

テクノロジーの進化と哲学的視座 ——なぜ自動運転車で倫理が問題となるか?——

岡本裕一郎 玉川大学文学部客員教授

20世紀後半から始まったデジタル情報テクノロジー革命は、私たちの未来社会に何をもたらすのだろうか。この壮大なテーマを考えるため、ここでは具体的な事例に絞って、その可能性を探ってみたい。

問題にしたいのは、人工知能が搭載された「自動運転車 (Autonomous Vehicles)」をめぐる倫理的状況である。自動運転車について言えば、世界中で開発競争が行なわれ、今や公道でも少しずつ走り始めている。やがて、「運転手無しのクルマ (Driverless Cars)」が走る日も近い、と言われたりする。

しかし、このとき、技術的な問題もさることながら、哲学的な難問も控えている。それはいったい何か、あらためて考えてみることにしよう。



1. トロッコ問題から自動運転車へ

まず、次のような状況を想定してみよう。これを「衝突問題」と呼んでおく。

衝突問題：大きな自動運転車が凍った路面に気づかず、スリップしてしまった。見ると、その先には、小さなミニバンに5人の乗員が乗っていた。このままでは、衝突は避けられず、衝突すればミニバンは大破し、5人全員が死亡するだろう。ところが、大きな自動運転車は、何とか左に進路を変えることができそうである。しかし左前方には、小型のオープンカーが止まっていて、一人の運転手が乗っていた。さて、この状況で、自動運転車はどのようにプログラムされるべきだろうか？

この状況を読んで、「トロッコ問題だな！」と気づかれた人も多いだろう。「トロッコ問題」というのは、1967年にイギリスの哲学者フィリッパ・フットが人工妊娠中絶を議論するために、提起した思考実験である。その後、アメリカの哲学者ジュディス・トムソンによって定式化されて、いろいろなバージョンが作り出された。

「トロッコ問題」については、マイケル・サンデルの「白熱教室」でも紹介されているので、あらためて繰り返すことはしない。ここで想定した「衝突問題」は、やや人工的な「トロッコ問題」の現実的なバージョンと見なされている。では、この状況で、いったいどう対処したらいいのだろうか。このまま何もしなければ、ミニバンにぶつかり、その乗員5人が死亡する。進路を変えて、オープンカーにぶつかれば、一人の運転手が死亡する。この状況で、「自動運転車」はどのようにプログラムしておくべきか。

この状況が「現実的なトロッコ問題」と呼ばれているのは、5人の命か1人の命かの選択を迫るところにある。この状況では、人物の特性や関係などはまったく捨象され、考慮すべき要因が、人数のみに限定されている。こうした抽象的な場面設定という点で、オリジナルの「トロッコ問題」の延長上にあるのだが、さらにもう一つ指摘しておけば、選択する立場の人（「大きな自動運転車の乗員」）に危険がないことも共通である。

「大きな自動運転車」に何人乗っているかわからないが、少なくともその乗員はどちらが選択されても、無事なのである。しかし、現実的に考えて、そうした状況はあるのだろうか。そこで今度は、乗員そのものにも危険が及ぶ事例を想定してみよう。



2.「乗員ファースト!」の原則は貫かれるか?

自動運転車は、「トロッコ問題」とは違って、乗員そのものの命にかかわることが少なくない。むしろ、そうした場合の方が多い、と言うべきかもしれない。とすれば、新たな原則が必要になるかもしれない。

それを考えるために、「トンネル問題」と呼ばれる状況を想定してみよう。やや技巧的な状況だが、「トロッコ問題」と関連づけるために、次のような思考実験を作ってみよう。

トンネル問題：自動運転車に乗員が一人で、山道を走っていた。本を読んだり、風景に見とれたりしていたら、前方にトンネルが迫ってきた。そのとき、突然5人の子どもが車の前に飛び出してきた。ブレーキでは間に合わず、唯一可能な回避手段はハンドルで方向を変えることだった。ところが、方向を変えると、車はトンネルに激突し、乗員の死亡は避けられない。この状況で、自動運転車はどのようにプログラムされるべきだろうか?

こんな状況を想定するのは、何ともやり

きれないが、しかしその可能性がまったく
ない、とは言い切れないだろう。むしろ、
条件をいくつか変えてみれば、現実的な状
況とも言える。とすれば、大いに議論すべ
きではないだろうか。

単純な人数からいえば、功利主義的に考
えて、5人の子どもの命を優先し、車はト
ンネルにぶつかるようにプログラムすべき
だ、と言えるかもしれない。しかも、相手
は子どもなので、世論の反響を想像すれば、
間違いなくこの選択がベストのような気が
する。しかし、自分が自動運転車の所有者
であり、乗員であったらどうだろうか。そ
の車を購入する前に、あらかじめこのプロ
グラムについて説明を受けたとすれば、は
たして購入する気になるだろうか。まして、
自分ではなく家族が乗車していると想像し
ただけでも、結果は見えているのではない
だろうか。

おそらく、乗員の安全を守ってくれない
ような車は、購入する気にならないのでは
ないだろうか。高いお金を払って車を買う
のに、その所有者の安全を第一にしないよ
うな車なんて、誰も買う気にならないはず
である。

だからこそ、某クルマメーカーのエンジ
ニアは、あるとき雑誌で「所有者・乗員
ファースト!」の方針を打ち出したのであ

る。ところが、こんな主張をメーカーの方
針として打ち出したら、社会的に非難され
るかもしれない。5人の子どもよりも、裕福
な1人をクルマメーカーは優遇するのか? 1
人の子どもの命さえ重視すべきなのに、5人
の子どもの命を無視して、利益優先の方針
でいいのか? そんな非人道的な会社は認め
られない、等々。

そのことを察知して、某クルマメーカー
は、「エンジニアの発言は会社の公式的な方
針ではない」と声明を出したのである。し
かし、これで問題が解決しないのは誰の目
にも明らかだろう。

そもそも、方針はいったい誰が、どのよ
うに決めるべきだろうか。メーカーか、政
府か、所有者か。また、方針は同じにすべ
きか、それとも多様性を認めるか。そのさ
い、何を最も重視すべきか。問題は山積し
ているのに、現状は手つかずの状態なので
ある。

◆ 3.人物特性はどこまで考慮すべきか?

今度は、「トンネル問題」より、もう少し
複雑な状況を想定してみよう。それを「タ
ンクローリー問題」と呼ぶことにする。こ
の問題で重要になるのは、それぞれの人物

特性である。自動運転車の場合には、人物認証システムが備わっているはずなので、歩行者や乗員の人物認証は可能であろう。

MITの「モラル・マシーン」というウェブ・サイトを見ると、自動運転車にまつわるさまざまなシナリオが設定されているが、そのときポイントになるのが、登場人物のさまざまな特性である。これは、従来の「トロッコ問題」では、意図的に排除された要因だったが、現実の自動運転車の状況を考えると、無視できないのである。

たとえば、信号を無視した犯罪者と、信号を守った妊婦では、同じようには対処できないはずだ。その他、多様な状況が考えられるが、取りあえず、ここではまず場面を確認することにしよう。

タンクローリー問題：自動運転車に、一人の乗員が乗っている。前方からタンクローリー車が近づいてきたが、横転してしまい前方の道をふさいでしまった。左に曲がれば、子どもが一人いて、右に曲がれば老人が二人いる。ブレーキでは間に合わず、止まることができないので、まっすぐ進んでタンクローリーにぶつかるか、子どもか老人を轢いてしまうことになる。この状況で、自動運転車は、どのようにプロ

グラムされるべきだろうか？

被害者の数で考えてみると、乗員や子どもは1人、対して老人は2人なので、乗員か子どもを犠牲にすべきだろうか。それとも、老人は2人だとしても、残された人生の時間からすると、犠牲になっても構わないのだろうか。

このとき、歩行者の子どもが自動運転車の所有者の子どもだったら、子どもと乗員はぜひとも守るべきとなるかもしれない。あるいは、老人が2人ではなく、5人のときも、同じような方針でいいのだろうか。

今までのように、自動運転車が実用化されていない時代であれば、おそらくこうした状況は問題とならず、ドライバーのとっさの判断で済ませていただろう。ところが、自動運転車では、あらかじめクルマをどうプログラムするのが「よい」のか、という「倫理的設定」が問題になるわけである。これが、何とも厄介な課題を私たちに突きつけてくるのだ。

◆ 4.自動運転車が自己学習(self-learning)したら？

自動運転車については、もう一つ、あるいは根本的な難問が控えているように思わ

れる。それは、自動運転車の「自律化」の問題である。自動運転車に人工知能が搭載されているならば、この問いは避けられないのではないだろうか。これはもちろん、人工知能の自律化が技術的にどこまで可能か、という問題にかかわってくるが、議論は技術的な次元にとどまらない。

従来、自動運転車については、倫理的問題が論じられるとき、あらかじめ人間がどうプログラムすべきか、という観点で解明されてきた。しかし、人工知能の可能性を考えると、もしかしたら人間の範囲を超えることになるのかもしれない。

人間が自動運転車の動きをひとつひとつ決定するというのではなく、自動運転車自身がさまざまな状況から「自己学習」して、人間の想定とは異なる答えを出すようになるのではないだろうか。最初の基本設定は人間が行なうとしても、そこから後は、自動運転車自身が最適解を出すかもしれない。

囲碁や将棋の人工知能と同じように考えていいのか留意が必要ではあるが、可能性としては、人間のドライバーがそのつど瞬間的に判断するように、自動運転車もみずから判断するようになるのではないだろうか。そうなったとき、何が起きるのだろうか。この問題を考えるために、「お迎え問題」という思考実験を作ってみよう。

お迎え問題：車の所有者を迎えに行くために、自動運転車が無人で動いている。そのとき、突然子どもが一人道路にとび出してきた。ブレーキでは間に合わないので、そのまま子どもを轢いてしまう。それを回避するには、進路を変える必要があるが、左にはコンクリートの壁があり、右は崖になって海へ墜落する。このとき、自動運転車はどうすればいいだろうか。

人間の常識的な発想では、自動運転車は直進して子どもを轢くよりは、壁に激突するか、崖から落ちてしまう方が望ましいだろう。「人間第一（ヒューマンズム）」の観点からは、それ以外には考えられない。しかし、人工知能を搭載した自動運転車が、そうした人間第一主義を選ぶかどうかは、疑わしいだろう。「自己保存」の原理にもとづけば、車にとって損傷の軽微な方法（子どもを轢く）を選ばないとも限らない。

人工知能が「自己学習」すれば、あらかじめ人間がプログラムするにとどまらず、やがては人工知能そのものがみずから判断し、自分（人工知能）にとって最もよい方法を選択する可能性も否定できないだろう。このとき、私たちは、人工知能にどうかかわればいいのか。

自動運転車は今まさに始まろうとしているところであり、自動運転車の自律化がすぐさまやってくるわけではない。しかし、自律化どころか、その手前の段階でさえも、解決すべき問題は数多く残っている。手遅れになる前に、準備しておかなくてはならない。

*参考文献

本稿では以下のものを参考にしたが、事例については筆者がアレンジして作成している。

Patrick Lin, Ryan Jenkins, & Keith Abney (ed.); *Robot Ethics 2.0 From Autonomous Cars to Artificial Intelligence*, Oxford, 2017.岡本裕一朗『人工知能に哲学を教えたら』(SB新書、2018年) 岡本裕一朗『答えのない世界に立ち向かう哲学講座』(早川書房、2018年)。

プロフィール.....

おかもと・ゆういちろう 玉川大学文学部客員教授
1954年生まれ。九州大学大学院文学研究科修了。博士(文学)。九州大学文学部助手を経て玉川大学教授。2019年4月より現職。西洋の近現代思想を専門とするが、興味関心は幅広く、人工知能や遺伝子工学など、現代社会のさまざまな問題にも領域横断的な研究をしている。参考文献で挙げた著書のほか、『フランス現代思想史』(中公新書)、『12歳からの現代思想』(ちくま新書)、『ヘーゲルと現代思想の臨界』(ナカニシヤ出版) 他多数。