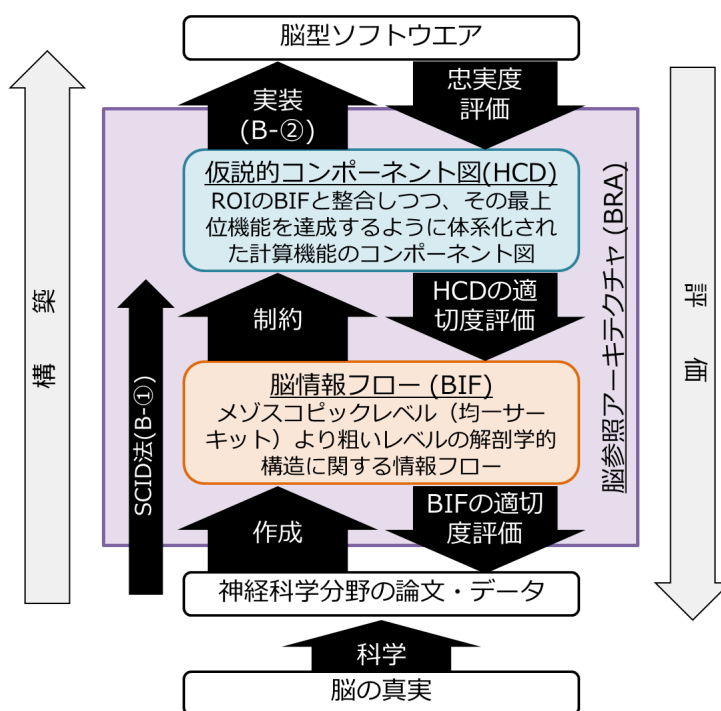


図3

ため、LLMを使用して神経投影情報を自動的に抽出し、データベースに組み込む方法についての研究を行い、この成果を共有しました [6] (行動変容)。

さらに、神経科学の文献から解剖学的な神経投射を網羅的に抽出するためのツール環境の構築に取り組み、報告しました [7] (行動変容)。この研究は脳型AIの基盤となる画像から解剖学的な神経投射を抽出する作業をさらに進めることを目的としており、その進展を発表しました [8] (行動変容)。

[設計：BRA-DB]

BRA (脳参照アーキテクチャ) のためのデータベースの構築、運用、およびマニュアル整備を行ってきました。このプロジェクトでは、神経科学の文献から神経投射と脳領域の記述を抽出するためのデータベース (BDBRA) の開発に成功し、その概要を発表しました [9] (MbSC)。このデータベースは対話プロンプトを使用して情報を抽出し、神経科学の知識を脳型AI開発に活用することを目的としています。

さらに、神経科学の文献から解剖学的構造を抽出し、その情報の信頼性を評価する手法と、その結果を蓄積するデータベースの提案にも取り組み、成果を共有しました [10] (MbSC)。データベースの構築と運用を通じて神経科学文献からの情報抽出を自動化し、研究者が容易にアクセスできる形で提供することに成功し、脳型AI研究における新たな進展を促進するための重要なインフラストラクチャを提供します。

[設計：HCD設計]

さまざまな脳器官を中心にHCD (Hypothetical Component Diagram) データの作成に着手しました。この取り組みの一環として、新皮質コネクトームに従って正準新皮質ループ回路を連結させ、ヒトの脳の高次機能を実装する方法に関する研究を行い、その可能性を探求しました。この研究成果を発表しました [11] (行動変容)。

さらに、神経科学の文献から機能不全な動物に関する情報抽出技術の調査を行い、その成果を共有しました [12][13] (行動変容)。また、大規模言語モデルを活用して情感豊かなインタラクションを実現するためのロードマップを提案し、扁桃体による恐怖や動機付けのメカニズムを中心に検討しました [14]。

脳領域をシミュレートする大規模言語モデルの連結による認知アーキテクチャの構築にも取り組みました。このプロジェクトでは、脳領域レベルで解釈・介入可能な会話AIのプロトタイピングに成功しました [15]。また、文脈依存の恐怖消失と再発に関する生物学的に妥当な扁桃体回路モデルの開発にも着手しました [16] (MbSC)。

他方で、これまでは、BIFに整合した計算機能は、階層化されたHCDによって記述されていました。しかしその記述では柔軟性が低く、しかもリバースエンジニアリングには不適切でした。そこで、要素的なサーキットである Uniform Circuit をコンポーネントした HCD の上に、新たに機能階層図を記述するための機能実現グラフ (Function Realization Graph: FRG) の検討を進め、NC研究会で発表しました [17]。

[実装：アーキテクチャ実装]

ヒトやその他の動物が解決可能な特定タスクを、神経活動との相関を含めて、脳と同様の計算機構で実行可能であるかどうかを検証するための研究開発に取り組みました。このプロジェクトの一環として、海馬体に着想を得た確率的生成モデルの実装とその有効性の検証に成功しました。この成果は神経科学の理論と計算モデルを融合させた新しいアプローチを示しています [18] (行動変容)。

この実装研究は、特定の脳領域から学んだ原理を用いて、複雑な認知タスクを解決するアルゴリズムを開発することを目指しています。私たちのアプローチでは脳の構造と機能を模倣することにより人工知能技術

の進展に寄与することを目標としています。この研究は脳型アーキテクチャの実装における重要なステップを表し、将来的には人間のような柔軟な思考や学習能力を持つAIシステムの開発に繋がる可能性があります。

[運用：BRAES]

BRA (Brain Reference Architecture) データベース、投稿審査フロー、情報公開ポータル、自動審査ツールの構築、運用、マニュアル整備、およびデータのバージョン管理を含む BRAES (BRA Editorial System) の開発に取り組みました (MbSC)。このシステムは神経機能性に関する仮説を蓄積し、共有するためのプラットフォームを提供することを目的としています。この開発活動を通じて、神経科学の知見と人工知能技術を統合し、研究コミュニティに貴重なリソースを提供します。

具体的には、BRAESの開発により研究者が神経機能に関する仮説を効率的に投稿し、審査し、公開するプロセスを最適化しました。このシステムは神経科学研究の進展を加速させ、脳の理解を深めるための新しい道を開くことを目指しています [19]。

[その他：AIアライメント]

脳型AGIを含むAGIが人類社会に適合するようにするための研究開発に取り組みました。このプロジェクトでは、AGIの行動や決定過程を人間が理解しやすい形で解釈可能にすることを目指し、脳に基づく解釈可能性に関するアプローチを探求しました [20]。この取り組みは AGI技術が社会に受け入れられ、信頼されるためには、その行動が人間にとって理解しやすく予測可能である必要があるという考えに基づいています。私たちの研究は、脳の機能や構造を模倣することで AGIの意思決定プロセスをより透明にし、その結果を人間が解釈しやすくする方法を模索しています。このアプローチにより AIのアライメントという技術的にも倫理的にも重要な課題に対する新たな解決策を提案しています。

Openな研究開発コミュニティの形成

Slack (開発者用SNS) 上のコミュニティでWBAIに関する情報交換を行っており、半定期的にオンラインミーティングも行っています。また立命館大学とは、上記のアーキテクチャ実装で述べたように、海馬体に着想を得た確率的生成モデルの実装とその有効性の検証を行っています。

活動とボランティア

2023年の活動は基本的に無報酬の正会員とボランティアによって行われました (事務局の業務委託職員を除く)。とりわけイベントの開催はボランティア主導で行われました。

財務状況

2023年度の貸借対照表 (別表1) と活動計算書 (別表2) を文書末に示します。

決算収入 (当期) は約84万円。支出は管理部門で約80万円、事業部門で約99万円、計約178万円でした。このため約94万円の赤字となりました。

経常収益のほとんどは賛助会員年会費から来ています。雑収入は、勉強会の参加者から徴収した参加費を含みます。

事業費のうち外注費は事務局費用の50%を按分したものです。業務委託費は、研究開発促進のための調査費用を一部WBAIが負担したものです。

管理費のうち、事務局の件費は外注費の形で支出しています（事業費と50%按分）。地代家賃は、ガラム合同会社のご厚意で無料となっています。支払報酬は経理事務を一部委託している会計事務所に支払っているものです。

2024年度の活動にむけて

私たちは2024年も、BRA駆動開発を軸として脳型人工知能の研究開発を促進するために以下の教育事業と研究開発促進事業をおこなう予定です。

教育事業（人材育成事業）

本事業は、長期的に全脳アーキテクチャアプローチに基づく研究開発に必要な、人工知能、神経科学、認知科学、機械学習など異なる専門性を持つ学際的な人材を育成し、増加させることを目的としています。2024年も例年通り複数の勉強会や年次のシンポジウムの開催などによって、社会的関心の喚起も含めた教育事業を継続してゆきます。

国内外の外部学術イベントへの参加・協力、外部学術団体との協力・情報交換などを行うことを通じ、当法人の進展状況をアピールしてゆきます。また、外部の研究者・技術者およびグループとの協力により人材育成、巻き込みをはかります。

WBAI奨励賞の応募を行い、脳型AGIの（国内外の）技術開発の促進において、波及効果の高い開発成果を残した方々を評価することでコミュニティの活性化をはかります。また、当法人が開発促進するヒト脳型AGIをより人類とAIとを調和させることに役立つように促進する活動も継続する予定です。

研究開発事業

研究開発促進事業は、NPO法人である利点を活かし、AGI研究に携わる他の研究機関等との競合を避け、BRA主導の開発を推進することで、オープンプラットフォーム上での民主的に全脳アーキテクチャアプローチによるAGI研究の活性化と加速を目指します。具体的には以下の表に示した活動を中心に実施する予定です。

表：WBA研究開発カテゴリー一覧

大カテゴリ	中カテゴリ	説明(新)
方法論	BRAデータ管理システム	BRAデータ、投稿審査フロー、BRAES、審査ツール等の方法論策定
方法論	BRA可視化	BRAデータ可視化ツールの開発とレイアウト・表示項目の検討
方法論	HCD設計	BRA設計における特にHCD設計のための方法論構築
方法論	BRA全般	BRA駆動開発の方法論に関する研究
運用	BRAデータ管理システム	BRAデータ、投稿審査フロー、BRAES、審査ツール等の運用（マニュアル整備・バージョンアップ等を含む）
運用	BRA可視化	BRAデータ可視化ツールの運用とレイアウト・表示項目の最適化
運用	WholeBIF管理	WholeBIFへの追加項目の審査と登録（BDBRA含む）
運用	データアップデート	データのバージョン管理実施
設計	BRAデータ管理システム	BRAデータ、投稿審査フロー、BRAES、審査ツール等の設計
設計	BIF構築	BIFの作成と関連作業の実施
設計	HCD設計	HCDの設計と関連作業の実施
設計	WholeBIF構築	脳全体を統合するBIFデータの構築作業（BDBRA含む）

設計	WholeHCD構築	分散構築されたHCDの統合作業
評価	BIF信憑性評価	BIFの信憑性評価実施
評価	HCD評価	HCDの機能性・構造整合性評価と審査フロー・自動審査の実施
評価	忠実度評価	実装の評価（異常系含む）、活動再現性評価、機能評価の実施
評価	異常系評価	パフォーマンス変動の評価（人間もしくは動物の機能不全）
評価	活動再現性評価	ソフトウェアの振る舞いと実験で得られた神経活動の比較評価
評価	構造忠実度評価	実装コードの構造類似性評価
評価	実行環境整備	脳型ソフトのタスク実行・テスト環境の整備
実装	実装環境整備	ソフトウェアプラットフォーム（BriCAなど）の構築
実装	アーキテクチャ実装	HCDを参照したソフトウェアフレームワークへのアーキテクチャ実装（構造忠実度評価可能な形式）
実装	BRAデータ管理システム	BRAデータ、投稿審査フロー、BRAES、審査ツール等の実装
実装	コンポーネント実装	HCDのCircuit毎プロセス記述を参照したソフトウェアコンポーネント実装（活動再現性評価可能な形式）
実装	実装環境整備	ソフトウェアプラットフォーム（OpenAI Gymなど）の構築
BRA以外	新機能の検証	未知の計算機構探索のための実装（ヒトを含む動物が解決可能な特定タスク対象）
BRA以外	新機能の設計	未知の計算機構の設計（ヒトを含む動物が解決可能な特定タスク対象）
その他	人材育成	実装プロセス実行可能な技術者を育成するための教育的な実装
その他	AIアライメント	AIと人類の調和を目指す活動の実施

以下ではこれらの活動の中でも、BRA駆動開発において中心的な部分として図4（次頁）のロードマップに記されている BIF構築と HCD構築について詳しく述べます。

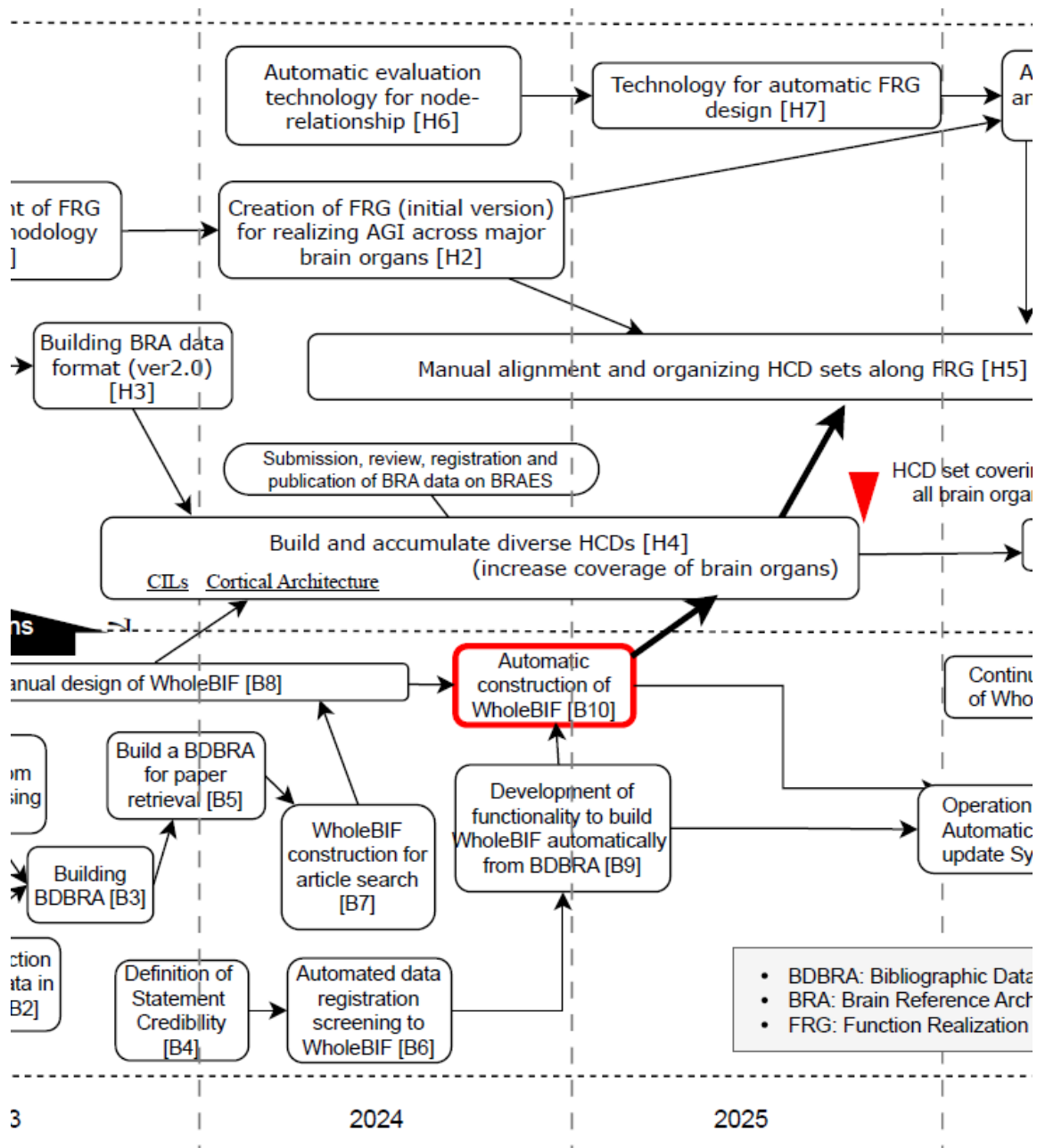


図4 WBA技術ロードマップにおける2024年の活動予定

BIF構築

BIF構築の取り組みとして、2024年末には WholeBIF を完全自動で構築する目標を掲げています。この目標に向けて私たちは以下のステップでBIF構築活動を進めていきます。

まず前半期に BDBRA を用いた論文検索システムを完成させること [B5]、そしてこのシステムを利用して論文検索に特化した WholeBIF の構築を行い完成させる [B7] ことが予定されています。さらに、文献からの情報の信頼性を定義する作業 [B4] を前半期に完了し、それを基に WholeBIF への自動データ登録のためのスクリーニングプロセスを構築します [B6]。このプロセス [B6] を用いて BDBRA から WholeBIF を自動的に構築する機能の開発を始めます [B9]。

年末には、この自動構築機能 [B9] の成果を用いて WholeBIF の自動構築に本格的に移行し [B10]、これにより2025年以降の WholeBIF の手動設計 [B8] がほぼ不要になり、プロセスの自動化と効率化を図ることができます。

HCD構築

2026年に全脳参照アーキテクチャ（Whole Brain Reference Architecture）を実現することを目標に、2024年のHCD（Hypothetical Component Diagram）構築においては以下の活動を進める予定です。

FRG（機能実現グラフ）に対応するために2023年から進行中のBRA（Brain Reference Architecture）データフォーマット（バージョン2.0）[H3]の完成を目指し、これを多様なHCDの構築と蓄積[H4]の推進に活用します。

また、主要な脳器官を横断してAGIを実現するためのFRGの初版の作成[H2]を進め、その成果をFRGに沿ったHCDセットの手動での整理と組織化[H5]の推進に利用します。

2024年中にはノード関係性の自動評価技術[H6]を完成させ、これを2025年の自動FRG設計技術[H7]の開発に繋げる予定です。

Openな研究開発コミュニティの形成

全脳アーキテクチャアプローチを実現するための研究を支援するためのソフトウェアなどの研究インフラストラクチャを整備する活動と、それらを利用して研究を進める活動を他の組織とも連携・協力しながら行っています。

脳型ソフトウェアの開発に役立つBRAを構築するためにBIF形式に従って解剖学的知見の蓄積を進めた脳情報フローの構築を進め、それを用いて特定のタスクや機能を前提とした機能仮説の付与をSCID法によって行うことを促進します。また、そうして構築された脳型ソフトウェアの実装を評価するためにBRAを活用しながら脳型AGIの神経科学的妥当性を評価する技術の開発も進めます。

特に本年はこうした開発促進の一環として全脳アーキテクチャ国際ワークショップの実施を進めます。

人類と調和したAIのある世界へと向かうための活動

WBAアプローチで構築されたヒト脳型AGIはヒトとの親和性が高いという特性を活かして、私たちは人類と調和したAIのある世界へと向かうことを促進します。

一つの重点は対人インタラクションが必要な応用分野の発展にあります。扁桃体による情動メカニズムとLLMを組み合わせた情感豊かなインタラクションの研究は既に進展しており、これを今後も推進していく予定です。

また、親和性の中でも特に「脳に基づく解釈可能性」を持つヒト脳型AGIが人類と超知能の架け橋となり、人類が超知能から受ける負の影響を軽減する可能性を探求しています。そのために必要に応じてAIアライメントに関連する組織との連携をはかりたいと思います。

2024年度の予算

予定収入は約54万円で、主に会費収入からなります（2023年度の会費収入は約70万円でした）。当期予定収入と前期繰越金約737万円を合計すると約791万円となります。

支出では、管理費に約81万円、謝金、賞金、通信費を含む事業費に約82万円、計約163万円を予定しています（当期予定収入との関係では109万円の赤字になります）。なお、2023年度の予算では管理費に約81万円、事業費に約108万円の計約189万円の支出を予定していました。

私とWBAI

竹中工務店技術研究所 上林厚志

竹中工務店技術研究所の上林です。私の主な研究テーマは鉄筋コンクリートの構造解析ですが、過去には橋梁補修に関するエキスパートシステムの開発やニューロシステムの応用、最適化システムの開発などの人工知能に関連する研究にも携わってきました。

第3次人工知能ブームの到来とともに、深層学習による画像認識の進歩や囲碁のAlphaGoの出現を目の当たりにし、建築会社においても新たな可能性を追求することを考えました。そこで、画像認識を用いた安全確認システムの開発を企画しました。ただし、建設現場は日々変化するため、固定カメラでは十分な視野を確保できませんでした。そのため、ロボットにカメラを装備し、現場を巡回することを検討しました。2016年頃、私はヒューマノイドロボットの研究者と協力し、仮設足場の沈み込みなども考慮した二足歩行プログラムの開発に取り組みましたが、その複雑さと困難さを痛感し、ロボットの動作も人工知能による学習に置き換える可能性を探るようになりました。

多機能なロボットを操作するための人工知能を探していたところ、汎用人工知能研究会（人工知能学会のSIG-AGI）に出会いました。また、目を持つロボットというアプローチも私に合っていると感じ、松尾先生の講演やNHKなどのテレビ特集、AI関連番組の影響もあり、WBAIのシンポジウムや勉強会に積極的に参加するようになりました。

SIG-AGIでは、定期的に海外の文献を紹介し合う輪読会が行われており、WBAI代表の山川先生やWBAI事務局の荒川さんなど、輪読会主催者の方々と親交を深める機会もありました。WBAIとの関係も一層深まってきました。WBAIでは、ハッカソンが開催され、脳を参考にしたシステムの開発を競い合っています。私自身も興味はありましたが、プログラムの作成やデバッグは60歳を超える私にとってはかなりのストレスとなるため、WBAI副代表の高橋先生と話し合った結果、2017年からWBAIの賛助会員となりました。

また、2020年からは2年間、WBAIの監事として責務を果たしました。WBAIの運営会にもオブザーバーとして参加し、全脳アーキテクチャを用いた人工知能開発を間近で目にすると同時に、汎用人工知能に関する情報や研究の動向について学ぶ機会も得ました。

私自身は汎用人工知能の開発を行うことはできませんが、その生まれる様子を間近で見続けられる現在の環境に非常に満足しています。ChatGPTの登場により、汎用人工知能の実現が現実味を帯びてきたと感じます（実際、この原稿の添削もChatGPTにお願いしました）。また、WBAIを用いた人工知能の開発もChatGPTを活用して大きく進展し、その実現がより近づいていると感じています。

現在まで、全脳アーキテクチャの研究を進めることは将来的に精神医学や行動変容生物学の研究にも貢献することになると考えてきました。同様に今後は現在の人工知能が汎用人工知能になるために何が 필요한のか、どういう仕組みが必要なのかを示唆できるのではと考えています。今しばらくはWBAIとともに汎用人工知能の誕生を期待に胸を膨らませながら見守りたいと思います。

おわりに

冒頭に述べたように、現在の高度なAI技術は脳を参照しない非脳型AIとして発展しています。対して私たちはヒト脳型AGIを構築するために全脳アーキテクチャアプローチを継続的に提唱しています。このアプローチには、AGI開発の加速と対人親和性の向上という2つのメリットがあります。

特に、超知能へと到達しうる高度なAIのリスクが強く懸念される現在においては、脳型AGIにおける「脳に基づく解釈可能性」が新たに重要性を持ちはじめています。このヒト脳型AGIの設計情報として、私たちはWhole Brain Reference Architecture (WBRA) の構築を目指しています。WBRAは、ヒトの脳の構造と機能を参照し、脳型AGIを実装するためのソフトウェア設計情報を提供するものです。最初のAGIがどのような形で完成したとしても、WBRAが存在すれば、それを利用することで必要に応じて脳型AGIを構築することが容易になるでしょう。

近年、AGIが2027年頃に完成するかもしれないという見方も出ています。このような技術的展望を踏まえ、WBAIとしても2027年までに「WBRAを半自動で生成できるシステム」の完成を目指すロードマップを作成しました。

現在のところ、ヒトの認知機能を十分に再現できる脳型AGIのソフトウェア設計情報の提供を目指している組織はWBAIにおいて他にはありません。この脳型AGIの実現に向けた取り組みの一環として、2023年夏には新たなミッションを「全脳アーキテクチャアプローチによる汎用人工知能のオープンな開発を促進する」と微調整し、開発促進されるAGIが脳型であることを強調しました。

AI技術の急速な進歩が人類の生存を含むリスクを増大させている現状を踏まえると、解釈可能性の高いヒト脳型AGIの構築に必要な技術的条件を整備しておく私たちWBAIの活動は、そのリスクの低減に貢献できる可能性を持っています。こうした点について皆様のご理解と継続的なご支援を心よりお願い申し上げます。

発表論文など

- [1] 山川 宏, 田和辻 可昌, 芦原佑太, 布川絢子, 荒川 直哉, 高橋恒一, 松尾豊: 脳参照アーキテクチャ駆動開発による全脳アーキテクチャ完成にむけた技術ロードマップ, 第25回汎用人工知能研究会, 2023
- [2] 山川 宏, 芦原佑太, 田和辻 可昌, 荒川 直哉: BRA駆動開発による全脳アーキテクチャ完成にむけた技術ロードマップ, *JNNS* 2023, 2023
- [3] Hiroshi Yamakawa, Ayako Fukawa, Naoya Arakawa, Yoshimasa Tawatsuji, Yuta Ashihara, Yutaka Matsuo: Proposal of a function-oriented SCID method for reverse engineering a wide range of brain regions, *Computational Neuroscience* 2023, P127, 2023
- [4] 山川、松尾: 人間レベルAIの到達点を定義するための包括的技術地図 一般情報処理能力リスト ×Entification能力リスト, *JSAI2023*, 2023
- [5] Yudai Suzuki, Yuta Ashihara, Hiroshi Yamakawa: Estimating Feedforward and Feedback Hierarchies in the Human Neocortex, 第3回「行動変容生物学」領域会議, 2023
- [6] Iriya Horiguchi, Yuta Ashihara, Hiroshi Yamakawa: Investigation of information extraction techniques for neuroscience literature on dysfunctional animals, 第3回「行動変容生物学」領域会議, 2023

- [7] 芦原、堀口、山川: 神経科学文献から解剖学的神経投射を網羅的に抽出できるツール環境構築に向けて, NC研究会, 2023
- [8] 堀口、芦原、山川: 脳型AIの基盤となる画像からの解剖学的神経投射の抽出, *JSAI2023*, 2023
- [9] Yuta Ashihara, Hiroshi Yamakawa, Iriya Horiguchi, Yutaka Matsuo: BDBRA: Database for extracting descriptions of neural projections and brain regions from neuroscience literature with dialogue prompts, *Computational Neuroscience*, 2023
- [10] Yuta Ashihara, Hiroshi Yamakawa, Iriya Horiguchi: Proposing a Database Extracting Anatomical Structures from Neuroscience Literature and Evaluating Its Information Credibility, *INCF Assembly*, 2023
- [11] 田和辻 可昌, 山川 宏, 荒川 直哉: 新皮質コネクトームにしたがって正準新皮質ループ回路を連結させてヒトの脳の高次機能を実装する: すべての(新皮質-X-新皮質)回路を総体とした正準新皮質ループ回路の可能性, *JNNS 2023*, 2023
- [12] 堀口、芦原、山川: BRA駆動開発のためのLLMを用いた神経投影情報の自動抽出とデータベース構築, *JNNS 2023*, 2023
- [13] 堀口維里優, 芦原佑太, 山川宏. 大規模言語モデルを用いた機能不全動物実験論文からの情報抽出プロンプトの考案. *IEICE Conferences Archives IEICE-123, IEICE-NC-49-IEICE-NC-52*, 2024
- [14] 大森、田和辻、宮本、芦原、荒川、山川: 大規模言語モデルによる情感豊かなインタラクション実現のためのロードマップの提案-扁桃体による恐怖や動機付けの仕組みを中心に-, *JSAI2023*, 2023
- [15] 宮本 竜也, 田和辻 可昌, 山川 宏: 脳領域をシミュレートする大規模言語モデルの連結による認知アーキテクチャの構築:脳領域レベルで解釈・介入可能な会話AIのプロトタイプング, *JNNS 2022*, 2023
- [16] Tatsuya Miyamoto, Yoshimasa, Hiroshi Yamakawa: A Biologically Plausible Amygdala Circuit Model for Context-Dependent Fear Extinction and Relapse, 第3回「行動変容生物学」領域会議, 2023
- [17] 山川宏、田和辻可昌、荒川直哉、芦原祐太、布川絢子、高橋恒一、松尾豊: 脳計算機能の効率的なリソースエンジニアリングのためのデータ記述形式. *IEICE Conferences Archives IEICE-123, IEICE-NC-53-IEICE-NC-58*, 2024
- [18] 大竹 俊輔¹、前山 功伊¹、長谷川 翔一¹、中島 毅士¹、谷口 彰¹、谷口 忠大¹、山川 宏²: 海馬体に学んだ確率的生成モデルの実装と有効性検証, *JSAI2023*, 2023
- [19] Yoshimasa Tawatsuji, Yuta Ashihara, Naoya Arakawa, Hiroshi Yamakawa: BRAES: Developing Brain Reference Architecture Editorial System for Accumulating Hypotheses on Neural Functionality, [INCF Assembly](#) (also on 第3回「行動変容生物学」領域会議), 2023
- [20] Hiroshi Yamakawa, Yoshimasa Tawatsuji, Yuta Ashihara, Ayako, Fukawa, Naoya Arakawa: Brain-based Intepretability, 第3回「行動変容生物学」領域会議, 2023

別表1：

貸借対照表

(2023年3月31日現在)

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

単位：円

科目	金額		
I 資産の部			
1. 流動資産			
現金預金	7,427,783		
未収入金	26,520		
流動資産合計		7,454,303	
資産合計			7,454,303
II 負債の部			
1. 流動負債			
前受金	80,000		
預り金	4,084		
流動負債合計		84,084	
負債合計			84,084
III 正味財産の部			
前期繰越正味財産		8,314,896	
当期正味財産増減額		▲944,677	
正味財産合計			7,370,219
負債及び正味財産合計			7,454,303

活動計算書

2023年4月1日から2024年3月31日まで
 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

単位：円

科目	金額	
I 経常収益		
1. 受取会費等		
正会員年会費	142,000	
賛助会員費	560,000	702,000
2. 受取寄付金		
受取寄付金	26,640	26,640
3. その他収益		
受取利息	71	
雑収入	111,409	111,480
経常収益計		840,120
II 経常費用		
1. 事業費		
(1) 人件費		
給料手当		
法定福利費		
人件費計	0	
(2) その他経費		
業務委託費		
外注費	529,597	
講師等謝金	90,000	
広告宣伝費		
交際費		
会議費		
旅費交通費		
通信費	267,110	
消耗品費		
新聞図書費		
支払手数料		
地代家賃		
奨励金・賞金	100,000	
寄付金		
雑費		
その他経費計	986,707	
事業費計		986,707
2. 管理費		
(1) 人件費		
給料手当		
法定福利費		
人件費計	0	
(2) その他経費		
業務委託費		
外注費	528,000	
支払報酬料	264,000	
交際費		
通信費	970	
消耗品費		
諸会費		
支払手数料	5,120	
地代家賃		
租税公課		
雑費		
その他経費計	798,090	
管理費計		798,090
経常費用計		1,784,797
当期正味財産増減額		▲944,677
前期繰越正味財産額		8,314,896
次期繰越正味財産額		7,370,219