

# 「人工知能の哲学の アップデートのために」

鈴木貴之  
(東京大学大学院総合文化研究科)

1

## 第2次ブーム期の哲学的議論

- 中国語の部屋（サール）
- フレーム問題（デネット）
- ドレイファスのAI批判
- ウィノグラードのAI批判

2



3

## サール「心・脳・プログラム」

「プログラムされたコンピュータは...**実際には何も理解していない**ということを以下において論じたい。コンピュータの理解は（私のドイツ語の理解のように）部分的なもの、あるいは不完全なものですらない。それはゼロなのである。」

ホフスタッター、デネット『マインズ・アイ（下）』(p.186)

4



5

## デネット「コグニティブ・ホイール」

「ではどのようなシステムが必要であろうか。それは、自分が知っていることの大半を純粋に無視し、いつでも自分の知識の一部をうまく選び出してそれによって作業するようなシステムである。...複雑な行為状況における実に多様な状況のもとで**無視すべきことを確実に無視するようなシステムを設計するにはどうすればよいのか。**」

『現代思想』1987年4月号 (p.140)

6



7

ドレイファス『コンピュータには何ができないか』

「予言、マスコミの記事、映画、警告にもかかわらず、人工知能は約束であって、達成された事実ではない。」 (p.152)

「伝統哲学の先入見を去って記述的あるいは現象学的な証拠に頼れば、**どんな形の知的振舞いにもプログラム不可能な人間の能力が含まれている**、という示唆が得られる」 (p.487)

8

「失敗したのは何らかの個別的説明ではなく、**概念枠全体**である。つまり、人間の振舞いの説明は物理学的説明で成功したプラトンの形式をとりうるし、とらなければならないということ、状況は物理状態と同じように扱うことができるということ、人間の世界は物理的宇宙のように取り扱うことができるということ、これらを前提とする概念枠が失敗に終わったということである。」 (p.397)

9

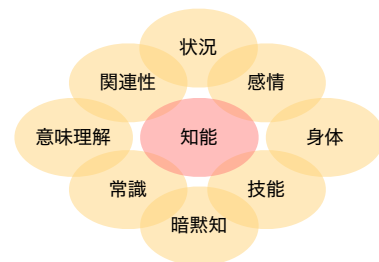


10

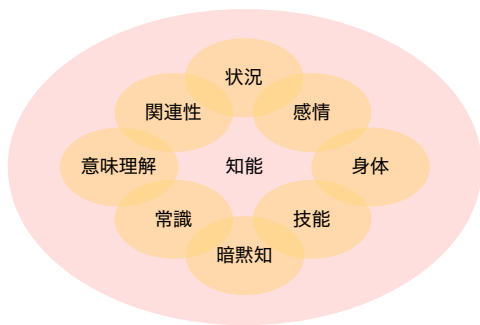
おもな論点

- 意味、状況、関連性、直観、常識、暗黙知、技能、身体、感情...

11



12



13

### しかし...

- 概念間の関係は？
- 説明項は十分に明確か？
- 知能が記号計算でないとしたら何なのか？

14

### サール「心・脳・プログラム」

「私が英語を理解することができ、他のタイプの志向性をもちうるのは、私がコンピュータ・プログラムの具体例だからではない。...そのようなことが可能になるのは、われわれが知る限りでは、私が**ある種の生物学的（すなわち、化学的、物理的）構造**をもったある種の有機体であるからなのであり、この有機体は、ある条件のもとでは知覚、動作、理解、学習、そしてその他の志向的現象を産出することができる。」

『マインズ・アイ（下）』(p.201)

15

### ドレイファス『コンピュータには何ができないか』

「知能は理解を要求し、理解はコンピュータに常識という背景を与えることを要求するが、その常識を成人した人間が持っているのは、彼が**身体**をもち、**技能**を通じて物質世界と相互作用し、ある文化へと教育されるからだ」(p.5)

16

### しかし...

- 概念間の関係は？
- 説明項は十分に明確か？
- 知能が記号計算でないとしたら何なのか？

17

### その後の展開：人工知能研究

- ハードウェアの性能向上
- 技術的進展（深層学習など）
- 具体的な成果

18

ドレイファス『コンピュータには何ができないか』  
(第3版)

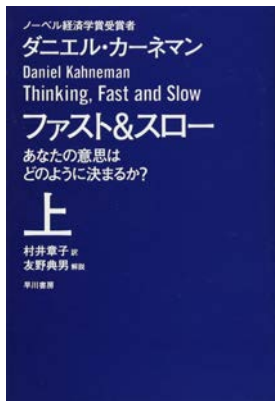
「脳内には理解可能なメカニズムがあり、それをハードウェアにおいて複製することで領域限定的な人工知能を作りだすことが可能である。そして強化学習は正しい方向への第一歩なのだ、と考えることはもっともなように思われる。しかし、いまのところ脳についてよくわかっていないことと、コンピュータのメモリサイズには実際的な制約があることを考えれば、この種の脳にヒントを得たAIに**近い将来実質的な進展が見られることは、まずありえない**と考えることもっともだろう。」 (pp.xliv-xlv)

19

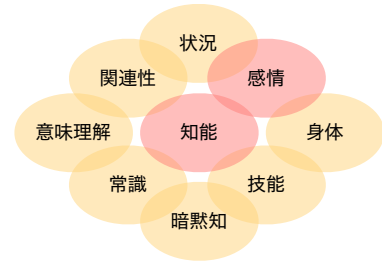
その後の展開： 認知科学

- 感情と知能
- 感情と身体
- 身体と知能

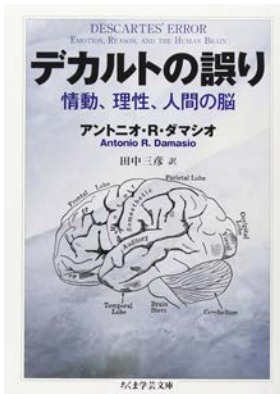
20



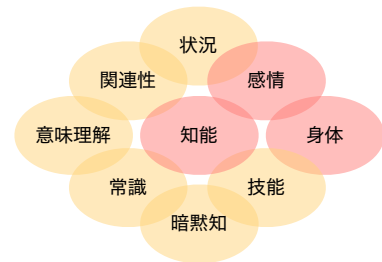
21



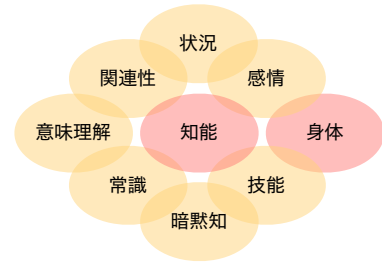
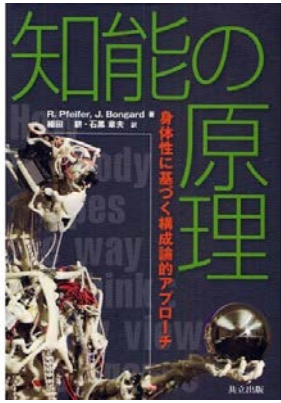
22



23



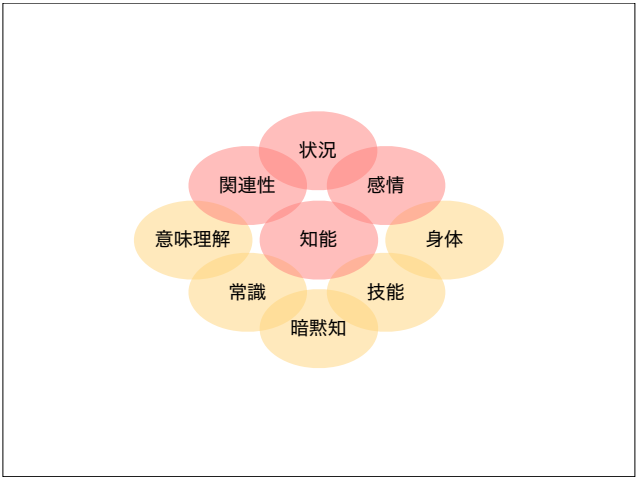
24



その後の展開：哲学

- 認知科学の哲学
- 現象学と認知科学の接近
- 徳倫理学・徳認識論



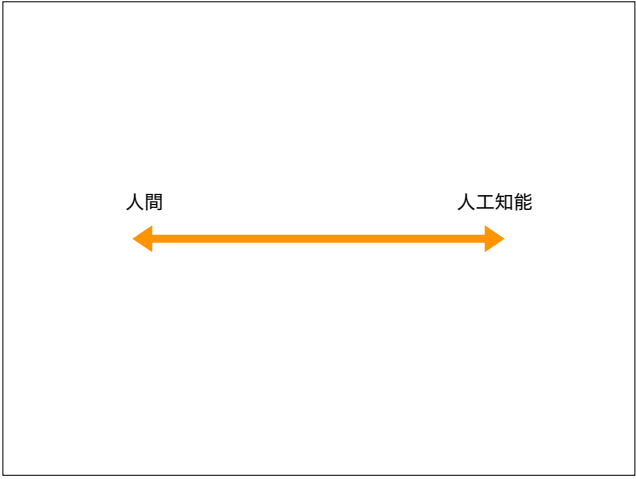


31

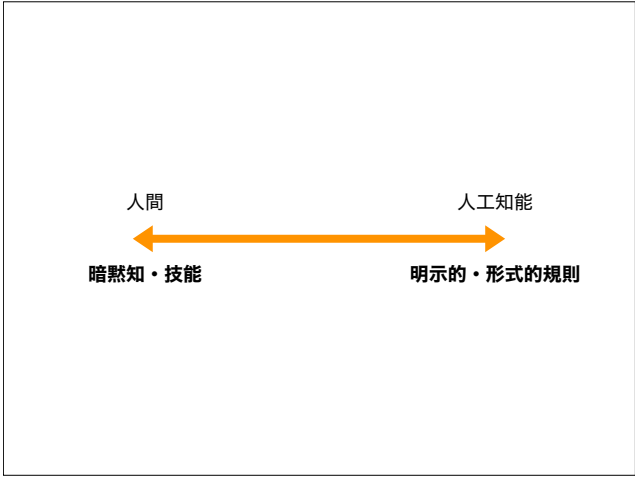
**アップデートに向けて**

- 各論点の関係は？どの論点が本質的か？
  - 関連性、身体、汎用知能...
- 議論の構図の変化

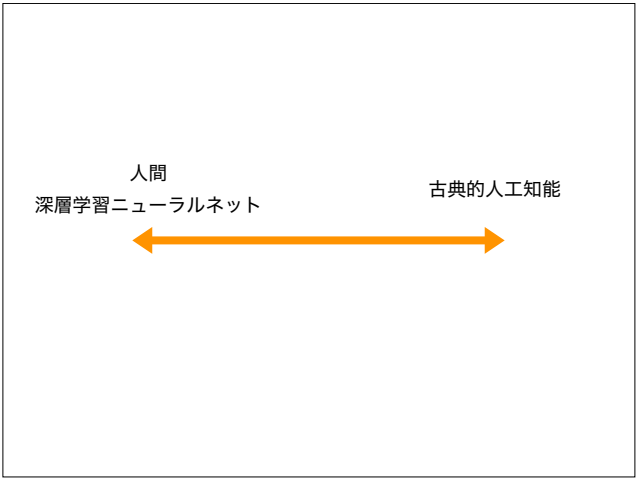
32



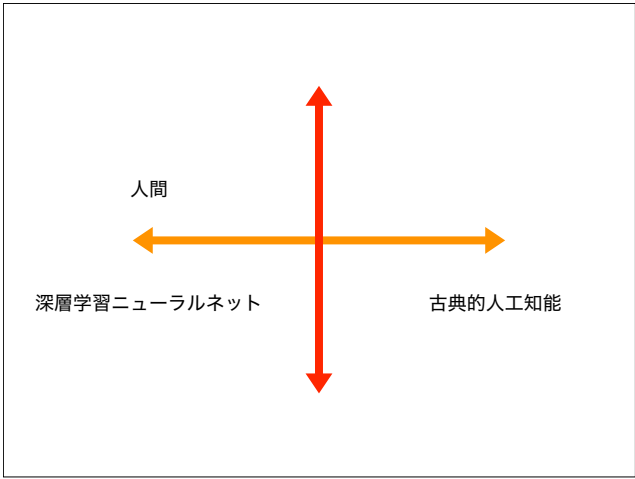
33



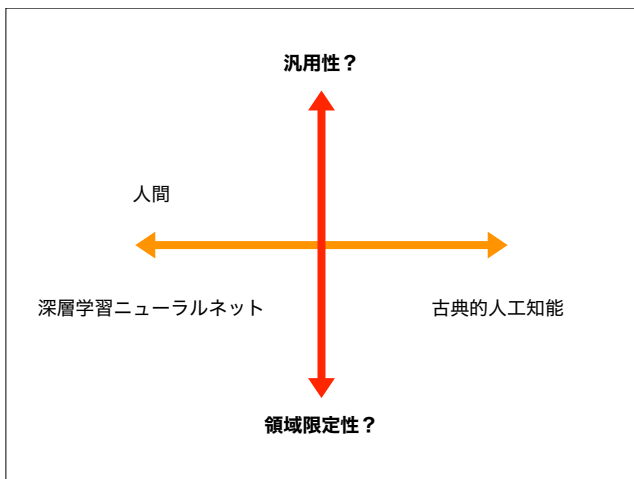
34



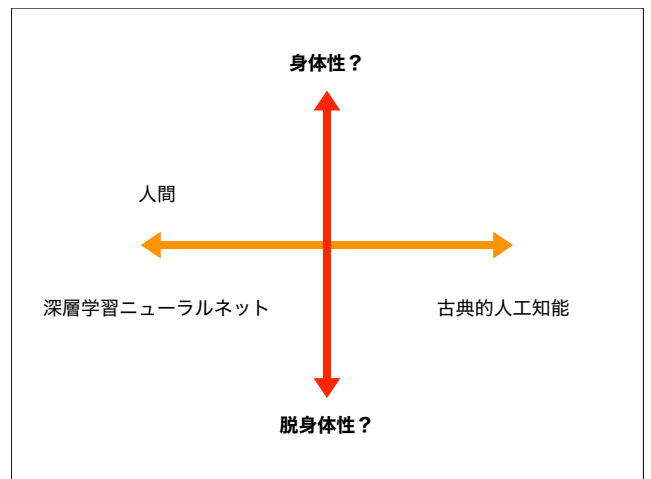
35



36



37



38

- よりよいリサーチ・クエスチョンとは？
- 深層学習にかんして：
- 従来のニューラルネットと深層学習ニューラルネットの違いとは？
  - 脳と深層学習ニューラルネットの違いとは？
  - 深層学習ニューラルネットが苦手な課題とは？

39

- 深層学習ニューラルネットの働きは、人間が理解可能な形で説明可能か？
- 深層学習ニューラルネットだけで汎用人工知能は実現可能か？
- 知能の本質は記号計算ではなく統計分析か？

40

- 感情にかんして：
- AIに感情は不可欠か？
  - AIは感情をもちうるか？
  - 感情をもたないAIには、どのような限界があるか？

41

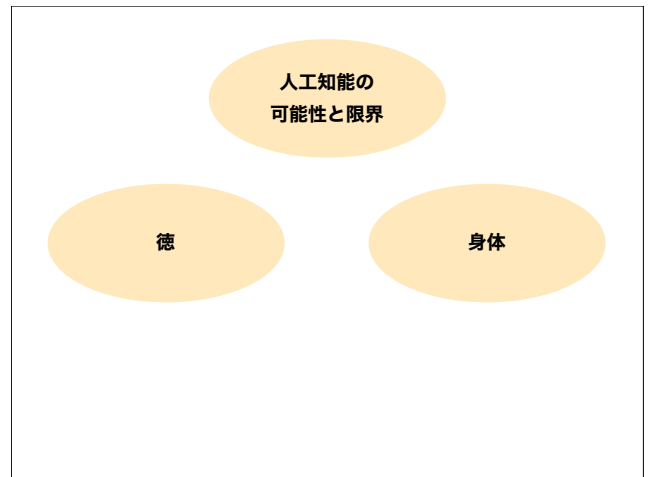
- 身体にかんして：
- AIに身体は不可欠か？
  - 身体をもたないAIには、どのような限界があるか？

42

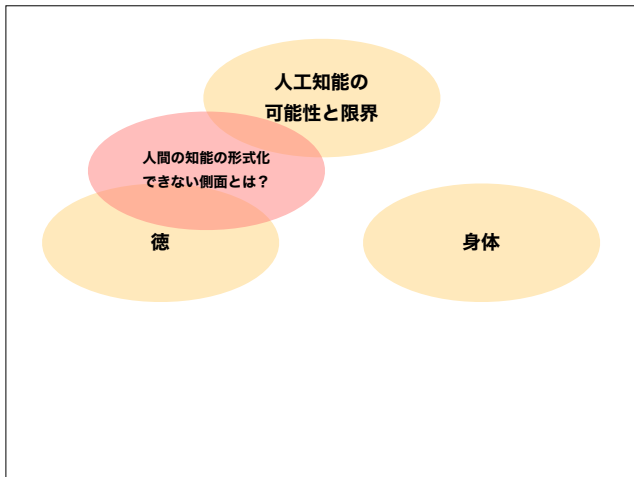
より一般的な問い：

- 人間と重要な点で異なり、かつ人間と同程度の汎用性をもつ知能は実現可能か？
- AIは人間とは異質な知能か？ そうだとしたら、その長所・短所とは？
- 人間の知能とAIを組み合わせることで、どのような可能性が開けるか？

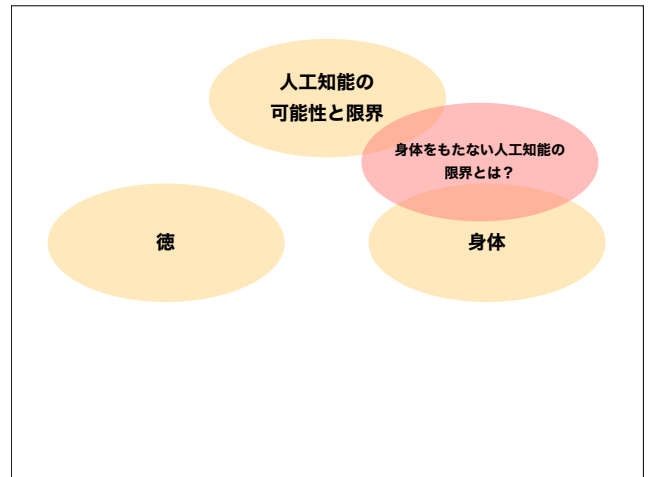
43



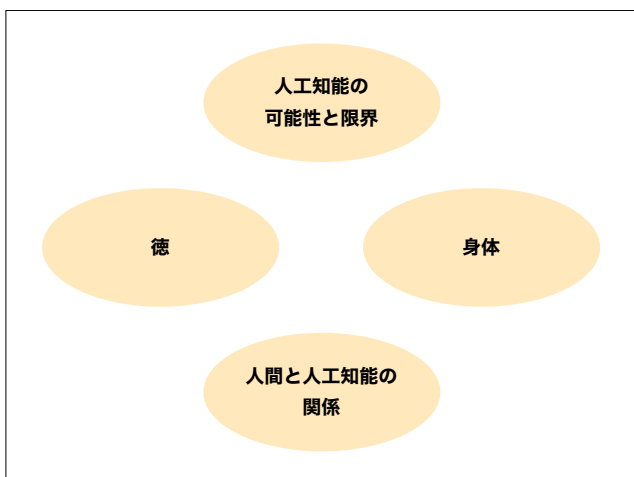
44



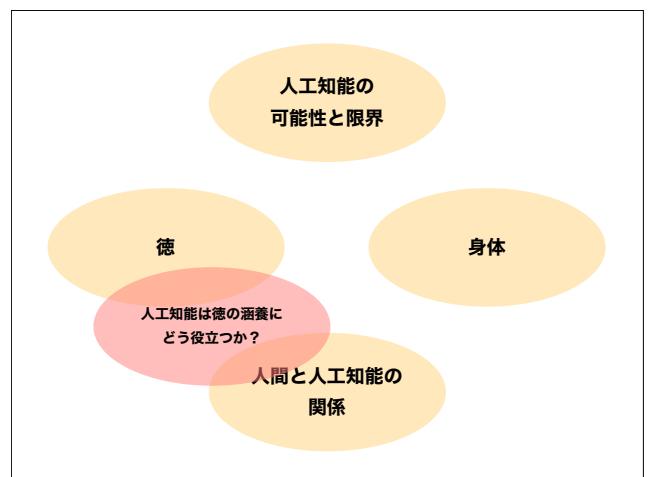
45



46

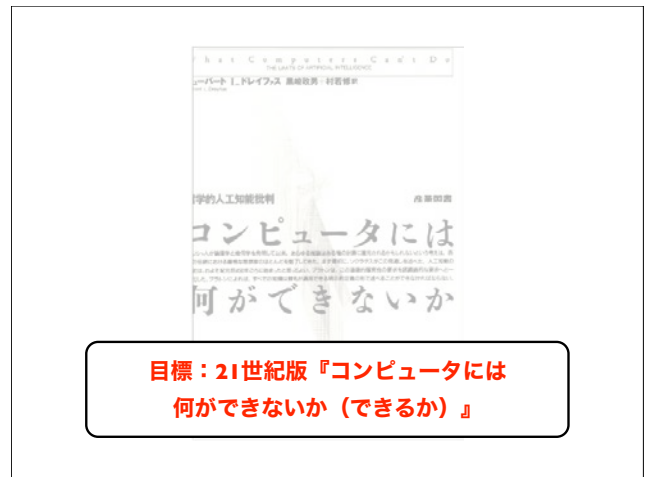
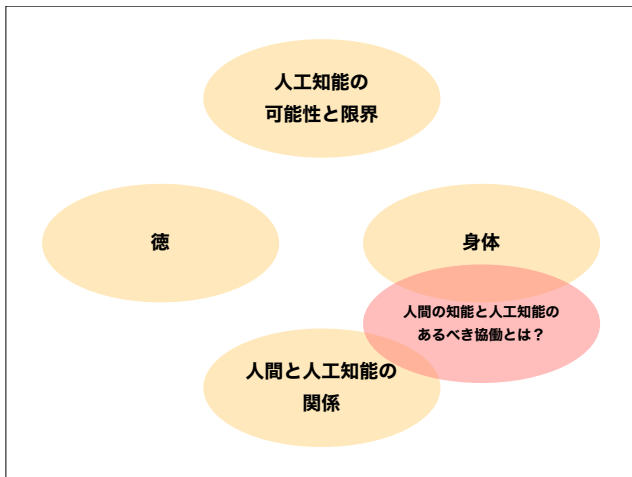


47



48





狭義の研究だけでなく...

- 文献リスト、読書ガイド、コンセプトマップなどの作成
- 研究者へのインタビュー
- 哲学研究者と人工知能研究者の対話イベント

など

プロジェクトのウェブサイト：  
<http://updatingphilosophyofai.net>