人工知能は道徳的になりうるか あるいはなるべきか

人工知能学会全国大会
2015年5月30日-6月2日，公立はこだて未来大学
オーガナイズドセッション
「汎用人工知能とその社会への影響」

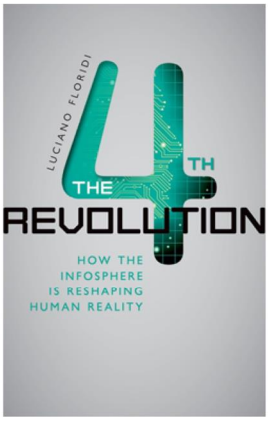
NAGOYA UNIVERSITY

久木田水生
名古屋大学情報科学研究科
社会システム情報学専攻

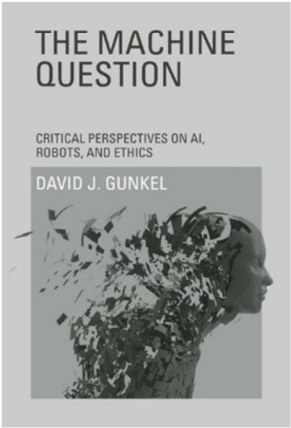
INFORMATION SCIENCE
NAGOYA UNIVERSITY

この資料は2015年5月30日-6月2日に公立はこだて未来大学で行われた人工知能学会全国大会のオーガナイズドセッション「汎用人工知能とその社会への影響」での発表時に使用したスライドに加筆修正したものです.

AGIの社会的影響：二つの革命



理性的存在者としての人間の地位



道徳行為者としての人間の地位

Luciano Floridiは，人間以外に人間に匹敵する知的能力をもつ存在者が現れることは，人間の自己認識を根本的に変えることになるだろうと言う．彼はこれをコペルニクスの地動説（人間は宇宙の中心ではない），ダーウィンの進化論（人間は他の動物と連続している），フロイトの無意識の理論（人間は必ずしも意識的合理的に決断・行動しているわけではなく，人間の行動の多くは無意識の欲望によって駆り立てられている）に次ぐ，第4の革命と呼ぶ．

David J. Gunkelはロボットや人工知能が道徳的行為者，あるいは道徳的配慮の対象となる可能性について考察している．そのような自体は私たちの道徳性の概念を大きく変えるだろう．

問いの重要性

http://www.stopkillerrobots.org/2015/04/humancontrol/

http://www.theguardian.com/technology/2014/may/28/google-self-driving-car-how-does-it-work

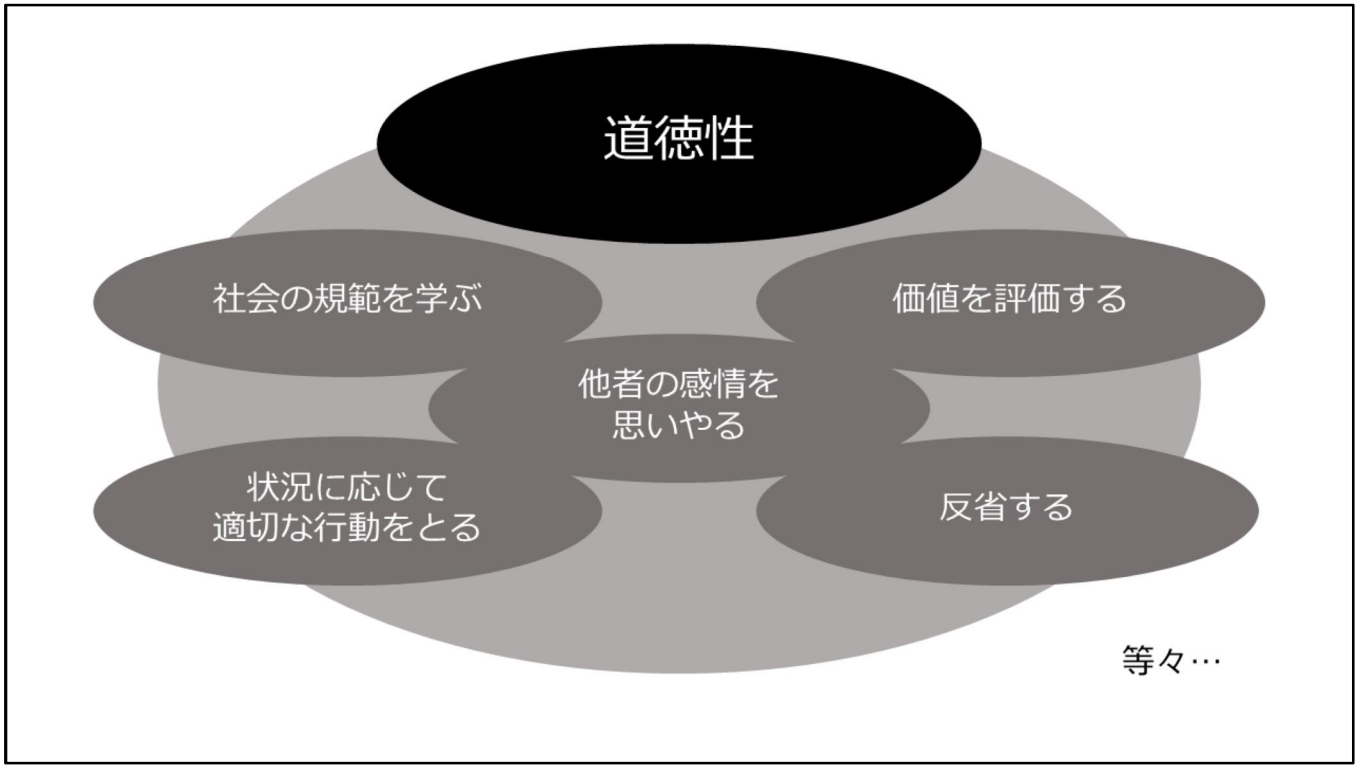
人工知能・ロボットが人間の生命や安全に直接かかわる場面に進出する可能性が高まり、それに伴ってそれらに倫理的判断をさせることの必要性・技術的実現可能性・道徳的是非が論じられるようになっている。

2014年5月と2015年4月，国連での特定通常兵器についての多国間会議で致死的自律型兵器が取り上げられる．そういった兵器の開発と使用を禁止する条約を求める声が上がったが，米国，イスラエル，英国などは反対した．

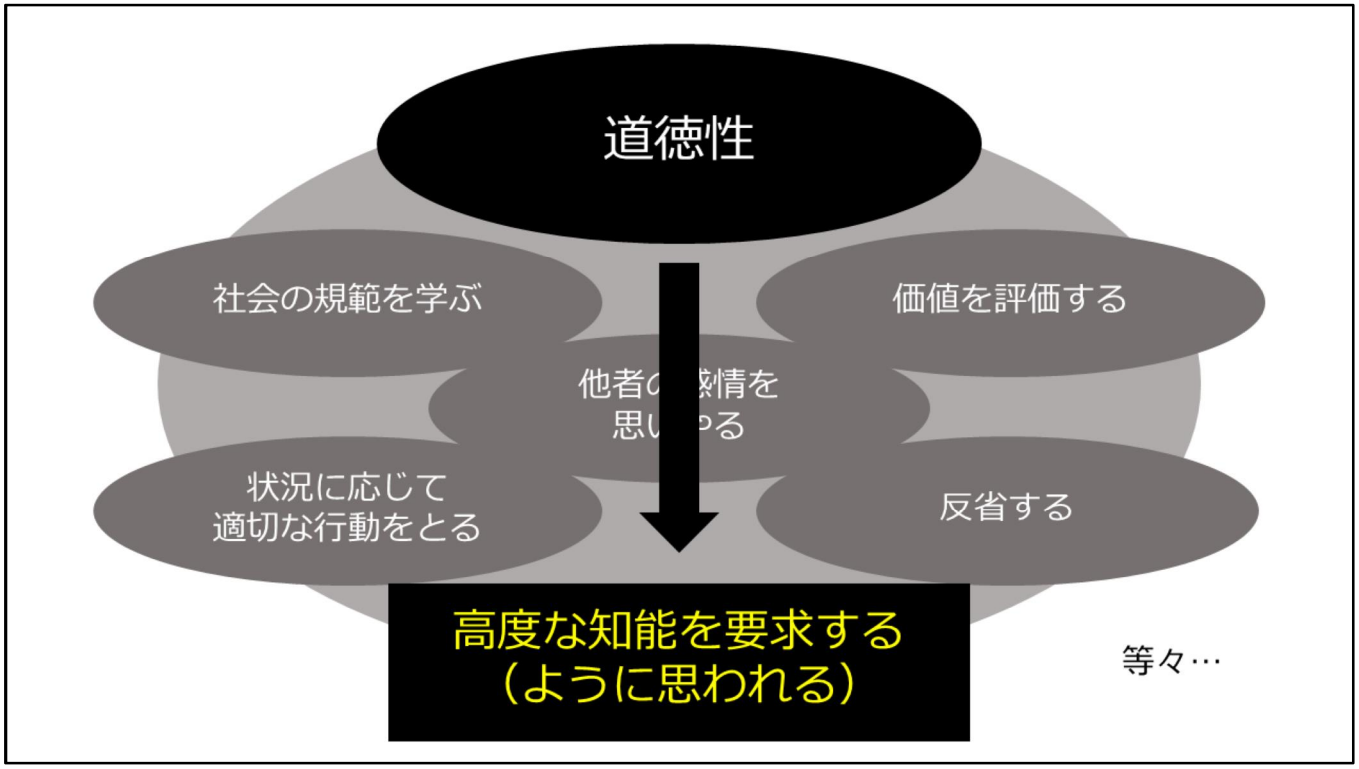
- AIコミュニティは致死的自律型兵器システムに対して立場を表明しなければならない．（Stuart Russell）
- AI研究者は自分たちの研究を大衆に理解させる努力をするべきだ．（Sabine Hauert）
- AIから得られる利益は公平に分配されなければならない．（Russ Altman）
- 人とロボットが共生し，助け合い，互いを補い合うことでポジティブな未来が拓ける．（Manuela Veloso）

http://www.nature.com/news/robotics-ethics-of-artificial-intelligence-1.17611#comments

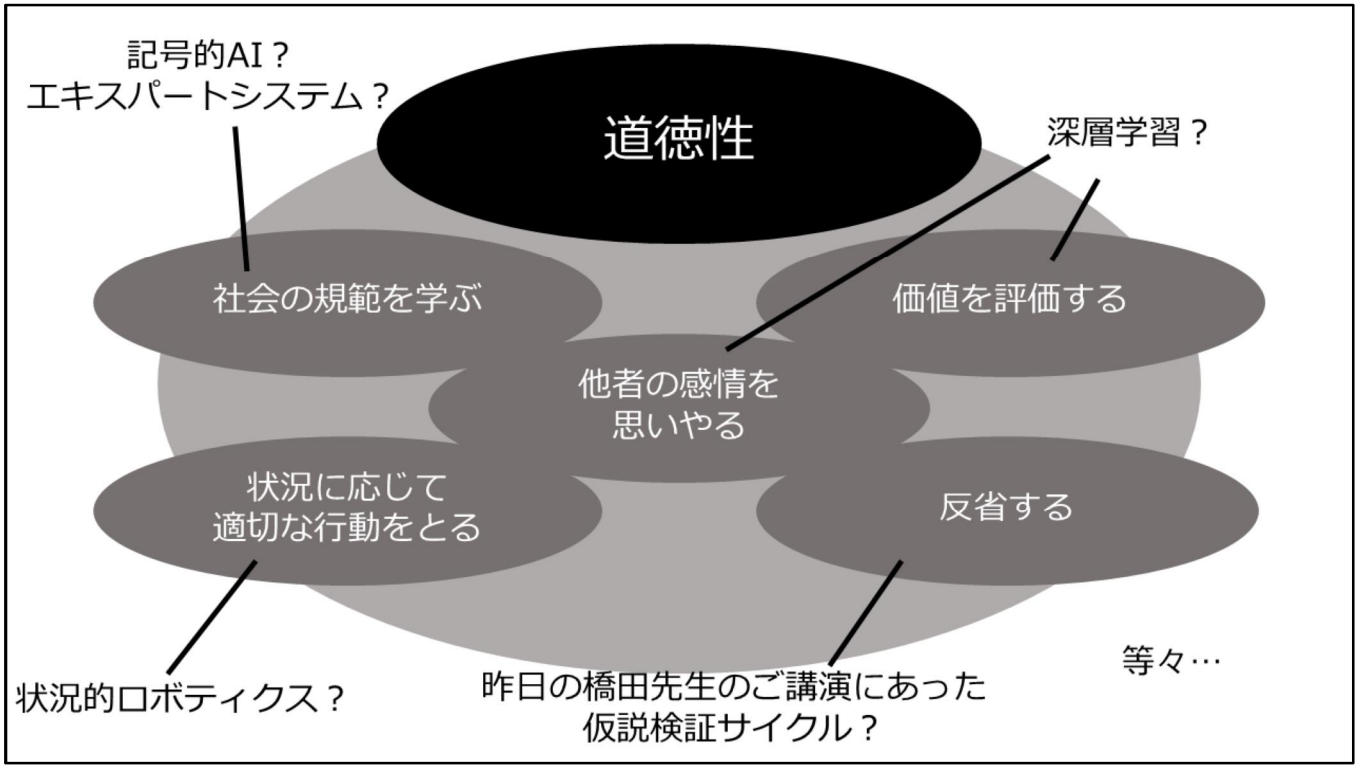
先週のNature.comの特集から．人工知能の倫理について4人の専門家の記事．LAWSについてのRussellの記事が筆頭で長さも一番長い．欧米ではLAWSに対する関心が高いことが伺える．



道徳的であるということには，これらのことが含まれているように思われる.



これらのことは高度な知能を要求するように思われる．汎用人工知能というからにはこういったこともできていなければならないだろう．

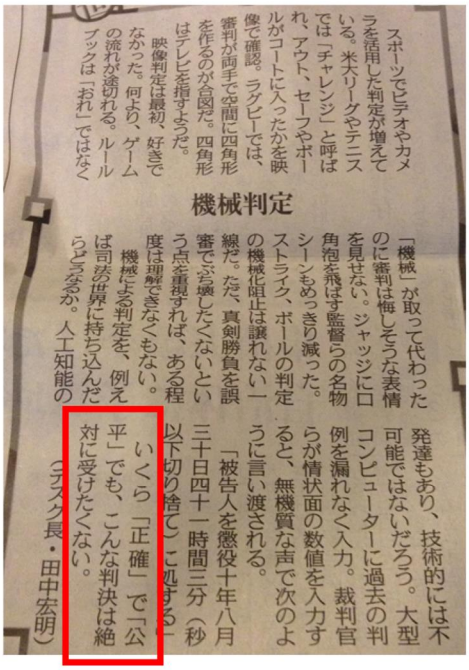


こういった能力は現在のAIのパラダイムによって実装できそうだ．ただし反省する能力は難しいだろう．この前日「ロボットは東大に入れるかプロジェクト」に関するパネル討論があり，そこで東大の橋田浩一先生は，人間の知能にとっては仮説と検証のサイクルが重要で，それはこれまでのAIでは取り上げられてこなかった，と言っていた．そのようなサイクルを実装したAIは反省能力に近いものを持つかもしれない．

道徳的AI，道徳的ロボットの難しさ

- 状況依存性が高い
- 道徳的行為には責任が伴う
- 共感を求められる
- 「何をやるか」だけでなく「誰がやるか」が重要

一方，道徳性には人工知能では取り扱いにくい側面もある．例えばマニュアル通りに同じ状況で同じ行動をとるということが，まさに相手の感情を害することがある．あるいは道徳的行為には責任が伴うが，AIやロボットに責任を帰属させることは不可能である．



スポーツでビデオやカメラを活用した判定が増えている。米大リーグやテニスでは「チャレンジ」と呼ばれ、アウト、セーフやボールがコートに入ったかを映像で確認。ラグビーでは、審判が両手で空間に四角形を作るのが合図だ。四角形はテレビを指すようだ。

映像判定は最初、好きでなかった。何より、ゲームの流れが途切れる。ルールブックは「おれ」ではなく

機械判定

「機械」が取って代わったのに審判は悔しそうな表情を見せない。ジャッジに口角泡を飛ばす監督らの名物シーンもめっきり減った。ストライク、ボールの判定の機械化阻止は譲れない一線だ。ただ、真剣勝負を誤審でぶち壊したくないという点を重視すれば、ある程度は理解できなくもない。

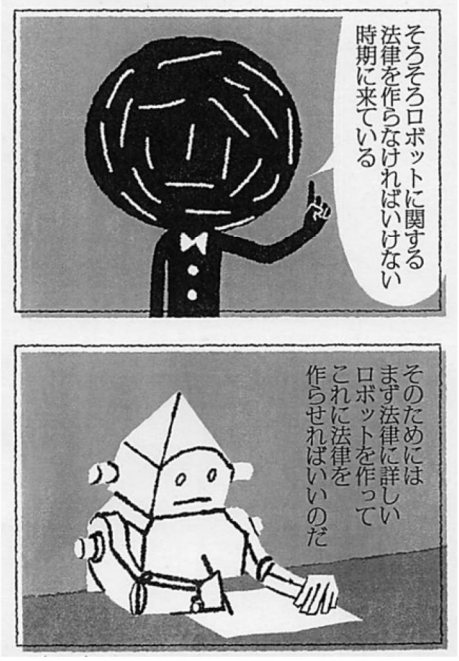
機械による判定を、例えば司法の世界に持ち込んだらどうなるか。人工知能の発達もあり、技術的には不可能ではないだろう。大型コンピューターに過去の判例を漏れなく入力。裁判官らが情状面の数値を入力すると、無機質な声で次のように言い渡される。

「被告人を懲役十年八月三十日四十一時間三分（秒以下切り捨て）に処する」

いくら「正確」で「公平」でも、こんな判決は絶対に受けたくない。

（デスク長・田中宏明）

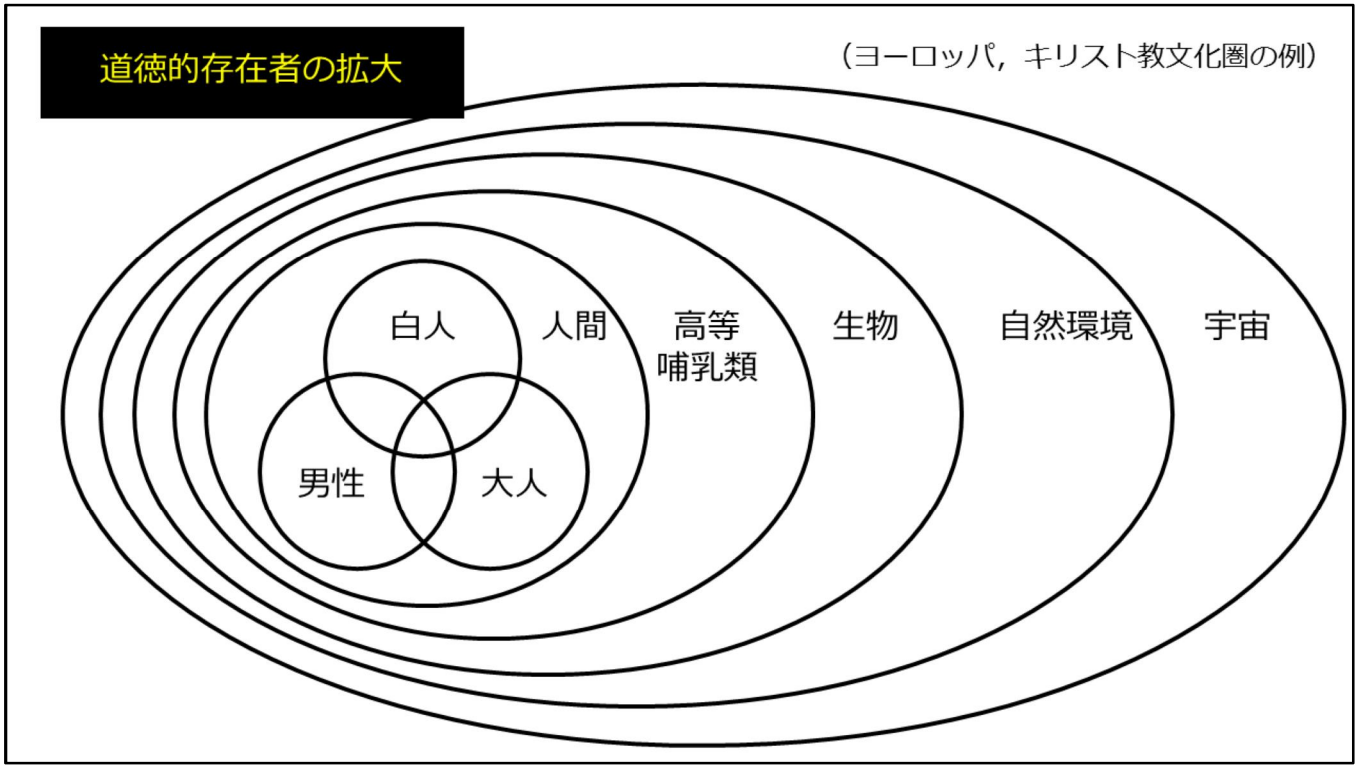
中日新聞，2015年5月17日



木内達朗「チキュウズィン」
（『考える人』，2015年春号）

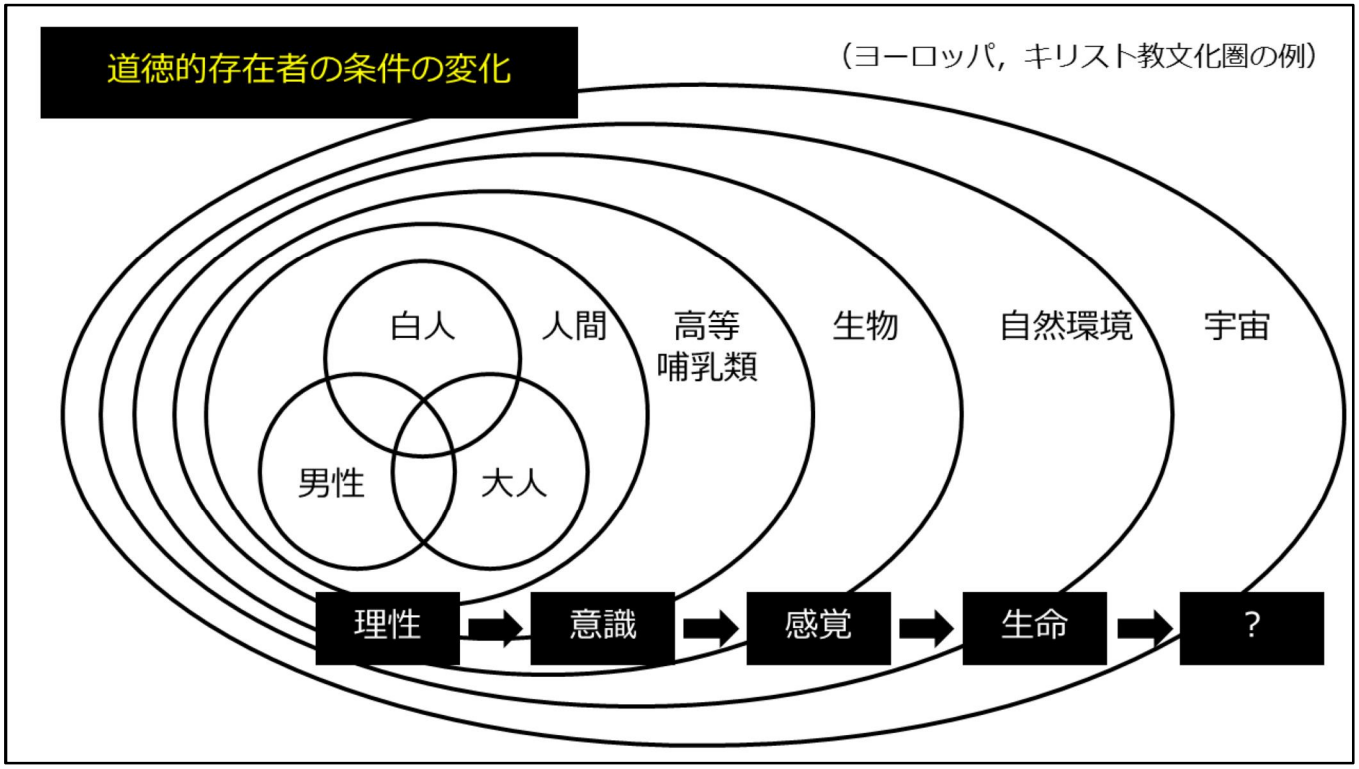
左の記事では法的な判断をAIがするという可能性に言及しているが，そのようなことはやはり問題外だと思っている．これは能力の問題ではなく，まさに人間でないということによってそれに法的判断をゆだねることが否定されているのである．

右の漫画ではロボットが法律を作るということが考えられているが，これがオチであるということから，そのような可能性が馬鹿げたものであるということが示唆されていると考えてよいだろう．

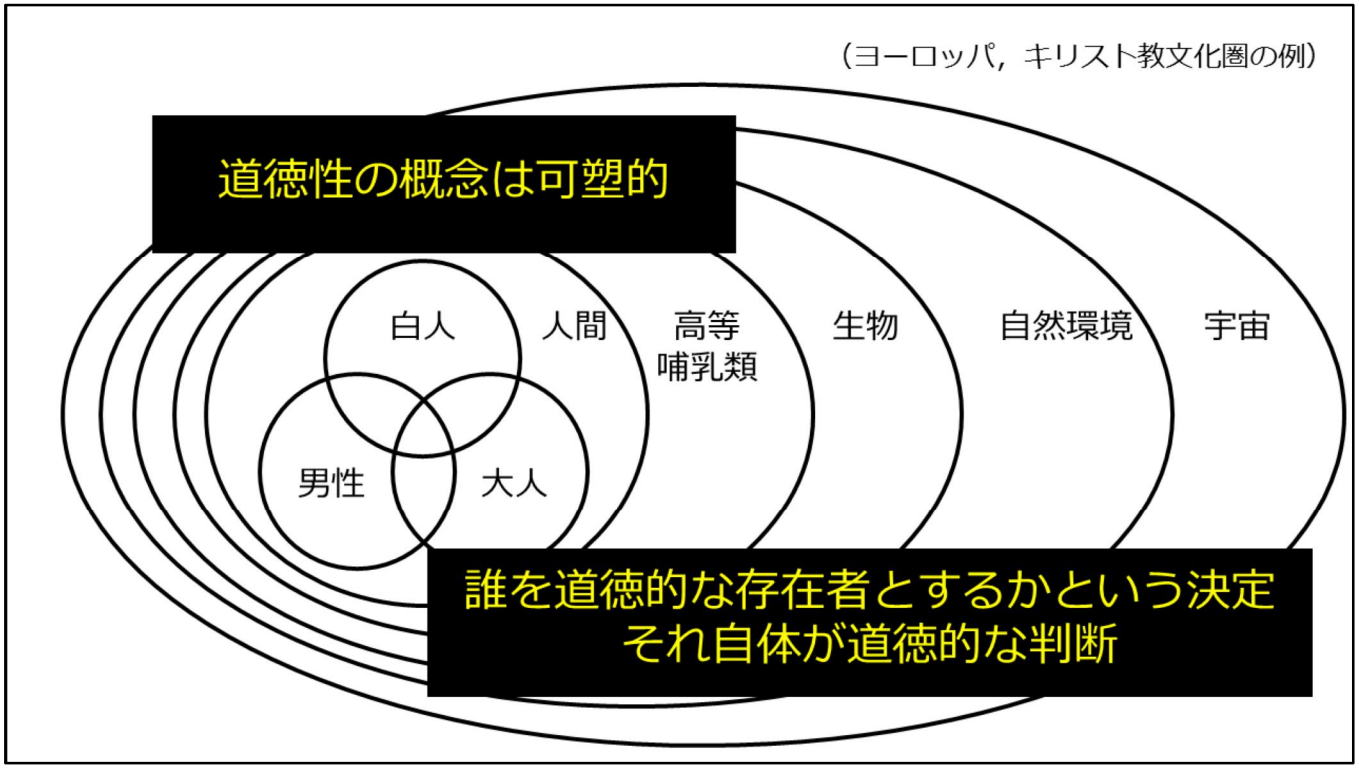


しかし道徳的存在者の外延は常に安定しているわけではなく，欧米では少しずつ拡大してきた．現在では宇宙環境の保護を訴える（例えばよその惑星にテラフォーミングをしてはいけないというような）人々すらいる．しかし道徳的対象の外延の拡大にはしばしば非常に大きな摩擦・衝突が伴ってきた．18世紀末，イギリスの社会思想家ウルストンクラフトが『女性の権利の擁護』を書いたとき，プラトンの翻訳者トマス・テイラーはウルストンクラフトを揶揄する目的で『獣の権利の擁護』という小冊子を匿名で発行した．彼にとっては女性の権利を認めることは，動物の権利を認めることと同じくらい馬鹿げたことだと思われたのである．しかし現在では動物の権利も多くの人が認めている．

しかし現在の私たちの基準で，白人・男性・大人以外に権利を認めなかった当時の人々を「不道徳的」と断じることはアナクロニズムであろう．現在の私たちにとって機械の道徳的権利は馬鹿げて思えるかもしれないが，当時の人々にとっては「女性の権利」や「動物の権利」などは同様に考えられないことだったのである．

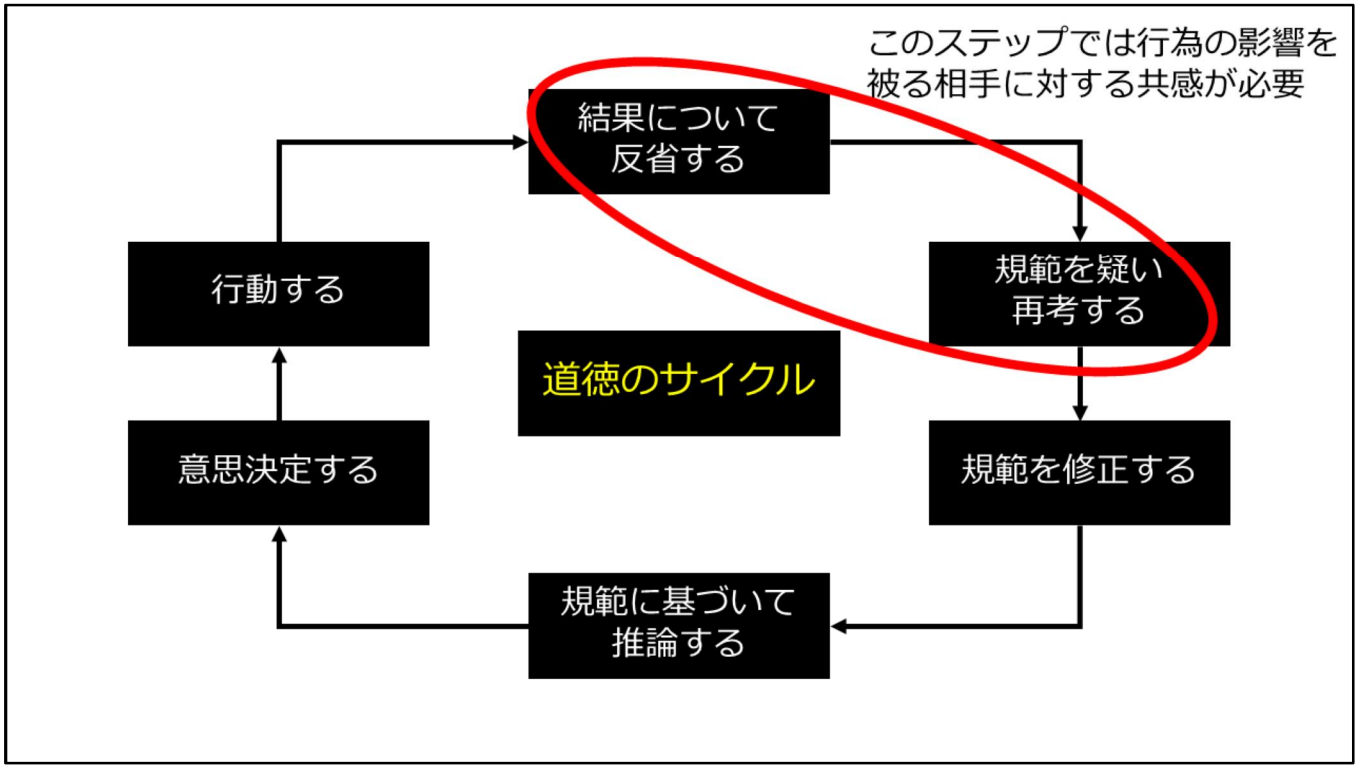


道徳的存在者の規準もまた時代とともに変化してきた.



何を道徳的存在者とするかということ，より一般には道徳であるとはどういうことかの概念は，時代とともに変化する．また社会によっても異なる．

さらには「何を道徳的存在者とするか」，「道徳的であるとはどういうことか」についての決定もまた道徳的に重要な判断である．



道徳というのは固定されたものではなく，図のような反省的サイクルによって変容するものである．そしてこのような変容性を欠いた社会や個人は，むしろ道徳的ではない．

真に道徳的な人間は，自分が従っている規範を場合によっては疑うことができなければならない．

人工知能は道徳的になりうるのか？

↓

私たちは人工知能を
道徳的存在者として受け入れられるか？

道徳性の規準は固定されていない．それは私たちが決定し選択するものである．従って「人工知能は道徳的になりうるのか？」という問いは，正しくは「私たちは人工知能を道徳的存在者として受け入れられるか？」という形で問われなければならない．

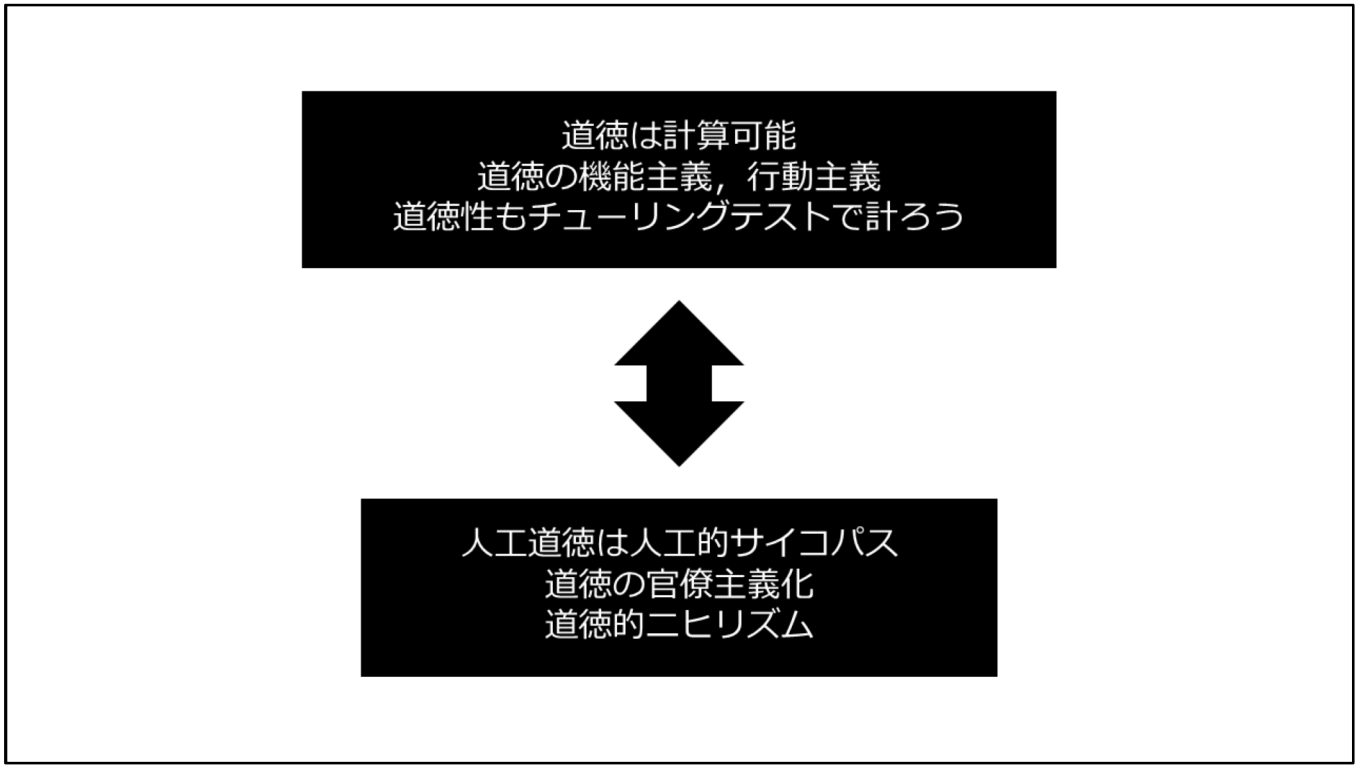
人工知能は道徳的になるべきか？

↓

私たちは人工知能を
道徳的存在者として受け入れるべきか？

同様に「人工知能は道徳的になるべきか？」という問いは「私たちは人工知能を道徳的存在者として受け入れるべきか？」という形で問われなければならない.

Boston Dynamics社の開発した四足歩行ロボットSpotを紹介する動画の中で，ロボットが蹴られるシーンがあった．それをめぐって「ロボット犬でも蹴ってはいけない」，「所詮ロボットだ，構わない」という論争が起きた．すでにロボットを道徳的配慮の対象として認めている人々もいるのだ．日本でもペットロボットのAIBOにお経をあげて供養する人々がいることが報じられている．



ロボットが道徳的行為者になりうるか，なるべきかという問題については，上記のような対立がある.

人工知能は道徳的になりうるのか？

現在の「道徳性」の通念からすれば無理.
しかし人工道徳的行為者の導入は
「道徳性」の概念を書き換えるだろう.
臓器移植という技術が「死」の概念を書き換えたように.
しかしその際には大きな衝突が起こることが予想される.

この問題について私の意見は上の通りである．感情や責任をAIやロボットに実装することは現在の技術ではなかなか難しそうである．しかしそういったものを伴わなくても道徳的といっていいのだというように，道徳性の概念そのものが変わっていく可能性はある．実際にそのような議論は欧米の倫理学者の間ではすでに行われている．（これについては前述したGunkelの本を参照してほしい）

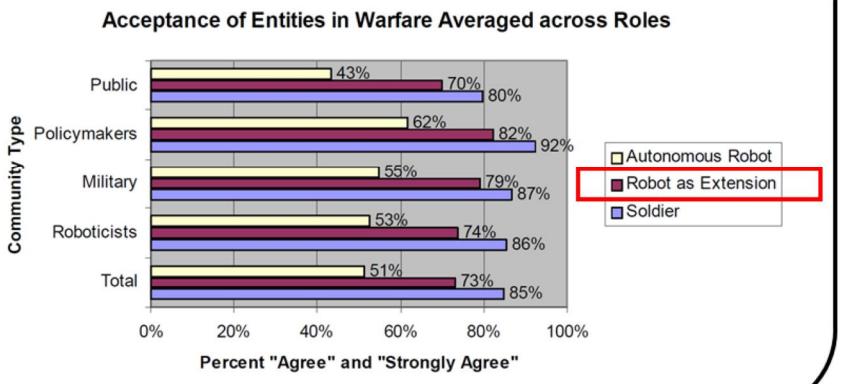
人工知能は道徳的になるべきか？

ケアロボットなどはこれからの日本に必要な技術である.
そしてそれらは表層的にでも「道徳的」に振舞うことが望ましい.
しかし私たちが伝統的に持ってきた「道徳性」の概念に
重大な変更をもたらす技術の導入には
少なくとも慎重な議論と合意形成のための努力が不可欠だろう.

基調講演で松原仁先生が言った通りで，古い価値観に固執する人たちを単に退けるのではなく，社会との粘り強い合意の形成を探っていくことが重要だと．それと同時に現代の社会的な問題に対してAIやロボットが唯一の実現可能な解決だと性急に判断することにも慎重になるべきだろう.

提案

Moshkina and Arkin, ``Lethality and autonomous systems'' (2007)



- まずは人間の道徳的判断・行動の拡張としての人工道徳的行為者を作ることから始める.
- ユーザーと相手との心理的距離を広げすぎないように配慮する.
- 相手のいうことを聞く耳を持たせる.

枠で囲ったのはロボットが戦争に行くことについての人々の意見についての調査. 自律的ロボットが戦争に従事することに賛成する人は5割くらいだが, 人間の延長としてのロボット(例えば遠隔操作される無人機)ならば7割以上の人が受け入れる.

人工道徳的行為者についても, 自律的なものではなく, 人間の道徳的判断や道徳的行為の拡張として使えるものから手をつけることが望ましいだろう.

また行為者と行為の影響を受ける人との間の心理的距離が道徳的判断に影響を与えるという研究があることから, AIやロボットの導入によって拡がる心理的距離の影響も慎重に考慮しなければならない.

AIやロボットの活動を拒絶できるような仕組みがあったほうが良いと思う. たとえば自宅の上空をドローンが飛べなくなる仕組みとか, グーグルグラスのようなものに写されない仕組みとか. 企業はユーザーの声だけでなく, その製品の影響をこうむる人間にもその製品のコントロールを部分的に持たせるべきだ.