

AIの倫理と法

組織のダイナミズム研究会

2021年5月22日(5月例会)

伊東俊彦

組織のダイナミズム研究会代表

目次

- 1. はじめに 2
- 2. 先行研究 4
- 3. AI倫理(西垣・他) 10
- 4. AI倫理に対する各社取り組み 16
- 5. AIの時代と法(小塚) 17
- 6. AIロボットは人格を持つか(西垣・他) 23
- 7. AI・ロボットと法(弥栄・他) 27
- おわりに 44
- 補助資料 46
- 参考文献 49

1. はじめに

• 前回発表について

- 発表期日: 2019年8月ODSG例会(8月24日)
- 発表タイトル:「AIと倫理」
 - AIの基本認識
 - AIと倫理を並列で捉える
 - AI使用のリスク増加の認識
 - 「AI倫理」をメンバーと議論するきっかけとしたい

• 今回発表について

- 発表タイトル:「AIの倫理と法」
 - AIの倫理面を中心に考察
 - AIシステムとしてのロボットにも焦点
 - 自律型兵器システムにも焦点
 - AIの利活用における法の概要を考察

2. 先行研究-1

- 前回発表と重複あり(確認のため)
- AIの**誕生**(松尾、2016a)
 - 1956年:米ダートマス大の研究集会で、ジョン・マッカーシーらがコンピュータを「思考」に使うこと「**Artificial Intelligence : AI**」を提唱
- AIの**定義**(松尾、2016a)
 - 明確な定義はないが、**人間のような知能**を、コンピュータを使って実現することを目指した技術あるいは研究分野
- AIと人間の**創造性**(松田、2020)
 - 人間は敢えて「創造的になれ」と言われなくても日々生きるという物語を創造している
- **哲学**と人工知能(長尾真、1992)
 - 哲学の道具としては紙と鉛筆しかない ⇒ 人工知能研究者は**コンピュータ**という強力な道具をもっている

2. 先行研究-2

- **倫理**とは(大辞泉、1998)
 - 人として守り行うべき道
 - **善悪・正邪**の判断において**普遍的な基準**となるもの
 - **道徳**とは:人々が**善悪**をわきまえて正しい行為をなすために、守り従わねばならない**規範**の総体
 - **道徳**という言葉と**倫理**という言葉は、ほとんど**同じ意味**である(種村、2019)
 - **倫理**(**ethics**)の原語 ⇒ **ethos**(習慣:ギリシャ語)に由来する**習慣的性状**(種村、2019)
 - アリストテレス「ニコマコス倫理学」の**倫理的徳**(ethike arete)

2. 先行研究-3

- AI倫理とは

- 人工知能技術を応用したシステムに関する倫理(芳田、2019)

- AI倫理の範囲

- AIシステムを開発する技術者の倫理
- AIシステムが実現するシステムの倫理的対応
- AIが身体性*を持つと、倫理がより重要となる(IPA、2019)
- AIがAIシステムを作るときの倫理(強いAI、汎用型AI)

*身体性: 行動体と環境との相互作用を身体が規定すること、及びその内容

2. 先行研究-4

- ロボット倫理-1

- ロボットの定義

- 一定の自律性をもち、環境内を移動して所期の**タスクを実行する作動メカニズム**(JIS B8445:2016)
 - しかし**身体性***をもつとは限らない⇒ロボット掃除機

- **ロボット倫理**の議論-1(久木田、2017)

- 機械は**道徳的存在**になりうるか
 - **道徳的な機械**はいかにして可能か
 - ロボットに対し**道徳的配慮**が必要か⇒ロボットの**廃棄問題**

2. 先行研究-5

- ロボット倫理-2

- アシモフのロボット工学三原則の限界

- AIの分野では、論理的アプローチには本質的な限界がある

- 知能とは論理的アプローチで実現できるものだけでなく、もっと多様なものである

- *アシモフのロボット工学三原則: 1. 人間に危害を加えてはならない
2. 人間の命令に従わなくてはならない
3. 自らの存在を守らなくてはならない

- AI倫理の問題 (人工知能学会、2017b)

- 1 AI技術がもたらす社会的影響としての問題

- 2 AI研究・開発者に要求される専門職倫理に関連する問題

- 3 一般消費者を含めた利用者側で生じる倫理問題

- 4 AI技術そのものの責任を含めた倫理性の問題

2. 先行研究-6

- 内閣府『人間中心のAI社会原則』
(内閣府、2019)

- 基本理念

- 人間の尊厳が尊重される社会
- 多様な背景を持つ人々が多様な幸せを追求できる持続性ある社会

- ビジョン

- Society5.0実現に必要な社会変革「AI-Readyな社会」

- 人間中心のAI社会原則

- 人間中心の原則: AI悪用がないよう、AIは人間の補助、情報弱者・技術弱者の排除
- 教育・リテラシーの原則: 幼児・初中等教育のリテラシー教育、インタラクティブ教育
- プライバシー確保の原則: 個人の自由、尊厳、平等の確保
- セキュリティ確保の原則: リスクの評価と低減、持続可能性に留意
- 公正競争確保の原則: AI資源の集中による支配による不公正な競争の排除
- 公平性、説明責任制及び透明性の原則: 人々の不当な差別の排除と公平な取扱い
- イノベーションの原則: AIの発展にあわせ人も進化していく継続的イノベーション



3. AI倫理(西垣・河島より)-1

- 近代の倫理思想-1 (pp.33-47)
 - AI倫理を論じるための最小限の整理としての4つの倫理思想
 - 功利主義: 最大多数の最大幸福(ベンサム)
 - 共同体の集団的な価値の最大化を目指す
 - 最大幸福: 共同体メンバーの幸福度の総和
 - 集団の利益が優先 ⇒ 個人が犠牲になることもある
 - 自由平等主義: カントの倫理思想
 - 人間は理性を持つ特別な存在であり基本的人権が守られるべき
 - 人類中心主義は環境倫理学から見れば独善的となる
 - 自由至上主義: リベラリスト(ロック)
 - 公平な市場経済により個人が得た財貨の所有権は絶体である
 - 王様とか国家といった公権力が勝手に個人の財産権をおかしてはならない

3. AI倫理(西垣・河島より)-2

- 近代の倫理思想-2 (pp.33-47)
 - 共同体主義: コミュニタリアン(サンデル)
 - 個人の属する**集団の伝統的な共通善**に注目する
 - 自由市場主義への反論: 自由至上主義は個人の自由を尊重しているといっても**富裕な人々の権利**だけをを守っているのではないか
 - **他の共同体の倫理は誤り**として否定される
⇒ そのため、**多様な共同体**におけるそれぞれの倫理規範を寛大に**尊重**せざるを得ない
- AI倫理と情報倫理 (pp.52-53)
 - AI倫理を情報倫理の一部と見なすわけにはいかない
 - AIが**擬似的な人格**を持つという観念があるからである
 - AIロボットを疑似人格として**責任**をとらせることができるか

3. AI倫理(西垣・河島より)-3

- AI倫理をいかに**構築**するか-1
 - **AI倫理**とは(pp.132-133)
 - **AIが作動中**に守るべき倫理規範だけを指すのではない
 - 社会における**多様な倫理的諸問題**を考察することが必要
 - 倫理のフレームワーク(pp.133-140)
 - **道德観**:個人的な**正義の基準**であり倫理的に適切かを定める
 - **社会規範**:慣習的な礼儀や罰則をともなう法律など
⇒ 各個人の**道德観と一致するとは限らない**
 - **行動**:自らの**道德観と社会規範の緊張関係**をもとにうまれる
 - **道德観と社会規範が合致/近接していないと矛盾軋轢**が発生
 - 「**共感**」から、他者の人格の幸福や尊厳をまもる**社会規範**が設定

3. AI倫理(西垣・河島より)-4

- AIエージェントの使命(pp.140-144)
 - **AIエージェント**は決して道徳的主体になれない
 - 倫理的/道徳的な**責任を負える**のは人間以外にはない
 - ただしAIエージェントに**社会規範を守らせる**ような設計は重要
 - **人間**が他者の心に宿る道徳観を乗り越える**相互作用が共感**である ⇒ 他者の身になって意味解釈をおこなうこと
 - AIエージェントに**共感はない** ⇒ 新規の社会規範の生成は不可能
 - ロボットが道徳観をもつ人格と見せかけると**倫理的危機を招く**
 - AIは**社会規範**をまとめるために有益な**補助計算手段**になる

3. AI倫理(西垣・河島より)-5

- AIシステムの**保守維持**の重要性(pp.158-163)
 - 保守維持がAIの**有用性と倫理性**の鍵となる
 - AIの予測を超えた行動
 - ① 人間の思いつかない**有用で新規性**のある出力をもたらす
 - ② プログラムの設計ミスによる**誤動作**
 - ③ **外部操作**による設計者の**意図に反する**作動
 - 保守維持のために第三者の独立した監視機関を設置
 - 監視機関の**構成**
 - AIとICTに通じた**理系専門家**
 - 公共哲学をふまえて人間のためのAIの活用に明るい**人文系研究者**

3. AI倫理(西垣・河島より)-6

- AI倫理の思想(pp.167-168)
 - AIEージェントやロボットは自律的に思考しているように振る舞っても、自律的に意味解釈をしているわけではない
 - 固有の道德観はもたず、組み込まれた社会規範にもとづいて作動
 - 人間は自分で世界を観察し、意味解釈し、他者に共感しつつ、固有の世界の中で生きる ⇒ その経験からはじめて道德観がうまれる
 - 個々の道德観をあわせ、集合知として社会規範がつくられていく
 - 人間の行うこのプロセスをAIEージェントが援助できるようにしていくことが人間の知性である
 - このときAIはIA(Intelligence Amplifier: 人知増幅機械)となる

4. AI倫理に対する各社取り組み-1

- AIの**利用が問われた**事例(日経新聞,2021)

年月	内容
2016.3	米マイクロソフトのAIロボット「Tay」が 差別的な発言 をするようになり公開中止
2018.1	米アマゾンのAIによる 人材採用 システムで、 女性に不利な 傾向が出て開発中止
2019.11	米アップルのクレジットカードの 与信アルゴリズム が 男性を優遇 すると指摘され、カード運営のG.サックスを調査
2020.4	日本IBMのAI 人事評価 の情報開示を求め関係労組が都労働委員会に救済申し立て
2020.6	米IBM、アマゾン等が警察向けの 顔認証AI に関する事業の停止や 撤退 を表明

4. AI倫理に対する各社取り組み-2

- AI倫理の取り組み(日経新聞,2020a,2021,2022、富士通,2019)

年	内容
2017	米マイクロソフトはAI倫理委員会を発足し、AIの透明性や公平性を担保する
2018	米グーグルは、AIの軍事利用への懸念などを受けAI倫理憲章を策定
2019.3	富士通は、「人と協調する、人のためのAI」を主張し、AI倫理への取り組みを「富士通グループコミットメント」として発表
2019.4	NECは、AIの活用に関する公平性、透明性やプライバシー等を盛り込んだ7つの指針を策定
2021.2	日立は、AIに関する3つの行動基準と7つの実践項目からなる倫理原則を策定
2021春	ソニーは、製品がAI倫理の基準に沿っているか品質管理の仕組みの中で審査⇒AI倫理を定めたガイドラインを品質管理に反映

4. AI倫理に対する各社取り組み-3

- グーグルのAI倫理の7原則(日経新聞,2020b)
 - 社会的な有益性
 - 不公平なバイアスの防止
 - 安全性確保を念頭においた開発と試験
 - 人々への説明責任
 - 自社のプライバシー原則の適用
 - 科学的卓越性の探究
 - 基本理念に沿った利用への技術提供

5. AIの時代と法(小塚2019)-1

- 法制度に対する期待と不満-1
 - 「**トロッコ問題**」と責任ルール(p.14)(太田、2020,pp.77-81)
 - トロッコ問題:解決が困難な哲学上の問題(英のフィリッパ・フット教授)
 - **ブレーキのついていないトロッコ**と線路上の**5人**の作業員
 - 手前のポイントを切り替えれば別の**1人**の作業員を轢いてしまう
 - ポイントを切り替えて**犠牲者が少ない方**を選択する(ひとつの解)
 - **法律家**から見たトロッコ問題(pp.15-17)
 - トロッコを**AI**に置き換えた場合
 - **犠牲者の遺族**から相談を受けた**弁護士**の対応
 - トロッコの**運行者**(線路を含む施設の設置・管理者)は、線路上に人が立ち入れないようにしておく**義務**がある
 - **ブレーキのないトロッコ**の機器自体が安全性を欠いた欠陥製品として**製造物責任**がある
 - 当事者として、**被害者**と被害を防げたかもしれない**管理者**や**メーカー**、**行政**や**警察**の**責任**も問題になるかもしれない

5. AIの時代と法(小塚2019)-2

- 法制度に対する期待と不満-2
 - 社会の変化と法の考え方(pp.17-21)
 - 法が技術に追いついていないからではなく、技術がどのように発展しても、想定外の事態が出現しても、社会の中で納得されるような解決を導くための筋道としてのルールを定めているのが法なのである
 - 新しく発生してくる問題に対して、適切な解決を導き出すための枠組みを用意しなければならない
 - AIを中心とする新しいテクノロジーの進化は3つの変化をもたらす
 - ① 経済活動がモノの取引からサービスの取引へと移行
 - ② 取引の対象としての財物からデータへの変化
 - ③ 取引のルールが技術的な仕組みにより決まってしまう
⇒ 契約から技術への変化

5. AIの時代と法(小塚2019)-3

- 間違わないAIの問題(pp.151-160)-1

- AIと差別

- 与えられた「教師データ」それ自体にバイアスがあれば、AIはバイアスのかかった法則性を学習してバイアスのかかった判定をする
- AIの「適正学習の原則」が要請されるが、現実社会にバイアスが存在しているとそれが学習内容に反映されてしまう可能性がある
- そうしたバイアスが差別を生むことにつながる
⇒ AIの判断を人間が確認し、差別とみられるようなバイアスのかかった判定は修正するようなステップを踏むほかはない

5. AIの時代と法(小塚2019)-4

- 間違わないAIの問題(pp.151-154)-2

- 信用スコアリング

- 中国アリババ子会社の「芝麻信用」は、個人の「身分特質」、「履行能力」、「信用歴史」、「人脈関係」、「行為偏好」を点数化している
- 点数が上がるほどローンの金利が低くなり、数々の恩恵が受けられる
⇒ エアビーアンドビーも芝麻信用のスコアを採用している
- スコアリングの対象となる詳細情報は公開されないため、ユーザーは自身の選択が「望ましい」ものかどうか考えながら行動する
- 色々なものがスコアリングに含まれるようであれば、精神の自由さえ保障されない監視社会にさえなりかねない
- 中国の消費者に便利な生活をもたらした芝麻信用は、反面でそんな不気味さをも持っている

6. AIロボットは人格を持つか(西垣・他)-1

- 自律型機械は存在するか(p.54)
 - AIロボットを自律型機械(autonomous machine)と見なせば疑似人格として責任をとらすことができるか ⇒ 最大の焦点
- ロボット三原則(pp.55-57)-1
 - 多くのロボット工学者はアシモフの三原則でロボットを設計
 - 軍用ロボットでは、敵は「非人間」と定義できるのか
 - 実際の戦場では敵と味方の識別は困難
 - 複数の矛盾した命令を受けとったロボットは行動が困難
 - 倫理的判断が困難 ⇒ 自由意志による道徳的主体の「人格」が問われることになる

*アシモフのロボット工学三原則: 1. 人間に危害を加えてはならない
2. 人間の命令に従わなくてはならない
3. 自らの存在を守らなくてはならない

6. AIロボットは人格を持つか(西垣・他)-2

- 自律性を問い直す (pp.59-62) (西垣、2018,p.170)-2
 - 自律性 (autonomy) とはなにか
 - 広義の自律性: 他者の指令を受けず行動を可能にする特性
 - 理論的自律性 (theoretical autonomy) ← 前提条件
 - 狭義の自律性: 社会的な責任をとれる主体のもつ特性
 - 実践的自律性 (practical autonomy)
 - 理論的自律性は実践的自律性の前提である
 - AIロボットは実践的自律性をもつか
 - 人は自由意志をもち責任をとれるから実践的自律性がもてる
 - AIロボットは自由意志を持ち合理的に判断し責任が取れるのか
⇒ AIロボットが実践的自律性を持つことは期待できない

6. AIロボットは人格を持つか(西垣・他)-3

- 生物と機会の境界線 (pp.64-66)
 - 生物: オートポイエティック(自己-創出的)な存在
 - AIロボット: アロポイエティック(他者-創出的)な存在
 - 行為を予測することができる存在
 - 生物の行動は完全には予測できない ⇒ 自由意思をもつ必要条件
 - 人間には理論的自律性と自由意思が認められる
 - ⇒ 責任を問われる道徳的主体が出現する
 - ⇒ 実践的自律性をもつといえる
 - AIやロボットは他律(heteronomous)システム
 - 自律型機械は不正解な表現
 - ⇒ 適応型機械(adaptive machine)と呼ばれるべき

6. AIロボットは人格を持つか(西垣・他)-4

- ネオサイバネティクスへ (pp.70-73)
 - サイバネティクス (N.ウィーナー)
 - 動物と機械における制御と通信
 - 生物と機械を**統一的に扱う**工学技術 ⇒ しかし本質的**相異**がある
 - 生物は**情報を解釈して主観的世界**を構成
 - AI(コンピュータ)は外部から情報をそっくり入力できるが生物のように自分で**自分を創出**することはできない
 - 人格をもつという**誤解**
 - 人間はAIやロボットが主体的人格をもつと**思い込む**ことはありうる
 - しかし、AIやロボットが(主体的)**人格をもつ**ことはない
- 西垣・他への**反論** (松原、2018、p.142)
 - AIに**心は宿るのか**
 - **汎用AI**が実現したら我々はAIに**心が宿ると認識**する
 - 「心がある」と認識した方が**便利であれば**、心があると考えるよい

7. AI・ロボットと法(弥永真生・他より)－1

- **AIの概念・定義** (pp.6-7) (総務省、2017)
 - AIとは**AIソフト**及び**AIシステム**を総称する概念
 - **AIソフト**とは、**データ・情報・知識の学習**等により、利活用の過程を通じて自らの出力やプログラムを**変化させる機能**を有するソフトウェアをいう(例：**機械学習ソフトウェア**)
 - **AIシステム**とは、AIソフトを**構成要素**として含むシステム(例：AIソフトを実装した**ロボット**や**クラウドシステム**)
- **法の観点の必要性** (p.i)
 - ロボット・AIが人や社会にもたらす**影響**は無視できなくなっており、その結果、**法学の観点からの考察や検討が欠かせなくなっている**
 - 法制度をそのまま維持することは、人と人との関係や社会にとって**不都合な結果、不幸な結果**を招きかねない

7. AI・ロボットと法(弥永真生・他より)ー2

- **強いAIと弱いAI**(p.7) (サール,1984、井元,2017、松尾,2017、小塚,2019)
 - 現在研究されているAIは、意識や自我を有する人間と同じように見える振る舞いをするけれども、実はそのような意識や自我をもっていない**[弱いAI]**である
 - 人間と同じ意識や自我をもっている**[強いAI]**は**実現されていない**
 - **フレーム問題**(西垣,2018,p.23、松原・他,1989)
 - マッカーシーとヘイズが提唱したが定義は多数ある ⇒ AIが与えられた問題を解決しようとするとき、問題の**枠組み**(フレーム)をうまく確定できず、考慮すべき事柄が**どこまでも広がり**お手上げになりがちなこと
 - **記号接地問題**(西垣,2018,p.23)
 - AIにはコンピュータ内部の**記号とそれが表す意味内容**を結びつけられない

7. AI・ロボットと法(弥永真生・他より)ー3

- 汎用AIと特化型AI(p.7) (井元,2017、西垣,2018)
 - AI技術は特定タスクを行う特化型のAI技術であり、あくまで人間の能力を補完するものである
 - 囲碁や将棋で人間の知能がすでに及ばないAIも、..人間が目的や機能を設定する「特化型AI」だ ⇒ 自ら機能を汎化する「汎用AI」ではない
 - 汎用AIは概念的には「強いAI」と同じ範疇の用語とみてもよい
 - 前記のフレーム問題が解決されないと、汎用AIは実現出来ない(筆者記)

7. AI・ロボットと法(弥永真生・他より)－4

- **ロボットの定義**(p.5) (JIS B0134:2015) ⇒ (平野,2017,p.55)
 - 2つ以上の軸について**プログラムによって動作**し、ある程度の自律性を持ち、環境内で動作して**所期の作業**を実行する運動機構
 - **現在の定義**:<感知/認識>+<考え/判断>+<行動>の循環を有する機械(人造物)
- **AI開発ガイドライン**(pp.19-22) (総務省,2017) -1
 - 開発者がガイドラインを**自主的に遵守**しつつ、それによってロボット・AIが**社会的に受容**されることを**国際的な方向生**として目指す
 - **AI開発原則-1**
 - **連携の原則**:開発者は、AIシステムの**相互接続性**と**相互運用性**に留意する
 - **透明性の原則**:AIシステムの入出力の**検証可能性**及び判断結果の**説明可能性**に留意する

7. AI・ロボットと法(弥永真生・他より)-5

- AI開発ガイドライン(pp.19-22)(総務省、2017)-2
 - AI開発原則-2
 - 制御可能性の原則: AIシステムの制御可能性に留意する
 - 安全の原則: AIシステムがアクチュエータ等を通じて利用者及び第三者の生命・身体・財産に危害を及ぼすことがないように配慮する
 - セキュリティの原則: AIシステムのセキュリティに留意する
 - プライバシーの原則: AIシステムにより利用者及び第三者のプライバシーが侵害されないよう配慮する
 - 倫理の原則: AIシステムの開発において、人間の尊厳と個人の自律を尊重する
 - 利用者支援の原則: AIシステムが利用者を支援し、利用者選択の機会を適切に提供することが可能となるよう配慮する

7. AI・ロボットと法(弥永真生・他より)－6

- AI開発ガイドライン(pp.19-22)(総務省、2017)－3
 - AI開発原則－3
 - アカウンタビリティ(説明責任)の原則:利用者を含むステイクホルダーに対しアカウンタビリティを果たすよう努める
 - AIネットワーク化と知連社会
 - AIネットワーク化:人間とAIネットワークが共生し、人間社会のあらゆる場面においてシームレスに連携していくこと
 - 知連社会:人間がAIネットワークと共生し、データ・情報・知識を自由かつ安全に創造・流通・連結して「知のネットワーク」を形成することにより、あらゆる分野におけるヒト・モノ・コト相互間の空間を超えた協調が発展し、創造的かつ活力ある発展が可能となる人間中心の社会像

7. AI・ロボットと法(弥永真生・他より)ー7

- AIと人間社会に対する**法的論点**(p.25)(総務省、2017)
 - ① 人工知能技術による**事故等の責任分配**の明確化と**保険**の整備
 - ② **個人情報とプライバシーの保護**も含めた**ビッグデータ**利活用
 - ③ 人工知能技術を活用した**創作物の権利**の検討
 - ④ 法解釈、法改正、法に関する**基本的概念の再検討**の可能性
- **ロボット・AIによる法のパラダイムシフト**(p.26)
 - **自律的AI**(もし登場したら)やそれを**組み込んだロボット**にも**法人格**や**人権(ロボット権)**を認めるべきという議論
 - AIは**予測不能**なので**使用者**には**結果責任**を追及すべきという議論
 - 自律的AIは**フィクション**なので、ロボット・AIの問題にも同じように対応すれば**足りる**という議論

7. AI・ロボットと法(弥永真生・他より)ー8

- 「ロボット・AI法」の**独自性**(p.27)
 - 「**ロボット・AI法**」のような独自の法(学)分野が必要になるのかそれとも**既存の法**(学)分野にロボット・AIのもたらす法的問題は**解消**されるのかの論争
- **人格なき社会**への展望-1(p.74)
 - 近代の**法システム**が前提していた**<意思ー行為ー責任>**という連関が、ロボット・AIの登場によって**揺るが**されているという問題
 - AIが**自己決定にゆらぎ**をもたらし、ロボットが社会を構成するメンバーの**行為者性を不明確**にすることによって、**・・・責任とは何を基礎とし・どのようなものとして・どの範囲に生じるのか**という問題に**影響**が生じる
 - **責任問題**の解決策として**結果からのみ責任**を導こうとする方向がある

7. AI・ロボットと法(弥永真生・他より)-9

- 人格なき社会への展望-2(pp.74-75)(山本、2018,p.405)
 - 動機や意図、故意・過失の存在を問うことなく、いかなる結果が生じたかのみを基礎とした責任追及システムを考えること
 - 純粹結果責任という考え方は刑事法の分野で、あまり喜ばしくない記憶として知られている
 - AI・ロボットの自己決定と行為責任を問題にしない立場は法領域全体にわたる結果無価値論の復権・拡大を意味することになる
 - 行為主体の内在的な傾向性を問題にする新派刑法学が国家による介入の早期化をもたらしたことを想起するべきだろう
⇒危険の未然防止へ、さらには危険の存在を予想させるに過ぎないリスクの段階で介入しそれを根こそぎに除去しようとする事前配慮へと進行することになる
 - AI・ロボットの登場を通じて展開されるのは、我々の社会の構造的変化である

7. AI・ロボットと法(弥永真生・他より)-10

- 自動運転と民事責任(pp.169-173)
 - 交通事故に関する民事責任
 - ドライバーの一般不法行為責任
 - 交通事故による人身損害では、自動車損害賠償保障法によって運行使用者責任と自動車損害賠償責任保険の締結強制を中心とした被害者保護の仕組みが整備
 - 自動運転車についても民法上の不法行為責任と自賠法上の運行責任者責任による民事責任の枠組みを維持できるのかが問題
 - 自動運転車メーカーの製造物責任をどのように考えるかも問題
 - 自動運転と法
 - 自動運転レベル4(特定領域での完全自動運転)以上は現法律の改定が必要
 - 自動運転は本発表の中心ではないのでこれ以上は省略

7. AI・ロボットと法(弥永真生・他より)-11

- **ロボット兵器と国際法**(p.286)
 - **ロボット兵器の使用理由**(4D任務)
 - **危険な任務**(Dangerous)
 - **汚い任務**(Dirty):核、生物、科学兵器で汚染された環境での任務
 - **単純で単調な任務**(Dull):長時間勤務の疲労、精神的弛緩に繋がる任務
 - **縦深性のある任務**(Deep):可能な活動範囲を超える任務
 - **ロボット兵器の使用目的**
 - 4D任務をロボット兵器に代替させることにより**犠牲者が出ない**こと
 - 自国兵士の**死傷に嫌悪感**を示す国内事情に望ましい事態であること
 - **操縦士用の空間や防護装置が不要**となること
 - 兵器の**小型化、軽量化**による**製造経費の削減**と**発見される危険性が低下**すること
 - 人間より**慎重に武力行使**することが可能であること

7. AI・ロボットと法(弥永真生・他より)-12

- ロボット兵器への不安(p.287)
 - ロボット兵器の自律性技術の進化
 - 米無人機の空母からのカタパルト発艦:2013.5
 - 同無人機の空母への着艦:2014.7
 - 自律型四足歩行ロボット(LS3)の物資輸送の実施:2014.7
 - 完全自律兵器(殺人ロボット)が出現するのではないか
- ロボット兵器の現状(p.289)-1
 - 遠隔操作ロボット:操縦士による遠隔操作
 - 爆発物処理ロボット(小型UGV*)、目標物攻撃ロボット(大型UGV)
 - 無人偵察車両
 - 軍用無人水上艦艇
 - 無人戦闘機(ミサイル搭載のドローンなど)

*UGV:Unmanned Ground Vehicle(無人走行車両)

7. AI・ロボットと法(弥永真生・他より)-13

- ロボット兵器の**現状**(pp.289-290)-2
 - **半自律**ロボット:人間の監視下で標的の探知・選択・攻撃を自律的に実行
 - 対ロケット・**野戦砲**・**迫撃砲**システム
 - **ミサイル**防御システム
 - 海上の**イージス**・システム
 - 陸上の**パトリオット**・システム
 - 終末高高度**領域防衛**システム
 - 上昇カプセル**魚雷型機雷**
 - **無人戦闘機**(自爆する特殊な**武装ドローン**)
 - 上記の兵器は人間を攻撃対象としない**対物破壊兵器**
⇒ 国際法上の**合法兵器**

7. AI・ロボットと法(弥永真生・他より)-14

- ロボット兵器の法規制動向(pp.291-293)
 - 2009年: NGOロボット軍備管理国際委員会(ICRAC)を設立し、武装自律無人システムの開発、配備及び禁止を提唱
 - 2012年: 完全自律兵器は国際人道法を遵守できず、拘束力ある国際法文書で完全な自律兵器の開発、製造及び使用を禁止すべき(ハーバードロースクール)
 - 2012年: 米国防総省は自律・半自律兵器システムは、指揮官及び操縦士が武力行使に対する適切なレベルの人間の判断を行使できるように設計すべき
 - 2017年: 上記国防総省の指令が恒久的方針となる
 - 致死性自律ロボット(LARs)(p.293)(小林、2017)
 - 2013年: 国連文書に致死性自律ロボットが明記
⇒ すべての国家に、致死性自律ロボットに関する国際法的枠組みが確立するまで、その開発、生産、取得、使用に関する履行の延長を宣言・履行するよう勧告

7. AI・ロボットと法(弥永真生・他より)-15

- ロボット兵器の国際法上の議論(pp.295-296)(岩本、2014)
 - 自律兵器システム、致死性自律兵器システムの国際的合意の定義はない
 - 自律兵器システム:一旦起動すれば、人間の操縦士によるさらなる介入なく標的を選択し攻撃することができる兵器システム
 - 自律とは、人間による監視または制御に依存することなく一連の行動を決定することができる能力
 - この定義に部分的に該当する自律兵器システムはすでに存在
⇒ 完全自律兵器システムはまだ存在しない
- 国際人道法の適用(p.296)
 - 致死性自律兵器システムは国際法、特に国際人道法の適用を受ける
 - 国家は、新兵器の研究、開発、取得または採用にあたり、その使用が国際法に禁止されているか否かを評価する義務がある

7. AI・ロボットと法(弥永真生・他より)-16

- 倫理的考慮(pp.301-302)

- ロボット兵器が倫理的に受け入れ可能か否かという根本的な非法的問題がある
 - ⇒ 倫理的観点から合法的な標的(戦闘員など)であっても、その標的を殺害する方法が問われる
- 人間の生死を決定するのは機械ではなく人間によってでなければならない
 - ⇒ 倫理的判断を含む人的関与のない機械には殺されたくないという命題
- ローマ法王庁は、生死の決定における人的要素は決して置き換えることができないと主張

7. AI・ロボットと法(弥永真生・他より)-17

- 自律兵器に対するアンケート(p.302)
 - 世界54か国実施のアンケート(2013年)
 - 67%: 致死性自律兵器システム(LAWS)を国際的に禁止すべき
 - 56%: LAWSは開発または使用すべきでない
 - 85%: LAWSは攻撃目的に使用すべきでない
 - 71%: LAWSでなく遠隔操作兵器システム(ROWS)を使用すべき
- 倫理的考慮の議論(p.302)
 - 倫理的判断のできないロボット兵器に人的関与または人間による制御の必要性に繋がる

おわりに

- AIについて倫理的側面から検討する本報告では、人の倫理とは異なる要素の考慮が必要であること(以下)がわかった
- それはAIを内蔵したシステムの自律性ということである
⇒ 完全な自律を達成しているAIシステムはまだないが、半自律のAIシステムはある
- AIシステムが犯罪を起こした場合、AIシステム自身に責任をとらすことができるか
⇒ 完全な自律でない限り責任を問うことはできない
- しかし、自律型ロボットシステムでは、完全自律ではないが、戦闘における責任には議論がある
- 結局、自律型ロボットシステムを兵器として開発、使用すべきではないという議論が過半数を占める
- 戦闘における兵器もAIシステム技術の発展で確固たる結論にはいたっていない
⇒ しかし、事前に議論しておくことが大切である

ご清聴いただき
どうもありがとうございました

コロナに負けずにガンバりましょう！



補助資料-1

- 人工知能に関する理事会報告(“Recommendation of the Council on OECD Legal Instruments Artificial Intelligence”, 2019)(OECD)
 - 5つの勧告
 - ① AIの研究開発への投資
 - ② AIのためのデジタル・エコシステムの整備
 - ③ AIを推進するための政策課題の形成
 - ④ 信頼できるAIのための国際協力
 - 信頼できるAIの責任あるスチュワードシップ(受託責任)のための原則
 - ① 包摂的な成長、持続可能な開発及び幸福
 - ② 人間中心の価値観及び公平性
 - ③ 透明性及び説明可能性
 - ④ 頑健性、セキュリティ及び安全性
 - ⑤ アカウンタビリティ

補助資料-2

・欧米日AI倫理ガイドラインの比較（上村、2018）

No.	国	組織・団体	ガイドラインの名称	時期
1	英	工学物理科学研究会議	ロボット5原則	2010年9月
2	英	英国規格協会	ロボティクス規制のガイドライン	2016年4月
3	英	下院 科学技術委員会	ロボティクスと人工知能	2016年9月
4	欧	欧州委員会	ロボティクス規制のガイドライン	2014年9月
5	欧	欧州議会 法務委員会	ロボティクスにかかる民主規則の欧州委員会への提言	2016年5月
6	米	スタンフォードAI 100	2030年のAI	2016年9月
7	米	パートナーシップ・オン・AI	信条	2016年9月
8	米	ホワイトハウス	人工知能の未来に備えて	2016年10月
9	米	IEEEグローバル・イニシャティブ	倫理的設計Ver.1	2016年12月
10	米	ホワイトハウス	AI自動化、そして経済	2016年12月
11	米	Future of Life Institute	アシロマAI 23原則	2017年2月
12	日	総務省 AIネットワーク化検討会議	報告書2016	2016年6月
13	日	人工知能学会 倫理委員会	人工知能学会 倫理指針	2017年2月
14	日	内閣府 懇談会	人工知能と人間社会に関する懇談会報告書	2017年3月
15	日	総務省 AIネットワーク社会推進会議	報告書2017	2017年7月

補助資料-3

• AI倫理規定の現状(芳田、2019)

No.	組織・団体	規定・指針の名称	公表時期
1	Future of Life Institute	アシロマの原則 (Asiloma AI Principles)	2017年2月
2	人工知能学会	人工知能学会 倫理指針	2017年2月
3	総務省 AIネットワーク 社会推進会議	AI開発ガイドライン案 AI利活用に関する原則案	2017年7月 2018年7月
4	IEEE (米国電気電子学会)	倫理的に配慮されたデザイン第2版 (Ethically Aligned Design Ver.2)	2017年12月
5	欧州委員会	Ethics Guidelines for Trustworthy (AI倫理指針)	2018年12月 Draft 2019年3月 Final
6	内閣府 人間中心のAI社会 原則検討会議	人間中心のAI社会原則	2018年12月 Draft 2019年3月 Final

参考文献-1

1. IPA『AI白書2019』情報処理推進機構、2019.
2. OECD「人工知能に関する理事会勧告(総務省訳)」, "Recommendation of the Council on OECD Legal Instruments Artificial Intelligence", 2019.
3. 池邊純一『シンギュラリティに関する情報の提供』組織のダイナミズム研究会2018年5月例会発表原稿、2018.
4. 井元剛『平和を願う人工知能』日本工業新聞社、2017.
5. 岩本誠吾「致死性自律型ロボット(LARs)の国際法既製をめぐる新動向」『産大法学』47巻3・4号、2014.
6. 上村恵子、他「日米欧の地域特性に着目したAI倫理ガイドラインの比較」『人工知能学会32回大会配布資料』人工知能学会、2018.
7. 太田裕朗『AIは人類を駆逐するのか: 自律世界の到来』幻冬舎、2020.
8. 小塚壮一郎『AIの時代と法』岩波書店、2019.
9. 小林雅一『AIが人間を殺す日』集英社、2017.
10. サール,J.、坂本百大監訳「心・脳・プログラム」ホフスタッター編著『マインズ・アイ(下)コンピュータ時代の「心」と「私」』ティービーエス・ブリタニカ、1984.
11. 人工知能学会『人工知能学会 倫理指針』人工知能学会倫理委員会、2017a.
12. 人工知能学会「人工知能と倫理」『人工知能学大事典』共立出版、2017b.
13. 総務省「国際的な議論のためのAI開発ガイドライン案」『AIネットワーク社会推進会議』別紙1、2017.
14. 総務省『AIネットワーク社会推進会議 報告書2019(案)』AIネットワーク社会推進会議、2019a
15. 総務省『AI利活用ガイドライン』AIネットワーク社会推進会議、2019b
16. 総務省, "Comparison of AI guidelines", 『AI利活用ガイドライン報告書2019』別紙、2019c.
17. 種村剛「倫理/倫理学」『事項リスト』<http://tanemura.la.coocan.jp>、2019.7.29.
18. 内閣府『人工知能と人間社会に関する懇談会報告書』、2017.

参考文献-2

19. 内閣府『人間中心のAI社会原則』統合イノベーション戦略推進会議、2019.
20. 長尾真『人工知能と人間』岩波書店、1992.
21. 西垣通『AI原論』講談社、2018.
22. 西垣通・河島茂生『AI倫理：人工知能は責任をとれるのか』中央公論社、2019.
23. 日経新聞『NEC、AI活用の指針を策定』電子版、2019.4.2.
24. 日経新聞『ソニー、AIの倫理を審査 差別や乱用回避』電子版2020.12.22, 2020a.
25. 日経新聞『AI倫理とは 人類に悪影響あたえないための規範』電子版2020.12.23, 2020b
26. 日経新聞『AI倫理規制、欧州が先陣 日本も対応急務』電子版、2021.1.16,2021.
27. 日経新聞『日立、AI開発に倫理原則 外部意見取り入れ改善』電子版、2022.2.22,2022.
28. 平野晋『ロボットと法』弘文堂、2017.
29. 富士通「AI倫理への取り組み」『特集 研究開発最前線：デジタル時代の信頼』Vol.70,No.4,PP.14-20,2019.
30. 松尾豊編著『人工知能とは』近代科学社、2016a.
31. 松尾豊・他「人工知能と倫理」『人工知能学会共同企画第3部「技術紹介」』人工知能Vol.31,5号、人工知能学会、2016b.
32. 松尾豊「強いAIの前に弱いAIでできること」鳥海不二夫『強いAI・弱いAI』丸善出版、2017.
33. 松田雄馬『人工知能に未来を託せますか？』岩波書店、2020.
34. 松原仁・橋田浩一『情報の部分性とフレーム問題の解決不能性』1989.
35. 松原仁『AIに心は宿るのか』集英社、2018.
36. 弥永真生・宍戸常寿編『ロボット・AIと法』有斐閣、2018.
37. 山本龍彦編著『AIと憲法』日本経済新聞社、2018.
38. 芳田千尋「AI倫理に関する現状」『UNISYS TECHNOLOGY REVIEW 第139号』日本ユニシス、2019.