強化学習で行動・推論・対話の制御プログラムを獲得する脳型 AGI アーキテクチャの設計

Designing a brain-like AGI architecture that acquires control programs for behavior, reasoning, and dialogue through reinforcement learning

一杉裕志 *1 Yuuji Ichisugi　高橋直人 *1 Naoto Takahashi　竹内泉 *1 Izumi Takeuti　佐野崇 *2 Takashi Sano　中田秀基 *3*1 Hidemoto Nakada

*1産業技術総合研究所 National Institute of Advanced Industrial Science and Technology　*2東洋大学 Toyo University　*3順天堂大学 Juntendo University

ヒトの脳の情報処理を模倣した脳型 AGI アーキテクチャ設計の取り組みの現状を報告する。このアーキテクチャはヒトの脳の計算論に関するいくつかの仮説を前提として設計されている。脳の前頭前野の機能に対応する行動ルール集合がアーキテクチャの中心にあり、他のコンポーネントの状態の変更や、コンポーネント間の情報の流れの制御を行う。行動ルール集合は、教師なし学習と強化学習によって自律的に獲得できるように設計されている。個々の要素技術の動作確認はすんでおり、全体の統合も可能であるという見通しを得た。

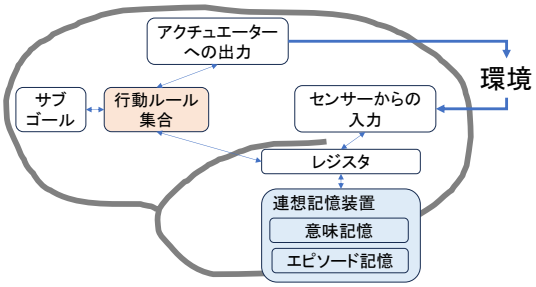

図 1: 設計中の脳型 AGI アーキテクチャの全体像。学習により獲得した行動ルール集合が全体の中心にあり、情報の流れを能動的に制御する。

1. はじめに

ヒトの脳は生存確率を高める目的で、狩猟、採集、調理、道具の作成などの物体操作、行動計画や未来予測などの推論、他者との対話などの機能を進化させた情報処理装置である。脳の情報処理原理を明らかにし、それを模倣すれば、ヒトに似た知能を持った汎用人工知能（脳型 AGI）が実現できるはずである。このような動機から、これまで我々は神経科学的知見を参考にしながら脳型 AGI アーキテクチャの設計 [一杉 21] を行ってきた。本稿ではこの取り組みの現状を報告する。

設計の中期的目標は、脳型 AGI の実現可能性を多くの人々が一目で確信できるようなデモを、小規模なもの [一杉 24b] でもよいから実現することである。アーキテクチャは、ヒトの脳の計算論に関するいくつかの仮説を前提として設計されている。それと同時に、強化学習・ベイジアンネット・数理論理学を主な理論的基盤としている。この理論的基盤により、アーキテクチャ全体の見通しがよくなり、神経科学的妥当性に縛られない効率的実装への道筋が付けられている。

2. アーキテクチャの全体像

図 1 は、設計中の脳型 AGI アーキテクチャの全体像であり、主要コンポーネントと脳の部位とのおおまかな対応も示している。**行動ルール集合**がアーキテクチャの中心にあり、他のコンポーネントの状態の変更や、コンポーネント間の情報の流れを制御する。行動ルール集合は**手続き的知識**（制御プログラム）であり、エージェントが自分自身の経験や対話・模倣を通じて、自律的に獲得できるように設計されている。脳の前頭前野が行動ルール集合を記憶すると我々は考えている。

脳の側頭葉の機能に対応する**連想記憶装置**には、**宣言的知識**（プログラムから読みだせるデータ）が記憶されている。手続き的知識と宣言的知識のみが長期記憶であり、他のコンポーネントの状態は短時間で変化し得る。

サブゴールはエージェントが自ら設定する行動の短期的目標であり、エージェントはできるだけ少ないコストでサブゴールを達成すべく、行動・推論を行う。

レジスタは小容量の記憶装置であり、エージェントによる推論結果や、センサー入力を通して環境の状態を認識した結果を記憶する。レジスタに入っている値は、いわばエージェントがその瞬間に「意識」している情報である。

エージェントは一定時間ごと（約 5Hz の θ 波のサイクルを想定）に、現在のレジスタの値とサブゴールの値にもとづいて、行動ルール集合の中から行動ルールを 1 つ選択し、そのルールが指定する行動を実行する。行動にはアクチュエーターを通して環境に働きかける対外行動と、レジスタやサブゴールの値を更新する脳内行動の 2 種類がある。推論は脳内行動の連鎖という形で実現される。

連想記憶装置は大容量の記憶装置であり、**エピソード記憶**の機構と**意味記憶**の機構の 2 つに分かれる。エピソード記憶機構には行動の 1 ステップごとに、更新されたレジスタの中身が次々と追記される。これはいわばエージェントの「意識」のログである。意味記憶は大量のエピソード記憶を圧縮・抽象化したものである。エピソード記憶は、随時意味記憶に変換されていくと同時に、適宜忘却されていく。

3. 設計の前提となる仮説

この節では、アーキテクチャ設計の前提となる、ヒトの脳の計算論に関するいくつかの仮説について説明する。

3.1 仮説：脳の目標は報酬最大化

脳の大脳基底核は強化学習に関与すると考えられている。そのことを踏まえ、我々は「脳の目標は生物が進化によって獲得した何らかの報酬関数の最大化である」という仮説を前提としてアーキテクチャを設計している。

連絡先: 一杉裕志, E-mail: y-ichisugi@aist.go.jp

生物は子孫を確実に残す方向に進化することから、脳にとっての報酬はおそらくエサの獲得、危険の回避、生殖行動、身体的・知的能力の維持、社会性に関することなどで与えられるだろう。工学的に有用な脳型 AGI 実現という観点においては、生物の報酬関数の詳細よりも、それを最大化する機構の方が重要であり、その設計に注力している。

3.2 仮説：大脳皮質は一種のベイジアンネット

大脳皮質はヒトの知能全般に最も関与する脳の部位であり、コラム構造、領野間の双方向の結合、深層学習のような階層構造といった解剖学的特徴を持つ。筆者は大脳皮質の解剖学的構造や機能的特徴を踏まえた上で、大脳皮質のマクロコラムを一種のベイジアンネットのノードと対応付ける仮説 (BESOM モデル) を提唱している [Ichisugi 07]。

この仮説では、大脳皮質は下位の領野からの入力を教師なし学習し、圧縮・抽象化した表現を獲得する。また、このベイジアンネットは記号処理と統計的機械学習を結びつける性質を持っている。ノードは他のノード間の接続のゲートを制御することが可能であり、それを利用してネットワークを設計することで、論理的推論や組み合わせ探索問題を解くことができる [一杉 16]。この機構により、自然言語の構文解析・意味解析の機能も実現できる見込みが立っている [一杉 18a][Ichisugi 19a]。

設計中のアーキテクチャにおける、パターンマッチを用いた行動選択 [一杉 20b] や、レジスタの値の保持・更新 [一杉 20c] などの主要な機能も、ベイジアンネット上の推論として実現可能になるように設計されている。また、行動ルール集合の獲得はベイジアンネットのパラメタ学習により行われ [一杉 18b][一杉 20b]、その後、強化学習によって各行動ルールの価値が学習される [一杉 20a]。

3.3 仮説：前頭前野は行動価値関数を表現

脳の前頭前野は、思考や行動を制御する脳の最高中枢である。前頭前野の解剖学的構造は、視覚野など他の大脳皮質の領野と基本的には同じである。前頭前野は大脳基底核との間のループ構造を持つことから、強化学習に関与していると思われる。

そこで我々は「前頭前野は強化学習における行動価値関数を圧縮・抽象化して表現している」という仮説に基づいてアーキテクチャを設計している。前頭前野も大脳皮質であるから、前節で述べた仮説に従うなら、下位の領野からの情報を圧縮・抽象化しているはずである。前頭前野が情報を受け取る領野には感覚連合野と運動野が含まれていることから、前頭前野が行動価値関数を表現するとするのは極めて自然な仮説であると考える。

3.4 仮説：前頭前野は再帰的強化学習を実行

ヒトは何か目的を達成するために、必要に応じて再帰的にサブゴールを設定する。我々はこの振る舞いをヒントにして、再帰的強化学習アルゴリズム RGoal を設計した [Ichisugi 19b][一杉 23a][一杉 24c]。そして我々は、「脳の前頭前野が再帰的強化学習のようなものを実現している」という仮説を脳型 AGI アーキテクチャ設計の前提とした [*1]。

再帰的強化学習を用いれば、経験で獲得した汎用サブルーチンの組み合わせにより体験したことのない新規タスクを解くことが可能になる。また、証明探索や行動計画のような複雑な推論も行える [一杉 19][一杉 23b]。

*1 RGoal は深さの制限のないスタックの存在を仮定してるが、脳内ではそれに代わる何らかの有限サイズの記憶機構を用いていると予想している。

我々はこの仮説にもとづいて、脳が行う行動・推論に関する実験・開発を容易にするために、 Pro5Lang というプログラミング言語の設計・実装を進めてきた [一杉 22b]。行動ルール集合は Pro5Lang のプログラムと見なすことができる。

3.5 仮説：前頭前野では命題が情報処理の基本単位

命題とは真か偽かを問うことができる文である。数理論理学は人間が行う推論や証明の過程を分析する学問であるが、命題はそこでの中心的な概念である。理論言語学においても、平叙文の意味は命題を表していると考えられている。脳の大脳皮質がベイジアンネットであるという仮説を 3.2 節で述べたが、ベイジアンネットのノードやノードの集合が持つ値は命題と見なせる。以上の背景から、我々は「前頭前野の情報処理の基本単位は命題である」と仮定し、アーキテクチャ設計の際の重要な前提の 1 つとしている [一杉 22b][一杉 24a]。命題に対する記号処理的な推論規則を学習で獲得できれば、強力な汎化性能を発揮すると期待できる。

レジスタの値と連想記憶装置に入れられる値はすべて命題である。命題の真偽値は自己完結的に決まるため、プログラム合成（行動ルールの獲得）、実行順序の柔軟な変更、記憶域管理が容易になるという利点がある [一杉 24a]。

行動ルールの学習が理想的に進んだ時、正しい推論だけが行われ、レジスタや連想記憶装置には証明済みの正しい命題だけが入れられるようになるべきである。それを保証する「証明システムの健全性」という要請が、アーキテクチャの詳細設計の際に強力な指針を与える [一杉 22b]。

3.6 仮説：前頭前野にある行動ルール集合がヒトの高度な知能の大部分を実現

行動ルール集合はエージェントの脳と身体を制御するプログラム（ソフトウエア）である。ヒトが持つ高度な知能、例えば複雑な行動計画戦略、推論戦略、文章理解の方法、宣言的知識の整理方法、知識獲得戦略などが、すべて生得的なものだとは考えにくい。むしろ、「これら高度な知能の大部分は、エージェントが長い時間をかけて経験や対話・模倣を通じて獲得したプログラムが実現している」と考えるのが自然であろう。この仮説に従うならば、脳型 AGI を実現するために必要なことは、ヒトが持つ高度な知能そのものを手作業で実現することではなく、複雑なプログラムを自律的に獲得するための仕掛け（プログラム合成技術）を実現することである。我々はそのような考えから、高度な知能を獲得するプログラム合成が効率的に行えるように、合成対象言語 Pro5Lang の言語仕様を設計している [一杉 18b]。

3.7 仮説：脳内では環境を信念状態として表現

3.7.1 信念状態と POMDP

信念状態とは、エージェントが推定する、環境の状態についての確率分布である。部分観測マルコフ決定過程 (POMDP) は、環境のモデルが得られていれば、信念状態を「MDP における状態」と見なして MDP で定式化しなおすことで、解けるようになる [Kaelbling 98]。POMDP を解いたエージェントは、必要な情報を得るために環境を能動的に観測したり、他者に聞くという行動をとるようになる。

我々のアーキテクチャも「脳は環境を信念状態として表現することで POMDP に対処する」という仮説に基づいている。設計中のアーキテクチャにおいては、信念状態は小容量のレジスタのみで表現される。また、値 unknown を使った命題表現により信念状態を大胆に近似表現している [一杉 22c]。これらのおかげで状態空間の次元が小さくなり、次元の呪いによる強化学習の破綻を防ぐと考えている。

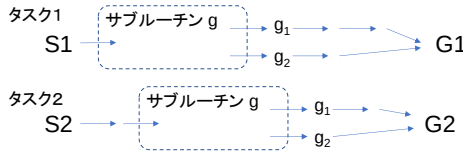

図 2: RGoal でのサブルーチンの考え方。異なる文脈でサブルーチンを共有することで学習効率を上げる。サブルーチンを新規タスクに用いることもできる。

3.8 行動・推論・対話の統一的制御

脳内行動はエージェントのレジスタの値を更新することで信念状態を変化させる。脳内行動の連鎖により環境の隠れた状態などに関する推論が実現される。推論のための行動ルールの「正しさ」は、行動ルールの価値（ゴールに到達するまでのコスト）として、強化学習で学習する [一杉 20a]。推論のための行動ルールは、環境に働きかけるための行動ルールと本質的に区別なく扱われる。推論はエージェントにとって、未知な環境の状態を知るための手段の 1 つにすぎず、自分で推論しないで直接環境を観測するとか、他者に聞くといった他の手段がある場合は、コストが低い手段ほど高い確率で選択される。また、複数の推論戦略があるときも、学習が進めば、状況に応じて効率的な戦略ほど高い確率で選択される。

行動計画もまた推論の一形態であり、行動ルール集合により実現可能である [一杉 23b][一杉 24c]。

対話もまた、通常の推論や行動と特別な違いはなく、行動ルール集合によって制御される [一杉 22c]。言語理解は話者が何を伝えようとしているのかという意図の推定であり、これは環境の隠れた状態の推論の一種である。発話は他者への知識伝達や行動の依頼などを目的としており、これは環境の状態を変化させる行動の一種である。

4. 主要な要素技術と実装の現状

4.1 BESOM

BESOM モデル [Ichisugi 07] を現実の情報処理に役立てるためには、大規模化可能な、効率的な推論・学習アルゴリズムの開発が必要である。現在はその研究は中断中であり、脳型 AGI アーキテクチャのデモに必要な他の要素技術の開発を優先してきた。

大規模化を可能にする方法の 1 つとしては、メッセージ計算をニューラルネットワークで近似する方法が考えられる。ベイジアンネットにおけるノード間のメッセージ計算は一般には指数関数的な時間がかかるが、ネットワークの結合やノードの活性度がスパースな場合はより少ない計算量で近似できる可能性がある。

小規模なデモを目標とするプロトタイプシステムの実装においては必ずしもベイジアンネットは必要ではなく、現在は行動ルール選択のパターンマッチは普通の記号処理として実装している。また、数値ベクトルの圧縮・抽象化も、小規模ベイジアンネットで動作確認する一方 [一杉 20b]、ベイジアンネットを使わない軽量なアルゴリズムの開発も試みている [一杉 22a]。

4.2 RGoal

RGoal [Ichisugi 19b] は再帰的なサブルーチン呼び出しを可能とする階層型強化学習アーキテクチャである（図 2）。サブルーチンは様々な異なる文脈で呼び出されるが、サブルーチン内部の最適方策が文脈に依存せずほぼ同じであれば、サブルーチンの共有によって学習効率を上げられるはずである。また、学習済みのサブルーチンをそれまで経験したことのない新規タスクに用いることもできる。

改良版 RGoal [一杉 23a][一杉 24c] ではサブルーチン終了後の報酬の情報がサブルーチン内部に伝搬する。これにより、複数の終了方法のうち、のちのちよい結果をもたらす可能性が高いものが選択されるようになる。例えば図 2 では、 S1 から G1 、 S2 から G2 に状態遷移する 2 つタスクで同じサブルーチン g が共有される様子を示している。いずれのタスクにおいても、g_1 より g_2 の方が低いコスト（少ない状態遷移のステップ数）でタスクを終了するので、 g_2 の方が価値の高い出口となる。改良版 RGoal のこの性質は、食料や道具の備蓄、知らない場所の探索、仲間同士の良好な関係の維持など、エージェントにとって今すぐは役立たなくても将来役立つ可能性がある行動を生み出すことになるだろう。

RGoal の学習アルゴリズムのいくつかのバージョンを実装し動作を確認ずみである。さらなる改良により、学習の効率化や安定性の向上が可能であると考えている。

4.3 Pro5Lang

RGoal の行動価値関数 $Q(s, g, a)$ を圧縮表現したものは、エージェントの行動を制御するプログラムと見なすことができる [一杉 18b]。我々はこのプログラムを表現するプログラミング言語を Pro5Lang [*2] と名づけ [一杉 22b]、その言語仕様を設計した。

Pro5Lang は論理型言語と機械語の特徴を合わせ持った手続き型言語である。エージェントは環境のなかで動作することで、Pro5Lang プログラムを自律的に獲得する。

Pro5Lang は、論理型言語の代表である Prolog 言語と比較すると、行動価値に応じた非決定論的な動作をする、矛盾を許容する、値 unknown を扱う、などの大きな違いがある。

Pro5Lang の言語仕様の基本部分の設計はほぼ終えたが、ある程度複雑なデモを行うためには、イベントの時系列を認識する機構、メタ認知の機構、タイムアウト・割り込みの機構などいくつかの機能拡張が必要である。また、現状の素朴な実装による推論機構は効率が悪く、ニューラルネットワークと組み合わせた効率化などが必要であると考えている。

5. 関連研究

ACT-R [Anderson 07] のようなプロダクションシステムは脳の振る舞いのモデルとして古くから使われている。Pro5Lang もパターンマッチによって行動ルールを選択するという点において、プロダクションシステムの基本構造を踏襲している。

グローバルワークスペース理論 (GWT) [Baars 07] は神経科学的知見をもとにした、いわゆる「意識」の機能に関するモデルである。Pro5Lang のレジスタは GWT のグローバルワークスペースに対応すると思われる。Pro5Lang は強化学習・ベイジアンネット・数理論理学を理論的基盤としているため、GWT をより理解しやすく効率的実装しやすいものにする可能性がある。

BRA 駆動開発 [Yamakawa 21] は脳の各部品の機能を再現するアーキテクチャを神経科学的知見と整合性を持たせて構築する開発方法論である。一方我々の研究構想 [一杉 21] では脳型 AGI の本質部分の小規模なデモを極力早期に動かそうとしており、BRA 駆動開発とは相補的な関係にあると思われる。

*2 Probabilistic Proven-Proposition Processing Programming Language （確率的証明済み命題処理プログラミング言語）の略。

6. まとめ

ヒトの脳の情報処理を模倣した脳型 AGI アーキテクチャ設計の取り組みの現状について報告した。個々の要素技術の動作確認はすんでおり、全体の統合も可能であるという見通しを得ている。

この方向で研究開発を続ける研究者がもし増えれば、現状の大規模言語モデルの延長では到達し得ないような、ヒトのような知能が実現可能になると考えている。

謝辞

本研究は JSPS 科研費 JP22K12188 の助成を受けたものです。

参考文献

[Anderson 07] Anderson, J. R.: *How Can the Human Mind Occur in the Physical Universe?*, Oxford University Press (2007)

[Baars 07] Baars, B. J. and Franklin, S.: An architectural model of conscious and unconscious brain functions: Global Workspace Theory and IDA, *Neural Networks*, Vol. 20, No. 9, pp. 955–961 (2007), Brain and Consciousness

[Ichisugi 07] Ichisugi, Y.: A Cerebral Cortex Model that Self-Organizes Conditional Probability Tables and Executes Belief Propagation, in *2007 International Joint Conference on Neural Networks*, pp. 178–183 (2007)

[Ichisugi 19a] Ichisugi, Y. and Takahashi, N.: A Formal Model of the Mechanism of Semantic Analysis in the Brain, in Samsonovich, A. V. ed., *Biologically Inspired Cognitive Architectures 2018*, pp. 128–137, Cham (2019), Springer International Publishing

[Ichisugi 19b] Ichisugi, Y., Takahashi, N., Nakada, H., and Sano, T.: Hierarchical Reinforcement Learning with Unlimited Recursive Subroutine Calls, in *Artificial Neural Networks and Machine Learning – ICANN 2019: Deep Learning*, pp. 103–114, Cham (2019)

[Kaelbling 98] Kaelbling, L. P., Littman, M. L., and Cassandra, A. R.: Planning and acting in partially observable stochastic domains, *Artificial Intelligence*, Vol. 101, pp. 99–134 (1998)

[Yamakawa 21] Yamakawa, H.: The whole brain architecture approach: Accelerating the development of artificial general intelligence by referring to the brain, *Neural Networks*, Vol. 144, pp. 478–495 (2021)

[一杉 16] 一杉裕志：疑似ベイジアンネットを用いた認知モデルのプロトタイピング手法の提案, 第 4 回 人工知能学会 汎用人工知能研究会 (SIG-AGI) (2016)

[一杉 18a] 一杉裕志, 高橋直人：脳における文の意味解析機構のモデル, 第 8 回 人工知能学会 汎用人工知能研究会 (SIG-AGI) (2018)

[一杉 18b] 一杉裕志, 高橋直人, 中田秀基, 佐野崇：単一化の機構を利用した階層型強化学習のテーブル圧縮手法の検討, 第 10 回 人工知能学会 汎用人工知能研究会 (SIG-AGI) (2018)

[一杉 19] 一杉裕志, 中田秀基, 高橋直人, 佐野崇：階層型強化学習 RGoal を用いた記号推論の実現手法の検討, 第 12 回 人工知能学会 汎用人工知能研究会 (SIG-AGI) (2019)

[一杉 20a] 一杉裕志, 中田秀基, 高橋直人, 佐野崇：推論規則の価値を階層型強化学習 RGoal を用いて学習する手法の提案, 第 14 回 人工知能学会 汎用人工知能研究会 (SIG-AGI) (2020)

[一杉 20b] 一杉裕志, 中田秀基, 高橋直人, 佐野崇：脳の自律的プログラム合成機構のモデルに向けて： 2 層ベイジアンネットによる記号処理命令の獲得・実行機構, 第 15 回 人工知能学会 汎用人工知能研究会 (SIG-AGI) (2020)

[一杉 20c] 一杉裕志, 中田秀基, 高橋直人, 佐野崇：物体操作に適したワーキングメモリを持つ汎用人工知能アーキテクチャの検討, 第 16 回 人工知能学会 汎用人工知能研究会 (SIG-AGI) (2020)

[一杉 21] 一杉裕志：報酬最大化原理にもとづく脳型ＡＧＩアーキテクチャの構想, 第 18 回 人工知能学会 汎用人工知能研究会 (SIG-AGI) (2021)

[一杉 22a] 一杉裕志, 中田秀基, 高橋直人, 竹内泉, 佐野崇：プログラム合成対象言語 Pro5Lang のための行動価値関数圧縮アルゴリズム, 第 22 回 人工知能学会 汎用人工知能研究会 (SIG-AGI) (2022)

[一杉 22b] 一杉裕志, 中田秀基, 高橋直人, 竹内泉, 佐野崇：汎用人工知能のためのプログラム合成対象言語 Pro5Lang のエピソード記憶機構, 第 20 回 人工知能学会 汎用人工知能研究会 (SIG-AGI) (2022)

[一杉 22c] 一杉裕志, 中田秀基, 高橋直人, 竹内泉, 佐野崇：報酬最大化ＡＧＩのための意思疎通機構の設計とプロトタイプ実装, 第 21 回 人工知能学会 汎用人工知能研究会 (SIG-AGI) (2022)

[一杉 23a] 一杉裕志, 中田秀基, 高橋直人, 竹内泉, 佐野崇：再帰的な階層型強化学習 RGoal へのサブルーチン例外終了機能の導入, 第 25 回 人工知能学会 汎用人工知能研究会 (SIG-AGI) (2023)

[一杉 23b] 一杉裕志, 中田秀基, 高橋直人, 竹内泉, 佐野崇：報酬最大化を目的とする行動計画・実行・対話・推論の統一的制御機構, 第 37 回 人工知能学会全国大会 (2023)

[一杉 24a] 一杉裕志, 高橋直人, 竹内泉, 佐野崇, 中田秀基：プログラム合成対象言語 Pro5Lang における知識表現形式, 第 27 回 人工知能学会 汎用人工知能研究会 (SIG-AGI) (2024)

[一杉 24b] 一杉裕志, 高橋直人, 竹内泉, 佐野崇, 中田秀基：複雑な知能を生み出す極限まで簡素化された環境の設計, 第 28 回 人工知能学会 汎用人工知能研究会 (SIG-AGI) (2024)

[一杉 24c] 一杉裕志, 中田秀基, 高橋直人, 竹内泉, 佐野崇：モンテカルロ版 RGoal アルゴリズムの改良, 第 26 回 人工知能学会 汎用人工知能研究会 (SIG-AGI) (2024)