制御可能性の原則についての意見

栗原　聡

電気通信大学大学院情報理工学研究科／
人工知能先端研究センターAIX

ドワンゴ人工知能研究所
DWANGO ARTIFICIAL INTELLIGENCE LABORATORY

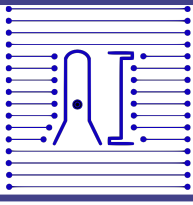
al+

2016 TSUGAIKE

UEC TOKYO

WBAI Whole Brain Architecture Initiative

透過性と制御可能性の確保という議論



① **透明性の原則**
ＡＩネットワークシステムの動作の検証可能性及び説明可能性を確保すること。

② **利用者支援の原則**
ＡＩネットワークシステムが利用者を支援し、利用者に選択の機会を適切に提供するよう配慮すること。

③ **制御可能性の原則**
人間によるＡＩネットワークシステムの制御可能性を確保すること。

④ **セキュリティ確保の原則**
ＡＩネットワークシステムの頑健性及び信頼性を確保すること。

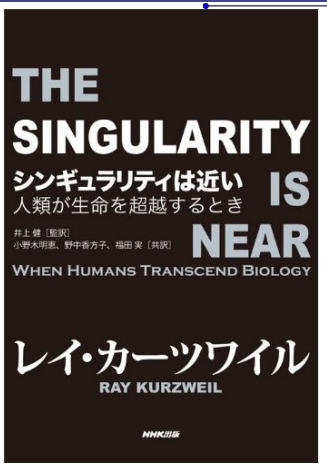
⑤ **安全保護の原則**
ＡＩネットワークシステムが利用者及び第三者の生命・身体の安全に危害を及ぼさないよう配慮すること。

⑥ **プライバシー保護の原則**
ＡＩネットワークシステムが利用者及び第三者のプライバシーを侵害しないように配慮すること。

⑦ **倫理の原則**
ＡＩネットワークシステムの研究開発において、人間の尊厳と個人の自律を尊重すること。

⑧ **アカウンタビリティの原則**
ＡＩネットワークシステムの研究開発者が利用者など関係するステークホルダーに対しアカウンタビリティを果たすこと。

指数関数型変化の変曲点にある現在



③そのあと間もなく、限りなく「垂直」に近い地点が訪れる（ここがシンギュラリティ）。

②ある時期から急に変化が感じられるようになり、一気に直線的変化を追い抜く。

①初めは「水平」的で無変化に感じられる時期が長く続く。直線的変化の方が大きいため騙される。

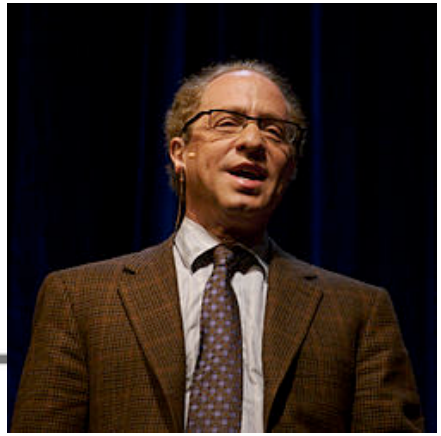
技術進化

時間

直線的変化

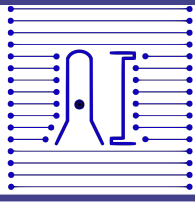
指数関数的変化

Deep Blue 1997
将棋終結宣言　2015
AlphaGO　2016



http://dentsu-ho.com/articles/3260

注視すべきことのレベル



可読性のある手法か？
可読性のない手法か？

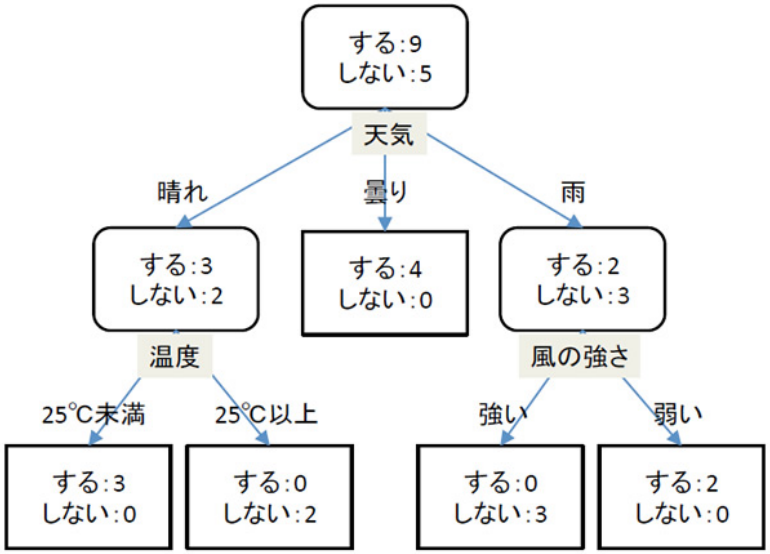
大規模システムか？
中小規模のシステムか？

難しくなる

トップダウン型のシステムか？
ボトムアップ型のシステムか？

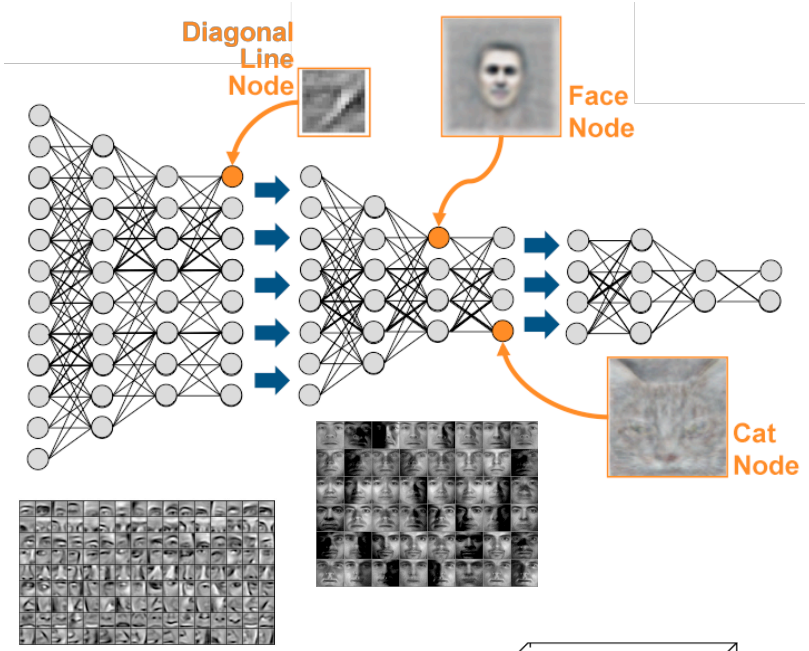
受動的なシステムか？　（道具としての性格）
能動的なシステムか？（自律性・共生する関係）

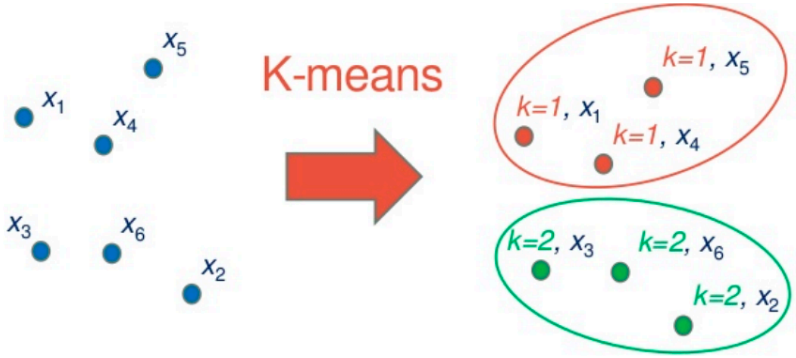
可読性の問題（人にとっての）



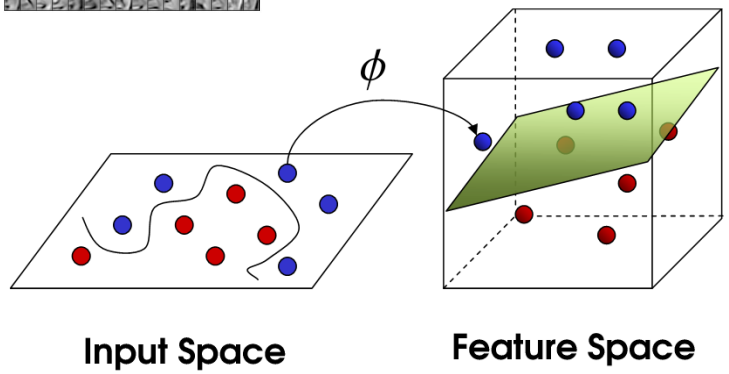
http://gihyo.jp/dev/serial/01/mahout/0006

http://theanalyticsstore.ie/deep-learning/





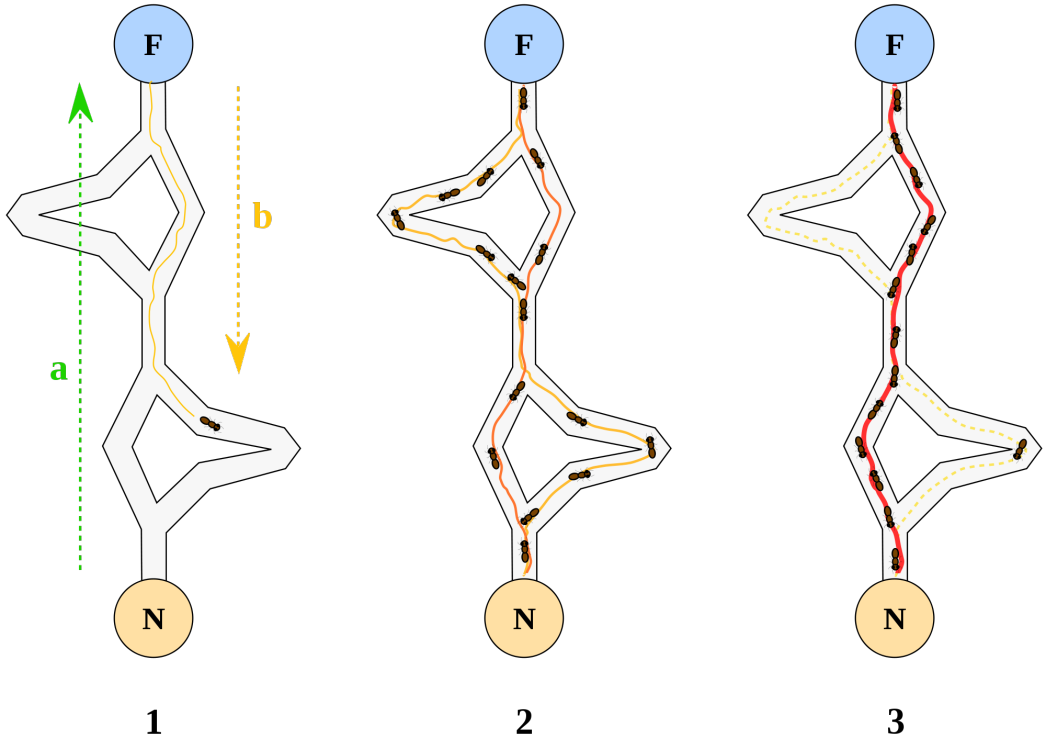
http://www.slideshare.net/oscillograph/6-51143088



http://www.tsjshg.info/udemy/Lec82-83.html

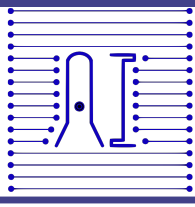
群知能における可読性

個の可読性はあっても群としての可読性はなし



http://www.sciencedirect.com/science/article/pii/S0142061515005840

システムの規模の問題



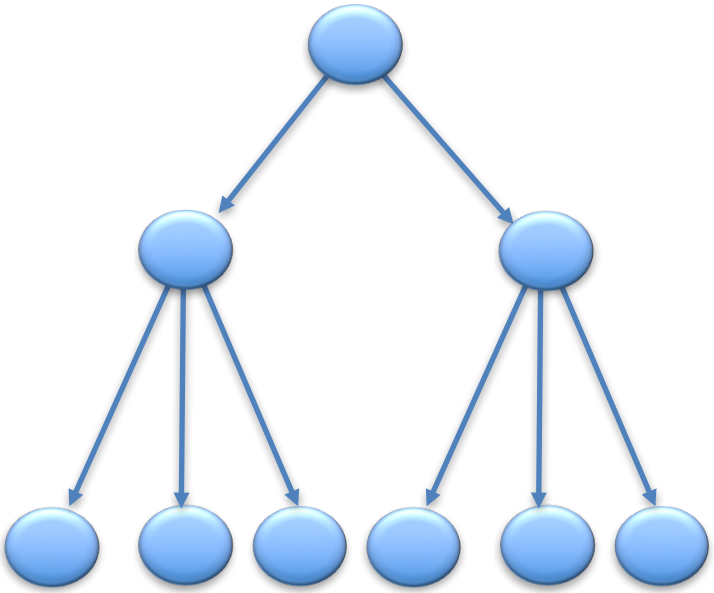
動作が予測できるレベルのシステム
- 家電，玩具用ロボットなど

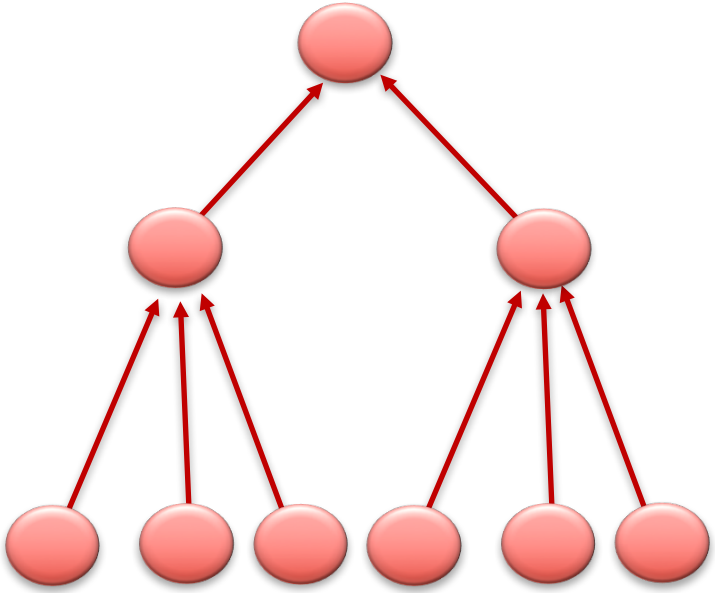
スケールした大規模システム
- OS, AlphaGO, ネット空間，Big Data，IoT関連システム，情報インフラなど

設計者にとって
大規模システムの全体を理解できてはいない
現在において，すでに透過性があるとはいえないシステムが多い
⇒OSにおける頻繁なアップデート，コンピュータウィルス．．．．．
制御可能性も怪しい状況
仮に個は制御可能性あっても群として機能するシステムの制御可能性は難しい
そもそも，
スケールすることでも透明性・制御可能性は難しくなる
スケールするとは人の認知能力を超えるということ

ユーザにとって
すでに透過性なし
制御可能性も100%あるとはいえない．
※意図とおり動作してくれればよい（受動的システムであることが必要）

トップダウンか？　ボトムアップか？





ほとんどの工学設計は
トップダウン型

大規模複雑システム構築
が困難

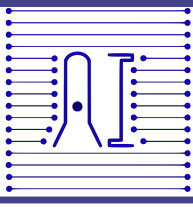
AI高度化には不向き

進化型，群知能型，脳型，深層学習

個の振る舞いと群の振る舞いには大きな開きがある.

こちらの手法の開拓が進む.

誰にとっての透過性・制御可能性？



人にとって？

そもそも人の認知能力の限界がある.

より便利なものの要求　⇒

認知能力を超えることが要求される.

心配だけと利便性が優先.　慣れ.

オートパイロット，自動運転.　.　.　.　.

MS Tayなど学習システムなどの制御可能性確保は難しい

システムを使う者次第（初等教育の重要性）

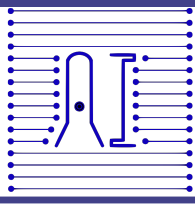
・学習機能に対する制御可能性の必要性

機械にとって？

機械に機械をチェックさせればよい！（AlphaGO）

ソフトウェア工学.　.　.　.　.

能動型システムへの対応が最大の壁

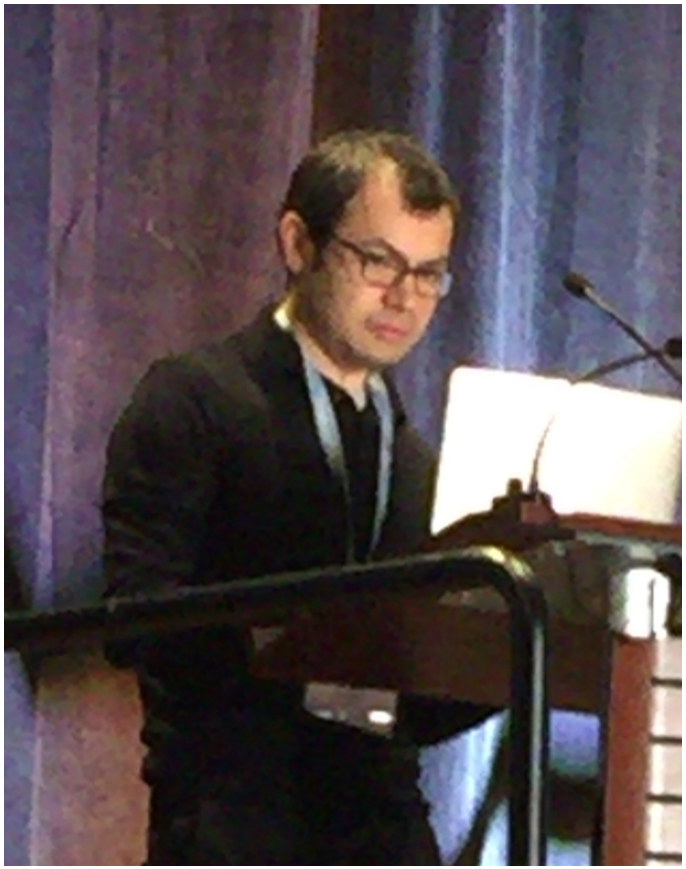


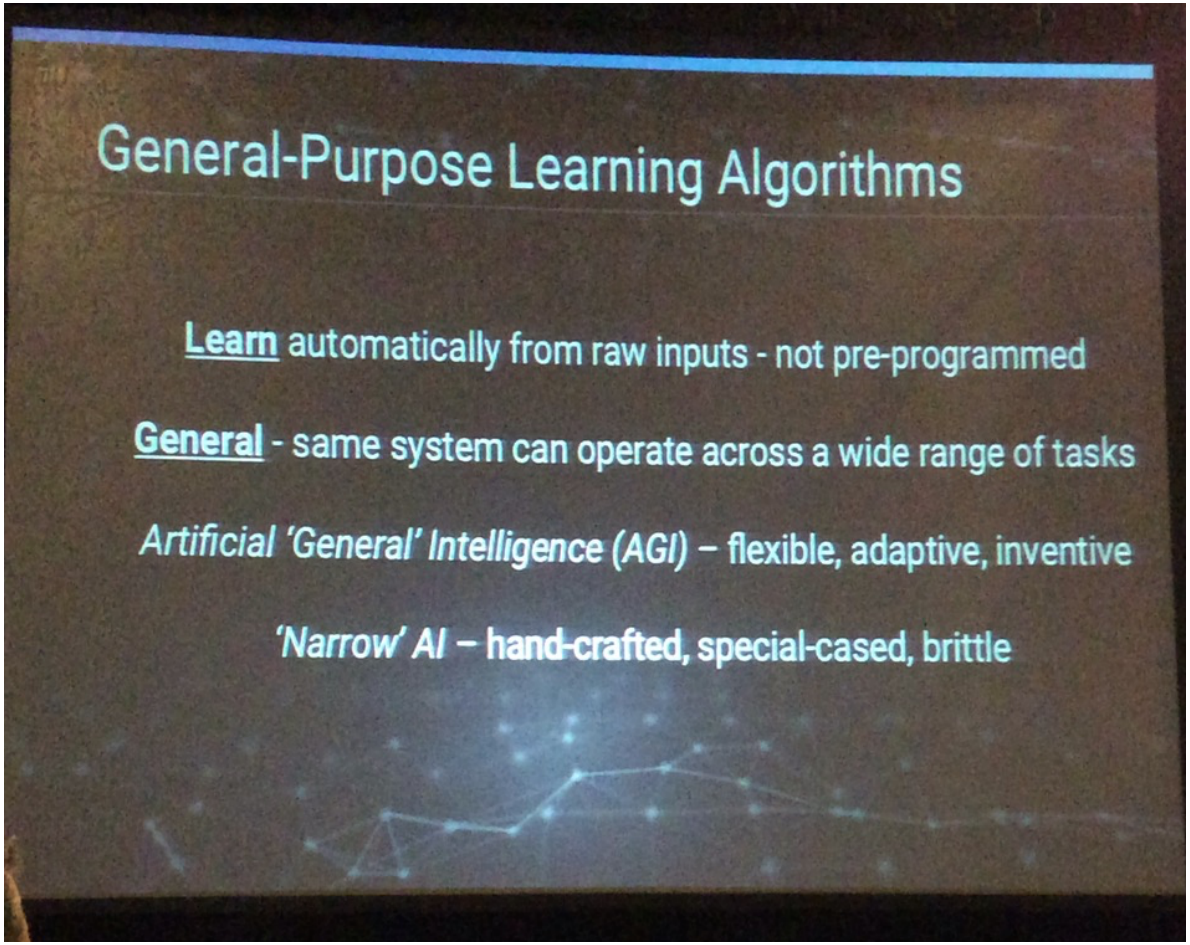
完全自律型システムの登場＝新生命体の登場

道具からの開放
システムが自ら目的を設定・成長する
透過性，制御性の議論は意味をなさなくなる
人にとっての究極の楽

ポイント！
奴隷ロボットからドラえもんへ！

AGIへ！！（追求をやめる気はなし）





方策

確実な透過性・制御可能性を前提での研究開発は極めて難しいであろう.

構成論的アプローチしかない

- 群知能型，ボトムアップ型システムに関する研究
- 箱庭での実験（AI牧場）※ネット環境では既に.
- スケールした実験可能な規模
- 特区や大規模シミュレーション環境
- 大規模仮想空間（人も投入）