

人工知能倫理に関わる社会的次元／個人的次元の峻別と交差

Distinction and Intersection of Mental Dimension / Social Dimension related to the Ethics of Artificial Intelligence

河島茂生^{*1 *2}
Shigeo Kawashima

^{*1} 青山学院女子短期大学現代教養学科 ^{*2} 東京経済大学情報コミュニケーション研究所
Department of Liberal Arts, Aoyama Gakuin Women's Junior College Information and Communication Research Center, Tokyo Keizai University

This report discusses the fundamentals of information ethics and examines ethics related to artificial intelligence with respect to the theory of neo-cybernetics. We have confirmed the differences between mental and social dimensions by manipulating viewpoint movements. Artificial intelligence has not yet entered the mental system as it is still operating as an autopoietic system. If this fact is ignored, then human beings become equivalent to machines. On the other hand, artificial intelligence has entered the social system. It is very important to maintain the ethics as society constrains individuals. Attention must also be paid to the fact that generation and promotion of discrimination will occur on the inclusion of artificial intelligence into the autonomy of the social system.

1. はじめに

国際的に注目を集めている情報倫理の基礎的議論としては Floridi の情報哲学が挙げられる。Floridi の情報哲学は、極限まで倫理の範囲を拡大した[Floridi, 10]。Floridi によれば、高等哺乳類や生物、自然環境にとどまらず、情動的に理解できる存在物すべてが道徳的価値をもっており、それらの全体を圏と捉えることで非生物も公正に扱われるという。しかしながら、生物も人間もコンピュータも「情報」という概念で記述できるからといって、これらを一緒に語ってよいのだろうか。人間を含む生物と機械との間には埋めがたい差がある。最先端技術の結集である人工知能でも、それは変わらない[河島 16]。機械は、後でいうアロポイエティック・システムであり、人間によって設計・製造・維持され、人間の指示どおり動くように要請されている。技術的人工物である機械にまで保護される権利を与えてしまえば、損壊したとしても捨てられない。あるいは人間が勝手に電源を切ったり改変したりすることが躊躇われる。人間が機械と対等に扱われる事態さえ招きかねない。情報倫理は、やみくもに範囲を拡大するよりも、むしろ生物のありかたに第一義的に目配りし、生物との異質性／関係性のなかで情報技術の果たす役割・機能を論じる姿勢を保つべきではないだろうか。

アロポイエティック・システムである情報技術の作動の責任はあくまで人間が負うべきである[Johnson 15, 河島 16]。そうであるならば、人工知能による意思決定支援が取りざたされている現在、その内的メカニズムの倫理性については十分に検討することが求められる。人工知能の開発・運用をめぐっては透明性や制御可能性、説明責任の確保が必要とされる[The White House 16, House of Commons Science and Technology Committee 16, 総務省 16]。これは、人工知能にまつわる倫理性に関連している。人工知能によって我々の社会はより公正になり、個人としての尊厳がより尊重されるようになるのだろうか。人工知能による分析によって抜け落ちる面はないのだろうか。

こうした問題意識のもと、本報告は、ネオ・サイバネティクスの理論に依拠しながら、生の根源的なありように目を向けた情報倫理の基盤づくりを図り、そのうえで人工知能との倫理的関わりについて検討する。人工知能倫理に関する研究は数多いが、このような理論的視座から立論された研究は見当たらず、本報告を行うこととした。

2. 概念装置

ここでは、ネオ・サイバネティクスの理論のなかでもオートポイエーシス理論および基礎情報学の視点移動の操作について簡単に整理する。

オートポイエーシス論は、単位体として生命を存立させる内部のメカニズムを定義づけるために立論された。そこでは、オートポイエティック・システム(autopoietic system)とアロポイエティック・システム(allopoietic system)が対比的に区分されている。オートポイエティック・システムとは、みずからで自己(auto)を制作(poiesis)するシステムを指す。内部でその作動を作り出し決定する自律システムである。具体的には、細胞や神経系、免疫系が例に挙げられ、これらは生命システムと総称される。

さまざまな研究者によってオートポイエティック・システムの概念は、生命の領域だけでなく、心理や社会の領域にも適用可能であることが指摘された。心理レベルのオートポイエティック・システム(以下、心的システム)は、構成素である思考が次なる思考に絶え間なく接続しながら存立する。個々の思考は志向性があり、常に「・・・に関する」思考がシステム内部で産出されていく。社会レベルでのオートポイエティック・システム(以後、社会システム)も確認される。社会システムは、人間それ自体ではなく、コミュニケーションを構成素として生起する。したがって、たとえ特定の人がいなくなったとしても、一定のコミュニケーションが接続すればよい。構成素の連続的な生起は、不確実性を抱えながらもしばしば安定した状態を形成する。

こうしたオートポイエティック・システムと対比されるのがアロポイエティック・システムである。アロポイエティック・システムは、自己産出するのではなく、他者(allo)によって生産(poiesis)される。そして、その動きは人間が設定した目的に規定される。人工知能の第3次ブームの中心をなす技術的人工物もアロポイエティック・システムである。なぜならば、人工知能の要求仕様を固め、どのような機械学習の手法を採用するか、いかなるハードウェアを用意するか、どういったデータを読み込ませるかといったことは、人間が多分に関与し、一定の入力に対して一定の出力がなされるように調整されているからである。

ここで留意すべきは、オートポイエティック・システムもアロポイエティック・システムも、そのようなシステムとして同定する存在が欠かさないことである。そうした存在を観察者という。観察者は、

言語を運用する能力を備えた心的システムであり、すなわち人間の心理が該当する(e.g., [西垣 03]). 観察者は、オートポイエティック・システムに内属する位置に立った場合、その内的作動に注意を払い、いかにして内部で現実が構成されているかに焦点を合わせる。

加えて心的システムは、視点移動の操作も可能である。オートポイエティック・システムとして社会システムを見て人間を交換可能とみなす立場に立つこともできれば、オートポイエティック・システムとして心的システムを捉える観点にも立つことができる。基礎情報学は、次の条件で考えたときに社会システム/心的システムの階層関係が成立するとした[西垣 03].

- (1)システム A に内属する観察者の視角から見て、システム A がオートポイエティック・システムとして出現するのに対し、システム B はアロポイエティック・システムとして立ち現れる。
- (2)システム B に内属する観察者の視角から見て、システム B がオートポイエティック・システムとして出現するのに対し、システム A は非明示的存在にとどまりアロポイエティック・システムとして出現しない。

これら(1)(2)の条件がともに成り立つとき、システム A とシステム B との間に階層関係が見いだせる。システム A が上位のシステムでありシステム B が下位のシステムである。社会システムと心的システムの間でいえば、社会システムが上位にあたり心的システムが下位にあたる。上位システムと下位システムとの間には拘束関係が働く。

念のために述べておけば、オートポイエシスであること自体が倫理的であるわけではない。すべてのオートポイエティック・システムを等価に扱うことは実践的な問題としては難しい。というのも、もし等価であれば、人間は動物のみならず植物でさえも食えることができなくなるからである。また、人間の生活を脅かす害虫や害獣、害鳥も駆除できなくなるからである。オートポイエティック・システムそれ自体が善であるとはいえない。それゆえ、次節では倫理学の思索を参考に論述を進める。

3. 視点移動の操作による倫理的領域の整理

倫理は後述する社会的次元で議論されることが多かった。けれども、社会システム内の倫理だけでは重要な側面が抜け落ちる。すなわち、心的システムに寄り添った倫理が欠落してしまう。心的システムに耳をそばだてる倫理は、一般的にはケアの倫理に似ているといえるかもしれない。ケアの倫理の学術的・社会的意義は、心的システムの内面を照らし出すことにある。

当事者研究などが示しているように、当事者と観察者との間に落差は生じる。当事者そのものの人生は、その本人だけが生きるものであり、他者がなりかわることはできない。当事者の心的システムの思考は、一人称のクオリアを覚え、私秘性を帯びる。観察者は、その内面に漸近することしかできない。しかしたとえそうであっても、呻きに似た痛みを耳をそばだてることは可能である。この能力が相手の身になって考える道徳的配慮の起点であるといつよい。

相手の心的システムへの道徳的配慮を伴う個別具体的な経験を通じて社会レベルの道徳的共同体のイメージが形成される。社会システムを観察する観点に立ち、道徳的共同体をイメージする段階である。互いにコミュニケーションを重ねながらともに社会システムを形成している存在は、相互に依存し合っているがゆえに共生の対象となる。

社会的次元である道徳的共同体の倫理は、正義の倫理の名のもとに、富の分配や機会平等などについて検討を加えてきた。

これらの正義の倫理は、ネオ・サイバネティクスの用語でいえば、社会システムを見ている観察者の観点から公正さについて検討しているといつよい。正義の倫理は、基本的には全体として偏りなく、非人称的・三人称的に公平性を担保する議論である。個人の内面を寄り添って眺める視点から立論されたものではなく、個人の心的システムは一義的な価値をもっていない。社会システムでは、コミュニケーションの結節点として個人が現出する。コミュニケーションの連続的継起こそが社会システムにとって一義的であり、個人は入出力関係で捉えられ、アロポイエティック・システムとして現れる。

看過してはならないのは、社会システムの次元で見ると、すなわち社会システムを観察している観点からすれば、個人は社会システムの拘束・制約を受けているということである。個人の振る舞いは社会から制限されている。

誤解を避けるために付け加えておけば、社会的次元の観察者から見ると、社会システムからの拘束を心的システムが受けるからといって、心的システムへの倫理的配慮が無に帰すわけではない。Slovic や Västfjäll らは、実験的調査をもとにして、統計的な数字で悲惨な社会的問題の全体像を示したとしても、人々は「心理的無感覚」(psychic numbing)に陥り、それほど深く社会的問題を捉えないと述べた[Slovic 07, Västfjäll 14]. むしろ個別具体的な困窮者のストーリーを前面に出したほうが人びとの注意を引く。つまり社会的次元で困窮者のようすを数字で示すよりも、困窮者の心的システムの内面を覗き見られるようにしたほうが他者の倫理的配慮を導きやすい。したがって、他者の心的システムへの共感がなければ、社会システム内の倫理は存立基盤を失いかねないことが見えてくる。明示的であると否とを問わず、他者の心的システムへの共感が働いてこそ、社会システムの倫理的制度が維持・改善されるのである。

4. 人工知能をめぐる倫理的課題

個人的次元でいえば、2017 年現在、これまでと変わらず心的システムは、その内部で構成素たる思考を接続していくオートポイエティック・システムである。すなわち、心的システムのなかに人工知能が直に組み込まれて動作している段階には達していない。もちろん、サイボーグ技術は人工内耳、人工眼、脳深部刺激療法など治療目的ですでに実用化されており、神経系にはすでに組み込まれている。したがって、心的システムと密接な関係にある領域にまで入り込んでいる。また脳波測定の精度が増し、各国で人間の脳のリバースエンジニアリングに向けた大規模な研究が着手されている。ただし神経系は、隣接しているとはいえ心的システムにとってあくまで環境(システムの外部)であり、また人工知能等によって心的システムの思考が直接的に操作可能になっているわけではない。今後、人間とコンピュータとの一体化がさまざまに進められていくだろうが、心的システムの閉鎖系にまで人工知能が及んでいないことは確認しておくべきことである。心的システムは、生命システムを基盤としながらそれ自身で思考を間断なく生み出す閉鎖系であり続けている。

また人工知能の技術が進展しても、心的システムの倫理が社会的次元に完全に回収されることはない。後に述べるように社会的次元では、人工知能等の計算により、人間は単なるデータを格納したりデータを入出力したりするアロポイエティック・システムに見える。しかし個人的次元では、人間は思いやりをもって接するオートポイエティック・システムである。唯一性や私秘性をもった自律システムであり、自分と想像上の立場交換を可能とする存在である。観察者たる自分だけでなく、相手もオートポイエティック・システムであり、相手の心情を顧慮しなければならない存在である。この心的システムの倫理があつてはじめて、社会

的次元の倫理が守られる。というのも前述したように、心的システムへの道徳的配慮が共生の基盤であるからである。危惧すべきは、社会的次元ばかりに注目が集まり心的システムの次元が省みられなくなってしまうことである。そうなれば人間は、単なるアロポイエティック・システムにすぎなくなり、機械と同一視されてしまう。人工知能によってその機能が代替されてしまえば人間の尊厳が脅かされてしまいかねない。人工知能時代であっても、心的システムの内面に配慮した思いやりが不可欠である。

社会的次元の倫理的問題はどうだろうか。いうまでもなく、ビッグデータ型の人工知能は人間が認識できないほどの大量のデータを処理し、実際のサービスに組み込まれることが増えてきた。すなわち、社会システムのコミュニケーション連鎖のなかでコンピュータ技術が組み込まれた「人間＝機械」複合系[西垣08]が成立している。そこでは人間は、データを生み出す存在、あるいはデータの集合体として捉えられる。また、さまざまな変数の相関関係のなかでどのような入力をすればいかなる出力を返すかが予測される存在である。そのような状況下で、人工知能が倫理的な問題を惹起することもしばしば見受けられる。たとえば2015年夏、Googleの写真自動認識サービスで、黒人の写真に「ゴリラ」とタグづけしてしまった問題が起きた。また2016年春、Microsoftの人工知能がTwitter上で人種差別的・性差別的な発言を連発し、すぐさまサービス停止に追い込まれた問題も起きている。あるいは、黒人と白人ではつけられる下の名前にははっきりとした違いがあるが、DeShawnといった黒人の名前を検索すると、白人の名前に比べて犯罪履歴の有無を調べられるサービスの広告がより目立って表示された[Sweeney 13]。実際に逮捕歴がなくとも「DeShawn, arrested?」といった広告が出てくると、それだけで負のイメージがつくられかねない。さらには、人工知能の進展によって失業者が増え経済的不平等が広がっていくのではないかという懸念も挙がっている。

人工知能は、意思決定支援にまで活用されている。コンピュータの予測の精度が人間のそれを上回るという調査結果は数多くある(e.g., [Grove 00])。先に述べたように、一般的に人々は統計的な数字にさして興味を示さない。心的次元の倫理だけでは場当たり的になってしまい、社会全体の最適化を図るには不十分である。計算モデルによって社会全体を俯瞰的に見ていくことが欠かせない。もちろん、「アルゴリズム嫌悪」(algorithm aversion)なる現象もあるだろう[Dietvorst 14]。この現象は、たとえアルゴリズムが人間よりも正確な予測をしても、人間は、アルゴリズムの予測を選ばず人間による予測を好み、アルゴリズムのミスを見るとすぐさまアルゴリズムに信頼を寄せなくなる事象を指す。とはいえ、これほどまで人工知能が高度化してくると、意思決定の場面で人工知能が人間にはない視野を提示してくれる期待が大きく高まっている。たとえば、ネット広告などを手がけるセプテーニ・ホールディングスは、「社員の実績や性格、相性などを数値化し、機械学習によって人材の採用や配置を決めている」[阿曾村 17]。そのことによって退職率が下がったという。

そうしたなか、人工知能の内部処理が社会システムのなかでの公正さや平等性を担保するものになっているかを検証しなければならない。ネオ・サイバネティクスの用語でいえば、人工知能が組み込まれた社会システムの作動を観察して、人工知能の働きによって倫理性が損なわれないように注視しなければならない。そうでなければ、歪んだ差別や排除の生成・助長につながってしまいかねないからである。犯罪の検挙や企業の人事、与信審査などに大きな影響を与えかねない。アメリカでは犯罪リスク評価のコンピュータ・システムがすでに法廷に持ち込まれているが、Propublicaの調査によると、黒人の被告は、実際よりも犯行が常習化すると推定されているのに対して白人の被告は

不当に低リスクが算定されていた[Larson 16]。遺伝的な肌の色で危険度評価が変わってしまっている。公正さが担保されていなかったといえる。また、顔写真データを用いた犯罪者の自動推定も試みられている[Wu 16]。この研究では、もっとも正確性が高かった判定法で10%以上の誤りがあった。また、あくまで確率論であり、顔が犯罪者のカテゴリに属すからといって、その人物が必ず罪を犯すとは限らない。このような自動判定プログラムが社会に広まると、重大な差別を生み出す恐れがある。顔は、生得的に形作られる要素が大きいからである。社会システムのコミュニケーションが倫理性を失い、特定の人間の人物像が歪んで構築されかねない。

現代社会は、人工知能を含めたコンピュータ技術によって人々が得点化される「スコア社会」(the scored society)でもある[Citron 14]。ビッグデータにより、たとえ断片的であっても従来よりも多面的に個人に関するデータが数多く集まり、それらを組み合わせることで人物像が形成され点数化される。コンピュータ技術による個人に対するスコアリングは、人生を左右するほどの潜在的能力をもっている。Citronらは、大学を出て就職に失敗した若い女性が仕方なくファーストフード店でパートタイムの労働についてにもかかわらず、その低賃金労働がさらに雇用可能性のスコアを下げることにつながり、最終的にはフルタイムの職を得ることができなくなってしまった仮想の例を出している。この例では、社会システム内のコミュニケーションの一形態である就職失敗という決定が連鎖的に彼女にとっての負の決定を誘発している。社会システムの自律性が負の方向に立ち上がり、人間はアロポイエティック・システムとしてのデータの集積体として扱われ、社会システムの作動に拘束されている。

第3次ブームを支えるビッグデータ型の人工知能は、素朴な見方ではこれまで以上に大量のデータを解析するので偏見が消えてなくなるとも思われるかもしれない。たしかに多種多様な大量のデータが処理されているのであり、その意味では人工知能のデータ処理を検証することは一部分とはいえ、非明示的で可視化されなかった不公平や偏見を明るみに出すことにつながるだろう。過去の人々と同様、我々もさまざまな偏見に囚われている可能性が高い。意識的にせよ無意識的にせよ、未だに確固たる根拠なしに顔や身長、姿勢、声色、住まい、性別などによって意思決定が大きく影響されることもあるだろう。そういった社会システムにはびこる不公平な偏見が人工知能によって明示化されることによって解消に向かうことが適切である。けれども当然のことながらビッグデータであっても、社会システムのありとあらゆるものを収集しているわけではなく、必ず偏向性が生じる[The White House 14, Federal Trade Commission 16]。アメリカの連邦取引委員会は、低収入や十分な社会サービスを受けられていない人々の排除につながらないように、ビッグデータを扱う企業に対して「データの代表性」「計算モデルに入り込んでいる偏見」「予測の正確性」「倫理的な問題の誘発」を考慮するように推奨している。これは、ビッグデータ型人工知能を使って社会システムを解析する際についても留意が求められる項目である。

ここで想起してもらいたいのは、社会システムを観察している視点からすれば、社会システムのありかたが個人を制約していることである。オートポイエティック・システムとして個人を観察して心の内面をケアしようとしても、社会システムの拘束が働いている。社会システムにおいて、人工知能が介在することによってそのコミュニケーションの流れが歪な方向に導かれているのであれば、問題発見・解決に至らない。ある顔貌や身長、姿勢が会社内の有用な人材に広く見られる特徴であることが人工知能の判定で出された場合、入社試験を完璧にこなしても落とされる人が出てくる。本人の意思がどうであれ、本人の不可知の

領域で採用の判断が決められ、人生の行方が拘束される。たとえ本人に公開されたとしても人工知能の計算処理は、科学的手法により計算結果がはじき出されているという信憑性を帯びるため、異議申し立てしづらい。人工知能が組み込まれた“人間＝機械”複合系の社会システムでは、これまで以上に個人が異論を唱えにくい面を抱えている。したがって、社会システムの倫理性を確保していくのはきわめて重要な意義を帯びている。

こうした人工知能の倫理性の検証や妥当性確認にあたっては、いかにすればよいただろうか。我々は、社会システムを観察することができる。人工知能が歪んだ倫理的問題を導いていることを観察したら、それをコミュニケーションの素材として記述することができる。観察／記述の役割は、専門家が多分に担うことになるだろう。ただしネット社会では一般の人々もみずからの観察を記述することが可能である。それらが社会システムにとって刺激となり、動態的变化を引き起こす契機となる。

そのためには、開発者の経済的活動・技術開発に十二分に配慮しつつ、人工知能が社会に信頼されるためにも透明性や制御可能性、説明責任が守られる必要がある[総務省 16]。もっとも人工知能の内部処理は、どのようなデータを読み込ませているかを含めて企業の競争力の源泉であり、またコストが増えることもあり、完全に公開されることはないと予想される。加えて、説明責任といっても機械学習の内部状態は、読み込むデータとともに変動していくものであり、すべての日時での内部処理の状態を保存しておくことは事実上、困難である。さらにいえば、航空機の運営に準じた監視体制の提案などもされているが、人工知能の定義が明確に定まらない状況下では、どこまでを対象としなければならないか判断とせず、その審査体制の実現性は乏しい。人工知能内部のメカニズムは、きわめて複雑であるゆえ、そのまま公開しても専門家以外には理解できない。簡潔で分かりやすい表示が求められる面もある。

したがって人工知能の倫理性を検討するにあたっては、開発者の経済的活動や技術開発を過度に阻害しないよう慎重に目配りしながら、一定レベルの抽象度を保って内部公開し、事故があった場合に詳細な検証が可能なコンピュータ・システムを構築することが望ましい。読み込んだデータの項目や量、期間、詳細度ならびに項目間の計量的な重みづけなどの概要を可視化し、事故対応に関しては監査証跡のような仕組みを導入することが候補として考えられる。また、人工知能の計算結果が確率論的推論であることも明示化しておくといよい。もっとも先に述べたアルゴリズム嫌悪なる現象が見て取れるため、人工知能の内部処理を完全なるかたちで公開していけば、ビッグデータ型の人工知能は確率モデルに則っているゆえ、人々からの信頼を獲得できない恐れすらある。けれども、科学的信憑性が伴った人工知能の判定が決定論的に受け止められると、偏見が生成・助長され個人の尊厳が脅かされる懸念が生じてくる。人工知能の判定が完全ではなく、たとえ社会システムの作動が安定していたとしても不確実性が内在していることに目を向けるようにしておくことが求められる。

5. 結言

ネオ・サイバネティクスの理論を参照すれば、人工知能はアロポイエティック・システムであり、その作動の責任は人間側にある。そうした観点のもと、本報告では、情報倫理の基礎的な議論を行い、人工知能に関わる倫理を検討してきた。具体的には、オートポイエーシス論や基礎情報学における視点移動の操作により社会的次元と個人的次元との倫理的ありかたの違いを確認した。心的システムの次元の倫理は、相手の心理への道徳的配慮であり、社会的次元の倫理的基盤をなすものである。この心

的システムの領域には、現在でも人工知能が直接入り込んでおらずオートポイエティック・システムのまま作動している。このことを忘れてしまえば、人間は機械と同等であり交換可能なアロポイエティック・システムのみ還元されてしまう。一方、社会システムでは人工知能が入り込んでいる。社会のありかたは個人を拘束するゆえにそこでの倫理性の担保はきわめて重要な意義を帯びる。社会システムの自律性に人工知能が加わることで差別的生成・助長が起こることも想定しなければならず、注意が要される。

本報告で取り上げなかったにもかかわらず、人工知能に関わる倫理的問題について検討すべき点は複数ある。たとえば情報収集のカスタマイズ化(フィルター・バブル)などの心的システムの環境の領域、人工知能を活用した軍事兵器の問題、個人情報統合の扱いなどである。こうした論点については本報告に残された課題であるといえる。

参考文献

- [阿曾村 17] 阿曾村雄太:「退職率下がった」AI 人事に偏見なし, 日本経済新聞, 電子版, 2017/1/30 発行 (2017)
- [Citron 14] Citron, D.K., Pasquale, F.: The scored society Washington Law Review, Vol. 89 (2014)
- [Dietvorst 14] Dietvorst, B. J., Simmons, J. P. and Massey, C.: “Algorithm aversion”, Journal of Experimental Psychology, American Psychological Association (2014)
- [Federal Trade Commission 16] Federal Trade Commission: Big data (2016)
- [Floridi 10] Floridi, L.: Information, Oxford University Press (2010)
- [Grove 00] Grove, W. M., Zald, D.H., Lebow, B.S., Snitz, B.E. and Nelson, C: Clinical versus mechanical prediction, Psychological Assessment, the American Psychological Association Vol. 12, No. 1, 19-30 (2000)
- [House of Commons Science and Technology Committee 16] House of commons science and technology committee: Robotics and artificial intelligence (2016)
- [Johnson 15] Johnson, D.G. and Noorman, M.: Recommendations for future development of artificial agents (2015)
- [河島 16] 河島茂生: ネオ・サイバネティクスの理論に依拠した人工知能の倫理的問題の基礎づけ, 社会情報学, Vol.5, No.2, pp.53-69 (2016)
- [Larson 16] Larson, J., Mattu, S., Kirchner, L. and Angwin, J.: How we analyzed the COMPAS (2016)
- [西垣 03] 西垣通: オートポイエーシスにもとづく基礎情報学, 思想, 岩波書店, No.951, pp.5-22 (2003)
- [西垣 08] 西垣通: 続 基礎情報学, NTT 出版 (2008)
- [Slovic 07] Slovic, P.: If I look at the mass I will never act, Judgment and Decision Making, Vol.2, No.2, pp.79-95 (2007)
- [Sweeney 13] Sweeney, L.: Discrimination in online ad delivery (2013)
- [総務省 16] 総務省情報通信政策研究所: AI ネットワーク化の影響とリスク (2016)
- [The White House 14] The White House: Big data (2014)
- [The White House 16] The White House: Preparing for the future of artificial intelligence (2016)
- [Västfjäll 14] Västfjäll, D., Slovic, P. and Mayorga, M.: Whoever saves one life saves the world (2014)
- [Wu 16] Wu, X and Zhang, X: Automated inference on criminality using face images (2016)