AIと倫理

組織のダイナミズム研究会

2019年8月24日(8月例会)

伊東俊彦（城西国際大学大学院、
組織のダイナミズム研究会代表）

AI_and_ethics9b.pdf(ダウンロードファイル)

目　次

1．自己紹介

- 生誕：1946年⇒ENIAC誕生
- 大学1．1969年：電気通信工学科(武蔵工業大学)卒業
- 就職1，1969年：日本NCR入社⇒技術員⇒テクニカルアシスタント
- 講師1．1982年：日本電子専門学校非常勤⇒論理回路初級･応用
- 就職2．1983年：日本DEC入社教育部⇒コンサルティング部⇒システム統括本部
- 学会：経営情報学会理事、監査人歴任、各研究部会代表、幹事歴任
- 講師2．1995年：学習院大学経済学部経営学科非常勤
- 就職3．1998年：コンパックコンピュータ入社（DECを買収）
- 大学教員1．1999年：学習院大学経済学部経営学科特別客員教授
- 大学2．1999年：MBA（青山学院大学大学院国際政治経済学研究科）
- 大学3．2001年：横浜国立大学大学院国際社会科学研究科博士後期課程(経営学博士)
- 大学教員2．2003年：愛知淑徳大学コミュニケーション学部⇒ビジネス学部教授
- 大学教員3．2005年：東北大学大学院経済学研究科教授
- 講師3．2005年−至現在：城西国際大学大学院ビジネスデザイン研究科ゼミ担任
- 顧問　2009年−至現在：エスツー（秋田、仙台）経営顧問
- 講師4．2013年−至現在：大和市シルバー人材センター・パソコン講師

2. 表タイトルについて-1

- 「AIと倫理」をタイトルにした理由
  - 現在、AI使用によるリスク増加が危惧されている（松尾、2016b）
    - たとえば、テスラの自動運転車の死亡事故
    - MSのチャットボットがヒトラーを礼賛
  - シンギュラリティ*はまだだが、AIの社会への悪影響の未然防止が叫ばれている（松尾、2016b）
    - EU、日本などではAIに対する倫理に関する制定が進められている
  - 企業組織でAI利用が日増しに増えている環境
    - 技術者倫理、プロフェッショナル倫理で無く「AI倫理」についてODSGメンバーの議論を進めるためのきっかけ発表としたい
    - まずはAIについて基本認識から始めたい

*シンギュラリティ（技術的特異点）：我々の生物としての思考と存在が、自らの作り出したテクノロジーと融合する臨界点、2045年（カーツワイル、2016）

2. 発表タイトルについて-2

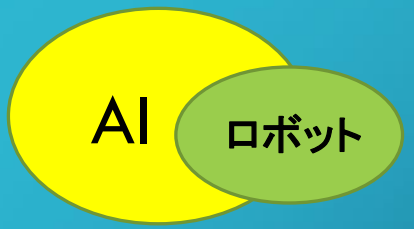
- 「AIの倫理」でなく「AIと倫理」なのか
  - 「AI*の倫理」をテーマにすると、倫理の範疇でAIを議論する傾向を強める⇒それでは抜けがあるかも
  - 「AIと倫理」と並列にすることで、AIと倫理の関係はどうなのか、AIの発展に伴い倫理も変えるべきか、
    など、広い視点でみることができるのではないか
  - 「AI倫理」と「AIの倫理」は本稿では同等のものとして扱う

*用語の「AI」と「人工知能」は本稿では同等のものとして扱う

3．AIとはなにか-1

- AIの誕生
  - 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
  - 1958年：MITにAI研究所が創設（ミンスキー、マッカーシー）
  - 1963年：スタンフォード大にAIラボが創設（マッカーシー）
- AIの定義（松尾、2016a）
  - AIの定義について研究者の中で明確な定義が定まっていない
  - 世の中の定義：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（松尾、2016a）
  - AIが普及すると人工知能とは言われなくなる（松尾、2016a）
    - 自動翻訳、検索エンジン、文字認識、音声認識、将棋・チェスのプログラム、推薦エンジンなど

3．AIとはなにか-2



- AIとロボット
  - ロボットは身体性*をもつことが多いが、人工知能とイコールではない
- AIの歴史
  - 第1次ブーム1960年代：推論・探索により特定の問題を解く研究
  - 第2次ブーム1980年代-90年代前半：エキスパートシステムによる実用化
  - 第3次ブーム2010年代以降：機械学習とディープラーニングによる発展⇒AlphaGo（2015年 Deep Mind社）

*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、環境相互作用に構造を与え、認知や行動を形成する基盤となること

３．AIとはなにか−2

- AIの種類−1
  - 弱いAI−強いAI、特化型AI−汎用型AI
    - 弱いAI（井元、2017、サール、1980）
      - 人間が設定したルールに従ってデータを処理するもの
      - ルールベース・アプローチによるAI
      - 機械学習（ディープラーニングを含む）、エキスパートシステム
      - ディープラーニングは強いAIそのものにはならない（松尾、2017）
    - 強いAI（井元、2017、サール、1980）
      - 機械が自己学習するもの
      - 機械自身がルールを決め、ニューロモデルを使い自己学習するもの
      - 人間を超える普遍的な知・・・それが強いAIということになる。しかし、神を冒涜する罪深い業という怖れがある（西垣通、2017）

3．AIとはなにか−4

- AIの種類−2（井元、2017）
  - 特化型AI
    - 特定の用途やタスクに特化したAI
    - 現在のすべてのAI（弱いAIとほぼ同様）
  - 汎用型AI（AGI）
    - あらゆる面で人間と同等かそれ以上の知性をもつAI（強いAIとほぼ同様）
    - シンギュラリティ*が起こる時点と微妙には違う
    - AIが自ら汎用型AIを作る
    - まだ実現していないが、いずれ機械に意志が宿る

*シンギュラリティ：我々の生物としての思考と存在が、自らの作り出したテクノロジーと融合する臨界点（R. Kurzweil)

4．AIのリスク-1

- AI使用によるリスクの例
  - 2016年：テスラ・モーターズの自動運転車がフロリダで世界初の死亡事故（松尾、2016b）
  - 2016年：マイクロソフトのチャットボット”Tay“がヒトラーを礼賛（松尾、2016b）
  - 人種や性別などの偏ったデータをAIが読み込み、差別的な分析が増える懸念も出ている（日経新聞、2018）
  - 技術は開発者が意図しない使われ方により、社会に深刻な損害をもたらす危険性がある（松尾、2016b）
  - 日本政府はAI人材25万人/年育成を目標に（日経新聞、2019）

4．AIのリスク–2

- AIのもつリスク
  - 機械が部分的に人間を超えることはすでに起こっている
    - コンピュータの計算能力、インターネットによる知識の量（松尾、2016b）
  - 自らを改変しさらに良いものを生み出すAIへの脅威
    - 人工知能が自ら目的を持ち、人類に危害を加えるような世界が現実となる蓋然性は今後数十年は極めて低い（内閣府、2017）
    - 将来シンギュラリティが起こることはほぼ間違いない（井元、2017）
    - AIは核と同じ、2面性をもつテクノロジー（バラット、2015）
- AIに関わる人間のリスク（松尾、2016b）
  - 科学技術のデュアルユース問題
    - 爆弾処理ロボットに爆弾を仕掛け、立てこもり犯人を爆死
  - AIを使う･研究開発する人間がもつ価値観（倫理観）が重要

4．AIのリスク-3

- 失業などの社会的インパクト（松尾、2016b）
  - AIにより今後10年でなくなる仕事が約半数
  - 職がなくなるのではなく、タスクがなくなることによる仕事内容の再定義がされる
  - 反論：ディープラーニングとものづくりの掛け合わせによる日本の産業競争力向上の可能性もある
  - 教育のあり方が変わる
    - 学んだ知識･スキルの通用する期間は益々短くなる
      ⇒人生全体を通じて学び続けることが重要

4．AIのリスク-4

- 法律や社会のあり方に関する問題（松尾、2016b）
  - AIが創造性をもった場合
    ⇒膨大な知的財産の独占が起きるかもしれない
  - 機械学習で他人が権利をもつデータを使い学習して得られたモデルの権利は誰のものになるか
  - 自動運転車の責任は運転者にある
    ⇒自動運転車に搭載されるAIを「運転手とみなす」米運輸省
    ⇒事故の責任が製造者責任とされる場合もある
  - AIを「人工人」とする議論がある⇒「人工人格」
  - AI技術が倫理学、道徳心理学、法哲学などを巻き込んだ議論を引き起こしている
  - 「AI時代の憲法論」⇒人工知能に人格はあるか

4．AIのリスク–5

- AIの究極目標（松尾、2016b）
  - AIを人類のサステナブルな未来のために活用すること
- AIリスクの悪影響を未然に防ぐ（松尾、2016b）
  - AIや新技術による負の側面による社会への悪影響を未然に防ぐ議論が必要
  - 新技術導入以前の研究開発の段階から、なんらかの規則やガイドラインを設けること
- AIと人間の議論
  - 自動運転AIの方が人間より事故回避できる確率は高い（浅川、2017）

5. AI倫理とはなにか-1

- 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）
- 倫理の原語（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
  - アリストテレス「ニコマコス倫理学」の倫理的徳（ethike arete）

5．AI倫理とはなにか-2

- 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倫理に関する議論
  - 技術者倫理、プロフェッショナル倫理の中で議論すべき、
    ⇒AI倫理として独立すべきでない
  - AIに固有の倫理問題はどれほどあるのか（人工知能学会、2017b）
    - 例：完全に自律的な兵器が登場した場合、当兵器の誤爆で死者が出たら、兵器自身以外に責任を負わせうる行為者が存在しない
      ⇒しかし人工知能の責任などという問題を倫理学の観点で扱うことは少なくなった（人工知能学会、2017b）

5. AI倫理とはなにか-3

- ロボット倫理-1
  - ロボットの定義
    - 2軸以上がプログラム可能で、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）
    - 身体性*をもつとは限らない⇒ロボット掃除機
  - ロボットの倫理の議論-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
    - ロボットにも道徳的配慮が必要か⇒ロボットの廃棄問題

*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、環境相互作用に構造を与え、認知や行動を形成する基盤となること

5. AI倫理とはなにか-4

- ロボット倫理-2
  - ロボットの倫理の議論-2（久木田、2017）
    - アシモフのロボット工学三原則
      - 1. ロボットは人間に危害を加えてはならない
        ⇒人間が危害を受けるのを黙視してはならない
      - 2. ロボットは人間の命令に従わなくてはならない
        ⇒ただし、第1原則に反する命令はその限りではない
      - 3. ロボットは自らの存在を護らなくてはならない
        ⇒ただし、第1・第2原則に違反しない場合に限る
    - ロボット工学三原則の限界
      - AIの分野においては、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

5. AI倫理とはなにか-5

- ロボット倫理-3
  - 倫理学の三種の下位分野とロボット
    - 規範倫理学：倫理上の正や善の基準にかかわる理論を扱う
      ⇒ロボットに実装するアルゴリズムのよりどころ
    - 応用倫理学：医療や環境のような現実の社会問題を扱う
      ⇒生命倫理学、環境倫理学
      ⇒医療などの実践的場面で使われるロボットの設計
    - メタ倫理学：正や善などの基本的概念についての分析や検討
      ⇒ロボットのセンサーで正しさなどを計測する際

5. AI倫理とはなにか-6

- AIで検討しておくべき倫理問題の分類（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

5. AI倫理とはなにか-7

- AI倫理の範囲
  - AIシステムを開発する技術者の倫理（弱いAI、強いAI）
  - AIシステムが実現するシステムの倫理的対応（弱いAI、強いAI）
  - 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、環境相互作用に構造を与え、認知や行動を形成する基盤となること

6. AI倫理の現状-1

- AIと倫理の議論
  - 2014年：人工知能学会倫理委員会の議論開始（松尾、2016b）
  - 2016年：「人工知能と人間社会に関する懇談会」開始（内閣府）
  - 2018年：EUで「GDPR：一般データ保護規則」の施行開始
  - 技術的な進歩でサービスがよくなるとデータが流動化
    ⇒そのデータの悪用を防ぐためにGDPRは必要（藤井、他、2019）
  - EU委員会は2018年末までにAIの倫理指針を策定
- 当研究会の対応（筆者の認識）
  - 企業組織として、AIと倫理の議論の概要をおさえる必要がある

6. AI倫理の現状-2

- 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 Traustworthy （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

6. AI倫理の現状-3

- AI倫理規定の現状（特徴）（芳田、2019）
  - 1 FLI（Future of Life Institute）
    - CA州アシロマで人類にとって有益なAIとは何かの議論のまとめ
  - 2 人工知能学会
    - 研究者である会員が守るべき指針
  - 3 AIネットワーク社会推進会議（総務省）
    - AIの開発に関する9原則案、利活用に関する10原則案
  - 4 IEEE(米国電気電子学会)
    - 技術者が人間に恩恵をもたらすテクノロジーの進歩のため倫理的に配慮されたデザイン案をもとめた
  - 5 AIリスクの最小限とAI恩恵の最大限には
    - 人間中心アプローチで信頼できるAIの実現を目指す⇒10の必要条件
  - 6 人間中心のAI社会原則検討会議（内閣府）
    - AI利用は基本的人権を守るべき⇒人間中心の原則⇒7原則

6. AI倫理の現状-4

- AI倫理規定の現状（共通点）（芳田、2019）
  - 1. 人間中心のAIの実現
  - 2. 人間の社会的・道徳的規範に相反しないAIの設計・実装
  - 3. 安全性の確保
  - 4. AIシステムの判断過程についての説明責任
  - 5. 個人データやプライバシーの保護
- OECDのAIに関する原則（OECD、2019）
  - 2019年5月：OECD42ヵ国が新原則を採択

6. AI倫理の現状−5

- 欧米日AI倫理ガイドラインの比較−1（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	IEEEE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

6. AI倫理の現状-6

- 欧米日AI倫理ガイドラインの比較-2（上村、2018）
  - 欧州のガイドライン
    - 人の権利・責任に重点
    - ロボットのガイドラインで人間が責任主体
      - 電子人間という法的地位
    - GDPR(一般データ保護規則)：データ、プライバシーの厳格化
  - 米国のガイドライン
    - AIによる社会便益の最大化を推進
    - 自律型兵器についての政策策定
    - AIの軍拡競争の禁止
    - 自律システムの透明性
    - データ、プライバシーの処理
    - AIの進展に適応させた規制

No.	国	組織・団体	ガイドラインの名称	時期
1	英	工学物理科研究会議	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	IEEEE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

6. AI倫理の現状-7

- 欧米日AI倫理ガイドラインの比較-3（上村、2018）
  - 日本のガイドライン
    - 人々の不安解消を目的とした倫理原則に重点
      - 透明性、制御可能性、安全性、プライバシー保護
    - 13では、研究者倫理に焦点
      - 作られた人工知能側にも人間同様に倫理指導を遵守
    - 自律型兵器についてふれていない
    - AIの適合性評価の認証制度提案は発展を妨げる理由から削除
    - 人型ロボットへの宗教的な抵抗が少ない
    - 人手不足の解決策として技術的対応に偏る
    - ルール作りが官主民従⇒ルールがないと慎重になる

No.	国	組織・団体	ガイドラインの名称	時期
1	英	工学物理科研究会議	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	IEEEE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

7．AI倫理への対応

- AI倫理を制定してもそれを守ることは困難
- AI技術者を倫理的な設計に巻き込む（関口、2018）
  - AI倫理ライブラリの実装
    - 各種のAI倫理の言説が持つ構造の違いを明確化
    - AI倫理の言説相互の意味的な距離の提示
    - AI倫理を繋げるようなシナリオパスの推薦
    - AIライブラリ技術：エディタ、クラウド環境、探求用エンジン
  - AI倫理ライブラリの提供
    - AI技術者が自身の研究開発とAI倫理を繋げることにより、AI技術者を倫理的な設計に巻き込むことができる
    - 今後具体例を増やし、AI倫理専門家との対話も進める

おわりに

- 企業組織でAI利用が増えている状況（RPA、DXなど）で、AI使用のリスクも重要視されている
- シンギュラリティは先としても、現時点からAIの使用やシステム開発におけるAI倫理への配慮が大切である
- 上記のような理由から、本日は「AIと倫理」の基調発表をした
- AI倫理の具体的な詳細内容は、アシロマAI 23原則などを挙げるのに留めたが他のAI倫理もこれに似たようなものである
- 最後に「AI倫理への対応」として「AI倫理ライブラリの構築と提供」を挙げたが、今後もAI技術者側支援システムの充実が必要である
- AI倫理という人間存在そのものを問われる哲学的なテーマに対し、我々個人の課題としても取り組むべきであると考える

- ご清聴いただきどうもありがとうございました

補助資料−1

- アシロマAI 23原則−1(東京海上、2017)
  - 1. 研究目標: AI 研究の目標は、方向性の定まらない知能ではなく、有益な知能の創造である
  - 2. 研究資金:AIへの投資には、AIの有益な利用のために必要な、コンピューターサイエンス、経済、法律、倫理、社会学などに関する、厄介な問題に関する研究費用も含めるべきである
  - 3. 科学と政策のリンク:AI 研究者と政策立案者の間で、建設的で健全な意見交換が行なわれるべきである
  - 4. 研究文化:AI 研究者や開発者の間で、協力、信頼、透明性の文化が育まれるべきである
  - 5. 競争の回避:AI システムの開発チーム同士は、競争のために安全基準を省略することがないよう、積極的に協力しあうべきである
  - 6. 安全性:AI システムはその運用期間を通して、可能な限り安全で堅牢で、検証可能でなければならない
  - 7. 障害の透明性:AI システムが障害を起こしたときは、その原因を確認できるようにするべきである
  - 8. 法的透明性:自動システムが法的判断に関わる場合、有能かつ権限を持つ人間が監査し、納得のいく説明ができるようにする
  - 9. 責任:先進的な AI システムの設計者と開発者は、システムの使用、悪用、結果に倫理的な関わりがある当事者であり、その関わりを 形作る責任と機会がある

補助資料−2

- アシロマAI 23原則−2(東京海上、2017)
  - 10. 価値観の一致:高度に自律的な AI システムは、目標と行動が倫理的に人間の価値観と一致するようデザインされるべきである
  - 11. 人間の価値:AI システムは、人間の尊厳、権利、自由そして文化的多様性に適合するよう設計・運用されなければならない
  - 12. 個人のプライバシー:AI システムが個人のデータを分析し、利用する力を持つ以上、データを生成する個人は自らのデータを閲覧、管理、コントロールする権利が与えられなければいけない
  - 13. 自由とプライバシー:AI による個人情報の利用は、人間が持つ、あるいは持つと思われている自由を不合理に侵害してはならない
  - 14. 利益の共有:AI 技術は、可能な限り多くの人々に利益と力をもたらすべきである
  - 15. 繁栄の共有:AI によって作られた経済的な繁栄は、人類すべてに利益をもたらすために、幅広く共有されなければならない
  - 16. 人間によるコントロール:人間が選択した目標を達成するために、AI システムに決定をどのように委ねるのか、あるいは委ねるか否 かは、人間が判断しなければいけない
  - 17. 転覆活動の防止:高度な AI システムをコントロールすることによって得られる力は、健全な社会に不可欠な社会的、市民的プロセス を転覆させるのではなく、尊重、促進するために使われなければならない
  - 18. AI の軍拡競争:破壊的な自動兵器による軍備の拡大競争が起きてはならない

補助資料−3

- アシロマAI 23原則−3(東京海上、2017)
  - 19. AI の能力:意見の一致がない以上、将来の AI の能力の上限に関する強い前提を置くことは避けるべきである
  - 20. 重要性:発達した AI は地球上の生命の歴史に重大な変化を及ぼす可能性があるので、相応の注意と資源をもって計画、管理されな ければならない
  - 21. リスク:AI システムによるリスク、特に壊滅的なものや存亡の危機をもたらすものに対しては、その影響に相応した慎重な計画と緩和 対策を行うべきである
  - 22. 再帰的自己進化:自己進化、または自己複製によって質的・量的に急激に拡大をもたらすよう設計されたAIは、厳格な安全管理対策 の対象とすべきである
  - 23. 共通の利益:超知能は、特定の国や組織のためではなく、広く共有されている倫理的な理想や、人類全ての利益に資するためにの み開発されるべきである

- アシロマAI 23原則のポイント(東京海上、2017)
  - 1. むやみに技術開発を競うのではなく、今開発している AI は人類全体にとって本当に有益かを考える
  - 2. AI の目標と行動は、人間の倫理観・価値観と一致するようデザインしなければいけない
  - 3. AI によってもたらされる経済的利益は、全世界で広く共有されなければいけない
  - 4. AI によって、人間の尊厳、権利、自由、文化的多様性が損なわれてはいけない
  - 5. 自己増殖機能を持つような AI の開発には、厳重な安全管理対策が必要である

補助資料−4

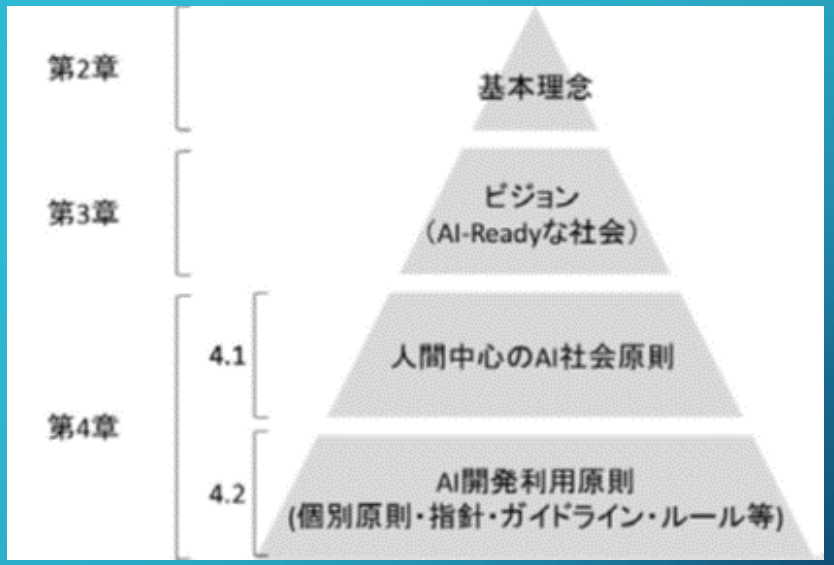
- AIに関するOECDの原則−1：その概要
  - AIは、包摂的成長と持続可能な発展、暮らし良さを促進することで、人々と地球環境に利益をもたらすものでなければならない
  - AIシステムは、法の支配、人権、民主主義の価値、多様性を尊重するように設計され、また公平公正な社会を確保するために適切な対策が取れる－例えば必要に応じて人的介入ができる－ようにすべきである
  - AIシステムについて、人々がどのようなときにそれと関わり結果の正当性を批判できるのかを理解できるようにするために、透明性を確保し責任ある情報開示を行うべきである
  - AIシステムはその存続期間中は健全で安定した安全な方法で機能させるべきで、起こりうるリスクを常に評価、管理すべきである
  - AIシステムの開発、普及、運用に携わる組織及び個人は、上記の原則に則ってその正常化に責任を負うべきである

補助資料-5

- AIに関するOECDの原則-2：各国政府に対する提言
  - 信頼できるAIのイノベーションを刺激するために、研究開発への官民投資を促進する
  - デジタルインフラとテクノロジーでAIエコシステムとデータと知識の共有メカニズムの利便性を高める
  - 信頼できるAIシステムの普及に道を開く政策環境を創出する
  - 人々にAIに関わる技能を身につけさせるとともに、労働者が偏りなく転職できるよう支援する
  - 情報を共有し標準を開発し、責任あるAIの報告監督義務を果たせるように、国際的、産業部門横断的に協力する

補助資料-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社会
    - 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
      - 校正競争確保の原則：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



補助資料-7

- 総務省『AI利活用ガイドライン』(総務省2019b)
  - 基本理念
    - 人間中心の社会を実現
    - 利用者の多様性の尊重
    - 様々な課題の解決と持続可能な社会の実現
    - 便益とリスクの適正なバランスの確保
    - 指針やベストプラクティスの国際的な共有
    - ガイドラインの普段の見直しと柔軟な改定
  - AI利活用原則
    - 適正利用の原則:適切な役割分担と適正な範囲と方法でのサービス利用
    - 適正学習の原則:学習に用いるデータの質に留意
    - 連携の原則:AIサービス相互間の連携
    - 安全の原則:第三者の生命・身体・財産への危害の排除
    - セキュリティの原則:セキュリティに留意
    - 尊厳・自律の原則:人間の尊厳と個人の自律
    - 公平性の原則:AIサービスの判断のバイアス留意、個人が不当に差別されない
    - 透明性の原則:検証可能性、説明可能性
    - アカウンタビリティの原則:ステークホルダへのアカウンタビリティを果たす

補助資料-8

- 『AI時代の憲法論：人工知能に人権はあるか』(木村、2018)より
  - AIは人間に準ずる尊厳をもつべきか
    - もつべきかどうかは皆さんで考えるべき
    - AIの尊厳は地域や宗教観で違う
      - キリスト教では、人間は特権的な存在⇒機械や動物は人間による支配の対象
        ⇒AIの所有者の権利としてAIの尊厳が限定
      - 日本は八百万の神や付喪神(長い年月を経た道具に宿る神)の世界
        ⇒古いまな板にも人格を感じる⇒AIに尊厳を見るのになんの躊躇もない
  - 意識をもった機械に人権はあるか
    - 機械に人を殺す権利を与えていいか
      - 完全自律型キラーロボットの危険性
      - 機械に人間を殺す判断をさせてよいのかという自律性
      - AIが判断ミスをした場合、システムエラーが起きた場合⇒どう責任をとるか
      - 論理的に冷静な判断ができるAIの方がベター
    - 自動運転と自律型兵器
      - 自動運転は人的被害の最小化⇒自律型致死兵器は敵側被害の最大化
  - AIが倫理観を学ぶには
    - AI同士、さらに人間の織りなす社会生活の中での経験をとおし倫理観を学ぶ
    - AIが十分かつ人間的な倫理性を身に付けたら⇒倫理判断をAIまかせに
      ⇒人間の倫理観の減退に

参考文献-1

1. IPA『AI白書2019』情報処理推進機構、2019.
2. OECD『42カ国がOECDの人工知能に関する新原則を採択』2019.5.22 https://www.oecd.org/tokyo/newsroom/、2019.
3. 浅川直輝「無知で未成熟な「AI脅威論」3つの誤解と本当の課題」『日経コンピュータ 2017.5.15』日経BP社、2017.
4. 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.
5. 井元剛『平和を願う人工知能』日本工業新聞社、2017.
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. カーツワイル,R、井上健監訳『シンギュラリティは近い：人類が生命を超越するとき』NHK出版、2016.
8. 木村草太編著『AI時代の憲法論：人工知能に人権はあるか』毎日新聞出版、2018.
9. 久木田水生、他『ロボットからの倫理学入門』名古屋大学出版会、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. 関口海良、他「AIはAI技術者を倫理的な設計に巻き込むことができるか?」『人工知能学会32回大会配布資料』人工知能学会、2018.
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

参考文献-2

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. 種村剛「倫理/倫理学」『事項リスト』http://tanemura.la.coocan.jp、2019.7.29.
17. 東京海上「AIの安全ガイドライン「アシロマAI 23原則」」『東京海上研究所ニュースレターNo.35 2017/03』東京海上研究所、2017.
18. 内閣府『人工知能と人間社会に関する懇談会報告書』同左懇談会、2017.
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. 西垣通「解説 シンギュラリティ仮説の背後にうごめくもの」ガナシア,J.G.伊藤直子監訳『そろそろ、人工知能の真実を話そう』早川書房、2017
21. 日経新聞『EU、AIに倫理指針 人種・性別の差別防ぐ』電子版、2018.11.6.
22. 日経新聞『政府、AI人材年25万人育成へ、全大学生に初級教育』電子版、 2019.3.27.
23. 野田ユウキ『図説 シンギュラリティの科学と哲学』秀和システム、2019.
24. バラット,J.、水谷淳訳『人工知能　人類最悪にして最後の発明』ダイヤモンド社、2015.
25. 松尾豊編著『人工知能とは』近代科学社、2016a.
26. 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.
27. 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.
28. 三宅陽一郎『人工知能のための哲学塾』ビー・エヌ・エヌ新社、2016.
29. 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.