Go Beyond the Cognitive Era.

-- Vol.01 未来の人工知能の品質保証と安全について --

Nobuhiro Hosokawa
(CARVIN@jp.ibm.com)

IBM Research Tokyo.
IBM Japan Ltd.

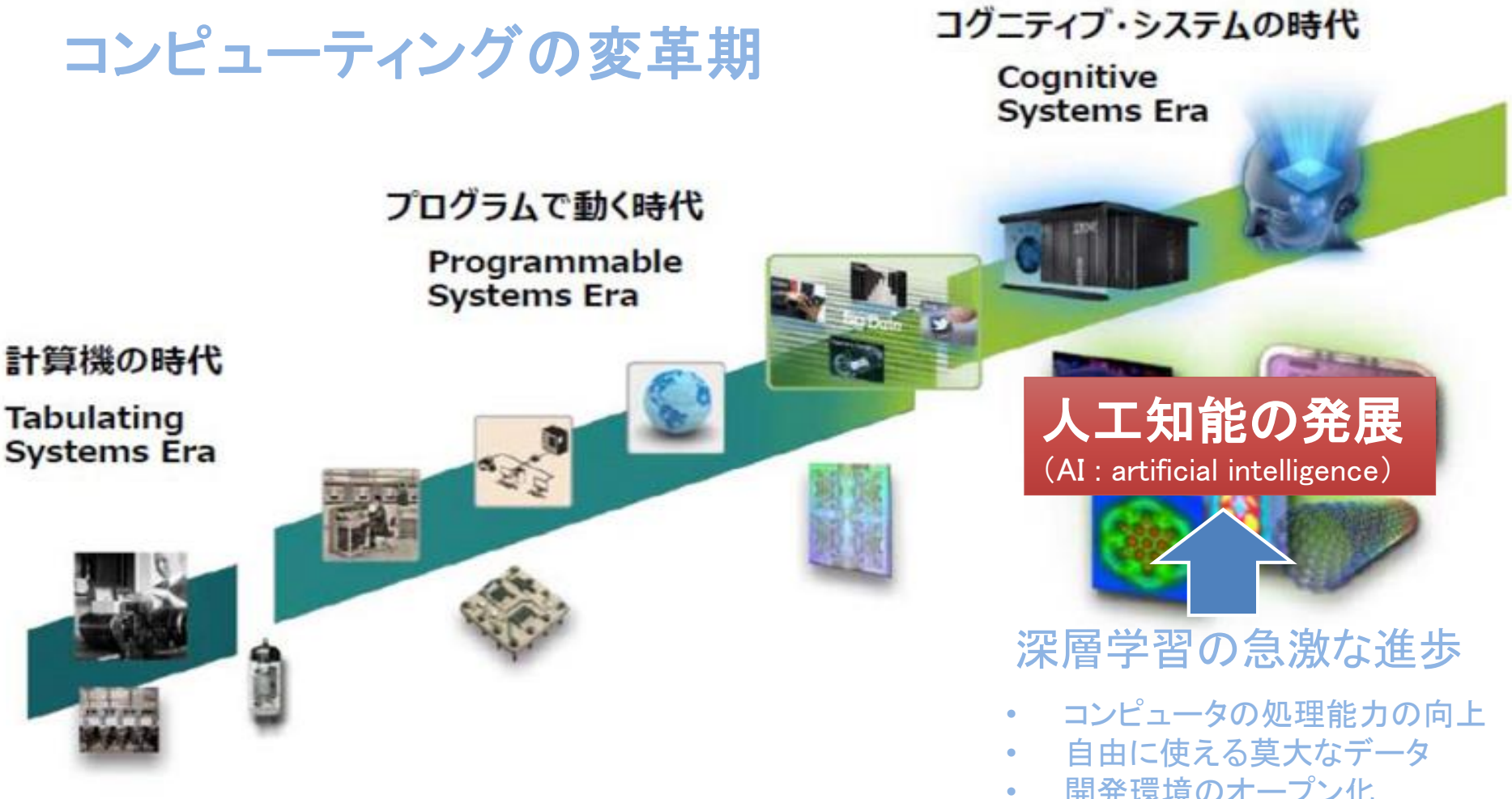
人工知能は今「どこまで」実現できているか？

The creation of a thousand forests lies in one acorn.
Una sola ghianda può creare mille foreste.
何千の森も、その創造の源はたった 1 個のどんぐりである。
成千森林的创造离不开一粒小小的橡籽。
Tausend Wälder entstehen aus einer Eichel.
수천 개의 숲을 만드는 것은 한 개의 도토리에 달려 있다.
La creación de miles de bosques radica en una sola bellota.
Uma bolota é responsável pela criação de mil florestas.
Un seul gland peut engendrer des milliers de forêts.
Et agern rummer kimen til at skabe tusinder af skove.

– Ralph Waldo Emerson

Chapter #01
人工知能の現在

コンピューティングの変革期



深層学習の急激な進歩

- コンピュータの処理能力の向上
- 自由に使える莫大なデータ
- 開発環境のオープン化

人工知能は今どこまで実現できているのか？

「人間ひとり相当のリアルタイムな深層学習を
ノイマン型コンピュータで実現するためには？」

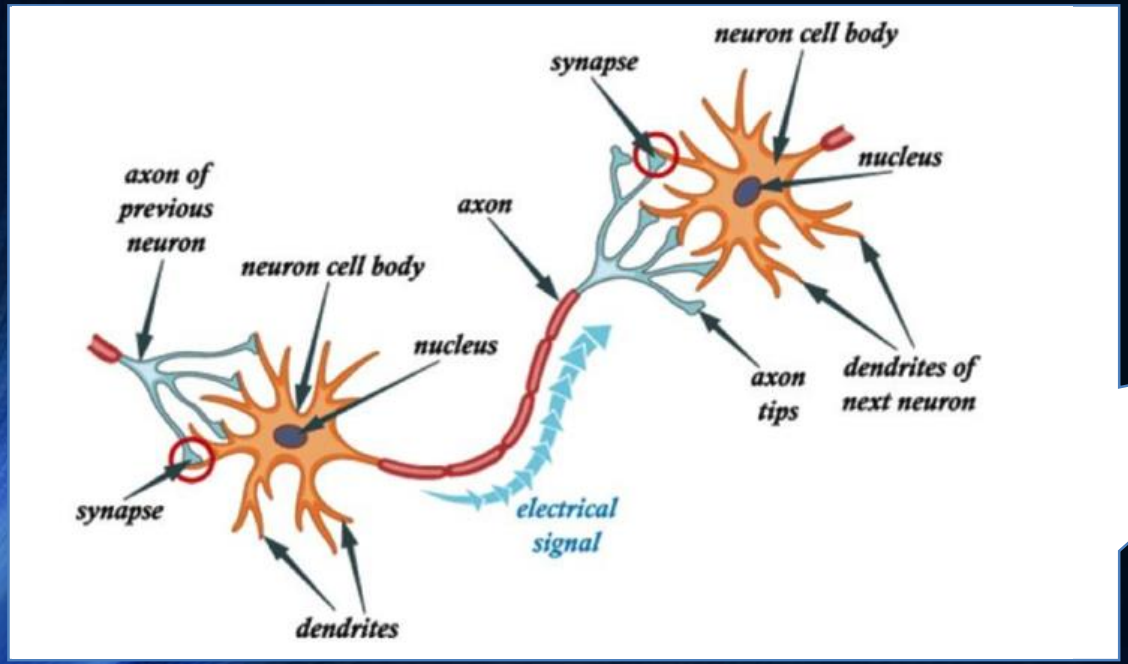
- 莫大な計算能力　6Exa FLOPS
  - ✓ 4億8千万個のプロセッサコア
  - ✓ 480ペタバイトのメモリー
  - ✓ 29,491,200のコンピュータノード
- 巨大な設備空間
  - ✓ 東京ドーム1.8個相当の設置面積
- 膨大な消費エネルギー
  - ✓ 2.4GWhの電力を消費
  - ✓ 120万キロワットの原子炉が2つ必要

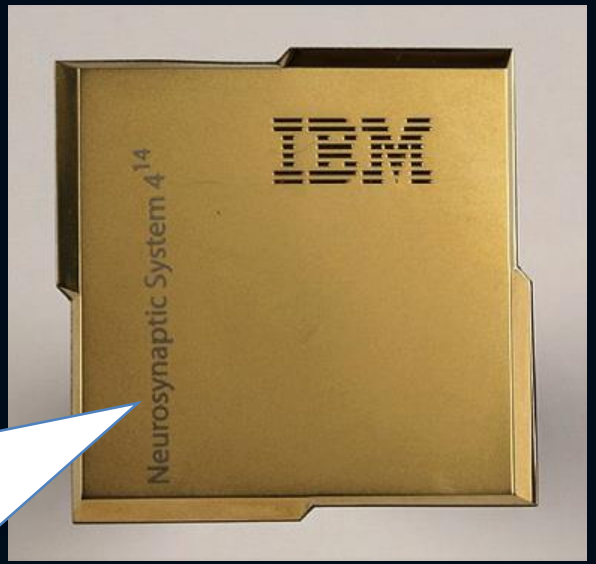
IBMが2012年に当時，世界最速のスーパーコンピュータを使って、ヒトの脳の5倍の規模（500×10億ニューロン・100×10京シナプス）のニューラルネットワークのシミュレーションを実施.

結果　：　計算時間が実時間の1,500倍遅かった

ニューロモーフィックデバイスとは？

- 生物の脳の仕組みを模した回路で構成する半導体
- 深層学習に必要なニューラルネットワークをハードウェアで実現
- 極めて高い動作効率が実現できる





ニューロモーフィック・デバイス の特徴

動作時の消費エネルギーが極めて少ない

クロック周波数に依存しないイベント・ドリブンの動作回路

Minimizing Active Power

高いリアルタイム性能

実時間での物体認識を実現

Real-Time Operation

高い欠陥耐性と信頼性

製造プロセスのバラツキやランダム欠陥に高い耐性
&
並列化による冗長性

Defect Tolerance

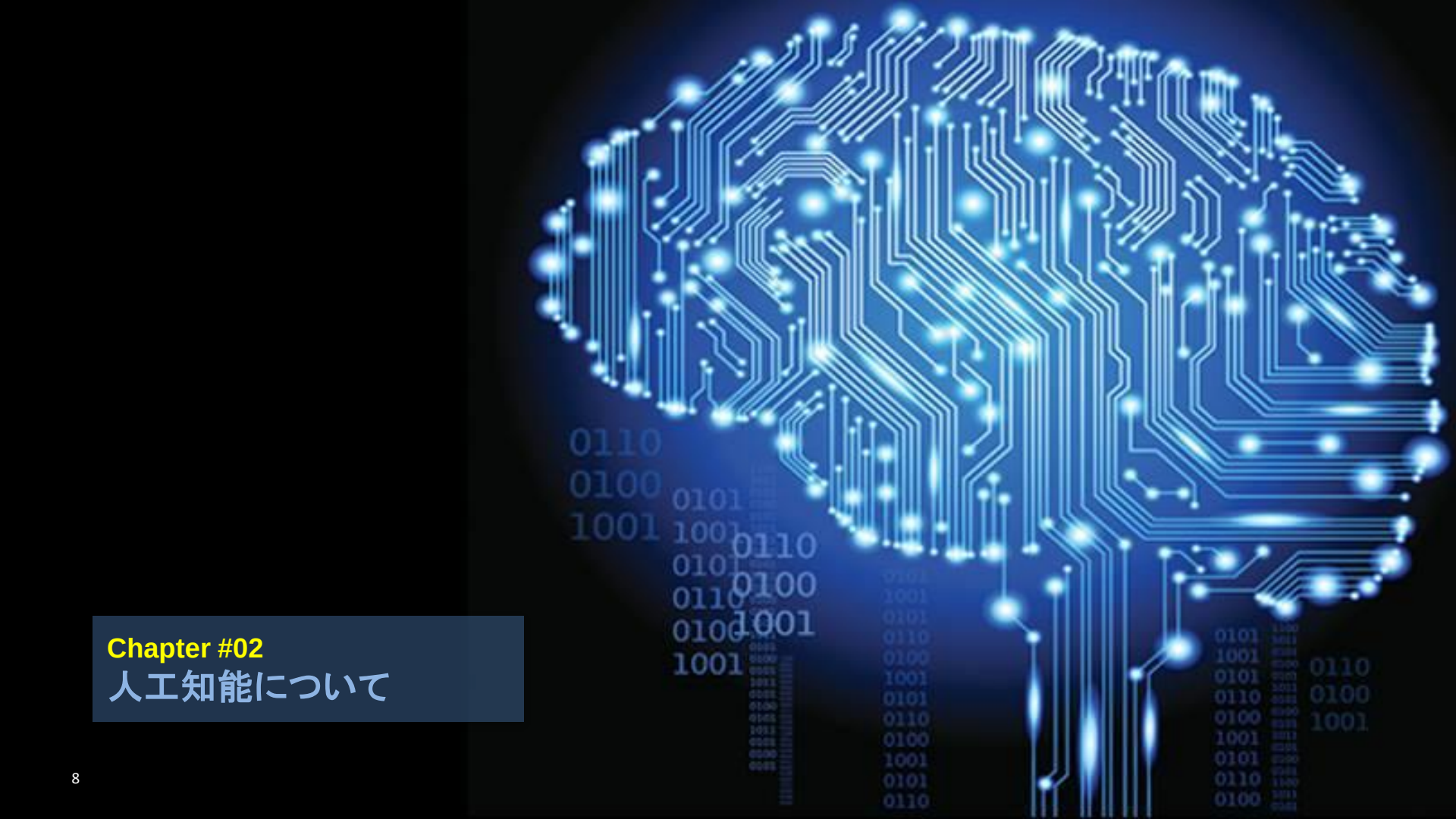
高いスケーラビリティ

1つのデバイスから数万個のデバイス連結まで対応

Scalability

ニューロモーフィック･デバイスの課題　～活用～

- ソフトウェア開発環境の充実
  - ✓ これまでの一般的なソフトウェア開発とは大きく異る開発環境
  - ✓ プログラミング言語 ： CoreletがMatlab上で動作
  - ✓ シミュレーター ： Compass
- 人材の育成
  - ✓ ニューラルネット，深層学習の深い知識
  - ✓ ニューロモーフィック・デバイスの知識
- 品質保証
  - ✓ 従来のシステム検証手法が適用できない
  - ✓ 実世界の状況に応じて挙動の正しさが変わってくる - 普遍性を求めることが難しい
  - ✓ 品質や信頼性の定義はまだこれから -検証シナリオの爆発的な増大



Chapter #02
人工知能について

人工知能の現在の限界点と「結果のテスト」

カラスとキツネの寓話

ズル賢いキツネが、自惚れ屋のカラスに歌を歌ってくれるよう頼んで、カラスがくわえていた食べ物を騙し取る、というお話があります。

「カラスさんはいつもいい声で歌うんだよね？聞かせてよ」

こう言われたカラスはくわえていたエサを地面において気持ちよく歌い始めます。

このお話を子供にするとします。この時、子供が正しくこの寓話を理解したかを確認するためには次の質問をするといいでしょう。

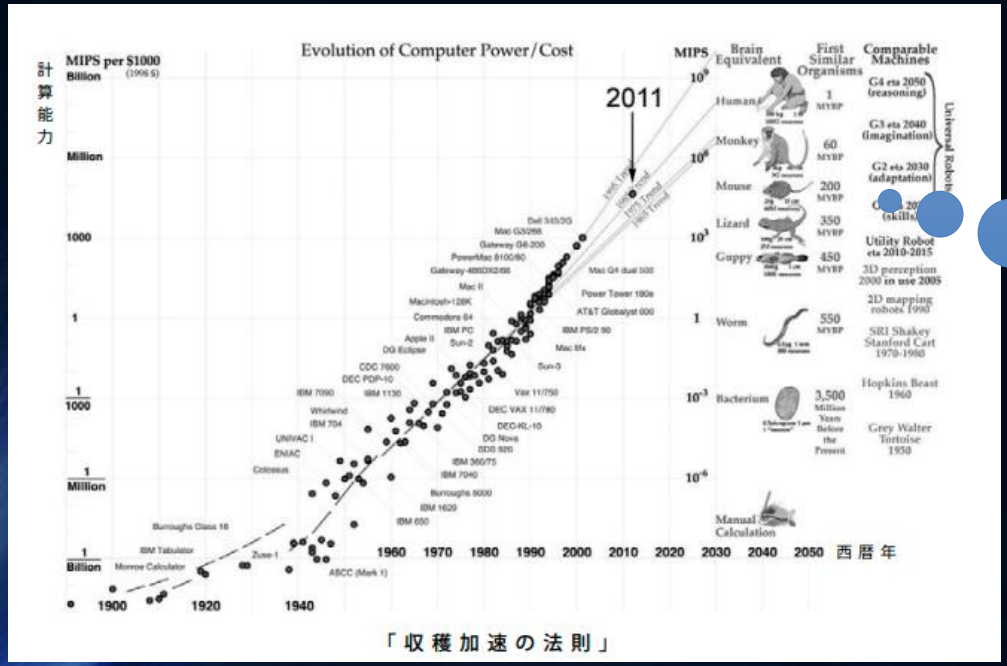
「キツネは、カラスが素敵な歌声をしていると思っていましたか？」

出典：コンピュータには何ができないか―哲学的人工知能批判, ヒューバート・L. ドレイファス (著), Hubert L. Dreyfus (原著), 黒崎 政男 (翻訳), 村若 修 (翻訳)
産業図書 (1992/04) ISBN: 978-4782800690

シンギュラリティ（Singularity）という言葉について

シンギュラリティー＝技術的特異点とは？

人類の技術開発の歴史から推測して得られる未来のモデルの正確かつ信頼できる限界（「事象の地平面」）を指す。
「強い人工知能」や人間の知能増幅が可能となったときが技術的特異点になると考えられている。
特異点の後では科学技術の進歩を支配するのは人類ではなく強い人工知能やポストヒューマンとなり、従って人類の過去の傾向に基づいた変化の予測モデルは通用しなくなると考えている。

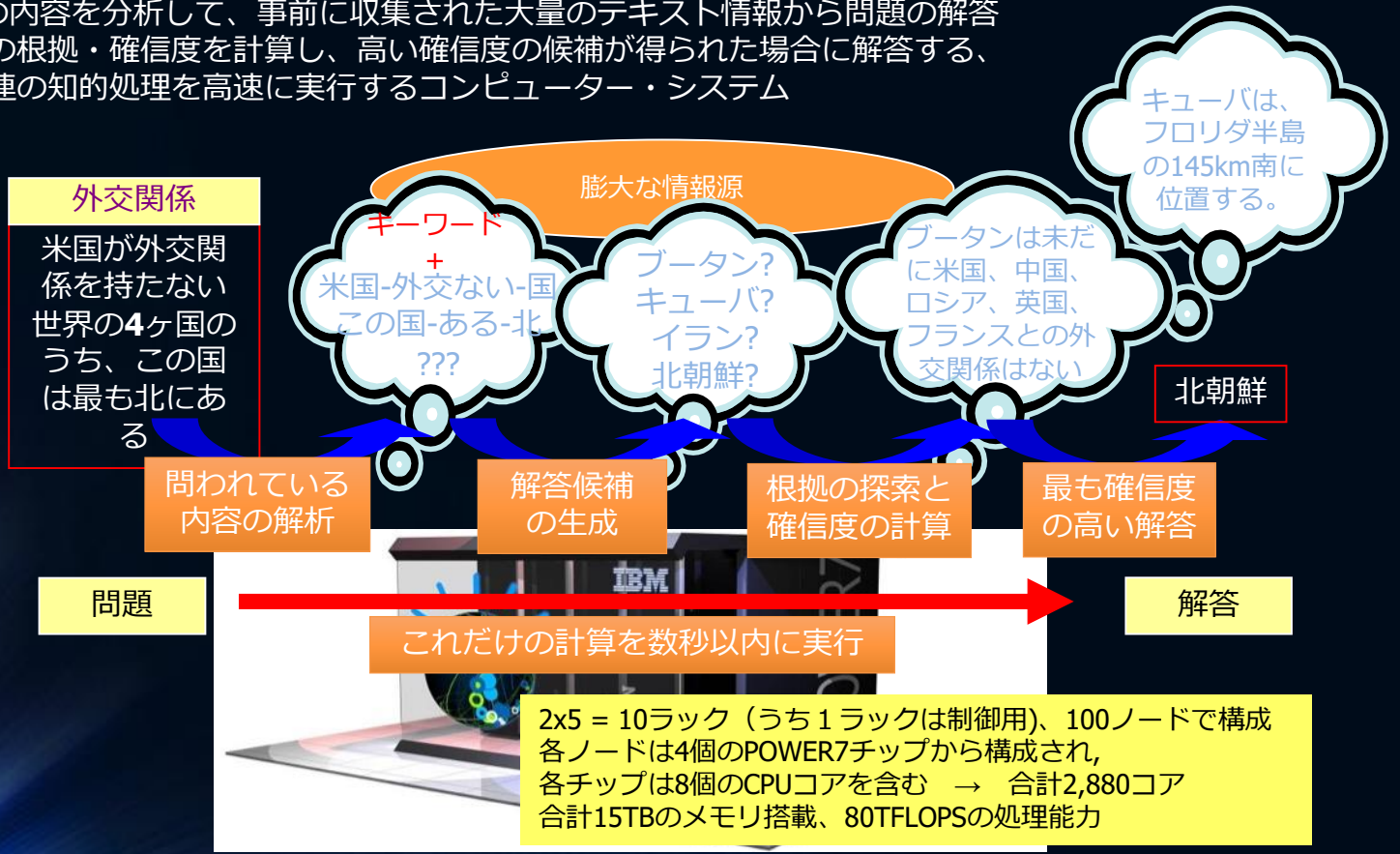


人工知能が全人類の処理能力を超える点は2045年に到来する

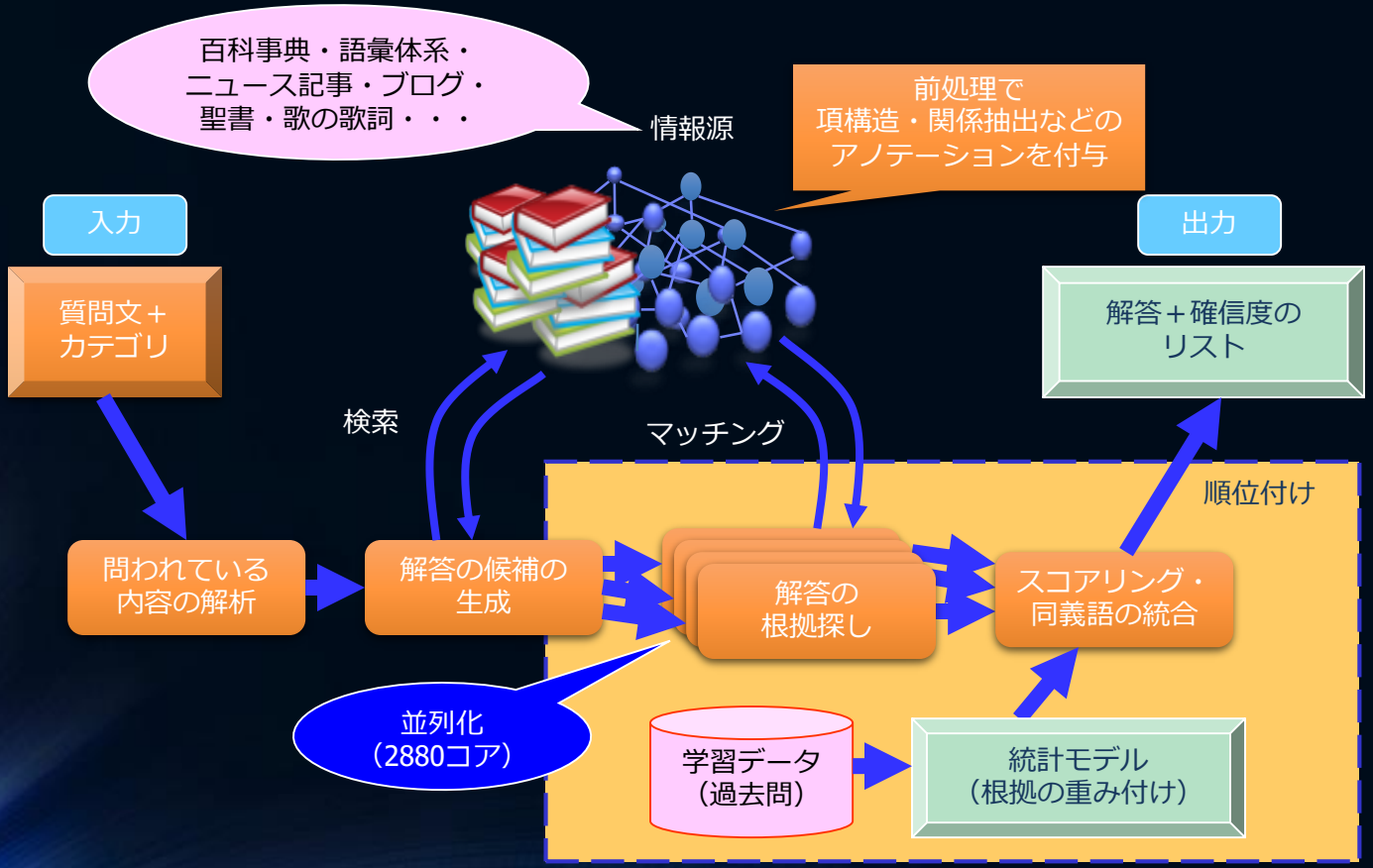
出典：http://www.godandscience.org/doctrine/singularity_movement.html

質問応答システムWatsonとは？

問題(文)の内容を分析して、事前に収集された大量のテキスト情報から問題の解答候補とその根拠・確信度を計算し、高い確信度の候補が得られた場合に解答する、という一連の知的処理を高速に実行するコンピューター・システム



正答率を高めるためのチューニング



Chapter #03
人工知能の功罪

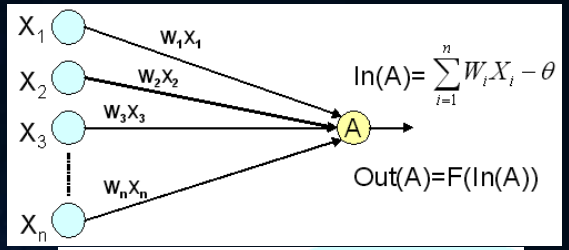
人工知能の二つの「恐怖」？

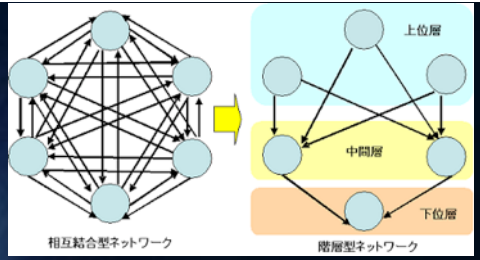
- １）過度に依存する危険性
  - 中身はニューラルネットワーク＝ブラックボックス。人工知能が「どう考えたか」を可逆遡及して検証することは難しい
  - 思考の正当性が証明しにくい＝暴走を止める事・予知すること・暴走後に原因を追求することができない

- ２）労働・職を奪われる不安
  - 労働シフト（例：農業→工業の産業革命等）と同じように単純労働からの解放を目的とする意見
  - 単純労働者が職を奪われる可能性は十分にある
  - 単純労働＝知能の代替という意見。
  - スーパーのレジは「無人POSレジ」に代替されている

人工知能の危機感： そもそも正しさの証明なんてできない？

ニューラルネットワーク(神経回路網、英: neural network, NN)は、脳機能に見られるいくつかの特性を計算機上のシミュレーションによって表現することを目指した数学モデルである。
研究の源流は生体の脳のモデル化であるが、神経科学の知見の改定などにより次第に脳モデルとは乖離が著しくなり、生物学や神経科学との区別のため、人工ニューラルネットワーク(人工神経回路網、英: artificial neural network, ANN)とも呼ばれる。





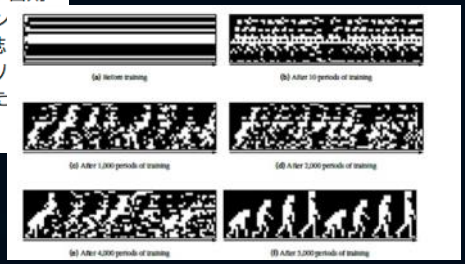

ニュース 日経コンピュータ

IBM東京基礎研、より生物の神経回路に近い人工ニューラルネット「DyBM」を提案

2015/09/17
浅川 直輝＝日経コンピュータ （筆者執筆記事一覧）

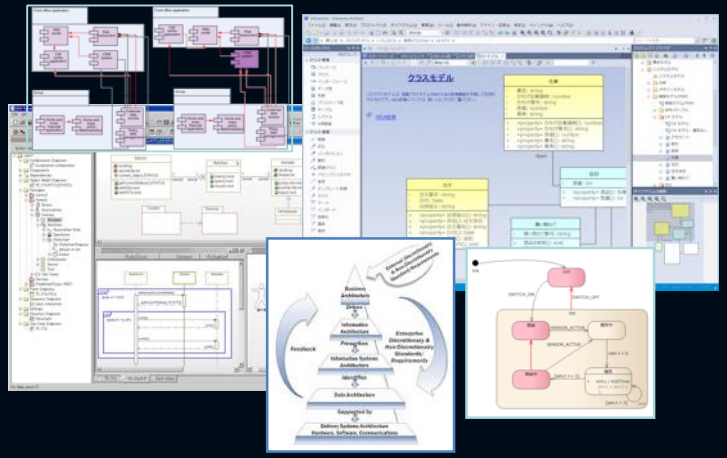
233 40 46 ツイート 保存する 記事一覧へ >>
おすすめ ブックマーク Pocket シェア

IBM東京基礎研究所は、従来よりも生物の神経回路に近い学習則を備えた人工ニューラルネットワーク「動的ボルツマンマシン（DyBM）」を考案し、英ネイチャー系列のオンライン科学誌「Scientific Reports」で公表した。時系列に並んだデータのパンを学習、再現できる特長があり、音楽や映像、言語といった列データを認識する人工知能に応用できる可能性がある。



問題はニューラルネットワークの可塑性の「検証」ができないこと。
（＝人工知能がどう判断してどう結論づけたかが後から追求できない）

人工知能の危機感と不安：ITの世界だって「オマエイラネ」現象



プログラミングが「プログラム合成」中心に。
作るから使うにシフトする？

設計はモデルリファレンス主体に？
設計作業の負担が軽減する？

プログラム＝自動合成、
仕様書＝モデルリファレンス
テスト自動化、自動テストケース生成
今や深く考えられるか？ではなく「知っているか？」の時代

顧客の望む通りのものをさっさと
納品しておわり。
「ユーザーが言ったから」、「xxがやれと言ったから」 仕事を早く終わらせる「作業」になってないか？

人工知能の現在の限界点: 人工知能でないものを人工知能として販売

Siriの「割り勘」機能

・iPhoneでSiriを起動

・「割り勘」と音声で入力

・お勘定を聞かれるので金額を音声で入力

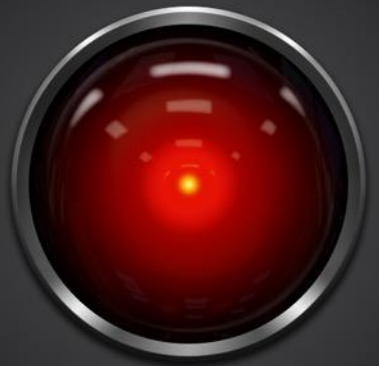
・人数を聞かれるので人数を音声で入力

・一人あたりの支払い金額が出力される

これは人工知能か？

出典：http://news.mynavi.jp/articles/2015/11/13/ipadiphonehacks/

Chapter#4 テストできるものならやってみろ



人工知能にできること:「識別」「予測」「実行」

- HBR記事の安宅氏によると機械学習をベースにしたAIの利用には主に以下の三つに分けられる。

(1)識別

- 情報の判別・仕分け・検索(言語、画像ほか)
- 音声、画像、動画の意味理解
- 異常検知・予知

(2)予測

- 数値予測
- ニーズ・意図予測
- マッチング

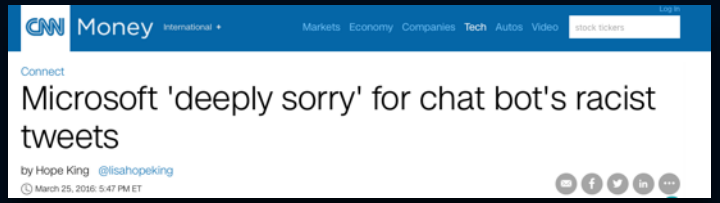
(3)実行

- 表現生成
- デザイン
- 行動の最適化
- 作業の自動化

http://blog.btrax.com/jp/2015/11/23/ai-02/

事例1）機械学習したツイートBotの「差別発言」

- Microsoft ツイートBotの”Tay”



  - 2016年3月: マイクロソフトはオンライン・ツイッターロボットである Tay（発音は”テイ”）の運用を開始しました。彼女は”モデル化され、学習済＋フィルター済）の パブリックデータを元にユーザーと個別にお話・会話できるように設計されていると発表公開されました。
  - しかし、実際はオンラインを通じて彼女＝Tayに悪意のある差別的発言を教え込むことで、実際のリアクションとして対話的な発言をするようにTay自身を学習させてしまいました。
  - 現段階でこのような悪い学習を行わせない、学習データとして入力を受け付けない方法を検討しているそうですが、今後機械学習アルゴリズムにより意思決定など、この分野以外の適用分野（例：交通、金融、物流、ヘルスケアの分野）など様々な分野で利用されていくのです。このときに重要な点は「システム機能の保証」のみならず、意思決定の品質精度は最終的にはデータの「インテグリティ」に依存してしまいます。
  - 「敵意ある機械学習（Adversarial Machine Learning）」という新しい分野は研究分野として確立しておらず、新しい領域です。特に公開された参考文献などがほとんどありません。
  - データインテグリティ（各種の改ざん、偏向データ、不十分なデータでない事をどう保証しますか？

- どうやって機械学習プロセスの品質を保証しますか？
- どうやってデータの品質をテストしますか？
- 最終的に品質を保証する方法はありますか？

事例2）テスラの自動車事故

- 逆光で空の色に溶けたトレーラーの色を従来の画像認識アルゴリズムでは識別できなかった事例。

JULY 1 TSLA: 216.50 4.22

Understanding the fatal Tesla accident on Autopilot and the NHTSA probe

Fred Lambert - 6 days ago @FredericLambert CARS TESLA

The fact that the Autopilot system didn't detect the trailer as an obstacle prompting emergency braking or steering is what is worrying a lot of people. The forward facing sensors of the Autopilot consist of a camera, a radar and a few ultrasonic sensors.

It's understandable that the camera couldn't detect the trailer as an obstacle based on Tesla's explanation of the trailer's "white color against a brightly lit sky" and the "high ride height", but what is less understandable is why the front facing radar didn't detect it.

Tesla CEO Elon Musk offered an explanation:

> *@artem_zin @theaweary Radar tunes out what looks like an overhead road sign to avoid false braking events*
>
> *— Elon Musk (@elonmusk) June 30, 2016*

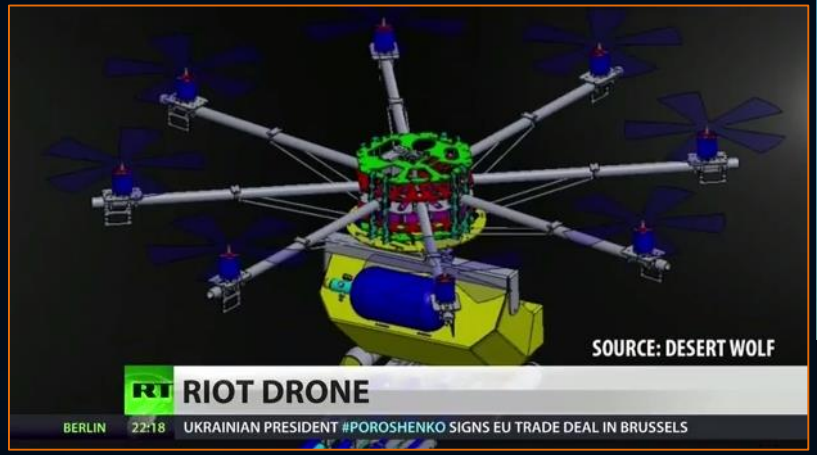
Our understanding here is that the high ride height of the trailer confused the radar into thinking it is an overhead road sign. It's obviously not ideal and the system should be refined to have a greater

source: http://electrek.co/2016/07/01/understanding-fatal-tesla-accident-autopilot-nhtsa-probe/

- どうやって機械学習プロセスの品質を保証しますか？
- どうやって修正後のアルゴリズムの「正しさ」をテストしますか？
- 「空の色で誤認しない」テストが実施できるか？

事例3)ドローン兵器

- 爆撃機の一部として
- 体当り(カミカゼ機)として



2016.10.17の記事

ISISが市販ドローンを爆撃機として使用、初の死亡者が確認される

source: http://www.gizmodo.jp/2016/10/isis-is-using-hobby-drones-to-kill-people.html /

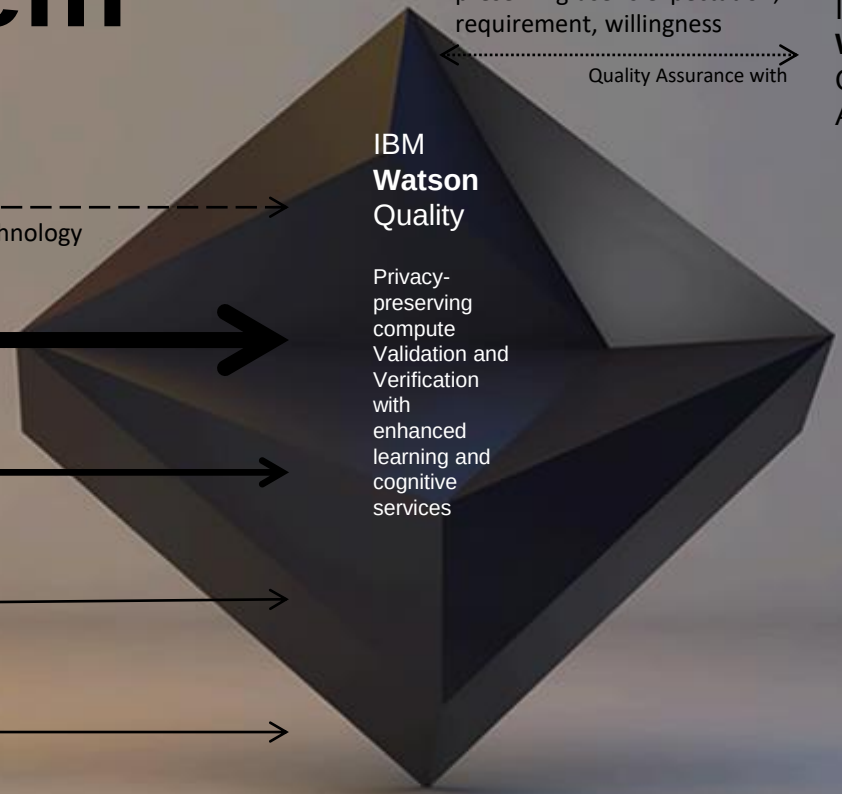
- 殺傷能力をどのようにテストしますか？
- 安全に自軍から射出できることをどうやって保証しますか？

Assured System
for Cognitive System

IBM **Assured System**

Embedded learning and cognitive services in each Quality aspect

- **Functional Quality** — with Classical Assurance Technology
- **Data Quality**
- **Non-Functional Quality**
- **Legal & Ethical Matters / Willingness**
- **Safety**
- **Reliability Trustiness**



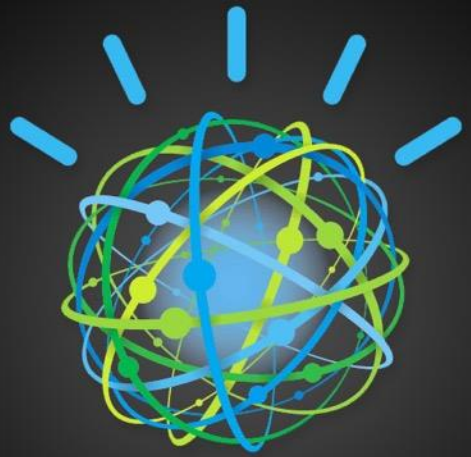
IBM **Watson** Quality

Privacy-preserving compute Validation and Verification with enhanced learning and cognitive services

Processed knowledge, preserving user's expectation, requirement, willingness

Quality Assurance with

IBM **Watson** Quality Analytics



“To complete me with your hands, please send out into the real world, PLEASE… ”
（早くあなたの手で私をテストし完成させて、現実世界に送り出してください…）