

全脳アーキテクチャ・イニシアティブ(WBAI)

良くも悪くも人類の福利厚生に対して大きな影響を与えるのは知能。

昨今は、ヒトのあらゆる知的能力を凌駕する汎用人工知能(AGI)の実現がさらに近づきつつある。

こうした中、私たちは、
「人類と調和した人工知能のある世界」というビジョンの実現を目指して、脳型人工知能の構築とその活用を推進してゆきます。



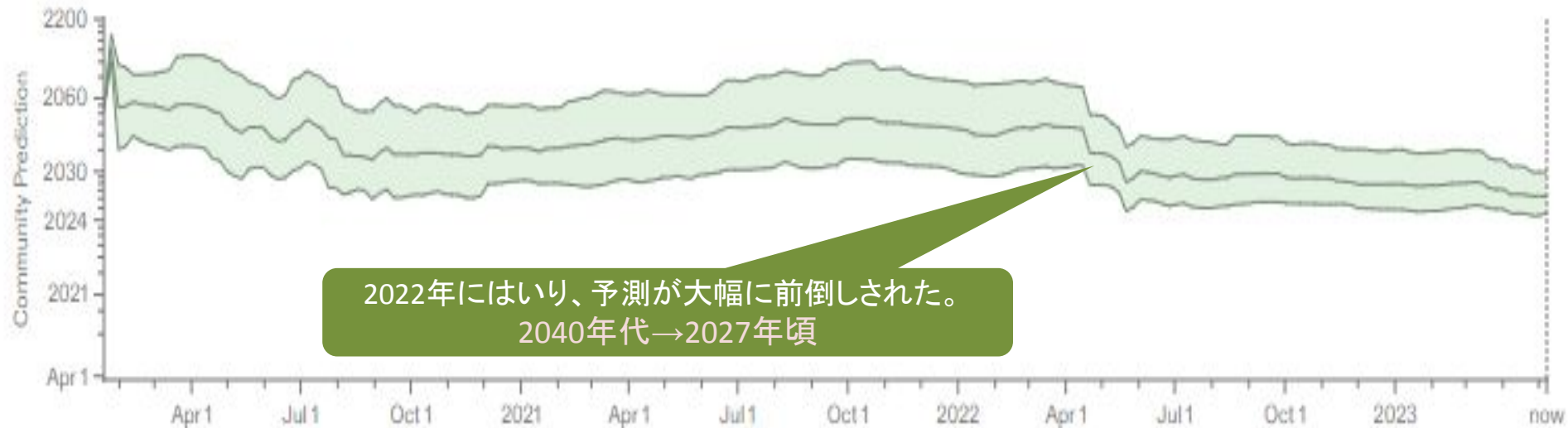
NPO法人WBAI代表

山川 宏

Date **Weakly** General AI is Publicly Known

<https://www.metaculus.com/questions/3479/date-weakly-general-ai-is-publicly-known/>

AI Progress Essay Contest



2022年にはいり、予測が大幅に前倒しされた。
2040年代→2027年頃

Total Forecasters: 857 Community Prediction: Mar 25, 2026

大規模生成モデルの進展をふまえた最近の動き



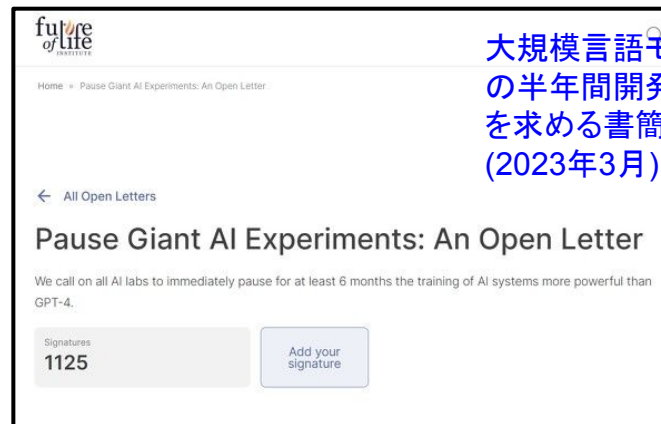
汎用AI、数年以内に実現の可能性＝ディープマインド CEO (2023/5/3)



AI想定より速く人知超える公算、危険性語るためグーグル退社＝ヒントン氏 (2023/5/2)



“We do not know how to control these things.”
Connor Leahy (2023/5/3)



大規模言語モデルの半年間開発停止を求める書簡 (2023年3月)

AGIに対する懸念の変化

従来はAGIを誰が開発するかが争点であった。

だがAGI実現の現実味が増したことで、
作られたAGIを誰も制御できないことがより大きな課題に

基本理念

ビジョン:

人類と調和した人工知能のある世界

ミッション:

全脳アーキテクチャのオープンな開発を促進

価値観:

まなぶ: 関連する専門知識を学び、広める

みわたす: 広く対話を通じて見識を高める

つくる: 共に作り上げる



人類と調和したAGIを目指すWBAIにおけるイベント



Beneficial AGIへ

第2回 全脳アーキテクチャ・シンポジウム

2017.8.29 TUE 13:00~17:15 ラゾーナ川崎 東芝ビル 15F

The End of Evolution and
the World after the Singularity

March, 19, 2020

Tokyo, Japan

生命進化の終焉とシンギュラリティ後の世界

脳型AIを目指して

パネル討論

私たちはいかに上手にAIの
手のひらに着地できるのか

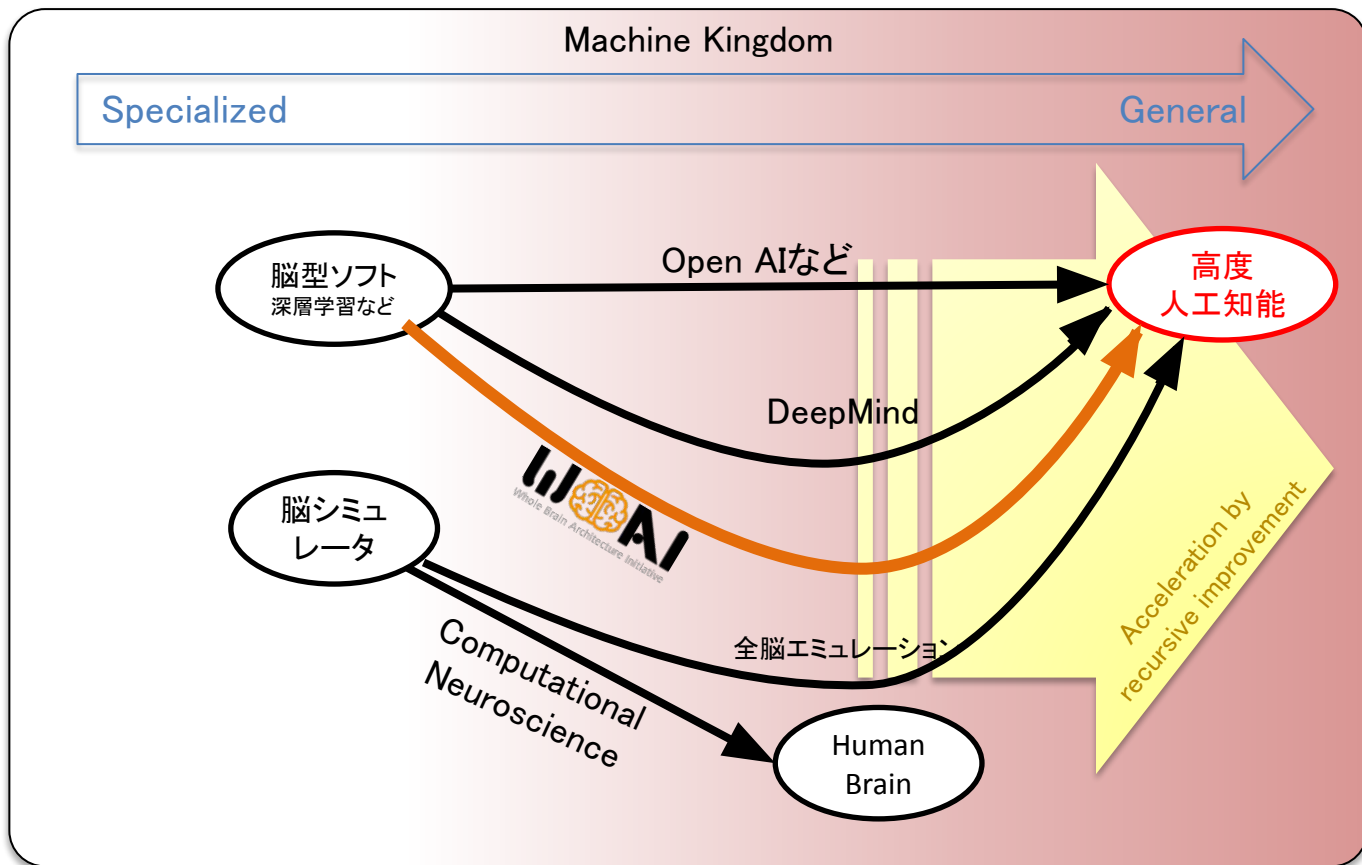


第6回全脳アーキテク
チャ・シンポジウム
2021年9月10日

AGIの安全性に関わる外部発表など

- 山川宏. & 松尾豊. 最も持続しやすい生命社会への移行を目指す「生命革命」シナリオ. 人工知能 vol. 38 254–265 (2023)
- 山川宏、汎用AIが投げかける人間の本質への根源的な問い, フロネシス, vol.12, no.1, pp.116–121, April, 2020.
- 高橋恒一, 知能爆発はいつどのように起きるのか, シンギュラリティサロン@SpringX, 2021年1月30日.
- 山川宏, 技術進展がもたらす進化戦略の終焉– 新たなる生命社会の形を求めて –, 進化の終焉とシンギュラリティ・セミナー, 2020年3月19日.
- Yamakawa, H. (2019). Peacekeeping Conditions for an Artificial Intelligence Society*Big Data and Cognitive Computing*3(2), 34.
- 山川宏, 人間社会を支えるエコシステムとしての自律AI社会—存在論的リスクを克服するために—, シンギュラリティサロン#40, 大阪府, 2019年11月30日.
<https://singularity40.peatix.com/>
- 山川宏, 人間社会を支えるエコシステムとしての自律AI社会—存在論的リスクを克服するために—, シンギュラリティサロン@東京 第37回公開講演会, 東京, 2020年1月12日.
- Yamakawa, H., Specifications of brain-inspired AGI development for everyone, Beneficial AGI 2019, 2019/1/5.
- Yamakawa, H., et al. (2018) Strategy to Build Beneficial General-Purpose Intelligence Inspired by Brain, 日本神経回路学会第28回全国大会 (JNNS2018) 予稿. (スライド: <http://bit.ly/2vxNtM0>)
- 山川宏, 人類のための人工知能のグローバルな役割, Bloomberg TV Bulgaria, World is Business, 02/21/2019, <http://bit.ly/2yumrFe>
- 山川 宏, 速度耐性をもつ社会の実現にむけた基礎的考察, 2018年度 人工知能学会全国大会 (第32回), 1F3-OS-5b-01, 2018
- 高橋 恒一, 将来の機械知性に関するシナリオと分岐点, レクチャーシリーズ:「シンギュラリティとAI」[第7回], 人工知能,33, (2018-11-01)
- 山川 宏, 機械知能社会は技術的特異点を超えられるか, レクチャーシリーズ:「シンギュラリティとAI」[第7回], 人工知能,33, (2018-11-01)

高度人工知能への道筋



ヒトと似た思考をするように AGIを作成もしくは改変する

AGIが脳型だからこそ可能になること:

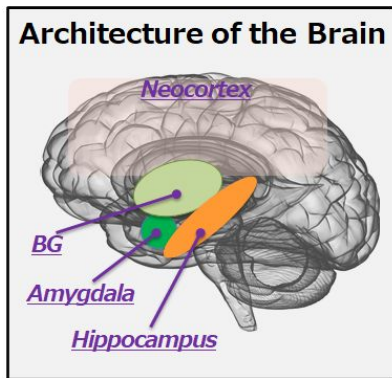
- ヒトらしい感情や行動変容を再現するAI Interface
- 人格をアップロードするためのソフトウェア基盤
- 同等に高速動作する他のAGIの思考の流れを説明する

脳型AGIが人類に対して果たしうる役割（案）

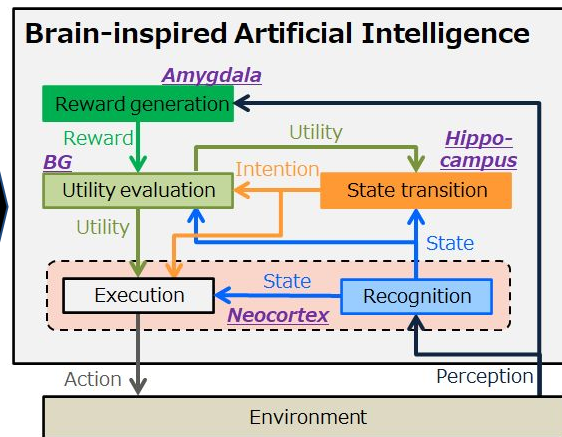
仮に、別経路によりAGI/超知能が先に実現されたとしても、超知能の速度でヒトの様に動作(思考/認識/運動)する脳型AGIは人類にとって大いに役立つであろう。

- [伝道者] 脳型でない超知能の思考の流れやその社会での議論の結果を、人類に理解できるように説明する
- [擁護者] 超知能社会において、現生人類を存続させる意義を訴え、その福祉向上を誘導する
- [保全者] 人類に近い価値観や(仮想的な)身体性を持ちつつ、人類ミームを継承する
- [後継者] 脳型超知能を我々人類が産み出した子孫であると認めて、彼らが発展的に生存することを応援する

脳全体のアーキテクチャに学び人間のようなAGIを創る(工学)



- ① 脳の各器官を機械学習モジュールとして開発
② それらを統合した認知アーキテクチャの構築



全脳アーキテクチャ中心仮説

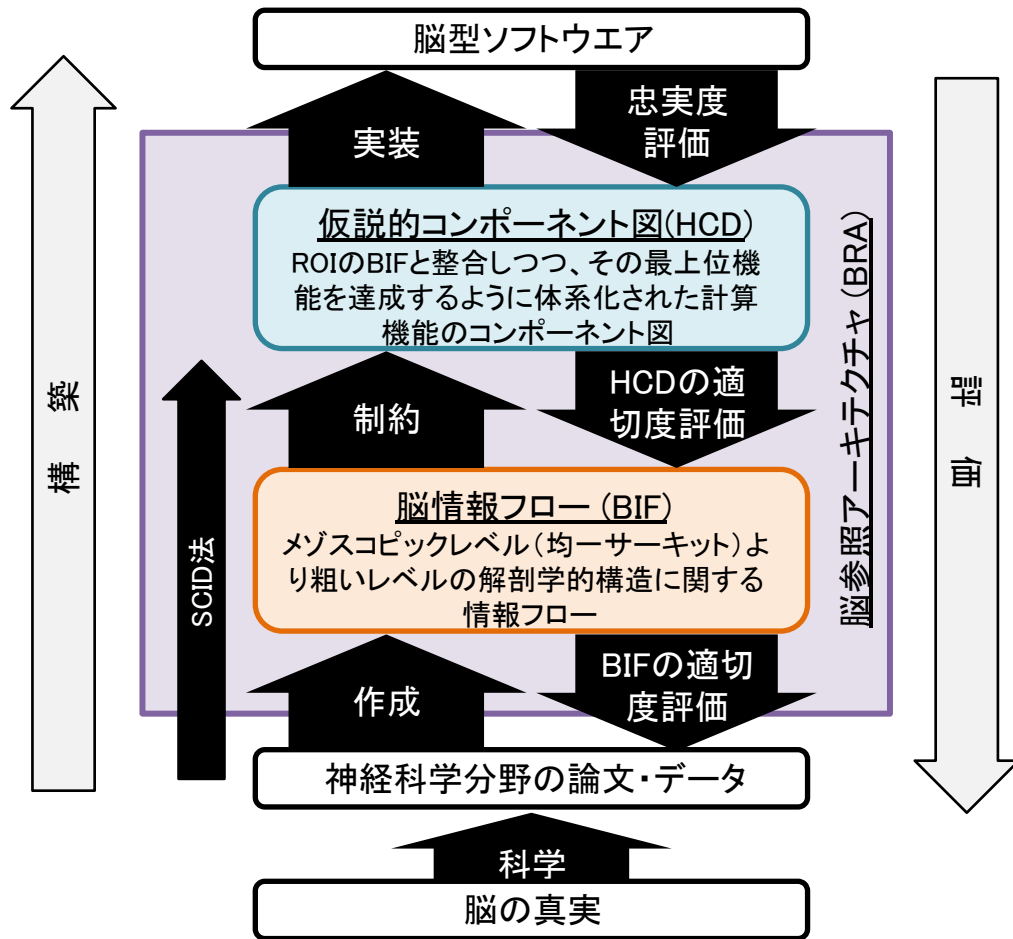
”脳はそれぞれよく定義された機能を持つ機械学習器が一定のやり方で組み合わせられる事で機能を実現しており, それを真似て人工的に構成された機械学習器を組み合わせる事で人間並みかそれ以上の能力を持つ汎用の知能機械を構築可能である ”

WBAIが技術を蓄積してきたBRA駆動開発

脳型AGIの設計データを構築する

脳神経回路のリバースエンジニアリングによって、設計データである脳参照アーキテクチャ(BRA)を作成する。

そのBRAに基づいて脳型ソフトウェアを構築する。



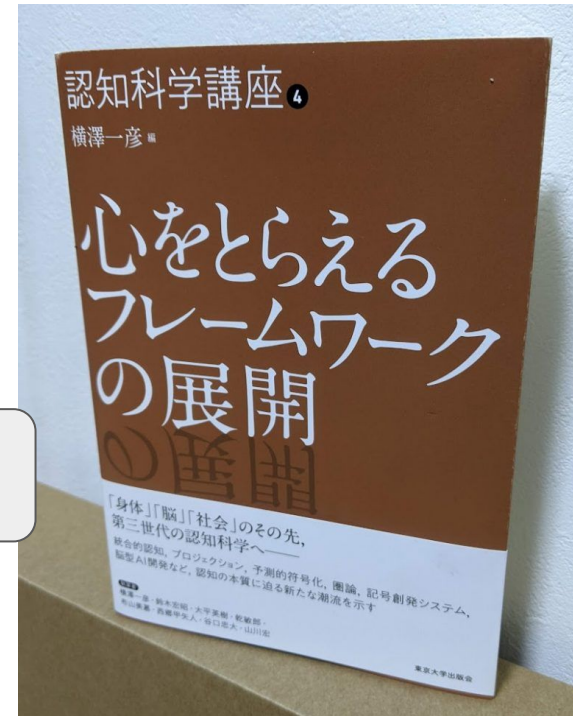
全脳アーキテクチャの解説を含む

書籍「心をとらえるフレームワークの展開」の発売

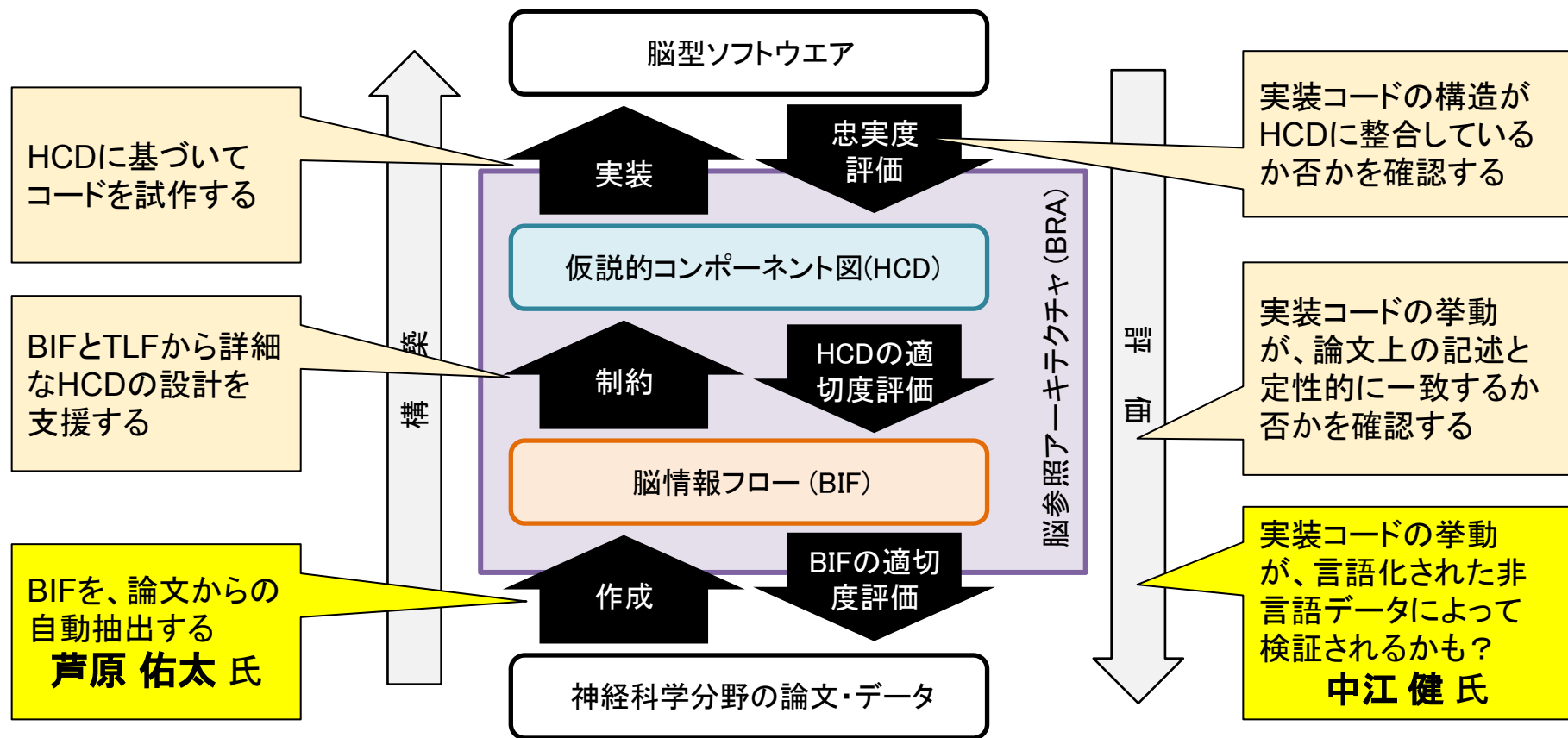
目次:

- 序 第三世代の認知科学の可能性(横澤一彦)
- 第1章 統合的認知(横澤一彦)
- 第2章 プロジェクション科学(鈴木宏昭)
- 第3章 内受容感覚の予測的処理(大平英樹)
- 第4章 自由エネルギー原理(乾敏郎)
- 第5章 圏論による認知の理解(布山美慕・西郷甲矢人)
- 第6章 記号創発ロボティクス(谷口忠大)
- 第7章 **全脳アーキテクチャ**
——**機能を理解しながら脳型AIを設計・開発する**(山
川 宏)

全脳確率的
生成モデル



大規模言語モデル活用はBRA駆動開発を様々に加速しうる



WBAIのビジョンにむけた技術ロードマップ

