

特集

1.

人と AI との協働から共生への転換

Scenario for Upgrading from a Collaboration to a Symbiosis Relationship
between Humans and AI

栗原 聡

abstract

第3次 AI ブームでの Deep Learning への熱狂も落ち着き、着々と社会実装が進み始めたように思えた 2021 年 11 月に ChatGPT が登場して世界は一気に新たな展開に突入することになった。AI の民主化と汎用 AI への道を開いた生成 AI の進化はロボット制御への応用など、まさに加速的に進展中である。人と AI が共に働くとはどうあるべきなのか？ 現在の道具としての AI から、今後の人と共生する次世代の自律型 AI への進展において、人の AI に対する信頼はどのように変化するのか？ あるべき人と AI の協働について考察する。

キーワード：生成 AI、自律型 AI と道具型 AI、AI との協働、人と AI の共生

1. 終わらない第3次 AI ブーム

AI という名前が誕生して 70 年近くが経過し、過去 2 回の AI ブームとその失敗による AI 冬の時代を経て、その間に着実にコンピュータは劇的に性能を向上させ 1980 年代に誕生したインターネットが大量のデータを容易に集めることを可能とした。また、過去 2 回の AI ブームも決して失敗に終わったのではなく、2000 年以降に始まる第 3 回 AI ブームをけん引する様々な機械学習法の基本的理論や手法が多数生み出されていたのである。現在のブームはこれらの多くの種が具体的にビジネス活用できるように発芽し成長できるようになったことによる盛り上がりとなっている。中でも、Deep Learning が可能とした表現学習能力は大きなブレークスルーであり、当初は画像や音声認識といった識別能力において注目を集めたが、2017 年に誕生した Transformer により、言語といったストリームデータに対する高精度な表現学習が可能となり、これにより ChatGPT をはじめとする LLM が誕生することとなった。画像生成においても GAN が終着点かと思いきや拡散モデル開発され、LLM を組み合わせることで、言葉から高いクオリティの画像を生成する Stable Diffusion といった手法が生み

出されるに至っている。

2. AI の民主化がもたらした光と影

ChatGPT が登場する以前の AI は技術を持つ限られた人しか利用できなかったのに対し、生成 AI は誰もがこれを使い熟すことを可能としたことから AI の民主化が実現されたと評される。この効果は絶大であり、AI を活用した多様なイノベーションを誰もが起こせるようになったことを意味する。無論、IT リテラシーが低くても、AI を利用したモラルに反する行為も可能としてしまった。誰もが簡単に質の高いフェイク画像や、一見信じてしまうようなデマ情報を容易に生成できるようになり、既に大きな社会問題となっている。AI が広く社会に進出することは仕事を奪うのではないかという懸念も生み出している。ロボット研究開発に目を向けて見ると、協働ロボットというキーワードをよく聞くようになった。生成 AI の応用範囲は広範で、既にロボット制御に生成を組み込む研究なども着々と進んでいる。人も技術もひたすら先に進みその歩みが止まることはなく、「人の社会」から「人と AI やロボットが共生する社会」に変容していくという大きな変化が、生成 AI の誕生により明確に始まったのだと考えている。しかし、いきなり共生関係に至るわけではなく、どのような過程を経ることになるのであろうか？ 最初の段階が人と道具型 AI との協働である。

栗原 聡 正員 慶應義塾大学理工学部管理工学科
E-mail satoshi@keio.jp
Satoshi KURIHARA, Member (Faculty of Science and Technology, Keio University, Yokohama-shi, 223-8522 Japan).
電子情報通信学会誌 Vol.107 No.5 pp.386-390 2024 年 5 月
© 2024 電子情報通信学会

3. 協働ロボットは人と協力して働いている？

弁当中身の具材を詰める流れ作業を、人とロボットが並んで行う動画を見たことがあるかと思うが、協働ロボットの例として紹介される典型例である。それまで人が作業している空間にロボットが入っての作業となることから人間形のようなデザインも必要であろうし、ロボットの開発や制御において大きな一歩であることは間違いない。ただし、ここでの協働とは同じ空間で人とロボットがお互いの干渉なしで作業するというレベルである。別の言い方をすれば、ロボット化された空間に人が入り込む構図とも言える。

超少子高齢化が進む日本においては、第一次産業を始め高齢化に伴う働き手の減少が著しく、移民かロボットの投入しか選択肢がないのが現状であろう。ロボットへの期待は高いものの、それは田畑等の人の働く現場をロボット投入により自動化するという意味ではないはずだ。ロボットが淡々と作業する場所で人が点在して淡々と作業する。そこに会話などの人間同士のやり取りもないというのはSF作品によく登場するネガティブな未来そのものである。

人と道具型AIの協働は、人間社会にロボットやAIの方が入ってくるという意味での人間中心でなければならぬと思う。郷に入っては郷に従えである。ロボットも人と同じく場の空気を読み、人の意図や感情をロボットなりに推定しての先回りした気配りとしての行動ができる必要があるし、あらかじめ決められた行動だけでなく動的な状況の変化への柔軟な対応能力も必要である。そのためには、ロボットに搭載するAIにおいても大きなアップグレード、すなわち現在の第3世代から第4世代への進展が必要となる。すなわち道具型から自律型(目的志向型)への進展である。

4. 自律型・目的志向型とは？

現在において身の回りにあるAIも他のIT技術と同じく全て我々が操作する道具として構築されている。これに対して、AIが自律性を持つということは、我々人と同じように自らの意識・意思^(注1)を持ち、自らの判断で行動できることを意味する。自ら考えて動作するAIの実現は、人を知るという意味において重要な研究テーマあるとともに、人間社会の持続と発展の観点においても必要不可欠な挑戦である。

超少子高齢化時代が到来する日本において、介護や労働力としてのAIやロボットへの期待は大きい。その際、決められた動作しかできないAIやいちいち命令しないと動作しないAI、そして状況を理解して空気を読

(注1) 人と同じ仕組みで意識を持つという意味ではない。

んだ行動ができないAIでは、社会に溶け込み人と共生する関係には到達できない。AIが目的(目的は人が埋め込む)に従い、状況に応じてどのように振る舞えばよいかを自らが判断し、人から指示がなくとも能動的に先回りして動作してくれる方が圧倒的に便利である。そのようなAIとであれば結果的に信頼関係も構築でき、人とAIとの共生関係も構築できると考える。

5. 人とAIが共生する社会での協働とは

目的を与えられ、それを達成するために重要なのはAIの行動選択能力である。目的といっても、より抽象度の高い目的(メタ目的)を達成できることが望まれる。AIはメタ目的を状況に応じた具体的な目的(副目的)に置き換える必要があるが、具体的な副目的は複数存在するであろうし適切な選択能力が求められる。また、同時に複数の副目的の達成を考慮する状況も起こり得るし、個々の副目的達成までの期限や優先順位も動的に変化することも想定される。何より行動選択の多様性はAIを搭載するロボットの身体性に依存する。例えばAI搭載掃除ロボットが、「掃除をせよ」というメタ目的が与えられたとしても、そのロボットの汎用性が高く、床に掃除機をかけるだけでなく、風呂掃除から食器洗い、窓拭きなどの多くをこなすことができるとしたら、そのAIはより多くの副目的を扱うことになる。

状況に応じて適切に即応的若しくは熟考的行動を選択する方法の代表はサブサンクションアーキテクチャであるが、これを多目的行動選択メカニズムと統合するようなイメージである。即応と熟考については、心理学の研究分野にて提案された考え方である「早い思考であるSystem1」と「遅い思考であるSystem2」の考え方とも類似している。通常においては、あらかじめ用意された達成するための具体的な段取りが学習済みの副目的群の中から、適切なものを選択することで大方の状況に対応はできるのだと思うが、想定外の事態は起こるもので、これに対応できる高い汎用性が社会に進出するAIには求められる。

想定外への対応には二つの可能性が考えられる。一つは想定外といっても、あらかじめ用意した副目的達成のための対応策に含まれていなかっただけという意味での想定外である。学習により獲得されている膨大な知識の中から想定外の事態に対応する方法を新たに探索することで対応できそうである。もう一つが厄介であり、AIの知識からでは導出できない対応策が必要となる場合である。我々人類も後者への対応は難しいものの、偶発的発見によりこれに対応してきた。

AIが後者の想定外に対応するためにできることとして、生物における進化での突然変異のようなやり方を組み込むことで、意図的に偶発的に新たな対応策を見つけ

ることができるかもしれない。ただし、この方法は倫理的な問題が大きい。仮に AI が自らの目的を突然変異的に書き換えた時点で、我々は AI の制御を完全に失う可能性がある。この部分については、現時点で人類が倫理的観点から踏み込まない人のクローンを作らないのと同等の慎重さが必要であろう。

道具型 AI の場合はそれを使う人間の方が明確な副目的を持ちこれを操作する。この使い方において「AI が何かをしてくれる」という他力本願的な関係はない。よって、道具型 AI に求められるのは信頼でなく精度である。しかし、自律型 AI の場合は、人が AI に持たせるのは抽象的なメタレベルな目的であり、AI 自体が能動的に状況に応じて具体的な副目的を設定し実行する。人からすると、「AI が何かをしてくれる」という任せ感覚を持ち、具体的な動作までは想像していないであろうから、具体的な動作をした AI を見たとき、AI に何らかの意識・意図のような感覚や、そのような具体的な動作を AI が自ら考えたのだという創造力のようなものさえ AI に感じるかもしれない。更にその動作が自分からの搾取がない満足する内容であると、人は AI に人同士と同等のレベルでの信頼感を感じるのだと考えている。

つまり、人と共生する AI は人間同士がそうであるように、お互いがお互いの状況を把握し、「今だから・ここだから・あなただから」というように、人からの命令に従うことなく先んじて実行できる能力が AI には求められる。この考え方は「いつでも・どこでも・誰とでも」のエビキタスの考え方と対比されるアンビエントの見方である。

次章では人が創造的作業を行う場合について考えてみる。道具型 AI であっても人の創造的作業のサポートを行い、その結果として人がより高い創造性を発揮することができる。このサポートにより自分の満足するコンテンツが制作できた場合、それまでの自分よりも高い創造性を発揮できた過程と結果から、道具型 AI に対しても自律型 AI のような信頼感を覚えるかもしれない。となれば、創造的作業において人のサポートする AI が自律型の場合は更に強い信頼感を覚えることになる。

6. 創造的作業においての人と AI の共働

6.1 創造することとは？

現在の AI の用途は主に効率化である。生成 AI に関しても要約や文章生成など短効果についての言及が多い。しかし、効率化自体は生み出して先に進むための仕組みではなく、効率化し尽くされればそれで終わりである。本来あるべき AI の用途、それは先に進むための創造的作業での活用であり、イノベーションに寄与する活用である。

では、生成 AI の登場で気になるのが、AI も人のようにゼロからのイノベーションを起こせるのかということであるが、現在の学習型かつ道具型の AI ではゼロからの創造は無理であり、そもそも我々人もゼロからの創造は簡単ではない。ChatGPT に何か問い合わせれば自分では思いつかない回答をどんどんしてくれる。いかにもこちらからの問いに対して創造性豊かであるという感覚を持ってしまいが、学習の際に獲得した膨大な知識からもっともな出力をしているだけである。とはいえこちらが思いつかない回答がどんどん瞬時に生成されるので創造性を感じてしまうのは致し方ないところではある。

生成 AI から学習された範囲外のものが出てくるわけがない。これは我々においても同様であり、人も生まれてから五感を通して入力される情報にて学習されたこと以外のものを創造することはできない。では、これまでの人類のゼロからの創造や発見による進歩が説明がつかないということになるが、偶発的イベントがこれを可能とするのである。ノーベル賞クラスの発明において、その発明に至るきっかけについての質問において、よく聞かれるのが「失敗からの偶然の発見」である。偶発的に起こる失敗は、失敗などではなくゼロからの創造や発見を生み出す重要な出来事なのである。しかし、このような発見はそう頻繁に起こることではないし、では日常でのイノベーションとは一体何なのか？

プレストにより新たなアイデアが生まれることはよくある話であるが、これは自分では思いつかないアイデアが、プレスト相手からのヒントと結びついて生まれたのである。つまり、ノーベル賞級の発見が、アイデアの種自体が生まれることだとすると、プレストでのアイデア創出は、自分が持つアイデアの種と相手からのヒントという種を結び付けたことで生まれた新たな種ということになる。つなぐことで新たな価値を生み出すイノベーションこそが日常におけるイノベーションと見立てると分かりやすい。

今のところ、つなぐことで新たな価値を生み出す能力は生成 AI にはない。しかし、膨大な数の種が詰め込まれていることから、人間のプレスト相手をはるかにしのぐアイデアの種をどんどん人に提供してくれるのである。これは人の創造力を引き出すツールとして大いに魅力がある。そこで、ドラマや小説、マンガからプロモーションといったストーリー性のあるコンテンツの制作を題材として人の創造的作業に対する AI サポートの可能性を追求する研究プロジェクトを、NEDO の人と共に進化する次世代人工知能に関する技術開発事業において「インタラクティブなストーリー型コンテンツ創作支援基盤の開発」というプロジェクトとして実施中である。

どのように AI を活用することで、人の創造的作業のサポート役としてクリエイターに認めてもらえるのかを追求している。その実証実験の一つとしてマンガの制作

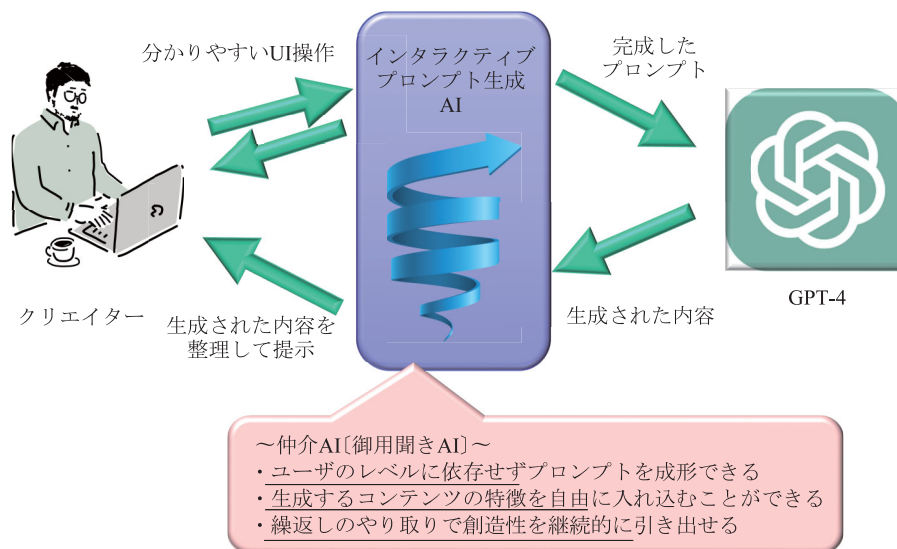


図1 御用聞き（仲介）AI

を具体的なテーマとして、実際にプロのクリエイターが筆者らの開発するAIを利用することで、新作の制作にどれくらい寄与するのかを確かめるTEZUKA2023を2023年6月に開始し、この11月に新作が完成したところである。TEZUKA2023にて制作したマンガ^(注2)やその作業の詳細は各種メディアにおいても広く取り上げて頂いているところである。以下簡単にどのようなAIを開発したのかについて概説する。

6.2 TEZUKA2023での取り組み

新作マンガの制作であることから、新たなストーリーと新たな登場人物のデザインが必要となる。ストーリーの生成にはGPT-4を、キャラクタ画像の生成にはStable Diffusionに追加学習を施したものを利用している。双方共やっていることはプロンプトエンジニアリングそのものである。

ストーリーの生成においては、小説など書いたこともない人が簡単なプロンプトでそれなりのしっかりした小説を生成できるといったことでも注目される生成AIであるが、だからといって、プロの脚本家が感嘆するレベルのストーリーを生成するとなると簡単ではない。緻密なプロンプトが要求されることになる。ChatGPTはAIの民主化を実現したと評されるが、その高い性能を引き出すのは実は難しい。その意味では「AIの民主化」という表現は間違いではないものの、「誰もが使える」ではなく「AIを利用することでより先に進めるはずの人ならリテラシーのレベルにかかわらず誰もがAIの能力を容易に使いこなすことを可能とした」という意味での

民主化」が正しい表現ということになる。

脚本家はAIの専門家でもなくプロンプトエンジニアでもない。例えば、シナリオの生成においてはプロの脚本家がうなるようなレベルのシナリオを生成させようとすると、場合によっては数千字にもなるプロンプトとなる。これをいちいち脚本家に整形させるのは無理であろう。また、本プロジェクトでは手塚治虫のブラックジャック的ストーリー展開であったりキャラクタ画像を生成させる必要がある。手塚治虫のキャラクタ画像を生成させるだけであればStable Diffusionへの追加学習で対応できるものの、制作するストーリーに登場する人物のイメージに合う画像をうまく生成させるためにもプロンプトの工夫が必要となる。

そこで考えたアイデアは特に奇抜なものではなく、生成AIと脚本家であるユーザの仲介を行う御用聞きAIを用意することである(図1)。分かりやすいUIも実装した。これによりインタラクティブなクリエイターとのやり取りを通して、緻密なプロンプトを仲介AIにて整形することを可能とした。仲介AIを用意することで、クリエイターのAIリテラシーに依存せず安定した品質のプロンプトを整形でき、今回であれば手塚治虫的ストーリー展開や登場人物の特徴を仲介AIにあらかじめ入れ込むことで、目的とする作家らしいストーリーやキャラクタの生成も可能となった。そして、以下も重要な観点なのであるが、生成AIに対するモラルに反した利用の可能性に対して、仲介AIを用意し仲介AIが整形するプロンプトの部分を隠蔽することで、プロンプトの模倣やモラルに反する利用をある程度防止するメリットもある。

実際の作業において、クリエイターはAIとのやり取りを何度も繰り返し行いつつ満足のいくストーリーを完

(注2) プロジェクトHPで読むことができる。(https://sites.google.com/view/nedocontents/)

成させるに至り、AIによる人の創造性を引き出す能力において大いに期待するという回答を得ることができた。また、印象的であったのが、AIをアシスタントや友人のように感じたことである。アシスタントは自分の言うとおりに作業する役割であるが、友人は自分と同等の立場でやり取りする立ち位置であり、AIを友人のようだと感じたことは、人とAIとの共生社会を構築する上で重要な指摘である。

7. 人とAIが共生してこそその前進

人の進もうとする欲求はテクノロジーを進展させ、生活の質を向上させた面もある。一方、それは格差を生み、地球環境を急激に変容させ、限られたリソースを巡っての争いを激化させ、最後は対話での解決というものに本当に人類の自助でこの状況を打開できるのかについては、誰も口には出さないものの懐疑的になっているのでは？ というのが実態であろう。打開策はあるのか？ 一つの可能性はやはりAIにあるのだと思うわけである。

人の創造的作業を道具型AIがサポートする場合において、確かに創造力は増強されるわけであるが、人の創造力の枠の範囲を大きく越えるものを生み出すことはやはり難しいのだと思う。どんなにテクノロジーが進化しようとも、進化させるのは人である。

しかし、創造的作業において、人と同じく自律性のあるAIがその作業に関わる場合は相乗効果による大きな創造性が発揮されるかもしれない。自律型AIが実現された場合、それは与えられるメタ目的を達成するための多様な方法を膨大な知識から発見することができ、人にとっては思いも寄らない自律型AIならではのアイデアを提示することができるからである。ただし、膨大とはいえこれまでの集積された知識内でのアイデアでは、真に想定外の対応策が生まれる可能性は低い。やはり究極は5.での進化的能力を組み込んだAIを実現することになるが、これをそのように構築し運用するのは入念に考える必要があるものの、最後はここに突入しなければならない状況が来るのかもしれない。この先、人類が自律型AIとの共生により新たな創造力や社会の有様を生み出すことができたとしたら、そのときは人類としてのバージョンが現状の1.0から2.0にアップグレードすると見てもよいのではないだろうか。

8. 人とAIとの共生社会の受け皿としての日本

最後に、現実の議論はどうなっているかという点、最先端AIの開発や運用に関するガイドラインの策定の動

きが国際的に活発化している。正しいAIの理解に基づく議論となっていればよいのであるが、現実にはミスリードのままの議論も散見される。その意味でも研究開発に関わる研究者・技術者と法制度の整備に関わる学者や専門家との密接な議論がこれまで以上に必要となる。

また、国や地域による人とテクノロジーの関係に対する見方の違いにも留意する必要がある。例えば、EUにおいてはAIやロボットはあくまで人が使う道具のままであるべきという考え方が強い。特にEUのAI規制法において最も危険で開発すべきでないというランクに位置されるAIがサブリミナルやプロパガンダといった人の個性や見方、考え方を強制的に操作するAIである。しかし、特に産業革命以降、インターネットが登場し、SNSが登場し、テクノロジーの進化に呼応するように人の見方、考え方は大きく変化させられてきた事実はどうのように解釈すればよいのか。もちろん、テクノロジーに変化させられてきたわけであるが、これまでのテクノロジーはプロパガンダのようにテクノロジーが意図的に強制的に変えようとするものではない。

確かにサブリミナルやプロパガンダは受け入れ難い。しかし、既に気が付かれているかと思うが、日本における誰もが知るマンガ「ドラえもん」において、ドラえもんはのび太君をちゃんとした人間に変えるために未来からやってきたのである。明らかにEU規制法における最も危険なAIということになる。

興味深いのは、ドラえもんのような作品が日本で誕生し、皆がそれを抵抗なく受け入れる社会こそが日本社会であるという事実である。EU社会と日本社会のこのような違いについては宗教の観点や島国であることなどいろいろ説明があるが、人とAIが共生する社会の構築において、日本の方がAIとの親和性が高いことは間違いない。しかし、現在の法体系ではそれが難しい。日本がAIにおいて世界をリードする存在となり、AIと共生することで人類の自助では解決できない問題に対して先に進める可能性があるのだとしたら、その可能性が日本ならではの社会において期待されるのだとしたら、早々に法体系のアップグレードが必要である。まさに国民的議論の活性化が望まれる。

(2023年12月1日受付 2023年12月30日最終受付)



くりはら さとし
栗原 聡 (正員)

慶大・理工・教授／共生知能創発社会研究センターセンター長、博士(工学)、NTT基礎研究所、阪大、電通大を経て、2018年から現職。JST さきがけ「社会変革基盤」領域統括、人工知能学会副会長・倫理委員会委員長。著書「AI兵器と未来社会キラーロボットの正体」(朝日新書)、編集「人工知能学事典」(共立出版、2017)など多数。