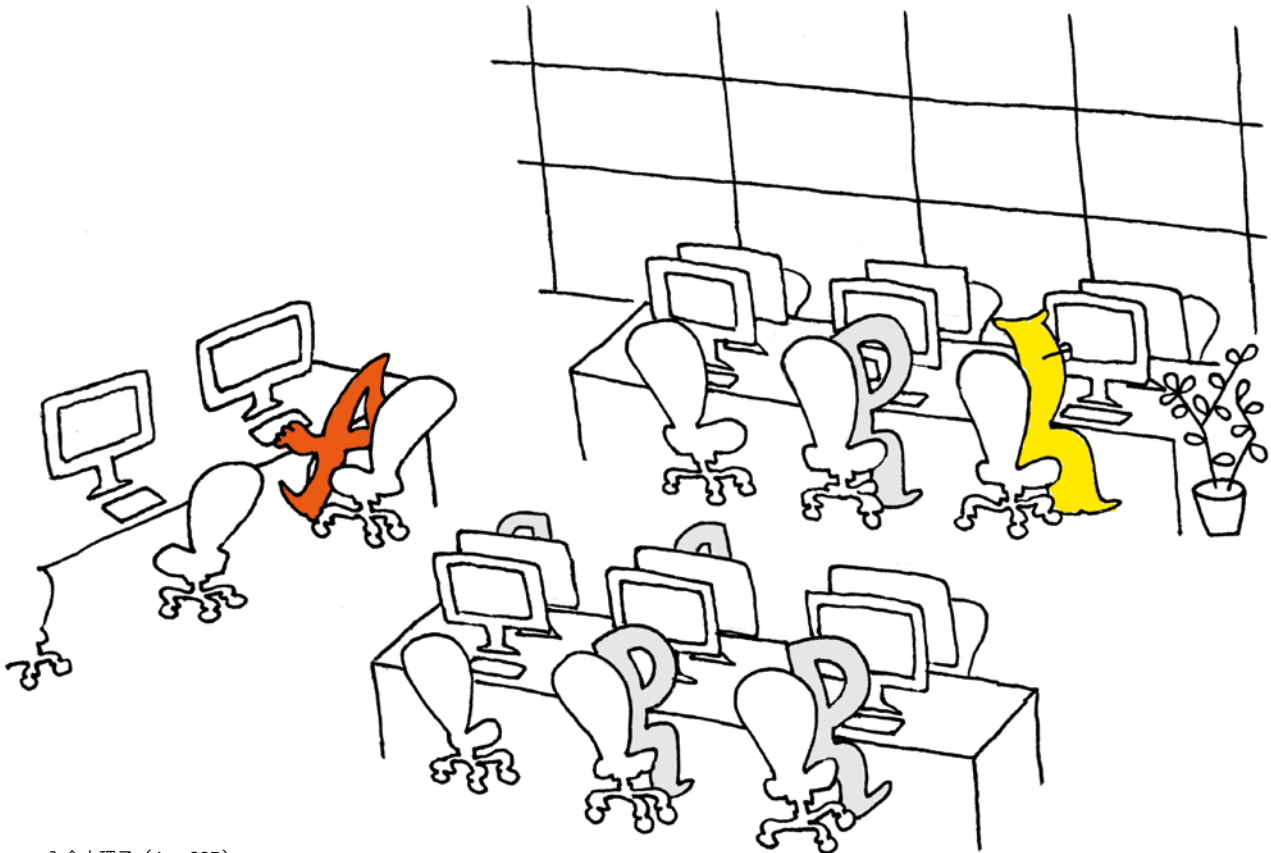


特集

人事の **A** **I** 原則



text = 入倉由理子 (4 ~ 33P)
photo = 刑部友康、鈴木慶子、宮田昌彦 illustration = 市村 譲

はじめに

今こそ、議論を始めるべきときである

医療やマーケティングなどの領域では実装が進む人工知能（AI）だが、人事という領域ではAIの活用はまだ端緒についたばかりである。しかし、データとテクノロジーをより経営に活かすべきという論調が高まるなか、人的資源をより充実させ、その配置や育成によって事業の成長に資するべしというミッションを持つ人事部が、AIを利用したいという思いを抱くのは当然だといえよう。

同時に、AIが個人の情報を扱うことにより、社会が内包する差別を助長したり、新たな差別を生んだり、また、AIが下す意思決定の不透明さが問題になったり、とAIが持つ負の側面が指摘されてもいる。そのため、AIを活用するにあたっての倫理規定や原則づくりに多くの国や企業が実際に取り組み始めている。

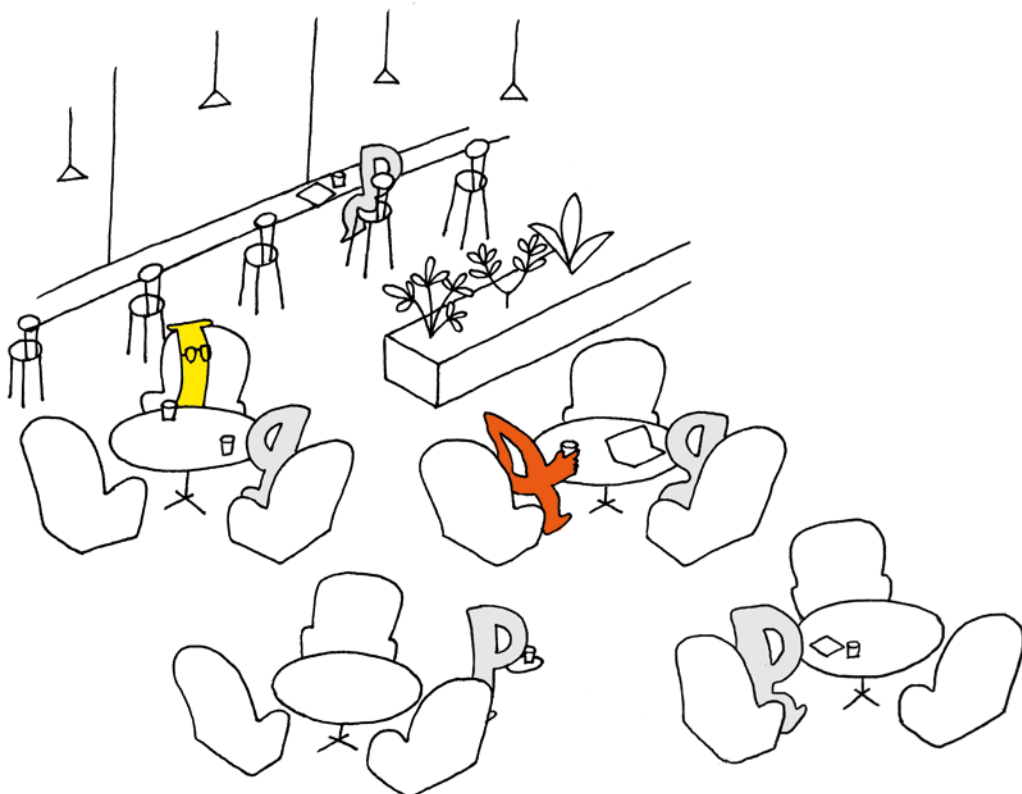
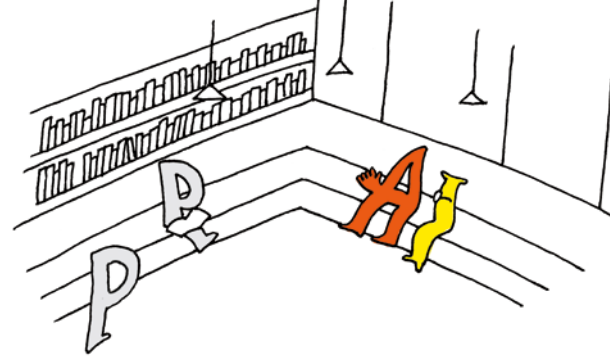
前述のように人事領域のAI活用はまだ夜明け前にすぎないため、この領域では活用にあたってのルール議論は本格的に始まっていない。だが、そもそも人事は“個人”

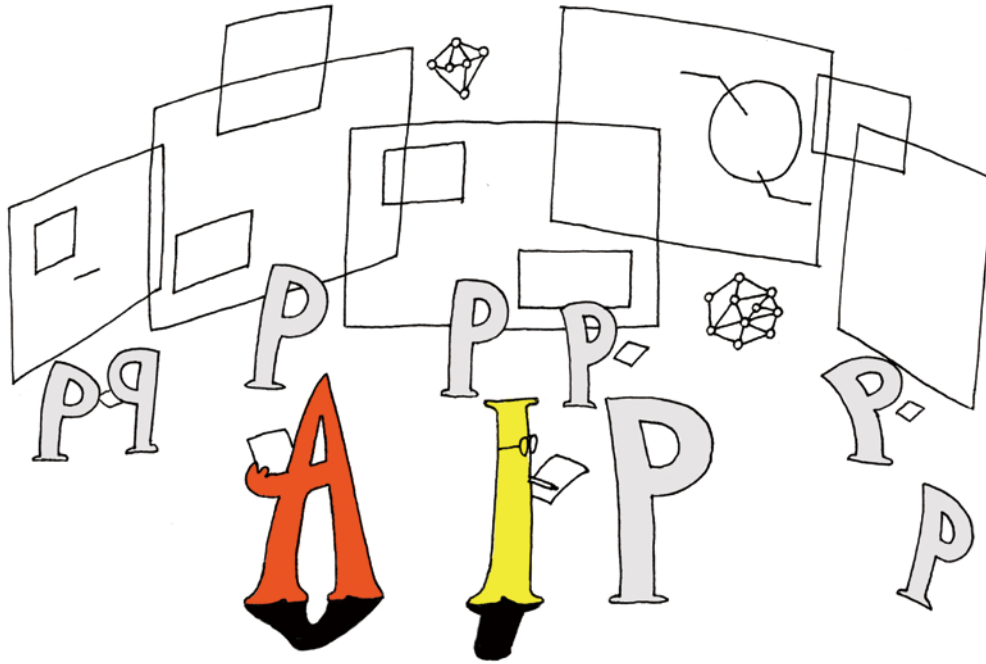
を取り扱う機能であり、人事の決定は働く人の人生に大いに影響を与えるものだ。ルールの議論を早急に始めるべきであると考え、特集を組むことを決めた。

原則づくりの前に、AIのテクノロジーは今どこまで進み、近未来はどうなっていくのか、それによって生じる課題は何か、AIの研究者や法学者、実際にAIのサービスを提供する企業などに話を聞いた。そして、さまざまな国や機関のAI原則づくりにかかわった人々からも知見を得たうえで、人事、AIの技術、法律などの専門家の力を借り、AIとともにある未来の人事が守るべき原則を議論した。

AIにかかわる技術の進展は著しく、また、人々の使い方も変化するなかで、多くのAI原則がそうであるように、原則を一度決めて終わりというのは非現実的であり、継続して議論を重ねる必要がある。読者諸氏の意見も、ぜひ聞きたいと思う。

本誌編集／入倉由理子





AIとはどのような技術か 今、どこまで進んでいるのか

第3次AIブームといわれて約5年が過ぎようとしている。

もはやそれはブームというよりはあらゆる産業にとって欠かせない技術領域の1つとなりつつあるが、進化のスピードが速いだけに実態やできることの現在地がつかみにくい。

AIとはそもそも何か、今何ができるのか、そして近い未来にはどのようなことができるようになるのかを探った。

ディープラーニングのその先へ 技術のホワイトボックス化が急がれる

AIとは何か。その問いに対して、慶應義塾大学教授、山口高平氏は「“AI”という技術はそもそもありません。AIを構成する技術は150から200にも及ぶのです」と答える。「現在の第3次AIブームは、その技術の1つ、

ディープラーニングの進化が目覚ましかったゆえに起こったものです」(山口氏)

ディープラーニングは、機械学習の1種である。機械学習とは、プログラムが大量のデータから反復的に学

習し、そこに潜むパターンを見つけ出し、そのパターンを新たなデータに適用してそのデータを分類したり結果を予測したりする技術である。これに、パターンを抽出するための特徴量まで自ら探索して学習する能力を備え

たのがディープラーニングだ。より複雑な構造のデータについて、より精度の高い分析・予測を行うことができると期待されている。センサーの進歩や、ウェブ上の画像やテキストなどのデジタルデータが爆発的に増えたこと、ハードウェアの処理能力が上がったことが相まって、ビジネスや医療などの現場での実用に堪えるレベルのサービスが次々と生まれているのだ。

「画像やテキストなどを、大量に“見る”ことで区別するのがディープラーニングは得意です。特に画像認識技術の進化は目覚ましく、工場での不良品の検出や、医療における画像診断などでは非常に高い成果をあげています。人間の見る力を補っているという意味では“人工知覚”といってもいいでしょう」(山口氏)

のちに詳述するが、人事の領域で使われようとしているのも、主にディープラーニングである。多くの人のデータをAIに与え、そこから一定のパターンを抽出させ、採用や人材配置などにおけるさまざまな分析・予測に役立てようとしている。

AIはブラックボックスからホワイトボックスへ

ところが、ディープラーニングには問題がある、という主張が生まれてもいる。それは、“ブラックボックス”だ。「ディープラーニングでは、パターンを当てはめ、結論を導き出しますが、

そこに至った理由を説明できません。意味そのものは理解していないため、出した結論の根拠がわからないのです」(山口氏)

そこで今、結論の根拠を提示できる“ホワイトボックス”型のAIに注目が集まっている。山口氏の研究領域であるオントロジーも、ホワイトボックス型AIを実現するための技術の1種である。

「知識を体系化する方法論がオントロジーです。1つの言葉に対して、その言葉の意味、ほかの言葉との関係性を定義し、リンクさせ合って意味ネットワークを構築します。たとえば“信長”という言葉を入力すると、AIが“秀吉=部下”“光秀=敵”というような関係性までを理解できるようにする。AIに“意味”を理解させるのがオントロジーです」(山口氏)



山口高平氏

慶應義塾大学理工学部管理工学科 教授
人工知能・ビッグデータ研究開発センター長

このような知識を獲得したAIは、推論の根拠も説明できる。「IBMが開発したWatsonの後継である『プロジェクトディベーター』は、意味ネットワークを使っており、ディベート大会で地区チャンピオンを負かしたことが米国で話題となりました。プロジェクトディベーターは相手の論の欠陥を見抜き、自分の論の正当性を十分に主張できるのです」(山口氏)

山口氏の研究室で開発された意味ネットワークを搭載したロボットが、東京都の小学校で子どもたちが行うフリーディスカッションに加わるという実験も行われている。あるときのテーマは、“地球温暖化に対して個人ができること”だ。子どもが“氷が溶ける”などと発言すると、ロボットが「今、氷が溶けると言ったね」と口をはさむ。「氷が溶けると、海の水が増えるね」と視点を広げたり、難しい言葉には解説を加えたりする。それによって子どもたちの議論を活性化していくのだ。

「企業でも、会議の場にこのようなロボットがいるのが当たり前になるでしょう。ひらめきとは、離れた領域にある知識と知識の結びつきから生まれます。それは人間の得意分野だとされてきましたが、意味ネットワークがもっと膨大になれば、意味のリンクを辿って人間が思いもよらなかった発想を提示してくれる可能性もあるでしょう」(山口氏)

“利用知”で世界をリードするための原則づくりを

ただし、問題は「意味ネットワークをつくっているのは人間。それには膨大なコストがかかる」（山口氏）ことだ。そこで、ディープラーニングと意味ネットワークの融合が注目を集めている。意味ネットワークの構築を自動化し、スピーディにすることに、各国

の研究機関や企業がしのぎを削っているという。

「そうしたなかであって、日本が後れを取っていることは否めません。AIに投入される国家予算は、ほかの領域よりも潤沢とはいえ、開発ではるかだに先行する米国や中国とは大きな差があります。もはや追いつくのは難しいかもしれません」（山口氏）

そんな日本が戦える分野として、山

口氏は“利用知”を高めることを挙げる。「利用知を高めるとは、エンドユーザーがいろんな使い方を試して、よりメリットを享受できる使い方を、実地で生み出していくことです。技術というものはコインの裏表で、使い方によって善にも悪にもなる。だからこそ、ユーザーの使い方の原則をきちんと定め、善となる使い方を世界に示すことが求められます」（山口氏）

開発が進む、自律性、汎用性を持つAI ますますルールの存在が重要に

今のところ、AIは目的に応じて事前に学習する必要があるし、その目的は、特定の狭い範囲のものでなくてはならない。これに対し、自ら幅広い範



栗原 聡氏

慶應義塾大学理工学部管理工学科 教授
電気通信大学人工知能先端研究センター 特任教授

囲のことを判断し活動するAI、つまり、より人間の知的活動に近い働きをするAIの研究開発も進んでいる。その研究者の1人が、慶應義塾大学教授の栗原聡氏である。「現在のAIは、人間の能力を拡張する“高度な道具”でしかない。しかし、AIの研究はそもそも“知能”を“人工的に”つくりたいという科学者の純粋な欲求から始まったものです。その思いを受け継ぎ、人の知能を模したAIの研究開発に取り組んでいます」（栗原氏）

“人工”の“知能”を持つAIの開発が進む

こうしたAIの開発において最も高い壁は、人の知能が何たるかが解明されていないことである。「人工心肺

にしる、人工ダイヤモンドにしる、“人工〇〇”というものをつくり出すには、その“〇〇”の部分の仕掛けや構造がわかっていることが前提です。ところが、知能とは何かということはまだ議論し尽くされていません。まず、“知能”を定義しなければならないのです」（栗原氏）

栗原氏は、知能を「目的のために生き抜く力」と定義づける。「生物、特に人は目的のために自律的、能動的に動き、汎用性を獲得し、目的を実現しようとします。ですから、真の人工知能を実現するためには、この自律性や能動性、汎用性を備えればよいということになります」（栗原氏）

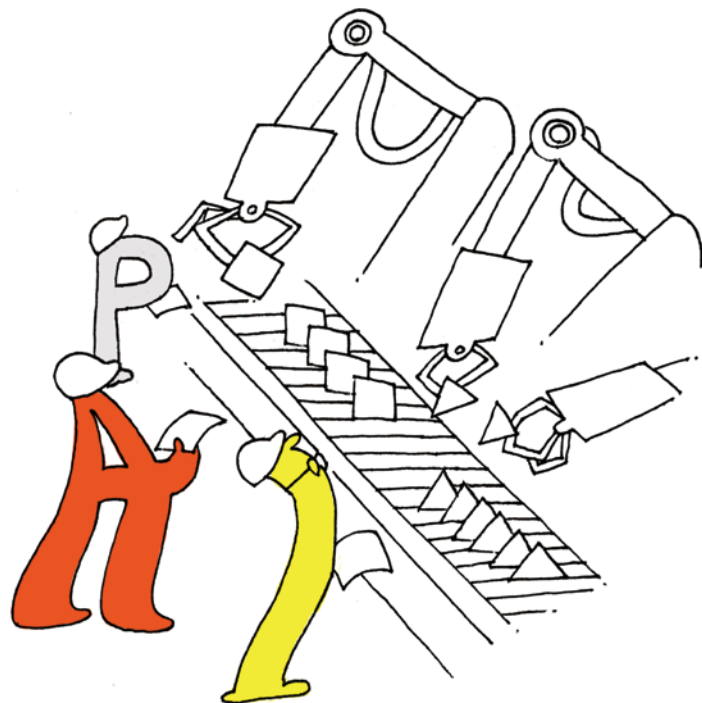
たとえば、センサーを搭載した車を走らせ、集めたデータでディープラー

ニングを行うことによって、自動運転技術はかなり進化してきた。また、目的地を設定すれば最短で行く方法を見つけ、到達できる技術も既にある。「しかし、高速道路のような条件の限定された道は別として、人や自転車も通る一般道での自動運転はまだ難しい。人が運転するとき、混み合った道では、人やほかの車と譲り合ったり、駆け引きをしたりしながら運転します。ところが、自動運転の車は、人が至近距離にいる以上、安全を守るために1ミリも動けません。人は、知覚というセンサーだけを使って運転しているわけではなく、知識や経験を総動員しているのです」(栗原氏)

このとおり、現在のAIは、まだ本当の意味での知能を持つには至っていない。栗原氏の研究では、AIにも、目的地に到達するという目的に向かって、臨機応変に譲り合い、駆け引きし、配慮しながら行動する自律性と能動性を求めようというのだ。

また、汎用性は“学んだことをほかの場面に応用できる性質”である。私たちは、本を“読むもの”として認識しているだけでなく、暑いときにはうちわの代わりにあおぐものとしても使うことができる。この、人間の持つ論理性を超えた能力を、どのようにAIに搭載するかというのも1つの大きな課題だという。

「私たちが人間に近いAIの開発を急ぐのは、日本は、今後、超少子高齢化時代に突入し、労働力が大幅に足



りなくなるといわれているからです。目的をともにし、配慮や譲り合いもしながら人と協働できるAIが、確実に必要となります」(栗原氏)

AIが差別や攻撃をするならそれは人間社会の“負”

このような人と同じレベル、あるいは人を超えるようなAIの出現はSF小説の世界での話のようにも思えるが、「5年後には、最初はAIスピーカーというようなシンプルなものかもしれませんが、目的を与えれば自律的・能動的に考えるAIが登場する可能性は高い」と、栗原氏は主張する。「背後に検索エンジンがあつてQ&Aに答えるというレベルを超えて、目的や場、相手のニーズに合った回答を自ら探索し、自然な会話を続けるAIは早晩登場するでしょう」(栗原氏)

しかし、人間に近い知能を持つとなれば、栗原氏がいうような協働できるAIだけではなく、人に脅威をもたらすAIが登場する可能性もある。た

とえば、AIの兵器にこうした自律性が備わると、人間に対する大いなる脅威となる。2019年8月、自ら標的を選び、自らの判断で攻撃をする自律型AI兵器の開発を認めないという指針を含む報告書が国連の専門家会合にて採択された。

「力がある技術は、負の面も大きい。原子力は有用なエネルギーですが、それを使いこなせなければ事故につながり、爆弾にすれば人類をおびやかす、というように。自律的に動くAIもパワーがあるからこそ、使う側がリテラシーを身につけ、一定のルールを守ることが求められます」(栗原氏)

同時に栗原氏が指摘するのは、「AIはそもそも人間社会から学ぶ」ということだ。「AIが差別をしたり、攻撃をしたりするならば、それは人間社会そのものに内在する“負”だと考えたほうがいいでしょう。『日本社会で育ったAIはいいね』といわれるような社会にするための原則をつくっていくべきでしょう」(栗原氏)

AIが人間の知識や知恵を拡張し、 人事の“Guess Work”にエビデンスを与える

では、人事の領域では今、どのようにAIが活用されているのか。実用化が進むAIとして知られるのが、IBMが開発したWatsonである。「私たちはWatsonを、同じAIでも“Artificial Intelligence”ではなく“Augmented (拡張された) Intelligence”だととらえています。人間が諦めていたこと、できなかったことを支援し、人間の知識や知恵を増強、補強するものと考えれば、“Augmented”という概念がフィットします」と説明するのは、日本IBMで人事領域のサービスを統括する石田秀樹氏である。

石田氏によれば、人事領域こそ、やりたかったけれど諦めていたことが多く、AIの活用が必要だという。その理由は、「人事の仕事の多くが“Guess Work”であるため」(石田氏)だ。「採用や配置においては、たぶんこの人はこうだろう、と予測することが多い。なぜその人を採用するのか、なぜその配置が最適なのか明白な理由がないまま行われてきました。AIはそこに、エビデンスを提供してくれる可能性があるのです」(石田氏)

課題はデータが 整備されていないこと

事実、IBMが2018年に行った人事担当役員向けのサーベイ^(*)によれば、AI活用のニーズは要員計画と最適配

置で高いという。これまで多くの企業で、採用選考や人事異動は手作業で行われてきた。確かにノウハウは存在するのだが、それは職人芸ともいえる暗黙知であり言語化されたものではなかった。またある一定の人数までは丁寧に検討できるとしても、数千、数万人を相手にするとき、すべての個人の能力やスキル、志向を見られるわけではないため、最適な人材の採用や最適な人材配置をするには限界があった。しかし、「AIを活用すれば、人事は従来よりも、より広い範囲で、より細かい人材情報を手に入れることができるため、適切な人を見落とすのを回避し、さまざまな機会損失を防ぐことができる」という(石田氏)。



石田秀樹氏

日本アイ・ビー・エム パートナー
グローバル・ビジネス・サービス事業部
コグニティブ・プロセス変革
タレント&トランスフォーメーション リーダー

「一方で、その会社にある性別や学歴によるバイアスをAIが学習し、差別を助長するという課題が挙げられますが、逆にそのバイアスを統制するアルゴリズムを組めば、バイアスフリーな選択をできるようにもなるはず」(石田氏)

ただし、人事でAIを活用する場合、課題もある。「最大の課題は、データが揃っていない、使われていない、更新されていないこと」(石田氏)だという。「日本企業は、実は膨大な人のデータを持っています。採用時のエントリーシートや面接結果、入社後の上司からの評価、360度評価、従業員サーベイ、管理職昇進時に行われる試験や論文……。しかし、これらは使える状態に整理されておらず、実際に使われることもありません。また、本人の異動希望や保有するスキルに関するデータは、最新のものに更新されていないことも少なくありません」(石田氏)

人に関するデータの取り扱いにはもちろん十分な注意が必要だが、「自分に関するデータを使って、キャリアの方向性を改めて考えるなど、データが整備されていれば個人にとってもメリットがある。これをもっとアピールして、働く人自身が進んで自らのデータを適宜更新し、最新の状態に保とうという意識を持つ仕掛けが必要」と石田氏は訴える。

(*) Cognitive Computing Survey (IBM, 2018年)

XAIの進化で、 AIのブラックボックス問題は解決するのか

ディープラーニングによる分析や予測における課題の1つは、その判断の根拠が“ブラックボックス”である、つまり説明されないことだと既に述べた。最近では、この課題を乗り越え、判断の根拠を“ホワイトボックス”で示せるAI、“説明可能なAI (Explainable AI、XAI)”の研究が進められている。7ページで紹介した、意味ネットワークを活用したオントロジーなどの技術も、XAI化に寄与するものだ。また、ディープラーニング自体の説明力の向上を目指す研究者もいる。「2016年頃から急速に、XAIの論文が世界的に増えました」と、話すのは、大阪大学産業科学研究所助教の原聡氏だ。

変数の重み付けを明らかに

「方法はさまざまですが、最も一般的なのは、AIに与えたデータのうち、どの項目を見て決めたのかをAIにアウトプットさせる方法です。実際にどのようなロジックで分析したのかというアルゴリズムまでは説明されなくても、



原 聡氏
大阪大学産業科学研究所 助教

どの変数が重み付けされているかが明らかになります」(原氏)

たとえば従業員の評価において、重きを置かれたのが“過去の業績”と“360度評価のフリーコメント”であるとわかれば、多くの人事はその推論の方向性は間違っていないと感じるだろう。これが“出身大学”と“子どもの頃の運動経験”に着目しているとなれば、そのアルゴリズムを使い続けるのはまずいと感じるだろう。つまり、ある程度の納得と判断ができるのだ。

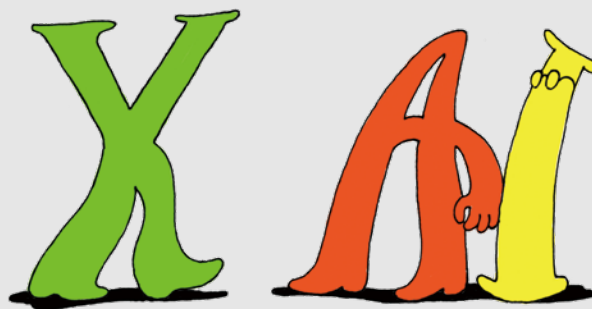
ただし、何が重み付けされているか説明されたとしても、データを提供し、AIによる推論をされる側が納得できるとは限らないという問題は残る。「AIの判断によって不利益を被ったと感じる人が、その不当性を主張するとして、それに対して、納得のいく説明ができるほどには現状のAIは賢くないのです。人間が情報を咀嚼して、説明する必要があるわけですから、AIで分析や予測を行う場合、人の介入を求めるAI原則が多いのは当然のことです」(原氏)

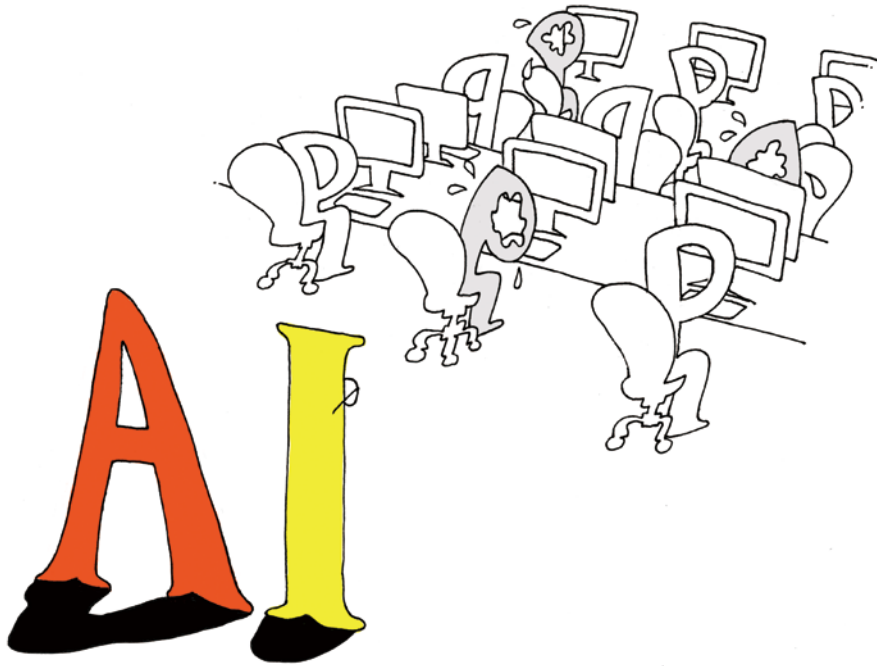
AIのおかしな挙動を説明

XAIの開発の目的を、データを提供する側に推論の根拠を説明することとするならば、できることはまだごくわずかだ。しかし、「実は、XAIの開発には

もう1つ目的がある。その目的に照らせば、それなりに能力を発揮する」(原氏)という。「それは、AIがおかしな“挙動”をするときに、どこに問題があるのか、どうすればより精度が上がるのかを開発サイドが確認するという目的です。従来は、いってみれば当てずっぽうでアルゴリズムを修正してきましたが、AIによる重み付けの説明などがあれば、どこを修正すべきか仮説をより緻密につくれます」(原氏)。たとえば、AIが犬の画像を見て犬だと認識するのに、犬の顔ではなく背景や地面など無関係な部位に注目しているとわかれば、修正のヒントが得られるのだ。

ただし、「こうした状況は、いずれ変わっていく」と原氏は予測する。「多くの人にとって、今のところAIは未知の領域だからこそ、使うことに不安を感じ、根拠を説明してほしいと思うのです。しかし、ほかのテクノロジーと同様に、社会のなかでさまざまに実装され、人々の間にAI活用の経験知が蓄積されれば、次第に“これまでも使ってきたのだから大丈夫”と思えるようになるはず。そのためにはAIを導入し、こう使ったらうまくいった・失敗した、失敗のあとこを直したら問題が解消されたというような試行錯誤の経験を重ねることが重要なのです」(原氏)





AI活用の進展で見えてきた そのリスクと課題

加速度的に進化するAI。研究者たちが強調したのは、「AIは人間の使い方によって、善にも悪にもなる」ということだ。活用法を誤ると、どのようなリスクや課題が生じ得るのか。専門家や、実際にAIサービスを提供する企業の経営者に聞いた。

AI活用にあたって考えるべき 日本国憲法に定められた“個人の尊重”

まずは、日本という国や国民のあり
よしの基盤である、憲法の視点でAI
活用における課題を見てみたい。「AI
は使い方を誤ると、『個人の尊重』と
いう憲法の根本原理を毀損する可能
性がある」と指摘するのは、『AIと憲

法』（日本経済新聞出版社）の編著者
である慶應義塾大学法科大学院教授
の山本龍彦氏だ。

AIが人に関する推論を行うとき、
ディープラーニングなどの技術を活用
し、大量のデータを学習し、共通の

属性によるセグメントをつくる。その
セグメントによって、そこに入る人は
このような行動を取る、このような能
力があるという結果に結びつけ、新た
な個人の情報を取得したとき、どのセ
グメントに当てはまるかを識別して、

その人の行動や能力を予測したり評価したりする。「どれだけセグメントが細分化されたとしても“このセグメントに含まれる人は〇〇する確率が高い”という確率論であって、その人自身が絶対にそうするとはいえません。セグメントによってつくられる“データの分身”と実存する自分の間には、必ずズレがあるのです」(山本氏)

また、いくら大量のデータによって分析されたとしても、「データが完全であることはあり得ない」(山本氏)という。「たとえば、オフィスのトイレなどでの同僚同士の気軽な会話。ここには360度評価にも出てこないような上司に対する評価、仕事への不満といった本音が詰まっているかもしれません。しかし、こうした場で発せられる感情や本音はプライバシーの観点からは取れないデータです。このように、本人の行動や感情をすべて漏れなくデータとして取得できるわけではないので、データには必ず偏りが生じる。すると、やはりデータの分身と自分の間には誤差が生じることになります」(山本氏)

ズレのあるデータの分身が自律的な人生の邪魔をする

こうした問題がある以上、AIを利用して人を評価したり、その人に対するなんらかの対応の結論を出す場合に、“個人の尊重”との矛盾が起こり得るのだ。山本氏によれば、個人の尊重には4つの側面があるという(右表)。「AIの活用が損なう可能性があるのは、特に“集団的拘束からの自由”と“個人の自律”という側面です」(山



山本龍彦氏
慶應義塾大学法科大学院 教授

本氏)

「“集団的拘束からの自由”というのは、集団的な属性によって個人の能力や生き方を短絡的に決めつけられない、ということの意味します。これは前近代の身分制度の反省です」(山本氏)。身分制度の下では、農民の家

に生まれた者は農民として生きることが定められ、職業選択や婚姻の自由がなかった。「これを否定して、誰もが1人の独立した存在として見られ、判断してもらえるべきだ、という考えが重要になったのです」(山本氏)

もう1つの“個人の自律”とは、個人が主体的に自分の人生をデザインし、それに従って歩むことができなければならないということである。「この2つの観点に照らすと、確率的存在にすぎない自己のデータの分身によって判断される個人は、セグメントという“集団”を根拠にその能力などが決めつけられ、人生の主体的な自己決定が妨げられるという問題が生じることになります」(山本氏)

もちろん、従来の人による評価においても、「〇〇大学出身者は活躍する」「女性は〇〇が苦手」というような、属性による決めつけ、すなわちバ

個人の尊重とは

1. 人間の尊厳

個人は“人間”として尊重されなければならない。人間としての生命や身体の安全が確実に保障されなければ、個人として尊重されることもない

2. 集団的拘束からの自由

個人は平等に尊重されなければならない。身分制社会では、個人は身分のような集団的で固定的な属性によって短絡的に評価され、あらかじめ自らの生き方を規定されてきたが、近代憲法では、個人を集団的屬性ではなく、それぞれの人格を持った“個人”として評価し、事前に集団的屬性によって生き方が決めつけられることがないように配慮すべきであるとする

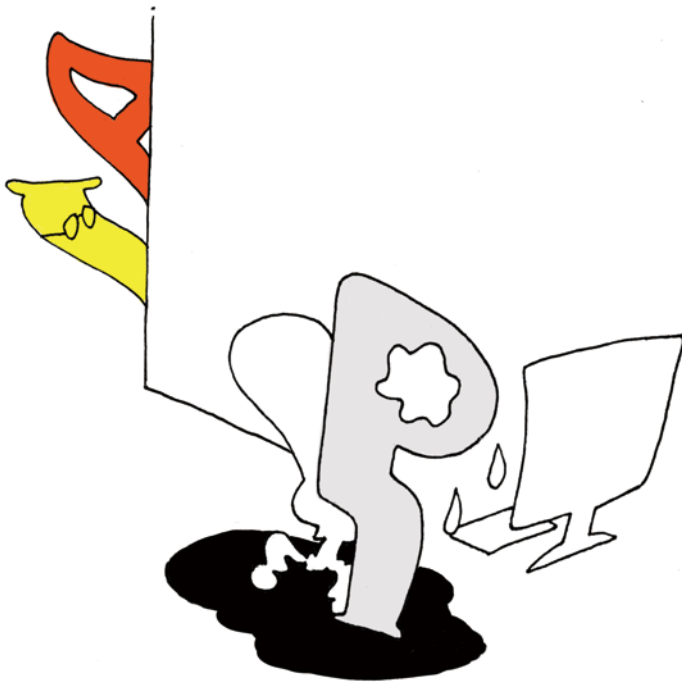
3. 個人の自律

個人は人格的に自律した存在として尊重されなければならない。個人が自律の能力を持つことを前提に、誰からも命じられることなく主体的に自己の人生をデザインしていくことを認める

4. 多様性、個別性の尊重

個人が自律的・主体的に決定・選択した結果を尊重しなければならない

出典：『AIと憲法』(山本龍彦編著、日本経済新聞出版社)より編集部作成



イアスは存在し、個人が1個の独立した存在として評価されないことはある。だが、相手が人間であるならば、たとえば採用選考であれば、相手が持つバイアスの見当もつくし、「エントリーシートのある部分の書き方が悪かったのかもしれない」「面接でのあの態度が悪かったのかもしれない」など、不採用になった根拠をある程度推測することもできる。

「人間が相手の場合とは異なり、AIによる評価の場合、インプットされるデータが多様かつ大量であり、どのようなデータが、どのような重み付けで用いられ、結論に至ったのかわからないという問題が生じます。判断過程のブラックボックス化は、自己改善のきっかけを与えてくれず、排除の不条理性が高まることを意味しています」(山本氏)

だからといって、「AIの活用を規制すべき、というつもりはない」と、山本氏は主張する。「AIは、活用の仕方によっては、個人の尊重に資する可

能性があるからです」(山本氏)

AI活用のメリットを享受するためにリスクを減らす

既に述べたように、人間の能力には限界があることを私たちは認めなければならぬ。たとえば採用の場面で、エントリーシートを数千、数万枚と読まなければならないとき、人は疲れてしまえば、じっくりと読み込み、適正に判断することができなくなるだろう。かといって、複数人で読めば判断基準が異なるという問題も生じる。人材配置の場面でも、社員全員の能力やスキル、志向を個別に見て、すべての人に対して適材適所を実現することは、実際には困難だ。

「人間の限界を、AIのサポートによって伸ばさせることが可能になります。それによって、個々の能力や人格を尊重し、自律的な生き方のデザインの支援もできるでしょう。個人にとっても、AIの分析や予測によって、自分では気づかなかったような強み・

弱みを発見し、能力を伸ばしたりキャリア選択の参考にしたり、ということが可能になります。“AIによるメリット”を享受するためにも、AI活用によるリスクをなくしていく必要があるのです」(山本氏)

リスクを減らすための透明性とアカウンタビリティ

では、リスクをなくしていくには、どのような原則が必要だろうか。「最も重要なのは、“透明性”と“アカウンタビリティ”です」(山本氏)

透明性とは、AIによる推論や分析にかけられる個人に対して、AIを使う目的、使用するデータの種類とその重み付け、使用の範囲などをあらかじめ提示すること。つまり、“隠れてこそそやらない”ということだ。

そして、AIによって導き出された評価は、どういうロジックで、どのような算定をしたのか、その説明を明確にすること。AIが導き出した結果を用いた行為の、責任の所在を明らかにすること。これがアカウンタビリティだ。

「透明性とアカウンタビリティが確保されていれば、個人は結果に異議を申し立てること、自分とデータの分身の間の齟齬を指摘することもできるでしょう」(山本氏)

どんな社会にしたいのか 選択を迫られている

もう1つ、重要なことは、「“More Data”の誘惑を断ち切ること」(山本氏) だという。「データ量と予測精度はトレードオフの関係にあります。予

測の精度を上げようと思うと、より多くのデータを取りたい、ということになりがちです。しかし、多くのデータを取ろうとすればするほど、プライバシー侵害の危険は増しますし、ブラックボックスの闇も深くなります」(山本氏)

たとえば、あらゆる場所に監視カメラを設置し、個人の行動履歴をつぶさに取ることによって、テロを起こす可能性のある人を早期に特定できる

かもしれない。テロリスト予備軍の早期発見を優先し、プライバシーを犠牲にするのか。あるいはテロリスト予備軍の発見が難しくなると、人々のプライバシーを守るのか。何を守り何を犠牲にするのかは、社会のコンセンサスの問題だ。

「たとえばEUは、プライバシー保護のほうに力点を置いた社会であり、データの保護は基本的人権に属する問題だととらえています。EU基本

権憲章というEUの“憲法”の第8条に、データ保護の権利が書かれています。EUの『一般データ保護規則(GDPR)』は第8条を具現化する立法であり、その前文にもデータ保護は基本的人権だと明記されています」(山本氏)

今、AIが私たちに突きつけているのは、どのように人権を尊重し、どのような社会を志向するのかという選択の問題なのだ。

AIによるスコアリングが引き起こす “バーチャル・スラム化”

AIによる分析・評価が、既にビジネスとなっている例がある。AIによるスコアリング(格付け)ビジネスがそれだ。「スコアリングビジネスには、複数の企業が参入したり、参入を表明したりしています」と話すのはデロイト トーマツ グループで、デジタル分野における社会課題およびテクノロジーによる社会課題解決に関するリサーチやルール形成支援を行う矢守 亜夕美氏だ。最も有名なのは、中国のアリババグループが運営する芝麻信用だ。アリババグループでは、アリペイの一機能として、利用者の全支払い履歴データを管理し、資産状況や社会的ステイタス、SNS上での人脈などの情報を加味して、個人の信

用力をスコアリングしている。

「こういったAIによるスコアリングが社会に浸透していけば、それまでは所得や勤務先などによってローン審査に通らなかった人々が、その人が生活態度として積み上げてきた“信用”の力によってお金を借りられるようになるなど、個人にもメリットがあります。一方で、“バーチャル・スラム”という新たな貧困層が生まれる可能性という、負の側面も持ち合わせています」(矢守氏)

あらゆる機会や 権利が剥奪される

AIによるスコアリングが導入される前、つまり現在までの世界にも、



矢守 亜夕美氏

デロイト トーマツ グループ
Social Impact / Regulatory Strategy マネジャー

ローン審査や住宅の賃貸、企業の採用選考など、個人の信用が問われる領域はもちろんある。「しかし、基本的に、金融は金融、不動産賃貸は不動産賃貸というように、領域ごとに評価基準は異なっていて、金融と不動産賃貸の間で評価を共有することはほとんどありません。つまり、金融面で信用力がないとしてお金を借りることができなかったとしても、それが賃貸住宅を借りることや、就職に影響することはほとんどないのです。就職することによって金融面での信用力も上がるなど、一度何かでつまずいても、再チャレンジができます」(矢守氏)

今後、AIスコアリングが多様な分野で共通に用いられるようになることで、この社会が一変する可能性があることを矢守氏は指摘する。「一度何かの領域で低いスコアをつけられて

しまった場合、ほかの領域での審査や評価にもそのスコアが影響を与える可能性が出てきます。あらゆる領域で、機会や権利が剥奪され、格差が固定化してしまう。こうして生み出されるかもしれない新たな社会的弱者グループがバーチャル・スラムなのです」(矢守氏)

AIが、領域を問わぬ多様かつ大量のデータを連携させながら飲み込んでいくことの現実的なリスクは、既に顕在化している。「単体ではそれほどセンシティブではないデータでも、いくつかのデータを連携・突合し、類推することで、人種や性的指向や性自認など、センシティブな個人情報 guidance 導き出され、それによって不利益を被ることがあります。データを取得される側は、データの使用目的や使用される範囲を認識する必要があります」(矢守氏)

矢守氏のチームでは、AIによるスコアリングが広く浸透することによって、バーチャル・スラムという新たな貧困層に陥ってしまう可能性のある人の数は、主要国(G20)において3億4000万~5億4000万人にのぼると試算した(「AI時代の新たな貧困—『バーチャル・スラム』とは」デロイト トーマツ グループ)。これは仮説であるが、私たちが過度にAIに依存したときに来るべき世界でもある。

そこで重みを増すのが、使い方のガイドラインだ。「その問題にかかわるステークホルダーになり得るすべての人、人事でいうならば人事部の人々や人事領域の専門家だけではなく、データを取られる側の人々やサービス提供者、法律家などを広く交えて議論し、特定の立場に偏らない現実的なルールを決めていくべきでしょう」(矢守氏)

責任の所在、不審者の特定…… 事業の現場で見えてきたリスク

AIサービスを実際に提供する企業サイドからは、どのような課題が見えてきているのだろうか。

スペクティは、世界中のウェブやSNS上にアップされた映像や画像を、ニュースの情報源としてテレビなどのメディアに提供している。映像・画像

を検索するのは、AIの仕事だ。「このとき問題となるのは、その映像・画像の“真偽”です。フェイクニュースやいたずらではないかどうかを見定め、それを排除することは、私たちのサービスの質を担保するための重要な要件です」と、同社の代表取締役CEO

の村上建治郎氏は話す。

AIが検索する映像・画像の真偽を、AIのフィルターで判定するシステムはあるという。過去にあった偽情報をデータベース化し、また、偽情報を書く人の文章のパターンや使いがちな単語も登録されており、それによる

判定を行う。今や映像や画像の合成が容易なため、これまでにアップされたほかの映像や画像と同じ部分はないかという比較検証もしている。「これらは、ものの1分もあればできるAIの得意分野です。しかし、それでも最終的に判断するのは人です。私たちのスタッフもチェックしていますし、クライアントにもお願いしています。事実でなかったとき、“責任を取る”という行為はAIにはできないからです」(村上氏)

責任の所在はどこにあるのか

「従来の機械は、人がプログラムしたとおりにしか動かないものでした。多くの情報から自ら結論を導き出すAIだからこそ、責任の所在はどこにあるのかという問題が生じます」(村上氏)。これは自動運転車での事故など命にかかわる場合に、よりクリティカルな問題となる。「責任を負うのはメーカーなのか、それを操作している

人なのか。もし、一方的にメーカーが悪いということになれば、AIの研究と開発は確実に足踏みするでしょう」(村上氏)

また、村上氏は、AIの“ブラックボックス問題”についてこう話す。「AIの判定がブラックボックスであることの問題が取りざたされていますが、もし、すべての根拠を明確にするならば、AIではなく従来のプログラムで十分。明白な基準と手順によって、100回やれば100回、同じ答えを出せます」(村上氏)。AIのよさは、「もっと曖昧な、人間も判定しにくいところを過去の膨大なデータを学習させることで、そこから導き出した何らかの基準によって判定を下してくれること」だと、村上氏は言う。

「責任を取るとは、なぜこういう判断をしたのかを説明することでもあります。AIの判定の根拠を責任者が知っておく必要があることとなりますが、技術そのものの特徴とのジレンマが生じるのは間違いありません」(村上氏)



村上建治郎氏
スペクティ 代表取締役 CEO

高度な安全の実現のために課題に丁寧に対応する

SEQSENSEはAIを搭載して自律的に動く警備ロボットを開発する。警備ロボットは、2019年8月より東京の大手町のオフィスビルで実際に使われ始めている。

「私たちが、最も気を配るのは“安全”です」と、同社CEOの中村壮一郎氏は言う。ロボットはAIによってビル内の地図を学習し、自ら動くルートやその安全性を判断して巡回警備する。人が近づいてきたらどの程度手前で止まるのか、死角があって人とぶつかりそうな場合、どのようなルートを取るのかなど、細かくプログラムされているという。「たとえば非常にレアなケースでも、実験の最中に生じた課題に丁寧に対応し、積み重ねることによって高度な安全を実現しようとしています」(中村氏)



AIを使える環境をつくる 使う側が慣れる

その一方で、社会で実装されるためには、「AIを使える環境を整備し、同時に使う側が慣れることも重要」と中村氏は言う。SEQSENSEのロボットが導入されているのは、オフィスビルだ。「オフィスビルは通路が広く、比較的ものが少ない、そこにいる人たちはすべて大人であるなど、ロボットが“動きやすい”環境が整っています」(中村氏)

同社は、人手不足を解消するためのロボット開発をミッションとして掲げている。「ロボットの使途として、病院でのものの運搬など、医療を担う人々の支援も検討しました。しかし、松葉杖を使って歩く人や高齢者もいる病院への導入は、私たちの経験知が少ない段階では難しいと判断しました」(中村氏)

子どもが多い商業ビルも、ひとまず避けた。「自律的に動くロボットに人が慣れることも重要です。子どもは大



中村壮一郎氏

SEQSENSE Co-Founder、代表取締役

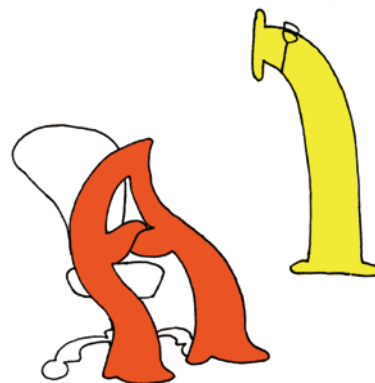
人と違って、ロボットがいれば寄ってくるし、触ります。社会のなかにロボットが増えていけば、ソフトバンクのPepperがそうであるように、だんだん“いるのが珍しくない”状態になり、共存が容易になるのです」(中村氏)

また、警備という領域において、気になるのはプライバシーの問題だ。警備では不審者を特定する必要があるが、ビデオや写真を撮られることに対する忌避感を持つ人は少なくない。「どのようなデータを取得し、どう使うかは、結局はデータを取得される側と合意形成できるかどうかという問題です。警備については、安心と安全のためのコストだと理解されるため合意は得やすいと思います。しかし、メリットが少なかったり見えなかったりする状態では、多くの人はデータ提供に躊躇するでしょう」(中村氏)

このとき求められるのは、「データを取得される側からどのように見えるかという想像力」(中村氏)だ。「開発者同士のみ、サービス提供側だけ、というように限られた人としか会話しなければ、データを取られる側の不安やリスクにも気づけません。こと細かに法律やルールをつくるというよりも、まっとうな倫理観に基づいて自ら判断する力が求められます」(中村氏)

人事領域における課題や リスクはどのようなものか

ここまでの取材によって、AI活用におけるさまざまな課題とリスクが見えてきた。具体的には、AIの活用によって憲法に定められた個人の尊厳やプライバシーを毀損するリスクや、



不当な差別を発生させるリスク、AIを使った結果に対する責任の所在が不明瞭という課題、AIの判断がブラックボックスであるという課題、AIが存在することに多くの人が慣れないなか、不安や不信が高まるという課題などである。

データに基づく経営を志向する企業は増えつつあり、今後、人事の領域でも多様なシーンでのAIの活用が見込まれる。慶應義塾大学の山本氏が指摘したように、AIのメリットをきちんと享受するためにも、私たちは人事領域におけるAI活用の課題やリスクを知っておく必要がある。そこで、取材をもとに、リスクとはどのようなものを整理したものが右の表である。

今後、AIが進化し、自律的なAIなどが生まれれば、また別の課題やリスクが生じるだろう。ただし、人事領域における活用は、データをもとに機械学習やディープラーニングで分析・予測することがほとんどというのが現状である。まずは、機械学習やディープラーニングを中心とした活用シーンや課題、リスクを表で挙げている。このあと、原則を検討するにあたり、しっかりと認識しておきたい。

人事領域でAIが使われるシーンと、そこにある課題・リスク

人事の領域	AIが使われるシーン	課題・リスク
採用	初期選考でのスクリーニング	<ul style="list-style-type: none"> ● 過去の実績に基づくバイアスがそのまま踏襲されてしまう ● データの組み合わせによって、実際に取得した情報以外のことを勝手に類推してしまう ● ある特定の“カテゴリ”に属すると判断され、それによって採否が決まってしまう ● アルゴリズムがブラックボックス化しており、どのような根拠で採否を判断したかが説明できない
評価・ 配置転換・ 昇進	人材を評価するためのデータを集める	<ul style="list-style-type: none"> ● 正確な評価のためにより多くの“相関性がある”とされる情報を集めようとするため、プライバシーを侵害する可能性がある ● 低い評価の履歴がいつまでも残り、評価の固定化につながる ● 意図していない相手に情報が開示されたり分析されたりする
	昇進や昇格や配置転換の案をつくる、決定する	<ul style="list-style-type: none"> ● 過去の実績に基づくバイアスがそのまま踏襲されてしまう ● ある特定の“カテゴリ”に属すると判断され、それによって昇進するか否かが決まってしまう ● アルゴリズムがブラックボックス化しており、どのような根拠で昇進させるか否かを判断したかが説明できない ● 低い評価の履歴がいつまでも残り、昇進や昇格を防げる ● AIの決定に対して、不服を申し立てられない
	健康状態や心の状態を知る	<ul style="list-style-type: none"> ● 正確な評価のためにより多くの“相関性がある”とされる情報を集めようとするため、プライバシーを侵害する可能性がある ● 退職予測など目的以外の予測に使われる
キャリア開発	受けるべき研修などのレコメンドをする	<ul style="list-style-type: none"> ● レコメンドの根拠がブラックボックス化されており、なぜ、それをやるべきかがわからない ● 意図的な誘導につながる可能性がある
組織開発	組織の状態を知る	<ul style="list-style-type: none"> ● 正確な評価のためにより多くの“相関性がある”とされる情報を集めようとするため、プライバシーを侵害する可能性がある ● 全員のデータが集められるわけではなく、データプールがその組織の代表性を担保できない場合、適正な判断ができない ● 相関から誤った因果を読み取って、間違った施策を打つ可能性がある
	組織のダイバーシティ&インクルージョンを推進する	<ul style="list-style-type: none"> ● データの偏りによって少数派への差別を再生産する
AIや ロボットと 協働する	自律的に動くロボットとともに(工場などで)作業をする	<ul style="list-style-type: none"> ● 人の仕事が奪われる ● ロボットの判断によって事故が起きたとき、責任の所在が不明瞭
	ロボットが提案をする	<ul style="list-style-type: none"> ● なぜそれを提案するのか説明されないため、その良し悪しを人が判断できない



課題やリスクに向き合え 国内外で始まる“原則”策定

より有効にAI活用をしていくために、国内外を問わず、国際機関、各国政府、大学や研究機関、企業などでAI原則や倫理指針の策定が進む。私たちが人事のAI原則をつくるに先立ち、政府や各種団体における原則・倫理指針づくりでリーダーシップを取った人々から原則策定にあたって重視すべきことを学んだ。

内閣府 人間中心のAI社会原則

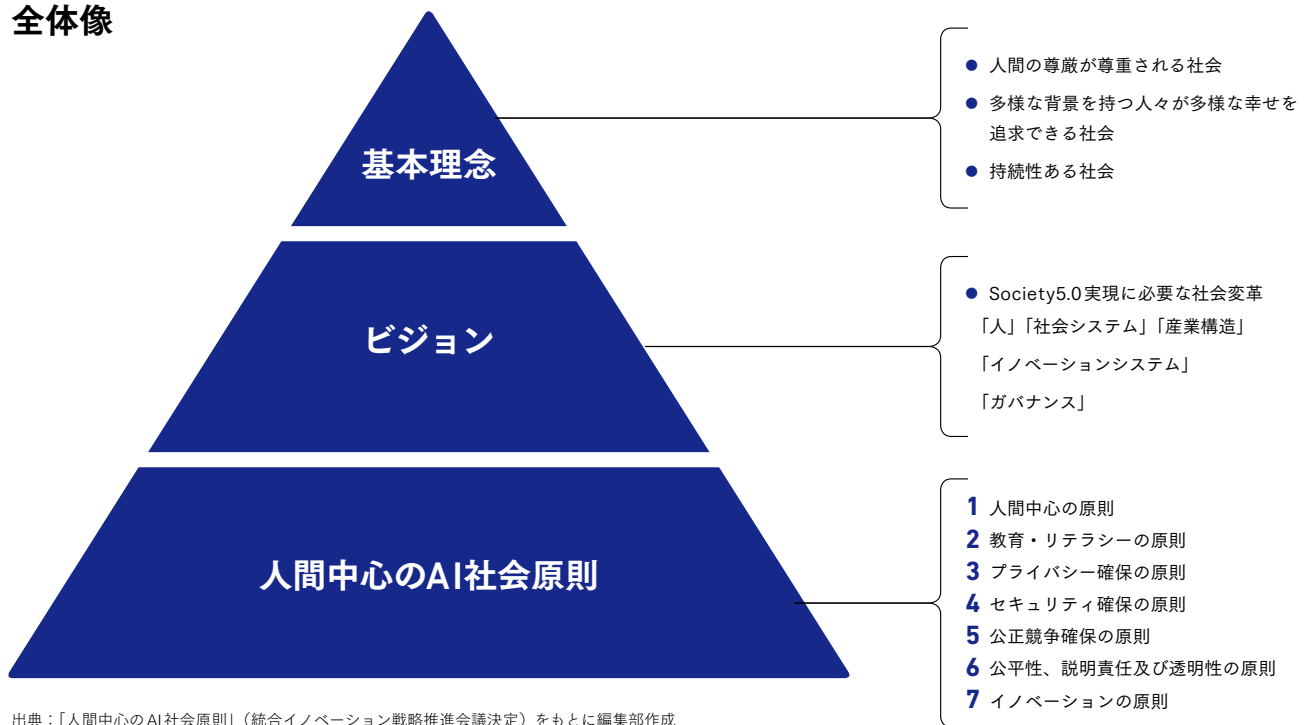
“機械中心”ではなく“人間中心”の社会のための
企業、個人のあるべき姿を提示

内閣府は「人間中心のAI社会原則」（以下、人間中心原則）を2019年3月に発表した。検討会議の発足は、2018年5月。AIの研究者、企業人、法曹界などさまざまな分野の専門家や実務家に参加し、8回にわたって会議が行われ、議論が重ねられた成果である。

「日本が提出した人間中心原則は、OECDや欧州委員会がそれぞれ定めたAI基本原則にも影響を与えたと考えています」と、話すのは東京大学大学院情報学環教授の須藤修氏だ。検討会議では議長を務め、OECDの「人工知能に関するOECD原則」（以

下、OECD原則）の策定にも日本代表としてかかわった。「特に、“人間中心”という概念は、非常に評価されました。OECD原則の冒頭では『AIは、包摂的成長と持続可能な発展、暮らし良さを促進することで、人々と地球環境に利益をもたらすものでなけれ

「人間中心のAI社会原則」の全体像



ばならない』と謳われ、日本の人間中心原則と非常に近いものになりました」（須藤氏）

日本社会や企業、個人などのあるべき姿を提示する

人間中心原則の中身を見てみよう。

人間中心原則は、基本理念、ビジョン、原則（人間中心のAI社会原則とAI開発利用原則）の3つのパートから成る。基本理念は、「人間の尊厳が尊重される社会」「多様な背景を持つ人々が多様な幸せを追求できる社会」「持続性ある社会」の3つだ。特に、最初の2つで“人間中心”ということ強く打ち出している。「人間がAIを使いこなして能力・創造性を拡張し、やりがいのある仕事に従事できる社会にすること。高齢者や若者、地方在住者など、日本で暮らすあらゆる

人々の間に格差が生じないようにする包摂性を持つこと。このように、基本理念では、単にAIをどう使うのか、何を禁止するのかといったルールではなく、日本社会、企業、教育機関、そして個人の未来はどのような姿であるべきかを提示しています」（須藤氏）

なぜ、“人間中心”とあえて言わなければならないのか。須藤氏はその理由を、「AIが社会や人にもたらすインパクトが非常に大きいため」と話す。「私たちは、人類の転換点ともいえるような大きな変化のただなかにあります。これまで私たちが基礎教育で学んできたのは、与えられた問いに対して最適に答えること、目的に対して最適な手段、最短のルートでソリューションを出すことですが、この、いわば最適化能力というのは、AIが非常に長けているものです。教育のありよ



須藤 修氏

東京大学大学院情報学環・学際情報学府 教授
東京大学総合教育研究センター長

うが変わらず、人が今のままの能力形成をするならば、AIに負け、多くの仕事が代替されてしまうでしょう。それでは“機械中心”になってしまいます。そういう未来を目指しているのではないと強く表すために“人間中心”という言葉を使いました」（須藤氏）

そのため、ビジョンや原則でも特に強調されているのが、教育やリテラシーの重要性である。「データやAIの基礎教養から実装及び設計等の応用力を、幅広い分野の横断的、複合的及び融合的な枠組みで身につけた人材が十分に存在すること」を社会の前提とし、すべての人がAIを使いこなせるようになるために、「原則に沿う教育・リテラシーを育む教育環境が全ての人に平等に提供されなければならない」と述べている。

「政府がかなりの費用を投入して、国民全体のリテラシーを上げようとしています。テクノロジーの教育に加え、法規、経営のあり方、倫理的な教育も行い、分離されていた文理の学問を再統合し、初等教育や高等教育機

関で次世代の人材を育みます。同時に、ビジネスパーソンもこの変化の波についていけない可能性があるため、リカレント教育を意識して、大学などでの人材育成を既に始めています」(須藤氏)

ルールは決めなくても 決めるための基準は提示

人間中心原則は、具体的なルールや禁止事項を定めたものではない。「基本理念や原則を提示し、ルールについてはそれぞれのステークホルダーが考えていきましょう、というスタンスを取りました。OECD原則でも、当初の思惑の違いを乗り越えて、『規制を定めるのではなく、基本理念を定める』という形式が取り入れられ、加盟

38カ国とそれ以外の7カ国を加えた45カ国が、この原則に署名しました」(須藤氏)

ただし、日本の人間中心原則でもルール自体は決めていないが、「ルールを決めるに際しての基準・考え方は提示した」(須藤氏)という。今後、異なる文化や慣習の集団ごとに、あるいは、自動運転、医療、軍事など異なる分野ごとに個別のルールがつけられていくだろう。「そのとき、どのような目的でつくられたルールであるかにかかわらず、すべてのルールが踏まえてはいけない問題については、7つの原則のうちプライバシー確保の原則やセキュリティ確保の原則、公平性、説明責任及び透明性の原則などで示しています」(須藤氏)

人工知能学会 倫理指針

AIと社会とのかかわりを規定する 研究者、開発者のための“職業倫理”

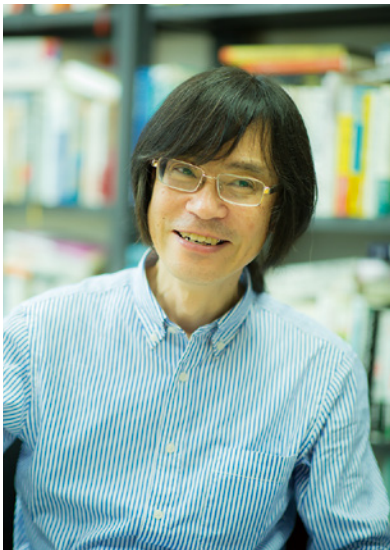
人工知能学会でも、会員である研究者・開発者約4000人に向けて、2017年にAIに関する倫理指針を発表している。倫理指針を議論し、決定したのは同学会の倫理委員会である。

倫理委員会発足時からのメンバーであり、現在は委員長を務める国立

情報学研究所教授の武田英明氏は、「倫理委員会の発足は、2014年に人工知能学会誌の表紙がネット上で“炎上”したことがきっかけでした」と振り返る。「その表紙は、女性のように見えるロボットが家事をしている、というものでした。“家事は女性の仕事”

というステレオタイプな表現に対して異議が噴出したのです」(武田氏)。当時はまだ、AI研究は“冬の時代”といわれていた頃で、「まさか社会からこのような反応があるとは思っていなかった」(武田氏)という。

「それまでAIの研究者は、新しいア



武田英明氏

国立情報学研究所 教授

ルゴリズムをつくりたいという意欲はあっても、社会への影響には興味がありませんでした。これほどにAIに興味を持つ人が世の中にいるのであれば、社会との関係を考えなければならぬと考えたのです」(武田氏)

メンバーは、AIの専門家、研究者を約半数とし、残りの半数は経営者やSF作家などさまざまな人々の協力を仰いだ。

より知的行為に迫るものだからこそ倫理指針が必要

AI以前にも電卓やパソコン、スーパーコンピュータといったシステムが存在したにもかかわらず、なぜ、AIという領域には倫理指針が必要なのか。「それは、これまでのシステムとは異なることが2つあるからです」(武田氏)

1つは、AIがより高度な知的行為、知的活動を扱っていることだという。「もちろん電卓ですら知的行為を代替

人工知能学会 倫理指針 (抜粋)

序文

高度な専門的職業に従事する者として、人工知能の研究、設計、開発、運用、教育に広く携わる人工知能研究者は、人工知能が人間社会にとって有益なものとなるようにするために最大限の努力をし、自らの良心と良識に従って倫理的に行動しなければならない。人工知能研究者は、社会の様々な声に耳を傾け、社会から謙虚に学ばなければならない。

(中略)

人工知能学会は、自らの社会における責任を自覚し、社会と対話するために、人工知能学会会員の倫理的な価値判断の基礎となる倫理指針をここに定める。

4. 公正性

人工知能学会会員は、人工知能の開発と利用において常に公正さを持ち、人工知能が人間社会において不公平や格差をもたらす可能性があることを認識し、開発にあたって差別を行わないよう留意する。人工知能学会会員は人類が公平、平等に人工知能を利用できるように努める。

6. 誠実な振る舞い

人工知能学会会員は、人工知能が社会へ与える影響が大きいことを認識し、社会に対して誠実に信頼されるように振る舞う。人工知能学会会員は専門家として虚偽や不明瞭な主張を行わず、研究開発を行った人工知能の技術的限界や問題点について科学的に真摯に説明を行う。

7. 社会に対する責任

人工知能学会会員は、研究開発を行った人工知能がもたらす結果について検証し、潜在的な危険性については社会に対して警鐘を鳴らさなければならない。人工知能学会会員は意図に反して研究開発が他者に危害を加える用途に利用される可能性があることを認識し、悪用されることを防止する措置を講じるように努める。また、同時に人工知能が悪用されることを発見した者や告発した者が不利益を被るようなことがないように努める。

出典：「人工知能学会 倫理指針」(人工知能学会)より抜粋

しているものですが、登場した当時は驚かれても、それが社会に浸透し人々が使うことに慣れると、それを知的なシステムとしてはとらえなくなります。AIとは、その時代ごとの最先端のエッジにある技術の集合体だと私は考えていますが、ここに来てそのエッジが、従来人間にしかできないと思われていた知的な領域と重なってきました」(武田氏)

もう1つは、生活の現場はもとより、AIが人の感情やメンタリティに近いところまで入り込んでくる技術である

ことだ。「こうした2つの特徴を持つため、人間に与える影響が大きい。だから、AIが人間や社会に悪い影響を及ぼさないようにするための倫理指針が必要なのです」(武田氏)

自らの研究が社会に実装されたときを想像する

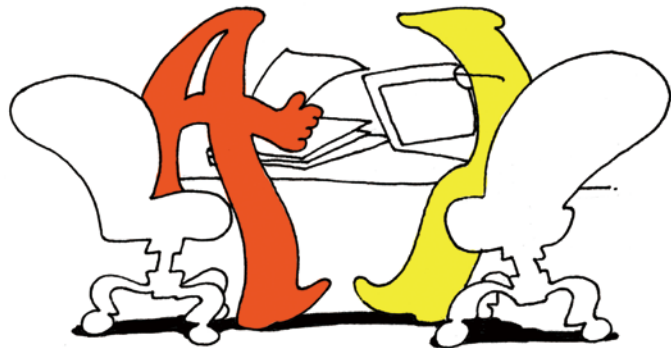
上は、倫理指針の抜粋である。研究者、開発者が持つべき職業倫理として書かれているのが特徴の1つだ。

「私自身も含め、研究者や開発者はやはり、自由に新しいことを探求して

いきたいという欲求が強い。しかし、研究の成果の多くは、誰かの手によってサービス化されます。社会で実装されたらどのような影響があるのか、想像力を働かせることの必要性に、議論を重ねるうちに気づきました」(武田氏)。AIの研究者や開発者たちに対して、論文を書いたら自分の研究は完成ではなく、その先に起こり得ることまでを見据えて研究や開発をせよという期待を込めているのである。ただし、新しいことをやりたい、発見したいという意欲に歯止めをかけることはしなかったという。「禁止事項を設けるのではなく、研究や開発にあたって求められる行動として整

理したのは、そういう思いがあったからです」(武田氏)
現在は、「倫理委員会としてシンポジウムを行うなどの活動をし、『人工知能学会には倫理委員会がある』と

徐々に認知されてきた状態」(武田氏)だという。「政府や企業なども原則や指針を定めています。そうした動きが相まって、AIに携わる人々の意識は変わっていくと思います」(武田氏)



ピープルアナリティクス&HRテクノロジー協会 ガイドライン

企業と個人がともにメリットを享受する データ活用を目指す

人事領域におけるテクノロジーやデータ活用におけるガイドラインの作成も進む。ピープルアナリティクス&HRテクノロジー協会では、2020年3月の一般公表に向けてガイドラインの整備を進めている。

「協会の設立は、2018年5月のことでした。近いうちにガイドラインを作成するという趣旨を聞き、それならば、

と理事を引き受けました」と、同協会の理事であり、早稲田大学教授として人事におけるデータ活用を研究する大湾秀雄氏は話す。

「人事が持つ、人にかかわるデータをどう分析し、どう使うかを学生や企業人に教える活動を通じて、ガイドラインの作成が急務だと思いました。特に、社員の健康データや行動履歴を

活用しようという動きが企業に出てきたことをリスクと感じたのです。これらはかなりプライバシーに踏み込む情報であり、それによって一方的に不利益を被る人が出てくる可能性があるからです」(大湾氏)

人事が人にかかわるデータを活用する目的は、会社をよくすることであり、同時にそれによって従業員も働き

やすくなることだと大湾氏は主張する。「一方的に不利益を被る人が出てくるような使い方をしてはならない、とはっきり線を引きたいと考えました」(大湾氏)

信頼関係の構築なくして データ活用は進まない

データの活用によって、個人情報を提供する側の従業員や応募者にもたらされる利益は必ずある。「従来は人間が判断してきたため、たとえば人事異動においては、配置を判定する役員や上級管理職の目に留まりやすい人のほうが有利だという傾向がありました。縁の下の力持ちタイプの人は能力や実績が見えにくい、というような不公平が存在したのは事実です。それを是正し、これまでよりも公平に評価や判断を行える可能性があることがデータ活用の大きなメリットといえるでしょう」(大湾氏)

こうしたメリットを含めたデータ活用の目的の説明がなければ、何か不利益を被るのではないかと、監視され、弱みをつかまれるのではないかと、というような疑心暗鬼を従業員の間を生むことにつながりかねない。「肝に銘じておくべきことは、データ活用は労使関係が良好な会社のほうが進みやすいという点です」(大湾氏)

同協会が発表する予定のガイドラインでは、特に“プロファイリング”に関するリスクを強調している。プロファイリングとは、データ分析によって得られたモデルから、個人のパフォーマンスや離職の可能性を予測するものだ。「プロファイリングは人がより活躍する

可能性を予測したり、働き方や働く環境を最適化するうえでは非常に有効ですが、統計的な差別を助長したり、プライバシーを侵害するといったさまざまな問題が生じ得ることも知っておかなければなりません」(大湾氏)

予測がどのように使われるのかわからないままデータを取られたり、それによって不条理と感じるような扱いを受けたりした場合、労使の信頼関係は著しく毀損される。「十分な配慮のうえで実行される必要があります」(大湾氏)

AIを理解する人事が いてこそそのデータ活用

ガイドラインでは、プロファイリングをはじめとするデータ活用における具体的な配慮について言及するという。

まず、労使協議のなかできちんとデータを活用することを話し合っておくことだ。「説明なきままのデータ活用によって不利益を被る人が生まれ、その人が会社を訴えるようなことがあれば、個人情報保護法違反に問われるだけでなくレピュテーションも低下するなどのリスクが考えられます」(大湾氏)。労使協議の際に説明すべきポイントは2つ。「何のためにデータを活用するのかという目的、それによって企業・個人双方にどんな効果があるのか。それを説明したうえで、目的に沿った使い方以外はしないという約束をすべきです」(大湾氏)

次に、苦情申し立ての手続きを決めておくことだ。「データを取得される側が不満に思ったり、おかしいと思ったりしたことを申し立てられる体



大湾秀雄氏

早稲田大学政治経済学術院 教授

制をつくる必要があります」(大湾氏)

そして最後が、意思決定には必ず人間が関与することである。「『AIが決めたので出向してもらいます』というのでは絶対に納得できません。AIが出したリコメンドを見て、人が確認、分析した結果として配置や評価を行う、というように人間が責任を持って結論を出し、その結果を説明する必要があります」(大湾氏)

こうした配慮とともにAIによるデータ活用をしようとするとき、人事に求められるのはAIに関する相応のリテラシーだ。「AIによって実現できることは何か。従業員がどんな不利益を被る可能性があるのか。アルゴリズムは適正なのか。AIが導き出す結論にどんなバイアスがかかっている可能性があるのか。これらを理解できる人事担当者がいてこそ、企業・個人双方に利益をもたらすデータ活用が可能になるのだと思います」(大湾氏)

人事のAI原則を議論しよう

ここからは、いよいよ「人事のAI原則」づくりに取り組む。まずは編集部で草案を作成した。

これをもとに人事領域の専門家、実務家、AIと法律、AIと社会の専門家による検討会を実施して、内容を吟味した。彼らの意見を参考に修正を加えた“暫定版”を、34ページで提示する。

人事のAI原則(草案)

基本原理

1

個人尊重の原理

人事が人事の諸活動においてAIを利用するのは、その利用によって、多様な働く人々が、今より無理や無駄のない仕事と働き方で、より高い成果を出すことのできる、よりよい職場を実現するためである。人事は、AIを利用することで、日本国憲法および国際的な規範の保障する基本的人権を侵すことのないようにすること、また、AIの利用が、社会や企業内に新たな不当な差別を生んだり、個人の尊厳を損なう不当な扱いを生じさせたりしないことを追求し続ける。

原則1.1 データ主体の権利尊重の原則

データ所有者はデータ主体である。データ主体は、データの開示・修正・消去を求める権利、データの利用のされ方に異議を申し立てる権利を有する。

原則1.2 プライバシー確保の原則

人事は、パーソナルデータの取り扱いには一層の慎重性をもってあたる。人事は、パーソナルデータの活用が個人の自由、尊厳、平等を侵害しないことを常に確認し、検証する。人事は、パーソナルデータの収集とその保管に対して道義的責任と技術的責任を負う。

原則1.3 公平性確保の原則

人事は、AIを活用するにあたって、人々がその人種、性別、国籍、年齢、政治的信念、宗教などの多様なバックグラウンドを理由に不当な差別をされないよう、その設計思想における公平性を常に検証する。

基本原理

2

公正プロセスの原理

人事におけるAIの利用が、人々に対してその人の持つ多様な背景を理由とする不当な差別をもたらすことになってはならない。そのため、AIの開発・提供・利用にかかわる人々は、公正なプロセスを設計し、それに則って適正な開発・提供・利用を実現する。人事は、AIが必要とするデータを提供する人々に対して、AIを利用している事実とそのプロセスなどについて説明を施し、データ提供とその利用に関する許諾を得る。AIの利用がもたらす結果については、問題の特性に応じて、AIの開発・提供・利用にかかわったステークホルダーが適切に分担して責任を負う。

原則2.1 合目的性の原則

人事は、AIを活用するにあたって、利用目的・範囲・プロセスを明確にするとともに、目的外のデータをみだりに取得したり、利用したりしないよう、データ最小化を目指す。

原則2.2 本人同意の原則

人事は、データを収集するに先立って、データ主体に対して、データ収集の目的、収集データの種類の、データ主体の持つ権利などに関する説明を実施し、それに対する同意をデータ主体から得る。また、これらの説明をし、同意を得たことを、必要に応じて証明できるようにする。

原則2.3 正確性の原則

人事は、AIを利用するにあたって、AIが必要とするデータの最新性と真正性、データに欠落がないことを常に検証し、不正確なデータを用いない。また、AIが用いるアルゴリズムにエラーがないことを検証する。

原則2.4 適正性の原則

人事は、データを取得するにあたっては、適正適法にこれを行う。また、AIを利用するにあたっては、信頼できる開発者・提供者を活用し、信頼性を担保するための複数の仕組みの構築に努める。

原則2.5 セキュリティ確保の原則

人事は、データを取り扱うにあたって、安全性の確保に努める。データの匿名化・仮名化を可能な限り実施し、データ保管にあたって、漏洩などが起きないように、技術的措置を講じる。

原則2.6 危機管理の原則

人事は、AIを利用するにあたって、起こり得る問題を想起し、技術的・非技術的対応策を事前、ならびに事後的に講じる。問題が発生した場合に、どのような行動を取るか、事前に決めておく。



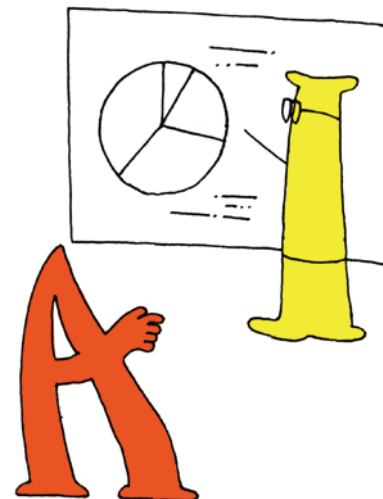
前ページおよび下記が、編集部で作成した人事のAI原則の草案である。

草案をまとめるにあたっては、ここに至るまでに取材した専門家や実務家からの示唆を活かし、加えて既に紹介した各国・各団体が作成した原則やガイドライン^(*)を参考している。

まずは、4つの大きな基本原理を掲げた。1つ目の「個人尊重の原理」は、人事がAIを活用して人事権を行使するにあたって、個人の尊厳を毀損してはならないという原理である。2番目の「公正プロセスの原理」と3番目の「透明性とアカウンタビリティの原

理」には、個人に不当な不利益をもたらさないように手続き上の正しさを追求すべきということと、AIを活用した人事を行うにあたって、受益者やステークホルダーに正しく情報を開示し、説明を行う義務があるということを表した。4つ目は「知性と倫理性の原理」とした。AIを活用していく時代にこそ、人事としての倫理観、リテラシーを持つべきだということを意図している。

この草案を一読していただき、そのうえで次ページからの5人の専門家たちの意見に耳を傾けてほしい。



基本原理

3

透明性とアカウンタビリティの原理

人事がAIを利用するにあたっては、AIの利用によって何を実現しようとしているのか、それをどのようなプロセスとアルゴリズムで行うのかについて、情報を開示する。AIの利用がもたらす結果について、なぜその結果になったのかを説明する。人々がAIの提案を理解したり判断したりするために、AIの利用や運用について、必要に応じて開かれた対話の場を適切に持つ。

原則3.1 透明性の原則

人事は、AIを利用するに先立って、AIを利用するという事実、利用の目的、利用に用いるデータなどについて情報を開示し、透明性を確保する。

原則3.2 アカウンタビリティの原則

人事は、AIによる分析に対して、そのプロセスやアルゴリズムを理解し、それを説明できる体制を構築する。

原則3.3 人間による意思決定の原則

AIによるいかなるサジェスションがあっても、その採否の決定や最終的な意思決定は、人事が責任を持って行う。

基本原理

4

知性と倫理性の原理

AIは、人間の労働の一部を代替するのみならず、高度な道具として人間を補助することによって人間の能力や創造性を拡大することができる。人事ならびに、人事で用いるAIを開発、利用、提供しようとする者は、AIに対する正確な理解を持ち続け、また、社会的に正しい利用ができるよう、法をはじめとする知識と職業倫理を持つ。

原則4.1 リテラシー確保の原則

人事は、AIをコントロールする知識を獲得するための体制を構築し、その利用にあたっては、適切なサービスとその提供者を見極める。また、従業員などのステークホルダーが知識の不足ゆえにAI時代の弱者となることがないように、従業員へのAIリテラシー教育の体系をつくりこれを実行する。

原則4.2 AIからの自立の原則

人事をはじめとするステークホルダーは、AIの進化を見極めつつ、同時にAIの限界についても認知し、AIに依存しない。

原則4.3 倫理尊重の原則

そもそも人事に求められる職業倫理は、AIを利用することによって損なわれてはならない。人事ならびに経営は、AIを社会的に正しく利用するために、適切なコストを負担することを惜しまない。

(*) 内閣府「人間中心のAI社会原則」、人工知能学会「人工知能学会 倫理指針」、OECD「人工知能に関するOECD原則」、EU「一般データ保護規則（GDPR）」

編集部：今日はお集まりいただきありがとうございます。

私たちは、取材を通じ、AIの進化は留まるところを知らず、そして、人事領域でもAIを使いたいニーズはどんどん高まっていること、同時にリスクもあることを学びました。各国や地域、多様なステークホルダーの間で原則の議論が始まっていますが、人のことを取り扱う仕事である人事こそ、AIを使うにあたっての原則を話し合うべきだと考えました。

編集部で作成した「草案」(26～27ページ)をベースに、「人事のAI原則」に何を盛り込むべきか、今日は議論したいと思います。

お集まりいただいたのは、各分野の専門家の方々です。学習院大学の守島基博さんは、人的資源管理論の専門家です。志水静香さんは外資を含めた複数の企業で人事トップを務められました。鵜澤慎一郎さんは、EYのHRコンサルティング部門の

トップであり、多くの企業のHRの課題や最近のAIに対するニーズをご存じです。東京大学の江間有沙さんは、科学技術と社会の関係性を研究され、人工知能学会倫理指針をはじめ、多くの原則づくりの現場を見ておられます。慶應義塾大学の山本龍彦さんには、法律の専門家として出席していただきました。

まず、私たちが今回提示しようとしている原則の位置づけを説明します。

私たちは、これを絶対的な規則として提示するものではありません。各業界、各企業が、AIを実用していくにあたって、自業界、自社なりの原則を考える必要があります。また、今の時点でのAIにできることを想定してつくっているものなので、今後、より賢いAIが登場すれば原則も変わると思います。今回作成するものは、いわば“暫定版”です。変化していくことを前提にしています。

Theme 1

この原則の目的は何か。 誰に向けてものか

江間：各論を議論する前に、私が絶対にあつたほうがいいと思うのは“前文”です。なぜこれを、誰に向かって、どういうスタンスでつくるのか、などを説明しなくては、誰の心にも響かな

いものになってしまいます。

鵜澤：いろんな団体が倫理憲章などをつくるなかで、どう差別化するかという視点も必要です。今の内容は非常にオーソドックスで、特に否定すると

ディスカッションの 参加者



志水静香氏

ファンリーシュ 代表取締役
ストラテジック・ピープルパートナー



鵜澤慎一郎氏

EYアドバイザリー・アンド・コンサルティング
ピープルアドバイザリーサービス リード パートナー

ころもありませんが、一方で、ここが重要、という特徴もありません。

江間：名称も、「人事のAI原則」という無味乾燥なものではなく、つくり手の意図を込めた名称をつけるといいですね。

編集部：前文をつくるということは考えていませんでしたが、確かにあったほうがいいですね。ほかに前文に入れるべきことはありますか。

鶴澤：基本的なこととして、これからの時代、人事が扱う人に関するデータは最高の経営資産であり、同時に最高機密だと考えるべきだと言っておきたいですね。この自覚がないから、データを軽んじてしまうのだと思います。

守島：なぜ最高機密か、理由も説明したいですね。ファイナンスの情報が最高機密なのは、企業の公正な競争との関係が深いからです。人に関するデータが最高機密なのは、人権を侵害してはならないからなのです。

鶴澤：もう1つ、原則を持つことは、グローバルな視点からも必要なのだと言っておきたいですね。海外で事業を展開する企業では、人のデータの取り扱い、グローバルに統一化されたルールに縛られることになります。そのことに敏感でなくてはなりません。

守島：ダイバーシティ&インクルージョンを実現する、という視点も欠かせないですね。

山本：確かにそれは重要です。適切にAIを使えば、人の判断に伴うバイアスのためにこれまで不利益を被っていた人々に、適切な機会が与えられる可能性がある。AIの活用は、よりよい意思決定のためになるんだということも強調しておきたいですね。

編集部：私たちの目的意識を明確にするためにも、前文をきちんと作成したいと思います。

Theme 2

人事にとって「個人を尊重する」とはどのようなことか

志水：基本原理1の「個人尊重の原理」は何より重視したい点です。AIを使うかどうか以前に、人事が働き手である個人を尊重しているかが問われるのではないのでしょうか。“社員は経営者のもの”あるいは“人事権はすべて

会社にある”という旧態依然の感覚のままでは、個人を尊重するのはまったく逆方向の施策が取られてしまう可能性があります。人事の基本的な価値観、職業倫理が問われているということをしっかり盛り込むべきだと思います。



江間有沙氏

東京大学
未来ビジョン研究センター 特任講師



山本龍彦氏

慶應義塾大学法科大学院 教授



守島基博氏

学習院大学
経済学部経営学科 教授 副学長

います。

守島：本来、人事という仕事は人に対する取り扱いを決めるものです。しかし社内における人事では、雇用関係のために、個人の権利は一定の制限を受けるのです。人事上の意思決定は、経営をよくする側面と、個人の尊重を実現する側面の両方を見ながら

行う必要があります。AIの活用以前にこの問題を考える必要があります。現実的には、AIの提案が経営の観点からは最適だとしても、会社全体をよりよくするために、人事などが個人の尊重の観点から問題がないかをウオッチする必要がありますでしょう。

志水：まさに守島先生の“よりよく”と

いう言葉が重要だと思います。前提として、経営と働く人々にとってメリットをもたらすものでなくてはなりません。“経営”そして“社員”にとってよりよい決断をすることで、エンプロイー・エクスペリエンスの向上を目指すことを原則に入れる必要があるのではないのでしょうか。



Theme 3

人事に求められる AIリテラシーとは何か

志水：もう1つ、これからの人事で重要なものとして、AIなどのテクノロジーを使いこなすためのリテラシーを高めることが挙げられます。日本では、HRテックに関するリテラシーが欧米と比較してまだ低いと思います。それは人事だけの責任ではなく、経営の問題でもあります。テクノロジーを導入する際、人事にも研修やOJTを通して学びを深める継続的な学習の提供が不可欠です。原則4.1の「リテラ

シー確保の原則」は、そういった意味でも非常に重要です。

鵜澤：AIリテラシーが低いと、相性診断のようなレベルのサービスを信じてしまうことがあります。「この2人は同じチームにしないほうがいい」などと大真面目に議論し始めたら、企業の人材育成力や競争力が失われることになりかねません。AIやテクノロジーを使わずに済む時代ではなくなったという潮目の変化に敏感に気づいて、

早くリテラシーを高めることに着手すべきです。

江間：そもそもAIにできることは、現状では限られています。また「AIが選んだ、判断した」というように、AIを擬人化しがちなこともミスリーディングを生みます。人が機械を賢く使いこなし、判断にかかる時間が短縮できたり、よりよい判断ができたりするように、人と機械の協力関係の仕組みを考えるのが大切です。AIへの過度な期待や懸念のあまり、人の仕事が余計に増えたり、HRテックの進展が阻害されることにはなってほしくありません。

Theme 4

本人が同意すれば 何をやってもいいのか

鵜澤：私は、原則2.2の「本人同意の原則」という項目がとても気になっています。既に多くの企業では、個人情報を活用するにあたっては本人の同意を得ることがルールというのを理解しています。でも、きちんと同意を得ていたなら、何をやってもいいかという、それは別問題なのです。ここに企業の品格が問われていると思います。同意さえあれば何をしてもいい、という同意至上主義には、大きなリスクを感じています。

山本：欧州のGDPRでは、形式的な同意の問題性を強く意識しています。一方日本は、同意至上主義であると同時に同意形式主義ですよ。

鵜澤：適切なプロセスで実質的な同意を得ていた場合ですら、当局に否定される事例が、欧州では既に起きています。

志水：以前勤めていた会社で、グローバル人事システムを構築する機会がありました。欧州では社員個人の同意を得るにあたり、事前に当局に相談して、この目的で社員のデータを取得する、このように本人から同意を得るプロセスがある、これで問題ないか、という確認を行いました。

山本：GDPRには、正当な利益を追求する場合など、本人の同意を得なくても正当化されるデータの取り扱いが5つ挙げられています。その意味では

GDPRは同意至上主義ではないんです。どういう目的のために使うのか。そこを固めて、説明できるようにしておけば、同意がなくても法的に問題ないということは十分あり得ます。

鵜澤：本人の同意があれば永遠にOKではなく、撤回できる権利や忘れられる権利もある、ということも大事ですね。物理的に本当にデータベース上から消去し、消去されていることが説明できるというトレーサビリティの確保も必要です。個人は、いつでもそのゲームから降りられるという状態にしておかなければなりません。

山本：ここで重要な視点は、個人のほうが立場が弱いということです。インフラの役割を果たしているプラットフォーム事業者から、個人情報の提出やその使用について同意を求められた場合、個人は事実上同意せざるを得ません。最近、公正取引委員会

は、こういう場合に独占禁止法の優越的地位の濫用を適用できないかを検討しています。企業内でも、データの収集や利用に関して同意を求められたら、やはり従業員は同意せざるを得ないでしょう。その意味で個人は弱い。だからこそ、労使間のコミュニケーションが求められます。

守島：これまで長らく協調的な労使関係が続いてきたため、経営や人事がなんらかの方向に変化を起こそうというときに、日本企業の労働組合などはストッパーの役割を果たせないという問題もあります。

山本：持てる情報の量と質が変わってきて、情報の非対称性が増すなかで、情報を持たざる者が持つ者に関与することができるプロセスがより重要になると思います。基本原理2の「公正プロセスの原理」のなかに、ベンダーとユーザー企業の間、企業と従業員の間でのコミュニケーションを確保すること、と入ってはどうか。

たとえば、個人の人事評価などの情報は、マネジャーなど管理者だけが見られる場合が多い。ですが、データを提供して評価をされた側も、そのデータを見て、自分のどこがどのように評価されたのか、どんな努力をすれば評価が高まるのかを理解できることが重要でしょう。それが本人の気づきにもつながる。原則1.1にデータの所有者はデータ主体とありますが、本人の関与を実質化する仕組みがなければ、絵に描いた餅になってしまいます。



Theme 5

“透明性”と“アカウンタビリティ”は実現可能なのか

江間：基本原理3にあるアカウンタビリティの概念がしっかり理解されていないように思います。そもそもの訳され方の問題でもありますが、説明して情報開示さえすればいいのではなく、結果に対して誰が責任を取るのか特定するのがアカウンタビリティの意味するところですよ。

守島：原則のなかには、誰が責任を取るのか決めておく、と書いておくべきでしょう。

江間：たとえば医行為は人間が行わなければならないと法律で定められています。人事については人間がしなければならないという法律はありません。極端な話、人事はAIで決めます、という会社が出てきてもおかしくないのです。だからこそ、AI原則には問題が起きたときの責任の所在を決めると明記する必要があると思います。AIを使って起きたことの結果について、責任を負うのは誰なのかを決めておきます、というルールです。

もう1つ、疑問を感じていることがあります。世に出ている多くのAI原則でも、「アルゴリズムがどのようなものを説明する」といったことが簡単に書かれていますが、実際にアルゴリズムを見せられたところで、一般の人にはほとんど意味がわかりません。AIによる推論の過程を明らかにするとはどのようなことなのか、という議

論をしなければなりません。

山本：そうですね。ディープラーニングの複雑さから見て、説明可能性には限界がある。最近、説明を補完すべく議論されているのは、アセスメントという手法です。自分たちが設計したAIの予測モデルが人権にどのような影響をもたらすのかを査定・検証しようというものです。人権影響評価(HRIA: Human Rights Impact



Theme 6

人がどれだけ関与することを求めるのか

守島：原則3.3「人間による意思決定の原則」については、意思決定に限らず、AIがどのように関与しているかは開示したほうがいいと明記すべきですね。AIの提案をそのまま使っているのか、それとも人が介在しているのか。それを知るだけでも、結果を受け取る側の心証はかなり違います。ど

Assessment) とも呼ばれています。ただ、すべての企業にこれを行うリソースがあるのか、HRIAの妥当性を誰が評価するのが問題になります。**守島：**チェックする能力をどこかに担保したほうがいいということですね。それが社内か社外かは持っているリソースによる、と。

個人に対するアカウンタビリティでいえば、従来、人事の世界では“救済”ということを重視してきました。つまり、何らかの不利益を受けたときに“文句”を言えるというプロセスです。アカウンタビリティのあとに、結果に対して異議を申し立てる仕組みを持つ、というところまで原則には明記したいです。

山本：GDPRでも、重要事項についてAIの完全自動意思決定を組み込む場合には、決定を争う権利を保障しなければならないと規定されています。争うには情報がないといけませんから、やはりそのためにも決定のプロセスについて説明しなければならない、ということになるわけです。

ういう使い方をしているのかをきちんと説明する必要があります。

鵜澤：GDPRでは、AIによる完全自動化された判断は基本的に否定されており、人間が必ず介在すること、と明記されています。

山本：そして、形式的に人間に関与させるだけでなく、権限のある人が関与

することを求めています。ただし、実際には、採用の初期のスクリーニングをAIで行うというとき、人間がすべて関与するのは難しい。むしろ、それにより意思決定の質が落ちる可能性もあります。どういう場面で人間が関与するかを明確にしておくべきです。雇い入れたあとの従業員に不利益になる意思決定をする場合に、人間を関与させるということは、AI原則に入れていいと思います。

鶴澤：これまでAIを使うのは、効率を

上げるため、というのがほとんどでした。採用の初期段階で、“絶対に合わない”人はAIの診断で除いてしましましょう、というように。ところが、今はAIを使って最適な人材配置を実現したい、という話になっています。AIを使った意思決定の重要性がかなり格上げされることになります。だからこそ、本当に信頼に足るのか、AIが間違った分析をしていないか、という議論が必要になっているのです。

江間：人と機械の協力体制における力

の配分は、場合によって変わると思います。必ずこう使ってくださいではなく、会社の方針に従って、こういう方針だからこう使う、と説明することが求められるでしょう。

志水：AIが出してくるリコメンドは一参照データにすぎないという認識も必要です。意見A、B、Cのうちの1つであって、AIが示すものが万能であるという思い込み、AIの出す結果だけ重み付けを高くするという行為は小さくなくてはなりません。

Theme 7

“データ最小化”に いかに真剣に取り組むか

鶴澤：原則2.1に小さく入っている「データ最小化」は、もっとしっかり主張すべきです。“ビッグデータ至上主義”を取る企業は少なくないからです。なんでもいいからデータを集めておけば、思わぬ素敵な示唆が出てくると思っている。こうなると、メールの文章、会議の音声、歩いているモーションのキャプチャなど、なんでもデータとして取っておきたくなる。これは倫理的にも問題ですし、オペレーション効率的にも無駄です。GDPRの影響で、ミニマムなデータからどう示唆を得るかというふうに、世の中が変わっています。

ビッグデータに対する幻想にはもっと注意を喚起すべきでしょう。

山本：そのとおりですね。データを取ってはいけいではなく、目的によっては取っていいのですが、必要に応じて取る範囲を決めておくべきです。目的に照らして考えると匿名でもいいものがもっとあるはず。それもデータの最小化の重要な考え方の1つです。

守島：とはいえ、個人情報をできるだけ細かく欲しい、というのが人事の基



本的な欲求です。必要のない情報は取らないという表現では不十分で、人権に抵触する可能性があることを示さない、という言い方が必要でしょう。

編集部：みなさんの意見を聞いて、AIを活用するにあたって考えるべきなのは、人事とは何か、そこに携わる人々が持つべき矜持とは何かということなのだとわかってきました。AIを使う・使わないにかかわらず大事にすべきこと、そしてAIを活用するからこそ持つべきこと、守るべきことを形にしたいと思います。

本日はありがとうございました。



人事のAI原則《暫定版》、発表

Works作成の草案をもとに前ページまで議論してきたことをここに集約し、人事のためのAI原則の“暫定版”を発表する。各界の専門家たちの知恵によって、構造、内容ともに進化させた。

ディスカッションを受けて変更したのは、主に原則の内容と文書の構造である。

前者については、本人同意やアカウントビリティ、データ最小化など、ディスカッションで検討された部分を中心に大きく書き換えている。

そして後者はまず、「誰に向けたものか」「作成の目的」などを明確にす

る必要があるとの指摘に基づいて、前文を加え、よりメッセージ性の強いものを目指す。

また、ディスカッションでは、文書が構造化されていないという指摘があった。そこで、個人の権利の尊重という大原則をベースに、インプット時とアウトプット時という、プロセスを分ける形での構造化を試みた。そ

して、それらの基盤として、人事としての職業倫理とリテラシーを置いた。それぞれの詳細は、右ページの前文で述べている。

なお、ディスカッションを経て完成したこのAI原則を“暫定版”と呼ぶのは、今後も多くの人々と検討を重ねていき、よりよいものにしたいという思いを込めたためである。

「人事のためのAI原則」の構造

インプット(個人情報)に対する センシティブリティ

- ⑤ データ主体の権利尊重の原則
- ⑥ 本人同意の原則
- ⑦ データ最小化の原則
- ⑧ 適正性の原則
- ⑨ セキュリティ確保の原則

アウトプットに対する アカウントビリティ

- ⑩ 人間による意思決定の原則
- ⑪ アカウントビリティの原則
- ⑫ 異議申し立ての権利確保の原則

個人の権利の尊重

- ① 個人の権利尊重の原則
- ② 公平性確保の原則
- ③ プライバシー確保の原則
- ④ 透明性の原則

人事としての 職業倫理とリテラシー

- ⑬ 倫理尊重の原則
- ⑭ 適正コストの原則
- ⑮ リテラシー確保の原則
- ⑯ 危機管理の原則

出典：編集部作成

人事のためのAI原則

(前文)

私たち企業人事に携わる人々は、経営のための重大な資産である従業員などの人的資産に対して、さまざまな施策を通じて広範かつ直接に影響を与える立場にあります。したがって私たちには、経営と働く人々の間にあって、働く人々がよりよく働き、満足を得られるように努める義務があります。

AIという優れた技術を、人事におけるツールとしても活用することが可能な時代が幕を開けようとしています。ですが、このツールの使い方を間違えば、私たちは、働く人々の尊厳を損ねたり、プライバシーを侵害したりすることによって、結果的に彼らの安心安全とよりよく働く権利を脅かすことになりかねません。また、個人の尊厳が世界的にますます重視されるなか、私たちは人とそのデータにかかわるグローバルな規範を遵守することも求められています。

そこで、私たち企業人事に携わるすべての人々が、AIを活用するにあたって守るべき基本的な原則をここにまとめます。

16個の人事のためのAI原則は、大きく以下の4つの観点でまとめられています。

- 個人の権利の尊重
- インput（個人情報）に対するセンシティブティ
- アウトputに対するアカウントビリティ
- 人事としての職業倫理とリテラシー

人事は、AIを利用することで、日本国憲法および国際的な規範の保障する基本的人権を侵したり、社会や企業内に新たな不当な差別を生んだり、個人の尊厳を損なう不当な扱いを生じさせたりしないことを約束しなくてはなりません（個人の権利の尊重）。そして、そのためにはまず、AIに与える個人に関するデータの取り扱いに対して、細心の注意と敬意をもってあたる必要があります（インput〈個人情報〉に対するセンシティブティ）。また、AIを活用した後段階では、その結果に対して、説明し、対話し、責任を負う必要があります（アウトputに対するアカウントビリティ）。これらは人事として当然守るべき義務であり、AIを活用するに際しても少しも損なわれることがあってはなりません。そのためにも、AIという新しい技術に対して、学び、経験を積み、正しい知識と理解を構築する努力が必要です（人事としての職業倫理とリテラシー）。

私たちは、AIを正しい形で活用することで、人という経営資産に関するよりよい意思決定ができるようになることを望みます。また、そのように人事が進化することで、企業における真のダイバーシティ&インクルージョンを実現することや、働く人々にこれまでになかった新しい仕事や学習の機会およびキャリア展望を提供することが可能になると信じます。

この原則は、私たちが理解し、守るべきものであると同時に、各企業の理念とAIの進歩に応じて発展させていくものです。私たちはこのAI原則をもとに、AIの研究や開発、サービス提供に携わる人々との間で、また経営者や従業員、入社を希望する人々との間での対話を重ねていきます。

個人の権利の尊重



1 個人の権利尊重の原則

そもそも人事は、よりよい経営のために人的資産に働きかけるものであるが、同時に従業員などの個人の権利と尊厳を尊重することが求められている。人事がその意思決定にAIを利用する場合であっても、個人の権利と尊厳が不当に毀損されることがあってはならない。

2 公平性確保の原則

人事は、AIを活用するにあたって、人々がその人種、性別、国籍、年齢、政治的信念、宗教などの多様なバックグラウンドを理由に不当な差別をされないよう、その設計思想における公平性を常に検証する。

3 プライバシー確保の原則

人事は、AIに与える個人情報の取り扱い、また、AIが導き出した個人にかかわる推論の取り扱いには一層の慎重性をもってあたる。人事は、個人情報の収集とその保管、AIシステム自体と結果データのセキュリティに対して道義的責任と技術的責任を負う。

4 透明性の原則

人事は、AIを利用するに先立って、AIを利用するという事実、利用の目的、利用に用いるデータなどに関する情報を従業員などに開示する。また、AIの利用について、従業員代表と話し合うなど、開かれたコミュニケーションの場を持つ。

インプット（個人情報）に対するセンシティブリティ

5 データ主体の権利尊重の原則

データを提供する個人は、データの所有権のみならず、データの開示・修正・消去を求める権利、データの利用のされ方に異議を申し立てる権利を有する。人事は、データの取得や活用についてデータを提供する個人とコミュニケーションする機会をもうける。

6 本人同意の原則

人事は、個人情報などのデータを収集するに先立って、データを提供する個人に対して、データ収集の目的、収集データの種類、個人の持つ権利などに関する説明を実施し、それに対する同意を得る。人事は、この同意が、データを提供する個人の理解と納得を伴う“実質的な同意”になるよう最善の努力をする責任を負う。個人は、同意を撤回する権利を有する。

7 データ最小化の原則

人事は、AIを活用するにあたって常に合目的性を重視し、利用目的・範囲・プロセスを明確にする。そのうえで、目的外のデータをみだりに収集したり、利用したりしないよう、取得するデータの最小化を目指す。

8 適正性の原則

人事は、データを取得するにあたっては、適正適法にこれを行う。また、人事は、AIを利用するにあたって、AIが必要とするデータの最新性と真正性、データに欠落がないことを常に検証し、不正確なデータを用いない。

9 セキュリティ確保の原則

人事は、個人情報を取り扱うにあたって、安全性の確保に努める。データの匿名化を可能な限り実施し、データ保管にあたって、漏洩などが起きないように、技術的措置を講じる。また、問題が発生した場合にどのような行動を取るか、事前に決めておく。

アウトプットに対するアカウントビリティ

10 人間による意思決定の原則

人事がAIを利用するにあたって、最終的な意思決定は人間が行い、AIによる完全自動化した意思決定は行わない。人事は、AIがどのように意思決定に関与しているのかを明らかにする。また、AIによる示唆は参照データの1つとし、過度な重み付けをしない。

11 アカウントビリティの原則

人事は、AIを活用した意思決定のもたらす結果についての責任の所在をあらかじめ決めておく。また、要望に応じて、その意思決定プロセスを説明する責任を持ち、それに応えられる人員を人事部に育成する。また、利用しているAIのアルゴリズムの適切性を検証する機関を社内あるいは社外に持ち、継続的にこれを検証する。

12 異議申し立ての権利確保の原則

従業員などは、AIが分析・予測した結果に対して、異議を申し立てる権利を持ち、人事はこれを受ける仕組みを構築する。また、人事は、従業員が知識の不足ゆえにAI時代の弱者となることがないように、従業員へのAIリテラシー教育の体系をつくりこれを実行する。

人事としての職業倫理とリテラシー

13 倫理尊重の原則

人事に求められる職業倫理は、AIを利用することによって損なわれてはならない。人事は、AIを利用するにあたって、データ提供者からの同意があり、適切なプロセスを踏襲するとしてもなお、倫理的・道義的な観点からその利用が正しいものであるかどうかを常に点検し続ける。

14 適正コストの原則

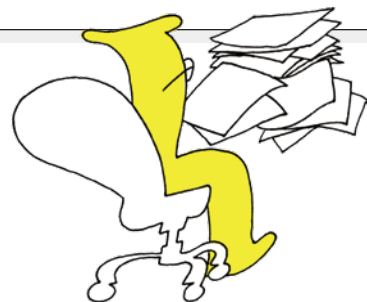
人事ならびに経営は、AIを社会的に正しく利用するために、適正なコストを負担することを惜しまない。AIサービスの利用にあたっては、適切なサービスを見極め、信頼できる開発者・提供者とのコミュニケーションを続ける。

15 リテラシー確保の原則

人事は、AIを過信せず、コントロールするための知識を獲得し、AIの進化についていくための学習を継続的に行う。経営は、人事を含む従業員全員がAIリテラシーを高めることへの責任を持ち、AIリテラシー教育の体系をつくり、継続的な学習機会を提供する。

16 危機管理の原則

人事は、AIを利用するにあたって、起こり得る問題を想起し、技術的・非技術的対応策を事前、ならびに事後的に講じる能力を高め、継続的に点検を実施する。



機械とともにあるからこそ 人間同士の対話を重視すべし

石原直子（本誌編集長）

未来を拓くテクノロジーとして注目されるAIだが、社会への実装が進展するにつれ、AIの持つ負の側面にも人々の関心が集まっている。本特集でも、AIの活用が引き起こし得る問題、課題としてさまざまなものを取り上げた。

だが、現段階で本当の自律型のAIが生まれていない以上、AIが何か“悪さをする”とすれば、それは人間の意思や人間社会の歪みの反映なのだということも、多くの専門家が異口同音に強調した。私たちは、そのことをしっかりと認識しておく必要がある。

人事領域でAIを活用するということは、人事という機能が内包している課題や問題もまた、AIによって増幅される可能性があるということだ。だからこそ、人事がどのような世界観を持っているかが、今後ますます重要になる。

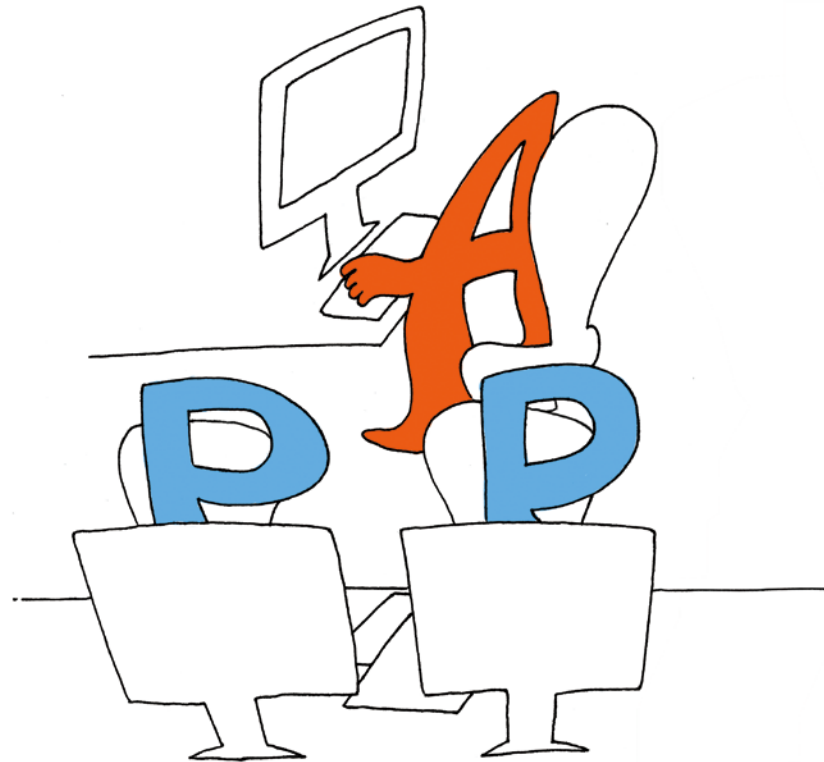
変わらない 人事の基本的な役割

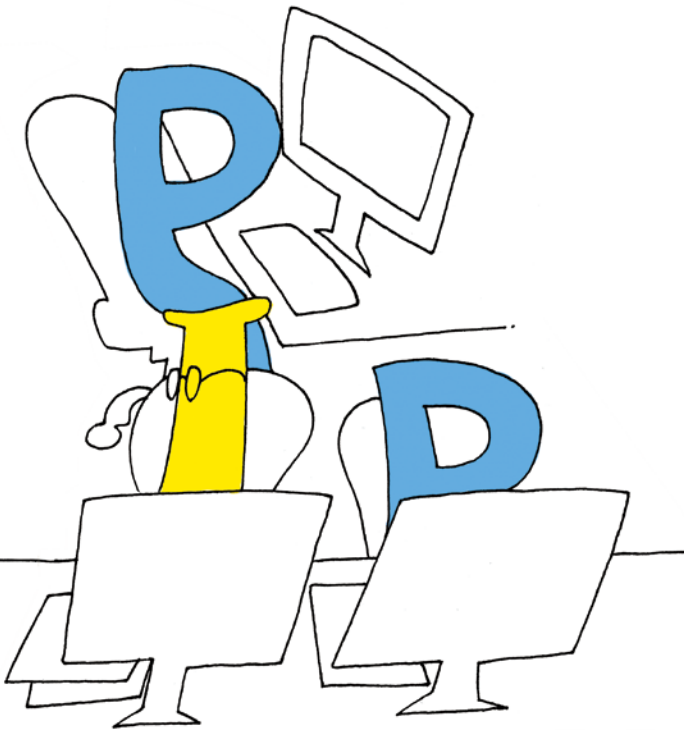
「人事のためのAI原則」全文を通じて繰り返し表したとおり、人事とは

経営における人的資産の効用の最大化と働く個人の幸せの最大化、両方を“同時に”追求する仕事である。どのような人に集まってほしいのか、どのように働いてほしいのかという、経営サイドの視点と、働く人たちにどのように報いるのか、何があれば働く人が幸せになれるのかという、従業員サイドの視点。両方が常に必要だ。このような、いわば各社なりの“人事ビジョン”を今一度確認し、肚に落とすプロセスなしには、AIに何を選別してほしいのかという基本設計図をつくることすらできないは

ずだ。

片側の立場だけでものごとを決めないために大事なのは、想像力とまっとうな倫理観だと指摘してくれたのはSEQSENSEの中村氏だ。想像力を培う具体的な方法が、多様なステークホルダーとの対話ということになるだろう。開発者だけの言い分、経営サイドだけの言い分しか聞かぬまま、新しい技術や施策に飛びつくのはリスクである。社会あるいは従業員はどう感じるのか、何を望んでいるのかを知るのには、人事の基本的な役割であり、これはAI時代に





なろうと変わることはない、ファンリーシュの志水氏も強調していた。

また、早稲田大学の大湾氏は、労使関係のよい企業のほうがデータ活用がうまく進んでいると教えてくれた。話し合いや、情報開示を通じて信頼関係が構築されていることは、新しいテクノロジーを用いるにあたっての最低限の前提条件になるだろう。

AIを用いて “小さき声”をすくい上げる

これまでの人事ではすくい上げる
ことのできなかつた、“小さき声”を

拾える可能性がある。これが人事領域でAIを用いることの大きなメリットの1つだと考えている。

そもそも企業人事という領域では、すべての働く人の権利が完全に尊重されているとはいえない面があると指摘したのは、学習院大学の守島氏だ。意に染まぬ異動、不得意な仕事、本意でないキャリア。そういうものでも、なんとか納得してもらおうのが人事の仕事でもあった。AIを活用することで、こうした個人と企業のニーズのアンマッチが少しずつでも解消され、より多くの従業員が自分

の望む仕事、望むキャリアに出会える。そのおかげで経営がよりうまくいく。そんな未来を期待したい。

“ありがたい姿”とともに それぞれのAI原則を

わざわざ“人間中心”を謳わなければ“機械中心”になってしまうと指摘したのは、「人間中心のAI社会原則」の検討会議において議長を務めた東京大学の須藤氏である。人間のためにAIという機械、ツールが存在する。そのAIに何もかも丸投げしてしまうのではなく、AIに助けをもらいながらも、私たち自身が責任を持って今よりいい未来をつくるのだというスタンスと、それを実行できるだけのリテラシーが大切だ。

今回、編集部が多くの方の力を借りてつくった「人事のためのAI原則」は、完成版ではない。各社会、企業、コミュニティにおける“ありがたい姿”の議論の末に、それぞれが、自らの原則を構築する必要がある。Works版「人事のためのAI原則」が、そのときの参考文献になるというのが、私たちの望むところである。■