

AIと安全工学

— AIの歴史と軌跡 —

横浜国立大学 准教授

杉本 千佳 Chika Sugimoto

1. はじめに

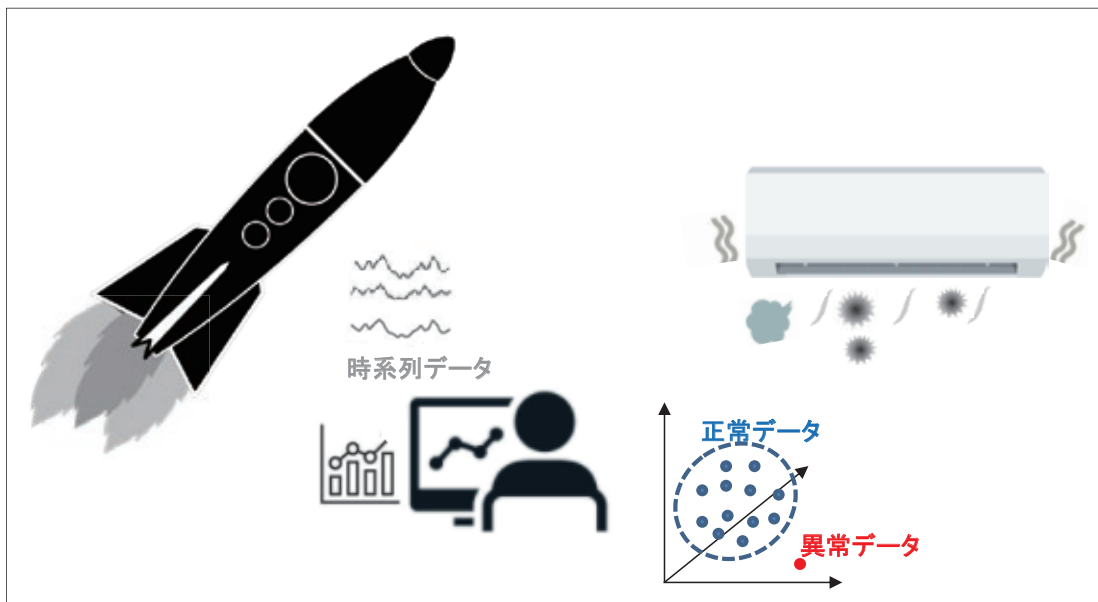
人工知能（AI）という言葉はすでに社会に浸透し¹⁾、エレクトリックコマース^{*1}（EC）、製造業、金融、医療など、多くの産業分野でAIが実際に活用されている。IoT（Internet of Things）の進展とともに多種多様なデータの取得・蓄積が可能になり、データの共有や活用が進む中でAIの重要性が増している。こうした動きは様々な産業に広がっており、多方面の領域にわたって安全の実現を目指す安全工学の分野でも、AIの重要性は高いと考えられる。

では、AIをいかに活用すればよいのか。AIは画像認識による異常検知・予知保全や、大量データに基づく異常値検出が得意である。まさに安全工学に必要とされる技術であ

り、道路やトンネル、橋や線路などの異常検知や異常予測、製品の製造における品質検査による不良品チェックや製造工程での機械などの異常・故障検知、金融取引での不正検知などに応用されている。自動車産業では安全な自動車部品のデザイン設計にもAIが活用されるなど、AI技術の導入が進んでいる。世界的な開発競争が行われている自動運転技術では、AIは脳の役割を果たす根幹技術として、安全確保に重要であることは広く知られている。

安全工学の対象範囲は宇宙開発から身近な生活領域まで幅広く多岐にわたる。これまで積み重ねてきたデータの蓄積や経験値があると考えられるが、安全工学での安全性の調査・改善には、データの取得と分析・評価が重要といえる（図1）。自然環境や人工環境を考

図1 データに基づく異常検知・安全性評価



慮して多様な対象を多面的に分析するために、AIが有用と考えられる。では、未知のことや想定外のことが起こり得る中で、AIはどのような役割を果たすことができるのだろうか。

第1回目の本稿では、AI技術について具体的に説明する前に、AI活用が近年飛躍的に進展するようになった背景を解説する。

なお、関連する用語については**表1**としてまとめて最後に示す。

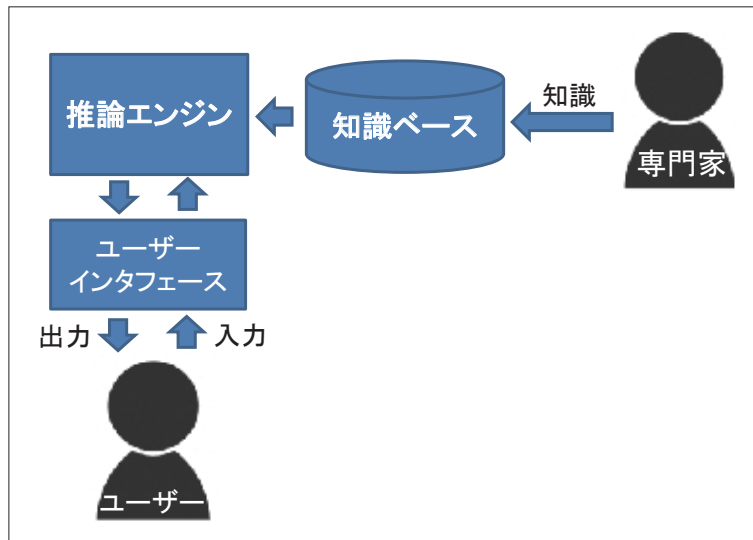
2. AIの歴史

AIの歴史の中で、AIブームとその後に続く冬の時代が繰り返され、現在はまさに第3次AIブームの中にあるとされる¹⁾。AIの概念を考えると、イギリスの数学者アラン・チューリングが1950年に出した著書「計算する機械と人間」に遡る。アラン・チューリングは、コンピュータの概念を初めて理論化した科学者であり、映画「イミテーション・ゲーム／エニグマと天才数学者の秘密」で描かれたように、ドイツの暗号機「エニグマ」を解読したことで有名である。その著書の中で「機械は思考できるのか」という問いを唱え、「機械」や「知能」といった言葉の意味を定義して考える代わりに、「機械が人間的であるかどうか」を判定するためにチューリングテストを考案した。このテストの具体的な方法の説明は割愛するが、注目を集めた2014年の疑惑の合格をはじめ、チューリングテストについては様々な議論がなされ、AIの発展に大きく寄与してきたと言われる。また、1956年に開催されたダートマス会議で、アメリカの計算機科学者であるジョン・マッカーシーが「人間のように考える機械」を「Artificial Intelligence」と定義し、AIに関する研究が活発化していく。これにより、1960年代に第1次AIブームが起こる。中で

も、1964年から1966年にかけてマサチューセッツ工科大学のジョセフ・ワイゼンバウムによって開発された自然言語処理プログラム「ELIZA」は有名である。しかし、この時代のAIは、コンピュータによる「推論」や「探索」の技術により、明確なルールが存在する問題に対してのみ高い性能を発揮するものであった。現実社会の複雑な問題を解くことができないという限界が見えると研究開発は失速し、1974年頃を境にAIは冬の時代に入る。

この後、第2次AIブームが起こることになるのは1980年代初頭になり、実世界の問題を解く実用的ツールとして多くのエキスパートシステム^{※2}が登場したことによる。エキスパートシステムは専門家の持つ知識を取り込み、その知識に基づき推論を行う専門家のように判断するコンピュータシステムである。**図2**に示すように、アーキテクチャ^{※3}は推論エンジン^{※4}と知識ベース^{※5}から構成され、推論エンジンが知識の集合である知識ベースを使って推論を行う。ブームより前の1970年代初めにスタンフォード大学で開発された、細菌感染を診断して適した抗生物質を推奨する「Mycin」が広く知られている。Mycinは優れた性能を示したにもかかわらず、医療現場では実際に利用されなかった。この要因は現代のAIの問題にも通ずるものであり、医療においてコンピュータ診断を利用することへの倫理・法律面の整備ができていなかったことや、技術に対する専門家や社会の受容性が低かったことにある。一方で、企業業務などに利用されるエキスパートシステムは多数開発され、広く商用利用された。安全工学分野でも、火災原因調査のための出火判定エキスパートシステム²⁾や工場防災診断エキスパートシステム³⁾などが開発されている。安全工学では、事故や災害の分析、リスクの調査・診断などにおいて、専門家の長年にわたる経験値を知識として蓄積し活用し

図2 エキスパートシステムのアーキテクチャ概略



ており、まさにエキスパートシステムが有用な分野といえる。このため、対象領域に応じて専門家のノウハウを知識ベースに格納することでシステムを構築し利用してきた。つまり、AIが活用されてきたといえる。しかし、お気づきのとおり、知識を言語化し明示的に表現して知識ベースを作成するのは大変な作業である。また、専門家が持つ知識には「暗黙知」が多くあり、全てを言語化することは難しい。コンピュータは人間と異なり、知識を与えられずに一般常識を理解することはできない。また、必要な情報を自ら収集・蓄積し、知識を学習により獲得することができなかった。このため、人が手動で常識レベルの膨大な知識をコンピュータに記述しなけりならなかった。また、知識が膨大になるとルール同士の矛盾や一貫性の欠如が生じ、当時のコンピュータはこうした矛盾や例外処理に対応できなかった。よって、知識を特定領域に限定して活用する必要があったのである。このような限界が明らかになり、ブームは終息しAI研究は再び冬の時代に入る。しかし、エキスパートシステム自体はその後2000年代になりビジネス分野で広く利用されるようになり、レコメンドシステム^{※6}などの我々に

身近なシステムとして現在も導入されている。

次のブームである第3次AIブームに至るまでには、大きな技術発展が二つあった。一つは機械学習の実用化であり、もう一つはディープラーニング^{※7}の登場である。第2次AIブームにおいて大きなネックとなっていた知識の獲得を、大量のデータを基に学習によりコンピュータ自身ができるようになったことがブレークスルーとなる。さらに、ディープラーニングにより知識を定義する要素を人手でコンピュータに与えることなしに、学習データから自ら特徴量を抽出し、認識や判断の精度を飛躍的に向上させることが可能になった。こうして、IoTによるビッグデータとAIの技術革新が、第4次産業革命のコアと呼ばれるに至る。

冬の時代が明けて現在に続くブームが来るまでには、多くの研究開発がなされてきた。1997年にIBMのチェス専用コンピュータ「ディープブルー」がチェスの世界王者に勝利した際は、コンピュータサイエンスの歴史に残る「世紀の対決」と言われ脚光を浴びた。さらに、2016年には囲碁プログラムの「アルファ碁」が世界トップレベルのプロ棋士に

初勝利した。囲碁はコンピュータが人間に打ち勝つことが最も難しい競技と考えられてきたため、世界的な注目を集め、AIの能力と有用性を広く世に知らしめた。一方、アカデミックでは2006年にトロント大学のジェフリー・ヒントンが深い層を持つニューラルネットワーク^{※8}、すなわち「ディープラーニング」の実用方法を提案する⁴⁾。2012年にはジェフリー・ヒントンらが世界的な画像認識コンペティション ILSVRC (ImageNet Large Scale Visual Recognition Challenge) でディープラーニングである深い畳み込みニューラルネットワークを用いて圧勝したことが有名だ。そして以降 AI は日進月歩であり、現在も世界中で AI 研究が進められ新たな AI 技術が生み出されている。

3. 今後の AI

アメリカの未来学者で AI 研究の権威であるレイ・カーツワイルは、2029年に AI が人間並みの知能を備え、2045年に技術的特異点、すなわちシンギュラリティ^{※9}が来ると提唱している⁵⁾。この問題は2045年問題とよばれ、AI が自ら人間より賢い知能を生み出し人類の知能を超えることによって起こるだろう様々な問題を指す。シンギュラリティが来る・来ないの議論は別にしても、AI は我々の暮らしを大きく変化させる可能性を持つ。こういった状況の中で、人間はどのように AI を扱い活用していけばよいのだろうか。唯一の解を導き出すことは難しいが、単に拒絶したり訳もなく恐れたりすることなく、AI を正しく理解し対応していくことが必要といえよう。新たな技術が社会に実装されるとき、制度が追い付かず倫理および法律面の整備ができていないことがある。AI についても判断とその責任、倫理的基準を引き続き議論していくことが必要である。

AI の分類の仕方の一つに、「汎用型 AI」と「特化型 AI」がある。人間のように広範な課題を処理できるかどうかという視点で分類されている。汎用型 AI は人間と同じように様々な課題を処理可能な AI である。既に示したように、レイ・カーツワイル博士はその著書で2029年にこの汎用型 AI が誕生すると述べた⁵⁾。一方、特化型 AI は限定領域の課題に特化して自動的に学習、処理を行う AI である。画像認識技術などが代表例であり、現在実用化されている AI はこの特化型である。

これに対し、アメリカの哲学者ジョン・サールは、人間の意識や知能を持つかどうかという観点から、「強い AI」と「弱い AI」という分類を提唱した⁶⁾。強い AI とは、人間のような自意識を備え、あらゆる認知能力を必要とする作業も可能な AI を指す。一方、弱い AI は人間の知性の一部のみを代替し、特定のタスクだけを処理する AI を指す。シンギュラリティではこうした強い AI が登場し、予測不可能な変化が起こることになるのだろうか。我々はこうした事態を念頭に置きながら様々な立場から考え、変化に柔軟に対応できるよう備えていくことが必要であろう。

4. まとめ

AI は今、ビジネスでの活用や様々な分野への導入が進められている段階にある。身近なところで AI が活用されるにつれ、利便性や効率性を享受する一方、人間の仕事や役割を取って代わられることや技術の信頼性に対する危惧が生まれる。AI によってもたらされる結果に対する工学の立場からの安全確保とリスク対応は、まさに安全工学の研究対象といえる。社会の安全・安心を確保するために、AI を活用したシステムの評価を進めるとともに、安全工学としての AI の活用方法を確立していくなど、安全工学が担う役割は

表 1 用語説明

No.	用語	説明
1	エレクトリックコマース (Eコマース、EC)	ネットワークを介して電子的に契約や決済などを行う取引形態のことで、インターネット上で商品やサービスを売買することの総称
2	エキスパートシステム	特定分野の専門家の知識をデータ化し、専門家のように推論や判断をできるようにしたコンピュータシステム
3	アーキテクチャ	コンピュータシステムの論理的構造や設計思想
4	推論エンジン	知識ベースの知識を利用し推論を実行し答えを導き出す仕組み
5	知識ベース	知識を組織化しコンピュータで検索・利用できる形式にしたデータベース
6	レコメンドシステム	利用者の興味のあるモノやサービスを推薦するシステム
7	ディープラーニング	ニューラルネットワークを多層に結合し、表現・学習能力を高めた機械学習手法
8	ニューラルネットワーク	人間の脳内にある神経細胞のネットワーク構造を模して数理モデルで表現したもので、機械学習手法の一つ
9	シンギュラリティ	AIが人間の知能を超える転換点、または、それにより世界に大きな変化をもたらされるという概念

大きいといえる。

参考文献

- 1) 総務省：人工知能 (AI) の現状と未来、平成28年版情報通信白書第1部、232、2016。 <https://www.soumu.go.jp/johotsusintokei/whitepaper/ja/h28/pdf/n4200000.pdf>
- 2) 山瀬敏郎：出火判定のためのエキスパートシステム その1 エキスパートシステムの概要、季刊消防防災の科学、No.017、46-50、1989。 https://www.isad.or.jp/pdf/information_provision/information_provision/no17/46p.pdf
- 3) 小暮英夫、他：工場防災診断エキスパートシステム、日立評論、70 (11)、1166-1170、1988。 https://www.hitachihyoron.com/jp/pdf/1988/11/1988_11_15.pdf
- 4) G. E. Hinton and R. Salakhutdinov : Reducing the dimensionality of data with neural networks, Science, 313, 5786, 504-507, 2006.
- 5) R. Kurzweil : The Singularity Is Near : When Humans Transcend Biology, Viking, 2005.
- 6) J. Searle : Minds, brains, and programs, Behavioral and Brain Sciences, 3 (3), 417-424, 1980.

すきもとちか

生体計測工学、知覚情報処理、アフェクティブ・コンピューティングなどの研究分野において、生体・行動・環境情報の認識とその応用に関する研究に従事。東京大学大学院新領域創成科学研究科助教を経て、2010年から横浜国立大学大学院工学研究院准教授。