

# 脳に寄り添って 汎用人工知能を目指す

山川宏



Whole Brain Architecture Initiative



## 研究促進事業

### 促進型研究開発

- 神経科学への接地
- 統合プラットフォーム
  - BriCA言語, BriCAコア
- AGI評価
- 学習環境構築

ドワンゴAI,  
理研らと共同

### 個別研究の支援

- 若手研究者の支援
  - メンター, 財政支援
- 既存研究者との連携

年間予算規模: 650万円程度  
(主な収入源は賛助会員会費)

## 人材育成事業

- イベント開催
  - WBA勉強会, ハッカソン
- 教材作成
- 専門勉強会

主にボラン  
ティア

## 広報事業

- インターネット広報
- 賛助会員リレーション
- ボランティアリレーション

主にボラン  
ティア

## 事務局

- 総会等の集会
- 財務, 経理
- 戦略のとりまとめ

兼任職員2名

# AGIの構築はなぜ脳から学ぶことが有効か



※AGIは必ずしも脳に学ぶ必要はないが、超えるまでは役立ちうるだろう

## 人との親和性が高いAI

人と同様に振る舞うAIに近づきやすい。マインドアップローディングの基盤技術になる

## AGIへの到達が保証されている

段階的に詳細化すれば、いずれはAGIへ到達する。

## 方向性の確認

研究開発が正しい方向に進んでいるかを検証できる。

## 汎用技術設計の困難さに対処

時に「作り込むほどにAGIから離れる」という性質をもつ。目的機能を限定しないままのAGI設計は人間には困難

## 未解決課題のヒントが得られる

例えば、「柔軟な再利用性」や「カリキュラム学習」のヒント。少なくとも、パラメータの組合

## 分散共同開発

個別に開発すべき作業を特定することで、多くの研究開発を統合できる。また不足している機能を特定できる。

## 知識を集約するための足場

神経科学や認知科学の知識を集約し役立てることができる。

AGIの開発において、脳はガイド：  
「生物学的な真実との一致性」は**目的ではなく手段**



# AGIに必要な再利用性



## 現在の特化型AI

- 個別の領域において知的に振る舞う
- 既に人以上の能力が数多く実用化されている。例えば
  - コンピュータ将棋/チェス
  - Googleカー(自動運転)
  - 医療診断

専門性を設計する

## 目指す汎用人工知能 (AGI)

- 多様で多角的な問題解決能力を自ら獲得する
  - 設計時の想定を超えた新しい問題を解決できる。
  - 自己理解／自律性
- AI創世記からの夢でありつつ、実現の困難さから取り組みは少なかった。

専門性を学習する



## 遭難するルンバ



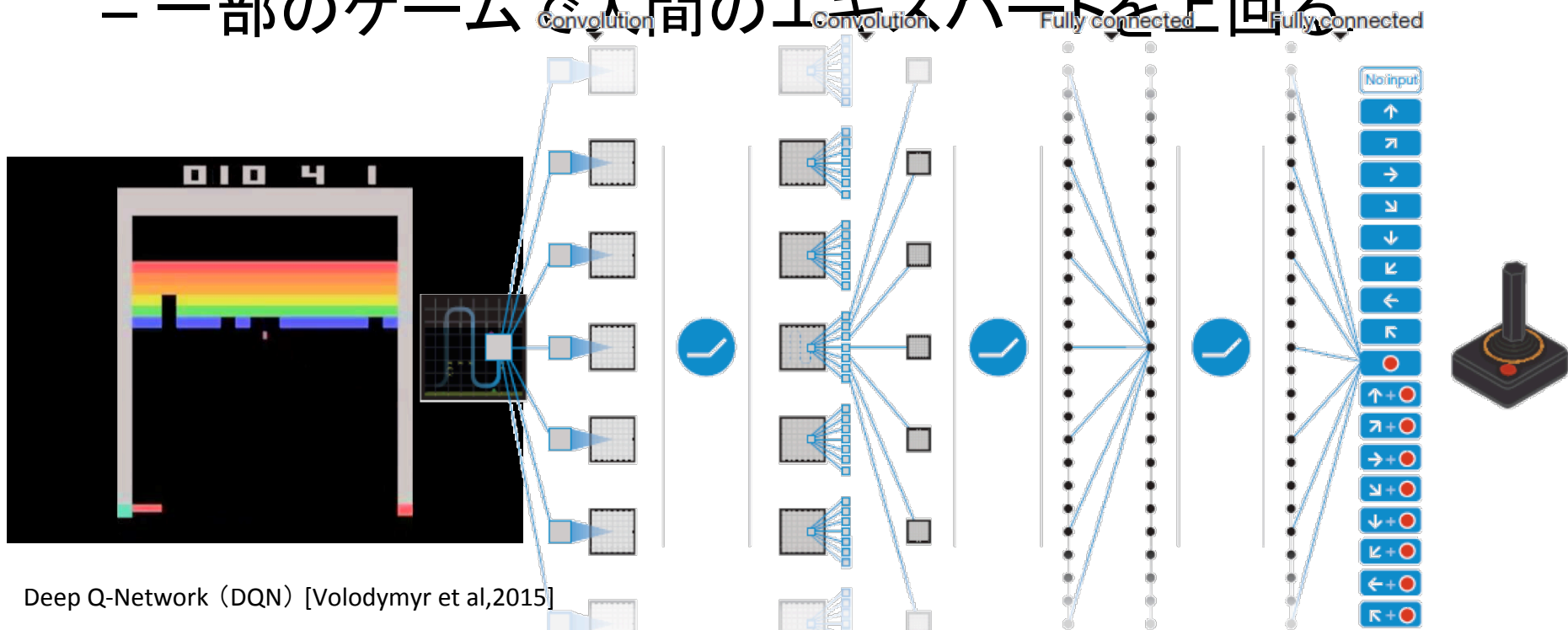
少なくとも実世界で活動するエージェントにおいては、物理環境内で生き延びられる汎用知能が必要である。

**特化した作業は設計して効率化したいが、例外状況への対応は汎用知能でカバーして欲しい**

# いま現実となった「強化学習 + 深層学習」



- 画像をそのまま入力する
- 行動価値関数  $Q(s,a)$  を深層学習で近似
- Atari 2600ゲームで、既存手法を圧倒
  - 一部のゲームで人間のエキスパートを上回る



Deep Q-Network (DQN) [Volodymyr et al, 2015]



- 5年前：設計特化
  - タスクごとに人間が設計する必要があった
- 現在：学習特化
  - 汎用的な学習
    - 一つのアルゴリズムで多くのタスクを学べるようになった
  - 技術的に見ると
    - 強化学習は汎用
    - 深層学習で一般物体認識ができるようになった
- 次のステップ：本当の汎用
  - 一つのAIが、複数のタスクを学ぶ
  - そこでの知識を転移／流用する





## 表現の獲得能力

- 遅いプロセス
- 領域内共有知識の学習
- 深層学習である程度で  
きるようになった

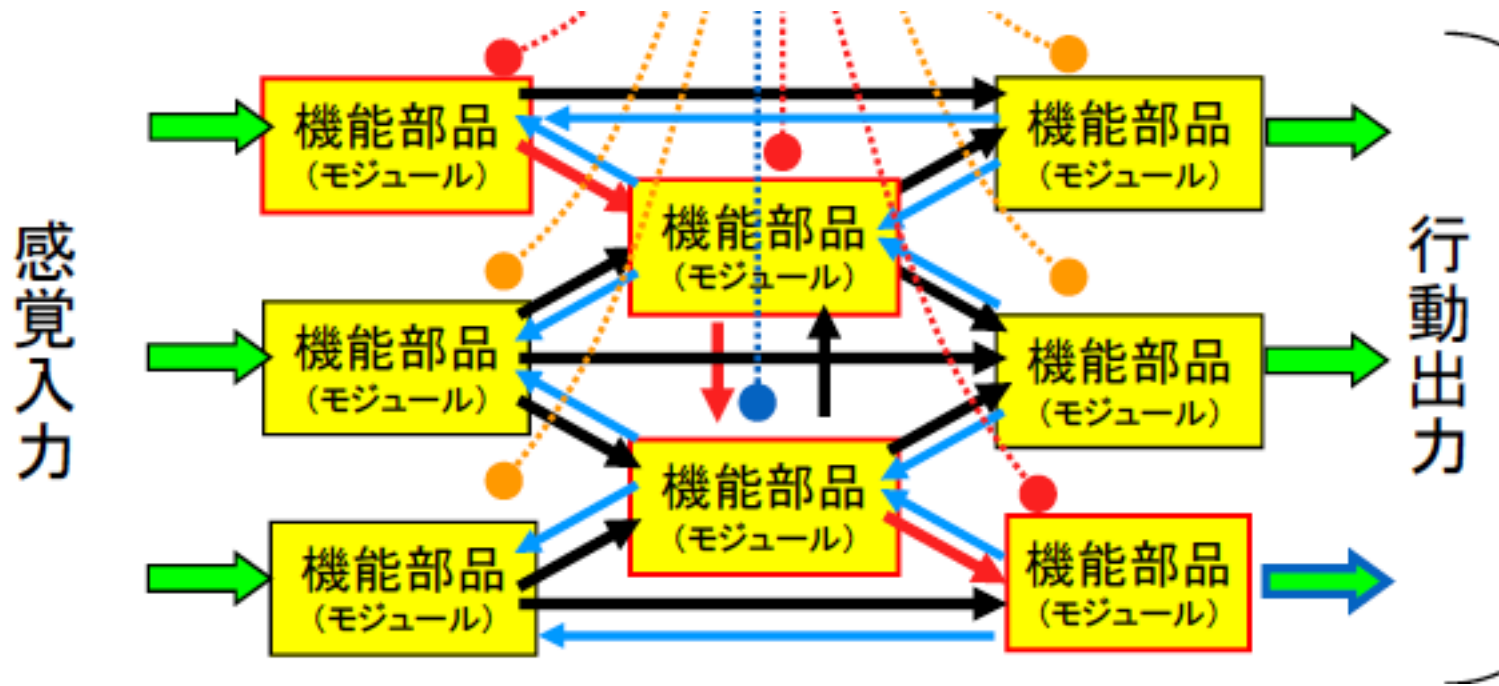
## 知識を組合せる能力

- 素早いプロセス
- ソフトな再利用性が課題  
– 脳に学びうるだろう

# AGIのための認知アーキテクチャに必要なこと



AGIを支える認知アーキテクチャは、再利用可能な部品を動的に組合せ、予期しない／未知の事態にも何らかの振舞いを行う必要がある。



機能部品組合せ仮説のシステムデザイン(小川, 大森)

脳においては複数の異なる機能を持つモジュールを状況に応じて組み合わせられて構成される神経回路の動作で、これを実現しているだろう。



# 設計の難しい汎用知能



試作

生物等  
に学ぶ

領域知  
識に基  
づく設計

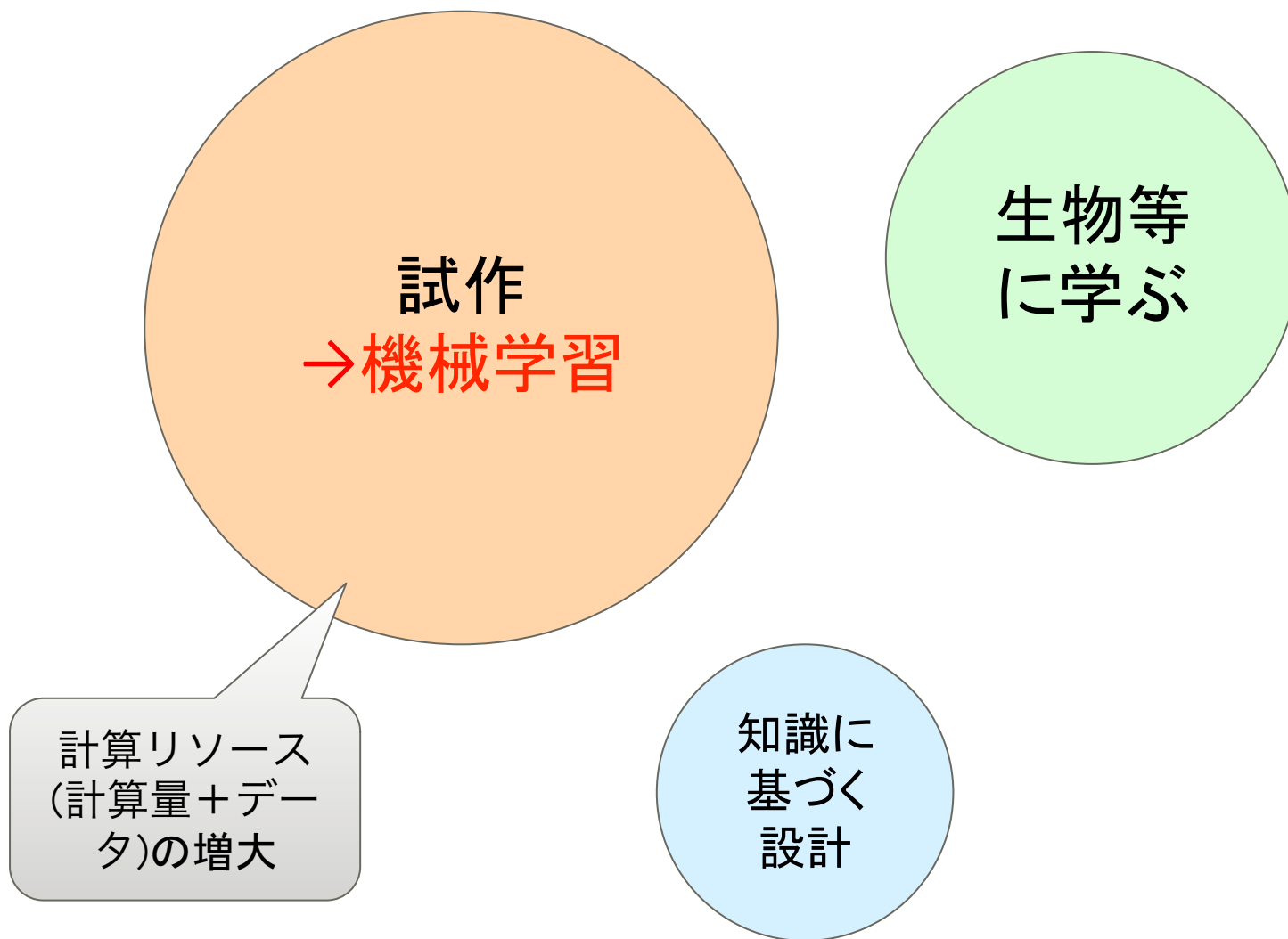


試  
作

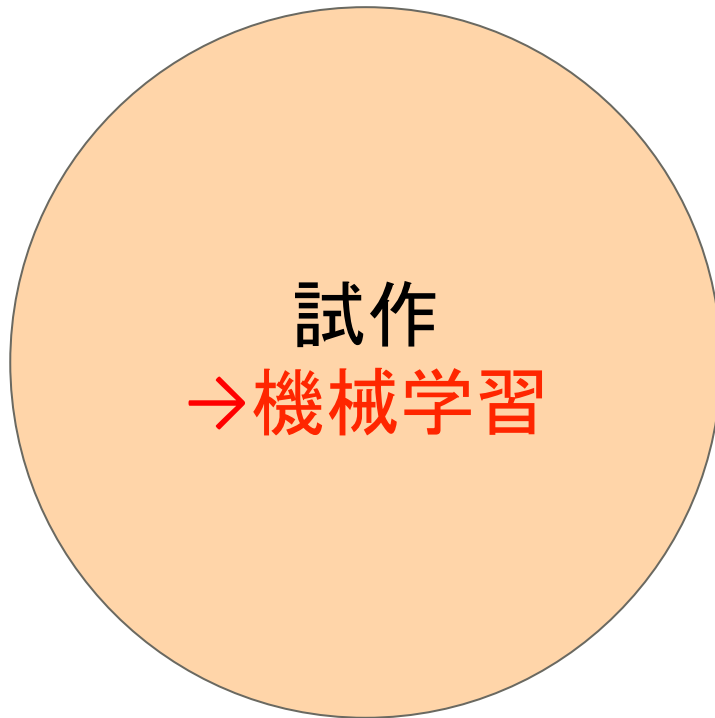


領域知識に  
基づく設計

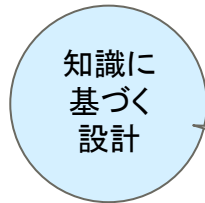
理解に基づく設計： 実現したい機能を分解して機構に割り当てる



# 汎用人工知能の難しさ



自然淘汰で生き残ったアーキテクチャ



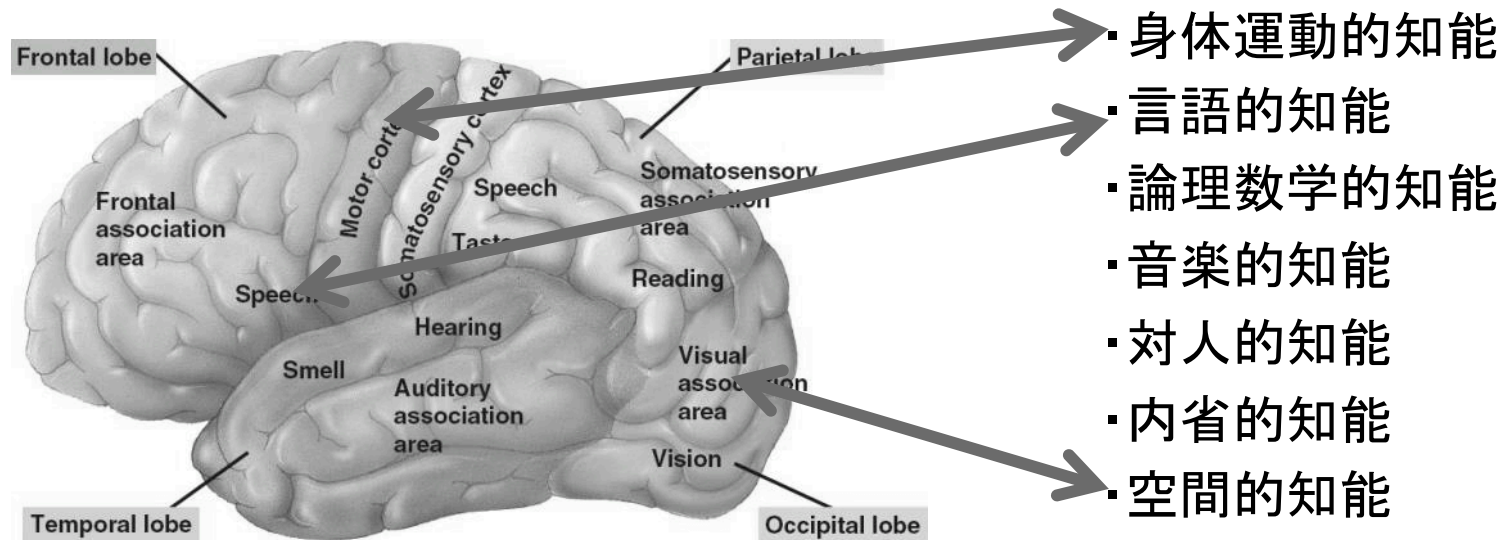
目的が不明確なため、人は機能に基づいた設計ができない



脳に学んで、  
理解し得ない汎用性をつくる



# 学習する汎用性の基盤は大脳新皮質の一様さ



図の出典: [http://bio1152.nicerweb.com/Locked/media/ch48/48\\_27HumanCerebralCortex.jpg](http://bio1152.nicerweb.com/Locked/media/ch48/48_27HumanCerebralCortex.jpg)

- 一様な機構で、多様な機能を獲得→汎用性
- 大脳新皮質において実現されている
  - 工学的なモデルが存在する(深層学習)

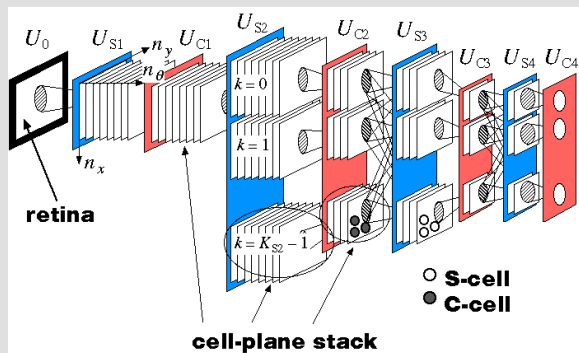


一般物体認識を実現した  
ネオコグニトロン(視覚野)

脳に学んだことは何か？

複雑細胞と単純細胞から構成される基本単位

階層的アーキテクチャによる繰り返し利用



汎用人工知能を実現する  
全脳アーキテクチャ

脳に何を学ぶべきか？

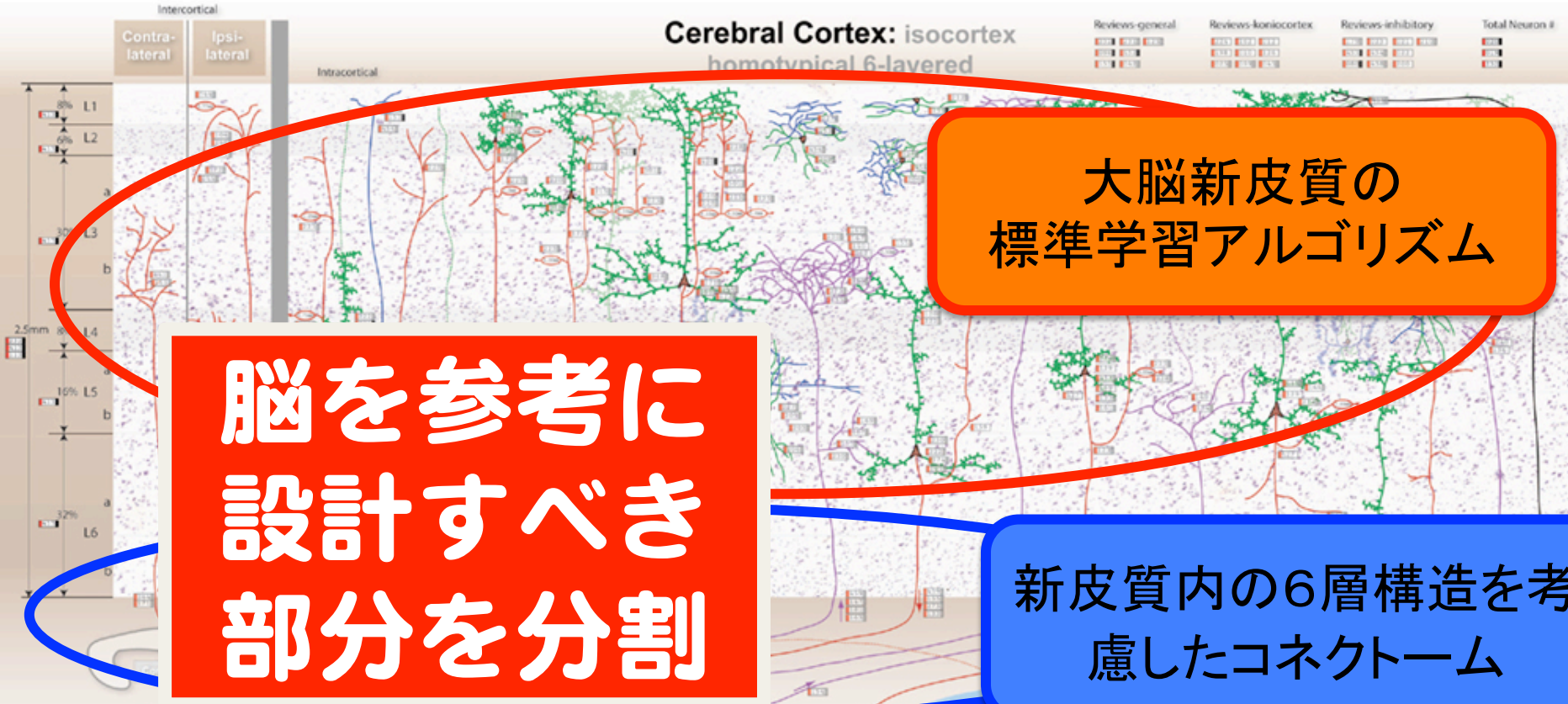
大脳新皮質の  
標準学習アルゴリズム  
認識だけでなく注意や運動も含んだ学習の単位モジュール

新皮質内の6層構造を考慮したコネクトーム  
標準新皮質の単位モジュールを結合したアーキテクチャ



# Cerebral Cortex: isocortex homotypical 6-layered

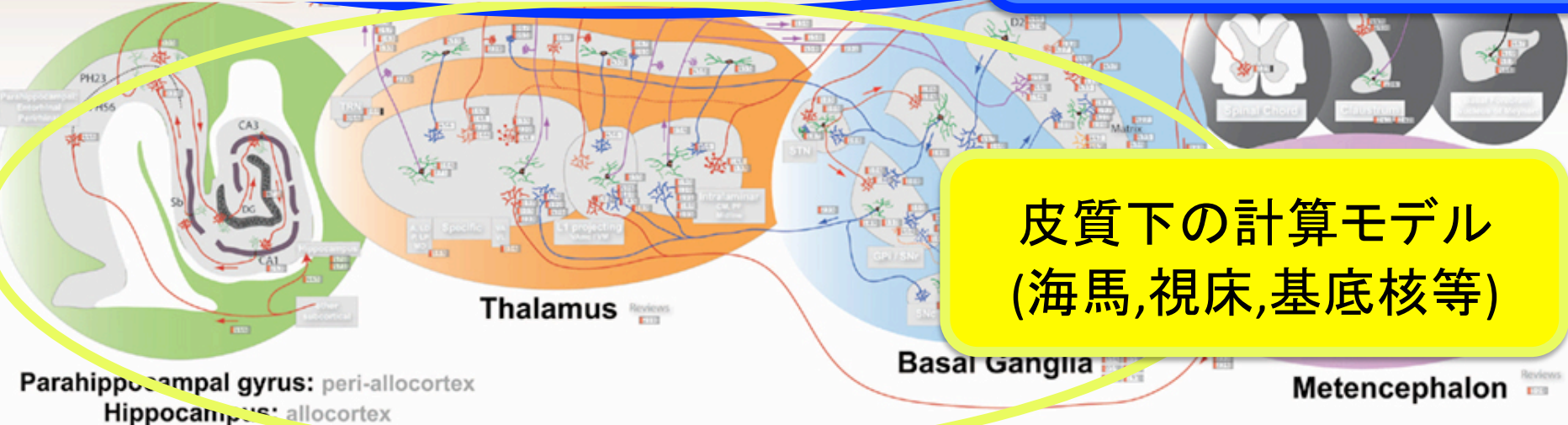
Reviews-general	Reviews-koniocortex	Reviews-inhibitory	Total Neuron #
100 100 100	100 100 100	100 100 100	100
100 100 100	100 100 100	100 100 100	100
100 100 100	100 100 100	100 100 100	100



大脳新皮質の  
標準学習アルゴリズム

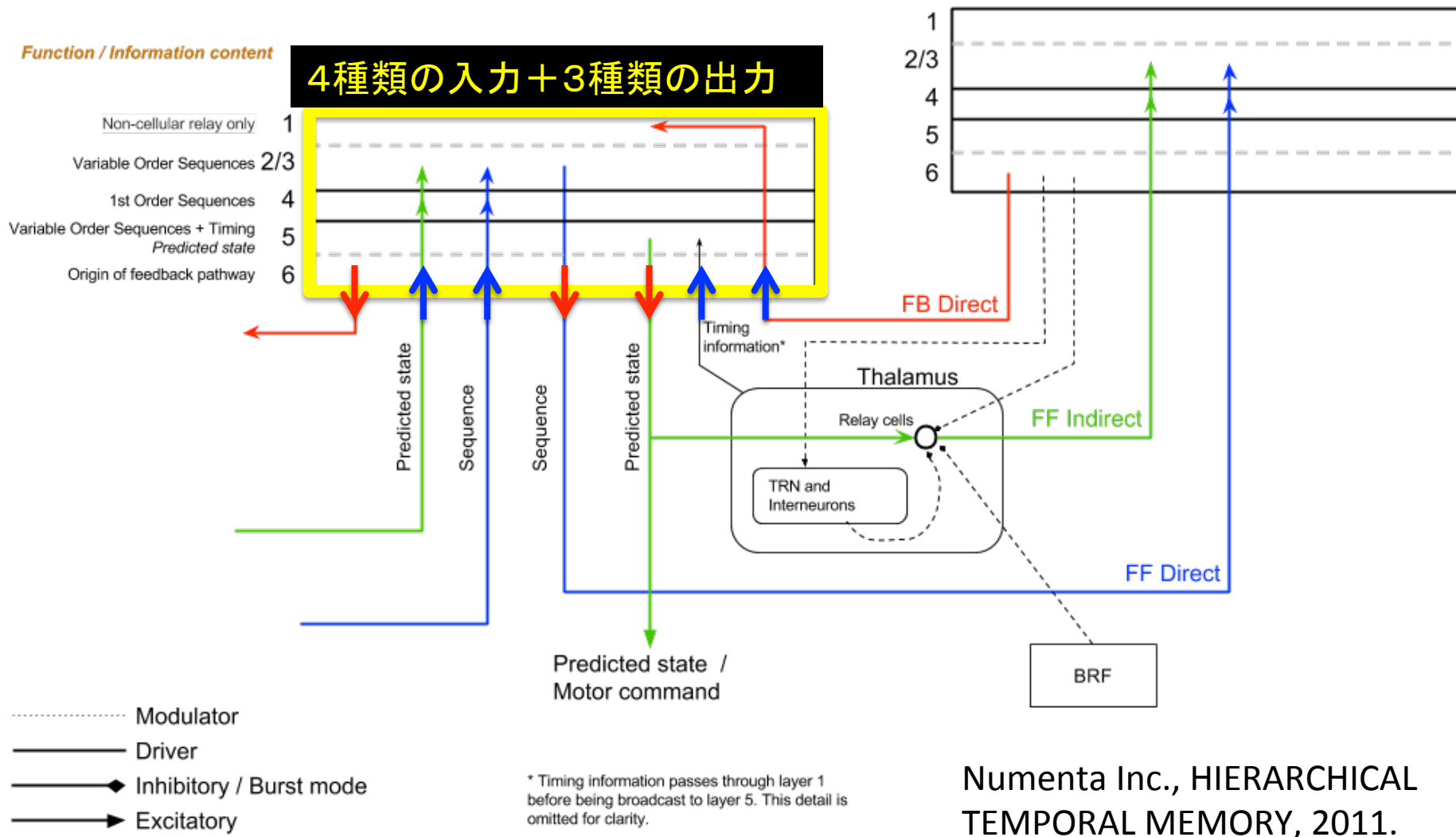
脳を参考に  
設計すべき  
部分を分割

新皮質内の6層構造を考  
慮したコネクトーム



皮質下の計算モデル  
(海馬, 視床, 基底核等)

# 大脳新皮質の標準学習アルゴリズム



具体的学習アルゴリズムはこの入出力あわせて設計すれば良い



様々な候補が検討されている

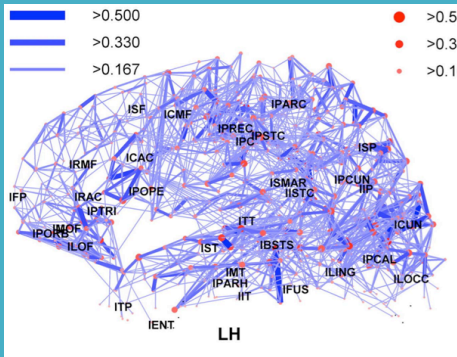
- BESOM@一杉
- Hierarchical Temporal Memory
- CNN (認識部分のみ)
- などなど



# 新皮質内の6層構造を考慮したコネクトーム

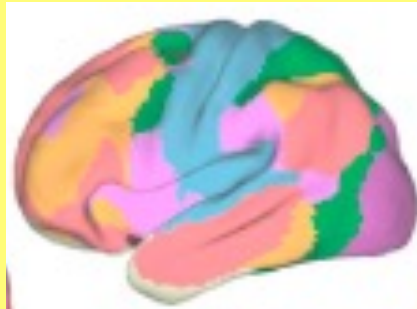
## 様々な意味で“Network of Networks”である

### 構造的ネットワーク



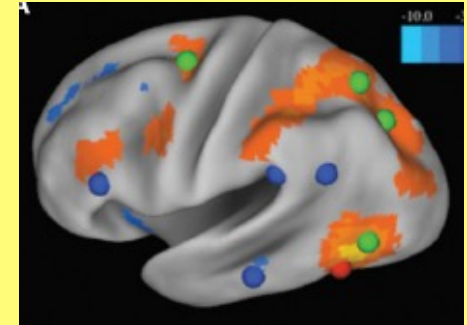
一貫したトポロジーを維持

### RSFC 安静時機能結合



様々なコミュニティ: ①モダリティ固有, ②情報統合, ③課題の制御

### 課題に依存した 機能的ネットワーク



コンテキスト、刺激、そして課題に依存して再構成

(Steven E. Petersen, Olaf Sporns, Brain Networks and Cognitive Architectures, Neuron, 2015)

構造的ネットワークに制約された形で、動的な神経活動が展開することで多様な機能を実現

→ 汎用性 = 理解不能な再利用性



# 最新のAIを脳に接地させるために



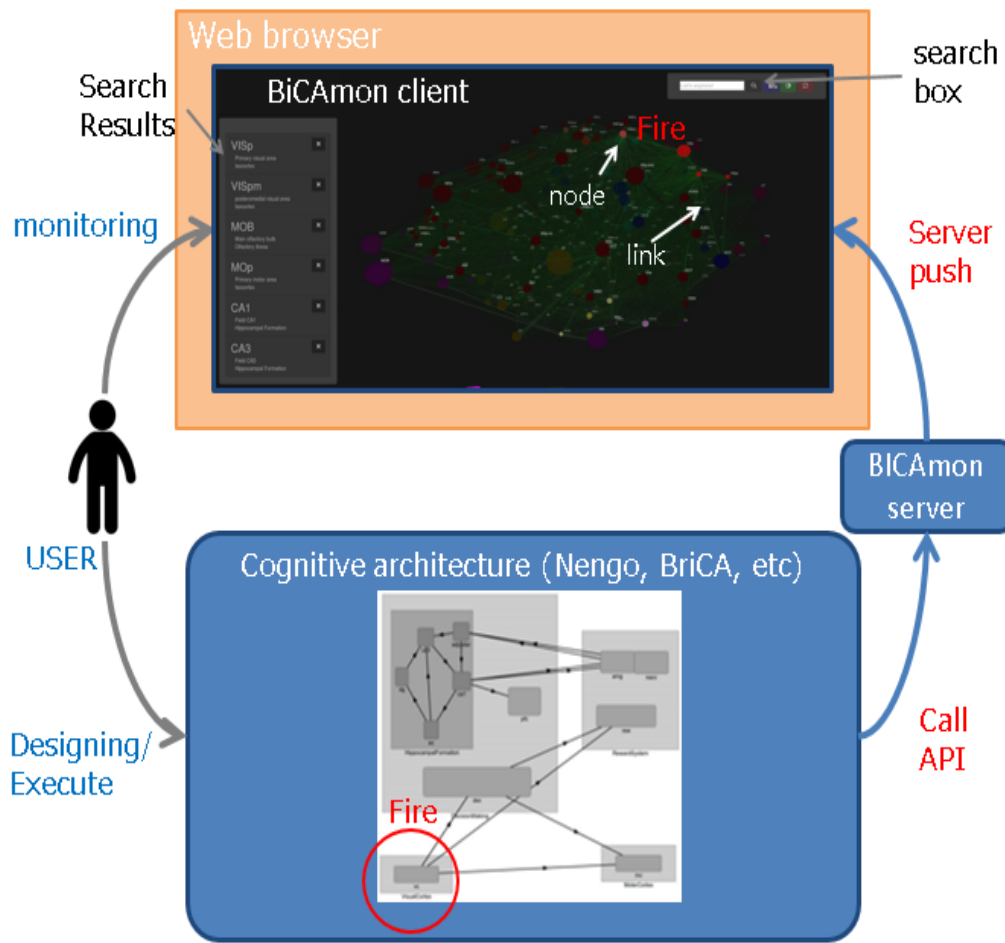
認知アーキテクチャの活動状態をコネクトーム上でモニタリングするツール

## 目的

- 学習アルゴリズムをコネクトーム（地図）に接地する
- WBA研究の理解促進
- 機械学習モデルを神経科学実験との対応づけるツール
- 神経科学者との接点

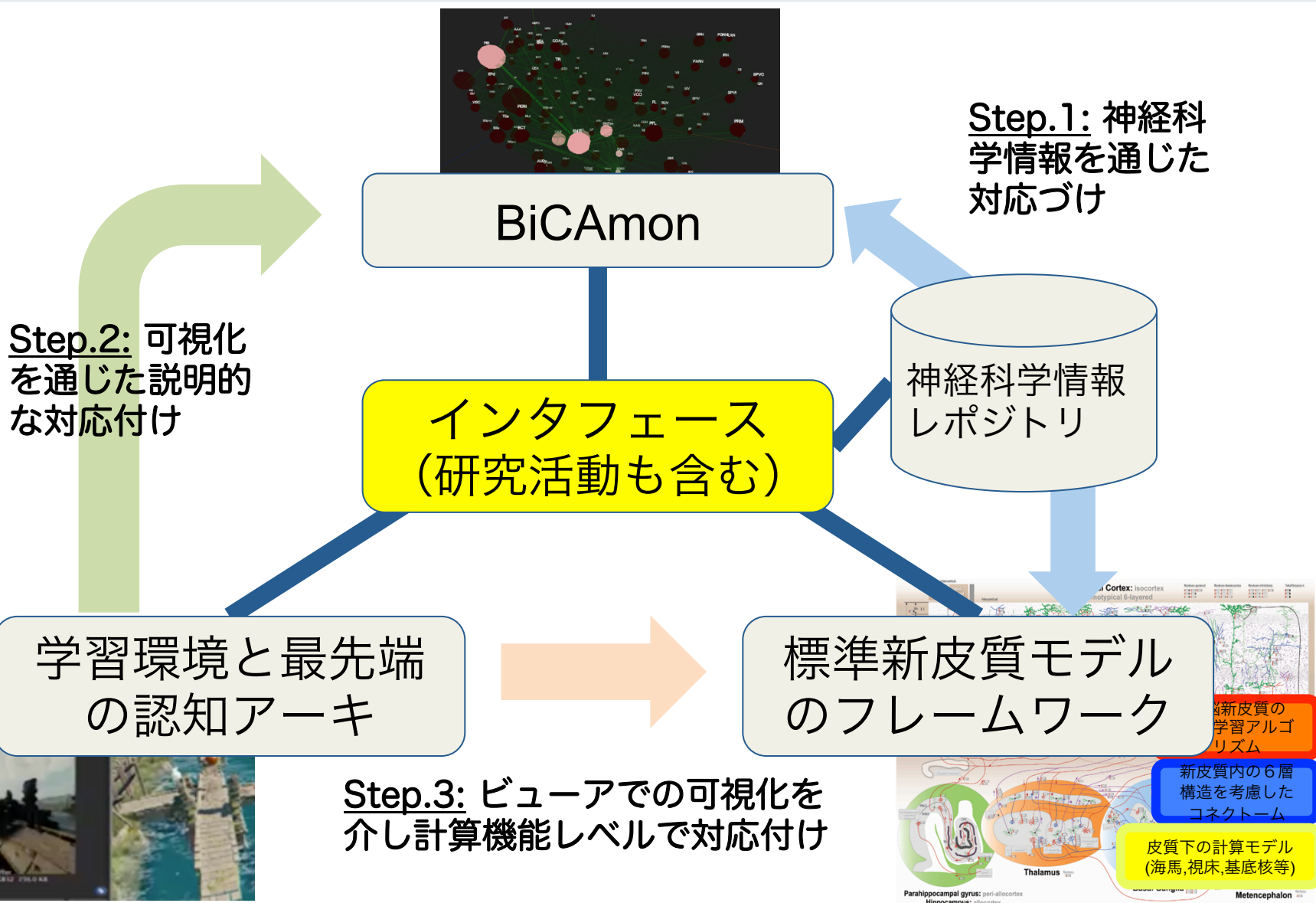
## グラフビューアとしての特徴

- 脳内位置に応じた3D表示
- 脳内の大きさに
- 活動状態のダイナミックな表示



ビデオ: <https://www.youtube.com/watch?v=1-iJccDBTsM&feature=youtu.be>







LISと標準新皮質フレームワークから想像される:

**三歳児WBAを完成させる100日間**



## ○実験設定

(まずは大脳新皮質のみを考える)

- 標準学習アルゴリズムの候補(組合せ含む) =  $10^2$ 個
- 領野毎のパラメータセット =  $10^5$ 個
- コネクトームパラメータセット =  $10^5$ 個

組合せの合計:  $10^{12}$ 個

## ○計算速度

- リアルタイムの1000倍で動くAIと学習環境  
三歳児までの1000日→1日
- 動的探索による枝刈り =  $10^3$   
(\*  $10^2$ イテレーション)
- 並列度:  $10^5$  台

探索可能な組合 =  $10^{10}$

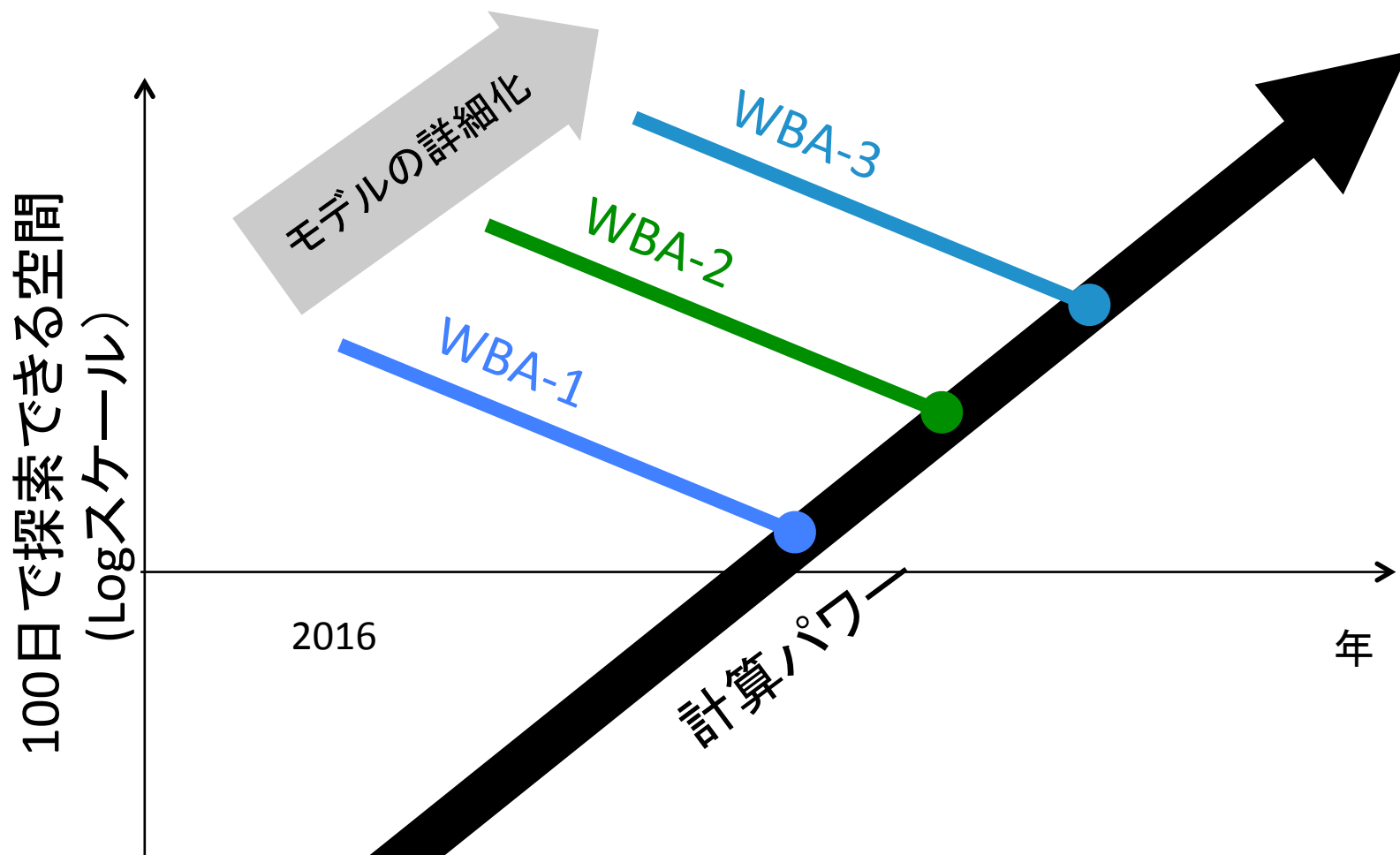
必要な研究期間:

$$10^{(2+5+5)} / (10^{(2+3+5)}) = 100 \text{ 日}$$

# 100日完成プロジェクトはいつ可能になるか？



神経科学や機械学習の進展で、探索すべきWBAも年々小さくなる



こうした見積もりを真面目に考える時期に来ている