

遡ることはできない。しかし、機械学習では何のデータが元になり何が学習されたかが追跡し得る。他人に権利のあるデータを使って学習して得られたモデルの権利は誰のものになるのだろうか。あるいは、モデルの二次利用はどう考えればよいのだろうか。そうしたことに関する議論も始まっている\*25。欧州では検索エンジンの進展に対して、2012年から「忘れられる権利」（現在では消去権）が認められたが、こうした新しい技術に対応した権利はもっとあるのかもしれない。例えば、防犯カメラが街中に設置されたときに、どんな些細な交通違反でも検出されてしまう。我々は少々の交通違反を「見逃される権利」があるのではないか、あるいは少しくらい「悪いこと」をする権利があるのではないか。あるいは、データが取得されたとしても、ある人には良いところだけを見せ、別の人には（嘘をつかないまでも）そのことを黙っている権利があるのではないか。こうしたことは、既存の法律の概念でいえば、プライバシー権、愚行権、自己決定権、黙秘権などの拡大概念などと解釈することもできる。技術の進展に伴い、こうした議論も必要になるのかもしれない。

個人情報保護法では、特定の個人を「識別できる」情報を含む情報をいう。しかし、我々は「識別器（分類器、classifier）」の能力によって、識別できるかどうかの結果が大きく変わることを知っている。意匠の類否判断はどうだろうか。意匠の類似性は、主に物品の類似と形態の類似から行われるが、我々は、類似度は「類似度関数」によって大きく異なることを知っている。すなわち「識別器」や「類似度関数」といったアルゴリズムの概念を入れないままに、人工知能が普及する時代に、こうした法律的概念がこのまま扱えるのであろうか。あるいは、道路の法定最高速度は、規制速度算出要領により、道路沿いに住宅や店舗があるか、車線数、交差点の数などによりポイントで決められる。しかし、仮に自動運転が導入され、細かい気象条件や人通りの多さが時々刻々と計算されるなかで、より本質的な「安全性」に従った条件を設定することも可能ではないだろうか。

自動運転に関しては、すでにさまざまな議論が行われている。冒頭に書いたようにすでに米国では死亡事故も起きているが、現状のレベルの自動運転車においては、事故を起こした場合の責任は運転者にある。米国運輸省高速道路交通安全局は、2016年2月に、Googleに自動運転車に搭載される人工知能を「運転手とみなす」と伝えた\*26。国内でも警察庁で自動運転における事故の責任や法律上・運用上の課題に関する議論が始まっている[警察庁 16]。自動運転のレベル1～3までは現行法と

あまり変わらないと思われるが、レベル4の段階になると、きめ細やかな判断が必要になり、ある場合には事故の責任が製造者責任となり、また保険の仕組みも整備される必要があるだろう[損保 16]。倫理委員会ではさらに進んで、人工知能が法人格に似たある種の格をもつ「人工人」（自然人、法人と対比的に）という概念ができるのではないかと議論した。これと同じ議論が[カプラン 16]でも行われており、将来は、何らかの人工知能に対応する格ができ、（法人が利益を競うのと同じく）安全性を競うようになるのかもしれない。

自動運転の話でよく出てくるのがトロッコ問題（あるいはトロリー問題）である[エドモンズ 16]。サンデル教授で有名な議論であり、他の人を助けるために別の人を殺してもよいのかという思考実験である。こうした議論が意味するところを単純化していうと、人間の本能・感情からしてどちらが正しい「ような気がするか」\*27という点をベースとしながら、「どういった社会規範を上位に置くと社会が安定するか」を考える設計の問題と捉えることができるだろう。前者は、どういったルールを採用するとどのくらい人の心に合うか・反するかというコストの設定の問題、後者は、社会の安定や発展という目的に従って諸制度をどのように設計するかという問題である。トロッコ問題でいうと、両方の状況が不可避なときに失う命の数の少ないほうを選ぶのだが、全く事件に関与しない第三者の命を脅かすことは、社会の安定性からして危険である。例えば、五人を守るためといって、いきなりある日殺されてはたまらないので、なくなる命の数が少ないほうが良いという規範より、善意の第三者は命を脅かされないという規範を上にもっていったほうが社会の安定性という観点からは良い。こうした話はこれまでのように単なる思考実験ではなく、人工知能技術の進展によりややもすれば現実味を帯びてきている。人工知能の技術が、倫理学や道徳心理学、法哲学などを巻き込んだ議論を引き起こしており、言い方を変えれば、人工知能の進展が人文社会学系のさまざまな学問分野に対して、これまで積み上げてきたそれぞれの学問分野での体系に挑戦状を叩きつけ、破壊と創造を迫っているともいえるかもしれない。

そして、こうした議論の先には、結局は、我々はどういった社会をつくりたいのかという問題に行き着くのではないだろうか。社会全体での目的関数が決まれば、人工知能の活躍する範囲は広い。しかし、社会全体の目的関数と簡単にいっても、非常に難解で複雑な議論が必要であることは論をまたないし、そのことは人類の歴史が雄弁に物語っている。一方で、人間が社会的な「動物」であることを考えると、人類という種を存続させることは、（ひとまず総論としては）人類全体の合意になり

\*25 「学習済みモデル」は知財か？一産業技術総合研究所、AI市場形成で価値・権利認めるルールの検討開始、日刊工業新聞（2016年5月25日）

\*26 「人工知能は運転手」米当局、Google自動運転に見解、朝日新聞（2016年2月14日）

\*27 進化倫理学あるいは神経倫理学などの分野での議論に近い。

得るのではないかと思う。IJCAI-13では、環境や経済、社会的需要に対する持続可能な発展と未来のための人工知能技術がテーマであった [Song 13]。人工知能を、人類の持続可能な未来のために活用するというのは、人工知能と倫理という文脈の上では、一つの究極の目標ではないだろうか。

やや空想的な議論になることをあえて恐れずにいうと、人工知能技術を活用した新しい形の社会が誕生する可能性もあるかもしれない。例えば、ディープラーニングによる認識技術により、労働者の「努力」が正当に評価されるようになれば、努力に応じた富の配分も可能になる。それは社会主義国家の致命的な欠陥であったフリーライダーを防げないという問題を解決し、理想の社会主義的な国家を実現してしまうかもしれない。格差を減らし、あるいは資源の消費を抑え、人々の多様性を重視しながら、人々が自由に生き生きと活躍するような持続可能な社会を実現することも可能になるかもしれない。これは少し楽観すぎる物語だろうか。

最後はかなり楽観論になってしまったが、人工知能の技術により我々の未来をどうしたいか。これは、研究者が自ら自覚と責任をもって正しい情報を社会に発信し、人工知能のもつ可能性とリスクを多くの人に正しく理解してもらい、そして社会全体を巻き込んで議論していかなければならないのではないだろうか。

## 6. おわりに

SF作家のIsaac Asimovは、小説のなかでロボット三原則を考案した。ロボット三原則は必ずしも完璧なものではなく、小説の中でもさまざまな問題とともに描かれているが、コンセプトとしてはシンプルでわかりやすい。それに対抗するような、シンプルでわかりやすい「人工知能三原則」はないだろうか。ずっとそう思っているが、本稿での議論を見てもわかるとおり、論点が多岐にわたり、とても簡単にまとめられそうもない。しかし、この議論が成熟し、より多くの人にわかりやすい形で情報発信ができればと思う。

人工知能と倫理の話は、究極的には、我々がどういう社会をつくりたいのかという話に帰結する。そういった責任ある議論を、人工知能という技術が土台となっていてきていることをうれしく思うと同時に、今後は、多くの人文社会学系の研究者も巻き込んで議論を進めていく必要があるだろう。そして、人工知能に関わる多くの研究者が、社会の在り方に関してのこうした議論に少しでも加わっていただければ幸いである。

## ◇ 参考文献 ◇

- [Amodei 16] Amodei, D., et al.: Concrete Problems in AI Safety, arXiv:1606.06565 (2016)
- [Barrat 13] Barrat, J.: *Our Final Invention, "Artificial Intelligence and the End of the Human Era."*, Thomas Dunne Books (2013)
- [Bostrom 14] Bostrom, N.: *Superintelligence*, Oxford University Press (2014)
- [知財 16] 知的財産戦略本部：次世代知財システム検討委員会報告書 (2016年4月)
- [Chui 16] Chui, M., Manyika, J. and Miremadi, M.: Where machines could replace humans — and where they can't (yet), *McKinsey Quarterly* (2016)
- [エドモンズ 16] デイビッド・エドモンズ：太った男を殺しますか？ — 「トロリー問題」が教えてくれること, 太田出版 (2016)
- [江間 16] 江間有沙, 長倉克枝: 公開討論「人工知能学会倫理委員会」(特集「2016年度人工知能学会全国大会 (第30回)」), 人工知能, Vol. 31, No. 6 (2016, 掲載予定)
- [Frey 13] Frey, C. B. and Osborne, M. A.: *The Future of Employment: How Susceptible are Jobs to Computerization*, [http://www.oxfordmartin.ox.ac.uk/downloads/academic/The\\_Future\\_of\\_Employment.pdf](http://www.oxfordmartin.ox.ac.uk/downloads/academic/The_Future_of_Employment.pdf) (2013)
- [Gatys 15] Gatys, L. A., Ecker, A. S. and Bethge, M.: A Neural Algorithm of Artistic Style, arXiv preprint arXiv:1508.06576 (2015)
- [伊庭 16] 伊庭齊志: 人工知能と人工生命の基礎, オーム社 (2016)
- [井上 16] 井上智洋: 人工知能と経済の未来 2030年雇用大崩壊, 文藝春秋 (2016)
- [カプラン 16] ジェリー・カプラン: 人間さまお断り—人工知能時代の経済と労働の手引, 三省堂 (2016)
- [警察庁 16] 警察庁交通局: 自動運転をめぐる最近の動向と警察庁の取組について (2016年6月)
- [Kurzweil 05] Kurzweil, R.: *The Singularity is Near*, Duckworth Overlook (2005)
- [野村 15] 野村総合研究所: 日本の労働人口の49%が人工知能やロボット等で代替可能に, プレスリリース (2015年12月2日)
- [松尾 15a] 松尾 豊 ほか: 人工知能学会倫理委員会の取組み (アーティクル), 人工知能, Vol. 30, No. 3, pp. 358-364 (2015)
- [松尾 15b] 松尾 豊 ほか: 公開討論「人工知能学会倫理委員会」(〈特集〉2015年度人工知能学会全国大会 (第29回)), 人工知能, Vol. 30, No. 6, pp. 754-755 (2015)
- [松尾 15c] 松尾 豊: 人工知能自動運転: ものづくりが変わる日本の製造業に勝機あり, エコノミスト (2015年10月6日号)
- [西田 15] 西田豊明: 人間力・社会力を強化する情報通信技術 人工知能を中心に, 情報管理, Vol. 57, No. 8, pp. 517-530 (2014)
- [Orseau 16] Orseau, L. and Armstrong, S.: Safely Interruptible Agents, <https://intelligence.org/files/Interruptibility.pdf> (2016)
- [Pistono 16] Pistono, F. and Yampolskiy, R. V.: Unethical research: How to create a malevolent artificial intelligence, *Proc. IJCAI-2016 Ethics for Artificial Intelligence Workshop* (2016)
- [Russell 15] Russell, S., Dewey, D. and Tegmark, M.: Research priorities for robust and beneficial artificial intelligence, *AI Magazine*, winter (2015)
- [Song 13] Song, J.: Addressing sustainability via AI — Report from the 23rd International Joint Conf. on Artificial Intelligence, *Bulletin of the Chinese Academy of Sciences*, Vol. 27 (2013)
- [損保 16] 日本損害保険協会: 自動運転の法的課題について (2016年6月)
- [若田部 16] 若田部昌澄: 経済学者は人工知能の夢を見るか: 第2次機械時代の経済社会構想, 総務省 ICT インテリジェント化影響評価検討会議 (2016年2月2日)
- [Walsh 16] Walsh, T.: The singularity may never be near, *Proc. IJCAI-2016 Ethics for Artificial Intelligence Workshop* (2016)

2016年8月22日 受理

## 著者紹介



松尾 豊 (正会員)

1997年東京大学工学部電子情報工学科卒業。2002年同大学院博士課程修了。博士(工学)。産業技術総合研究所、スタンフォード大学を経て、2007年10月より、東京大学大学院工学系研究科准教授。2012～13年度本学会編集委員長、2014年より本学会倫理委員長。専門は、Webマイニング、Deep Learning、人工知能。



西田 豊明 (正会員)

1977年京都大学工学部卒業。1979同大学院修士課程修了。1993年奈良先端科学技術大学院大学教授。1999年東京大学大学院工学系研究科教授。2001年東京大学大学院情報理工学系研究科教授を経て、2004年4月京都大学大学院情報学系研究科教授。現在に至る。会話情報学、原初知識モデル、社会知のデザインの研究に従事。本学会会長(2010～11年度)、日本学術会議連携会員(2006年～)、情報処理学会フェロー、電子情報通信学会フェロー、JST CREST「共生社会に向けた人間調和型情報技術の構築」研究領域総括(2014年～)。AI & Society誌 Associate Editor。



堀 浩一 (正会員)

東京大学大学院工学系研究科教授。2008～09年度本学会会長。2010年～本学会顧問。知識処理システム構築のための知識獲得の困難に直面したのち、研究すべきは、知識の獲得ではなく知識をつくり出すことの支援であると考えようになり、創造活動支援システムの研究を行ってきた。人間がもっと創造的に仕事するための道具としての人工知能の研究を行ってきたが、単なる道具では済みそうになってきたので、最近では人工知能倫理の問題にも興味をもっている。



武田 英明 (正会員)

1986年3月東京大学工学部卒業。1988年3月同大学院工学系研究科修士課程修了。1991年3月同博士課程修了。工学博士。ノルウェー工科大学、奈良先端科学技術大学院大学を経て、2003年5月より国立情報学研究所教授。2006年4月～2010年3月同学術コンテンツサービス研究開発センター長。2005年12月～2010年3月東京大学客員・特任教授。知識共有、セマンティックWebなどの研究に従事。AAAI、Design Society、情報処理学会各会員。



長谷 敏司

1996年3月関西大学文学部卒業。2001年角川スニーカー大賞にて作家デビュー。2009年よりSFを本格的に書き始め、2009年『あなたのための物語』、2012年『BEATLESS』で日本SF大賞ノミネート、2014年『My Humanity』で日本SF大賞受賞。作品は近未来を舞台にAIが登場するものが多く、2014年ニコニコ学会参加をきっかけに、AIが現実社会に実装されるとき、社会の変化を望まない人々にまで触れてしまうインパクトについて考えるようになる。日本SF作家クラブ会。



塩野 誠

1998年慶應義塾大学法学部政治学科卒業。2008年ワシントン大学ロースクール法学修士。株式会社経営共創基盤取締役。国内・海外にて企業、政府機関に対して戦略立案・実行のアドバイスをを行っている。テクノロジー関連の事業開発、企業提携やM&A、企業危機管理の実績を多く有する。



服部 宏充 (正会員)

2004年3月名古屋工業大学大学院博士課程修了。博士(工学)。2004年4月～2007年3月日本学術振興会特別研究員(PD)。リバプール大学、およびマサチューセッツ工科大学客員研究員を経て、2007年4月～2014年3月京都大学大学院情報学系研究科助教。現在、立命館大学情報理工学部准教授。マルチエージェントシステム、社会シミュレーション、デザインに関する研究に従事。本学会の「表紙問題」における社会の反応から、人工知能と社会の関係に関心を持ち、有志とともにAIR (Acceptable Intelligence with Responsibility) の活動を開始。



江間 有沙 (正会員)

2007年東京大学教養学部卒業。2012年同大学院総合文化研究科博士課程修了。博士(学術)。2012年京都大学白眉センター特定助教。2015年4月より東京大学教養学部附属教養教育高度化機構特任講師。2014年より人工知能と社会の関係について考えるAIR (Acceptable Intelligence with Responsibility) 研究会を有志とともに開始。専門は科学技術社会論。



長倉 克枝

2007年北海道大学獣医学部卒業。同年、日本経済新聞社入社(科学技術部、証券部記者)。2011年退社。文部科学省系独立行政法人勤務などを経て2015年からフリーのライター・編集者。テクノロジーと社会、医療などを中心に取材・執筆。朝日新聞出版AERA編集部所属(記者)。