レクチャーシリーズ「シンギュラリティとAI」〔第1回〕

人を超えるとはどういうことか？

What Does It Mean “Exceeding Human”?

栗原 聡 電気通信大学，人工知能先端研究センター
Satoshi Kurihara The University of Electro-Communications./ Artificial Intelligence eXploration Research Center.
skurihara@uec.ac.jp, http://www.ics.lab.uec.ac.jp/, http://aix.uec.ac.jp/

Keywords: singularity, artificial general intelligence（AGI）, consciousness, autonomy, meta-planning, multimodal.

1. は じ め に

「シンギュラリティ」と聞くと辟易（へきえき）する読者もおられるかと思うが，本誌にていよいよシンギュラリティをテーマとするレクチャーシリーズの開始である．このテーマは山川 宏編集長にしかできないと思っていたし，実際にこの時期にこのシリーズが企画されたことはうれしい限りである．そして，恐れ多いことにトップバッターを仰せつかることになった．次号以降，ビッグネームがどんどん登場することになることから，少しでも議論のきっかけになるような内容としたい．

この3度目の人工知能（AI）ブームにて，著者もいろいろな場で講演させていただくのであるが，総じて「シンギュラリティは訪れるのか？」，「本当に人は人工知能に抜かれるのか？」，「人類が人工知能に支配されるときは来るのか？」辺りが最も気になるところのようだ．「2045年にAIが人を凌駕し，映画「ターミネーター」のごとく人類を支配する！」といわれれば誰もが不安になるのはもっともであろう．果たしてそのような世界は本当に訪れるのであろうか？ 現在のAIブームを牽引するDeep Learning技術の進展がシンギュラリティを引き起こしてしまうのであろうか？ 研究者の間でもいろいろな考えがあり，つまりは一般の方々とAI研究者ではその捉え方には大きな差も見受けられる．まず，想像しているAI像からして異なっているともいえる．なぜそのような差異が生じてしまったのか？ そして，上記の疑問にAI研究者の端くれとして，ここでは著者の主観的な考察を書かせていただきたいと思う．

2. 用途限定型 vs 汎用型

シンギュラリティについての話題の際に，これまでのAIと今後のAIという文脈において必ず引合いに出される言葉が，「用途限定型AI・汎用型AI」もしくは「弱いAI・強いAI」である．一見わかりやすい区別のようにも思えるが，よくよく考えるといろいろ疑問がわいてくる．例えば，掃除ロボットは用途限定型AIである．この掃除ロボットに不審者検知機能も搭載されたとしても，やはり用途限定型であろう．しかし，仮に，もちろん現在の掃除ロボットとは外見は一変するであろうが，さらに家電操作機能やペット見守り機能，電子秘書機能，電話機能……など，より多くの機能が搭載されるとなると，これでも用途限定型と呼べるであろうか？ ロボットアームが搭載され，モノをつかむ機能が搭載されればさらに多くの機能が可能になる．もはや汎用型であろう！ 用途限定型と汎用型に明確な線引きは難しいのである．しかし，単に多くの機能が搭載されれば汎用型である，と考えるのも短絡的である．

より多くの機能を搭載すればするほど，今度はそのときどきにおいてどの機能を実行するかの選択をする必要が出てくる．部屋が汚れていれば掃除機能を発動させ，モノが落ちていればピックアップ機能を発動させる．掃除中にユーザからテレビ予約依頼が入れば掃除を一時中止して録画機能を発動させるといった具合である．その際，録画開始時刻が5時間も先であれば，そのまま掃除を済ませたほうが効率が良いと考えるかもしれない．それこそ，「家を片付けておいて」といった命令に対して，何をどの順番で実行すればよいのであろうか？ もちろん，搭載される機能がそれほど多くはなく，開発者が「どのようなときにはどの機能を実行させる」というように選択部分をつくり込むことが可能なレベルであれば，そのAIは多くの機能を開発者の意図どおりに適切に使い分けることができるであろう．もちろん，この方法であっても多機能ロボットは有用に機能するであろう．そして，このレベルの汎用性をもつAIを「低汎用型AI」と呼ぶことにしよう．しかし，開発者が想定しない状況に対応することはできない．乳幼児が怪我や事故で病院に行くことになる大半は実は家庭内で起きている．それだけ予期せぬことが起こるということである．人であっても想定外の状況に直面することがあるということは，開発者が，想定されるすべての選択ルールを記述することは不

可能であることを意味する．

いずれにせよ，これからのAI開発においてはまず低汎用型AIへの取組みが加速することが予想される．しかし，上述するように低汎用型AIはあくまで開発者の意図どおりの動作に限定された汎用性は発揮できるものの，想定外の状況に適切に対応できる保証はない．つまり，搭載される機能の増加は，それだけ機能を選択する部分の負荷の増大を招くことになる．そして，この機能選択部分の開発が今後課題になるわけだが，そう簡単な話ではない．人の意思決定メカニズムを実装することと同義だからである．仮に100種類の機能が搭載されたAIを開発するとした場合，100種類から一つを選択すればよいわけではない．新たな状況に対応するために，複数の機能を組み合わせる場合もあれば，さらにその際に組み合わせる順番やそのタイミングを考慮するケースも想定される．現在のところ，人の意思決定レベルと同等の機能選択メカニズムは実現されてはいない．まさに今後のAI，すなわち「高汎用型AI」を実現するには，このメカニズムの創出が重要な課題となると考えている．

3. 強いAIと意識

強いAI・弱いAIは哲学者ジョン・サールがつくった用語である．弱いAIは用途限定型AIと同じ解釈でよいものの，強いAIは汎用型AIという意味ではない．サールによれば，高い知能をもつAIは「意識」をもつとされる．そもそも，意識とは何か？　とかAIは意識をもつことができるのか？　といった質問についてはAI研究者の間でもいろいろな意見があるが[人工知能16]，著者は，上記「高汎用型AI＝強いAI」と考えている．高い汎用性をもつAIは意識をもつということである．ちなみに，ここでの意識は顕在意識，つまり自分として意識して行動する，というときの意識のことを指している．

書籍「人工知能とは」[人工知能16]においても，意識とは「計算のプロセス」や「大脳の動きを大脳が認識する再帰的状態」などさまざまな意見がある中，著者としては，意識を，「生命活動を細胞レベルの時間粒度を超えたスケールにて時間軸方向で安定化させるために神経細胞ネットワークにて創発される（生み出される）現象であり，脳において同時並行に処理される膨大なタスクにおいて，ごく一部の重要なタスクが顕在意識という形で俗に言うところに意識として認識される」と捉えている．何しろ，五感や外界とのインタラクションからのリアルタイムな大規模データと，経験・記憶といった膨大な情報に基づいて動作する，およそ1 500億個の神経細胞が形成する大規模複雑ネットワークが生み出す意識空間は超巨大であり，このすべてが顕在化すれば我々は混乱してしまうであろう．そこで，他者との関係を持続させるに必要な最低限度の意識が顕在化するように進化したのではないかと考えている．なお，顕在化しない意識は潜在意識と呼ばれる．

我々の身体を構成する細胞は数か月ですべて新しい細胞に入れ替わる，つまり，昨年と今年の自分の身体は，実際同じではないにもかかわらず，「自分」という一貫した意識をもち続けることができる．この機能が，人が種を維持させるために有効であったということなのであろう．人が他者との関わり，そしてお互いに社会的な存在としてその関係を維持するためには，各自が一貫して自分であるという意識をもち続けることが必要であり，人が自然界にて絶滅することなく生き残るために重要な機能だったのだと考えられる．別の言い方をすれば，一貫性を維持するには時系列的な記憶，すなわちエピソード記憶を利用できることが必要であり，それを扱うための機能が意識なのだと考えている．この考え方の詳細は[前野10]を参照いただきたい．

では，なぜ高汎用型AIは意識（ここでは顕在意識のこと）をもつのか？　であるが，意識とは要するに「重要な振舞いや行動のモニタ」である．我々が，相手も意識をもつことを実感するのは，相手の行動に意図を感じるからである．意図とはその行動を行う動機であり，それは人が自律システムであるからである．つまり自らがある目的に基づいて自発的に行動するとき，そこには行動を起こすための意図があるはずで，我々はそこに意識を感じるのだと考えられる．我々が明らかに限定された行動パターンしかもたない家電などに意識を感じないのは，それが大きな理由であろう．家電が真にユーザのことを気遣って巧妙に自律的に動作することを想像するに，恐らくは我々は家電が意識をもっているように感じるであろう．システムが高い自律性があるかないかが，我々がシステムに何らかの意図を感じ，システムが意識をもつという感覚をもつかもたないかを分けるのだと考えている．その意味では高い自律性をもつ高汎用型AIに対して我々は意識を感じるはずであり，これこそサールの定義する強いAIと見ることができる．整理すると，用途限定型AIと低汎用型AIは弱いAI，そして，高汎用型AIが強いAIという関係であるという主張である．

4. 現在のブームに見るずれ

これまでの議論から，現在のAIブームにおける研究・開発側と一般社会とのAIに対する見識のずれの正体が見えてくる．明らかにこれまで，そして現在のAIブームの中核であるDeep Learningに関わるAIは用途限定型か低汎用型である．自ら意識をもつことを想起される高汎用型AIをつくりだすことはそう簡単な話ではない．そして，用途限定型や低汎用型のAIでは人を支配することはできない．これに対して，映画「ターミネーター」に見るSF作品などでのAIは明らかに高汎用型AIである．鉄腕アトムやドラえもんも高汎用型である．そして，究極の高汎用型AIは生命と同じく自らが自らを生み出

す能力ももつであろうし，まさに「ターミネーター」の世界に登場するスカイネットと同じ能力である．ここまでのレベルになればもはや新しい生命体であるが，このレベルのAIが登場すれば，明らかにシンギュラリティの到来なのであろう．しかし，だからといって人を超えたということになるのであろうか？

5. 超えられたと感じるとき

そもそも，「2045年にAIが人を超える」という簡潔な表現が上記のずれを含むさまざまな憶測（AIが職を奪うやAIに支配される）を生み出しているのだと思われるが，では，具体的にどのような状況であれば我々はAIに超えられた，またはAIに支配される感覚を抱くのであろうか？

すでに現在において，人はほとんどの能力において情報処理技術や機械に超えられている．計算，記憶，認識，移動，正確さ，駆動時間など，個々の機能においてはほぼすべてにおいて機械のほうが圧倒的に優れている．そもそも，より便利な，そして人が楽をできるようにすることが科学技術の目的であるのだから当然であろう．しかし，現在身の回りにあるさまざまな機器に対して我々が驚異を感じる類いは存在しない．対話システムやチャットボットは，間違いなく個々人のもつ知識をはるかに凌駕する量をもっているであろう．にもかかわらず，我々はSiriに恐れを抱くことはない．我々は用途限定型や低汎用型システムには驚異を感じないようだ．なぜなら，システムがどのように振る舞うかが既知であり，想定外の振舞いをしないのであれば，そこに意識を感じることもなく単なる機械としか感じないからである．このレベルのAIには必ずそれを操作する人間が必要不可欠であり，人の良きサポータにはなれるかもしれないが，人と共生する関係にはなれない．よって，我々から職を奪ったり，ましてや人を支配することもない．ただし，道具であることから，道具として悪用される面をもつことはこれまでの科学技術と同様である．

これに対して，今後，我々が意識を感じる高汎用型AI（サールの定義における強いAI）が登場するとどうなるであろうか？　それが家庭内におけるお手伝いさん用途であり，徹底して人に寄り添うように動作する高汎用型AIに対してであれば我々は安心感を抱くのであろうか？　つまりはドラえもんのようなAIである．この問いについて，著者は「人は安易に安心感を抱く」と即答することはできない．

初対面の人とのやり取りの際，相手がいきなり殴りかかってくることは常識的にあり得ないであろうし，よって特に身構えることもない．しかし，もし，その相手がガタイが大きかったりすれば多少の不安を覚えることがあっても，話すことで安心した．こういった経験は誰しももっているのではなかろうか？　一見怖そうでも，体が小さく，いざとなれば力で押さえ付けることができる，と確信すればやはり安心する．AIやロボットに対しても同じなのだと思う．付き合い始めて時間をかけて徐々に信頼感が芽生え，AIは安心であるという社会的なコンセンサスが構築されていく．この過程は自動運転車が社会に受け入れられる場合でも同様であると考えている．

つまり，人と共生するAIが登場したとしても，人はまだ超えられた感覚は抱かないのではないかと思う．共生の関係から，そしてAIの高汎用性のレベルから，人がAIを自分より下と認識する状況においては抜かれたという感覚は抱かないであろう．では，いつその感覚を抱くのかといえば，それは社会的な立場における逆転が生じたときであろう．具体的には，高汎用性のある自律型AIに管理される状況が訪れたときである．すでにAIが適切な配属先を決めるAIシステムも登場しているが，人が操作しているのであり，人が人を管理している図式であることには変わりはない．しかし，高汎用型AIは自らの判断で人を管理する図式となる．このとき，我々は人ではないモノに指図される，という「超えられた」という感覚を抱くのではないだろうか．つまり，ハードウェアの性能としては，高い自律性をもつことが必須であり，その意味では，脳をコンピュータと見立てたときに性能を数字的に超えるスーパコンピュータを実現したからといって，人を超えるAIができたなどと短絡的にはいえない．逆にいえば，機能や能力は限定されても高い自律性をもてば人を超えることは可能だということである．高い汎用性をもつ自律型AIに日常生活において何かしら命令される立場になったとき，社会的な立場が逆転したときが，人がAIに超えられたときということになる．

6. 高汎用型AI実現への課題

では，高汎用型AIはいつ実現するのかというと，まだまだ多くの課題があるわけであるが，著者としては以下に述べる三つが主要な課題であると考えている．

6·1 メタプランニング能力

高い自律性をもつAIは置かれた状況においてどの行動や機能を実行するかの選択を能動的に行う機能が必須となるが，現在の環境状態を与えられた目的状態に変換するための行動の手順を求める技術がプランニングである．定番の古典的プランニング法といえばSTRIPSであろう．ロドニー・ブルックスの提案したサブサンプションアーキテクチャもプランニング技術の一つだし，リアクティブプランニングやリアルタイムプランニング，マルチエージェントプランニングなど，プランニング研究の歴史は深い．しかしメタプランニングはこれらと異なり，プランニングモジュールに対してどのような目的を

与えるのかを考えるのがその目的である．わかりやすくするため，ロボットを例として考える．バッテリ残存量が減ったので充電ポイントに移動するための移動プランを生成する場合の「充電ポイントへの移動」が従来のプランニングにおける目的であり，このような具体的な目的のことを「実目的」と呼ぶことにする．一方，ロボットに与えられた目的が「家の見守り」といった抽象的なものである場合，これを「メタ目的」と呼び，家の安全の維持という目的を達成するため，安全を脅かす状況が発生する度，それを除去する実目的を生成する．ロボットは駆動し続ける必要があることから，充電するという実目的を選択する場合があるかもしれないが，仮に充電という実目的を選択した直後に，異常を検知し，その異常を排除する実目的が生成された場合，ロボットはどちらの実目的を優先するのかを決定しなければならないし，実目的によっては複数の実目的による複合的な対応が必要なケースも考えられる．これがメタプランニングである．2章でも述べたように，AIが選択可能な振舞いが100種類と少なくても，それらを組み合わせることでさらに多くのことに対応できるかもしれない．それこそが生物が環境に適応するための重要な能力であり，AIが人に追いつき追い越すにはこの能力が必須だと考える．

そして，メタプランニングはロボットの行動といった身体的な動作に限るものではなく対話にも当てはまる．Siriなどの対話AIやチャットボットなどの開発が加速しているが，残念ながら人同士のような生きた会話とはならない．対話AIが利用できる語彙力や知識量はすでに個人のレベルを大きく上回っているであろう．それにもかかわらず，人同士のような生きた会話や，場の空気を読んだやり取りができない理由は何なのであろうか？

一つは，AIが人と会話する状況において，その場の雰囲気やそのときの社会状況，そして，会話相手の現在の状況といった背景を知らないで会話しようとするからである．そして，もう一つが，そのような背景に基づき，例えば，その場の雰囲気や会話相手の平常心の維持といったメタ目的を達成するための会話生成機構を備えていないからである．

身近な例について考えてみたい．現在の対話システムに「喉が渇いた」と話しかければ，直近のコンビニや自販機の場所が回答として返ってくるであろう．しかし，人同士の場合，「今は我慢して！」などと返答する場合もある．この発言は喉の渇きを潤すための返答ではない．理由は，直近の自販機には水以外の高カロリーなジュースしかなく，相手の糖分取り過ぎによる健康への悪影響を防ぐための発言だったのである．これは相手の体型や好み，健康状態，そのときの季節や気温などを把握していない限りそのような返答はできない．つまり，「相手の健康を気遣った」，別の解釈をすれば「相手の幸福度を向上させたい」というメタ目的を達成するために「今は我慢して」という発言をしたのである．相手への気遣い以外にも，「その場の雰囲気を維持したい」とか「自らの欲望を達成したい」など，我々はさまざまな目的をその場その場の状況で自分なりの価値判断にて選択し相手との会話や振舞いを行っている．しかし，現在の対話システムには，このような目的指向性がなく，単に与えられた質問に解答するのみであることから，そもそも人同士のような会話の成立が困難なのである．

6·2 マルチモーダルネットワーク処理技術

Deep Learning研究においてもマルチモーダルデータを対象とする論文を見掛けるようになってきた．画像や音声といった異なるデータを利用することで認識精度を向上させようという枠組みであるが，データの種類ごとに特徴抽出を行った結果を統合する方法が主流である．

しかし，この方法では重要な情報を利用できていない．それは，異なるデータの種類間の関係，すなわちつながりの情報である．著者はマルチモーダルデータにおいてつながりこそが重要であると考えている．かなり強引な例であるが，1種類のデータが100個あり，得られる情報量が100であるとしよう．そして，10種類のデータがそれぞれ10個あれば，単純に情報量は100となる．しかし，実際は各データそれぞれ7個くらいあれば情報量100となる，という主張である．情報量30が足りないが，これがつながりから得られる情報量という意味である．脳という限られたリソースを効率的に利用するには，五感からの情報を独立に処理するのではなく，互いに関連させることによる効果を利用するほうが効率的である．つまり，脳は五感で得られる情報を互いに関連させて利用することで高い認知能力を実現させるエコシステムだという見方である．では，ここでのつながりは何かというと，同時刻に経験したという，時間的なつながりである．この画像を見たときにこの音を聞き，そのときの天気は快晴で気温は30℃で場所は○○……など，我々は同時刻に入力される五感からの情報を関連させて記憶している．もちろん，その直前に入力された情報とも関連させているであろうし，入力された画像に対して，過去に類似した画像があればその画像とも関連させているであろう．

このように入力されるマルチモーダルデータが時間軸やデータ間の類似性に基づき大規模かつ複雑なネットワークを形成しているとすれば，このネットワークを対象とした特徴抽出を行うことで，つながりの情報を利用することができる．そのためには，例えば関係グラフである複雑ネットワークを対象とするDeep Learning法の確立が望まれる．画像や音声はピクセルの位置や音の順番に意味がある空間グラフであり，スモールワード性におけるショートカットも存在しない．よって，CNNなどの多層ニューラルネットワークが効果的に機能するが，関係グラフである複雑ネットワークは，いわゆるス

ケールフリー性・スモールワールド性という典型的な特徴をもつ．スモールワールドネットワークでは，空間グラフのように個々のノードの近隣ノードのみに着目してしまうと，ショートカットによる遠方のノードの効果を見落としてしまい，スケールフリーネットワークでは，ハブノードという特徴的なノードの取扱いが課題となる．一方，複雑ネットワーク分析研究により中心性や同質性といったさまざまな指標が提案されているが，これらの人が経験的に定義した素性を利用する方法は即効性があるものの，真に筋の良い素性が獲得されているかはわからない．その点，多層ニューラルネットワークの表層学習能力を利用するアプローチへの期待は大きい．

6·3 因果の鎖の発見

「風が吹けば桶屋が儲かる」ということわざがある．ある事象の発生により，一見すると全く関係がないと思われる場所・物事に影響が及ぶことの喩(たと)えである．バタフライ効果も似たような表現であるが，メタプランニングはこの逆問題を解くということである．「桶屋を儲けさせる」というメタ目的に対するメタプランニングを実行した結果，「風を吹かせる」という実目的を生成できればよい，ということである．「大風で土ほこりが立つ→土ほこりが目に入って盲人が増える→盲人は三味線を買う→三味線に使うネコ皮が必要になりネコが殺される→ネコが減ればネズミが増える→ネズミは桶をかじる→桶の需要が増え桶屋が儲かる」，という因果の鎖であるが，これを遡る能力が必要となる．今風の表現をすれば「Deepアブダクション」と呼ぶべき方法であろうか．そして，この課題については，メタプランナが実目的を生成するために発見・抽出する因果の鎖が，必ずしも人にも理解できる表現である必要はない．これはDeep Learningなどの可読性のない手法での議論と同じである．因果の鎖が長くなればなるほど人には難解となる．もちろん，メタプランナにせよDeep Learningにせよ，入力に対して出力があるということはそれぞれの内部においては因果関係がちゃんと計算されているということである．しかし，人が理解できる保障はない．この場合，人が理解できる形に表現を変換したり，場合によっては省略や人が理解できるようにゆがめるといった新たな作業が必要になる．省略などしたらシステムがしていることを100％理解することができない，というご指摘もあるかと思うが，もはや人の理解力を超えたレベルでの作業を人が100％理解することはできないのであるから，致し方ないといえよう．

7. どうやって実現するのか

そもそも人を超えるようなモノを人がどうやってつくるのか？ いわゆる工学的な方法であるトップダウン型の方法では難しい．トップダウン型の設計では，まず完成させたいモノをイメージし，それを分割し，個々のパーツを組み立て合体させる．当たり前であるが，人を超えるモノを具体的にイメージすることはできないわけで，この方法では人を超えるAIの実現は難しい．もう一つの方法がボトムアップ型である．まさに生命はこの方法で進化してきた．ボトムアップ型では上記トップダウン型におけるパーツレベルがまず先に設計される．あとは個々のパーツの相互作用や自己組織化など，パーツ同士のインタラクションによる創発に委ねるのである．この方法であれば人が設計するのはパーツレベルであっても，そのパーツ同士のインタラクションで個々のパーツの能力を超える能力の創発が期待できる．今年の本学会全国大会での松尾 豊氏の講演で興味深い講演があった[松尾 17]．進化という方法は，要はいろいろな可能性をランダムに試し，より良い方法を発見することから離散的な方法であるのに対し，Deep Leaningや情報処理技術での山登り法などは微分的な方法であり，そのほうが効率が良い．そこで，従来は離散的な処理しかできなかったデータ処理などをどんどん微分可能にすることでより科学技術が進化するという内容である．さすがである．これに対し，中島秀之氏から微分だけだと局所最適に落ちるのでは？ という指摘もあり，これもそのとおりだと思う．進化はあえて離散的な方法を選択したのかもしれないが，効率が悪い，人類が手にした微分的方法とうまく組み合わせることができるかどうか，ということかもしれない．

8. 人が試されるとき

今後，AIの知的レベルは向上し，高い自律性を備えるAIも実現されるであろう．そのようなAIに対しても「道具をどのように利用するかはユーザの問題」という姿勢は許されるのであろうか？ 包丁はもちろん料理のための道具であるが，人に危害を与える能力ももち合わせている．では，包丁職人は，包丁をつくる際に，料理には使えるが悪用されない仕掛けを組み込む必要があるのであろうか？ これまでの科学技術においてマリ・キューリーにせよ，ロバート・オッペンハイマーにせよ，アラン・チューリングにせよ，皆葛藤があった．真理を追究したいという純粋な科学者としての立場と，それが悪用されたときの影響に目を向けるかどうかの葛藤である．これまでの歴史において唯一人類が追求を踏みとどめているのは「人のクローンをつくってはならない」という事例のみだという．命に対する畏怖であり，神の領域には立ち入らないということである．

しかし，科学技術はそもそも人が使う道具という立場であること，そして生命とは明らかに異なる素材であることから，命に対する畏怖の感覚を抱くことが難しい．もちろん高汎用型AIは人に大きな利便性をもたらし，実際に社会に浸透していくであろう．しかし，必ず反社

会的に利用することを考える人間が存在することも事実である．高い能力をもつ技術はそれが悪用や想定外の事態となってしまったときの影響も大きい．3.11 をはじめ，特に日本はそのことを思い知っている．現在，総務省にて議論されている「AI ネットワーク社会推進会議」における AI 開発におけるガイドライン策定に向けた動きや，世界規模な組織である非営利団体 Future of Life Institute（FLI）など，AI を平和利用するための開発指針策定に関する世界的な動きが加速していることは好ましい流れであり，もちろん，この動きに対する期待は大きいものの，我々がまだ実際に目にしたことがないモノに対しての対策は難しい．我々が試されるときが迫っている．

9. 開発指針と倫理

上記「AI ネットワーク社会推進会議」は著者も委員として参加させていただいており，皆が納得できるガイドラインの策定を目指して議論を重ねている．この会議は AI 研究者だけでなく，法律や知的財産の専門家，IT 企業トップなどさまざまな委員から構成されており，それは多様性の観点からは正しいものの，「人工知能」に対する捉え方がバラバラな人を集めて「AI 開発指針を決める作業をせよ」という，かなり無茶な話でもある．会議の最初の頃は，本稿前半でも述べたように，Deep Learning の延長線に高汎用型があるという見方の方もおられれば，この開発ガイドラインは用途限定型に特化したほうがよいと主張される方がおられたり，そもそも，高汎用型 AI に必須な「自律性」ということについてはほとんど認知されていなかったと思う．また，制御可能性や透明性の確保についてもその必要性が当たり前のように主張されていた．我々が使うシステムの制御可能性や透明性を確保することは当たり前だし，その必要性を問われれば誰もが 100％必要と答えるであろう．しかし，すでに Deep Learning にしても，人が理解できるレベルでの透明性はなく，高汎用型 AI を 100％制御することはできない．そもそも制御できる保障がないのが高汎用型なのである．つまり，これから開発される AI が制御可能性も透明性も確保できないのだ，という認識を皆が共有することが重要な作業だったと思う．そのうえで，そのような AI を開発するための指針をどのように考えるかという建設的な議論がされている．現在は AI に対する共通認識のもと，最初の頃のガイドライン案とはかなり変わり，皆が納得できる内容に精緻化されつつある．議論の経緯やガイドライン案の具体的な中身については「AI ネットワーク社会推進会議」ホームページ[*1] を参照いただきたい．

以下は著者の試案であるが，100％の制御可能性が保障されない高い自律性をもつ AI をどのように市場に投入すればよいのであろうか？　高汎用型 AI は人にとってはこれほど便利な機械はない．しかし，万に一つでも誤動作したら，という不安があっては安心して使用することはできない．ずばり，ある小さな島全体を汎用型 AI 特区などとし，可能な限りのあらゆる状況にて長期間にわたる徹底的な実環境での動作テストを行うといった箱庭的な方法しかないと考えている（Jurassic Park をもじって AIssic Park とでも呼ぼう）．

また，開発指針と同じく議論され始めているのが AI 倫理である．特に本学会倫理委員会において深く議論されており，倫理指針といった文書もとりまとめている．山川編集長と著者も倫理委員会[*2] のアドバイザを拝命しているが，特に指針 9 条「（人工知能への倫理遵守の要請）人工知能が社会の構成員またはそれに準じるものとなるためには，上に定めた人工知能学会員と同等に倫理指針を遵守できなければならない．」は，AI 研究者だからこそ生み出すことができた条文だと思う．AI 自体が倫理指針を守る必要があるという内容であり，これはまさに自律型 AI のことを意味していると著者は捉えている．

10. そもそも主役は人か知能か

日本におけるミスターシンギュラリティこと，宇宙物理学者にして神戸大学名誉教授の松田卓也先生によれば，AI についての考え方で「宇宙派」，「地球派」という分け方がある．

AI は当たり前であるが人のために人が開発するものである，という立場が地球派である．これに対して，宇宙派は，主役を「人」ではなく「知能」と考えるのである．そして知能の目的はその知的レベルの向上である．この見方だと，人類が登場するまでにおいて，他の生物によりそれなりに知能のレベルが向上し，人類の登場によりレベルが格段に向上し始める．そして人類が生み出した AI，そして，人を超える AI の登場によりさらに知能のレベルが指数関数的に向上していくという流れである．人類は知能という主役がその知的レベルを向上させるある一時代を担ったという見方である．まさに宇宙的スケールであるが，安易に宇宙派的な立場を否定することもできない．本稿の流れに従えば，人が高汎用型 AI と共生できれば，AI は人を超えることなく地球派が正解となり，AI に支配される方向になってしまった場合は宇宙派の正解という見方もできるからである．地球派でありたいという自分と，知能がどこまで進化するのかへの興味から宇宙派を完全否定することはできないのが正直な気持ちである．

---

*1 http://www.soumu.go.jp/main_sosiki/kenkyu/ai_network/

*2 http://ai-elsi.org

11. 予　　言

最後に，今回いろいろ勝手に書かせていただいたが，山川編集長からは，それがいつ起こるのか？　といった時間的な予測についても書くように，という難しい指示がされている．本稿であれば，低汎用型AIや意識をもつ高汎用型AI，そして生物のように自己創出能力をもつレベルのAIがいつ登場するのか，ということかと思う．

まず，低汎用型AIはすべてがつくり込みであることから，技術的な壁は低く，現在の用途限定型AI開発の延長線として徐々に登場してくると思う．むしろ，実環境で多くのタスクをこなすためのハードウェア，つまりロボット技術の進化のほうが重要であろう．特に人の手の指の動きなどは極めて複雑でロボットとして実現するのは難しいと聞いている．高汎用型AIについては，本稿で述べたように，メタプランニングやマルチモーダルデータをネットワークとして処理する技術についての革新が必要なものの，こなせるタスクの数や自律性の高さは低く，おもちゃレベルであったとしても，真に自らの判断にて行動選択を行うタイプのAIが5年くらいで登場するのではないだろうか．もちろん，環境認識レベルや対話レベルなどの個別の能力も用途限定型AIの進化とともに向上するであろうから，高汎用型AIのコアモジュールに，これら用途限定型AIで高度化された各AI要素技術を組み込むことで，おもちゃレベルからいきなり実用レベルに進化させることは容易であろう．これが実現するのが10年後であろうか．これは，10年後にはAIと人の社会的地位の逆転が起き始めることを意味している．そして，自己創出系としてのAIの登場であるが，自己創出系としてのAIをそもそも人類が必要とするか，という問題もあるし，開発ガイドラインしだいかもしれないが，20年後という予測としておきたい．自己創出系となると，もはやSFの世界と思われるかもしれないが，現在のコンピュータの骨格を発明したフォン・ノイマンが，なぜコンピュータを発明したのかというと，生物のように機械が機械を生み出すシステムをつくりたかったからなのである．今後，自己創出型AIが本当に実現するとすれば，それはフォン・ノイマンの壮大な夢がようやく実現されることを意味する．

12. 最　　後　　に

本稿を書き始めた段階では，あれとこれを書けば十分などと安易な考えでいたものの，書けば書くほど何をどう書けばよいのか悩むことになった．まだまだ書き足りないし，ストレスがたまった状態である．次回以降も続く先輩研究者の方々の考察が楽しみであるし，著者とは真逆の考え方かもしれない．それでこそ議論が生まれ，このレクチャーシリーズが起爆剤となり，今後のAI研究開発に対する好ましい取組み方が生まれる可能性もある．まずはトップバッターとして好き勝手に書かせていただく機会を得たことに感謝したい．著者のストレス解消はこれから登場する先生方にお願いすることにしよう．

◇ 参 考 文 献 ◇

[人工知能 16] 人工知能学会 監修：人工知能とは，近代科学社（2016）
[前野 10] 前野隆司：脳はなぜ「心」を作ったのか「私」の謎を解く受動意識仮説，ちくま文庫（2010）
[松尾 17] 松尾 豊：ディープラーニングと進化，第31回人工知能学会全国大会（2017）

2017年6月7日　受理

著 者 紹 介

栗原　聡（正会員）

慶應義塾大学大学院理工学研究科修士課程修了．NTT基礎研究所，大阪大学大学院情報科学研究科・産業科学研究所を経て，2013年より電気通信大学大学院情報理工学研究科教授．同大学人工知能先端研究センターセンター長．博士（工学）．（株）ドワンゴドワンゴ人工知能研究所客員研究員．人工知能，複雑ネットワーク科学，ユビキタスコンピューティングなどの研究に従事．著書『社会基盤としての情報通信』（共立出版，2000），『人工知能とは』（近代科学社，2016），翻訳『群知能とデータマイニング』（東京電機大学出版局，2012），『スモールワールド』（東京電機大学出版局，2006）など．（前）本学会理事・編集委員長．