

特集「研究会紹介」

汎用人工知能研究会 (AGI)

AGI: Artificial General Intelligence

山川 宏
Hiroshi Yamakawa

NPO 法人全脳アーキテクチャ・イニシアティブ
The Whole Brain Architecture Initiative, a non-profit organization.
ymkw@wba-initiative.org, <http://wba-initiative.org/>

市瀬 龍太郎
Ryutaro Ichise

国立情報学研究所
National Institute of Information.
ichise@nii.ac.jp

嶋田 悟
Satoru Shimada

エアロセンス
Aeronsense Inc.
satoru.shimada@aeronsense.co.jp

ジェプカ・ラファウ
Rafal Rzepka

北海道大学, 理化学研究所革新知能統合研究センター
Hokkaido University / RIKEN Center for Advanced Intelligence Project.
rzepka@ist.hokudai.ac.jp

Keywords: style file, template, a list of keywords, etc.

1. 汎用人工知能 (AGI)

1.1 汎用人工知能とは何か

汎用人工知能研究会 (SIG-AGI) (<http://www.sig-agi.org/>), 主査: 山川 宏, 主幹事: 市瀬龍太郎) を紹介する前に, 汎用人工知能について述べる.

汎用人工知能 (Artificial General Intelligence: AGI) とは, 人間のように十分に広範な適用範囲をもち, 設計時の想定を超えた未知の多様な問題を解決できる知能をもつ人工知能である. これは現状で実現されている特化型 AI, つまり, 個別のタスクごと (例えば囲碁, 自動運転など) に問題解決を行える AI に対置する概念である.

AGI は, 技術として実現されておらず, むしろそれを目標として研究を行う分野である. そういった意味では, 知能に関わる「未解決の課題に挑む」という, AI 分野元来の文化を継承した分野であるともいえる.

AGI 研究が目指す知能の姿は必ずしも人のようなものではないが, しばしば「人間レベルの汎用性」が目標となる. そのメリットは, AGI を人間観点から評価でき, 社会実装時の有用性が高く, 研究開発の過程で認知科学や神経科学などの知見を活用しやすいことなどである.

一方, 人のような知能の実現は, AI 分野における当初からの夢である. その点から見て, いまだ実現されず, かつ, 比較的评价を行いやすい「汎用性」は, 研究分野を規定するのに適した側面であるといえる.

1.2 汎用人工知能の意義

人間レベルの包括的なレベルに達せずとも, AI の汎用性が継続的に高まることは価値を生む. つまり, ある範囲のタスク集合に対して継続的に利用される AI は, 新規のタスクの解決において, それ以前に経験したタスクで獲得した知識を部分的に再利用し得る (継続学習, ライフロング学習) ので, 開発コストを削減できる. 汎用技術の性能が市場ニーズを満たしていれば, その性能が特化型 AI より若干低くても, 経済的優位性をもつことになる. さらに近年の機械学習の進展により, 学習で得られる性能が設計を上回るケースも増えつつある.

世界に関する多様な常識を AI が獲得することで, しだいに AGI に近づき, その応用範囲を広げる. まず, 物理世界の常識を獲得すれば, 災害時や宇宙探査などの人間の支援が難しい状況において活動するロボットなどが自律的に生存する能力を高められる. 次に心理学的な常識をもつことは, 接客などの対人インタラクションを行う AI にとって必須である. また上記を兼ね備えることで, ホーム・オフィスロボットなどの実現性が高まる. さらに広範な社会常識を獲得するに至り, 政治家・管理職などへの適用に道が開けるだろう. なお, 広範な常識の獲得には, 好奇心をもって世界を探索してモデル化する自律性が必須である. こうして汎用人工知能は, 知能技術におけるインフラとしての地位を確立するであろう.

長期的にみれば AGI が人類社会に果たす最も大きな貢献は, 人智を超えたレベルで総合的に世界を理解し, 知的フロンティアを開拓することである. この背景として, 近年の機械学習を含むデータ中心の科学においては,

すでに人間の理解を超えた形で予測性をもつ知識（生体内シミュレーション、高次元の物理学など）の蓄積が、個別分野で進みつつある。今後は、さまざまな領域の知識を統合する（組み合わせる、転移させる）AGI技術の進展により、人間の理解を超えたレベルで、研究のディレクションが行えるAGIの出現が期待される。こうして、ノーベル賞級の科学技術の進展の自動化や、気候変動などのグローバル課題の解決策、世界を変革するビジネス・イノベーションなどを主導することが期待される。

いったん、AGIがこうした創造性をもつレベルに到達すれば、それ自身が新たな人工知能を設計・製造（再帰的自己改修）できる。よって生物進化や、人類による技術開発よりも格段に速く知を蓄積できるため、便益とリスクの両面において社会に大きな変化をもたらす。これはしばしば技術的特異点と呼ばれる[山川 15]。

2. AGIに関わる研究動向

2.1 伝統的な認知アーキテクチャ研究

汎用知能のハイライトである、設計時の想定を超えてさまざまな問題を解決できる能力は、知識の柔軟な再利用（組合せ、転用）を基盤とすると考えられる。人間の振舞いを模して、こうした知能を実現するために、以前より認知アーキテクチャの研究が進められていた。これは、知的エージェント（動物や機械）の認識行動の全体を統合的に理論化、実装するためのアーキテクチャであり、知的エージェントを構成するモジュールの配置と、それらモジュールの組合せ（相互作用）を通じて、多様な認知機能をモデル化するものである。古き良きAI（GOFAI）では、世界を記号によって記述し操作し得ると想定する「物理シンボル仮説」を前提としていた。ここでは、記号で記述された知識を総合的に扱うACT-RやSOARなどの認知アーキテクチャが設計された[荒川 14]。

だが、設計者の想定を超えた未知の問題解決能力を、人が設計したドメインごとの知識の組合せから得られる局面は限られる。そもそも、タスクのドメインに依存した知識をつくり込まれたAIにおいては、さまざまなタスクを自律的に解決できるAGIとは言いがたい。目指すべき姿は、個々のドメインにおける知識を獲得・教育できる枠組み（それを支える一般性の高い知識）のみがつくり込まれたシステムである。そうしたAGIであればドメインから得られた知識を利用しつつ、必要に応じて修正しながら、さらなる未知の知識を獲得できるだろう。

こうした機能の実装は困難であるにもかかわらず、AI研究が本来目指していた目標への回帰の動きは根強く、2000年代に入りベン・ゲーツェル氏がAGIという概念を提唱し始めた。ほぼ同じ時期にレイ・カーツワイルが「Singularity in Near」を出版し、人を超えるAIへの期待感が高まった。この時代には、Web上における知識が

大量に蓄積され始め、それを活用して知識を獲得する可能性にも期待が寄せられた。しかし、それだけでは決定的な突破口は得られず、ほとんどの専門家はAGIの実現は遠い未来のものと考えていた。

2.2 深層学習の発展

AI研究で主流であった、記号を中心としたアプローチとはほぼ独立に、非記号的あるいはサブシンボリックな計算を用いる技術も着々と進展していた。特に2012年以降に大きく発展した深層学習は、それ以前に2回のブームがあったニューラルネットワークと、1990年代から進展した統計的機械学習の合流点として形成された分野である[山川 16]。

深層学習技術の発展により、従来に比べて、多種多様な知識を学習データから獲得することが可能になった。こうしてタスクドメインごとの知識を手で設計する状況は大きく変わった。潤沢なデータから獲得した知識を、直接的に利用して帰納推論が行えるAIでは、現在でもしばしば人の能力を凌駕している。

一方で、AGIで目指すべき汎用性を実現するには、さまざまな観点から知識を柔軟に活用（組合せ、転用）することが必要になる。しかし機械学習で得られる知識はデータの統計的な振舞いの再現に留まり、データが少ない未知の問題を解決する能力には至っていない。

2.3 近年の機械学習からの進展状況

タスクをまたぐ汎用性に着目した研究テーマはAGIと関連が深く、以前より、マルチタスク学習（multi-task learning）、継続学習（continual learning）、ライフロング学習（life-long learning）などの問題設定が存在した。またニューラルネットワークの分野では新タスクへの適応時に過去のタスクで得られた知識の破壊を防ぐ研究などが進められていた。技術的進展を背景に、2010年代後半となった現在において、最先端の機械学習分野では、こうした問題設定に対して本格的に取り組む状況になってきた。

具体的なアプローチとしては、個別のタスクの学習を制御するメタパラメータの学習、さらに、再利用可能な形で知識を獲得する研究と、知識を活用する研究も始まっている。後者としては、多様なモダリティにおいてデータから得られる知識を再利用性が高い部品として整理するためのDisentanglement、グラフ表現といった研究やオントロジーとの対応付けなどがある。次にニューラルネットワーク上において知識を組み合わせるため、注意の機構の導入や記号や言語機能との結合、また類推のように知識の転用を行う転移学習といった研究も進んでいる。

こうした流れの中で改めて、さまざまな知識を組み合わせるためのアーキテクチャの重要性は再び高まりつつある。つまりGOFAIで目指していた知識の再

利用を実現する認知アーキテクチャ研究が、データから獲得した知識に対して展開しはじめている。現在もアーキテクチャの設計は、人の試行錯誤に負う場合が多いが、設計空間は膨大である。そこで、次々に自動生成したアーキテクチャを比較評価して進化させる、アーキテクチャ探索も試みられている。また、神経科学の進展を背景に、現存する汎用知能である（人間の）脳を参考にして、設計を行う脳型 AGI も見直されている（例えば、全脳アーキテクチャアプローチ）。

AGI 完成までに残された課題のすべては見通し得ないが、少なくともアーキテクチャ上で行われるさまざまな知的情報処理をメタな視点から捉え、新たな情報処理の観点自体を探索する能力は重要である。こうしたメタ処理では、膨大な可能性の中から、有力な候補を効率的に見つけ出す必要がある。そこで例えば、データ内に潜む、普遍性、対称性、等価性、などを捉えようという基礎的な研究も始まっている。

なお AGI の開発状況については、SSRN の 2017 年の報告 [Baum 17] が網羅的である。

3. 研究会の紹介

3.1 助 走 期

当研究会の発端は、現主査の山川が 2012 年の冬に英国オックスフォード大学で開催された二つの国際会議である AGI 2012 および AGI Impact に研究発表のために参加したこと (2013)、およびほぼ同時期に、現主幹事（前主査）の市瀬が関連する活動に着目していたことがある。

そこで、長らく認知アーキテクチャ研究について議論を継続していた山川・市瀬が、前章で述べた深層学習の発展が始まったタイミングにおいて、日本でも AGI 研究を進めるべきであると考え、2013 年の春から、AGI に関わる研究活動の立上げについて検討を進めた。まず出発点としては、日本国内の人工知能分野において AGI という学際研究の動向を把握しながら、この分野の中核コミュニティを形成する必要があると考えた [山川 14]。

他方で、山川らは、2012 年 9 月 21 日に行われたトボス会議「人間の知性とコンピュータ科学の未来—2040 年コンピュータは人間をこえてしまうのか—」の書き起こしを行い、当学会誌上において 2013 年 5 月に、人工知能学会誌 (Vol. 28, No. 3, pp. 424-471) での特別企画「シンギュラリティの時代：人を超えゆく知性とともに」の記事として発行した。

3.2 輪読会の開始

こうした流れのなかで、まずは AGI の世界的なリーダーである Ben Goertzel 氏らの著書 [Goertzel 06] を対象とする輪読会を立ち上げる提案が市瀬よりなされた。なぜならば、AI 研究者自身が、AGI の概略について理

解を深めることが重要と考えたからである。

そこで本学会のメーリングリストを中心に参加メンバーの募集を行ったところ、我々を含む計 14 名の研究者が集まり輪読会を開始した。その後もほぼ毎月 1 回のペースで、国立情報学研究所を拠点として輪読会を開催し、現在も継続している。

初回 (2013 年 7 月 3 日) は、輪読会の開催趣旨確認や自己紹介に引き続き、オーガナイザーである山川と市瀬が序章と第 1 章の紹介を行うことで、AGI の全体像と輪読書の内容を概観した。

第 2 回 (2013 年 8 月 26 日) では、現幹事のジェプカ・ラファウの仲介により、ちょうど来日していた上述の本の著者 Ben Goertzel 氏を招き、AGI の研究動向全体について 1 時間の講演をいただいた。

こうして収集した輪読会での情報や、交流を深めた Ben Goertzel 氏の寄稿、さらに 2012 年の AAAI Magazine に掲載された「Mapping the Landscape of Human-Level Artificial General Intelligence」、第 6 回 AGI Conference (2013) のレビューなどを含める形で、本誌の 2014 年 5 月号に特集「汎用人工知能 (AGI) への招待」(人工知能, Vol. 29, No. 3, pp. 226-325) を企画した。

その直後には、シンギュラリティの瞬間を映像化した映画を題材として JSAI 2014 で実施された、「映画『トランセンデンス』共同特別企画～エヴリンの夢～」を企画した。

その後も輪読会は、毎月一度のペースで継続し、2019 年 5 月時点までに 58 回開催している。その間に雑誌 Neuron の「Cognitive Architectures」特集 [Dehaene 15]、書籍「Theoretical Foundations of Artificial General Intelligence」[Wang 14]、書籍「The Measure of All Minds. Evaluating Natural and Artificial Intelligence」[Hernández-Orallo 17]、書籍「Lifelong Machine Learning, Second Edition」[Chen 18]などを始めとして、AGI 関連のさまざまなサーベイ論文などの輪読を行ってきた。

コミュニティ活動にも力を入れており、Facebook グループ「汎用人工知能について語ろう」を運営しており、2019 年 5 月時点で 2400 人ほどのメンバーが登録している。同時期から、Web サイト「汎用人工知能と技術的特異点」(<http://www.sig-agi.org/>) の運営も行い、当初は全脳アーキテクチャ勉強会の情報も合わせて発信していた。これは現在は、「汎用人工知能研究会」のサイトとして運営されている。また、AGI 関連のイベントなどを広報するメーリングリストも運用しており、2019 年 5 月時点で、300 名以上の方が登録されている。

3.3 全国大会 OS

本学会の全国大会において 2014 年から 2017 年までの 4 年間連続で「汎用人工知能とその社会への影響」と題するセッションを企画・開催した。このセッションで

は、AGIを技術的・社会的側面の両面から議論を行った。招待講演としては、2014年は松田卓也氏による「2045年問題」、2015年は久木田水生氏による「人工知能は道徳的になり得るか、あるいはなるべきか」、2016年は村上祐子氏による「人工知能と共存する人間」、2017年は井上智洋氏による「経済的特異点—汎用人工知能が経済に与えるインパクト—」を企画した。

関連してJSAI 2018では、OS「人とAIが織りなす新たなエコシステム」および物理学会連携「理解」企画セッションに協賛した。

3.4 研究会の設立と活動 (2015年)

「汎用人工知能研究会 (SIG-AGI)」は、そのAGIの研究活動を支える母体となることを目的として2015年7月に発足した。

また、このメンバを中心として、本分野の教科書となる書籍の執筆・刊行をゴールとするプロジェクトを実施することを目標に含め、2017年度までの3年を年限とする第3種研究会 (プロジェクト指向研究会) とした。

その年の9月30日には、招待講演者としてBen Goertzel氏を招いて設立記念シンポジウムを実施した [日塔 16]。氏からは「Robotics & AGI」と題した講演において、“ハイブリッド”と呼ばれるシンボリックとパター

ンを組み合わせたアーキテクチャである「OpenCog」の開発状況の解説をいただいた。

研究会の設立後は、2015年12月の初回の研究会を皮切りに、毎年3回ほどの研究会を実施している。

研究会で扱われている主な技術分野は、「認知アーキテクチャ」、「認知ロボティクス」、「機械学習・推論・知識」、「脳の計算の仕組み」、「人工知能の評価方法」、「記号創発 (パターンと記号の関係)」などであるが、「人工知能の社会的影響」についての発表もある。以下の表1に2019年3月までの実施実績を、招待講演を中心に示す。

関連した特別企画として、2016年11月11日には、人工知能学会創立30周年記念事業「みんなで作る認知アーキテクチャ」において、第2回全脳アーキテクチャハッカソン (同年10月開催) 中の優秀な選抜チームに対して当研究会がプレゼンコンペを行い、最優秀チームを表彰した。

また2017年10月12日には、AGIの理論研究者として著名なオーストラリア国立大学のMarcus Hutter教授とAIXIの最新動向を中心に活発な意見交換を行った。

第3種研究会としての期間が満了後の2018年4月以降は、継続的に研究会や輪読会の実施などを行うために、第2種研究会に変更し、活動を継続している。

4. 関連する学術コミュニティ

学術コミュニティとして、伝統的にはAGI Societyが2008年からAGI国際会議を例年実施している。AI関連の大きな会議としては、以前よりAAAIにおいてはAGIに関する活動があり、近年はIJCAIでも汎用性に関わるワークショップなども企画されている。また、より生物を参考にする立場からはBiologically Inspired Cognitive Architectures (BICA) Societyが2009年から国際会議を継続している。また2017年からは、より神経科学寄りの会議としてCognitive Computational Neuroscience (CCN) が立ち上がっている。国内では、機械学習機の統合による脳型AIの構築を目指し、研究開発を促進するNPO法人全脳アーキテクチャ・イニシアティブがある。現在は、先端的な深層学習コミュニティもAIの汎用性に着目しており、代表的なNeurIPS、ICLRなどの会議においても本質的にはAGIに関連する研究が増えている。

◇ 参考文献 ◇

- [荒川 14] 荒川直哉, 山川 宏, 市瀬龍太郎: 汎用人工知能の研究動向, 第28回人工知能学会全国大会論文集, 2C4-OS-22a-1 (2014)
- [Baum 17] Baum, S.: A Survey of Artificial General Intelligence Projects for Ethics, Risk, and Policy, Global Catastrophic Risk Institute Working Paper, 17-1 (2017)
- [Chen 18] Chen, Z. and Liu, B.: *Lifelong Machine Learning*, Second Edition, Morgan & Claypool (2018)
- [Dehaene 15] Dehaene, S., Dudai, Y. and Konen, C.: Cognitive

表1 研究会の実施状況

回	開催日	招待講演	ロング	ショート
1	2015/12/16	—		12
2	2016/03/28	小林亮太: 脳における計算の仕組み: ハードウェアとアルゴリズム		4
		大澤博隆: 汎用人工知能とHAI: 人らしさを作る/使うことの意味		
3	2016/09/12	栗原 聡: 汎用人工知能を目指す人工知能先端研究センター (AIX) の紹介	1	4
4	2016/12/15	西田豊明: 人間調和型人工知能をめざして	1	5
5	2017/03/27	巻 俊宏: 自律型海中ロボットAUV ~魚のごとく~	1	7
6	2017/09/15	宮部賢志: Solomonoffの万能推論・アルゴリズムの確率	1	8
		相澤彰子: 強化学習と汎用エージェント		
7	2017/11/24	中川裕志: AGIへの道程: 2017年版	2	5
8	2018/03/20	野波健蔵: 自律型空中ロボットUAV ~鳥のごとく~	3	5
9	2018/08/30	片山 晋: General AI Challengeに参加して	3	3
10	2018/11/22	谷口忠大: 記号創発ロボティクスによる汎用人工知能への挑戦	3	2
11	2019/03/15	熊谷 亘: 汎用性の獲得に有効な機械学習手法	2	5

- architectures, *Neuron*, Vol. 88, Issue 1 (2015)
- [Goertzel 06] Goertzel, B. and Pennachin, C.: *Artificial General Intelligence (Cognitive Technologies)*, Springer (2006)
- [Hernández-Orallo 17] Hernández-Orallo, J.: *The Measure of All Minds*, Cambridge University Press (2017)
- [日塔 16] 日塔 史: 研究会報告: 汎用人工知能研究会 (SIG-AGI) 創設記念シンポジウム, 人工知能, Vol. 31, No. 1, pp. 109-110 (2016)
- [山川 14] 山川 宏, 市瀬龍太郎: 汎用人工知能輪読会の発足とその後の活動, 人工知能, Vol. 29, No.3, pp. 265-267 (2014)
- [山川 15] 山川 宏, 市瀬龍太郎, 井上智洋: 汎用人工知能が技術的特異点を巻き起こす, 信学誌, Vol. 98, No. 3, pp. 238-243 (2015)
- [山川 16] 山川 宏: 人工知能学会共同企画—人工知能とは何か?: [人工知能のホットトピック] 3.1 汎用性の創発を脳に学ぶために, 情報処理, Vol. 57, No. 10, pp.981-984 (2016)
- [Wang 14] Wang, P. and Goertzel, B.: *Theoretical Foundations of Artificial General Intelligence*, Atlantis Press (2014)

2019年7月8日 受理

著者紹介



山川 宏 (正会員)

1992年東京大学大学院工学系研究科電子工学専攻博士課程修了。工学博士。同年、株式会社富士通研究所入社。1994年から2000年まで通産省RWCプロジェクトに従事。2014年から2019年3月まで株式会社ドワンゴドワンゴ人工知能研究所所長。現在、特定非営利活動法人全脳アーキテクチャ・イニシアティブ代表。専門は人工知能、特に、汎用人工知能、

全脳アーキテクチャ、概念獲得、意見集約技術など。産業技術総合研究所人工知能研究センター客員研究員、電気通信大学大学院客員教授、理化学研究所革新知能統合研究センター客員研究員、東京大学医学部客員研究員。共訳書に、「パターン認識と機械学習」(丸善出版、2012)、共著書に「人工知能とは」(近代科学社、2016)、「宗教と生命」(角川学芸出版、2018)、「AI時代の憲法論」(毎日新聞出版、2018)などがある。本学会汎用人工知能研究会主査、電子情報通信学会、日本認知科学会、日本神経回路学会などの各会員。



市瀬 龍太郎 (正会員)

1995年東京工業大学工学部情報工学科卒業。2000年同大学院情報理工学研究科計算工学専攻博士課程修了。博士(工学)。同年より国立情報学研究所助手、助教授を経て、現在、同研究所情報学プリンシプル研究系准教授。総合研究大学院大学准教授。産業技術総合研究所招聘研究員を併任。機械学習、知識発見、知識処理などの研究に従事。電子情報通信学会、

情報処理学会各シニア会員、AAAI、日本認知科学会各会員。本学会理事、本誌編集委員長。



嶋田 悟 (正会員)

2008年セント・アンドリュース大学(スコットランド)にて哲学を修めた後、ペイン・アンド・カンパニー(外資系経営戦略コンサルティング会社)、株式会社経営共創基盤(IGPI: 富山和彦 CEO)を通じて、ハイテク業界を中心に、新規事業開発、カーブアウト、事業再建などに従事。2015年自動運転を手掛ける株式会社ZMPに参画、エアロセンス株式会社取締役COO。



ジェプカ・ラファウ (正会員)

2004年北海道大学大学院工学研究科博士後期課程電子工学専攻修了。同年、北海道大学大学院情報科学研究科メディアネットワーク専攻助手。2007年同助教。現在に至る。言語学修士、工学博士。常識的知識獲得、感情処理、機械倫理、集合知などの研究に従事。オーストラリア国立大学および理化学研究所革新知能統合研究センターの客員研究員。Lifeboat

Foundationの諮問委員、言語処理学会、日本認知科学会、AAAI各会員。