

医療機関における意思決定者に対する人工知能教育

研究分担者 亀田 義人

（千葉大学医学部附属病院 病院長企画室 特任助教）

研究要旨

本分担研究では医療機関における病院長、その候補や支援人材等、人工知能導入に際する意思決定者に向けた教育プログラムを開発する。今年度においては、種々の関係者からのヒアリングや教育講座の情報収集と再編成を行い、病院執行部向けの教育プログラムの作成を試みた。

その結果、病院管理者や執行部としてAIを活用していく上で重要なポイントとして、以下の項目を抽出した。すなわち、(1) 人工知能とは何か、(2) 人工知能を活用する意義、(3) 人工知能の歴史とトレンド、(4) ディープラーニングの理論的な背景、(5) 医療における人工知能の活用領域、(6) 人工知能の限界と適切な活用、(7) 人工知能を活用した製品の調達について、の7項目となる。

来年度においては、必要な知識や能力を整理することでコンピテンシーの設定を図る。また、病院執行部が広く受講できる様な環境整備及び、試験的講座の実施と受講者からのフィードバックによる教材の質の向上に取り組むたい。

A. 研究目的

本研究班では、医療用人工知能(AI)分野において、技術革新と国際競争力向上に資する人材育成プログラムを開発することを目的としているが、本分担研究では医療機関における病院長、その候補や支援人材等、人工知能導入に際する意思決定者に向けた教育プログラムを開発することを目的とし、もって、AIを搭載した技術への理解と医療機関への研究開発協力への理解と社会実装を推進することを目的とする。

B. 研究方法

今年度、プログラムの開発に際して、以下の3点に取り組んだ。

① AI教育プログラムの現状・情報把握

まず、AI教育の現状やAIに関する情報の把握を目指した。そのために、各民間および国公立やそれに準ずる機関が提供から提供されているオンサイト/オンラインAI教育講座について実際に参加し、必要なリテラシーの取得と抽象化・概念化を行った。また、参加を通じて、AI系情報技術者との関係構築を図った。

② AI研究者・企業のAIシステム開発者へのヒアリング

AIのプログラミング教育提供者、AI研究者や企業のAIシステム開発者へヒアリングを行い、必要なリテラシーやAIを活用するためのコンピテンシー、および、それらの評価方法の開発を試みた。

③ プログラム開発の管理体制の構築

プログラムの開発に際しては、千葉大学医学部附属病院が展開中の「病院経営スペシャリスト養成プログラム」の協力を得た。「病院経営スペシャリスト養成プログラム」は、持続的な病院経営システムを構築できる実践的な医療人材を養成することを目的とした病院経営に関する研修である。これは、平成19年の学校教育法の改正以降実現した大学としての正式な履修証明書を発行する研修となっている。そのために、プログラムの進捗管理及び報告・改善の場として、内外の委員からなるプログラム運営委員会を有している。本研究分担における教育プログラムの開発に際しては、収集した情報をもとに教育プログラムを作成し、その概要を公私立の病院長を外部委員として擁するプログラム運営委員会に諮り、意見を反映する形式にて進めた。

C. 研究結果

東京大学大学院情報工学系研究科が主催する「実データで学ぶ人工知能講座」、公益財団法人医療・病院管理研究協会が主催する「病院管理研修 医療における人工知能とビッグデータの活用と影響」、株式会社ARKが主催する「AIエバンジェリスト養成講座」、株式会社キカガクが主催する「人口知能脱ブラックボックスセミナー」、「ディープラーニングハンズオンセミナー」及び株式会社メドピア主催の「Health 2.0 Asia-Japan 2017」にて、人工知能周辺の知見を収集した。また、人工知能学会理事の千葉大学荒井幸代教授よりヒアリングを行い、病院管理者として人工知能とどのように向き合うべきか等についてまとめた。

病院管理者や執行部としてAIを活用していく上で重要なポイントとして、以下の7点を作業仮説として提示した。

- (1) 人工知能とは何か
- (2) 人工知能を活用する意義
- (3) 人工知能の歴史とトレンド
- (4) ディープラーニングの理論的背景
- (5) 医療における人工知能の活用領域
- (6) 人工知能の限界と適切な活用
- (7) 人工知能を活用した製品の調達

文末に、各トピックの要旨を整理する。

D. 考察

人工知能の社会実装に向けて、病院経営陣が備えるべき知識について整理したが、そもそも各個人の基礎知識にばらつきがある中で、どのような項目を学習するべきか、コンテンツベースの教材作成には限界があると考えられる。医学教育や他分野の教育と同様、コンピテンシーを設定して、その習得に向けて各個人が主体的な学び (Active learning) を行う Outcome based Education が必要であると考えられる。

E. 結論

今年度においては、種々の関係者からのヒアリング、教育講座の情報収集と再編成を行い、病院執行部向けの教育内容の概要案を作成した。来年度においては、これを更に成熟させ、必要な能力を抽象化しコンピテンシーの設定を図る。また、病院執行部が広く受講できる様な環境整備及び、試験的講座の実施と受講者からのフィードバックによる教材の質の向上に取り組みたい。

F. 研究発表

なし

補遺－医療機関トップを対象とした AI 教育教材要旨

(1) 人口知能とは何か

人工知能の定義の前に、知能に関する統一的な定義を考える必要がある。知能とは、一般的には知的な活動の能力であり、知的な活動には問題解決・推論・学習などの情報処理能力や抽象化、一般化などが一面として挙げられる。

一方、人工知能は、一般に「何らかの知的動作が可能な計算機システム」と言われるが、この「知的動作」にも定義が無い。また、人工知能では知能をシミュレートすることになるが、知能のシミュレーションが知能といえるかという疑問もある。表 1 に示されるとおり、本邦の主要な人工知能研究者においても、人工知能の定義は様々であり、統一的な定義はなされていない。

人工知能の分類の一つに、強い AI、弱い AI という分類があり、前者は汎用型人工知能と呼ばれる。いわゆる人間のような振る舞いをする知的コンピューターであり、設計した時の想定を超えた新たな問題にも対処する。後者は特化型人工知能と呼ばれ、人間が設定した特定の問題(画像の識別等)に対して、主に機械学習を用いて解決する。強い AI の実現は未だ課題が多く、通常活用されるのは後者である。

人工知能の持つ機能は人体の構造・機能になぞらえて、目、耳、言語に例えられる。すなわち、画像認識、音声認識、自然言語処理である。これらの情報を数値化し、機械学習を通じて規則性を見つけ出し、推計あるいは分類した結果として出力するというのが基本的な挙動となっている。

人工知能と機械学習及び、近年注目されるディープラーニングの関係性として、「人工知能⇒機械学習⇒ディープラーニン

中島秀之	公立ほこだて未来大学	人工的につくられた、知能を持つ実態。あるいはそれをつくろうとすることによって知能自体を研究する分野である
武田英明	国立情報学研究所	
西田豊明	京都大学	「知能を持つメカ」ないしは「心を持つメカ」である
溝口理一郎	北陸先端科学技術大学院	人工的につくった知的な振る舞いをするためのもの(システム)である
長尾真	京都大学	人間の頭脳活動を極限までシミュレートするシステムである
堀浩一	東京大学	人工的に作る新しい知能の世界である
浅田稔	大阪大学	知能の定義が明確でないので、人工知能を明確に定義できない
松原仁	公立ほこだて未来大学	究極には人間と区別が付かない人工的な知能のこと
池上高志	東京大学	自然にわれわれがペットや人に接触するような、情動と冗談に満ちた相互作用を、物理法則に関係なく、あるいは逆らって、人工的に作り出せるシステム
山口高平	慶應義塾大学	人の知的な振る舞いを模倣・支援・超越するための構成的システム
栗原聡	電気通信大学	人工的につくられる知能であるが、その知能のレベルは人を超えているものを想像している
山川宏	ドワンゴ人工知能研究所	計算機知能のうちで、人間が直接・間接に設計する場合を人工知能と呼んで良いのではないかと思う
松尾豊	東京大学	人工的につくられた人間のような知能、ないしはそれをつくる技術。人間のように知的であるとは、「気づくことのできる」コンピュータ、つまり、データの中から特徴量を生成し現象をモデル化することのできるコンピュータという意味である

表 1 国内の主な研究者による人工知能の定義
(総務省 平成 28 年度情報通信白書より)

グ」という関係がある。ディープラーニングは機械学習の一手法であり、機械学習は人工知能を構成する一要素であると表現できる。弱い人工知能による問題解決の手法は、現在、二つのアプローチが主流となっている。一つは、回帰問題の解決(結果の推計・推定)であり、もう一つは分類問題の解決(分類)である。

このようなアプローチで人工知能に問題解決をさせる上では、大量の利用可能なデータ、高い計算機能力、機械学習アルゴリズム、3つの要素が重要となる。近年の人工知能ブームには、これらの3要素がそろったことになり、研究がなされてきた技術が現実的な問題に適用できるようになった背景がある。

(2) 人工知能を活用する意義

人工知能が注目を集めるようになった社会的な背景には、技術的側面のみならず、人口構造の変化及び産業構造の変化がある。

日本においては、1960年頃から1990年代半ばまで人口の生産年齢割合が高く、人口構造が経済にプラスになる時期があった。これは人口ボーナスと呼ばれ、安い労働力を背景に世界中の仕事を受注する一方で、高齢者割合が低く社会保障にかかる費用が比較的少ない時代であることを活かし、インフラ投資を進めることができた。経済発展に関しては、高度成長期となる時代である。

高度成長期が訪れると富裕層が子供に教育投資し、高学歴化による人件費の上昇と晩婚化・晩産化が進み少子化の傾向が生じる。医療や年金制度が充実するため平均寿命も延び、やがて高齢化社会へと移行する。

人口ボーナス期が終えんを迎え、次に来るのが人口オーナス期である。オーナスとは「重荷・負担」という意味で、人口構造が経済の重荷になる時期である。ボーナス

期に比べ働く人よりも支えられる人が多くなる状況であり、ボーナス期のような人を大量に投入して行うような経済発展の仕方は通用しなくなる。人口オーナス期に生じる典型的な問題は、労働人口が減少すること及び、働く世代が引退世代を支える社会保障制度の維持が困難となることである。高齢化は世界的な課題であるが、日本は特に主要国で最も早く少子高齢化が進行している。ことさら医療に関しては、限られた財源の中著しく増加する高齢者に対してサービスを提供していくことが求められており、生産性の向上が不可欠である。人工知能による産業構造の変化は、ドイツが人工知能含めてIoT(Internet of Things)の社会実装をIndustry 4.0(第4次産業革命)と銘打っており、人工知能の活用による生産性の向上が期待されている。米国では「米国人工知能研究開発戦略」を策定して、各分野の人工知能の実装に関するロードマップを策定しており、中国では新一代人工知能発展計画を立案し、自動運転、都市計画、医療映像、音声認識の4分野に特に力を入れ、百度、アリババグループ、テンセント、科大訊飛(iFLYTEK)の4社を指定して2030年までに世界トップ水準へ向上させ中国を世界の主要な「AIイノベーションセンター」にする事を目指している。このように各国が人工知能の活用を国家戦略に位置づけており、日本でも日本世界最先端IT国家創造宣言・官民データ活用推進基本計画(平成29年5月30日閣議決定)が策定されるなど、国家として取り組んで行くべき事項である事がわかる。

(3) 人工知能の歴史とトレンド

人工知能研究は1950年代から続いているが、現在までに3回のブームが生じてきたとされる。

第1次ブームでは、コンピューターによる「推論」や「探索」が可能となり、特定の問題に対する解決ができるようになった。しかし、複雑な問題の解決が困難な事が明らかとなり、ブームの終焉を迎えた。この頃の技術ですでにパーセプトロンや人工対話システム **ELIZA** などが作られている。

第2次人工知能ブームは、1980年代に生じた。推論に必要な知識を人手を用いて表現し、コンピューターに与えることで、実用的な「エキスパートシステム」(専門分野の知識を取り込んだ上で推論することで、その分野の専門家のように振る舞うプログラム)が実現した。また、エキスパートシステムの他に、入力された知識から確定的な手順の記述なしに結果を推定する「ファジィ」や「ニューロ」コンピューティングが開発された。しかし、コンピューターが十分な情報処理能力を持たない時代であったために、「ファジィ搭載」を謳った商品が市販される程度に留まり、ブームはやがて終焉を迎えた。

第3次人工知能ブームは、2000年代から現在まで続いている。発達したコンピューターの計算能力を活用し、また、大量に集積するようになったビッグデータを用いて、「ディープラーニング」(2006年)などの機械学習アルゴリズムによって様々な実用的課題が解決できる様になったことが背景にある。

こうした技術の多くやプラットフォームは、無料で公開されることが多い一方で、対象となるデータの量と質が課題解決に重要視されている。また、実際の商品・サービスとして社会に浸透するためには、実用化のための開発や社会環境の整備、課題を持つ業界と人工知能技術者間でのすりあわせなどの取組を進めていく必要がある。

(4) ディープラーニングの理論的背景

人工知能を活用する上での得意不得意を知るに当たって、その前提となる理論的背景を知ることは重要である。特に実応用が進んでいるディープラーニングの理論的背景の概要を説明する。

機械学習を利活用する場合、大きく二つの手法が存在する。一つは線形代数を基礎としたもので、もう一つは確率統計学を基礎としたものである。前者には個々で扱うディープラーニングやサポートベクターマシーン(SVM)と呼ばれるパターン認識手法などが含まれ、後者にはマルコフ連鎖モンテカルロ法(MCMC)や隠れマルコフモデル(HMM)などといったものが含まれる。

機械学習を活用して問題解決を図る場合、大まかな流れは以下の通りとなる。

1) 解決したい問題に対して計算モデルを決める

ディープラーニングは、計算モデルとしてはディープニューラルネットワーク(DNN)と呼ばれ、入力層、中間層(隠れ層)、出力層からなる(図1)。中間層を何層置くか、隠れ層のノードをいくつ置くかについてもヒトが設定する必要がある。パラメーター W の値ははじめランダムに設定する。DNNでは、入力 x についてパラメーターが掛けられた結果の値 U に対して、そのまま結果を用いるのでは無く、非線形データにも対応できるように非線形変換が行われる。非線形変換に用いられる関数を活性化関数と呼び、従来はシグモイド関数というものが使われていたが、現在、ReLU 関数というのが一番精度高く結果を出せるといわれており、これを用いる場合が多い。

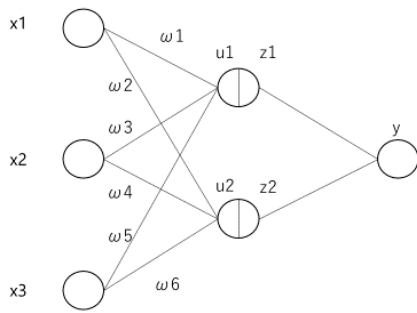


図 1 ニューラルネットワークの例

2) 評価関数(損失関数)を決める

DNN において、評価関数では、回帰問題では通常平均二乗誤差、分類問題ではソフトマックス関数という関数を用い、クロスエントロピー誤差という値を求める。設定したパラメーターを元に評価関数の値を求める事を順伝播という(図 2)。

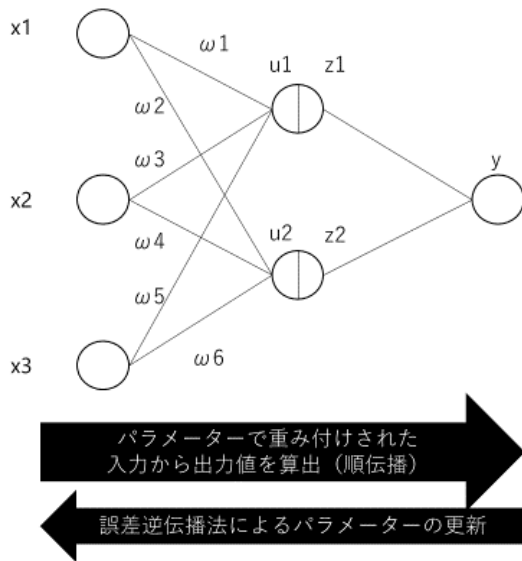


図 2 順伝播と誤差逆伝播

3) 評価関数を最大化もしくは最小化する

評価関数を最小化するために、評価関数をパラメーター W からなる式に整理して、その式を W で偏微分し、その値を元のパラメーター W から引く事により更新する計

算を繰り返す(最急降下法)これを誤差逆伝播法という。

(5) 医療における人工知能の活用領域

医療においても様々な領域で人工知能の活用が期待されている。一つの分類方法として、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」では、識別(音声認識、画像認識、動画認識、言語解析)、予測(数値予測、マッチング、意図予測、ニーズ予測)、実効(表現生成、デザイン、行動最適化、作業の自動化)に分類している。これらを医療に照らし合わせることで、AI を活用してどのような事が実現可能か考える手がかりになるかもしれない。

医療における人工知能の活用も多岐荷渡また現在進行形で増え続けているため、枚挙にいとまが無いが、個々では 3 つほど例を挙げる。

1) 医療画像処理

画像処理はディープラーニングが最も特異とする分野の一つである。現在の第 3 次 AI ブームは、Stanford 大学の Fei-Fei Li が教師付画像データ : ImageNet を創設した事から注目を浴び、画像認識で始まった経緯がある。人工知能を画像解析に活用する場合、畳み込みニューラルネットワーク(CNN: Convolutional Neural Network)というモデルを用いることが多い。

画像認識タスクでは、画像データとそのラベルを大量に学習させることによって、未知の画像に対してそれが何であるかのラベルを出すという枠組みで動作する。この技術を活用して、既に放射線領域では脳動脈瘤の高精度な検出に応用されており、そのほかの頭蓋内病変の診断への応用や、胸部 X 線写真から骨の特徴量を削除することにより鎖骨や肋骨が無い画像を提供し、肺

の病変を見やすくする、などの技術が開発されている。東京大学初ベンチャーの□ピクセルは富士フィルムと共同し、人工知能による診断支援を富士フィルムの医療用画像情報システムに搭載していくことを目指している。東芝デジタルソリューションズは千葉大学と共同し、胃がんのリンパ節生検検体に対し、人工知能を用いた術中迅速病理診断の研究を行っている。

十分な量の教師付データがあれば様々な分野に応用が可能であり、病理診断の自動化、内視鏡所見をリアルタイムに解析して、腫瘍らしき所見にアラートを発するシステムや、悪性黒色腫等の皮膚疾患の検出など様々な画像診断技術に人工知能の活用が取り組まれている。

2) がんの Precision Medicine

ゲノム医療において、患者個々人のゲノム情報から、より効果的で安全な治療法を選択する Precision Medicine が行われ始めている。既報の文献や既存の遺伝子・病態・治療に関するデータベースの情報をキュレーションする事により、個々人に最適化された診療について提案する。国立がんセンターは産総研や国内発人工知能ベンチャー企業である Preferred Networks と共同で Precision Medicine に取り組んでいる。

3) その他自然言語処理等の活用

患者の問診内容や身体所見をインプットとして、診断補助に活用する手法や、診療文書の作成補助などの活用方法が考えられる。文脈として理解する必要がある場合は、前後関係を踏まえた分析が可能なモデルが必要であり、主に Recurrent Neural Network が用いられる。

総務省次世代人工知能社会実装ワーキンググループの自然言語処理の医療応用の報告によれば、学会が行ってきた症例データ

ベースの蓄積を活用して、診断支援に取り組むことや、ツイッターのつぶやき内容を自然言語処理で解析し、感染症発生を推定するツールの開発、語彙から将来の認知症の発生を予測する技術など、様々な分野で実用化が図られている。

上記以外にも、人工知能の推計・分類機能が活用できる領域は様々にあると考えられる。人工知能の特徴を理解し、医療現場のどのような課題解決に、どのように人工知能を活用するか考えていく事が社会実装には重要である。

(6) 人工知能の限界と適切な活用

チューリングテスト

チューリングテストという、機械に応答させて人間が応答しているか機械が応答しているか、判断するテストがある。しかし未だ十分に人間らしい反応であるという成果を挙げてはいない。

不完全性定理

「自己言及のパラドックス」いわゆる「クレタ人は嘘つきだとクレタ人は言った」を、ゲーゲルが数学的に証明したもの。完全な人工知能はその完全性を証明できないという矛盾が生じる。外部から条件・定義を設定する必要が生じる。

中国語の部屋

中国語を知っているヒトがコンピューターに対して中国語を入力、別の部屋で中国語を知らないヒトが中国語を記号として認識して、ある記号に対してはこの記号、というように返答すると、見かけ上中国語を理解しているヒトが対応している様に見える。単に記号を操作するだけで知能をシミュレートしており、知能はヒトにも機械にもでっち上げられるというもの。

フレーム問題

現在、人工知能は思考すべき範囲や詳細度の決定を人工知能自身でできない。ある目的を達成使用とする場合、関係ある事項のほか、関係無い事項は無限に存在するため、それを洗い出そうとすると無限に時間を要し停止してしまう。

現在のところ、人工知能に対しヒトが「適切」な範囲を定める必要があり、「範囲を定める問題」の一般的な解決は困難である。

記号接地問題

記号を実世界の意味と結びつけられるかという問題。コンピューターは記号の意味を理解していないので、記号の操作だけで知能を実現できない。例えば、シマ+ウマ=シマウマという様な概念を一般化して獲得する事はできない。

トロツコ問題

トロツコが線路の上を走行しており、岐路が設けられている。線路の先には修理作業中の5名の作業員がおり、そのまま走ると5名が轢かれる。別の岐路の先には1人が作業している。A氏はたまたまトロツコの進行方向を変えるレバーの近くにいる。A氏がとるべき「正しい」行動とはどのようなものか。という問題。人工知能に「正しい」判断をさせる場合、何を持って「正しい」とするか。人工知能にどのような価値判断を与えるべきか、どのようにコンセンサスを得るのか、コンセンサスは得られるのか。

瑕疵があった場合の責任の所在

人工知能を搭載した製品を使用して起きた問題の中で、人工知能の判断が誤っていたために起きた問題の責任は、製品製造者

にあるか、製品使用者にあるか。人工知能が出した解答に従って起きた問題の所在は、どのような場合にどこに置くべきか。

Asilomar AI Principles (2017)

2017年1月、カリフォルニア州アシロマにて行われた、全世界のAIの研究者と経済学、法律、倫理、哲学の専門家が集まり議論された人類にとって有益なAIに関する原則。研究に関する5項目、倫理と価値観に関する13項目、長期的な問題に関する5項目計23項目が挙げられた。

AI ネットワーク社会推進会議報告書 2017

総務省情報通信政策研究所が、平成28年10月から「AI ネットワーク社会推進会議」を開催し、AI ネットワーク化をめぐる社会的・経済的・倫理的・法的な課題について検討を進めている。平成29年7月に報告書が発出され、その中で、国際的な議論のためのAI 開発ガイドライン案が提案された(図3)。

(7) 人工知能を活用した製品の調達

人工知能の用いた製品を開発・実装する場合、図4のような手順が必要になると考えられる。前述の通り、現状人工知能を活用する場合、問題の設定を解決モデルの設定はヒトが行わなければならない。適切なシステム開発のためには、現場の問題意識と、人工知能開発研究者と、ユーザーインターフェース開発ベンダーとの間で、密なコミュニケーションが必要となる。それを理解した上で、対応適任者を専任する必要が出てくる。また、精度向上のためには、利用環境で生まれるデータを学習させ続ける仕組み作りも重要である。

第2章 国際的な議論のためのAI開発ガイドライン案①

目的	基本理念	用語の定義及び対象範囲
<ul style="list-style-type: none"> AIネットワーク化（AIシステムがインターネット等と接続され、AIシステム相互間又はAIシステムと他の種類のシステムとの間のネットワークが形成されるようになること）の進展により、人間及びその社会や経済に多大な便益がもたらされる一方、不透明化等のリスクに対する懸念も存在 オープンな議論を通じ、国際的なコンセンサスを醸成し、非規制的で非拘束的なソフトローたるガイドラインやそのベストプラクティスをステークホルダー間で国際的に共有することが必要 AIネットワーク化の健全な進展を通じたAIシステムの便益の増進とリスクの抑制により、利用者の利益を保護するとともにリスクの波及を抑制し、人間中心の智連社会を実現 	<ul style="list-style-type: none"> 人間がAIネットワークと共生することにより、その恵沢がすべての人によってあまねく享受され、人間の尊厳と個人の自律が尊重される人間中心の社会を実現 非拘束的なソフトローたる指針やそのベストプラクティスをステークホルダー間で国際的に共有 イノベティブでオープンな研究開発と公正な競争、学問の自由や表現の自由といった民主主義社会の価値を尊重するとともに、便益とリスクの適正なバランスを確保 技術的中立性の確保、開発者に過度の負担を課さないよう配慮 不断の見直し、必要に応じた柔軟な改定、広範で柔軟な議論 	<ul style="list-style-type: none"> 「AI」: AIソフト及びAIシステムを総称する概念 ●「AIソフト」: データ・情報・知識の学習等により、利活用の過程を通じて自らの出力やプログラムを変化させる機能を有するソフトウェア（例: 機械学習ソフトウェア） ●「AIシステム」: AIソフトを構成要素として含むシステム（例: AIソフトを実装したロボットやクラウドシステム） ・AIシステムの「開発者」及び「利用者」は以下のとおり定義 ※ただし、「開発者」及び「利用者」は場面に依りて個別に決まる相対的な概念 ●「開発者」: AIシステムの研究開発（AIシステムを利用しながら行う研究開発を含む）を行う者（自らが開発したAIシステムを用いてAIネットワークサービスを他者に提供するプロバイダを含む） ●「利用者」: AIシステムを利用する者（最終利用者（エンドユーザ）のほか、他者が開発したAIネットワークサービスを第三者に提供するプロバイダを含む） ・対象とするAIシステムの範囲→ネットワーク化され得るAIシステム（ネットワークに接続可能なAIシステム） ・対象とする開発者の範囲→一定定義された開発者すべて ・対象とする開発の範囲→閉鎖された空間（実験室、セキュリティが十分に確保されたサンドボックス等）内での開発は対象とせず、ネットワークに接続して行う段階に限定
<p>関係者に期待される役割</p>	<ul style="list-style-type: none"> ・各国政府及び国際機関による多様なステークホルダー間の対話の促進に向けた環境整備 ・開発者、利用者等ステークホルダーによる対話やベストプラクティスの共有、AIの便益及びリスクに関する認識の共有 ・標準化団体等による推奨モデルの作成・公表 ・各国政府によるAIの開発者コミュニティの支援、AIに関する研究開発を支援する政策の積極的な推進 	

第2章 国際的な議論のためのAI開発ガイドライン案②（AI開発原則案の解説）

開発原則	開発原則の解説（留意することが期待される事項等）
<p>① 連携の原則 開発者は、AIシステムの相互接続性と相互運用性に留意する。</p>	<ul style="list-style-type: none"> ・相互接続性と相互運用性を確保するため、①有効な関連情報の共有に向けた協力、②国際的な標準や規格への準拠、③データ形式の標準化及びインターフェイスやプロトコルのオープン化への対応、④標準必須特許等のライセンス契約及びその条件についてのオープン・公平な取扱い、などに留意することが望ましい。
<p>② 透明性の原則 開発者は、AIシステムの入出力の検証可能性及び判断結果の説明可能性に留意する。</p>	<ul style="list-style-type: none"> ・採用する技術の特性や用途に照らし合理的な範囲で、AIシステムの入出力の検証可能性及び判断結果の説明可能性に留意することが望ましい。（※アルゴリズム、ソースコード、学習データの関係を想定するものではない。）
<p>③ 制御可能性の原則 開発者は、AIシステムの制御可能性に留意する。</p>	<ul style="list-style-type: none"> ・AIシステムの制御可能性について、あらかじめ検証及び妥当性の確認をするよう努めるとともに、採用する技術の特性に照らし合理的な可能な範囲において、人間や信頼できる他のAIによる監督・対処の実効性に留意することが望ましい。
<p>④ 安全の原則 開発者は、AIシステムがアクチュエータ等を通じて利用者及び第三者の生命・身体・財産に危害を及ぼすことがないよう配慮する。</p>	<ul style="list-style-type: none"> ・AIシステムの安全性について、あらかじめ検証及び妥当性の確認をするよう努めるとともに、AIシステムの本質安全や機能安全に資するよう、開発の過程を通じて、採用する技術との特性に照らし可能な範囲で措置を講ずるよう努めることが望ましい。 ・利用者及び第三者の生命・身体・財産の安全に関する判断を行うAIシステムについては、利用者等ステークホルダーに対して設計の趣旨などを説明するよう努めることが望ましい。
<p>⑤ セキュリティの原則 開発者は、AIシステムのセキュリティに留意する。</p>	<ul style="list-style-type: none"> ・AIシステムのセキュリティについて、あらかじめ検証及び妥当性の確認をするよう努めるとともに、開発の過程を通じて、採用する技術の特性に照らし可能な範囲で対策を講ずるよう努めることが望ましい（セキュリティ・バイ・デザイン）。
<p>⑥ プライバシーの原則 開発者は、AIシステムにより利用者及び第三者のプライバシーが侵害されないよう配慮する。</p>	<ul style="list-style-type: none"> ・AIシステムのプライバシー侵害のリスクについて、あらかじめ影響評価を行うよう努めるとともに、開発の過程を通じて、採用する技術の特性に照らし可能な範囲で措置を講ずるよう努めることが望ましい（プライバシー・バイ・デザイン）。
<p>⑦ 倫理の原則 開発者は、AIシステムの開発において、人間の尊厳と個人の自律を尊重する。</p>	<ul style="list-style-type: none"> ・AIシステムの学習データに含まれる偏見などに起因して不当な差別が生じないよう、採用する技術の特性に照らし可能な範囲で所要の措置を講ずるよう努めることが望ましい。 ・国際人権法や国際人道法を踏まえ、AIシステムが人間性の価値を不当に毀損することがないよう留意することが望ましい。
<p>⑧ 利用者支援の原則 開発者は、AIシステムが利用者を支援し、利用者を選択の機会を適切に提供することが可能となるよう配慮する。</p>	<ul style="list-style-type: none"> ・AIシステムの利用者のために、①利用者の判断に資する情報を適時適切に提供し、かつ利用者にとって操作しやすいインターフェイスが利用可能となるよう配慮、②利用者を選択の機会を適時適切に提供する機能が利用可能となるよう配慮、③社会的弱者の利用を容易にするための取組、に留意することが望ましい。
<p>⑨ アカウンタビリティの原則 開発者は、利用者を含むステークホルダーに対しアカウンタビリティを果たすよう努める。</p>	<ul style="list-style-type: none"> ・開発原則①～⑧の趣旨に鑑み、利用者等に対しAIシステムの技術的特性について情報提供や説明を行うほか、ステークホルダーとの対話を通じた意見聴取や、ステークホルダーの積極的な関与を得るよう努めることが望ましい。 ・AIシステムによりサービスを提供するプロバイダなどの情報共有・協力を努めることが望ましい。

図3 総務省によるAI開発ガイドライン案

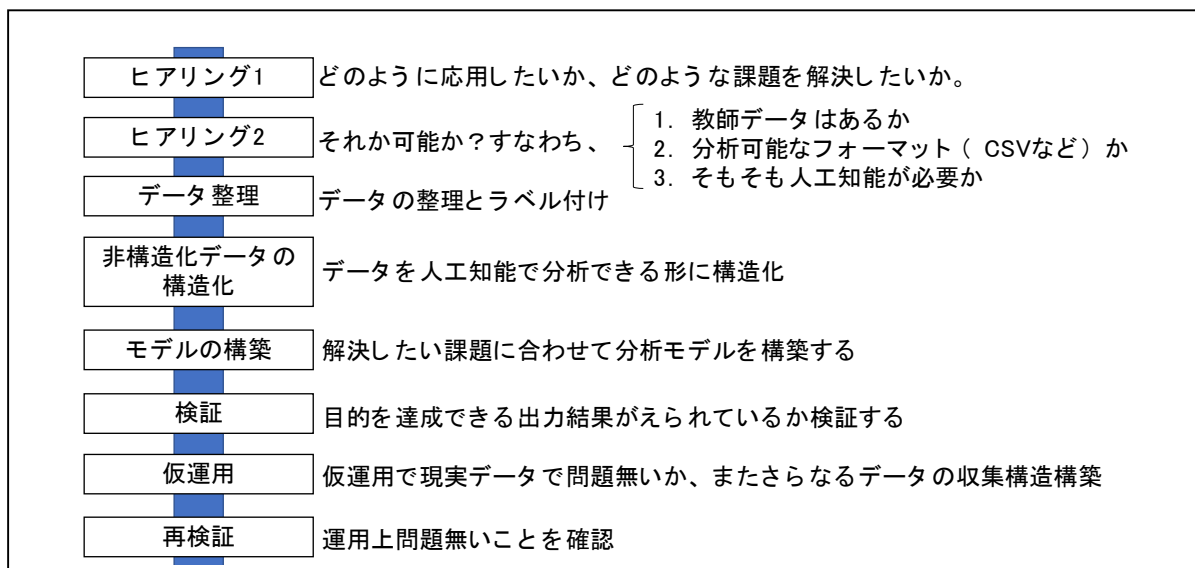


図4 機械学習を含んだ人工知能製品の開発運用フロー例

以上、人工知能の社会実装に向けて得ておくべき基本的知識を記載した。経産省発行のH29年通商白書によれば、日本のイノベーション能力は世界21位、研究開発における産学協業は18位と順位が低く、今後イノベーションの推進に向けて行動していくことが重要と考えられる。