

# AIの利活用における刑法上の諸問題（1）

——利用者と製造者の刑事責任を中心に——

日 原 拓 哉\*

## 目 次

はじめに

第1章 AI概念の明確化

第1節 AIの歴史

第2節 AIの定義への試み

第3節 強いAIと弱いAI

第4節 汎用型AIと特化型AI

第5節 AIと学習

第6節 AIの利活用と刑法上の問題 (以上、本号)

第2章 AI製品の利活用における刑法上の諸問題——生命・身体侵害

第1節 問題の所在

第2節 将来的な技術水準のAI製品における具体的検討

第3節 現状の技術水準のAI製品における具体的検討

第4節 AI製品による生命・身体侵害に対する刑法上の一般的考察

第5節 小 括

第3章 さらなるAIの利活用における刑法上の諸問題——財産侵害

第1節 問題の所在

第2節 経 済 犯 罪

第3節 コンピュータ犯罪——行為客体としてのAI

第4節 小 括

第4章 AI製品開発に対する将来的な刑法上の規制

第1節 問題の所在——強いAIとその現状

第2節 規制的措施

第3節 刑法上の保護

第4節 小 括

おわりに

---

\* ひらは・たくや 立命館大学大学院法学研究科博士課程後期課程

## はじめに

深層学習がその火付け役となり、21世紀初頭から現在に至るまで継続する第3次 AI ブームも佳境に差し掛かる。近時、自動運転車、産業用ロボット、介護用ロボットなど、社会生活上においてますます AI・ロボットが浸透してきている。日常生活を円滑にするために、これらに判断を委ねることが徐々に増えつつある。例えば、自動運転車、家事や手術室、軍が運営する監視室で用いられるロボットだけでなく、車の駐車や予定のリマインダーを設定することなどが挙げられよう。しかし、それに伴う人的損害も存在し、例えば、股関節手術にミリングロボットを使用したところ神経や筋肉に損傷を与えた事例<sup>1)</sup>、車線維持支援システムが原因で、母子が事故死した事例<sup>2)</sup>、バウナタールにある Volkswagen 工場で作業員がロボットによって死亡した事例<sup>3)</sup>がある。さらに、将来的に想定されうるものとして市場経済を脅かす事例も考えられる<sup>4)</sup>。具体的には、AI・アルゴリズムを利用した投資によって相場操縦に該当する取引が行われる事例や未公開情報をデータベースに記録し、AI・アルゴリズムにインサイダー取引をさせる事例等が考慮される。

---

1) *Caetano da Rosa*, Robotoc - Zukunftsvisionen und Risiken robotisierter Spitzentechnik AI im Operationssaal, Technikgeschichte 74 (2007), S. 291 ff.

2) 部分的自動運転の乗用車のドライバーが脳卒中になり、ハンドルを握れなくなった事例である。車線維持支援システムが車両を安定させ、意識を失ったドライバーの乗った車両を高速で町中に誘導し、そこで衝突した。システムの介入がなければ、車両はすでに町の入り口で停止していたとされる。*Hilgendorf*, Automatisiertes Fahren und Strafrecht - „der Aschaffener Fall“, DRiZ 2018, S. 66 を参照のこと。

3) FAZ v. 01.07.2015 (Roboter tötet Arbeiter bei VW in Baunatal). Volkswagen 社の工場勤務する22歳の男性が、電気モーターの新規生産ラインでロボットのセットアップに追われていたところ、ロボットにつかまり、金属板に押し付けられ、胸部に重度の打撲傷を負った。一命をとりとめたものの、その後、病院で亡くなったという。

4) アルゴリズム・AI の利用を巡る法律問題研究会「投資判断におけるアルゴリズム・AI の利用と法的責任」金融研究38巻2号（2019年）を参照。

AIの利活用をめぐつては、とりわけ、自動運転車の関与する交通事故事例が注目されてきた。例えば、Tesla社製の自動運転車(レベル2)の「オートパイロット」を搭載した自動運転車を運転中、眠っていた疑いがあるとして過失致死罪で訴追した事例<sup>5)</sup>や、Uber社製の自動運転車(レベル4)の公道実験中の事故(2018年)に関し、2020年8月27日、同車に乗車していたドライバーが過失致死罪で起訴された事例<sup>6)</sup>が挙げられる。後者においては、同人の裁判について2021年8月10日(現地時間)に第一審が開廷される予定であったが延期されている<sup>7)</sup>。このように、AIの利活用に関する利用者もしくは製造者の刑法上の問題は、主に「自動運転車」の事故事例を引き合いに、2017年ごろより検討されてきたが<sup>8)</sup>、裁判例な

---

5) *Hawkins*, (2020, September 18). Tesla owner in Canada charged with 'sleeping' while driving over 90 mph. THE VERGE. <https://www.theverge.com/2020/9/18/21445168/tesla-driver-sleeping-police-charged-canada-autopilot> (Data Accessed: May 31, 2022)

6) Uber's self-driving operator charged over fatal crash. (2020, September 16). BBC NEWS. <https://www.bbc.com/news/technology-54175359> (Data Accessed: May 31, 2022)

7) *Stern*, (2021, May 12). Trial Delayed for Backup Driver in Fatal Crash of Uber Autonomous Vehicle. PHOENIX New Times. <https://www.phoenixnewtimes.com/news/uber-crash-arizona-vasquez-herzberg-trial-negligent-homicide-charge-11553424>. (Data Accessed: July 5, 2021)

8) ドイツにおいては、いわゆる RobotRecht チームによる研究が知られている。*Hilgendorf/Beck* (Hrsg.), *Jenseits von der Maschine, Robotik und Recht 1, Nomos*, 2012を皮切りに、主にAIの利活用に関する複数領域(例えば、自動運転車の民事法問題・刑法上の問題やジレンマ問題、データ保護、AIの権利主体性など)の問題を扱う研究書が多数刊行されている。2022年3月現在では、その研究書も23巻まで刊行されており依然としてその研究は盛んである。

また日本でも、それに続く形で自動運転車をめぐる法的課題を中心に多数の論稿が上梓されている。ここではその全てをあげることはできないが、以下の文献が挙げられる。深町晋也「ロボット・AIと刑事責任」弥永真夫・穴戸常寿編『ロボット・AIと法』(有斐閣, 2018年) 209頁以下, 根津洗希「ロボット・AIに対して『刑罰』を科すことは可能か」法学新報(中央大) 125巻11号(2019年) 475頁以下, 川口浩一「ロボットの刑事責任2.0」刑事法ジャーナル57号(2018年) 4頁以下, 今井猛嘉「自動車の自動運転と刑事実体法——その序論的考察」『西田典之先生献呈論文集』(有斐閣, 2017年) 519頁以下, 佐久間修「AIと刑法・序説」名古屋学院大学論集社会科学編55巻1号(2018年) 107頁以下, 遠藤聡太「人工知能(AI)搭載機器の安全性確保義務と社会的便益の考慮」法律ノ

どの実務的見解が存在しないまま議論されてきた。

しかし、Tesla 社製のレベル2自動運転車（Model X）による東名高速道路での死傷事故（横浜地判令和2年3月30日 判例秘書 LLI/DB L07550489）のように、実際に AI を利用した製品に起因する判例が実際に現れるようになった。そのため現在は、既存の議論と実務的見解との妥当性について再検討すべきものと思われる。

自動運転車を基調とした既存の議論では、主に SAE 標準でのレベル1～2/3/4～5に段階分けをした検討がなされてきた。そこでの論調として、レベル1ないしは2においては利用者も製造者もレベル0（普通自動車）と同一の注意義務が課されるものとされ、レベル3では、当該自動車の利用者たる「運転者」は、普通自動車のそれに加えて、オートパイロット時におけるオーバーライドに対応できるように運転する義務（道交法71条の4の2参照）や点検義務・整備義務（「自動運行装置」の利用者・運行供用者の義務）が加わる。また、製造者に関しては明確な言及はなされていない

---

㍷時報91巻4号（2019年）19頁以下、今井猛嘉「AI時代の刑事司法」罪と罰222号（2019年）、坂下陽輔「人工知能の開発・利用における過失——自動運転車と過失を題材に」法律時報91巻4号（2019年）13頁以下、笹倉宏紀「人工知能の法規制における行政手続と刑事手続」法律時報91巻4号（2019年）41頁以下、石井徹哉「AIに関する刑法上の課題」罪と罰222号（2019年）5頁以下、稲谷龍彦「人工知能搭載機器に関する新たな刑事法規制について」法律時報91巻4号（2019年）54頁以下、稲谷龍彦「ロボット事故の刑事責任」日本ロボット学会誌38巻1号（2020年）37頁以下。

またドイツ法との比較を扱ったものとして、根津洗希「ロボットの処罰可能性を巡る議論の現状について」比較法雑誌51巻2号（2018年）145頁以下、伊藤嘉亮「エリック・ヒルゲンドルフ『ロボットは有責に行為することができるか？ 規範的な基本語彙の機械への転用可能性について』（文献紹介『ロボットと法』シリーズの論文紹介(1)）」千葉大学法学論集31巻2号（2016年）136頁以下、田村翔「サシャ・ツィーマン『機械の本性とは何であったか？ 機械刑法をめぐる議論について』（文献紹介『ロボットと法』シリーズの論文紹介(2)）」千葉大学法学論集31巻3・4号（2017年）87頁以下、根津洗希「スザンネ・ベック『インテリジェント・エージェントと刑法 過失、答責分配、電子的人格』（文献紹介『ロボットと法』シリーズの論文紹介(2)）」千葉大学法学論集31巻3・4号（2017年）105頁以下。米国の議論を扱ったものとして、松尾剛行「自動運転車と刑事責任に関する考察 ロボット法を見据えて」早稲田大学大学院法務研究科臨床法学研究会 Law and practice 11巻（2017年）73頁以下など。

いものの、普通自動車のそれと同一と結論づけるものが多い。レベル4ないしは5では、利用者については道交法上の「運転者」概念から外れるため、一見すると交通事故事例において利用者の道交法上の義務やそれにに基づく過失責任を問うことはできないように見える<sup>9)</sup>。それに対して製造者に対しては、普通自動車のそれと同一のスキームが妥当するという見解もある<sup>10)</sup>。

AI（ここでは自動運転車）の利用によって人に死亡結果や傷害結果が生じた場合、AIをめぐる主体——利用者、販売者、開発製造者、プログラマー——にどのように刑事責任が帰属されるべきであるのかが問題となる。この点につき、従前の製造物による法益侵害結果惹起との注意すべき相違は、例えばAIを搭載した自動運転車（レベル4以上）が死傷事故を起こした場合、その事故原因にかかる機序がブラックボックス化しうることにある。このことは、責任所在の証明を困難にする可能性がある。すなわち、製造者ないしは利用者が当該義務を履行していれば結果が発生しなかったということがいえず、客観的に当該結果が帰属される主体が存在しない可能性、すなわち「帰属の間隙」が生じるおそれがある。もっとも、そのようなリスクをはらむことをあらかじめ上記主体に加えて社会全体が許容しており、それがAIを搭載した自動運転車を利用することに対して通常有すべきリスクであるとされるならばこのような問題は生じない。しかし、そもそも自動運転車（SAE基準でのレベル4以上）は実用化されておらず、現状はそのような状況にあるかは不明である。もっとも、その自動運転車が普通自動車の有するリスク、すなわち普通自動車による交通事故件数よりも少なくなるというのであれば、自動運転車の有するリスクを社

---

9) もっとも、2022年道路交通法改正により、いわゆるレベル4の自動運転車の利用者は「特定自動運行実施者」と定義され、その者に対する義務も制度化された。この改正については、樋笠堯士「自動運転レベル4における刑事実務——道路交通法改正案の分析と提案——」捜査研究858号（2022年）25頁以下が詳しい。

10) 例えば、山下祐樹「AI・ロボットによる事故の責任の所在について」ノモス45巻（2019年）108頁。

会が許容しうる可能性は指摘される<sup>11)</sup>。むしろ、このことは現状では不明確なので、なおのこと現行法の解釈によって、誰にどのように答責を帰属させるのか、ないしは、帰属させることはできないのかを仔細に検討する必要がある。

考えられる帰属のモデルとして、例えばいわゆる「連帯責任」のモデルがある。これは、当該結果の発生に関与した主体に結果を分散的に帰属させる着想であり、このような分散的な結果帰属により、死傷結果に対する「帰属の間隙」を回避することはできるといえる。しかし、このモデルでは、とりわけ開発製造者側にとって、将来的な発生結果が自己に帰属されるリスクを背負ってまで製品を開発することを意味する。それは、製造開発者のモチベーションという点からみて、開発を阻害する要因となり、結果として科学技術の進展を阻むことにもなりうる。利用者側についても、果たして、自己や第三者に生じた損害につき、過失責任が帰属されるリスクを背負ってまでAIを利活用するのか、という疑問が残る。それゆえ、損害結果の発生につき、分散的に答責を帰属させるという構想に立つ場合<sup>12)</sup>、却って市民社会における市民の平穏な生活を法が脅かすおそれがあり、科学の発展を阻害するおそれがあるのではないかという懸念も存在する<sup>13)</sup>。

解決手法の一つとして、AIの利活用における刑法上の答責帰属を議論する文脈で、AIそのものに刑罰を科すことができれば、このような問題は生じないとする見解もある<sup>14)</sup>。しかし、AIそのものに刑罰を科すとす

11) 松宮孝明「自動運転をめぐる刑事法的諸問題」立命館法学395号（2021年）4頁以下。

12) これを問題にするのは *Beck, Die Diffusion strafrechtlicher Verantwortlichkeit durch Digitalisierung und Lernende Systeme, ZIS 03/2020, S. 41 ff.* など。

13) 例えば、経済産業省「GOVERNANCE INNOVATION Ver. 2: アジャイル・ガバナンスのデザインと実装に向けて」（2021年）66頁（<https://www.meti.go.jp/press/2020/07/20200713001/20200713001-1.pdf> : 2022/05/29閲覧）など。

14) 松宮孝明「自動運転と法」学術の動向25巻5号（2020年）59頁以下、松宮・前掲（注11）では以下に示す論点の検討を簡単に批評する。ただし、それを可能にしようとする、伝統的な刑罰理論にどこまで適合するのかを以下のプロセスで検討する必要性があるとされる。すなわち、① AIを人間と同レベルの権利主体として見なしてよいのか、➤

る考え方については、因果的に結果を惹起せしめた自然人を特定する努力を放棄することにつながるのではないかという懸念が指摘される<sup>15)</sup>。そのため、この見解を否定しつつ、AI の利活用に関与する人間の主体のうち、誰に当該結果について帰属されるのかを検討する必要がある。すなわち大別して、製造者と、管理者を含む利用者それぞれの立場に応じ、従前の製造物とは異なり、予測不可能な判断をなしうる AI 製品を利用し、これによって生命や身体といった法益が侵害された場合に、どのようにして刑法上の解決を図るべきかを検討する必要がある。

この背景には、AI 製品によって利用者ないしは第三者の生命・身体が侵害された場合、製造者の刑法上の責任（刑事製造物責任）の前提となる、「必要な注意義務」の内容は、AI を搭載していない製品のそれと同一であるか否かの検討を改めて行うべきではないかという問題意識がある。そのため、改めて製造者に課せられる注意義務の内容を確定させる必要がある。

また、少なくとも AI による動作が介在する以上、その利用者には、それが搭載されていない製品以上の注意が課されるのではないか、すなわち利用者は AI 製品をどの程度まで「信頼」することが許されるかということを検討しなければならない。例えば、Tesla 社製自動運転車の事故事例においては、異常な走行動作の原因が、センサーの欠陥ないしは限界なのか判然としていない。そうすると、このような事故が発生しないように、利用者に対しては当該「AI の誤動作に対応できるように利用する」という新たな義務が課されることになろう。もっとも、このことは AI 製品の利用促進を阻害する要因にもなりうるのではないかという疑念、およ

---

② AI を処罰することに意義はあるか、③ AI の処罰として考えられるものは「刑罰」であるのかというものである。なお、中国における AI と刑事責任の議論については、劉憲権（孫文訳、松宮孝明監訳）「人工知能時代における刑事責任の変遷」立命館法学396号（2021年）467頁を参照。そこでは AI に対する刑罰が肯定的に描かれている（486頁以下）。

15) この指摘は、山下・前掲（注10）106頁にもみられる。

びこれは利用者のみならずその背後に存在する保有者・運行供与者にも及ぶのではないかという問題もある。

以上のような問題意識のもと、本稿では AI 製品に由来する事故に関する刑事責任の検討を行う。そのような事例として、先に取り上げてきた自動運転車の事例のように、人間の生命・身体を侵害するものが想定されるが、そればかりでなく、近時その利活用が注目される投資アルゴリズム AI やソフトウェア・エージェントのような AI 製品は人間の財産を侵害することがありうる。具体的には、実体を持たないプログラムベースの AI の利活用によって刑法上の結果をもたらす事例が存在する。例えば、AI・アルゴリズムを利用した投資システムを利用したところ、利用者の認識なく、AI の判断により証券犯罪としての相場操縦取引やインサイダー取引が結果として行われてしまう事例や、さらに、ネットワーク化された経営判断に供される AI が協働してカルテルを締結する事例などについて、これらの経済犯罪には過失を処罰する規定が存在しないので、利用者の可罰性を肯定することは難しい。なぜなら、そもそも、不公正取引行為の過失処罰規定が存在しない以上、不公正取引をしているという認識のない利用者処罰を考慮すること自体に疑念があるからである。これらの問題については、自動運転車等の製品における解釈とは別に、立法上の提言も含めて検討を行う必要がある。

最後に、本検討ではあくまで現状の技術水準における AI 製品の利活用に関するものが主となるところ、将来的には AI を搭載した自律的兵器のように、人間社会を脅かすような利用形態に対する危惧も示されている。このような AI 利用に関する研究を事前に抑制する刑法上の規制の可能性をも考察すべきであろう。

その際、まずは AI に関連する用語法の定義及び実例を確認し、AI に関する概念の明確化および本論文で想定する AI を示す（第 1 章）。さらに、現状の AI 製品および、将来的な AI 製品を基調とした刑法上の問題の検討を、その解決手法に関する議論を敷衍しつつ、最終的には利用者の

行為についての検討、そして製造者の行為についての検討を刑事製造物責任の文脈に即して検討を行う(第2章)。さらに、人間の財産を侵害する特別な問題として、AIを通じた経済犯、そしてAIが行為客体となるコンピュータ犯罪についての問題を検討し(第3章)、最後に今後のAI開発における規制を刑法上の観点から検討するものとする(第4章)。

## 第1章 AI概念の明確化

AI製品の利活用をめぐる刑法上の諸問題に関する先行研究においては、AIの深層学習(ディープラーニング)が因果関係の機序の不明確化、それに伴う予見可能性認定の困難性をもたらすとされるが、その前提であるAI概念や深層学習の定義理解が論者によって異なり、その上で予見可能性認定の検討を進められることが多いようである<sup>16)</sup>。この点、概念を明確にしなければ正確に検討を進めることができない。このような問題意識のもと、本論文での検討にあたって、まずはAIやその周辺概念である機械学習や深層学習などの概念をAI研究の歴史から遡る形で明らかにし、その上で想定すべきAIを指定する。

---

16) この指摘は、石井徹哉「AIに関する刑法上の課題」罪と罰222号(2019年)6頁にもある。例えば、今井・前掲(注8)罪と罰222号(2019年)25頁はAIを「データ処理を超高度化させる」ものだとし、稲谷・前掲(注8)日本ロボット学会誌38巻1号(2020年)37頁以下は深層学習を「大量のデータを統計的に処理することで一定の法則性を見出し、自らそれに従って振舞う」ものとする(AIの定義はない)。また、例えば、佐久間修「AIの刑事責任——否定説の観点から」刑法雑誌59巻2号(2020年)159頁以下では、AI研究の展開を示しつつ、AIを「明文化しやすい行動原則にもとづくルールベースのAI」と、「統計・確率型のAIであって、『隠れマルコフモデル』とか『ベイジアンネットワーク』などと呼ばれる」AIと、「人間の脳機能を参考にしたニューラルネットワークであり、ディープラーニングによる各種のパターン認識をおこなう」AIの組み合わせであるとするが、深層学習については「多層的ニューラルネットワークによる情報処理が可能」なものであるとして、その効果のみが説明されるにすぎない。このように、論者によってAIや「深層学習」に対する理解には濃淡があることがうかがえる。

## 第1節 AIの歴史

人工知能という言葉は、1955年8月31日に開催されたダートマス会議における、*John McCarthy, Marvin Minsky, Nathan Rochester, Claude Shannon*, “A proposal for the Dartmouth Summer Research Project on Artificial Intelligence” という宣言書の中で初めて姿を表した<sup>17)</sup>。当時ダートマス大学の数学の助教授であった McCarthy はこの宣言書で、「言語を使い、抽象化し、構想を展開し、目下人間のみにしかなしうることのできない種類の問題を解決し、自己改善を続けることができる機械をいかに作ることができるか<sup>18)</sup>」という目標を掲げた。ここには、「推論」と「探索」を基軸に置く AI のみならず、ニューラルネットワークや人間の言葉のコンピュータ処理することも含まれていた。

この McCarthy らの宣言は、当時の AI 領域の研究において時代を先行していたという事実のみならず、ダートマス会議において AI が楽観的に理解されていたことも窺える。このようないわばユートピア的熱狂に駆り立てられる形で、McCarthy らは、専門家集団が数ヶ月間集中的に取り組めば、個々の研究分野での飛躍的進歩が達成されるとも述べた<sup>19)</sup>。もっとも、このダートマス会議において飛躍的進歩をもたらしたというわけではないとされる<sup>20)</sup>。とはいえ、人工知能という言葉が初めて定着し、それ以降、当時の研究をはるかに超えたセンセーションを巻き起こしたことは否定し得ない。これが1960年代に始まる第1次 AI ブームである<sup>21)</sup>。

ダートマス会議後も AI 研究はさらに進められた。例えば、Arthur Samuel は、常に自分自身と対戦することで、徐々にそのゲーム戦略上

17) *Russel/Norvig, Artificial Intelligence -A modern approach-*, 4th ed. 2021, p. 18.

18) *McCarthy, Minsky, Rochester, Shannon, A proposal for the Dartmouth Summer Research Project on Artificial Intelligence* (Aug. 31th 1955), p. 2.

19) *Ibid.*

20) *Russel/Norvig, supra* (fn. 17), p. 18.

21) 松尾豊『人工知能は人間を超えるか——ディープラーニングの先にあるもの』（角川 EPUB 選書、2015年）81頁。

の実力を向上させるようなコンピュータ・チェッカーを開発し、最終的にこのプログラムはチェッカー上級者のレベルに至ったという<sup>22)</sup>。また、Allen Newell と Herbert Simon は、Principia Mathematica における多くの数学上の定理の証明を可能にする記号処理プログラムである Logic Theorist を開発した<sup>23)</sup>。むろん、McCarthy 自身も、自らを改善するプログラムを書くことを可能にするプログラミング言語である LISP の開発者<sup>24)</sup>としては無視できない。しかし、これらの AI 分野の先駆的開発には決定的な問題点があった。それは、どのプログラムもトイ・プロブレムを主に扱っているため、適用領域が非常に限定的であったことである。例えば、米軍の研究機関である DARPA は、冷戦時代にロシア語の文章を英語に自動翻訳するプログラムの開発に着手したものの、その開発は難航し、結果として研究資金が大幅に削減されるということがあった<sup>25)</sup>。

AI が再び脚光を浴びたのは、1980年代であった (第2次 AI ブーム)。第1次 AI ブームとは対照的に、ここでは「知識」を用いたエキスパート・システムがその中心に置かれている。この背景にあるのは、実世界に対応するシステムを開発するにはシステムが現実世界における膨大な知識を有する必要があることが強く認識されたことである<sup>26)</sup>。これに対応する形で、ある専門分野の知識を取り込み、推論を行うことで、まるでその分野のエキスパートであるかのように振舞うプログラムの開発が進められた。その一例として、スタンフォード大学によって開発された MYCIN が挙げられる。これは、伝染病の血液疾患患者に対して質問に回答させる形で診断を行い、感染した細菌を特定し、そのうえで適切な抗生物質を処方す

---

22) See Scheffer, *One Jump Ahead: Challenging Human Spremacry in Checkers*, Springer, 2009, chap. 6.

23) *Russel/Norvig, supra* (fn. 17), p. 18.

24) *Russel/Norvig, supra* (fn. 17), p. 19.

25) *Russel/Norvig, supra* (fn. 17), p. 30.

26) 小林一郎『人工知能の基礎』(サイエンス社, 2008年) 4頁。

るように設計されていた<sup>27)</sup>。しかし、このシステムにも限界があり、莫大な知識を与えることによって、論理的矛盾や一貫性の喪失が生じた。とりわけ、「けだるさ」や「痛い」などの表現については個別に定義付けをしなければならなかった。このように、知識を与えることの困難性が露呈した結果、第2次 AI ブームはその終焉を迎えた。

その後しばらくは AI ブームに裏打ちされた AI 研究は下火であったが、それを覆す決定的な出来事があった。それは、1997年の IBM によるチェスプログラム Deep Blue が、当時のチェス世界チャンピオンであった Garri Kasparov に3.5：2.5で勝利を取めたことである<sup>28)</sup>。このコンピュータ戦で注目すべきは、Kasparov が、コンピュータは悪手を打たないことを信じて序盤戦を対局したことにある<sup>29)</sup>。しかし、第2局の Deep Blue の手で Kasparov は判断を誤り、最終的には Deep Blue が勝利したのである。また自動運転車の開発も重要なステップである。この関連では、先に触れた米国の研究機関 DARPA が再びその役割を担っており、例えば2004年には DARPA グランドチャレンジが初めて開催された。この競技は自動運転車のみを対象とし、最速かつ自律的にレースを完走した車両が勝者となるものである。2005年のネバダ州で開催された全長約212kmのレースでは、Volkswagen 社製の Stanley が時速35kmで、6時間53分58秒で優勝した<sup>30)</sup>。2004年の第1回大会では自律走行車がゴールに到着することすらかなわなかったことに鑑みれば、まさにこの結果はブレイクスルーといえるものであった<sup>31)</sup>。さらに2006年、CMU の自動車である Boss は米軍飛行場で交通規則を遵守しつつ安全に走行することに成

---

27) 松尾・前掲（注21）87頁。

28) *Russel/Norvig, supra* (fn. 17), p. 27.

29) 伊庭齊志『ゲーム AI と深層学習 ニューロ進化と人間性』（オーム社、2018年）5頁以下。

30) *Russel/Norvig, supra* (fn. 17), p. 28.

31) ローレンス・D・バーンズ・クリストファー・シュルガン（児島修 訳）『AUTONOMY 自動運転の開発と未来』（辰巳出版、2020年）49頁。

功し<sup>32)</sup>、そして2009年には初期モデルの Google Self Driving Car がカリフォルニアの高速道路を走行し、2015年には累計160万kmを公道走行するようになった<sup>33)</sup>。もう一つの重要な事例は、IBMが開発した Watson である。この Watson は、2011年に米国のクイズ番組である Jeopardy! のチャンピオン Ken Jennings と Brad Rutter の2人に勝利した<sup>34)</sup>。このクイズ番組の特徴として、問題に正解するだけでなく、異分野の知識を複合させる必要があった。Watson は人間の司会者の質問を正しく理解し、いわば皮肉的な言辞をフィルタリングして、正解を準備し、かつ、人間の対戦相手よりも速くブザーを押すことができたのである<sup>35)</sup>。あるプログラムが人間の相手を理解し、それに反応することが可能になったことは、AIの歴史における一種のマイルストーンと見なされるべきだろう<sup>36)</sup>。しかし、Watson では、問題を読み上げながら処理し、データベースから適切な回答を検索し、それを瞬時に出力するという、ソフトウェアの圧倒的なスピードがその成功の決め手となったことも言うまでもない<sup>37)</sup>。

そしてここ数年続いている深層学習に裏打ちされた第3次 AI ブームは、まだ沈静化したというわけではない。2016年、Google DeepMind は囲碁プログラムである Alpha Go を開発したが<sup>38)</sup>、このソフトウェアは2016年3月の対局で、当時世界チャンピオンの囲碁棋士であった Lee Sedol に勝利した<sup>39)</sup>。その理由は、過去の膨大な対局データのみならず、

---

32) *Russel/Norvig, supra* (fn. 17), p. 28.

33) その詳細は、日経 EXTEC 「Google 社の『Waymo』が自動運転開発に与えるインパクト」(2016年12月26日)を参照されたい。

34) 松尾・前掲(注21)19頁参照。

35) 金山博・武田浩一「Watson: クイズ番組に挑戦する質問応答システム」情報処理52巻7号(2011年)840頁以下。

36) 金山・武田・前掲(注35)842頁。

37) 金山・武田・前掲(注35)849頁。

38) 伊藤毅志・松原仁「AlphaGo の技術と対戦」人工知能31巻3号(2016年)441頁。

39) 日本経済新聞「囲碁 AI、プロに4勝1敗 最終局も熱戦制す」(2016年3月15日)。

ソフトウェアが自らと対戦することで常にその実力を向上させることができたことにあるという<sup>40)</sup>。

このように、AIの歴史を概観すると、AIは自己改善するシステムだけでなく、部分的に人間とのコミュニケーション能力を持つシステムや、人間に新たなソリューションを示すシステムも含まれていることがわかる。しかしその利活用の多様性は、「AI」自体の不明確さをももたらしている。AIの利活用における（刑）法上の問題を議論するにあたっては、検討の対象とすべきAIを明確にするために、まずはその「AI」の定義を明確にしなければならない。そこで次節ではこのことについて一定の指針を与える。

## 第2節 AIの定義への試み

AIの定義は多義的であり、例えば、「人間の思考プロセスと同じような形で動作するプログラム、あるいは人間が知的と感じる情報処理・技術」<sup>41)</sup>（総務省）や、「インテリジェントな機械、とりわけコンピュータプログラムを作るための科学技術。それは人間の知能を理解するためにコンピュータの類似のタスクを用いることと関連付けられるが、生物学的に観察される方法でそれ自体を制限する必要はない」<sup>42)</sup>（McCarthy）などが挙げられる。このように、「AI」概念には統一的定義が存在するというわけではない。むしろ、この概念の統一的定義を確立しようとする自体非常に困難ともいえよう。これは、ダートマス会議でMcCarthyが紹介した“Artificial intelligence”という概念の解釈が不正確だったことが原因とされる<sup>43)</sup>。そこで様々なアプローチが試みられた。

---

40) 松原仁「コンピュータ囲碁の進歩」日本ロボット学会誌35巻3号（2017年）192頁。

41) 総務省「令和元年度版 情報通信白書 第1部 第3節 2. AIに関する動向（1）」（2019年）。

42) McCarthy, WHAT IS ARTIFICIAL INTELLIGENCE?, Computer Science Department, Stanford University, 2003, p. 2.

43) Herberger, „Künstliche Intelligenz“ und Recht, NJW 2018, S. 2825 (2826).

まず、AI を人間の知能と対極の存在として定義する試みがなされてきた。これによると、人間の知的なものとして見なされる能力それ自体を発揮するシステムを人工知能と呼ぶことになる<sup>44)</sup>。しかし、人間の態度がどこから知的であるかを判断する統一された基準がないため、この定義はあまり意味を為さないようにも思える。とはいえ、概念定義が存在しないとすると、どのようなものを知的と見なすのかについて個別に異なる基準を設定することが重要となる。例えば、数学の素人にとっては、どんなに複雑な算数の問題でも頭の中ですぐに解けるような人のことを「知的である」というかもしれない。しかし、そのような計算能力だけで、あるシステムを AI システムとして理解するのであれば、それは否定されなければならない。なぜなら、この定義では単なる電卓でも AI という名称を与えなければならないこととなるが、それは間違いなく AI ではなからう。

これに対して、情報学者の Elaine Rich は、「現時点で人間が得意とすることをシステムが行うことができれば、人工知能と呼ぶことができる」という点に着目して定義を試みる<sup>45)</sup>。高速な計算能力を持つシステムは、その点において我々人間よりはるかに優れており、逆にこの分野で優れているのは人間ではないため、一見すると、この定義は当を得たものであり、さらに、我々人間が多彩なシステムより優れている分野は未だ存在しているので、この定義は当面の間は有効なものともいえる。もっとも、この定義を別の側面から見ると完全なものではない。例えば Alpha Go は、自分自身と対戦することで常に自らの実力を改善し、人間に勝利したが、この定義では、現在ではすでに人間より優れているという理由で、この囲碁システムは AI であるとされなければならない。しかし、自主学习によって達成したシステムなので、どのようにして思考過程を形成したのかも考えなければならず、この点が従来のシステムとは大きく異なる点である。

---

44) *Kaplan*, *Artificial Intelligence: What Everyone Needs to Know*, Oxford, 2016, p. 7.

45) *Rich*, *Artificial Intelligence*, McGraw-Hill, New York, 1983, p. 3.

そのため、Rich の定義が必ずしも AI システムを正確に分類できるわけではない。

さらなる AI の定義の試みとしては、数学者 Alan Turing が提唱したアプローチであるチューリング・テストがある。システムがインテリジェントであると言うために、彼はその定義ではなく、合格しなければならないテストを試みたのである。彼が模倣ゲームと呼ぶこのテストでは、被験者はコンピュータを用い、相手の声を聞くことも見ることもなく、もう 1 人の被験者及び機械とコミュニケーションをとる。そこで被験者は、いくつかの問題を投げかけることによって、会話の相手のうち、どちらが人間でどちらが機械なのかを区別する。5 分間の会話時間の後、少なくとも 30% 以上の被験者が、本来は機械である方を人間だと回答すればテストに合格したとみなされ、機械はインテリジェントであるということが出来る<sup>46)</sup>。

このテストの支持者は、チューリング・テストを用いると、AI を複雑かつ不正確な定義に当てはめる必要がないため、極めて容易に AI を特定することが可能であるといった利点があるという<sup>47)</sup>。というのも、これに合格するためには、AI は自然言語を処理することが可能であり、すでに述べられたものを再述するための膨大な知識量と記憶容量を持っているだけでなく、会話パートナーに適応することができなければならないからだとする<sup>48)</sup>。それゆえ、チューリング・テストでは、自然言語の使用と処理、および未知の状況への適応能力が、AI と判定されるための決定的な基準となる<sup>49)</sup>。しかしこのテストは、AI の特定領域のみと関連するものにすぎないし、そもそもこのテストに参加するためには、まず AI がチャット機能で人間相手にコミュニケーションすることが可能でなければ

---

46) *Russell/Norvig, supra* (fn. 17) p. 984.

47) *Lenzen, Künstliche Intelligenz Was sie kann & was uns erwartet?*, C.H. Beck, 2018, S. 25.

48) *Russell/Norvig, supra* (fn. 17) p. 2.

49) *Lohmann, Strafrecht im Zeitalter von künstliche Intelligenz*, Nomos 2021, S. 44.

ならない。そのような機能を持たない AI (例えば Alpha Go) は、テスト対象として最初から除外される。したがって、このテストは、AI の存在を確認するというよりもむしろ、いわゆるチャットボット<sup>50)</sup>の適性判断に相応するものであると言わざるをえない<sup>51)</sup>。また、AI に付随しうる検索機能が、AI としての決定的要素になり得るかどうかとも考慮しなければならない。入力された大規模なデータベースではパターン検索のみが行われる<sup>52)</sup>が、この機能は一般に AI にとっての最低条件と言える。仮に大規模なデータベースがなければ、AI は学習に際して困難を要し、このようなテストがインテリジェントな機械として特定するのに適切なかどうかの問題となる。人間の知性もまた、一つのテスト結果だけでは描ききれないほど多くの側面を持っており、このことは、機械知能のテストにも同様に当てはまる。しかし、チューリング・テストではあくまで自然言語処理と適応能力という二つの機能のみがテストされるため、画像処理などその他の側面は依然として考慮されない。したがって、一つのテストですべての AI を特定できるわけではない。

このように、AI の定義の確立においては、それぞれのアプローチにメリットとデメリットがあるため、普遍的な定義は存在しない。このことは、AI という言葉がいかに政治、社会、メディアで日常的に使われているかを考えれば、その現状と相反するもののように思われる。そのため、定義が困難な AI 概念をより明確なものにするために、次節では、AI の現象形態を見ることにする。

### 第 3 節 強い AI と弱い AI

現在、AI はさまざまな形で利用されている。例えば、産業分野だけで

---

50) チャットボットとは、人間と機械だけが対話に入るが、人間のコミュニケーションを可能にする対話システムまたは支援システムである。

51) *Calo, People Can Be So Fake: A New Dimension to Privacy and Technology Scholarship*, Penn State Law Review 2010, S. 830.

52) *Matthias, Automaten als Träger von Rechten, Logos*, 2008, S. 220.

なく、医療機器<sup>53)</sup>や刑事裁判<sup>54)</sup>でさえも AI の存在が見受けられる。もっとも、AI が突然、産業技術から医療、確定判決に至るまであらゆるものをコントロールできるようになったというような、一見すると圧倒的な力を持つように見えるということには、より詳細な検討が必要である。このような個別の適用方法ごとに、どのような種類の AI が問題となるのかを区別することが重要である。そのような AI の現象形態としては、「弱い AI」と「強い AI」の2種類に分けることができる。この分類は人間の認知能力をコンピュータ上で再現するにあたり考慮すべき AI の分類である<sup>55)</sup>。

「弱い AI」とは、厳密かつ精確な方法で仮説を立て、検証することを可能にする道具にすぎない AI である。例えば、先述した Alpha Go のみならず、Siri（音声認識）、Google 画像検索（画像認識）など、特定分野に限って利用できるものである。

それに対して、「強い AI」とは、正しくプログラムされたコンピュータが認識状態を持ち、そのプログラムが人間の認識を説明する AI である。この AI に基づいて正しくプログラムされたコンピュータは知能となるとされる<sup>56)</sup>。すなわち、心理的説明を検証する道具ではなく、むしろそのプログラム自体が説明となる。

---

53) 例えば、黒色皮膚がんを発見するための医療診断支援に AI を利用される可能性が指摘されている。*Haenssle, u.a., Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists, Annals of Oncology, Vol. 29, Issue 8, 2018, p. 1837* 参照。

54) 米国では Eric L. Loomis に禁固6年の判決が下された。この判決は、「Compas」という AI ソフトに基づいたものである。しかし、この判決の最大の問題は、裁判所も Loomis も、この判断に至ったアルゴリズムを立証できなかったことである。しかし製造者は、その背後にあるアルゴリズムを公表しておらず、取引上の秘密であると説明している。*Liptak (May 1<sup>st</sup>, 2017), Sent to Prison by a Software Program's Secret Algorithms, New York Times* を参照。

55) *Searle, MINDS, BRAINS, AND PROGRAMS, University of California, 1980, p. 2.*

56) *Ibid.*

この強い AI の開発は目下進行中のものである。これは、我が国の研究者の尽力に代表されるものであり、人間に近似したロボットの開発が進められている<sup>57)</sup>。ここでは、外観の類似性<sup>58)</sup>に価値を置くのみならず、ロボットの内面的な価値も重要視されている。そして、人間との機能的な一体化を目指し、感情を認識し、共感を示すものでなければならないとされる<sup>59)</sup>。例えば石黒浩教授は、演劇の観客もしくは講義を受ける学生が、誰が人間で誰が機械なのかを確認するのに困難なほど人間に類似したアンドロイドの開発を行っている。彼によると、子供から始まり、最後は老人に至るまで、完全な自律型ロボットの開発までを行うとする。これらはいずれ、多様な状況で人間と自由にコミュニケーションをとる能力を有し、その結果、外見上だけでなく、その行動の結果として人間と同レベルのものと類似するようになるという<sup>60)</sup>。

このような展開は、強い AI の開発がもはや遠くないことを示すものともいえるが、現在の技術状況を考えた場合「強い AI」は存在しないといわざるをえない。そのため、本稿の議論の対象とすべきは「弱い AI」である。

#### 第4節 汎用型 AI と特化型 AI

汎用型 AI と特化型 AI の分類は、人間のように広範な課題を処理できるか否かに関する AI の分類であり、「強い AI」と「弱い AI」の分類と

---

57) Asada, Towards Artificial Empathy. *International Journal of Social Robotics*, Vol. 7, No. 1, 2015, pp. 19-33.

58) 例えば、浅田研究室が開発した人間そっくりの幼児ロボット Afetto は、その表面が赤ちゃんの肌のような感触だとされる。Asada Research Group「子供アンドロイドの開発」([http://www.er.ams.eng.osaka-u.ac.jp/asadalab/?page\\_id=177](http://www.er.ams.eng.osaka-u.ac.jp/asadalab/?page_id=177)) (最終アクセス2022/05/31)

59) Eberl, *Smarte Maschinen: Wie Kuenstliche Intelligenz unser Leben veraendert*, HANSER, 2016, S. 291.

60) 石黒共生ヒューマンロボットインタラクションプロジェクト (<https://www.jst.go.jp/erato/ishiguro/outline.html>) (最終アクセス2022/05/31) 参照。

は一線を画する。

「汎用型 AI」とは、ある目標や状況からの知識を別の目標や状況に対して汎化する「転移学習」により目的や状況の変化に自ら適応する広範な能力をもつ AI である<sup>61)</sup>。この点について「強い AI」に存在する「認識の有無」は問題としないところに注意しなければならない<sup>62)</sup>。

他方、「特化型 AI」とは特定のタスクを行うように設計されているものであり、その限りにおいて人間以上の能力を発揮するが、それ以外のタスクは解くことができない AI である<sup>63)</sup>。これに当てはまる AI の例としては前述した Alpha Go の他に、ELIZA, Deep Blue, Tay, Alice & Bob が挙げられる。まず、ELIZA とは Joseph Weizenbaum (MIT) による簡単なパターンマッチング技法を使った自然言語処理プログラムのことである。患者役のユーザーが入力するスクリプトに対して、心理療法士を装う DOCTOR が来談者と診断会話をシミュレーションするものであった。次に Tay とは、Microsoft 製のチャットボットで、Twitter 上にそれを公開し、不特定多数のユーザーとのコミュニケーションを通じて言語を習得するものであった。しかし、公開後24時間以内で差別的発言を繰り返すようになり、わずか1日で公開停止となってしまった。さらに、Alice & Bob とは、Facebook 製のチャットボットであり、実験段階で「両者」の会話は通常の用語法では理解できないやりとりを始めたため、公開停止となったことで知られている。

現在の技術水準に照らすと、存在が認められるのは専ら「特化型 AI」である。そのため、本稿の対象とすべきは「特化型 AI」である。

---

61) ゲーツェル・ベン「汎用人工知能概観」人工知能29巻3号（2014年）228頁。なおこの用語は、Gubrud, “Nanotechnology and International Security”, Fifth Foresight Conference on Molecular Nanotechnology (November 1997) が初出であるとされる。

62) 鳥海不二夫「人工知能技術を俯瞰する」立法と調査405号（2018年）5頁。

63) 鳥海・前掲（注62）5頁。

## 第5節 AIと学習

ここでは、AI分野における「学習」概念について俯瞰する。より具体的には、機械学習と深層学習、そして特別な学習形態としてのフリース・ラーニングが挙げられる。

### 第1款 機械学習

機械学習とは、「明示的にプログラミングすることなく、コンピュータに学ぶ能力を与えようとする研究分野」であるとされ、さらにこれには「教師あり学習」、「教師なし学習」、そして「強化学習」という3つの学習手法が存在する。

まず、「教師あり学習」とは、事前に与えられたデータをいわば「例題(先生からの助言)」とみなして、それをガイドに学習を行う手法である。これは、入力されたデータに対して正しい出力を返すことを学習の目標にするものであり、徐々に入力と出力の関係を学習するものである<sup>64)</sup>。例えば、犬の写真10枚を「これは犬である」として見せ、次に猫の写真10枚を「これは猫である」として見せる。その後いずれかの写真を見せて、それが何かを当てる課題が挙げられる。

次に「教師なし学習」とは、学習データに正解を与えない状態で学習させる手法である。その理由は、入力データがあらかじめ構造化されてシステムに提示されるのではなく、システム自身が構造化してパターンを認識しなければならないことにあり、クラスタリングは、このような学習の典型例である。これは、新たな知見を得るため、あるいはそもそもシステムが利用できるようになった(入力)データのパターンを認識するために利用される<sup>65)</sup>。例えば、100人の漫画のキャラクターを見せ、そのキャラクターの類似性に基づいて10グループに分ける課題が挙げられる。

さらに「強化学習」とは、学習データに正解はないが、目的として設定

---

64) 谷口忠大『イラストで学ぶ 人工知能概論 [改訂第2版]』(講談社、2020年)205頁。

65) 松尾・前掲(注21)118頁。

された「報酬」を最大化するための行動を学習する手法のことである。ここでは、適切な目標を設定した後に環境と相互作用し、その結果が正しければ、一種の報酬である強化を獲得することでその経験から直接学習するシステムである<sup>66)</sup>。システムは最終的に行為の結果を記憶し、報酬を受け取ることを目的に行動しなければならない<sup>67)</sup>。そのため、自分の行動経過を分析し、どの行為が報酬もしくは罰となるかをフィルタリングしなければならない<sup>68)</sup>。これを繰り返すことによりシステムの機能は改善されていく<sup>69)</sup>。例えば、迷路を抜けた時にのみゴールに到着したことがわかり、それを褒められるが、途中の経路は教えてもらえないような学習が挙げられる。

## 第2款 深層学習と人工ニューラルネットワーク

伝統的な学習形態とは対照的に、近年登場したAIの学習形態として、人工ニューラルネットワークを用いたいわゆるディープラーニング（深層学習）がある。これは、「概念の階層から、コンピュータは、単純な概念から複雑な概念を構築することにより、複雑な概念を学習することができる。これらの概念がどのように相互に構築されているかを示す図を描くと、その構図は深く、多くの層がある。このため、このアプローチをAI深層学習と呼ぶ」<sup>70)</sup>と説明される。

深層学習が機能するためには、いわゆる人工ニューラルネットワークが必要とされる<sup>71)</sup>。人工ニューラルネットワークは、脊椎動物の脳のような実際の神経回路網を再現しようとするものである。モデルとする人間の

---

66) *Russell/Norvig, supra* (fn. 17), p. 789.

67) *Sutton/Barto, Reinforcement Learning: An Introduction* (Adaptive Computation and Machine Learning series), Bradford Books, 1998, p. 3.

68) *Russell/Norvig, supra* (fn. 17), p. 789.

69) 谷口・前掲（注64）129頁。

70) *Goodfellow, Bengio, Courville, Deep Learning*, The MIT Press, 2016, p. 1.

71) 小林・前掲（注26）138頁。

ニューロンは約 $10^{10}$ 個あり、脳内での情報伝達に中心的な役割を果たす<sup>72)</sup>。細胞体から軸索が伸び、その先端の終末ボタンという部位がその隣の細胞体から出る樹状突起とシナプス結合することにより、1つのニューロンから他のニューロンへと信号を伝達する。このとき、各ニューロンはシナプスと結合された他のニューロンから微弱な電気信号を受け、その総和がある閾値を超えると興奮状態となり、信号はシナプスを通じて他のニューロンに伝達される<sup>73)</sup>。このプロセスが人間の高度な情報処理製造者メカニズムを担う。この機能は、人工ニューラルネットワークに引き継がれ、そこでは、ニューラルネットワークを正確に再現しようとすることなく、抽象的にモデル化する試みがなされている<sup>74)</sup>。例えば、階層型ネットワークはノード(ニューロン)、エッジ(シナプス)、および複数の層で構成され、各層は階層的に接続されており、各層はその上下の層としか接続されていない。入力層は情報を受け取り、その重みに応じてそれを隠れ層に伝達する。

深層学習を可能にするためには、システムはまず、その訓練段階でそのタスクを処理できるようにしなければならない。これらの学習手順は、さまざまな方法で実施することができる。教師あり学習では、システムに入力値を与え、システムが出力値を計算する。そして、これらの出力値が実際に正しい出力値からどれだけ誤差があるかを検査し、人工ニューラルネットワークの重みと接続を調整することで誤差を最小化する。この方法は誤差逆伝播法とも呼ばれており、主にパターン認識の分野で応用されている<sup>75)</sup>。

深層学習方式のもう一つの種類は、いわゆる畳み込みニューラルネットワーク(Convolutional Neural Network)であり、これらは、畳み込み層、

---

72) 小林・前掲(注26)138頁。

73) 小林・前掲(注26)138頁。

74) 谷口・前掲(注64)229頁。

75) 小林・前掲(注26)143頁。

プーリング層、全結合層、出力層で構成されている<sup>76)</sup>。ここでは、入力画像に対してカーネル（重みフィルタである線形行列）を畳み込み処理することで、そのカーネルに対応した特徴マップを得る。次に、特徴マップをプーリングすることで特徴マップのサイズを縮小する。これらを通じて、入力画像の微小な位置ズレや回転などを補正することができる。この畳み込み処理とプーリング処理を繰り返し適用し、特徴マップを抽出する。抽出した特徴マップを全結合層へ入力し、最終的に各クラス確率を出力する<sup>77)</sup>。このネットワークでは、画像認識の各タスクに合わせた出力層を設計することで、画像分類だけでなく物体検出も同時に行うことができる<sup>78)</sup>。

深層学習の手法は、主にパターン認識の領域で用いられており、利用可能なデータ量が多いほど深層学習は有効に機能する。システムに膨大な量のデータが与えられるならば、深層学習方式は誤差に対応するという利点がある。つまり、人工ニューラルネットワークに何らかの異常が生じても、それがシステム全体の崩壊には至らず、結果としてシステム全体の性能が向上するため、個々の異常を補填することができる<sup>79)</sup>。深層学習は、主に自動運転の領域で、周囲の環境認識のために用いられつつある。

### 第3款 フリート・ラーニング

AIの特殊な学習方法として、いわゆるフリート・ラーニングがあげられる。フリート・ラーニングとは自動運転領域に由来する概念であるが、この場合のエンドデバイスは特定の車両である<sup>80)</sup>。AIが機能するためには、大量のデータが必要とされ、訓練段階中にもそのデータが利用可能な

76) 谷口・前掲（注64）233頁。

77) 藤吉弘亘「機械学習の進展による画像認識技術の変遷」計測と制御58巻4号（2019年）、293頁。

78) 藤吉・前掲（注77）294頁。

79) *Lenzen*, aa.O. (fn. 47), S. 63.

80) *Fraunhofer IAIS*, Maschinelles Lernen „on the edge“, 2019, S. 2.

状態になければならない。しかし、このデータを利用するだけでは、過去のデータが再現できるにすぎないという問題を伴う<sup>81)</sup>。そこで、AIによる将来の決定が過去の学習内容に影響されないようにするため、現在の状況をその判断に反映させるよう、定期的にデータを更新することが試みられなければならないが、その可能性の一つが、いわゆるフリート・ラーニングである<sup>82)</sup>。その背景にあるのは、学習と改善を可能にするように、当該システムはクラウドからデータを取得し、そこに新たなデータを常にアップロードするものの、そのようなアップロードが低帯域域などのために技術的に不可能な場合が多いことについて問題視されていることである<sup>83)</sup>。そこで、フリート・ラーニングという手法により、一定のネットワークのもと、データをクラウドに転送する必要なく、エンドデバイスそれぞれ自身で学習させることを可能にする<sup>84)</sup>。具体的には、あるデバイスでローカルに学習したモデルを集めて一つのモデルに統合し、最終的に再びすべてのデバイスに同期して共有するので、車両の集団が他の車両のデータまたは経験から利益を得ることができる。この学習プロセスにより、個々のシステムの継続的な自己改善が達成されるだけでなく、集団内の車両がすでにデータを収集し、伝達されている範囲では、少なくともデータの最新性も保証される。これは、それぞれのシステムが得た知識を交換し、協働するという形で、既知の問題を解決することを目的としている。ここでは、全体の問題をアルゴリズムで分解し、個々のシステムが個々の小さな問題を解決し、発見されたソリューションによって最後に再び全体の問題を解決する。しかし、フリート・ラーニングの場合、各車両が自らの下位問題を解決し、これらソリューションが最終的に道路上で安全に動作することを達成するために一括して集計されるわけではなく、個々の車

---

81) *Lenzen*, a.a.O. (fn. 47) S. 61.

82) *Lohmann*, a.a.O. (fn. 49), S. 51.

83) *Fraunhofer IAIS*, a.a.O. (fn. 80) S. 1.

84) *Fraunhofer IAIS*, a.a.O. (fn. 80) S. 1.

両が他の車両の学習記録から全体としての利益を得る<sup>85)</sup>。

なお近年では同様の用語として、いわゆる Car-to-Car communication（車両間の通信）や Car-to-X communication（他のシステムとの通信）もある<sup>86)</sup>。この通信形態では、ネットワーク接続された車両が、他のネットワーク接続された車両から情報を受信する<sup>87)</sup>。この車間通信は、WLANと同様の無線技術で動作する<sup>88)</sup>。この新種のネットワークをつうじて、道路交通はより強固で安全なものになるとされる<sup>89)</sup>。これは、緊急車両の接近、渋滞の解消、逆走運転に関する情報、路面凍結などの道路状況など、事故データから経路に関連する情報を、運転手自身がこれらを目視する前に受信し、反応できるようになるからである。さらにこの技術は、いわゆる Road Side Unit、すなわち、信号や道路工事、現在の交通渋滞に関する情報を提供するインフラ内のデジタル機器との通信に使用することができる<sup>90)</sup>。この新しい形式のコミュニケーションによって、道路交通が改善され、交通安全が向上し、運転がより快適になることが期待される<sup>91)</sup>。

このような車両間のコミュニケーションは、AIの一領域なのか、それとも単なる情報伝達の一形態なのかが問題となるが、正しいのは後者であ

85) もっとも、収集したデータには利用者の位置情報が含まれることもあり、場合によっては当該利用者の個人情報とみなされることもありうる。この点、当該データの取り扱いにつき、分析を施すなどプロファイリングを行うことは欧州では規制される。しかし、この論点については本論稿の対象からは外れるため、その検討は別稿に譲りたい。

86) *Jeschke*, in: *Jeschke/Isenhardt/Hees/Henning (Hrsg.)*, *Automation, Communication and Cybernetics in Science and Engineering* 2015/2016, p. 499.

87) *Mercedes Benz*, *Car-to-X Communication*. Mercedes-Benz is starting a Europe-wide cooperation project. <https://group.mercedes-benz.com/innovation/case/connectivity/europe-wide-cooperation-car-to-x.html> (Data Accessed: 2022/05/31)

88) *Car to Car Communication Consortium*, *Deployment of V2X communication based on IST-G5*. [https://www.car-2-car.org/fileadmin/press/pdf/CAR\\_2\\_CAR\\_Communication\\_Consortium\\_Statement\\_ITSG5.pdf](https://www.car-2-car.org/fileadmin/press/pdf/CAR_2_CAR_Communication_Consortium_Statement_ITSG5.pdf) (Data Accessed: 2022/05/31)

89) *Jeschke*, *supra*, (fn. 86), p. 499.

90) *Car to Car Communication Consortium*, *C2C-CC Manifesto*, 2007, p. 31.

91) *Car to Car Communication Consortium*, *Clear benefits for road safety and traffic efficiency*. <https://www.car-2-car.org/about-c-its/> (Data Accessed: 2022/05/31)

る。車両間の通信や車両と環境との通信は、AI の一領域には該当しない。その理由は、個々のシステムは常に新しい情報を受け取り、それに従って作動するものの、例えばオイルが漏れた道路に近づくと、運転手がすぐに対応できるように車両が警告を発するなど、システムは他の車両や環境との相互作用を通じて学習するように訓練されていないからである<sup>92)</sup>。従って、Car-to-Car Communication という概念は、一見するとフリート・ラーニングと同義に見えるとしても、それと同一視してはならない。AI の一領域と言えるのは、フリート・ラーニングの分野だけである。

#### 第 4 款 2 種類の AI システム

先述した学習プロセスにおいて、現在 AI システムには、クローズ AI とオープン AI の 2 種類があることを考慮に入れなければならない<sup>93)</sup>。前者は、訓練段階が終了した後は学習を継続しない AI システムのことを指すので、これらのシステムは訓練段階を終えた後に到達した水準にとどまっている。その一方で、オープン・システムの場合は、これらも製造者側から訓練を受けるが、訓練段階が終了した後も学習を続けるという点で状況が異なる。このことは、製造責任者が訓練段階から製品をリリースし、各利用者と具体的に学習し続けることを意味する。このようにして、個別の利用者に最適化された AI 製品となるが、その一例が、Apple 社が開発したアシスタントシステムの Siri である。Siri は、利用者から新たなスキルを学ぶわけではないものの、個々の操作や習慣によって、利用者個人に最適化される。もっとも自動運転分野では、オープン AI はまだ使われておらず、クローズ AI のみが用いられている。

---

92) *Lohmann*, aa.O. (fn. 49) S. 51.

93) この分類は *Lohmann*, aa.O. (fn. 49) S. 60 による。

## 第5款 人間とAIのインタラクション

近年、自動運転や医療などの領域では、さまざまなAI技術の浸透が進んでいる。その理由は、そのような領域でこそ、AIが特に効果的に人間をサポートし、高い成果を上げることが可能であるとされ、人間のリスク要因はAIによって最小化されるという。

しかし、それが効果的な方法であるように見えても、あらゆるケースでAIが人間の影響から切り離されて動作するわけではないので、たとえば欧州AIハイレベル専門家グループや日本の総務省は、人間がAIを監督するシステムを提案している<sup>94)</sup>。この提案の背景にはAIハイレベル専門家グループが「信頼できるAI (Trustworthy AI)」の確立を望んでいることがある。監視やインタラクションが可能となるために検査できるからこそ、人間が信頼するAIであり、換言すると、信頼できるのは、人間より先に行動するAIではない。人間が監督するモデルでは、人間がAIを観察し、場合によっては介入する可能性を与えなければならない<sup>95)</sup>。それは、ある意味で人間とAIとの協力関係を創ることになる。

これを実現するためには、HITLモデル、HOTLモデル、そしてHICモデルという3つの可能性がある。まず、HITL (Human in the Loop) モデルには、このような、システムがすでに稼働中であるときに人間があらゆる決定に介入できるような協力が含まれる。“in the loop”という表現にあるように、人間はAIシステムの一部である。AIの利点を十分に活かせるかどうかは別個の問題ではあるものの、AIは個別の適用領域の改善に資するべきであることは言える。しかし、人間が常に介入すると、改善は遅々として進まないか、あるいは、およそ改善さえなされないこともありうる。

---

94) 総務省「AIネットワーク社会推進会議 報告書2021～『安心・安全で信頼性のあるAIの社会実装』の推進～」(2021年)102頁。

95) 以下のモデルの説明は、以下の説明は *HLEG on AI* (High-Level Expert Group on Artificial Intelligence), *Ethics Guidelines for trustworthy AI*, 2019, p. 16. を参照。

これに対して、いわゆる HOTL (Human on the Loop) モデルとは、人間がシステムの設計サイクルの形成に専ら従事しつつ、システムの作動中のみ監督できるモデルである。ここで人間は、開発段階ではシステムが機能するように障害を取り除くが、設計段階が終了した後は一步後退し、単なる監督者としての役割を果たすことになる。このモデルは、自動運転の分野でも頻繁に見受けられる。例えば、深層学習を行うカメラを搭載した車両で様々なテスト走行を行う際、この開発段階において「誤学習」を回避するために、カメラが学習したことを確認し、場合によってはそれを消去する形で人間が介入する。このようなシステムの開発プロセスを経てカメラが車両に搭載されると、人間は単なる観察者の役割を担うだけであり、システムは人間の影響を受けずに自立的に動作するようになる。

さらに、いわゆる HIC (Human in Command) モデルは、AI システムが計画領域で使用可能かどうか、使用可能な場合は個別状況に応じてどのように利用されるべきかを、人間が(予備的に)決定するものである。例えば、AI システムが融資の可否を決定するようなものが挙げられる。人間がまず、個別状況に応じて AI システムを利用するかどうか、利用する場合はどのようにするかを決定する。つまり、人間がそのシステムを利用することを決定した限りで、AI システムは人間によって正確なパラメータを与えられ、これらに基づいてシステムが独自に判断する。そして、その決定を受容もしくは拒否するかを責任者である人間が選択することができる。

これらの説明は、AI は自立的に行動する可能性はあるものの、必ずしも人間の存在が無視されているわけではないことを意味している。

## 第6款 小 括

上記で紹介した学習プロセスは、AI が様々な方法で学習可能であることを示している。それに応じて、最終的に AI が利用される分野によって、その各々の学習プロセスが目的に適したものとなる。AI は、これま

で人間が行っていたタスクを引き継ぎ、人間に比肩できないほどのスピードでこれを実行することで、多くの領域で人間の生活を便利にするとされる。このように AI は利便性を有する反面、ブラックボックス化というデメリットも無視することはできない。ブラックボックス化とは、AI の学習プロセスは分かっているとしても、個々の学習プロセスにおいて、なぜそのような結果になるのか、人間には理解できない場合が生じることである。特に人工ニューラルネットワークの領域では、これらのネットワークの内部で一体何が起きているのか、とりわけ AI がノードになぜある重み付けをしたのかを確実に説明することはできない<sup>96)</sup>。ニューラルネットワークの領域ではこの不明確性から、AI はブラックボックスとも呼ばれる<sup>97)</sup>。まとめると、AI は、設定された課題を解決するためのアプローチが必ずしも人間には理解されえないという点で、学習に人間のコントロールが及ばないのである。

このような性質を抱えた AI の利活用をつうじて、本論文の議論の対象でもある、人間の生命・身体・財産が侵害された場合における刑事責任の検討に当たり、たとえ「弱い AI」ないしは「特化型 AI」であるとしても、その学習手法によっては人間には予期し得ないような思考過程を構築することがありうる。このことは、結果との因果関係における不明確性をもたらしうるため、次節では、まさにこのような問題意識のもと、AI の特性を考慮しなければならないとされる事例を概観する。

## 第6節 AI の利活用と刑法上の問題

AI が搭載された製品は、その利用者が詳細には予測できない判断を下すこともありうる。例えば、自動運転車は、予測できない、部分的なスマート環境と協働しなければならず、それに搭載される AI は、内蔵されている学習方式に応じ、そのダイナミックな環境に対する作業予測を確立

---

96) HLEG on AI, *supra* (fn. 95) p. 13.

97) HLEG on AI, *supra* (fn. 95) p. 13.

するために、保存されているすべての情報と入力された情報を使用する。利用者は、AIが自身で取得した情報を独自に分析し、それに応じて自動的に作動することを知っているはずである。そのようなAIの特性を認識した、その利活用に関わる各主体に対し、発生した法益侵害結果を、刑法上どのように負責させるべきか、それとも負責させるべきでないのか。この問題が関連する事例は網羅的に検討する必要がある。それにあたり、まずはAIに起因する人間の生命・身体・財産の法益侵害事例について概観する。

### 第1款 自動運転車——アシャッフエンブルク事例と東名高速事例

海外の運転支援システムの事例については、冒頭でも触れたドイツのアシャッフエンブルク事例が挙げられる<sup>98)</sup>。2012年春、高性能の車線維持システムを搭載した自動車が、アシャッフエンブルク近郊のアルゼナウ村に進入したが、その入口で、運転手の60歳前後の男性が脳卒中となり意識を失った<sup>99)</sup>。その際、彼はハンドルを右に切り損ね、通常であれば村に入る前に茂みの中で停止するところであったところ、車線維持システムが車を道路に戻すように誘導したため、車は高速でアルゼナウ村に突っ込み、女性とその子供が死亡した。その子供の父親はジャンプすることで助かり、足を負傷しただけであった。この事件は、自律システムが人間の生命を「侵害した」事例であると考えられる。民法上は、ドイツ道路交通法7条の危険責任のために、特別な問題が提起されることはないが、より困難なのは刑法上の評価である。この事例においては、死亡事故を過失により惹起していない事故車両の運転手は答責的ではなく、2人の人間の死と1人の身体傷害に基づく過失致傷もしくは過失致死の答責主体として製造者が考慮されるという。

日本では、運転補助システム（レベル2）における交通事故（横浜地判令

---

98) *Hilgendorf*, a.a.O. (fn. 2), S. 66.

99) この運転手が死亡したか否かは不明である。Vgl. *Hilgendorf*, a.a.O. (fn. 2), S. 68.

和2年3月30日判例秘書LLI/DB L07550489)<sup>100)</sup>が発生している。後に検討を行うが、さしあたり簡単にその事例の概要を紹介する。これは、2018年4月28日、神奈川県綾瀬市内の東名高速道路上において、被告人はレベル2の普通乗用自動車を、運転支援システムを使用して走行中、仮睡状態に陥り、前を走行中の車が車線変更した後もそのまま進行し、進路前方に停車していた普通自動二輪車に加速した状態で衝突し、それにより、同自動二輪車を前方に跳ね飛ばして前方に佇立していた被害者3名に衝突させて、そのうちの1名を死亡させ、2名に傷害を負わせた事案であった。被告人には、運転中止義務違反に基づく過失運転致死傷罪により禁錮3年（執行猶予5年）が言い渡され確定した。

この事例はいわゆるレベル2の自動運転車<sup>101)</sup>による交通事故における利用者の注意義務を明らかにした初の判例であり、2010年代より盛んに議論されてきた自動運転車の事故事例をめぐる刑事責任の検討に大きな示唆を与えるものとなろう。もっとも、この検討から発展して、もし当該自動車がレベル3相当のものであったり、まだ実用化されていないがレベル4相当のものであったりする場合には事情は異なる。というのも、2020年道路交通法改正および2022年道路交通法改正においてレベル3ないしはレベル4相当の自動運転車に関する定義規定、及びそれに伴う利用者（特定自動運行実施者等）の義務規定が創設されたからである。むろん改正法の規定に従った、自動運転車の利用による利用者、ひいては製造者等に課せられる義務、及びそれに伴う刑事責任の帰属の妥当性に関しては別途検討を要する。

---

100) この判例を検討したものとして、中川由賀「具体的事故事例分析を通じた自動運転車の交通事故に関する刑事責任の研究② ～運転支援車（レベル2）の事故～」中京法学55号1巻（2020年）4頁以下や、樋笠堯士「自動運転（レベル2及び3）をめぐる刑事実務上の争点——レベル2東名事故を手がかりに——」捜査研究847号（2021年）46頁がある。

101) 以下、自動運転車のレベルに関しては特に断りのない限り、SAE基準を用いるものとする。

## 第2款 介護ロボット

近時の日本では、介護人材の不足が大きな課題となっている。介護分野の人材を確保する一方で、限られたマンパワーを有効に活用する解決策の一つとして、高齢者の自立支援を促進し、質の高い介護を実現するためのAIを搭載したロボット・センサー等の活用が期待される。その中で、介護ロボット<sup>102)</sup>を利用する老人ホームも見られるようになりつつあり、さらに、このようなロボットは将来的には、家庭のヘルパーとしても使われるようになるだろう<sup>103)</sup>。

このような介護ロボットの利活用における刑法上の問題として、例えば、AIを搭載した介護用ロボットが多世代で暮らす家庭に販売されるといふ、架空ではあるが、前述のように将来的には十分ありうるケースを想定する。例えば、ある家庭では、物忘れがひどくなってきた祖母Aのために、このロボットが定期的に飲み物や食べ物を運搬しており、その間母親Bは別の仕事に専念していた。このAIロボットが作動していたとき、予期しないセンサーの誤作動により毛布の上で遊んでいたBの子であるCを轢いて死亡させてしまった。しかし、このセンサーは、製造企業による定期メンテナンスが行われており、これまでそのような不具合は知られておらず、予期もされていなかった。

この事例において製造企業や利用者(管理者)<sup>104)</sup>の過失による刑事責任を検討するにあたっては少し注意が必要である。まず、製造企業に対する

---

102) 「介護ロボット」とは、ロボット技術が応用され利用者の自立支援や介護者の負担の軽減に役立つ介護機器である。その種類としては、移乗支援、移動支援、排泄支援そして認知症の方の見守りなどが挙げられる。(厚生労働省ホームページ <https://www.mhlw.go.jp/file/06-Seisakujouhou-12300000-Roukenkyoku/0000210895.pdf> (最終アクセス2022/05/31) 参照)

103) 日本における介護ロボットの現状の利活用例については、厚生労働省「介護ロボット導入活用事例集 2021」(2021年) <https://www.mhlw.go.jp/content/12300000/000928395.pdf> (最終アクセス2022/05/31) を参照。

104) 上記事例ではBを指すものとする。なお、上記事例については *Lohmann*, a.a.O. (fn. 49), S. 88 f を参照。

過失犯の成否につき、この AI ロボットの誤作動に対応する刑法上の注意義務の有無については、当該誤作動に対する予見可能性、そしてこの誤作動に対応する作為義務を事故当時に有していたか否かが問題となる。そして利用者（管理者）に対しては、この AI ロボットの誤作動による結果発生を予見し得たか否かが問題となる。

### 第3款 産業用ロボット——バウナタール事例

工場生産に供される産業用ロボットにも AI を搭載しているものはあり、それが人間の生命・身体を侵害することもありうる。実際に、次のような事例がドイツであった。

バウナタールの工場棟で、2015年6月に作業員Aが遮蔽されていないロボットアームにつかまれて死亡した。他の作業員Bが、誤って規定よりも早くこの機械のスイッチを作動させてしまったという<sup>105)</sup>。CNNでも報道されたこの事故のニュース<sup>106)</sup>は、ロボットによる人間の殺害と関連するものにみえるかもしれない。もっとも、Bが注意義務に違反したことが証明できれば、過失致死罪が成立しうる。ここ数年間でロボットが遮蔽された場所で働くのではなく、人間と直接接触することができるようになったため、この種の事例はありえないものではなくってきている。さらに、部門長や企業の安全管理者などが、労働者に対して保障人的地位にあった場合には、そのような安全管理者の発生結果に対する負責も考えられる。しかしながら、この事件で大きな注目を集めたのは、人間が自働するロボット（機械）に殺害されたように見えたことによる。これは、人間に危害をもたらしうる「人工的な存在」への深層的な恐怖や空想に訴えかけるものであったとされる<sup>107)</sup>。

---

105) <https://www.nha.de/kassel/kreis-kassel/baunatal-ort312516/toedlicher-roboter-unfall-bei-vw-kassel-in-baunatal-vor-gericht-8413531.html> (Data Accessed: 2022/05/31)

106) <https://edition.cnn.com/2015/07/02/europe/germany-volkswagen-robot-AI-lls-worker/index.html> (Data Accessed: 2021/09/30)

107) *Hilgendorf*, *Autonome Systeme, künstliche Intelligenz und Roboter Eine Orientierung* ↗

この事例は、AI が搭載されていない産業用ロボットが人間の生命を侵害したのではあるが、仮にロボットアームが AI の判断により作動するものであったとするならば、本稿で検討する問題に沿うものとなる。すなわち、バウナタール事件の変形事例として、ロボットアームがもっぱら自立学習を行う AI の統制のもとで作動していたものとするれば、A の死亡結果と B の行為の間にある因果関係の機序が不明確となりうるからである。さらにこの場合、B の背後にいる管理者の保障義務の検討についてはなお慎重になされなければならないだろう。

#### 第 4 款 過失犯処罰規定のない犯罪類型

さらに AI を搭載した機械がネットワーク上で、コンピュータ犯罪や、経済犯罪の構成要件が実現されることがありうる。特に、経済領域においては、近時ますますそのサービスが展開されているとされる AI・アルゴリズム投資における、相場操縦やこれを介したインサイダー取引のような証券犯罪、さらにはデジタル・カルテルのような競争法違反が挙げられる。これらの犯罪において問題となるのは、利用者ないしは製造者の予期しない AI の学習結果によって当該犯罪の構成要件が実現された場合の刑事責任の所在である。これらの犯罪類型には、過失犯処罰規定が存在しない。そのため、学習しない AI・アルゴリズムの使用であれば利用者ないし製造者にその刑事責任が帰属しうるところ、学習を行う AI が介入することによって、彼らの故意の認定が困難になるのではないかということである<sup>108)</sup>。

また、コンピュータ犯罪の領域では、学習によって利用者に最適化された AI に対する個人情報のハッキングをはじめとする不正アクセス罪（不正アクセス禁止法 3 条・8 条）の成否や、そのような AI の利用を妨害する場合の電子計算機損壊等業務妨害罪（刑法 234 条の 2）、不正指令電磁的記

---

↘ aus strafrechtlicher Perspektive, FS Fischer, 2018, S. 107.

108) このことについては、第 3 章で詳細に検討を行う。

録作成・供用罪（刑法168条の2）の成否、さらには学習を行う AI を搭載した電子商取引に供されるソフトウェア・エージェントに対して不正アクセスがなされ、虚偽ないしは不実の電磁的記録により当該ソフトウェア・エージェントの管理者の財産が騙取されたが、その電磁的記録が不正アクセス時における虚偽の情報もしくは不正な指令の供与によって作出されたのか、それとも不正アクセス後の AI の学習結果によって作出されたのか、が不明確な場合における電子計算機等使用詐欺罪（刑法246条の2）の成否<sup>109)</sup>などが考えられる。このように、人間が行為主体で、AI が行為客体となる類型を中心にした議論があまり見られないところ、これらの事例はサイバー・セキュリティ上の観点から不可欠なものであるといえる。

## 第5款 小 括

以上のような AI 製品に起因する人間の生命・身体・財産等の法益侵害についての刑法上の責任帰属問題は実際の事案から仮想事例まで含めて喫緊の課題ともいえよう。まず、従来の自動運転車の刑法上の議論について、レベル2相当の自動運転車については実務上の見解が示されており、同時に二度の道路交通法改正を経てレベル3ないしは4の自動運転車に関する利用者の義務が詳細に明文化されている<sup>110)</sup>。もっとも、上記で示したようなアシャッフエンブルク事件<sup>111)</sup>や東名高速事故のようなレベル2

109) この場合、不正アクセス罪は成立することを前提にする。

110) この状況はドイツでも同様である。例えば、StVO（ドイツ道路交通法）第8次改正において新設された1a条における「高度な又は完全な自動運転機能を有する自動車」とはレベル3,4,5の自動運転車を指し、1b条以下でこれら自動車の利用者・製造者に対する義務を明文化している。その概要については、泉真樹子「ドイツにおける自動運転車の公道通行——第8次道路交通法改正——」国立国会図書館（2018年）[https://dndl.go.jp/view/download/digidepo\\_11052071\\_po\\_02750004.pdf?contentNo=1](https://dndl.go.jp/view/download/digidepo_11052071_po_02750004.pdf?contentNo=1)（2022/05/31閲覧）を参照。

111) もっとも *Hilgendorf*, aa.O. (fn. 2), S. 68 では、利用者は「すでに被害惹起者としての明確に帰属可能である行為が欠如し、そもそも注意違反の存在を指摘することができない」という理由で刑事責任の検討から利用者を排除し、専ら製造者の刑事責任のみを検討している。しかし、なぜ利用者に帰属可能な行為が欠如するといえるのかの説明はなさず

相当の自動運転車に関する利用者や製造者の法律上の義務に関しては依然として普通自動車（レベル0）のそれと同一であるが果たしてこの状況は妥当なものであろうか<sup>112)</sup>。近時の動向を踏まえて再度検討を要するものと思われる。

その一方で、AIを利活用した製品は今や自動運転車に限定されず、介護現場で利用される介護ロボットや、製品工場においてオートメーションに供される産業用ロボットなど実体を持つAIから、AI・アルゴリズム投資のような実体を持たないAIも存在し、そこにも刑法上の問題をはらむので、自動運転車をベースにした議論が隆盛を極めていた頃に比べて、この問題はより多彩なものとなり、より具体的なものとなってきた。これらの問題は断片的にではなく、体系的に検討することが必要であり、このことこそが、AI製品の利活用における刑法上の問題を再び仔細に検討する意義である。次章では、そのなかでも、従前の議論の対象であった生命・身体侵害に対する検討を行う。その際、過剰な刑事責任を課すことによりAIの利活用や技術開発を萎縮させないように配慮しつつ、AI製品の挙動の予見不可能性を前面に出すことにより、法益侵害結果が、いかなる人間的主体にも帰属されなくなるという「帰属の間隙」は可能な限り生じるべきでないという立場から検討を行う。

---

ゝれていない。

112) 免許制度に関する指摘ではあるが、運転支援システムの監視制御にかかる訓練・技能はレベル2の自動運転車から必要であると古川伸彦「自動運転車、死傷事故、刑事過失責任」刑法雑誌59巻2号（2020年）176頁の脚注16は言う。