

安全への提言



人工智能に求められる安全問題

いた がき はる ひこ
板 垣 晴 彦†

自動車の自動運転が社会の話題となっている。これに関連して「トロッコ問題」と呼ばれる思考実験の話聞いた。何でも、哲学者や心理学者が道徳や倫理について考える時に例題としているようだ。

その内容は、線路を無人のトロッコが暴走している。その先ではトロッコの暴走を知らずに保線作業をしている作業員5名がいる。このままでは5名の死亡事故が起きるという状況について、次の問いが出される。

第1の問いは、その線路には分岐があり、引き込み線に進路を変えることができるとする。ただし、引き込み線でも作業員1名が保線作業をしている。「あなたは分岐の前にいます。レバーを切換えて進路を変えますか？」

この問いに対して、国籍や年齢に関係なく、ほぼすべての人は「はい」と回答するという。1名も死亡しないことが最善ではあるが、それが不可能であるとするならば、被害を少しでも小さくしたい、ということだろう。

第2の問いは、線路をまたぐ橋の上に巨漢の男がいる。もし、この巨漢の男が線路上に落ちれば、トロッコは巨漢の男に衝突し巨漢の男は死亡するがトロッコは止まり、その先にいる5名の作業員は助かる。あなた自身はやせているので、線路上に落ちてでもトロッコは止まらない。「あなたは巨漢の男の後ろにいます。この男の肩を押して突き落としますか？」

この問いに対して、ほとんどの人は「いいえ」と回答するという。作業員5名の命が助かり、犠牲者の人数は第1の問いと同じ1名であるけれども、危機的状況にはない1名を自分自身の手で犠牲にしてまでも感じるのだろう。「いいえ」なのだから作業員5名が死亡するということになってみても、行動をしたことによる危害は、行動をしなかったことによる危害よりも重く感じるという心理だそうだ。

トロッコを車の自動運転に置き換えれば、新しい

「トロッコ問題」ができあがる。

第1の問い。完全自動運転のワゴン車が片側1車線の道路を走行していたところ、渋滞している反対車線の車両の間から、自転車が飛び出してきた。ワゴン車の人工智能は、ハンドルを左に切って自転車をよける選択肢を考えた。しかし、左側の歩道には、ちょうど小学生が5名歩いていて、よけるとひいてしまう。人工智能はハンドルを切るかどうかを計算し、急ブレーキのみをかけた。しかし、減速が間にあわず、自転車に衝突して、自転車の男は死亡した。

第2の問い。ワゴン車が同様に走行中、対向車線からダンプカーがはみ出してきた。そのまま直進するとダンプカーと衝突し、乗員1名の命が危うい。左によければ歩道の小学生5名をはねてしまう。ワゴン車の人工智能は、ハンドルを切るのか？

第2の問いに対して、少なからぬ人が「5名の小学生をはねてしまう結果になるのならば、人工智能にもハンドルを切らないという『道徳回路』が必要である。ワゴン車の乗員が死ぬのはやむを得ないこと。」と思うだろう。アメリカの大学で実際にアンケート調査をしたところ、「乗員自身が犠牲になることも含めて、死者数を最小限にするように人工智能をプログラムすべきである。」と考える人が多かったという。

どちらを選択するか？という問いをするからパラドックスになる。実現させなくてはいけなのは、乗員1名と小学生5名の全員が助かる選択肢を作ることである。人工智能に自動運転の安全を担わせるのならば、他車との協調運転など高度な対策を完成させ、危機的状況に陥る前に、その危険を回避するという選択肢がある。リスクの捉え方が技術により変わる例だろう。

そうでなければ、道徳回路があると知った時、「あなたは、この自動運転の車を買わないのではありませんか？」

† (独)労働安全衛生総合研究所 化学安全研究グループ：
〒204-0024 東京都清瀬市梅園 1-4-6