

自動運転車の倫理的側面を考える

交通事故の大幅な減少をもたらすものとして期待が高まっている自動運転車について、事故発生時の責任の所在やAIの説明責任に関するルール作りなどが求められている。本稿では、自動運転車におけるAIの説明責任を題材に、近年議論されているAIの倫理的側面について考える。

1. はじめに：トロッコ問題とは

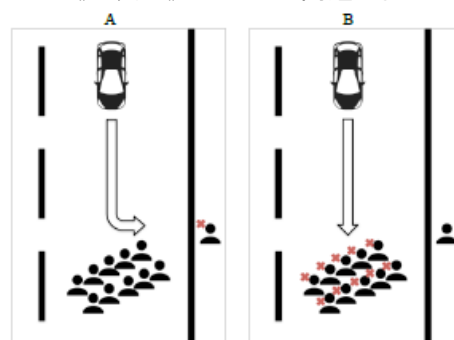
トロッコ問題とは、イギリスの倫理学者・哲学者 Philippa Foot が提唱した問題であり、例えば《図表 1》のように、ブレーキの故障した自動車がハンドルを切ると1人の命が失われ（A）、直進すると10人の命が失われる（B）、といった場合に、A/Bどちらを選択するかというジレンマに陥るような問題である¹。

また、この問題の派生形として、隣にいる人を自動車の前に突き落とすことで、その1人の命は失われるものの10人の命は助けられる（A'）、何もしなければ10人の命が失われる（B'）という「太った男問題」というものもある。

より純粋なトロッコ問題においては、例えば《図表 1》のようなケースではAを選択する人が多いだろうが²、同じ結果となっているはずの「太った男問題」においてはAという選択をすることに強い違和感を覚える人が多いだろう。人間の倫理的な判断とは、その時々状況の影響を強く受け、必ずしも一貫性がないという点において興味深いものである³。

2018年10月には、このようなある種の「究極の選択」に関する大規模な調査をもとにした分析結果がNature誌に掲載されるなど⁴、議論が活発化していくことが想定される⁵。本稿では、交通事故の大幅な減少をもたらすものとして期待される自動運転車を題材として、近年議論されているAIの倫理的側面について考えてみたい。

《図表 1》トロッコ問題の例



【出典】河島茂生ら「自動運転車の『トロッコ問題』などに関する意識：日本に居住する人に対する質問紙調査を通じて」

2. 自動運転車とトロッコ問題に係る、よくある疑問について考えてみる

自動運転車とトロッコ問題についての疑問 —なぜこんなことを議論しなくてはならないのか等— についていくつか取り上げ、考え方の参考となるものを紹介する。

(1) トロッコ問題は人間のとるべき行動としても結論が出ていない。にもかかわらず、自動運転車やAIの問題として議論する必要があるのか？

[考え方] トロッコ問題は、AIだからこそ問題になる。

このような究極の選択が求められる状況では、人間はパニックに陥ることも想定され、また瞬間的な「とっさの判断」をとらざるを得ず、それは必ずしも理性的、慎重な思考を経た判断とは言えない⁶。だからと言って、まさに「とっさの判断」なのであるから、その判断が倫理的に間違っている、倫理的責任が加重されるといった問題は起こりにくい。

しかし、AIやアルゴリズムは、究極の選択が求められる状況においてもパニックには陥らないし、ある意味で冷静に学習してきたことに基づいた判断を下すため、その判断は「とっさの判断」とは言えな

い。技術の進歩に伴い、AIはより複雑な状況を解釈できるようになってきていることから、まさに人間における「とっさの判断」という問題視されなかったものが⁸、AIにおいては倫理的かつ現実的な問題のものとなっており⁹、事前の学習・プログラミングが問題となっているのである¹⁰。

また、多数の死亡者を生み出している交通事故に係る現状に鑑みると、むしろ「急がず、ゆっくりと慎重に議論を重ねてから自動運転車技術を導入する」ということは非倫理的であるとも言えるだろう¹¹。

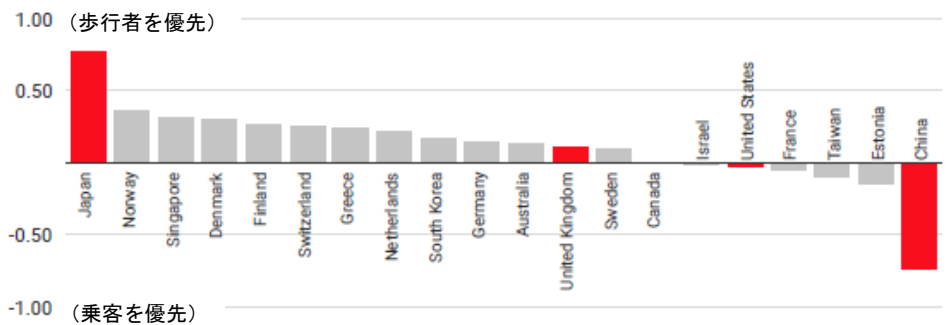
(2) AIの場合だからこそ倫理的側面が問題となるということであるが、だからといって結論が出るとは考えにくい。そもそもこのような究極の状況は起きにくく、考える必要はないのではないか？

[考え方] 結論を出すのが難しくても、規制・ルールを作る上で、我々はどのような価値観を持っているのか、どのような自動運転車であれば社会は受容するのか、といったことは考慮すべき。

何が正しい／間違っているということは、価値観をもとにしているために確かに結論が出にくい。しかし、何が正しい／間違いを問題にしているのではなく、社会に受け入れられるためには人々の考えというものが重要な役割を担うということを問題にしている¹²。例えば「乗客の命を最優先し、歩行者の命は全く考慮しない」という自動運転車が開発されるというのは、社会として許容できるのだろうか¹³。

このような極端なAIが社会に出てこないよう、事前に何かしらの規制・ルールを設けておくことは大事なことであろう。また、倫理的な考え方は、発生頻度に依存するものではなく、事前に合意すべきことである¹⁴。

《図表2》歩行者の命を優先する傾向（国別比較）



【出典】MIT Technology Review, “Should a self-driving car kill the baby or the grandma? Depends on where you’re from”. に当研究所にて一部補記

(3) そもそも「乗客の命を最優先にする自動運転車」でないと売れないと考えられる。AIがどのような判断を下すべきかについては、ある意味で既に結論が出ているのではないか？

[考え方] 市場原理というものが社会にとって最適なものであるとは限らない。むしろ「市場の失敗」という現象が起きるからこそ、規制が必要である。

人々や企業が自己の利益を追求することだけを考えていては、格差や環境問題といった社会的な問題は解決されない。だからこそ規制が必要なのであり、どのような規制を整備するかについて議論の必要性が問われているのである。

なお多くの方は、被害者数の最小化を企図する功利主義的な自動運転車（図表1におけるA）を望んでいるが、実際に購入したい自動運転車は乗客を優先的に守ってくれるものであるという調査結果も出ている。他人が利用する場合には被害者数が少ない功利主義的な自動運転車を望むが、自身が利用するにあたっては自分が犠牲になる可能性が高い自動運転車は買いたくないのである（功利主義的な自動運

転車だと、乗客が最も人数が少ないという場合に自分が犠牲になる)。

交通事故による犠牲者数に鑑みれば自動運転車の早期採用・普及が望ましい。しかし、功利主義的な自動運転車は購入されにくい。一方で倫理的に違和感のある(事故発生時に多くの歩行者を犠牲にしかねない)自己防衛的な自動運転車の方が、早期に採用・普及しやすいことから結果として交通事故犠牲者が少なくなるという、別のジレンマも存在している¹⁵。

(4) 全ての個人の価値観を反映させることは難しく、そのような規制・ルールを設ける必要性には疑問がある。また、そもそも開発者は規制を望んでいないのではないか？

[考え方] 規制・ルール化しなかった場合、開発者は「事前に考慮することができたにもかかわらず対処しなかった」として非難を受けることが想定され、しかもその非難は「非倫理的である」という性質のものとなる可能性がある。

社会で実際に使われる自動運転車のメーカーや技術者は、事後的に非倫理的であるとして非難され、自分たちだけが責任を負わなくてはならないという状況を望んでいない¹⁶。非難され責任を負うことになるくらいなら、事前に何かしらの規制・ルールを作成してもらうことを望むであろう。また、開発者たちは法的責任という具体的で予想しやすい問題は気にするだろうが、善悪・価値観といった問題にはそれほど関心を持たないことから、規制やルールによって決着させるべきという側面もある¹⁷。

(5) 開発者が「非倫理的である」と非難されることを避けたくとも、全ての人の価値観に合致する判断というものが存在するとは考えにくく、こういった「究極の選択」に関する学習をさせずにAIを開発することが解決策なのではないか？

[考え方] AIに係る問題の一つに、例えば自動運転車がどう運転時に判断しているのかわからないという「ブラックボックス」の問題がある。つまり、あらかじめ明確にしておかない場合、事故が起こってから「轢かれた理由」が明らかになる、またはそもそも「理由」が明らかにならないという事態になるが、果たして被害者やその家族に納得感はあるだろうか。

人間の脳が「意味のある重要な情報だけを選別して知覚している」ということは知られているが¹⁸、AIは人間なら無視するようなエラー・情報も無視せずに処理してしまっている点で人間の脳とは異なっている¹⁹。つまり、人間には考慮されない情報、例えば歩行者の服の色など、についてもAIは無視せずに学習していることから、そのような人間には考慮されない情報をもとに何かしらの判断を下すことが考えられる。意図的に学習させない、プログラミングしない行動は、仮にそのような行動をとったとしてもAIは何かしらの判断を下してくるために解決策になっておらず、むしろ「ブラックボックス」としての性質が濃くなってしまう可能性がある。どのような根拠で判断したのかわからないAIの判断結果を社会が受け入れるとは考えにくいだろう。

(6) 確かに「ブラックボックス」については好ましくないと考える。しかし、価値観について議論するのは難しく、いっそのことAIがランダムに判断すれば公平なのではないだろうか？

[考え方] 「倫理的議論は難しいから、議論しない」ということで、よいのだろうか。

ランダムに判断させるように AI を設計すれば、確かに差別的取り扱いを回避できるかもしれない。しかし、ヒューマン・エラーを解消するためのテクノロジーである AI・自動運転車の能力を意図的に用いない点、プログラミングの段階で悲惨な結果を回避しえたのに、その選択の機会を意図的に捨てきつているという点、難しいから議論しないということがそもそも倫理的に許容されないという点、などからの批判がある²⁰。

3. おわりに

本稿では、トロッコ問題を例に、AI の倫理的側面について、いくつかの疑問を整理してみた。整理すべき問題は上記だけでなく、「子どもをなるべく助けるべきかどうか」、「安全性の高い相手を標的にすべきか（例えばバイクと自動車なら、丈夫そうな自動車にぶつかるべきか）」、といったものもあり²¹、議論は尽きない。

しかし、消費者の自動運転車・AI のブラックボックスに対する不安感・抵抗感を解消し、自動運転車を普及させることは多くの人命を救うという観点からも重要である。今後、議論を重ねていくことで納得感のあるルール作りをすることが望まれる²²。本稿により、多少なりとも関心を持ってくれる方が増えれば幸いである。

【副主任研究員 吉田 順一】

¹ 多くの場合、被害者数を最小化する（ハンドルを切る）という「功利主義」と、意図的に人を害することは許されない（ハンドルを切るとは好ましくない＝直進する）という「義務論」の対立によって語られる。

² 我が国において、《図表 1》の状況で「A を選ぶべき」だとする人が全体の 79.9% を占めたという調査結果がある（河島茂生ら「自動運転車の『トロッコ問題』などに関する意識：日本に居住する人に対する質問紙調査を通じて」（2017 年社会情報学会 (SSI) 学会大会発表資料）。

³ WIRED 「人間の倫理は非理性的か：『トロッコ問題』が示すパラドックス」（2008 年 11 月 11 日）

⁴ Edmond Awad, et al. “The Moral Machine experiment”. *Nature*, Vol. 563, Oct 24, 2018.

⁵ ドイツでは 2017 年 6 月に Automation and Connector Driving に関する倫理委員会が 20 の倫理規則を発表している（Ethics Commission Automation and Connector Driving, “Report”. Jun 2017.）。

⁶ 時間的余裕がある場合、人間は一生懸命ブレーキをかけたり、通行人に避難するようメッセージを送ったりすることが想定される。トロッコ問題は、そのような余裕のない、究極の選択を求められる状況下での判断を問題としている。

⁷ 多くの場合「その状況では仕方なかった」と評価・解釈されることになると思われる。ただし、この場合の「仕方がない」は「とっさの判断」についてのものであり、例えば、法的責任が免責されるべきであるということの意味するものではない。

⁸ 「とっさの判断」に対する法的見解を示した判例として、エッカー特事件がある。目の前の線路に幼子が座って轢かれそうな場面に遭遇したエッカー特は、とっさの判断で幼子を救うも、自身は轢かれて死亡してしまう。賠償責任に関する裁判において NY 州最上級審はエッカー特の寄与過失について注意義務違反はないと評価、「とっさの判断」に過失を認定しなかった（平野晋『ロボット法』と自動運転の『派生型トロッコ問題』『NBL』No. 1083（2016 年 10 月 1 日））。

⁹ MIT Technology Review, “How to Help Self-Driving Cars Make Ethical Decisions”. Jul 29, 2015.

¹⁰ 規制当局トップである米国国家運輸安全委員会のクリストファー・ハート委員長（当時）は、ジレンマとなる事態についてもシステムに織り込む必要があり、倫理的選択は自動化システムには避けられないと語っている（MIT Technology Review 「完全自律自動車は実現しない 米政府当局トップが語る」（2016 年 9 月 1 日））。

¹¹ 前掲脚注 9

¹² MIT Technology Review, “Why Self-Driving Cars Must Be Programmed to Kill”. Oct 22, 2015.

¹³ 日本人は「乗客 or 歩行者」であれば、歩行者の命を優先すべきであるとする傾向が強いようである（MIT Technology Review, “Should a self-driving car kill the baby or the grandma? Depends on where you’re from”. Oct 24, 2018.

¹⁴ 前掲脚注 4

¹⁵ Jean-François Bonnefon, et al. “The social dilemma of autonomous vehicles”. *Science*, Vol. 352, Jun 24, 2016.

¹⁶ クリストフ・リュトゲ「自動運転車のための倫理：ドイツの事例から」『千葉大学法学論集』第 32 巻第 3・4 号（2018 年）

¹⁷ Bryan Casey, “Amoral Machines, or: How Roboticians Can Learn to Stop Worrying and Love the Law”. *Northwestern University Law Review*, Vol. 111, Iss. 5, Aug 2017.

¹⁸ このようなメカニズムを「効率的選択」という（理化学研究所 Web サイト『効率的選択』では脳は注意を向け集中を高める」（2011 年 12 月 8 日））。

¹⁹ MIT Technology Review, “The Dark Secret at the Heart of AI”. Apr 11, 2017.

²⁰ 平野晋『ロボット法』と自動運転の『派生型トロッコ問題』『NBL』No. 1083（2016 年 10 月 1 日）

²¹ 前掲脚注 20

²² 我が国でも「AI の 7 原則」をもとにした法整備が進められていくことになっており、原則の一つに「AI を利用した企業に決定過程の説明責任」を課すというものがある（日経新聞電子版「AI の判断、企業に説明責任 ルール作りへ政府 7 原則」（2018 年 11 月 26 日））。