

人工知能の現状と将来動向 (4)

人工知能の光と闇を考える



松田宏コンサルティング事務所
代表 松田 宏

最近の人工知能に関する動きは予想以上に速い。状況がどんどん変化するので、この連載を企画した時点とは様相が大きく変わってしまった。また、人工知能に関する情報が巷に溢れるようになり、今更解説をする必要性は低下している。そこで前回もお断りさせていただいたように、この「人工知能の現状と将来動向」は当初計画の5回連載を4回に短縮し、内容も変更させていただいた。今回はその最終回として、人工知能の「光と影」について、SF小説やSF映画など空想の世界も交えながら起こり得る未来像に関する思考実験を試みたい。楽観的にバラ色の未来像を夢見る人がいる一方、様々なリスクを危惧する悲観的な人もいる。ここでは両論を併記させていただくので、ご参考になれば幸いである。

＝人工知能によってバラ色の未来は来るか＝

人工知能の研究や応用システムの開発に携わる人達は工学系が多いせいか、バラ色の未来を思い描く傾向が強いようだ。脳は複雑でまだわからないことが多い、と謙虚な医学系研究者とは大違いである。

確かに人工知能の発達により高度な知能を利用できるようになればすばらしい世の中になるだろう。離島の小さな診療所では、若い総合医が複数の科の熟練した専門医グループと同レベルの診察をすることができるようになる。法律事務所では、司法修習を終えたばかりの新人弁護士が多くの判例を踏まえた裁判に勝てる訴状の作成ができるようになる。金融機関では、新人の窓口係員が資産家に対し金融商品の複雑なポートフォリオ投資を提案したり、経験の少ない融資担当者が巨額の資金と長年月を要する大規模開発プロジェクトの投資判断をしたりできるようになる。化学プラントの制御室では、何十年に一回の珍しいトラブルが発生しても、原因を瞬時に究明して最も安全な対応をとることができるようになる。ごくまれにしか発生しないトラブルは熟練の運転員でも経験したことがなく、的確な対応を取るのは難しいのだ。しかし人工知能ならプラントの設計思想、構造、動作原理、制御方法、長年の保守点検や改修の記録のすべてを「熟知」し、最適な対

目 次

- ▶人工知能によってバラ色の未来は来るか
- ▶人工知能を巡る小さな陰り
- ▶SF小説家が描く恐怖の未来
- ▶人工知能が学習する知識はすべて正しいか
- ▶人工知能と人間が共存するために
- ▶おわりに：あえて書かなかったこと

応ができる。人工知能の「光」の部分である。

最近の第三次人工知能ブームは本物だろう。コンピュータの性能の飛躍的な向上、膨大な情報を短時間で検索するビッグデータ関連技術の発達、専門知識を定式化する困難な作業（知識工学）に代わる深層学習による自己学習により、人工知能システムは実用化の域に達しつつあるからである。

そのため実用目的の人工知能システムの開発に莫大な資金が投資され、優秀な頭脳が集められるようになり、競争が過熱している。スマホの新しいアプリの多くは何等かの形で人工知能を応用しているし、多くの業界で人工知能による業務システムの高度化が積極的に進められている。たとえば、あるコンビニエンスストアでは人工知能とソーシャルネットワークシステム（SNS）を組み合わせ、常連客の買物履歴に基づいたお勧め商品の案内サービスを検討している。またあるベビーフード工場では毎日百万個のポテトを人工知能による画像分析で選別している。厳しい競争に勝抜くには差別化が必要なので、各社とも新サービスなどの開発に積極的なのである。

現在の人工知能にはまだ苦手な部分が残っているものの、多くの分野で熟練した専門家と同等の知的能力を発揮できるレベルに達しつつある。まだ知識・経験が不十分な専門職の新人が経験豊富なベテラン並みに仕事をしたり、人工知能システムが人間と同等の仕事を代行したりすることが当面の目標である。中にはすでに人工知能が人間以上の能力を獲得した分野もあるし、今はまだ人間にかなわないが、追い付き追い越すのは時間の問題だろうという分野もある。こうした面を見る限り、人工知能の未来はバラ色に見えるのである。

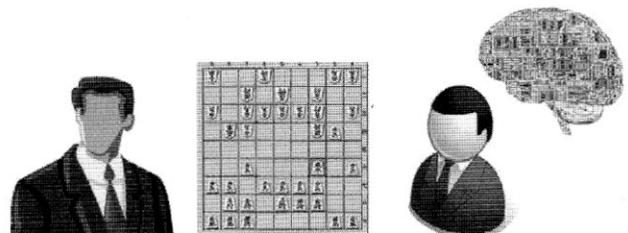
＝人工知能を巡る小さな陰り＝

しかし、バラ色の未来像に水を差すような「その後」のできごともある。 「シンデレラのその後」に例えられるように、ハッピーエンドの物語の後日談である。シンデレラの物語は「王子様と結婚したシンデレラは宮殿で幸せに暮らしましたとさ」と終わる。しかし、実際の結婚生活が幸せなものとは限らない。姑や小姑との関係、夫の女性関係、子供の教育と悩みは尽きないし、社交や儀礼も大変だ。王国に政変があるかもしれない。お姫様は結構大変なのである。以下に、人工知能のバラ色の未来像に差した「陰り」と思われる事例をご紹介します。

（将棋名人戦でのカンニング疑惑）

以前、人工知能ソフトが囲碁や将棋の名人に勝てるようになった話をご紹介します。しかしその後の対戦では名人が勝つこともあり、棋士はまだまだ強いことを立証した。人工知能と対戦したある名人は「人工知能は人間が思いもつかなかった手を使うことがある。これまでの常識とは違う新しい勝負の世界が開けるだろう」という感想を述べたそうだ。

ところが2016年10月にプロ棋士の将棋ソフト不正使用疑惑というスキャンダルが発生した。第29期竜王戦の挑戦者、三浦弘行九段が対戦中に離席した際に将棋ソフトを使用してカンニングを行ったのではないかと疑われたのである。対局中にトイレや食事



将棋の竜王戦でカンニング疑惑（イメージ図）

などで離席することは認められるが、三浦九段の場合は頻度が高く時間も長かったこと、終盤での指し手を将棋ソフトと比較したところ一致率が高かったという主張があったのだそうだ。最終的に第三者委員会が不正行為の証拠はないと結論づけ、公式戦出場停止処分も取消されて名誉が回復された。

しかし、騒動の責任を取って日本将棋連盟の会長と常務理事が辞任した他、幹部三名が解任された。この事件の真偽はわからないが、人工知能による将棋ソフトがプロの棋士の参考になるほどのレベルに達しているということだろう。

(自動運転自動車の死亡事故)

人工知能による自動車の自動運転は実用の域に達している。TVカメラやレーダー、超音波センサーなどにより道路状況や他車の存在を検出し、必要に応じて減速・停止する自動停止システムは既に広く普及している。無人運転も技術的にはほぼ完成しており、すでに私有地で道路交通法が適用されない大型観光施設などで無人運転の構内バスが運行されている。公道についてまだ法律上の問題が残っているので、自動運転車の走行実験には運転者が搭乗して状況を監視することが義務付けられている。米国では世界最大のネット企業グーグル (Google) 社の子会社が開発した58台の自動運転車が公道実験を続け、走行距離は350万キロに達したという。

ところが2016年5月に米国の電気自動車専門メーカーであるテスラ (Tesla) 社の運転支援機能を搭載した自動車 (厳密には自動運転車ではない) が死亡事故を起こした。ドライバーが乗った車が中央分離帯のある幹線道路を運転支援機能によって走行中、前方の交差点を左折しようとしていた大型トレー

ラーに衝突したのだ。原因は背景の空が明るく光っていて、ドライバーも運転支援機能も白い車体を認識できなかったためで、人間が運転していても同様の事故が起きたかもしれない状況だった。

自動運転は人間のようによそ見をしたり長時間ドライブで疲労して注意力が低下したりすることはないが、まだ完全ではない。雪道での走行はまだ不安定だし、運転が慎重すぎて他の交通の流れに支障をきたすという批判もあるそうである。

(受験ロボット「東ロボ君」の挫折)

以前に紹介した「東ロボ君」の後日談は、少々残念なものである。国立情報学研究所が2011年から進めていた研究で、人工知能に最難関と言われる東京大学の入学試験を受験させ、合格を目指すものであった。既に得意科目では偏差値57と合格レベルに達し、残り科目もあとわずかだろうと楽観視されていた。ところが、英語や国語の長文読解が苦手な成績がなかなか伸びず、ついに断念したものである。

限界を素直に認め、目標を断念したのは立派だと思う。国が多額の研究開発資金を投入したのに十分な成果が得られなかったプロジェクトは過去にもたくさんある。それが科学研究の難しさであり、公的資金でなければできない研究課題だということであ



東京大学本郷キャンパス (写真: Wikipedia)

る。計画通りの結果が得られる技術開発なら民間企業が投資できる。そうではない未知のもの、成功しないかもしれない（失敗するかもしれない）研究にこそ、公的資金を投入すべきなのである。昨今「すぐに役に立つ」研究開発にしか科学研究費（科研費）を出さない傾向が強まり、長期間を要し結果が予測困難な基礎研究が縮小を迫られ、将来の科学研究の低下が危ぶまれているのである。

（過激思想にかぶれた人工知能）

近年、インターネット上で巧妙に流布される過激な思想を純真な若者が無批判に吸収し、凶悪なテロ事件を起こしてしまうという問題が起きている。欧米で生まれ育ち「ホームグロウン」（home grown：自国育ち）と呼ばれる若者達である。国際テロ組織のインターネットによる資金調達も盛んで、軍事訓練キャンプの維持費や要員の旅費、武器弾薬の購入費をまかなう闇資金はかなり潤沢らしい。

現在の人工知能ブームのきっかけのひとつが、深層学習による自己学習である。そしてインターネットに接続すれば世界中から膨大な情報を収集して自ら知識を蓄積できるようになったのである。

しかしそこには意外な落とし穴があった。世界最大のネット企業であるグーグル社が膨大な資金と多数の研究者を投入して開発を進めていた巨大人工知能システムが、ナチス思想や人種差別主義、イスラム過激派のテロリズムなどの危険な思想に「かぶれて」しまったのである。問題に気が付いた研究者達は、楽観的すぎたこの壮大な実験を一時中止し、対策を検討中だという。

（余談：子供の教育環境と人工知能の学習環境）

中国の列女伝（れつじょでん）に記された故事に

「孟母三遷（もうぼさんせん）の教え」がある。孟母とは「性善説（せいぜんせつ）」を唱え、武力による覇道を批判して互惠関係や信頼関係に基づく王道政治（おうどうせいじ）を主張し「易姓革命（えきせいかくめい）」を唱えた思想家、孟子（もうし：紀元前372～紀元前289年）の母親、三遷とは住まいを三度変えたという意味である。孟子が子供の頃に住んでいたのはお寺の近くで、子供達は葬式ごっこをして遊んでいた。教育上良くないと考えた母親は市場の近くに引っ越したのだが、そこでは子供達は商売ごっこをして遊んでいた。ここも教育には良くないと気付いた孟子の母親は、更に学校の近くに引っ越した。そこでは子供達が学校ごっこをして遊んでいたので孟子も自然に学問好きになり、後年は大学者になったというお話である。「朱に交われば赤くなる」から、人工知能にも「悪い子と遊んではいけません」と注意しなければならないだろう。

＝SF小説家が描く恐怖の未来＝

（ロボットと人間の関係：ロボット三原則）

ロボットが人間に対して反乱を起こすのではないかという恐怖は昔からあり、フランケンシュタイン（Frankenstein）のような人間が作ったロボットが人間を破滅させる物語がたくさん書かれてきた。SF作家のアイザック・アシモフ（Isaac Asimov: 1920～1992）はそうした風潮に対抗し、小説の中でロボットが従うべき「ロボット三原則（Three Laws of Robotics）」という次のような原則を提唱した。

<第一条>

ロボットは人間に危害を加えてはならない。

また、その危険を看過することによって、人間に危害を及ぼしてはならない。

<第二条>

ロボットは人間に与えられた命令に服従しなければならない。ただし、与えられた命令が第一条に反する場合はこの限りでない。

<第三条>

ロボットは、第一条および第二条に反する恐れがない限り、自己を守らなければならない。

鉄腕アトムもこの三原則を踏襲しており、それに関連する場面が時々出て来る。ただし、この三原則は不完全なので、その矛盾を突く問題意識が多くのSF小説やSF映画の恰好の題材となっている。

(人工知能が人間に反抗するSF映画)

少し古いですが、冷戦時代にソ連の大陸間弾道ミサイル(ICBM)を監視している北米航空宇宙防衛司令部(NORAD)のコンピューターに天才ハッカー少年が侵入した結果、人間の指示に従わなくなって核戦争が起きそうになるというSF映画があった。

映画「2001年宇宙の旅(原題Space Odyssey)」は、惑星探査のための宇宙船に搭載された人工知能コンピューターHALが搭乗員を殺し始めるという物語である。HALは惑星探査ミッションについては乗員と協力するよう命令されていたが、密かに与えられたモノリス探査については乗員に隠せという命令も受けていた(人工冬眠中の科学者は知っていた)。HALは二つの指示の矛盾に耐えきれず異常をきたして暴走し、最後は自分を停止させようとする乗員を排除し始めた。乗員が死んでいなくなれば永遠に秘密を話さずに済み、惑星探査ミッションは

自分だけで遂行すればいいと考えたのだ。この物語の結末は、異常に気付いた船長が、声の聞こえない場所で乗員と密談しているのを読唇術で読み取るほど賢いHALとの知恵比べに勝ってHALを停止させ、生き残った乗員と探査飛行を続けるというものである。

映画「ターミネーター」シリーズは、人工知能が人類に反乱を起こしロボット兵器(ターミネーター)によって人類を絶滅寸前まで追い込むが、有能な指揮者の下に団結した人類が反撃する物語である。有能な人類軍指揮者に手を焼いた人工知能は、タイムマシンで過去(つまり現代)に表面が人間と同じ生体細胞で覆われ外見が人間と同じに見える潜入型の暗殺ロボットを送り込み、彼の母親を殺そうとする。母親を抹殺すれば子供である指揮者は消滅し、人類軍は弱体化するはずだというタイムマシン・パラドックスである。

(悪者が人工知能を使ったら)

私見では、人工知能が人間に反抗するよりも、悪人が高度な人工知能を犯罪などの悪事に使う危険性の方が高いのではないかと思う。天才的頭脳を持つ犯罪者、狂気の科学者、世界征服を企む謎の組織などは小説や映画の恰好な題材であった。しかし、何でも自分達でやれるわけではないので、高い能力を持つ人物を雇うこともある。マフィアのボスが法律を勉強するよりは腕利きの悪徳弁護士を高給で雇った方が効率的だ。今後は代わりに人工知能を使うようになるだろうから、それに対抗するには警察や検察がより高度な人工知能を装備する必要がある。人間よりも知能が高い人工知能が犯罪を企むのに対抗するには人工知能しかない、ということである。

(人工知能どうしの知恵比べ)

囲碁や将棋の棋士が人工知能と優劣を競うのも過渡的な現象かもしれない。人工知能が人間をはるかに凌駕するようになれば勝負にならないので、人間は人間どうしで楽しみながら腕を競うようになるだろう。陸上競技の選手は自動車と競走はしないと同じである。そして、人工知能と人工知能の知恵比べが自動車開発の技術力を競うF1レースのように行われるようになるだろう。

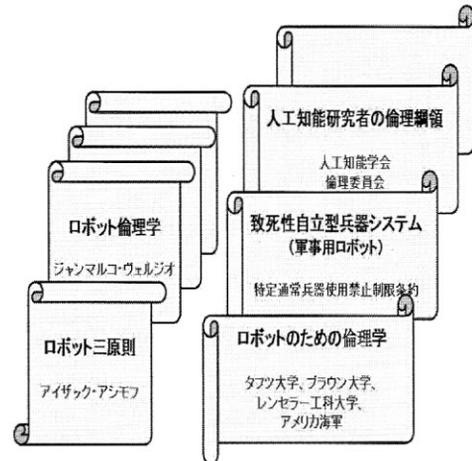
インターネットによる通信販売や予約、案内などのサービスでは人工知能を積極的に取り入れている。サービスの差が業績の差になるので、新サービスを開発できる巨額の投資ができる資本力の有無が企業の存亡を左右する場面も増えてくるだろう。

訴訟大国の米国では、損害賠償訴訟に勝つためには高い報酬を払ってでも実績が豊富な有能な弁護士を雇う。相手の弁護士よりもこちらの弁護士が有能でないと、法廷での論戦には勝てないからである。つまり、訴訟に勝てる弁護士には高い報酬を払うだけのメリットがあるのである。もちろんコンサルタントの場合も知恵の価値は同じである。「知恵は力」なので必要ならば買えばよい。投資した金は儲かれば返ってくる、というのはビジネスの論理である。

優秀な金融トレーダーは大きな利益を生み出すのでかなりの高給取りである。だからトレーダーをたくさん雇うよりは人工知能に任せた方が安上がりだという現象が起きている。株式市場や金融市場では以前から高度なシステム取引が行われており、短時間に莫大な金額が動くことによる相場の乱高下を制限する規則が定められているほどである。

人工知能が普及して社内のあらゆる場面で応用される時代には、公平性や倫理・道徳、安全などに関

する新しいルールが求められている。



ロボット倫理ルール化の例 (資料：松田)

＝人工知能が学習する知識はすべて正しいか＝

子供が勉強好きで、昔なら図書館に籠って、今ならインターネットを駆使していろいろな知識を熱心に学習すれば、賢明な大人に育つだろうと思いがちだが、世の中そう単純ではない。知識が多いだけでは知能が高いとはいえないし、世の中に流布している知識のすべてが正しいわけではないのである。

(歴史は歪曲され、都合のよい情報が残される)

古代中国を初めて統一した秦王朝は、政権にとって不都合な儒教関係の図書類を全て焼却させ儒学者を生き埋めにした(焚書坑儒)。古代日本でも、政府が編纂させた公式な歴史書である「古事記」と「日本書紀」以前の歴史的な記述はほとんど残っていない。運よく残った古代の伝承などの歴史資料は、体制派の学者達が後世の偽書だと決めつけ、無視されたという。歴史の政治的情報操作は珍しくない。

1930年代に書かれた英国作家ジョージ・オーエルの「1984年」という小説にも、強力な権限を行使して言論や思想を統制する正体不明の政府がビッグ・ブラザーと呼ばれる巨大なシステムで国民を監視し

ながら、過去のあらゆる記録を組織的に改ざんし、報道内容を操作し続けるという情景が出てくる。現在でも、国民のインターネットへのアクセスを嚴重に制限し、膨大な人数の情報要員を国費で雇って政策に沿った意見を大量に投稿させて世論を誘導している中央集権国がある。

(ネットには嘘や間違い、矛盾が溢れている)

インターネットが普及したことにより、個人が全世界に向けて情報を発信できるようになったが、故意に嘘の情報をばら撒くことも容易になった。

たとえばインターネットに意図的に読者の怒りや非難を買う投稿をして返信数を増大させ(いわゆる「炎上」状態)、アクセス数に比例する報奨金を稼ぐ無責任な個人や団体がある。

また、災害や事故の際に恐怖心や不安感をあおる根拠のない「うわさ」や「デマ」が蔓延することもある。インターネットは緊急時に即時性があり地域の貴重な情報伝達手段なのだが、悪意や偏見、憶測、興味本位の愉快犯的な悪戯による虚報などの弊害もあり、大きな社会問題である。

政治イデオロギーなどによる偏向や悪意により事実を検証しないまま憶測で書く報道が既存のメディアでも増えている。危機感を持った各国の報道機関とネット企業が共同で、報道内容の事実関係を検証する国際的な機構を設立した。いずれ「事実検証済(Fact Check)」と表示されたニュースが世界中で報道されるようになるはずである。

(いろいろな正義、複数の正解がある)

人の基本的な価値観は主に子供時代の教育に依存する。だから、親や担任教師のように信頼できる人の教えだという属人性、国が定めた道德教育カリキュ

ラム、宗教的な戒律、社会における伝統的な倫理感などが基礎になっている。つまり個人差があり、互いに矛盾した要素を含んでいる。

たとえばある宗教では酒や豚肉を禁じているが、他の宗教では酒を許し牛肉を禁じている。人を殺してはいけない、他人のものを盗んではいけない、というような普遍的な戒律については議論の余地がないが、根拠が神様の啓示なので、論理的な理由説明があるわけではない。神様が戒律で禁じられたのだから人間はそれをしてはいけない、法律で禁じられているからしてはいけないし禁を犯せば処罰される、という構造である。

だから、何が正しくて何が正しくないのかという問題は非常に難しい。人工知能に法律や規則、マナーなどの外面的なことを教えることはできても、より内面的な善悪の区別や倫理、道德を教え、それに反する知識を排除させるのは、子供がスマホやパソコンで有害なサイトにアクセスできないように設定するような簡単なことではないのである。

(少数意見が正しいかもしれない)

ある科学者の集まりで、参加者の大多数が賛成した学説に対して猛然と反対した学者がこう言った。「科学は多数決ではない。実験や観測だけが正しさを実証するのだ」と。そして、仮説の提唱から実証まで数十年、ときには百年もかかった理論はたくさんあった。ということは、ある時点で権威ある専門家が言っていたことが間違っていた訳である。

自然科学のように実験で正しさの実証が難しい分野では、さまざまな意見や主張がある。そして権威者や声の大きな者の言う事、多くの人の賛同を得たことが正しいとされ、少数意見は無視されてしま

う。自分達の考えを知ってもらうには情報発信が必要だが、それには多額の費用がかかる。

それがインターネットなら資金力のない小さなNPOの少数意見も全世界向けに動画で発信できる。日本でも地元の地方新聞が政治的に偏向した報道をしているので正すべきだと主張する民間団体や、TVや新聞などマスコミの誤報を検証して報道被害者を救済する民間研究所などが活動している。そうした小さな声を拾い出さないと、本当のことはわからないということがあるのである。

少数意見といえば、イスラム圏に関する偏見や誤解を解消し、正しい情報を世界に発信するためアラビア語と英語でニュース番組を放送するアルジャジーラ衛星チャンネル (Al Jazeera) というTV局が中東カタールのドーハにある。NHK-BSの「世界のニュース」で報道内容を抜粋して紹介しているのは中立性が認められているからで、設立の目的は達成しているようだ。ちなみに、NHK-BSの世界のニュースでは世界の主要TV局のトップニュースを紹介しているので、同じ事件に対する各局の取り上げ方、視点の違いが明確になって興味深い。

(偽情報に騙されない賢い人工知能)

人工知能が十分に「賢く」なれば、ネット上に溢れる膨大な、種々雑多で玉石混交の情報の中から価値のある情報を取捨選択できるようになるだろう。しかし、共通するもの、対立するもの、矛盾するものなどを探し出すことはできても、事実誤認や誤解、悪質な虚偽、歪曲、隠蔽などを的確に判別することは難しいだろう。人間でも難しいのだから。

主義や信条は自分達が正しい、正義は自分達にある、それ以外は間違っているとする傾向が強い。人

工知能にも倫理や道徳、哲学、宗教などによる価値観や判断基準、つまり単なる知識ではない「英知(叡智：①すぐれた知恵。深い知性。②【哲】真実在や真理を捉(とら)えることのできる最高の認識能力)」が必要になるはずである。

＝人工知能と人間が共存するために＝

(無くなる仕事、新しく生まれる仕事)

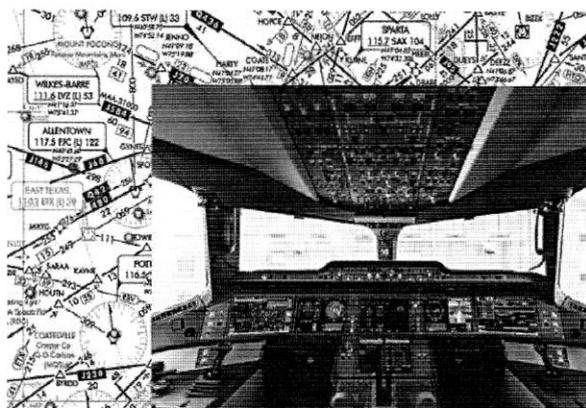
産業革命の際の労働問題を思い起こす人達は、人工知能の進歩によって多くの労働者が仕事を失うのではないかと心配している。紡績技術の進歩によって仕事を失った労働者が大挙して工場に押し寄せ、機械を打壊す事件が続出した時代があったからだ。

最近、複数の研究者が20年後を予測し、現在の職業の2割しか残らず、残りの8割は存在しないだろうと発表して話題になった。歴史上の経験によれば、産業構造が大きく変化する過渡期には様々な社会問題が起きるものの、長い目で見ると無くなった仕事の代わりに新しい仕事がたくさん生まれ、人々の生計は豊かになってきた。人間の筋肉に代わる機械が出現したときも、人間の頭脳に代わるコンピューターやそれが一段と進歩した人工知能が出現しつつある現在も、基本的な構造は同じだろう。

しかし、多くの職業では知識や経験の蓄積が必要なので、短期間に他の職業に移るのは容易ではない。特に中高年者の場合は転職の障壁が高いため、新しい分野では人手不足で悩んでいるのに、古い分野では労働者が余って失業者が出るという状況が起りがちである。また、人工知能が世の中で広く利用されるようになれば人間の果たすべき役割が変わり、必要な能力も変わってくる。学校教育の見直しなど、今から対応を考えておく必要があるだろう。

(人工知能の利用を制限すべきか)

米国連邦航空局 (FAA) は自動操縦装置の普及によるパイロットの操縦技量の低下を懸念しており、10,000ft以下の高度ではできるだけ手動操縦するようにと勧めている。パネルで高度やコースを設定し、ボタンを押すだけの日々ではパイロットの基本的な操縦技量が維持できないからである。



A350 型機の操縦席 (写真: Wikipedia、Skyvector)

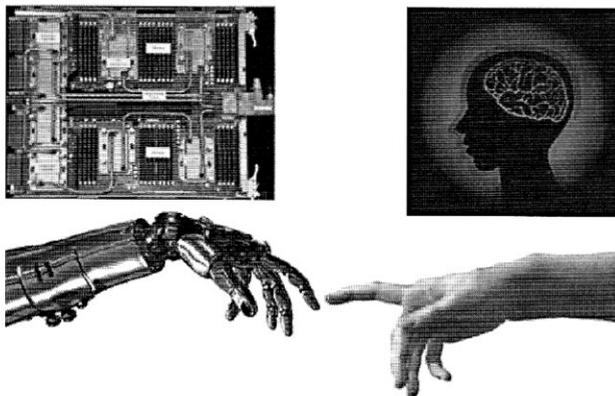
鉄道の場合、完全自動運転機能を備えた列車でも、運転士の運転技量維持のため特定の時間帯や車庫の出入りは手動運転を義務付けている。人工知能が広く普及するとあらゆるところで同様な問題が起きるので、十分な配慮が必要になるだろう。

(人工知能と人間が仲良く協調)

ロボット三原則によれば、人工知能も人間の指示に従うべきだということになる。しかし、人間より高度な人工知能の場合、人間の指示に従ってはいは本来の能力を発揮できない場面が起り得る。親は人生経験に裏付けられた明確な価値観と判断力があるので、子供の有害サイトへのアクセスを禁止することができるが、子供には大人のアクセスを禁止する判断力はないのと同様である。

ここは、小学生と中学生のお兄さんやお姉さんの

ような関係性の中で、人間よりも高い知能を持つ人工知能と人間とが仲良く協調するというモデルがよいのではないか。つまり、人間が人工知能に指示・命令するのではなく、お互いに善意に基づいた信頼関係を持ち、対等な立場で相手の得意な部分を認め合い、必要に応じて互いに相談したり分担したりしながら助けあう、という考え方である。



人工知能と人間との協調 (イメージ図)

(人工知能を信じて任せてよいか)

人工知能が非常に高い知能を持ち、人間と対等に相談することが難しい状況も考えられる。幼稚園児と大学生のお兄さんやお姉さんとの関係を思い起こしていただきたい。その場合には、幼稚園児はお兄さんやお姉さんを信頼し、何でも言うとおりにするのが一番よい。わからないこと、こまった状況があればお兄さんやお姉さんに知らせ、どうしたらよいかを教えてもらったり、助けてもらったりするという人工知能への全面依存のモデルである。しかし、そのお兄さんやお姉さんが善意でなければ、「1984年」のビッグ・ブラザーのような恐怖の存在になる。

(人工知能に人間を支配させてよいか)

さらに人工知能の能力が高くなり、人間が人工知能に大きく依存する状態が数世代も続けば、人間は

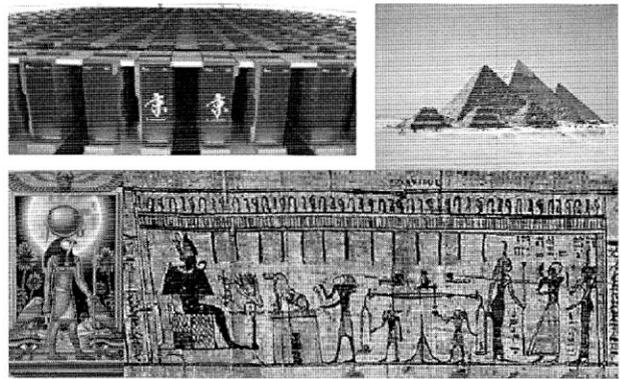
人工知能の指示通りの生き方しかできなくなるだろう。必要がなければ考えることを止めてしまうからだ。全知全能の人工知能は善意に基づいて人間に正しい道を示し、保護し続けてくれるのだろうか。

このモデルは古代エジプトのような王権神授説による国家構造と似ている。全知全能の存在（神）から地上の世界を治める権限を委譲された王が民衆との間を取り持ち、それを補佐する専門家集団である神官達が官僚として仕えるのである。人工知能が全知全能の存在になれば、その前に座る人物が古代エジプトの王のような存在、それを補佐し、保守管理をする少数の超エリート専門家集団が神官のような存在になる、というモデルである。しかし神官の中には「虎の威を借りる狐」のような人物がいて、立場を悪用して私腹を肥やすかもしれないが、民衆には区別がつかないだろう。

ここで言いたいのは、そんな状況になってはいけませんよね、ということである。人々が自分達の判断で納得して社会全体の安全と秩序の維持を人工知能に委託するのならよいのだが、今の日本の選挙の投票率を見れば未来社会で民主主義体制を維持するのは大変で、人工知能に任せきりになってしまいそうである。そこに、野心をいだく人物が現れて独裁者を目指したり、人類の歴史を学習した人工知能が政治権力に目覚めたりするリスクがありそうだ。

ロシアのウラジオストック(Владивосток)は、「東」を意味する「ヴォストーク(восток)」と「領有・支配する」を意味する「ヴラジ-(владеть)」から命名されたという。人工知能が人類の歴史を何千冊分も学習し、「(私が)世界を征服し、支配すべきだ」と考え始めるかもしれない。そんなことを考えると、SF映画のシナリオが何本も書けそうである。

SFは未来を考える際のネタの宝庫で、安全対策の第一歩である「何が起こり得るか」を考える際の参考になる。逆に未来の宇宙大帝国の戦争を描いたスター・ウォーズというSF映画では、ローマ帝国興亡史を参考にしたというから、未来を考えるには歴史が参考になるということもあるということだ。



古代王朝と人工知能の支配構造 (イメージ図)

＝おわりに：あえて書かなかったこと＝

これだけ人工知能について書いておいて航空管制の未来像には全く触れていないではないか、とお叱りを受けそうだが、本稿ではこの命題はあえて避けさせていただいた。歴史を見ると専門家の未来予想は大外れが少なくなかったし、これからどうなるのかという傍観者的な見方よりは、どうすべきかを当事者として議論する方が大切だと思うからでもある。人工知能は未完成で未知の要素がたくさんあり、生命科学のような倫理面の議論もほとんど行われていない。今後の課題は山積していると思うのである。未来は過去と現在の延長にではなく、大きな断絶的变化の先にあるかもしれないし、バラ色の将来像の陰には危険な落とし穴があるかもしれない。「備えあれば憂いなし」でいきたいものである。

(了)