機械学習のビジネス活用にかかわるリスク

昨今、AI（Artificial Intelligence：人工知能）という言葉が新聞等の紙面に現れない日はない、とまでいえるかもしれない。AI による技術革新は毎日のように語られ、その成果や好影響の傍ら、失敗例や危険性について議論されることもたびたびある。こうした現在の AI 流行の大きな理由は、機械学習手法の進展にあるといえるであろう。

本稿では、機械学習を活用したシステムの業務導入に関し、どのような点を考慮すべきかについて検討する。総務省「AI ネットワーク社会推進会議報告書 2017 [1]」をベースに、企業の業務活用として導入するケースを仮定し、特に従来の IT システムと比較して重要となる課題・検討事項について整理を行う。

1．AIは機械学習によって何が変わったのか

「AI を導入することで画像認識や翻訳、文意の抽出や要約、異常予測などが可能になった」というニュースは最近頻繁にみられる。だが、そもそも AI とは何であろうか。ごく単純に考えれば、AI とは“人工的に何らかの知性を表現したもの”といえる。だが“人工的な知性”という意味だけであれば、AI は古くから存在したことになる。例えば、コンピュータゲームに登場する敵キャラクターはプレイヤーが何らかの行動をとるのに合わせて様々な対処を行うことができる。これも一種の AI といえる。AI の大きな発展を広く印象付けた囲碁や将棋についても、家庭用ゲームの黎明期からソフトは存在した。

近年話題となっている AI と、古くからのごく単純な AI が決定的に違うのは、「機械学習」の広まりにある。古くからの AI は、“もし～ならば”で記述される条件節を積み重ねて表現されるアルゴリズムで動く。この条件節は開発者が設計し、コーディングするものであった。

この仕組みの場合、設計者は設計すべきものについて、十分な知識と結果の予測能力を持たねばならないことになる。将棋の AI は将棋の強い人がつくる方が望ましい。しかし、機械学習では異なるアプローチで正解を探し出す。

機械学習の場合、“設計者が作成するのは、どのように学習をするか”であり、与えられた学習ルールに従い、最も正解に近づく方法を何度も試行を重ねて探し出す。囲碁で有名になった AlphaGo は 2 台の AlphaGo がお互いに試合をし続け、最も強い打ち方を探し出していった。機械学習では、解きたい課題に対する知識もさることながら、知識以上に効果的な学習方法の設計を行うのが設計者の仕事となる。

機械学習を有効に行う方法が発見されていったことと、学習を行うための膨大な計算量に耐えられるコンピュータシステム、学習すべきデータが揃ったことがブレークスルーとなり、機械学習を利用した様々なサービスが生まれるきっかけとなった。機械学習は人間の“学習する”という能力の一面を再現することで、人間並みかそれ以上の知覚・判断能力を得ようとするものといえるが、“これを AI と呼んでいいのか”という点は議論の余地がある。実際には単純に高度な自動処理システムを指して AI と表現する例もあり、これが AI という言葉が多用される理由となっている。

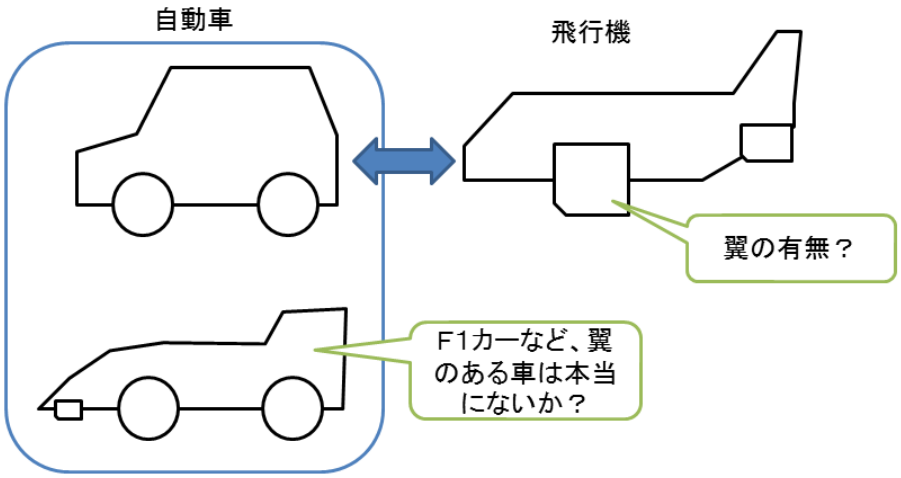
1 http://www.soumu.go.jp/menu_news/s-news/01iicp01_02000067.html

2. 機械学習の特徴

機械学習が期待されている理由は、高い判断精度と、その精度を得るための構築が従来に比して容易であることによる。ここでは機械学習の大きな成果としてしばしば取り上げられる画像認識課題で考えてみる。

例えば、ある画像が車なのか、飛行機なのかを判別するとする。従来の方法では、車と飛行機の間の差異を可能な限り考え出すことになる。翼があるかどうかは判断しやすい基準になるが、では、そもそも“翼”とは何か、という点も開発者は定義しなければならない。その要素は多岐にわたるため、この作業はすぐに大きく膨らんでしまう。また、本当に翼が基準でよいのか、といった点も議論はあるだろう。

■図１　車と飛行機の差異について



出典：弊社作成

機械学習においては、この判断基準を人が定めることはない。人が行うのは、飛行機と車の画像をとにかく大量に用意し、“どれが飛行機でどれが車である”という正解を学習するシステムに示すことのみである。画像を集める作業は時間がかかるものの、インターネットや IOT の発展で、作業のハードルは近年著しく低下した。画像を示されたシステムは、自動的に差異を探し、それを基準として学習する。機械学習では、判断基準となるアルゴリズムはデータに基づいてつくられるため、さらに画像を増やせば、より判断の精度を高めることが期待される。これも機械学習の利点である。

機械学習は対象のデータの特徴を調べることで自動的に最適解を見つけ出す手法群といえる。実際には機械学習に属する手法は無数に存在する（表１）。深層学習はその中でも比較的有名であるが、それも機械学習手法の一種に過ぎない。

■表１　代表的な機械学習手法の例

機械学習手法	例
深層学習系	深層ニューラルネットワーク 畳み込みニューラルネットワーク 再帰型ニューラルネットワーク
サポートベクターマシン系	サポートベクターマシン サポートベクター回帰 ガウス過程回帰
木系	回帰木 ランダムフォレスト XG ブースト
クラスタリング系	k 近傍法 k 平均法 適応共鳴理論（逐次学習）
強化学習系	Q 学習 Q 深層学習

※上記は一例。複数の手法が組み合わされ、単独の手法として分類できないケースも多い。

出典：弊社作成

分類や予測など目的に合わせ、開発者は最適な方法を選んで組み合わせる必要がある。手法によっては正解データが必要なかったり、データによって得意不得意があったりするケースもあり、どの手法を用いるかは開発者の判断が必要になる。ただ、大量のデータがある限りにおいては従来の手法に比べ高い成績を出すケースが多いため、その利用場面は広がっている。

ここまで示した用語を簡単にまとめると表２のようになる。本稿では、アルゴリズム生成において機器学習を用いたシステムを「機械学習システム」と呼ぶことにする。

■表２　AI、機械学習の用語解説一例

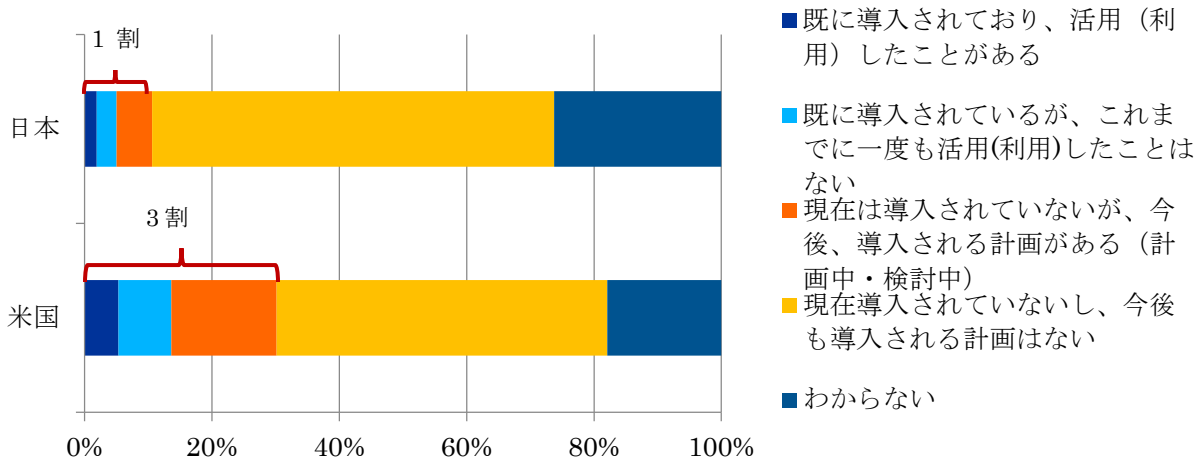
用語	解説
AI	定まった定義はなし。人間が行うことを代替できる何らかの IT システムを指すことが多い。
アルゴリズム	IT システムの中に組み込まれる、ある課題を解くための手順を定式的に表現したもの。
機械学習	分類・予測などを行うアルゴリズムを与えられたデータをもとに自動的構築する手法群。
深層学習	機械学習手法の一種。もともと人間の脳のネットワークを数学的に再現する手法から生まれた。

出典：弊社作成

3. AIの業務導入とその現状

機械学習の広まりをきっかけに、何らかの AI を用いたシステムは現在、急速に普及を始めている。ここまで示したように、AI は必ずしも機械学習を取り入れたものであるとは限らない。総務省の調査によれば、日本では 2016 年の調査の時点で、AI の職場での導入計画がある職場は 1 割ほどであるが、米国ではすでに 3 割ほどに上っている[2]。

■図２　職場への人工知能（AI）導入の有無および計画状況

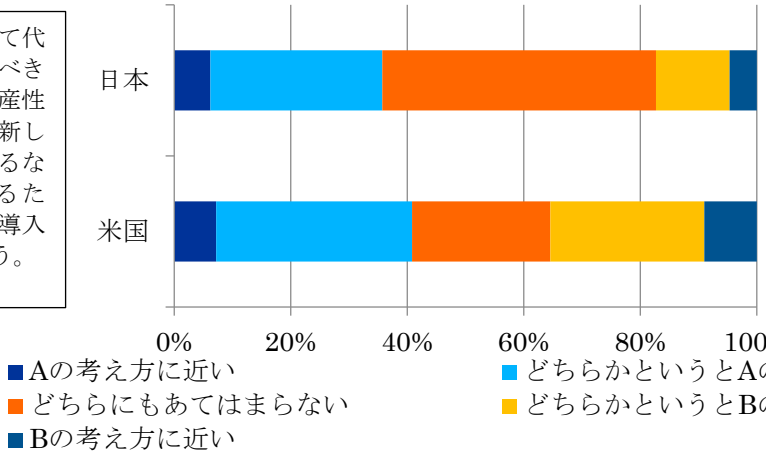


出典：総務省「ICT の進化が雇用と働き方に及ぼす影響に関する調査研究 報告書」をもとに弊社作成

ただし、AI の導入は常にうまくいくとは限らない。例えば、AI の導入にあたっては雇用への影響がたびたび語られている。調査によれば、職場への AI の導入に対する従業員の反応は二分されている。このように導入自体に反対する意見も根強く、導入のハードル、あるいは導入以後も活用のハードルとなると考えられる。

■図３　自分の職場への人工知能（AI）導入についての賛否

A)　業務や雇用が取って代わられるなどの憂慮すべき点はあるが、業務の生産性向上や高付加価値化、新しい業務の創造につながるなどの効果が期待できるため、人工知能（AI）の導入の流れは好ましいと思う。



B)　業務の生産性向上や高付加価値化、新しい業務の創出につながる効果については喜ばしいが、業務や雇用が取って代わられるなどの効果を考えると、人工知能（AI）の導入の流れは好ましいことではないと思う。

出典：総務省「ICT の進化が雇用と働き方に及ぼす影響に関する調査研究 報告書」をもとに弊社作成

[2] 総務省「ICT の進化が雇用と働き方に及ぼす影響に関する調査研究 報告書」（平成 28 年）
http://www.soumu.go.jp/johotsusintokei/linkdata/h28_03_houkoku.pdf

4. AI開発ガイドラインについて

このようにAIの導入が進む中で、機械学習を深く取り入れたシステムは、今後より増加するだろう。AIといっても本質的にはITシステムであり、開発の場で起きる課題については他のITシステムでも同様なものが多い。ただ、機械学習システムは他のシステムとは異なる特殊な課題を抱えることもある。そこで、特に機械学習システムを導入する場合に問題になりやすい点についてみていきたい。

本検討については、総務省「国際的な議論のためのAI開発ガイドライン案[3]」に示されている9つの基準をもとに示す（表３）。当該ガイドラインは研究開発者用のガイドラインという観点が強いが、導入する企業の立場からでも参考になる面が多い。

■表３　機械学習システム導入時に考慮すべき原則

①連携の原則	活用性を高めるため、システムの相互運用性を鑑み、通信方式や処理方式に配慮すること。
②透明性の原則	ログ取得や設計方式を通じて、入出力の検証可能性及び説明可能性を確保すること。
③制御可能性の原則	制御可能性を確保するため、人間や他のシステムによって監督・定期的な確認ができるよう、体制を整えること。
④安全保護の原則	利用者及び第三者の生命・身体の安全に危害を及ぼさないよう設計段階から配慮し、防護を備えること。
⑤セキュリティ保護の原則	使用するデータの保存について、あらかじめ検証確認するとともに、セキュリティ対策を確保すること。
⑥プライバシー保護の原則	プライバシーを侵害しないようデータ保護に配慮し、説明責任を果たすこと。
⑦倫理の原則	設計開発において、人間の尊厳と個人の自立を尊重すること。
⑧利用者支援の原則	利用者を支援し、利用者に選択の機会を適切に提供するよう配慮すること。
⑨アカウンタビリティの原則	設計開発において、ステークホルダに対するアカウンタビリティを果たすように留意すること。

出典：総務省「国際的な議論のためのAI開発ガイドライン案」をもとに弊社作成

なお、ここでは理解を進めるため、一例として工場等で活用されているセンサ値や稼働状況を監視し、異常検出を行う「センサ監視システム」を題材にする。以降、各原則について詳細に説明する。

① 連携の原則

機械学習システムが単体で動く、というケースはまれである。既存の社内システムと連携がしやすいよう、データの形式や転送の仕方に留意する必要がある。また、連携先のシステムに変化があったときに停止してしまったり、あるいは外部の連携先のシステムに対して悪影響を与えたりすることがないよう、どのように連携しているのかという点は、関係者全体で共有す

[3] http://www.soumu.go.jp/main_content/000490299.pdf

る必要がある。

センサ監視システムであれば、連携するシステムが使用しているデータの形式をあらかじめ調べ、スムーズな接続ができるようにするなどの検討が必要である。

② 透明性の原則

機械学習では大量のデータから自動的に判断基準を作成するが、機械学習の多くの手法では、その基準は何なのかということが複雑すぎて人間には理解できなくなってしまう。

この透明性の課題を回避するため、機械学習の手法については、その判断基準が理解できる手法を採用する方が良いケースがある。ただし、人が理解しやすい手法は精度があまり良くないことも多いため、理解しやすさを取るのか、予測の精度を取るのか、といった点が課題になる。

ただ、いずれにせよ、どのようなデータを与え、どのような手法で学習させているのかといった要素を利用者は理解しているべきである。これは何かトラブルが起きたときの解決や、またアルゴリズムの改良を検討する際にどんなデータが必要か、どのような手法が取り得るかといった検討を行うために必要になる。

センサ監視システムの例でみれば、必要な対応を判断するため、判断基準は重要である。ただ、危険物を扱う場合など、高い精度が優先される場面もあるだろう。“誤検知が許されるのか”“見落としが許されるのか”といった条件に応じ、適切な手法の選定を行う必要がある。また、そもそも機器異常の予兆検知を行う場合、“いかに早い段階で異常状態を正常状態と切り分けるか”という教師データ作成以前の課題もあり、事前の検討や定義が必要である。

③ 制御可能性の原則

そもそも機械学習では、どのような基準で判断をしているのかはほぼ定義できないため、複雑な学習手法になるほど、まれな場面ではまったく予想できない結果を出す可能性がある。このような事態にならないよう、導入前の試験にあたっては開発中に使ってきたデータとは異なるデータを使ったり、意図的にノイズを含むようなデータを使い、安全な環境での試験を行ってから本番導入に移るようにするべきである。ただ、従来のシステムのように全条件を網羅してチェックするということは不可能なことが多く、高い安全性が要求される場合は、従来以上に慎重なテストと、異常時の対策に関する事前検討が重要である。

また、機械学習を用いる場合、運用中にデータを追加しても、期待通りに精度が向上しないことも多い。例えばデータにノイズが入るのはよくあることであるが、そのデータをもとに学習を続けることで性能がむしろ低下してしまう可能性がある。

例えば Microsoft が作成した AI「Tay」は、利用者との会話を通じて学習して会話の幅が広がる、という触れ込みで登場したが、不適切な学習が行われたことにより問題発言をするようになり、停止させられた。これは利用者から問題のある言葉を教え込まれたためであり、ある意味、想定していなかったデータをもとに学習してしまった結果、望ましくない学習を深めてしまった、というケースである。機械学習システムは、必ずしも常時学習し続けるシステムである必要はない。むしろ、常時学習する機械学習システムを導入するのはトラブルが発生してもすぐに修正できる体制が必要であり、それが難しい場合は最低限の性能監視だけは行いつつ、学習の反映は定期的に行うようにして、問題があった場合は元に戻せるようにすべきである。

センサ監視システムの例では、センサの劣化などによって長期的なデータの特徴が変化して

しまうことがあり得る。これを学習してしまうことで異常検出の精度が下がるということもあり得るし、また学習をさせないと劣化したセンサからノイズがたびたび入ることで、誤検知が増えてしまうといった影響も考えられる。どちらにせよ問題があるため、定期的な監視と導入前の試験が必要である。

④　安全保護の原則

実社会においては、常に機械学習システムが未経験のデータに触れる可能性がある。この際、機械学習システムが機能しなかったり、想定外の動作を起こしたりする可能性は避けられない。このような誤判断はあり得るものとして、誤判断があった場合にどのような被害が起き、どのように損害の極小化やリカバリーが行えるかという点は、あらかじめ予測・計画すべきである。

センサ監視システムとしては、まずたとえ異常動作があっても危険に至らないように、危険なエリアに対する進入禁止措置を取ることや、異常動作を監視、警告し、停止させる別のシステムでカバーを図るなど、機械学習システムが動作させることのできる機器類の安全確保に留意する必要がある。また、定期的な安全環境のチェックができるよう、マネジメント体制を組み、もし何かあった際の業務代替が可能なのか、最大何日くらいなら停止を許容できるのか、といった BCP 面も検討しておくべきである。

⑤　セキュリティ保護の原則

機械学習システムは、実態として IT システムである以上、内部・外部からのハッキング等の攻撃の対象となり得る。攻撃によって攻撃者の意思通りに操作されるといった最悪のケースはもちろん、情報漏えいや機能停止など、多様な危険があり得るため、これらの場合への対処が必要となる。

センサ監視システムの例であれば、一般的に工場のネットワークは隔離されていることが多く、侵入が難しいものの、特に機械学習ではクラウドからデータを取得したり、工場内のセンサと連結しているなど、データの入出力経路が増えているため脆弱になりやすいという面もある。機械学習システムがどのようなネットワーク上にあるのか、計画時点で整理が必要である。

⑥　プライバシー保護の原則

機械学習を実施するために、非常にプライバシー性の高いデータを集めることがある。例えば、工場内従業員の行動や心拍等、生体データを取得する場合などがある。もちろん、従来から行われている個人情報保護のための対策の延長線上の話ではあるが、データ内容の示すものがよりプライバシーに踏み込むため、漏えいした際は問題になりやすいうえ、データの多量化などに伴い、管理が難しくなるという課題もある。

センサ監視システムの例であれば、従業員の安全行動を管理・監督するため、従業員にセンサ等を今後取り付けることがあるかもしれない。このデータを十分に保護することはもちろんであるが、保存する際に個人を特定できない形での保存方法にするなど、扱う情報を最低限に抑える対応の検討が必要である。

⑦　倫理の原則

システムは実装・学習されたアルゴリズムに従って動作するが、その結果が人間にとって倫理的に許容できなくなる場合がある。残念ながら実際環境のデータを用いた場合、これらの差別的なデータが取得されてしまう可能性はある。例えば、現時点で女性の管理者が少ないからといって、女性という理由で管理者に向いていないという判断を行ってはならないのは自明で

あるが、実際のデータをみれば、女性の方が管理者が少ない会社はいくらでもある。このようなデータを機械学習システムが学習した場合、「女性が管理者に向いていない」とする結論を導く可能性がある。それは学習してもよいデータなのか、望ましくない事実を反映した倫理的問題を含むデータなのではないか、といった点は留意すべきである。

センサ監視システムで倫理的要素が問題になることは多くはないと思われるが、例えば従業員の性別・信条などが機械学習システムによって学習され、その結果、差別的な待遇が生み出されることがないように留意しなければならない。

⑧　利用者支援の原則

機械学習システムの機能に限界がある以上、それだけで業務が完結するということはまれであり、周辺の人との連携が不可欠である。機械学習システムは、あくまで利用者の支援を行うことに留意し、利用者に適切な情報を与えつつ、判断を行うことができるようインタフェースをつくらねばならない。

機械学習システムは、関連する業務を行っている人に対して、今何が処理されているのかを伝えるようにし、人とのコミュニケーションに留意するべきである。これは何らかのトラブルが起きたりした際に気付くための要件となる。システム自体がミスをしなかったとしても、システムが間に入ることで、システムと人とのやりとりが新たなミスの原因ともなり得る。

センサ監視システムの例であれば、センサの値をすべて同じように示すのではなく、重要な要素だけ大きく示すことで問題の大小を人にわかりやすく伝えたり、単にアラームを鳴らすのではなく、何が普段と違うのかを示すことで異常の度合いを伝えるなど、その後の対応を迅速化できるよう、あらかじめ設計が必要である。

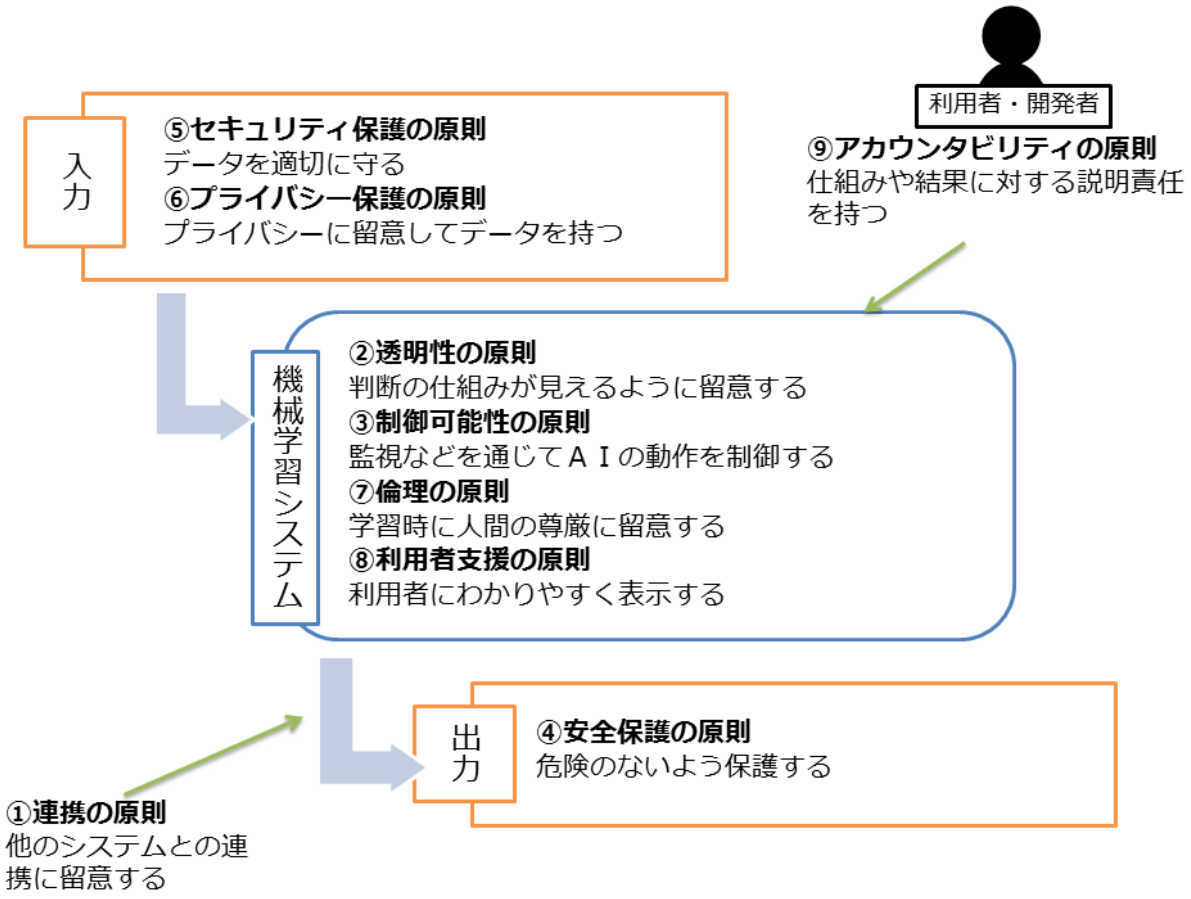
⑨　アカウンタビリティの原則

機械学習システムを開発し、利用する際にはその限界、課題を利用者自身が理解するとともに、その影響を受ける関係者・顧客などにもその限界、課題を説明する責任がある。

センサ監視システムであれば、使用する従業員がその仕組みを理解することは実務上必要であり、万が一トラブルがあった際にはその原因・対策などを関係者に説明する責任がある。これは従来のシステムでも同様ではあるものの、不透明なシステムになりやすいため、より意識しておく必要がある。

ここまでの概要を図４にまとめた。

■図４　AI 開発時に留意すべき原則の概観



出典：総務省「国際的な議論のための AI 開発ガイドライン案」をもとに弊社作成

5. 機械学習システムの導入に向けて

ここまで紹介したように、機械学習システムは従来のシステムとはやや異なる導入上の課題がある。そのため導入にあたって、“機械学習でなければならないのか”という点は改めて考える必要がある。機械学習の利点は、大量のデータさえあれば精度の高い結果を比較的容易につくり出せることである。一方で欠点はまさしく大量のデータが必要なことであり、あくまでデータに依存したシステムになってしまうことで中の仕組みがわかりづらくなってしまうことである。

従来の IT システムでは“予想のつかない動作”は、プログラムバグがなければそうそう起きるものではなかった。一方で機械学習はもともと動作の予測はつきづらく、常に予想外の動きを起こす可能性を内包しているので、動作の確実性を求めるようなケースで機械学習を用いるのはトラブルの原因になり得る。表４に従来型のシステムと機械学習システムの構造的な差異を示した。

■表４　アルゴリズムの設計方式による利点と欠点

	利点	欠点
従来のシステム	● 動作条件通りに動くため、トラブル原因の理解が比較的容易 ● 判定を行う場合、特定の条件を付加することが容易	● 複雑な予測判定を行おうとすると、相応に開発難易度が上昇する
機械学習システム	● データさえあれば、比較的容易に予測精度の高いシステムが作成可能 ● データを追加整備することで、精度向上を図ることができることが多い	● 制作に大量のデータが必要であり、データを集め、整備することに費用や時間がかかる ● 判定の基準が理解しづらく、トラブル対応などに差し支えることがある ● 精度が予想に反して出ない場合、精度を向上させる方法を特定しづらい

出典：弊社作成

機械学習システムを導入するうえでは、上記のような差異についても意識してプロジェクトを進めなければならない。

従来のシステムと機械学習は容易に切り替えられるものではなく、場面に応じて使い分ける、あるいは双方が補完し合うものである。設計の段階では機械学習の手法選定と同様に、そもそも機械学習が本当に適しているのかも検討が必要である。人に得手不得手があるように、システムにも得手不得手がある。“この課題にはこの種類の機械学習。逆にこの課題は従来型で、この部分は人手で”といった選別を行う必要がある。

「AI」という言葉の流行により、機会学習に興味を持つ方は大変増加している。ただ、漠然と機械学習に万能感を抱いたり、圧倒的な精度を期待すると、多くの場合、期待外れになる。機械学習はあくまで IT システムのアルゴリズムの構築法の一つに過ぎない。自社での導入を考えている場合、あらかじめ検討すべき要素として以下のような点がある。

■表５　機械学習システムの導入前に検討するべき問いの例

整理すべき点	主な関連する原則
どのような仕事を省力化したいのか、自動化したいのか？	①　連携の原則 ⑧　利用者支援の原則
機械学習を行うことでより優れた性能を期待できるのか？　どんな手法を用いるべきか？	②　透明性の原則 ⑨　アカウンタビリティの原則
教師データは用意可能なのか、外部から入手できるのか？　データはどのように保持するか？	⑤　セキュリティ保護の原則 ⑥　プライバシー保護の原則 ⑦　倫理の原則
判断基準が不透明になることは許容可能なのか？	②　透明性の原則 ⑨　アカウンタビリティの原則
継続的なメンテナンスなどを予定できるのか？	③　制御可能性の原則
精度はどのくらい必要か、そもそも精度は何で測るのか？	②　透明性の原則 ③　制御可能性の原則
見落とし、過剰検知はどの程度許されるのか？ 見落とし、過剰検知にはどのように対応するか？	③　制御可能性の原則 ④　安全保護の原則
どこまでを機械が、どこまでを人が行うのか？	①　連携の原則 ⑧　利用者支援の原則

※「主な関連する原則」の項目に関する内容の詳細は表３参照。

出典：弊社作成

例えば「教師データがない状態で機械学習に取り組もう」というのはかなりハードルが高く、そもそも自社のデータ整理から進めた方が良いといったケースはしばしばみられる。先の原則等も踏まえ、貴社の現状が機械学習の導入に適しているのかを検討していただきたい。

6. 機械学習システムの効果的な活用を狙って

機械学習は有用な技術であり、様々な形で労力を削減し、新たな価値を生み出す可能性がある。しかし、特定の技術がすべての問題を解決することなどあり得ないように、機械学習も当然万能ではない。課題に応じて適切な機械学習手法を選定するとともに、人・従来のシステムと連携して全体としての効率性・有用性をつくり出すことが必要である。

いずれにせよ、機械学習システムはまだ草創期である。過剰に恐れることなく、まずは小さく始め、経験を積み、本格導入を目指してみてほしい。

[2018 年 3 月 19 日発行]

東京海上日動リスクコンサルティング株式会社

To Be a Good Company

自動車リスク本部　主任研究員　駒田　悠一（専門分野：人間科学・安全工学）
〒100-0004 東京都千代田区大手町 1-5-1 大手町ファーストスクエア ウエストタワー23 階
Tel. 03-5288-6586　Fax. 03-5288-6628
http://www.tokiorisk.co.jp/

