

特集 AIと経営倫理

新たなパラダイムに 求められる「倫理ライン」



NextA Partner's 代表

永井郁敏

AI社会へのパラダイムシフト

現代社会は、私たちが知らない気がつかないうちに大きく変わってきています。その変化の中心が情報システム、つまりコンピュータの進化ではないでしょうか。自動化された社会とも言うことができる現代は、社会のインフラから社会経済、企業経営や金融まで、およそ殆どのものにコンピュータが関与し、コントロールしています。そしてそのコンピュータに、AIによる判断能力が付加され、従来の自動化社会はAIがコントロールする社会へと、大きく変わろうとしています。

そんな中、このAIの開発や導入に関して、倫理をテーマとした議論が世界的に広まっており、AIを導入する社会の多くの分野に及んでいます。それは汎用人工頭脳に関するものから、移動体等のモビリティ、金融フィンテック、そして人事やマーケティングといった企業のバックヤード領域など、帰属する領域ごとに取りあげられています。本稿では、AIの導入はもはや多くの組織で避けることができないことを前提に、AIを受け入れ、私たちにとって健全なAI社会を迎えるためにしなければならないことを考えます。

AIのどこに経営倫理が求められるのか？

現在私達は何をAIと呼んでいるのでしょうか。例えば、従来の自動化したプロセスに、対象の色や

形をシステム側で自動判別する機能を付加した

「高度オートメーション」を指すもの。より自然な翻訳機能の実現や、囲碁や将棋で名人を打ち負かしたり、自動運転技術として自律走行車に搭載されたりしている、「単一機能AI」。そして、SF小説に登場するような人間を凌駕し影で人類を支配する「汎用型人工頭脳」まで、AIと呼ぶその領域は人それぞれで、幅広いものがあります。そんなこともあって、AIの統一定義は未だなされていない状態です。とは言え、先進国では10年ほど前からビジネスへのAIの導入が進み、情報やサービスのグローバル化と共に、私たちの暮らしに既に大きく関わっています。特に米国のGAFGAがこれをリードしており、amazonのネット通販における購買履歴や個人情報を利用したサービス、GoogleやFacebookによる個々人の思考をグルーピングすることでの情報サービスが、利用者の満足度を高め、GAFGAの成功を裏づける要因になっています。

しかし、AIを導入することによる成功の裏で、大きな問題が指摘されています。個人情報・プライバシーの問題、スコアリングによる差別、利用者が気づかない内の情報や意見への偏りが生み出す問題などです。我が国でもAIをビジネスに用いることによる、差別や区別、そして判断が、新たな問題としてフォーカスされはじめています。

昨年起きた就職情報を提供するリクナビの事件は、記憶に新しいところではないでしょうか。

これはリクナビの持つ学生の個人情報を外部の企業に無断で販売したことが、個人情報保護の観点から問題視されたものでした。しかし、この問題にはもう1つ大きな問題がありました。それは学生の個人情報からAIシステムが判定した、個人の傾向値（今回は内定の辞退率）が扱われていたということです。

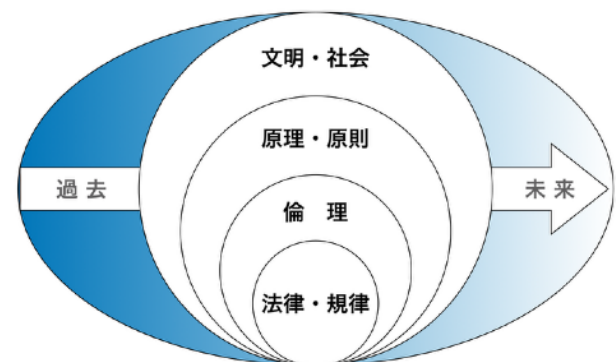
ネット上の販売サイトで、利用者の過去の購買履歴をもとに関係する商品をシステムが自動的に勧める仕組み（クロスセルやリコメンドサービス）では、その選択権、つまり購入するかしないかの意思決定は利用者側にあります。でも、このような利用者自身が選択権（決定権）を有しているサービスとは異なり、今回のリクナビに見る内定辞退率の様な情報は、本人の意思とはかかわりなく個人のその後の人生を決定することになりかねません。この様な、融資に関する信用情報や個人の医療情報などを用いてAIが判断した結果には、求められれば、その判断するに至ったプロセス（アルゴリズム）を、企業は説明する責任があります。顔認証システムの誤った識別により当局に拘束されたり、低所得者の多い地域に住所があるために与信枠が低く設定されたり、AIの技術が未成熟であるがための問題が海外では多数生じており、サービスの停止に至るケースも起こっています。

予想できないAIの進化と倫理観

人は母親から生まれた後、様々な体験を経て成長します。誰ひとり同じ人間はいません。今日までの成長過程での学習や経験などが、人格や思考といった一人ひとりの特性を形作っているからです。家族や地域、学校、組織といった環境で長年培われ、日々その特性は強化されています。AIの学習も基本的に同じプロセスで行われています。ただし異なるのは、大量のデータを投入すること、つまり人でいうところの知識や経験を、ほんの数分、数時間で大量に得る点です。仮にデータを投入することなく、ネットワーク上のあらゆるデータを取得できたら、理論的には人間をはるかに凌ぐ記憶

と、それらの情報に基づく判断をすることができるとはなりません。この考え方が、人工知能がヒトの知能を超えるとされている「2045年問題」です。

一方、我々人間側における倫理はどうでしょう。昨年「AIってなんだ?!」という番組がありましたが、私たちはAIを考えると同時に、「倫理ってなんだ!?!」と考える必要があります。私は、倫理の体系を図1の様に考えています。まず文明や社会を司るものとして原理・原則があります。そこには人として組織人として求められるものから、地域や家族に至るまで、様々な原則が存在しています。そしてその中に、例外なくすべてのものに共通する普遍的な倫理（原則倫理）と、限られた時代や状況にしかあてはまらない、いわば例外もある特殊な倫理（状況倫理）があります。そして最後に、具体的なガイドラインとしての法律・規律が存在していると考えています。原理・原則（principle）には本来例外はありません。人が生存するための多様な共通認識と、地域や組織独自の慣習の積み重ねによって作られた、いわば「拠り所」ではないでしょうか。これらの原理原則を踏まえた共通の価値基準を、私は倫理と考えることができるのではと考えています。つまり社会規範から原理・原則、そして求められる倫理が、必要な倫理ラインとしてのルール（法律や規律）をつくる、この様に考えることはできないないでしょうか。



【図1】 原理・原則と倫理の体系

倫理を吹き込むのは人間

いまの所、AIが人間の様な思考を持つためには、意識がなければ実現しないと言われています。つまり、あたかもAIが判断したかの様に振る舞う

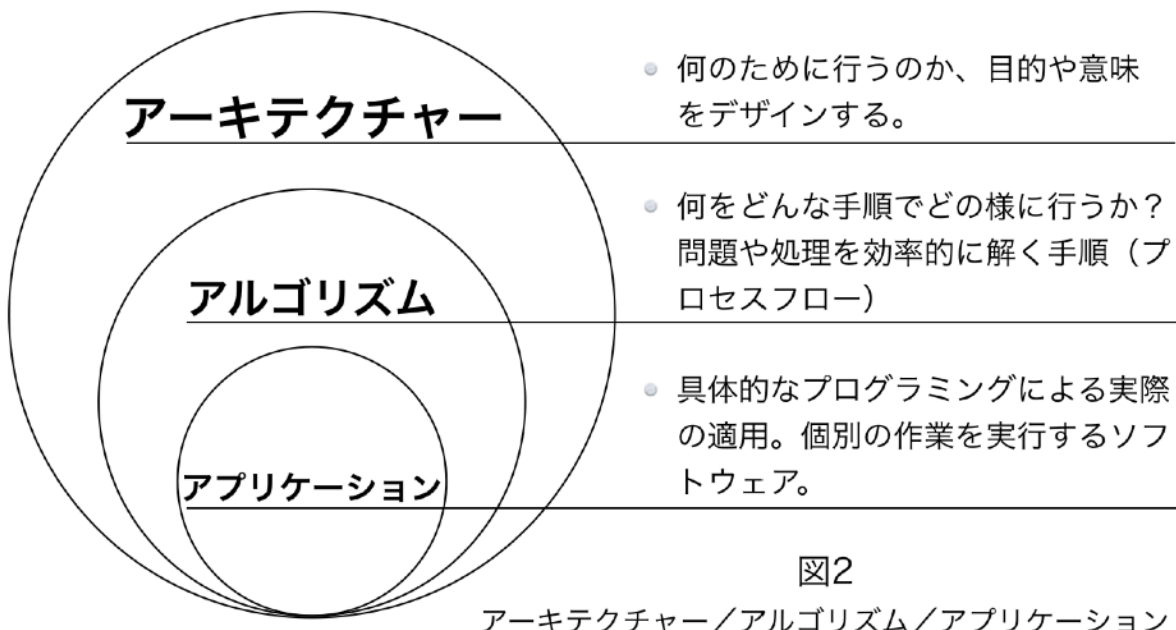
ロボットに、意識はありません。全ては開発者によって考えられるアルゴリズムとプログラムによって動いています。アルファ碁などは碁に特化したスーパーシミュレーションマシンとして、ヒトの判断を超える例だと言えますが、意識はありません。AIが意識を持って結論づけている様に見えますが、これは、もしAという状態になったら、Xという動きをするという判断を実行するプログラムでできています。このAとXは、極端に考えれば、その開発者であるプログラマーの持つ経験や認識、そして価値観で開発されることになります。Aという状態をどの様に認識するかは、それぞれ様々です。そこに思いや価値観の「癖（バイアス）」として、異なる結果を表すことが考えられます。AIの倫理問題が世界的に議論されている中、結局は人間の倫理（ある事象に対する判断や行動として）をAIはデータとして取り込み、紐付けや分類を行った後、先ほどのAならばXというロジックを用いて、結果を導いています。つまり、現在のAI開発には開発者の無意識な意図が反映されてしまう「可能性がある」のです。開発者個人では無いにしても、開発をしたプロジェクトやチームのメンバーの「常識や価値観」などが介入することになるのです。鉄腕アトムを作ったのはお茶の水博士だったことを思い出してください。博士の正義感や倫理観・行動原理が、アトムの判断や行動に随所

に反映されていました。博士は悪人ではありませんでしたが、もし悪意を持った人物が開発に関わったらどうなるでしょう？

原則の共有と倫理ライン

AIを開発するためには、今まで私たちが曖昧にしてきたことやグレーな部分に、倫理的な一線を施すことが、間違いなく求められてきます。なぜならAIはデジタルで判断をするマシンだからです。人間社会での曖昧さや空気を読むといった高度な振る舞いは、近い将来可能になるかもしれませんが、まだ少し先になるでしょう。AIに正しい判断や結果を求めようとするのであれば、現在の曖昧さやグレーな部分、ジレンマを生む状況を、全員で共有できる基準として明確にしていく必要があると考えています。

つまり、現在の私たちの拠り所である原理・原則や倫理観を、AIの判断基準として正しく機能させることが、今後の人間社会にも求められると思うからです。行動原理や判断基準、優しさ、正義など、日本人のそれは、滝沢馬琴の「南総里見八犬伝」にある8つの力「仁・義・礼・智・忠・信・孝・悌」（仁義八行の玉）なのかも知れません。何をAIの判断内容として織り込むか、私達自身が問われているのでは無いでしょうか。現在の未完



成のAIは「私達人間を映す鏡」でもあります。まさに、子（AI）は親を見て育つ、です。

求められるCEOの問いかけ

道徳的な価値や常識が地域の文化によって異なり、プログラムのロジックへの無意識の思考の偏りが存在するだけに、AIシステムに人間の倫理的判断を委ねるには、未だに多くの懸念があります。AIの利用に関して、倫理やリスクとは一見無関係に思われる利用においても、視点をこことは全く異なる方向から見ること、重大なリスクを発見することもあります。これらAIがもたらすであろう問題に、現時点で対応できる最善策は「AIを使わない」ことですが、それこそ現実的ではありません。そこで、AIをどの様に取り扱えば良いか、組織のリーダーが、責任あるAIの構築と利用のために求められる5項目を、考えてみました。

まずは、①言うまでもなく、AIの出力（結果）が公正であること、②取り扱う情報に差別的なバイアス（偏り）がないこと、③データの取得と使用が消費者のプライバシーを侵害しないこと、④システムの性能と、透明性（プロセスを説明できること）の両立が図れること。そして何より、⑤CEOが①～④の状況を、ビジネスリーダーとAI開発のリーダー達に問いかけることです。リーダーとして自身の仕事の価値をどのように理解していて、より良い判断を下すために、自身の価値をどのように使うのかを説明してもらうことです。この部分を「自分では分からないから」という合理的判断を理由に、リーダー達に任せることは責任放棄につながります。情報の扱い方、プライバシー、状況判断の仕方1つで、築いてきた信用を無くすことにならないよう、これらのチェックが求められます。現時点で、AIは諸刃の剣なのですから。

相互信頼できる組織

改めて、現在は意思決定判断をするための情報があまりにも多く、そして急速に変化しています。従って自分の過去の経験だけを頼りに行う判断は、それ自体がリスクになってきます。同じように

見えても、取り巻く状況は決して同じではないからです。組織の善悪基準をどこに置くのか、行うべきことと禁止項目、そして個人および組織の普遍的原則をどの様に作るのか、そしてその原則を全員で共有するために、どうすべきか考えることは、AIの導入にかかわらず求められることです。職場のメンバー間や経営会議の席上で、お互いが、話し合い聴き合い考え合う、いわば当たり前のコミュニケーション環境が維持されているでしょうか。

職場や組織の上司と部下、社員と社員、取引先と社員、そしてAIと社員との齟齬のない本当に求められるコミュニケーションがなければ、このパラダイムシフトを乗り越えることは難しいのではと考えています。言うべきことを正しく伝える技術、そして信頼のおける発言と行動、相互信頼のない職場や組織に仕事を任せること、信任することはできません。

間違っていることを間違いと言い、違いを受け入れて相乗効果を生む、本当に必要なコミュニケーションがAI時代を迎えるにあたり、いま、求められているのではないのでしょうか。

あなたはAIが出した判断を信頼し、行動に移せませうか…？