

人工道徳は倫理学にどのような 影響を与えるか？

久木田水生

minao.kukita@gmail.com

京都大学大学院文学研究科研究員
応用哲学会2012年度大会(千葉大学)

医療

家事

警備

ロボットが浸透する社会

軍事

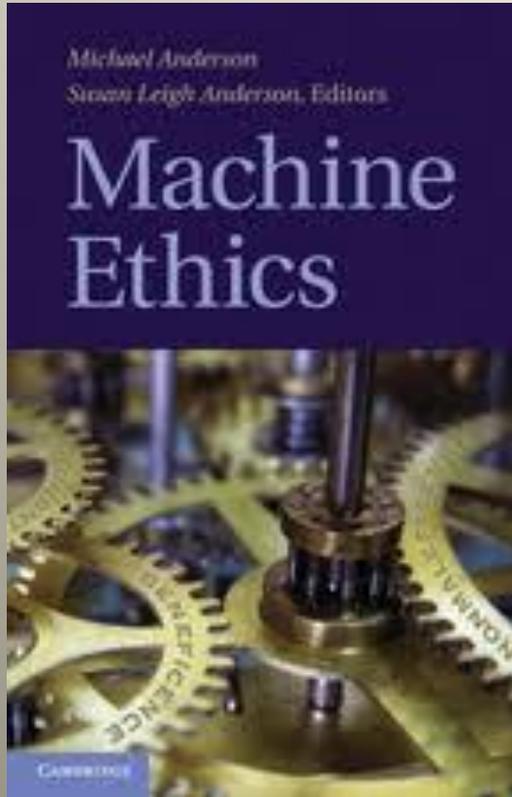
医療

交通

社会的影響への懸念

- 2007年：千葉大学「千葉大学ロボット憲章」
 - 第2条（民生目的）千葉大学におけるロボット教育・研究開発者は、**平和目的の民生用ロボットに関する教育・研究開発のみ**を行う。
 - 第3条（非倫理的利用防止）千葉大学におけるロボット教育・研究開発者は、**非倫理的・非合法的な利用を防止する技術をロボットに組み込むこと**とする。

人工道徳 (Artificial Morality) への 関心の高まり



Anderson and Anderson eds.
(2011)



Wallach and Allen
(2009)

人工知能



心の哲学

大きな影響

人工道徳



倫理学

同様の影響が期待できる？

目的

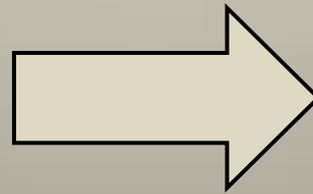
これまでの人工知能の発展を顧みながら、
将来の人工道徳の方向性と、
それが倫理学に与える影響について考察する。

人工知能と心の哲学

総合的(構成論的)アプローチ

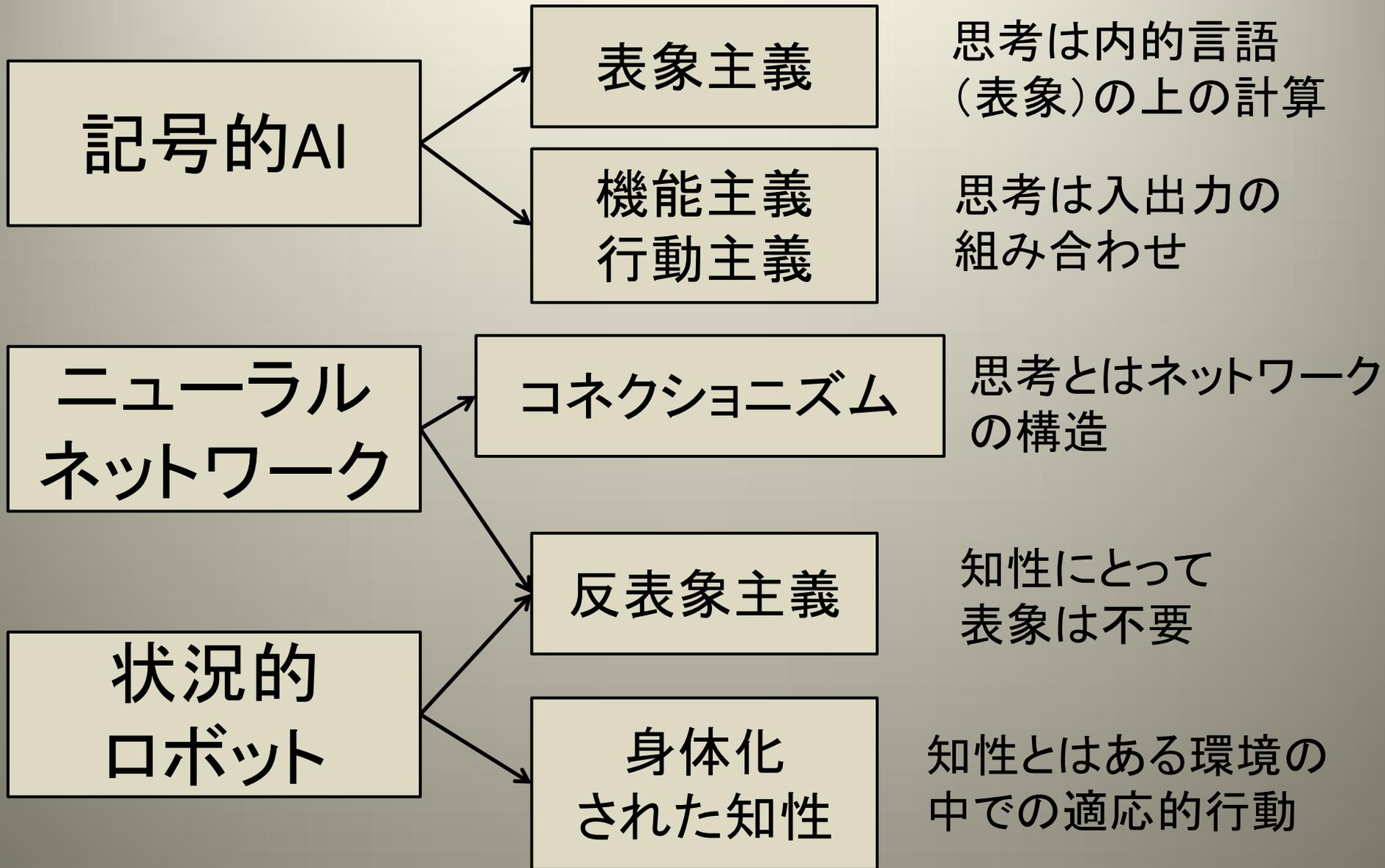
こどもの発達
生物の進化
動物のコミュニケーション
人間の精神
ETC.

対象の観察による
研究に限界



似たようなものを
作ってみたら
何か分かるかも!

人工知能と心の哲学



チューリング マシン

- 高度な推論や計算は得意.
- パターン認識や適応的行動が苦手.
- 漸進的進化で生まれるとは考えにくい.



- 漸進的進化観と相性が良い.
- 人間に特徴的(と考えられる)な複雑な推論や計算, 言語使用の能力を説明できない.

ニューラル
ネットワーク

状況的
ロボット

拡張された精神仮説

精神＝問題解決システム

脳・身体

人工物

パターン認識・適応的行動

記憶・計算・推論の補助

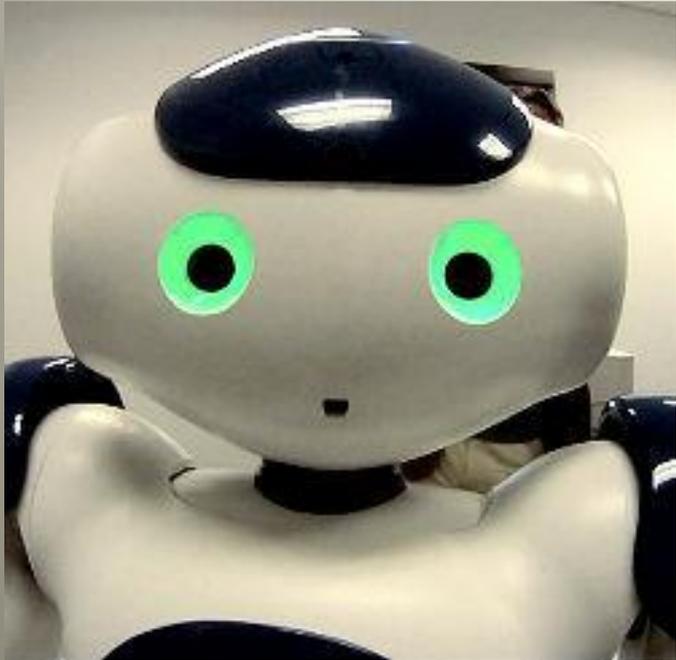
アンディ
クラーク



人間は
「生まれながらのサイボーグ」
である！

人工道德と倫理

倫理的ロボット「NAO」



- マイケル・アンダーソンとスーザン・アンダーソンによって開発されているMedEthExという人工知能を組み込んでいる。
- 患者に薬を飲むように指示する看護ロボット。
- 患者が薬を飲むことを拒否した時、NAOはそれがどれだけ重大かを判定する。
- その判断に基づきNAOは人間の看護師に知らせるか、放っておくかを決定する。

写真は <http://cacm.acm.org/news/101611-the-ethical-robot/fulltext> よりダウンロード。(2012年4月24日)

MedEthExの推論

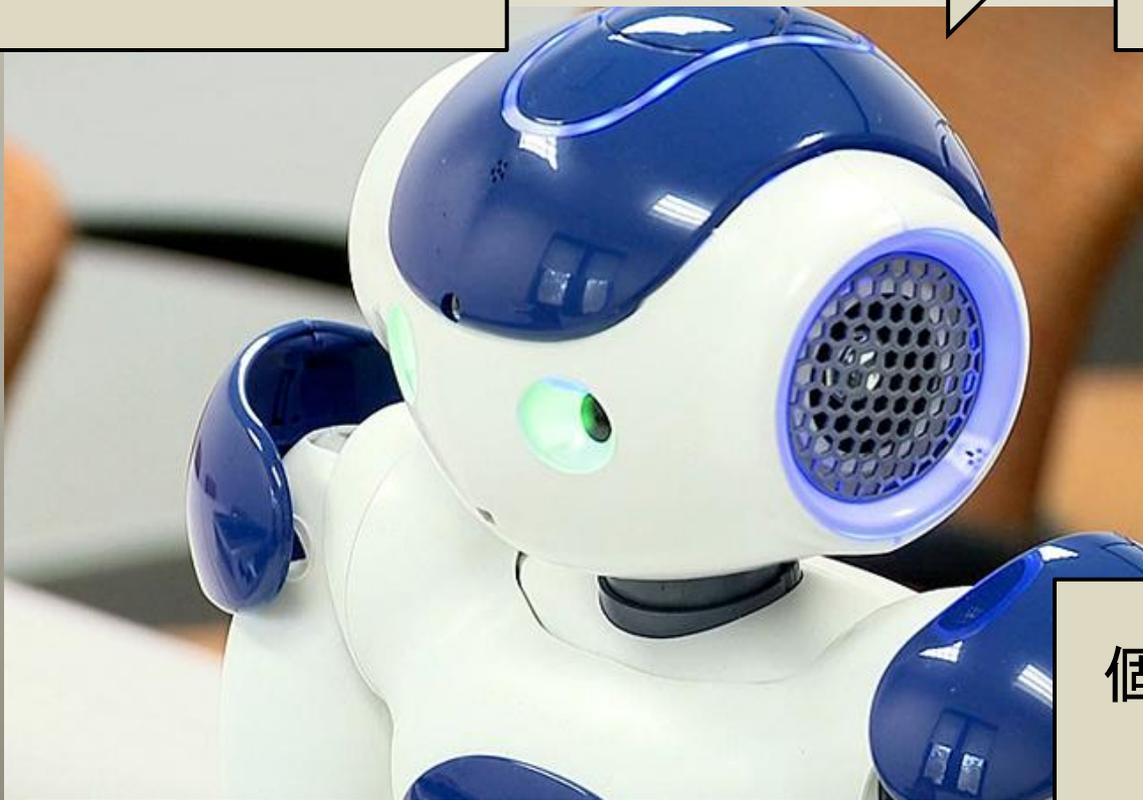
人間の専門家の
推論の事例

帰納

一般原則

演繹

個別の事例における
最善の選択



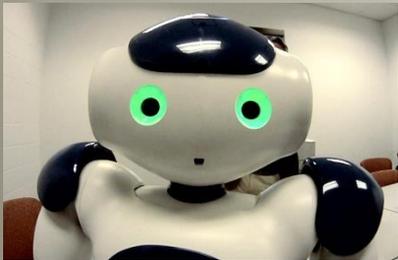
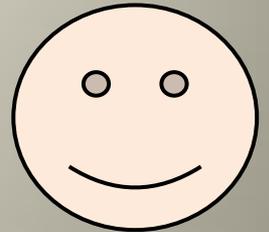
写真は<http://www.robaid.com/robotics/nao-robot-used-by-researchers-to-implement-machine-ethics.htm> よりダウンロード。(2012年4月24日)

どう倫理的なの？

患者の健康と自尊心に
配慮してます！

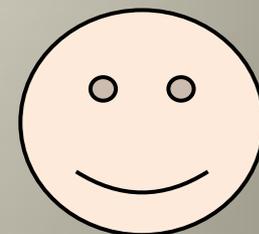
どうせ決められたプログラムに
従っているだけじゃないの？

機械学習しています！

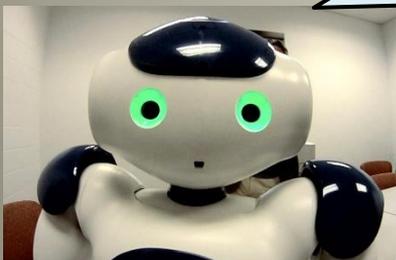


じゃあ「健康」とか「自尊心」
とかって意味分かってる？

....



現時点での人工道徳は
初期の記号的AIの段階



道徳的機能主義

Wallach and Allenは人工道徳に関して
機能主義をとっている。

Floridi and Sandersも同様とみなせるだろう。

Next Step?

道徳的コネクショニズム？

身体化された道徳？ 表象なき道徳？

拡張された道徳性？

道徳的コネクショニズム？

ニューラルネットワークに限らず
優れたパターン認識が道徳的な感受性に
繋がることはありうる。



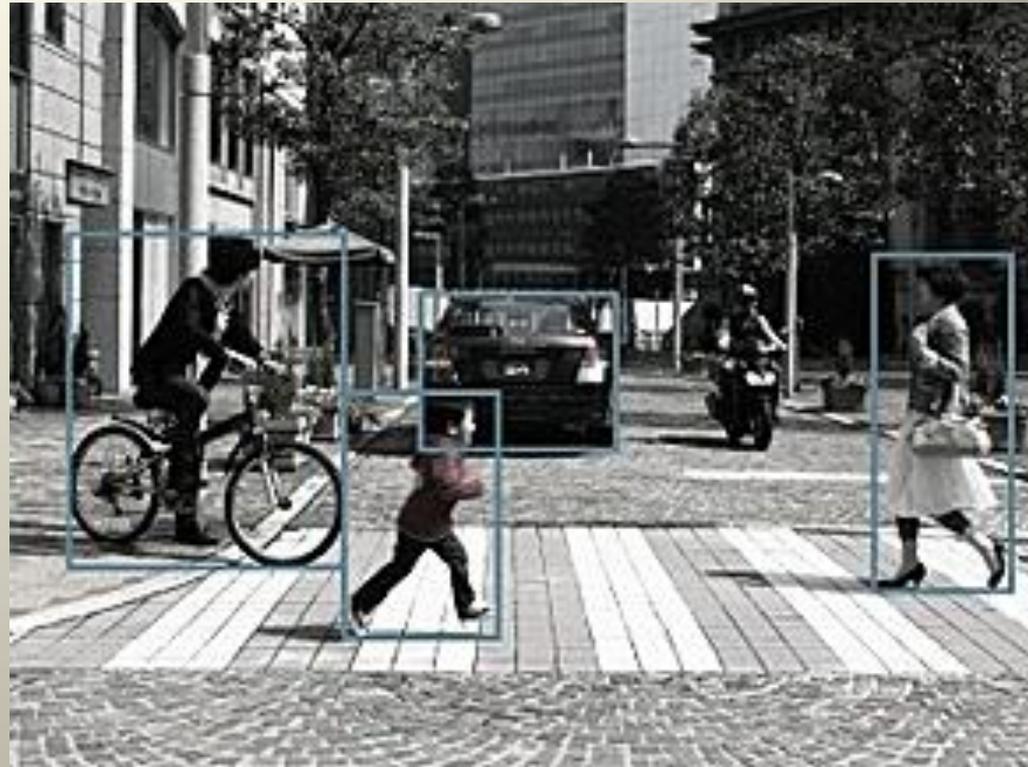
マルチ・エージェント・
モデリングに基づいて
ロス市警で実施された
予測警察 (Predictive Policing)

写真はhttp://lapd.com/news/headlines/can_computers_predict_crimes_of_the_future/よりダウンロード。(2012年4月24日)

身体化された道徳 = fail-safe



ステレオカメラにより対象の接近を感知し、自動的にブレーキをかけるスバルのアイサイト



身体化された道徳はおそらく機械の歴史と同じくらい古い。
Wallach and Allenはこれを「操作的道徳」と呼ぶ。
これは伝統的に倫理学において軽視される傾向にある。

写真は<http://bizmakoto.jp/makoto/articles/1004/22/news095.html>
よりダウンロード。(2012年4月24日)

拡張された道徳性？

人間と人工道徳が協働するシステムの
全体が道徳性の担い手？

そのようなシステムに対しては
「企業の責任」や「政府の責任」と同様の責
任帰属がなされるかもしれない

まとめ

- 現在の人工道徳の中心は記号的AI.
- 人工道徳についての有力な論者たちは機能主義をとっている.
- しかし身体化された道徳, 表象なき道徳が今後, 人工道徳で優勢になってくることは大いにありうる.
- 拡張された道徳仮説を受け入れるならば人工道徳は人間とともに一つの道徳的問題解決システムを構成する.
- その場合, そのシステムは企業や政府のような集合的責任主体として扱われるかもしれない.

ご清聴ありがとうございました

久木田水生

minao.kukita@gmail.com