〔1346〕

氏名（生年月日）	**根津洸希**（1989年10月31日）
学位の種類	博士（法学）
学位記番号	法博甲第141号
学位授与の日付	2021年3月17日
学位授与の要件	中央大学学位規則第4条第1項
学位論文題目	**AIの刑事責任**
論文審査委員	主査　只木　誠
	副査　鈴木　彰雄・安井　哲章

博士学位（甲）請求論文審査報告書

Ⅰ　本論文の主題と構成

根津洸希氏より提出された博士学位（甲）請求論文「AIの刑事責任」の構成は、以下の通りである。

序論

第一章　「AIと刑法」の問題は過失犯論の精緻化によって解決可能か
　　―いわゆる「ブラックボックス性」を手掛かりに―

Ⅰ．はじめに―AI と刑法を巡る二つのアプローチ

Ⅱ．「ブラックボックス性」の原因と解決の試み

1. ブラックボックス性の原因
2. 解決の試み

Ⅲ．「ブラックボックス性」の刑法的取扱い

1. ブラックボックス性の実質的意義：予測困難性と検証困難性
2. 予測困難性としてのブラックボックス性
3. 検証困難性としてのブラックボックス性

Ⅳ．おわりに

1. 本稿の検討の要約
2. AI 責任論の可能性

第二章　AI技術を巡る刑法的問題の概説と解決の試み
　　―(部分的)自動運転技術を一例に―

Ⅰ．はじめに ―（部分的)自動運転自動車の利用による事故

2. 「決定論」の定義と非両立性論：Inwagen の見解

Ⅲ. 両立性論の応答

1. 他行為可能性は責任の構成要素ではない：Frankfurt の見解

2. そもそも意思自由論や決定論は責任と無関係である：Strawson の見解

3. 行為のメカニズムを所有する：Fischer=Ravizza の見解

Ⅳ. 検討

Ⅴ.おわりに

第五章

Ⅰ. はじめに ― 4 つのテーマの提示

Ⅱ. 「刑罰」を受けるのは誰かという問題

Ⅲ. 「刑罰」はロボットや AI にとって苦痛となりうるかという問題

Ⅳ. 近代刑法における意味での「刑罰」と呼べるかという問題

Ⅴ. ロボットや AI を「処罰」することによる「刑罰」という語の意味変容という問題

Ⅵ. おわりに

終章

Ⅱ 本論文の概要

１．本論文の目的および構成

AI 技術が実用段階に達した現代において、その AI 技術の利用によって法益侵害結果が生じた場合、刑法上の過失論の精緻化のみによっては解決できない領域が存在しており、そこでは責任分配が主たる問題となるであろう。本稿は、そのような筆者の立論のもと、技術発展を阻害せず、過剰処罰にならないようにするため、AI に刑事責任主体性を肯定することにより妥当な結論を導くことができるかについて、AI の責任を否定する見解を詳細に批判的に検討し、同時に、いわゆる決定論に立ってもその責任を肯定できることの論証から、AI の刑事責任肯定説の基礎づけを試みようとするものである。

２.「第一章 『AI と刑法』の問題は過失犯論の精緻化によって解決可能か―いわゆる『ブラックボックス性』を手掛かりに―」の概要

第一章では、本研究における問題の所在を提示することが試みられている。すなわち、いわゆる「強い AI」という技術が実用段階に入った場合、人間の関与者のコントロールの及ばないところで法益侵害結果が生じうることが想定しうること、ならびにその事故原因の究明は AI のいわゆる「ブラックボックス性」によって困難を極めるであろうことを指摘する。そして、このブラックボックス性という AI 独自の問題性に対して、これまでの過失犯論の要件とされてきた結果予見可能性の

精緻化によって解決することはできないのではないかという疑義を提起するのである。

そのことから、まず、AI のブラックボックス性がいかなる技術的特性から生じるのか、それは技術的取り組みにおいて解消可能な問題なのかについて概観したうえで、筆者は、ブラックボックス性は完全には排除できないことを確認し、ブラックボックス性は刑法における検討枠組にどのような影響を与えうるのかを検討している。

ブラックボックス性は「予測困難性」と「検証困難性」という二つの要素から成っているが、前者については、AI の行動が予測困難であっても、過失犯論における結果予見可能性判断においては因果経過の基本的部分の認識があれば足り、結果発生の機序を厳密に認識している必要はないことから、現在の過失犯論を理論的に進化させることによって対応が可能であるとする。

他方で、検証困難性に関しては、筆者は、過失犯論ではなく責任主体の主体選定にかかわる議論が不可避となるとする。たとえば、過失の競合などにより結果原因が検証できない場合には、過失犯論からではなく、共同正犯論からのアプローチ(すなわち、過失の共同正犯を肯定して因果関係の認定を擬制するなど)によって解決が試みられるが、このように、検証困難性が問題となるような事例においては、責任主体の選定、ならびに責任主体間での責任分配が問われているとする。

そして、AI が法益侵害結果をもたらし、その結果原因が検証困難であるとき、そこには「答責の間隙」が生じており、刑法上の過失犯論よっては解決されず、答責の間隙に対する責任分配をめぐる議論が重要となってくるという。

3.「第二章 AI 技術を巡る刑法的問題の概説と解決の試み―(部分的)自動運転技術を一例に―」の概要

前章において論じた「答責の間隙」を解決するため、本章において、筆者は、現在、①過失犯論を修正して製造者・利用者に広く答責するという引き受け過失という法律構成による解決、②あらかじめ法律により答責主体を規定してしまうという立法的解決、③新技術の長所も短所も社会全体で引き受ける社会的受容という構想による解決、④AI に法的人格性・自由答責的な主体性を肯定することで関与した人間の答責を限定するという法律構成による考え方が提案されていることを確認する。

上記のなか、いずれも答責の間隙を人間の答責領域の拡張によって解決しようとする立場である①と②は、比較的現実的な解決策であるように思われるものの、人間の負担軽減のために開発された AI 技術の利用により、自らのコントロールが及ばない領域についても人間が責任を負わされることになるため妥当な結論とはいえず、また、③は理念的には正しいようにも思えるが、具体的な帰結が明らかではないことから、同理念は具体的な法律構成に落とし込まれることが必要であろう。検証を通してのこのような結論から、筆者は、AI に理論的フィクションとして責任を肯定し、責任分配の当事者と認めることで製造者や利用者の答責領域を限定するという④の見解が妥当であるとの結論に至る。

４.「第三章 AI の責任主体性を巡る諸見解」の概要

第三章では、「AI の責任」という新たな概念をめぐる議論を紹介するとともに、AI に責任を肯定するために障害となる問題について検討を加えている。

AI の責任主体性をめぐっての主張は、積極説、消極説、中間説という三つの説に大別することができるが、積極説も消極説も、現在のところ決定的な論証を提示するには至っておらず、筆者によれば、議論は膠着状態にあるという。

そこでかかる膠着状態に陥ってしまっている根本に遡るに、筆者は、AI の責任主体性をめぐる議論においては責任の基礎理論における争点形成がなされていないという点にその原因があるとしている。とりわけ、積極説も、また、消極説も、現状の責任の基礎理論を否定せずにこれを維持しようとすることから、ややもすれば人間の責任までをも否定しかねない「決定論」問題に踏み込むことを躊躇し、そのため迂遠な議論となって錯綜してしまったのではないか。そのような分析に立ち、筆者は、AI の責任主体性を論じるにあたっては決定論問題があらためて問われなければならないことを指摘する。

５.「第四章 AI 責任否定論と決定論問題」の概要

第四章では、第三章での帰結を受けて、問題をごく限定しつつ、筆者は、決定論問題に取り組んでいる。そして、本研究の目的は AI の責任を肯定することそのことではなく、AI に責任を肯定することが理論的に不可能なことではないということを示すこと、換言すれば、決定論と責任の両立可能性を示すことで十分であるという。

一般的には「決定論が真ならば、責任はない」という命題は比較的広く共有されている前提であるようであるが、それは、決定論は意思の決定性に基づいて他行為可能性を否定するゆえに、他行為可能性を中核におく規範的責任概念とは両立不可能だとされているのであり、しかし、非決定論も決定論も責任とは無関係であるとする見解や、責任にとって重要なのは「他行為が可能であった」という事実ではなく「その行為が『その行為者のもの』であった」といえるかであるとする見解には一定の説得力があるという。したがって、事実としての他行為可能性を責任の中核に据える必然性はないのであって、そうであれば決定論が事実としての他行為可能性を否定しても、それは必ずしも責任を否定する理由にはならず、それゆえ、仮に AI がプログラムに従って行動するのみで他行為の可能性が事実的にないとしても、それは AI の責任を否定する論拠にはならないということは明らかであると結論づけるのである。

６.「第五章」の概要

そして、最終章の第五章において、筆者は、では、AI に責任を認めるということからただちに AI の処罰が要請されるべきであるかを問題として提起する。

AI の処罰の如何を検討するにあたって、筆者は、①AI の人格の範囲、②AI に対する刑罰の実効性、③近代刑法における刑罰の意味、④刑罰の意味変容という問題を掲げて、それぞれに検討を加

えている。すなわち、AIがネットワーク化された場合、各個体としてのロボットが「一人」の人格であるのか、あるいは情報共有がなされた総体が「一人」であるのかが決定されなければならず、また、現行法がAIに対して有効な刑種を予定していないなか、それならば、再プログラミングを「刑罰」と呼ぶことが可能であるのか、とりわけ思想刑や科学的去勢の問題との関係で検討されなければならないとする。そして、それらを前提として、AIに処罰を科すという場合でも、それが単なる犯罪への不安から生じるのであれば、刑罰は責任非難としての意義を失い、市民の不安や不満の「ガス抜き」に堕してしまうことになるであろうことを指摘する。

そして、これらの問題が提起され、その解決策が求められるところ、AIに責任が否定されないということからただちにAIの処罰可能性が導き出されるわけではなく、処罰の可能性については、今後の技術発展の展開をまって、慎重な態度決定が必要であると結論づけるのである。

Ⅲ　本論文の評価

AI技術の実用化が進む現在、AIそれ自身が法益侵害結果をもたらすような事故を引き起こし、そのような場面においては、起こりうる結果の予測や事故が引き起こされた原因の検証のいずれもが困難であることが予想されるところ、本論文において、筆者は、まず、人間のコントロールが及ばないかかる領域における法益侵害結果について、製造者や利用者の処罰を限定する手立てとして「AIの責任」という概念の適用が有用であろうと思われる事例群、すなわち、筆者の言葉によれば「答責の間隙」が生じる領域を明らかにする。そして、この答責の間隙を埋める理論的な根拠は、これを、AIに法的人格性・自由答責的な主体性を肯定する考え方に依るべきであると説いている。本論文の結論ともいうべきこの見解はドイツの学説を範とするものであるが、独自の理論づけも加えられており評価に値しよう。

筆者は、また、AIの責任主体性を認めることの障害となり、また、AIの責任主体性の議論が低調であったところの所以は、責任を認めるための一定の能力や属性がAIに欠けているからなのではなく、たとえば、意思決定論問題の掘り下げなど、責任論の基礎研究における議論が意図的に避けられ、十分に検討されてこなかったことにあるとし、このような自説の丹念な立証作業を通して、AIの責任主体性を肯定するため積極的な基礎づけとAI処罰の可能性について明確な回答が導かれ得るとしている。この点、責任論の基礎研究についての筆者の分析に対しては、なお反論も予想されるところではあるが、一つの理論的仮説として首肯できるものであり、オリジナリティーも認められる。

そのような独自の問題意識のもと、続いて、筆者は、これまで主張されてきたAIの責任否定説に詳細な批判的検討を加え、そのうえで、意思決定論との関係において、AIの責任は理論的に当然に否定されるべきものではなく、肯定可能なのであることを論理的に基礎づけようと試みる。ドイツおよびわが国の学説の紹介に立って展開される同部分には、紹介論文にとどまらない価値を見て取ることができよう。

以上のような考察、検討を経て、筆者は、AIに責任主体性を認めることにより、設計者、製造者、

利用者など AI に関わる者の処罰が理論的に回避可能となり、同時に、それによって、AI 技術の発展も阻害されないようになるとの結論を試論として導くに至る。このような視点とそれにもとづく筆者の主張は、自動運転の発展という社会的な変容に即した刑罰論の醸成という点で意義があり、たとえば刑罰非難の道徳的価値と法的責任論をめぐる問題や、刑罰非難は個別行為のみに向けられているのかという刑法基礎理論上の問題の解決に向けて新たないとぐちを模索するという意味でも有用であるといえよう。

とはいえ、本論文には、以下のような問題点も指摘できよう。

まず、方法論的には、AI に理論的フィクションとして責任を肯定することが可能であるとしても、それによって導かれる結論が十分に検証可能であるのかについては、筆者の検討にもかかわらず、なお、疑念が生じよう。また、意思決定論を通して AI の責任を肯定することが理論的に全く不可能ではないということをもって、なにゆえに、AI の責任が肯定できるのかについて、この点も筆者なりの説明はあるものの、さらなる論証、基礎づけが必要であろう。さらに、解釈論のレベルでは、そもそも、AI による法益侵害結果が発生したとき、過失共同正犯理論を用いることで実際に答責の間隙が生じるものなのか、この点についても一層積極的な根拠づけが求められよう。この点は、最終試験でも問題とされた点である。加えて、紙幅の関係もあろうところ、やむを得ないことながら、意思決定論あるいは責任概念等の根本的な基本概念について踏み込んだ考察がなされていれば、一層魅力的な論文に仕上がっていたものと思われる。別ながら、一部には、AI 技術による運転のどの段階を前提とした議論なのかが不分明なところも見受けられた。

もっとも、以上のような点は、従来の理論を介して新たな課題に挑戦的に検討するに際し必然的に伴うところということができ、本論文の価値を大幅に低めるものではない。

AI が刑事責任の主体となりうるのかという問題については、いまだ議論の黎明期であり、わが国では自覚的な研究がほとんど見出せないなか、本稿は、ドイツにおける AI の責任主体性の議論を詳細に検討し、わが国における責任論の考え方を問い直し、その上で、今後の AI 技術の発展、すなわち、あるべき AI 技術の発展を最大減に効果あらしめようとし、それを可能とするために、AI 技術を利用する人の処罰を回避し、不要な萎縮効果を取り除くことを意図したものである。その点で、本稿は独自性を有する意欲作であり、そのアプローチが評価され、筆者に対していくつかの研究機関から共同研究者としての招聘のオファーが存するのもうなずけるところである。

Ⅳ　結論

以上を総合的に判断するに、審査委員一同の意見として、この度根津洸希氏より提出された本論文は博士（法学）の学位を授与するに値するものと思料する次第である。