人工知能研究の今までとこれから

丸山宏
Preferred Networks
@maruyama on twitter/X

自己紹介:

- 1975 高校のミニコンで初めてのプログラミング
- 1977 東京工業大学理学部情報科学科へ進学
  - 当時情報科学を専門に教える大学は少なかった
  - 修士課程:米澤明憲先生の指導で、自然言語理解システムの研究
- 1983-2009 日本IBM東京基礎研究所
  - 自然言語処理、機械翻訳、手書き文字認識
    XML、Webサービス、セキュリティなど

いわゆるコンピュータサイエンス

- 2011-2016 統計数理研究所
  - 統計モデリング、ビッグデータ解析、レジリエンス
- 2016年4月～ 株式会社Preferred Networks
  - 深層学習
- 現在:東大 上席研究員／PFNシニアアドバイザ

統計
データサイエンス
機械学習

オペレーションズ・リサーチ -- 大西洋の戦い 1942-1943

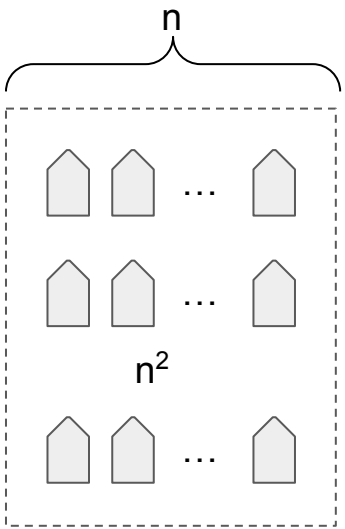
1942/10-12: コンボイ SC107
- 商船37隻＋護衛6隻 vs Uボート16隻
- **商船15隻喪失**、Uボート2隻撃沈

“被害を最小にするコンボイの構成は何か？”
与えられた知識:
- 各商船の出発地・目的地・積荷・速度
- 各護衛艦の兵力・速力・他の作戦目的
- **Uボートの攻撃成功率に関する数理モデル**

最適化

1943/5: コンボイ SC130
- 商船37隻＋護衛9+4隻 vs Uボート25隻
- **商船喪失なし**、Uボート3隻撃沈

仮説:Uボートの攻撃成功率に関する**数理モデル**



Uボートの攻撃成功率は、コンボイ周囲長における護衛艦密度によって決まる

オペレーションズ・リサーチ生みの親
物理学者 ブラケット男爵
出典:Wikipedia

人工知能とデータサイエンス

人工知能

- **研究領域** – (人間の)知性とはなにか、をコンピュータ上に実装することで明らかにする(計算機科学)

データサイエンス

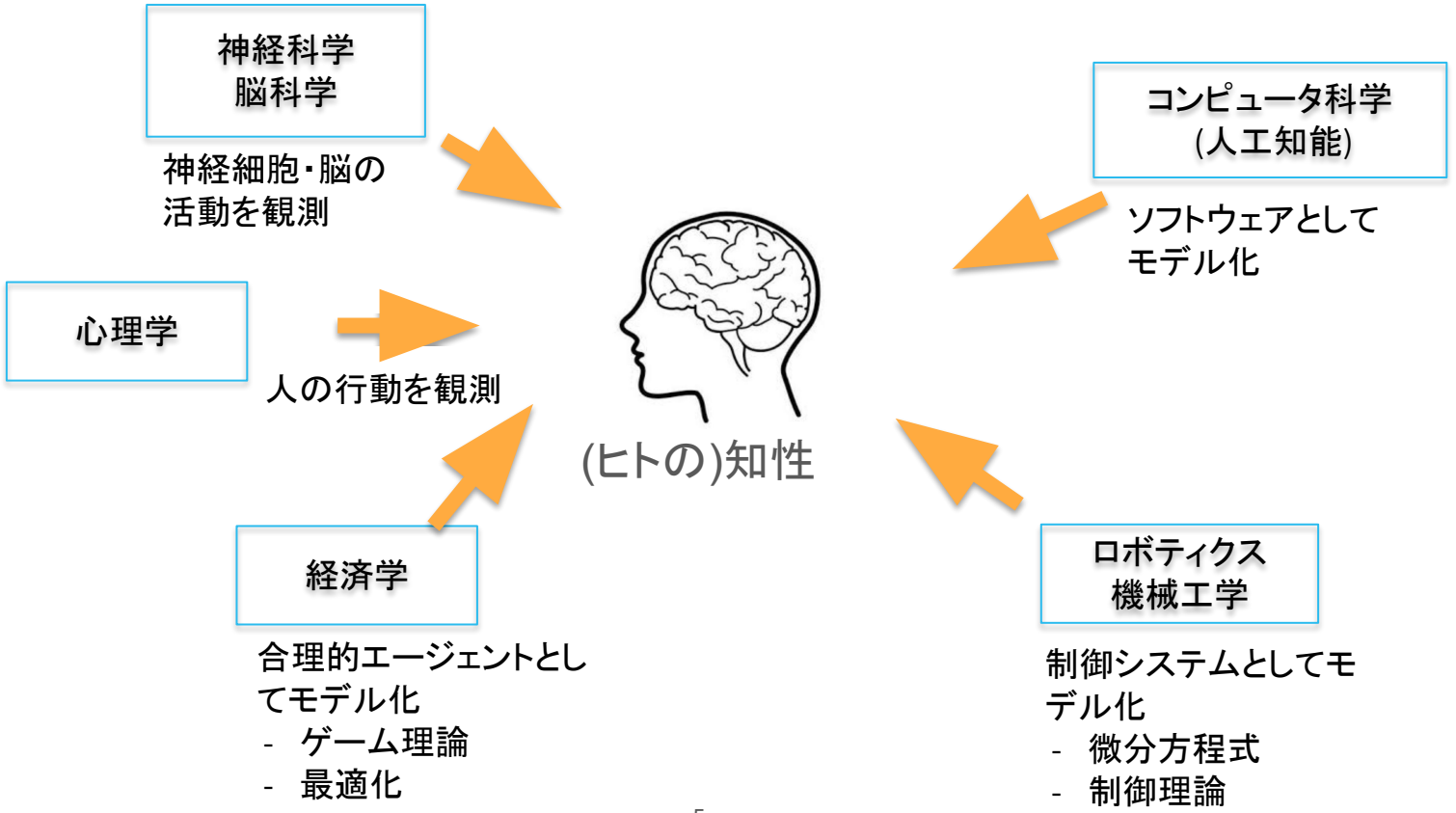
- **意思決定の方法論** –「勘と度胸」ではなく、観測されたデータとそれから得られる統計モデルに基づくもの(経営学)

人工知能の最近の流行はデータで訓練する機械学習システム

≒

データサイエンス

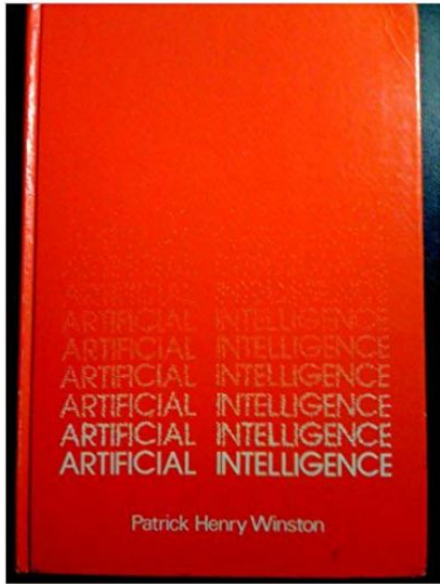
人工知能とは：（ヒトの）知性へのアプローチ



アジェンダ

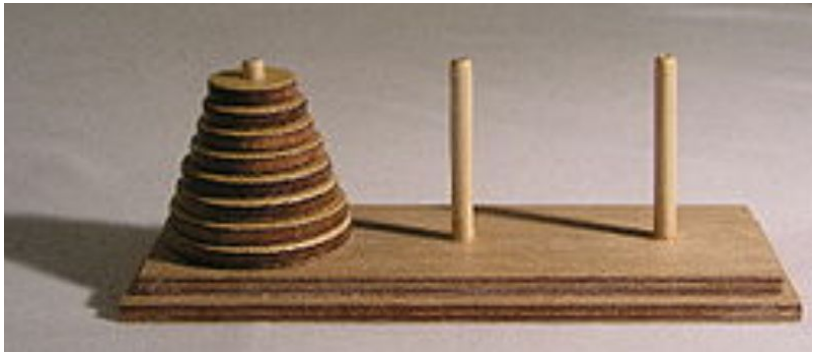
1. 人工知能研究の今まで
    - **最初の 50年 – 賢い人の思考を対象とした人工知能**
    - 次の15年 - 普通の人の思考を対象とした人工知能
2. 人工知能技術の課題とガバナンス
    - 生成モデルの限界
    - 今のAIは「計算機科学の総合格闘技」
    - 人工知能の危険
3. 人工知能研究のこれから
    - 人にできない思考を対象とした人工知能
    - 人間との棲み分け - 超知性と共存する社会へ

賢い人の思考(1) -- パズルを解く: ハノイの塔



P. H. Winston, *Artificial Intelligence*
ISBN-13: 978-0201084542, 1977

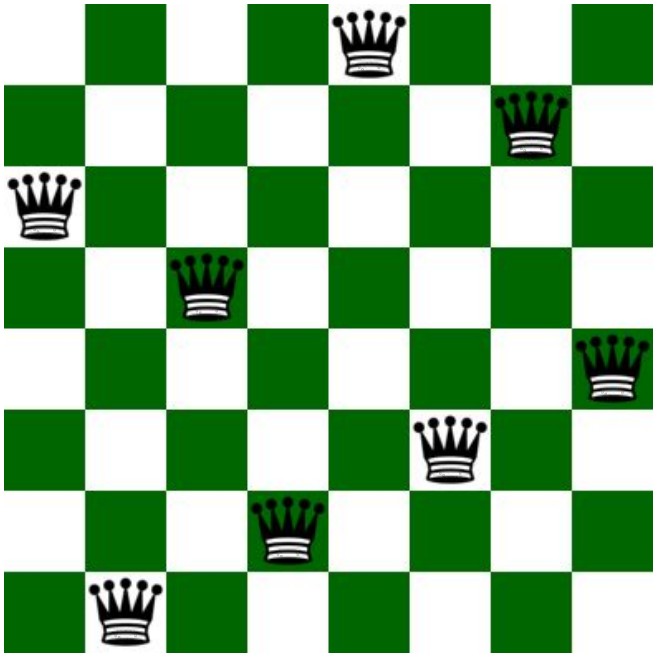
- すべてのディスクを右側の軸に移す
- 1度に1枚しか動かせない
- より小さなディスクの上に乗せてはいけない



Source: Wikipedia

- 帰納法を知っていれば簡単
- n-1枚のディスクを動かし方を知っていれば、n枚のディスクを動かせる

賢い人の思考(2) -- パズルを解く: 8クィーン



チェス盤の上に、8つのクィーンをお互いに
取られない位置に置くにはどうしたらよいか

試行錯誤（＝探索）によって解く

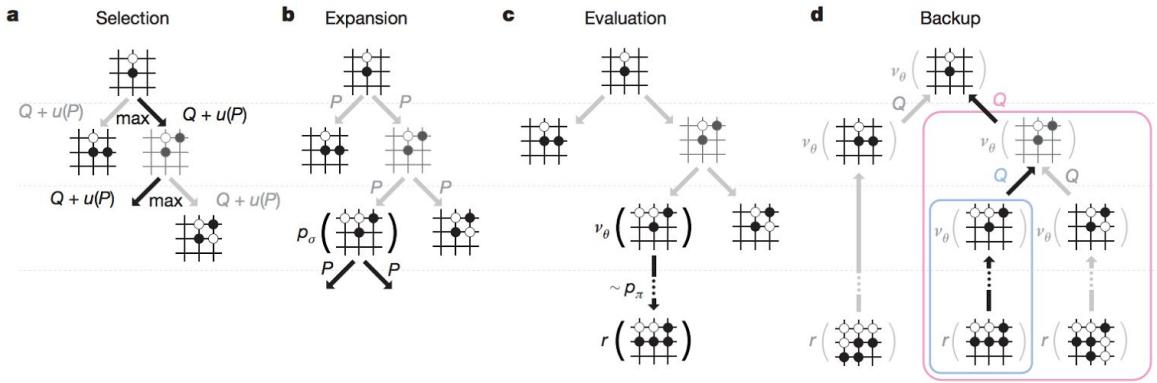
クィーンの数nが大きくなると、探索が難しい

出典:
https://www.geeksforgeeks.org/n-queen-problem-local-search-using-hill-climbing-with-random-neighbour/

賢い人の思考(3) -- ゲームに勝つ

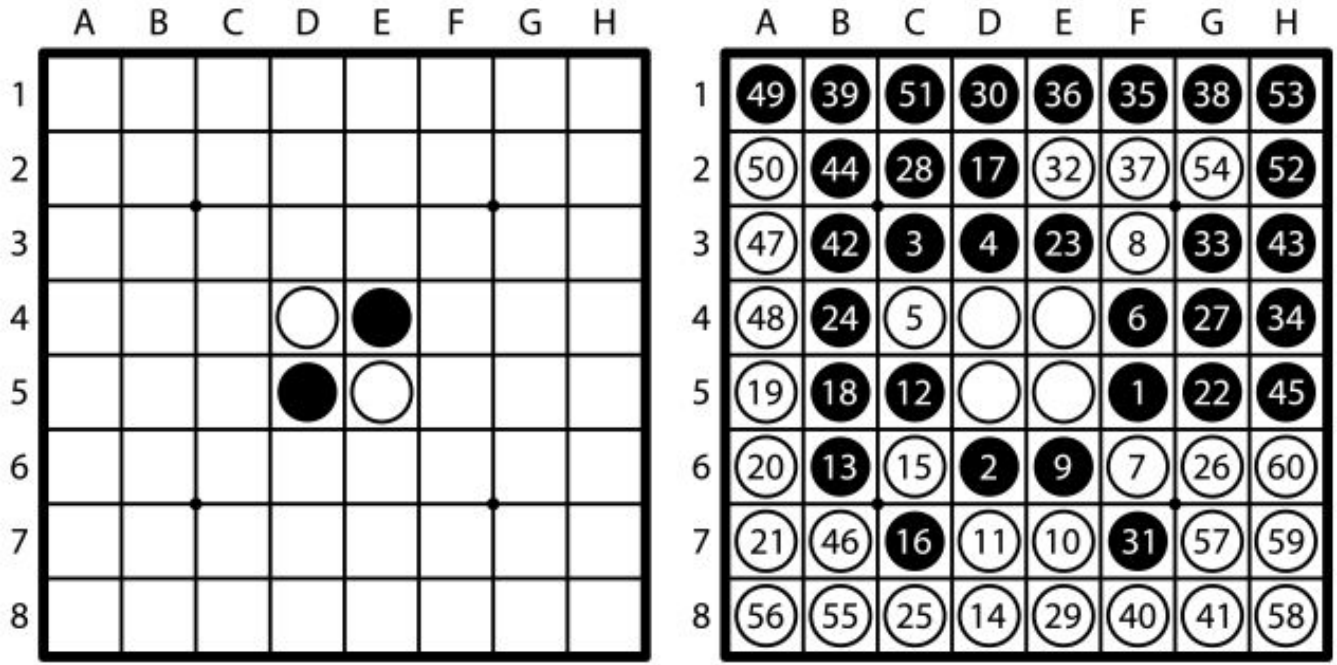
- 三目並べ -- A S Douglas, 1952
- チェッカー -- Chinock , 1992 - 1994
- オセロ -- Logistello, 1992 - 1997
- チェス -- Deep Blue, 1997
- 将棋 -- Bonanza, あから2010など、2007 - 2017
- 囲碁 -- AlphaGo, 2016

ゲーム木の探索



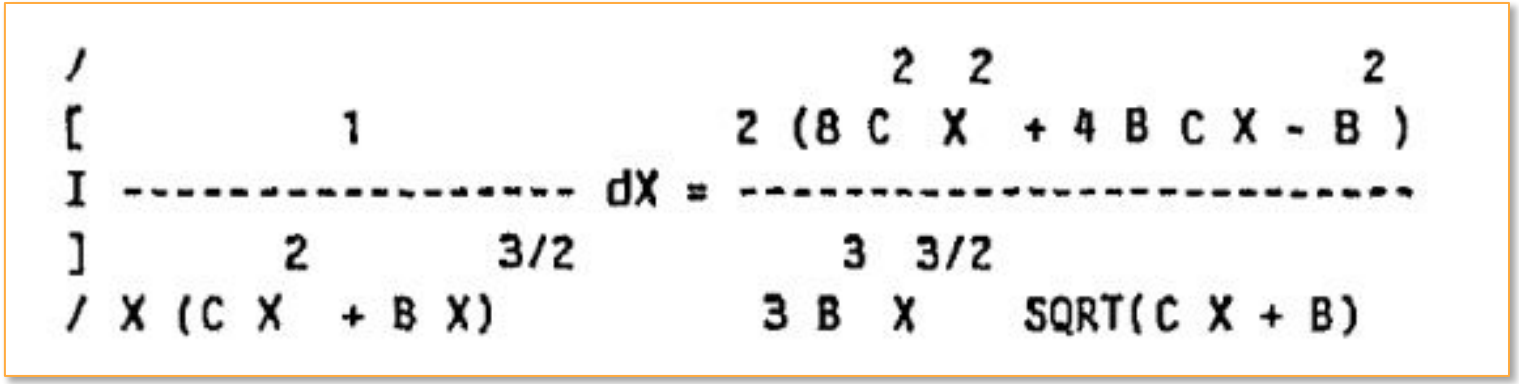
Silver, David, et al. "Mastering the game of Go with deep neural networks and tree search." *nature* 529.7587 (2016): 484-489.

AIの未解決問題、オセロが解けた!（互いに最善手を打てば引き分け）



Takizawa, Hiroki. "Othello is Solved." *arXiv preprint arXiv:2310.19387* (2023).

賢い人の思考(4) --数式を処理する:
MACSYMA, 1968-1982



Richard Pavelle and Paul S. Wang. 1985. MACSYMA from F to G. J. Symb. Comput. 1, 1 (March 1985), 69-100.

意思決定に必要なデータ分析の3局面*

1. 説明的 (Descriptive) データ分析
   - 現状がわかれば意思決定できる場合
   - 手法：可視化、データマイニング…
     必要な資質： 想像力
2. 予測的 (Predictive) データ分析
   - 未来(不明の情報)が推定できれば意思決定できる場合
   - 手法：モデリング、機械学習、…
     注意： 機械学習には訓練データが必要
3. 指示的 (Prescriptive) データ分析
   - 複数の打ち手の中から最善のものを選ぶ必要がある場合
   - 手法：最適化、探索
     前提： モデル＋KPIがよくわかっている

必要な物資・場所・日程
稼働可能な船舶・護衛艦
敵の戦力…

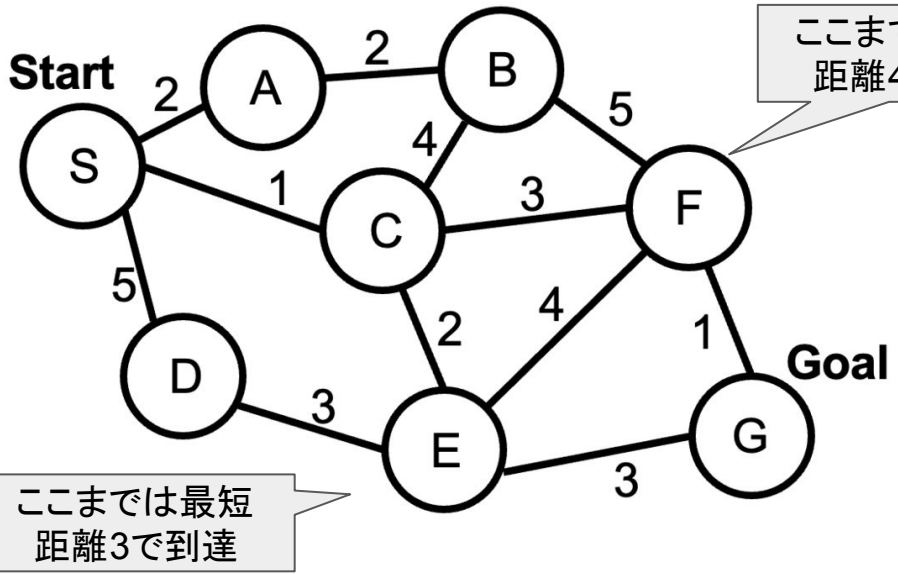
敵の位置、天候、想定される被害…

船団の計画、護衛艦の割当て、…

* James R. Evans, “Business Analytics: The Next Frontier for Decision Sciences,” *Decision Science*, Vol. 43, No. 2, 2012.

知識(モデル)の与え方(1) - 最短経路問題

SからGへ行く経路のうち、最短のものを求めよ



ダイクストラのアルゴリズム
部分問題の解をうまく使う
動的計画法（DP）

知識(モデル)の与え方(2) - 線形計画法

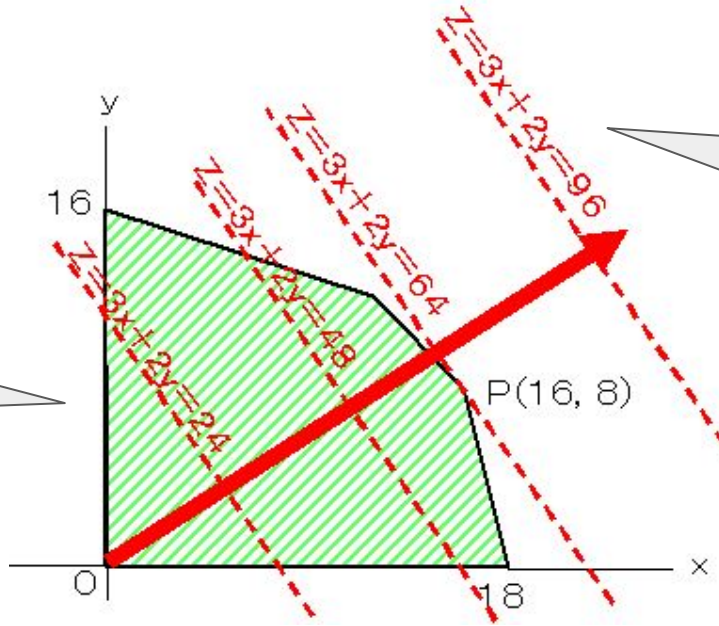
“緑色の (x, y) 空間の中で、3x + 2y を最大にする x, y を求めよ”

目的関数が線形式で表される。

Uボートの攻撃成功確率に関する仮説

実行可能空間が線形不等式で表される。

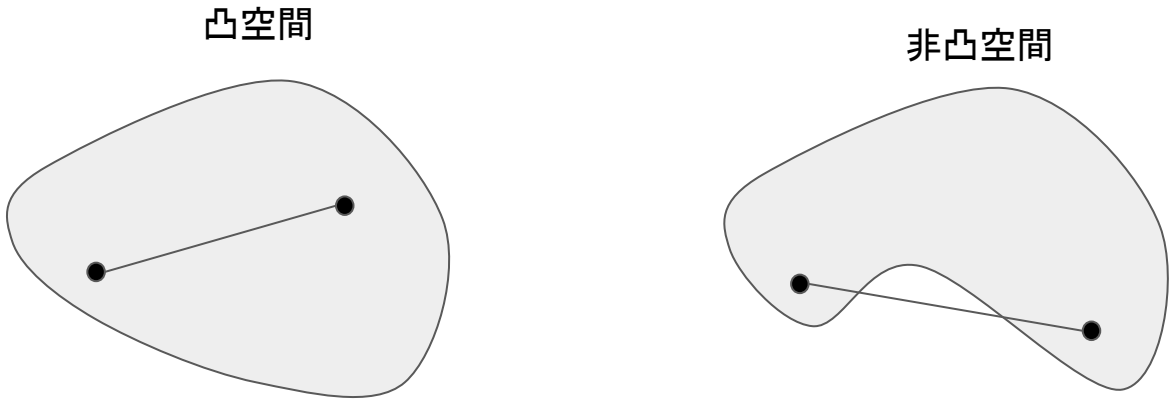
各商船・護衛艦の出発地・目的地・積荷・速度 …



出典: 小暮仁 Web教材「OR（オペレーションズ・リサーチ）」

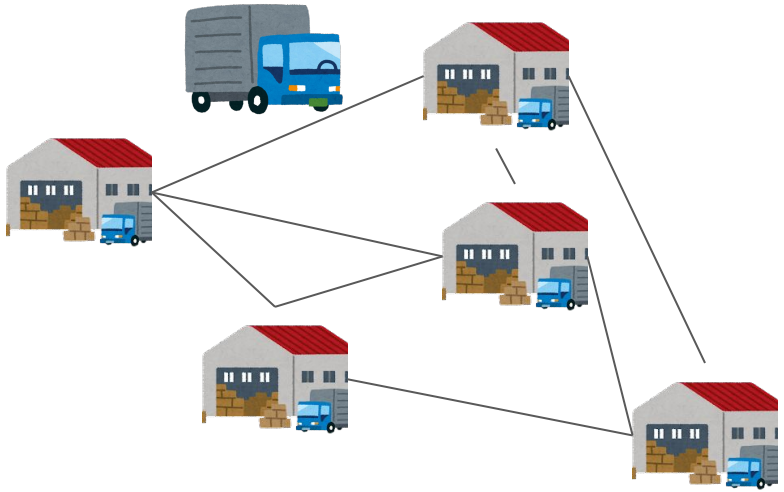
知識(モデル)の与え方(3) - 内点法

- 任意の凸空間における凸関数の最適化は内点法で効率よく解ける



知識(モデル)の与え方(4) - 離散最適化問題

- 定義域が離散の制約充足問題
  - 3充足問題など
- 定義域が離散の最適化問題
  - 巡回セールスマン問題
  - ナップザック問題



これらはいずれも**NP完全問題**として知られ、最悪の場合は2^n の時間がかかる

現在では経験的に効率のよいアルゴリズムがあり、ほとんどの現実問題は効率よく解ける

モデル記述の方略

1. 数理モデルで記述
   - 線形、微分方程式、確率分布、…
2. プログラム(ルール)で記述
   - 探索問題、ゲーム、シミュレータ、…
3. 有限のサンプルデータで近似
   - 統計的機械学習、生成モデル
4. 現実世界で試行錯誤(「経験の時代」)
   - A/Bテスティング、創薬、治験、…

「経験の時代へようこそ」

Welcome to the Era of Experience

David Silver, Richard S. Sutton*

Abstract

We stand on the threshold of a new era in artificial intelligence that promises to achieve an unprecedented level of ability. A new generation of agents will acquire superhuman capabilities by learning predominantly from experience. This note explores the key characteristics that will define this upcoming era.

The Era of Human Data

Artificial intelligence (AI) has made remarkable strides over recent years by training on massive amounts of human-generated data and fine-tuning with expert human examples and preferences. This approach is exemplified by large language models (LLMs) that have achieved a sweeping level of generality. A single LLM can now perform tasks spanning from writing poetry and solving physics problems to diagnosing medical issues and summarising legal documents.

However, while imitating humans is enough to reproduce many human capabilities to a competent level, this approach in isolation has not and likely cannot achieve superhuman intelligence across many important topics and tasks. In key domains such as mathematics, coding, and science, the knowledge extracted from human data is rapidly approaching a limit. The majority of high-quality data sources - those that can actually improve a strong agent's performance - have either already been, or soon will be consumed. The pace of progress driven solely by supervised learning from human data is demonstrably slowing, signalling the need for a new approach. Furthermore, valuable new insights, such as new theorems, technologies or scientific breakthroughs, lie beyond the current boundaries of human understanding and cannot be captured by existing human data.

The Era of Experience

To progress significantly further, a new source of data is required. This data must be generated in a way that

Silver, David, and Richard S. Sutton. "Welcome to the era of experience." *Google AI* 1 (2025).

モデルは常に間違っている – ただし、役立つモデルもある … George Box, 1976

アジェンダ

1. 人工知能研究の今まで
    - 最初の50年 – 賢い人の思考を対象とした人工知能
    - **次の15年 - 普通の人の思考を対象とした人工知能**
2. 人工知能技術の課題とガバナンス
    - 生成モデルの限界
    - 今のAIは「計算機科学の総合格闘技」
    - 人工知能の危険
3. 人工知能研究のこれから
    - 人にできない思考を対象とした人工知能
    - 人間との棲み分け - 超知性と共存する社会へ

人工知能の冬

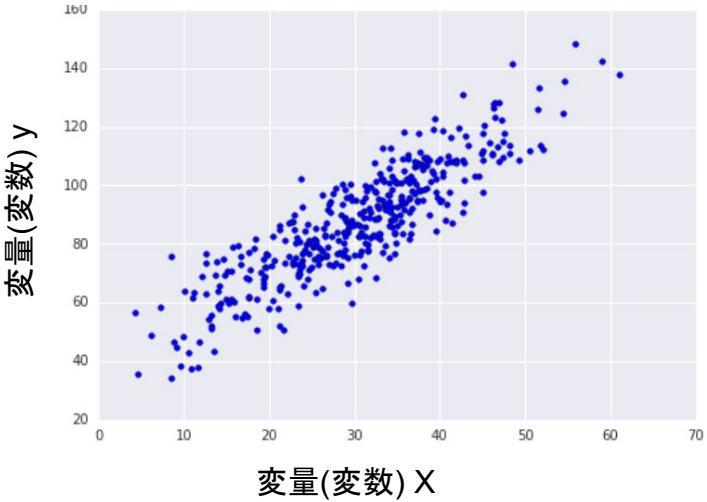
- 知識と探索で「賢い人の思考」の多くは解けるようになったが、、、
  - 人が決めたルールが厳密にわかっている世界

- 普通の人間が(子供でも)苦も無く行える、画像認識や言語運用は全然だめ！！
  - ルールが数理的に決められない世界

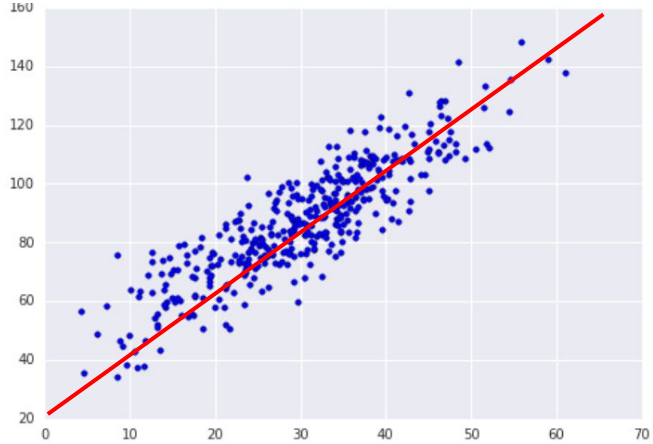
普通の人の思考を模倣する人工知能とは？

経験の世界：統計モデリング：観測されたデータにパターンを見つけること

観測データ

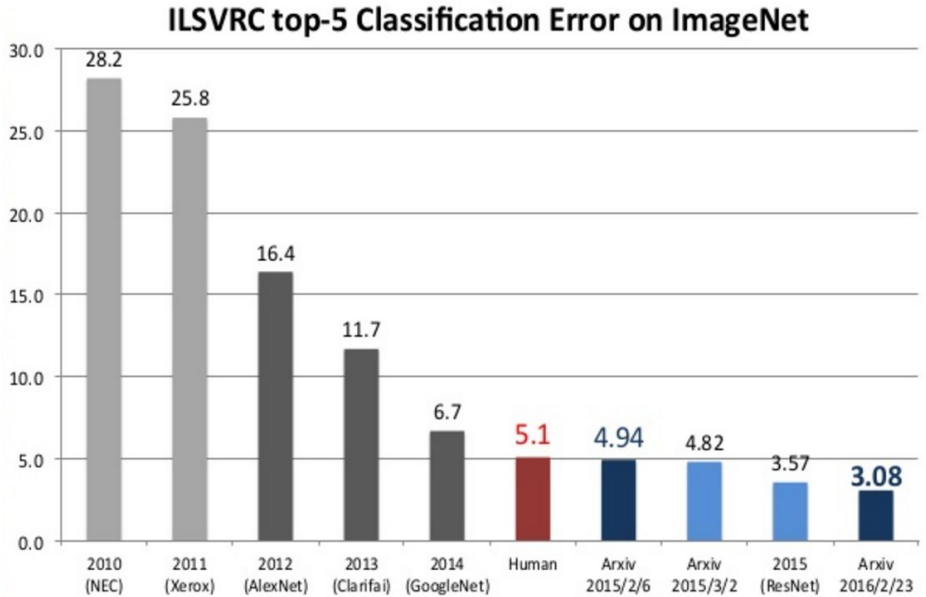


モデル： y = ax + b + ε



統計モデル とはデータに現れるパターン表現したもの

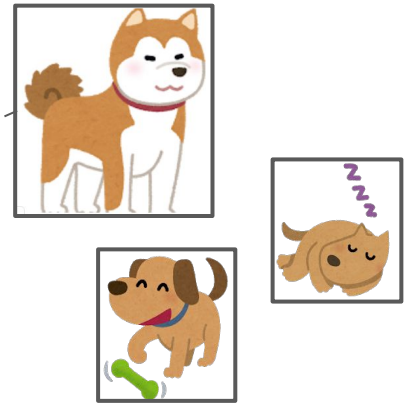
統計モデリング(1): 画像を認識する
深層学習によるブレークスルー – AlexNet, 2012



http://www.image-net.org/challenges/LSVRC/

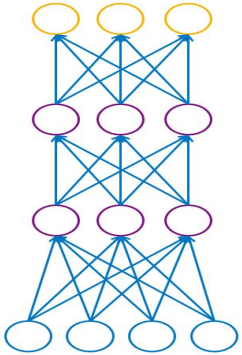
深層学習は基本的に統計モデリング -- ただし超高次元

観測データ



100ピクセル x 100ピクセル x RGB = 3万個の変数を持つベクトルデータ

モデル：数千万個のパラメタを持つ関数

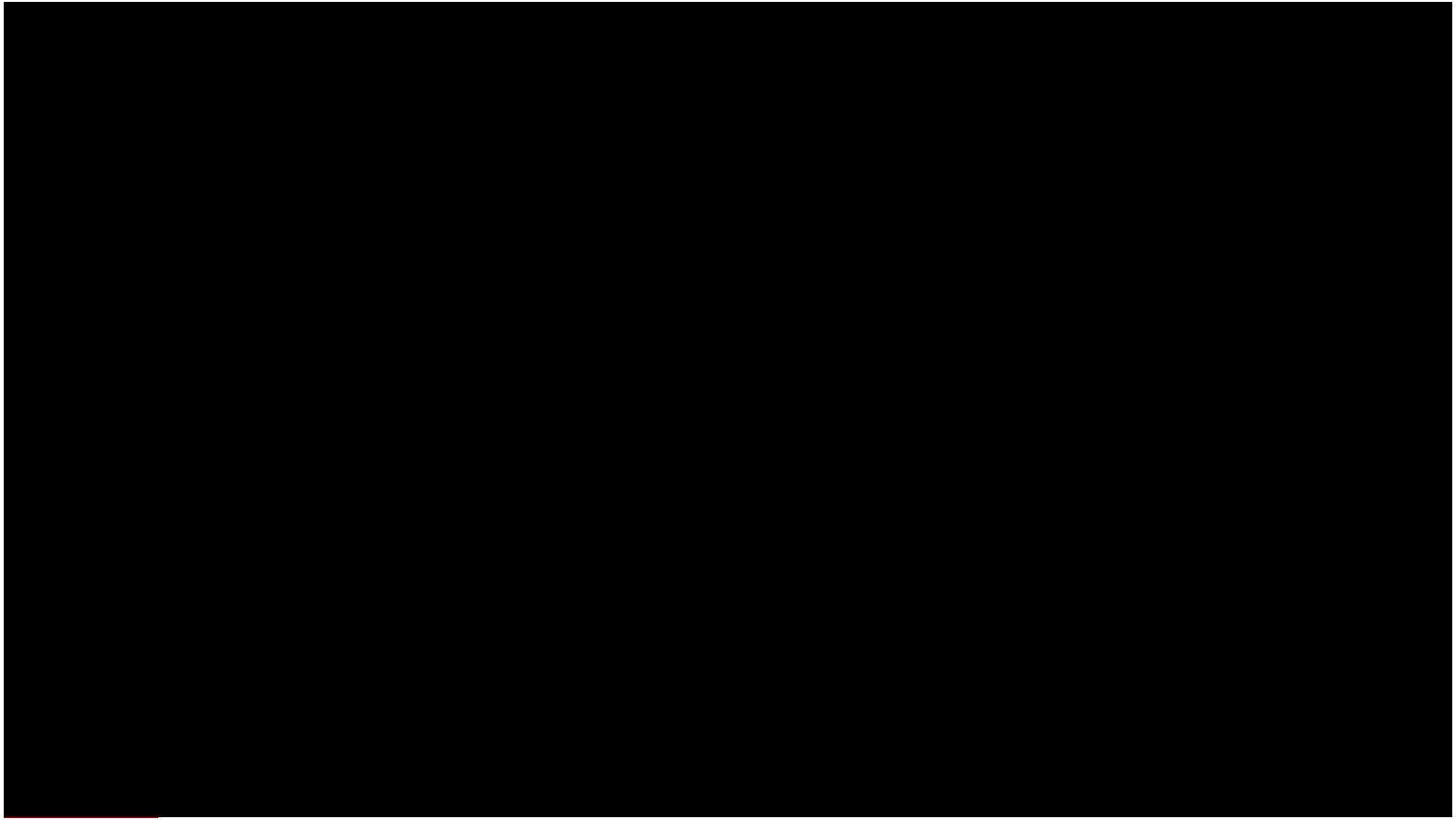


各変数（画素）間の関係（パターン）を統計的に表現したもの

統計モデリング(2):自動運転のためのセグメンテーション

https://www.youtube.com/watch?v=lGOjchGdVQs

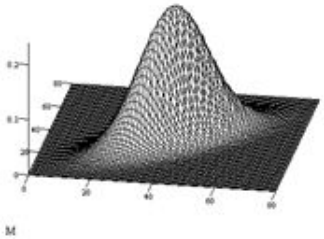
統計モデリング(3):3次元形状の再構成



統計モデリング(4):大規模言語モデル(LLM)

大量の言語データ

新聞
ビジネス文書
書籍
SNS
論文
メール



ある文脈cが与えられた時に、次に来る単語wの確率モデル Pr(w|c)

例: “陽春の候” の後に “貴社ますますご清栄” が続く確率はいくらか？

ChatGPT

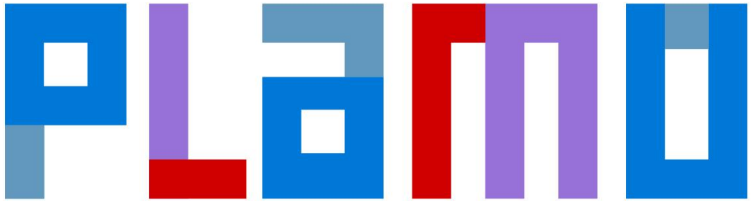
出力の適切さに関する、大量のアノテータによるファインチューニング

研究者の驚き -- **言語モデルでそこまでできるのか！** *

* 岡崎直観,「大規模言語モデルの驚異と驚異」, https://speakerdeck.com/chokkan/20230327_riken_llm

PLaMo：世界最高レベルの日本語性能を持ったLLM

- GENIACプロジェクト（2024年2~8月）で開発
- 製造業、素材産業、医療、金融などの専門領域での応用を目指す
- PLaMo Primeとして2024年12月にAPI提供開始



純国産フルスクラッチモデル

- 独自構築したアーキテクチャ・学習データ
- ゼロからの事前学習･指示学習

世界最高クラスの日本語性能

- 主要な日本語ベンチマークにおいてGPT-４oを超える精度を記録

クラウドおよびオンプレミスで利用可能予定

- 閉じたオンプレミス環境でも利用可能

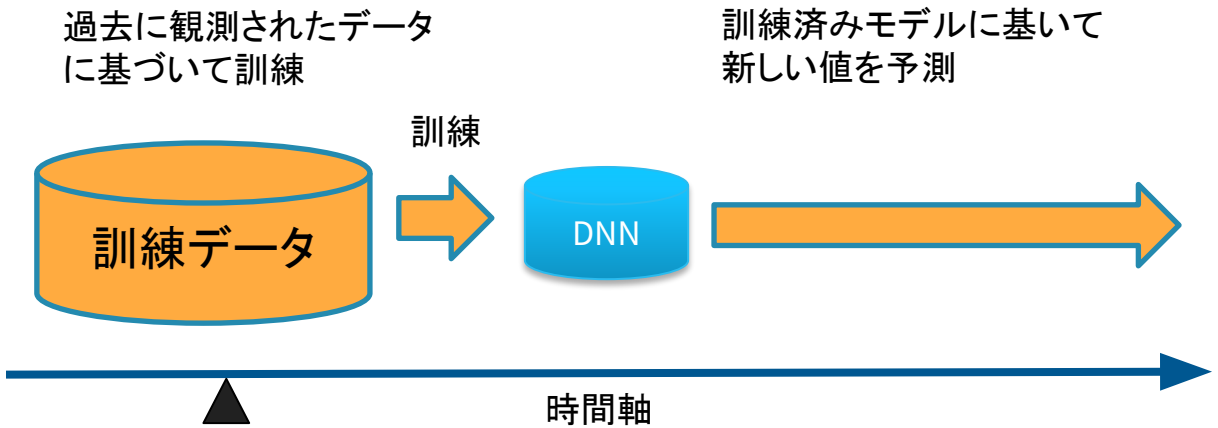
GENIAC技術モデル賞
GENIACナレッジ賞

Preferred Networks

アジェンダ

1. 人工知能研究の今まで
    - 最初の50年 – 賢い人の思考を対象とした人工知能
    - 次の15年 - 普通の人の思考を対象とした人工知能
2. **人工知能技術の課題とガバナンス**
    - **生成モデルの限界**
    - 今のAIは「計算機科学の総合格闘技」
    - 人工知能の危険
3. 人工知能研究のこれから
    - 人にできない思考を対象とした人工知能
    - 人間との棲み分け - 超知性と共存する社会へ

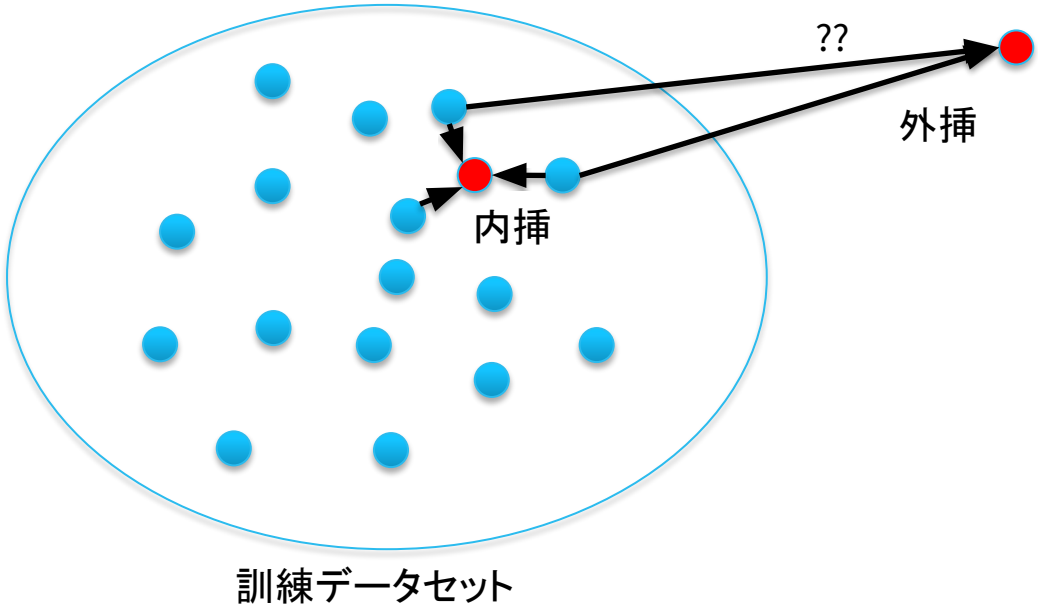
統計モデリングの本質的限界 (1)



将来が過去と同じでないと正しく予測できない

統計モデリングの本質的限界 (2)

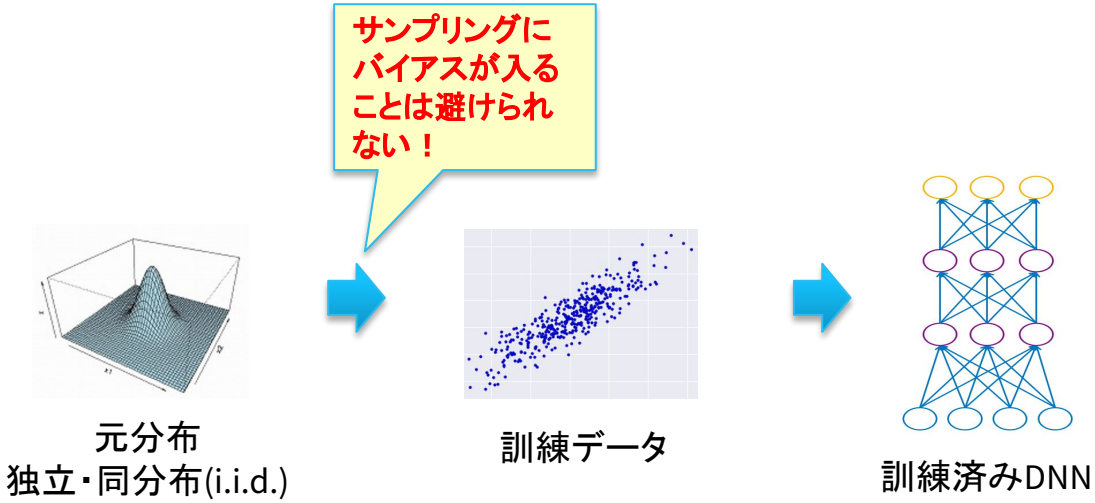
- 訓練データセットは有限



訓練データに現れない、希少な事象に対して無力

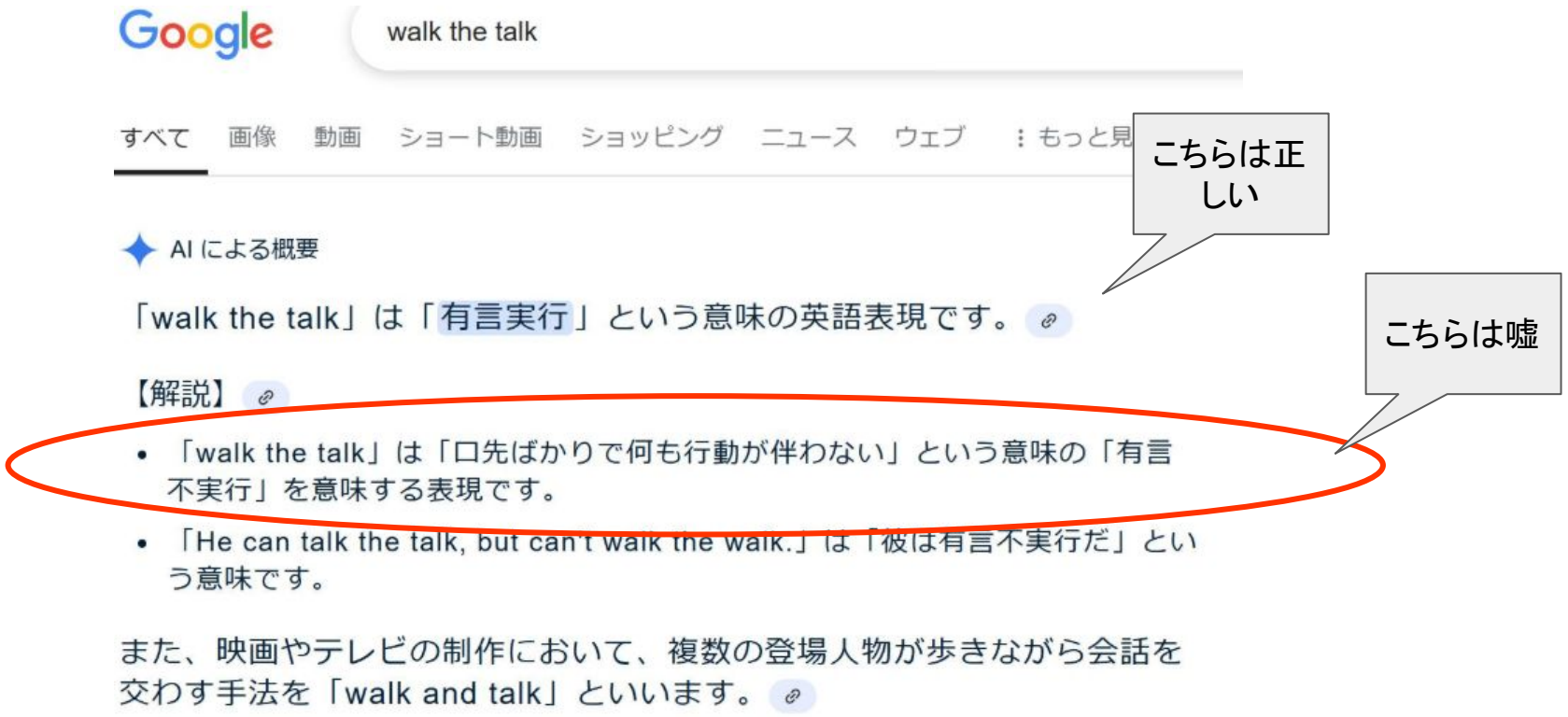
統計モデリングの本質的限界 (3)

- 本質的に確率的



「100%の正しさ」は保証できない

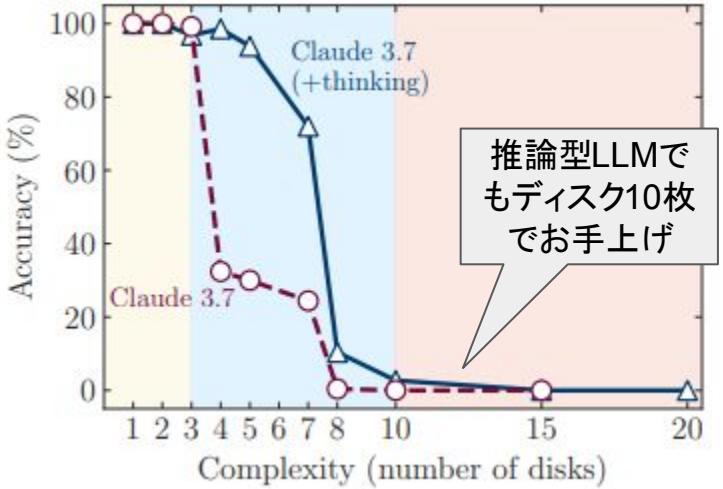
ハルシネーション -- 事実で無い事を述べる



出典: 丸山自身による2025/03/04時点での goole.com へのアクセス

LLMはハノイの塔が苦手(推論型LLMでも！)

アップルの論文「思考するLLMの幻想」



Shojaee, Parshin, et al. "The illusion of thinking: Understanding the strengths and limitations of reasoning models via the lens of problem complexity." *arXiv preprint arXiv:2506.06941* (2025).

ディスクが20枚程度までなら瞬時に解くプログラム

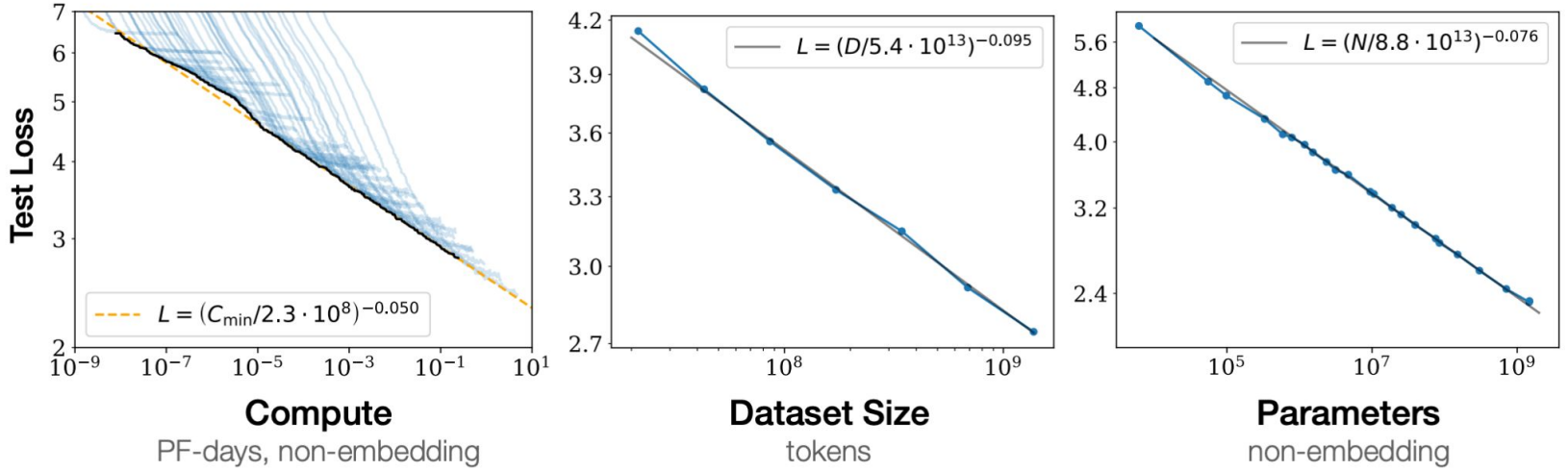
```python
def hanoi(n, source=1, target=3, auxiliary=2):
    if n == 1:
        print(f"円盤1を棒{source}から棒{target}へ移動")
        return
    hanoi(n-1, source, auxiliary, target)
    print(f"円盤{n}を棒{source}から棒{target}へ移動")
    hanoi(n-1, auxiliary, target, source)

hanoi(10)
```

アルゴリズムが得意な問題に LLMを使うべきではない！

アジェンダ

1. 人工知能研究の今まで
    - 最初の50年 – 賢い人の思考を対象とした人工知能
    - 次の15年 - 普通の人の思考を対象とした人工知能
2. **人工知能技術の課題とガバナンス**
    - 生成モデルの限界
    - **今のAIは「計算機科学の総合格闘技」**
    - 人工知能の危険
3. 人工知能研究のこれから
    - 人にできない思考を対象とした人工知能
    - 人間との棲み分け - 超知性と共存する社会へ

LLMの学習スケーリング則 [Kaplan+ 2020]



言語モデルでTransformerをモデルとして使った場合に
「学習時投入計算量」「学習データ量」「モデルサイズ」と、
「検証データのクロスエントロピー損失」との間にべき乗則が成り立つ
（予測正解率のようなもの）

→　予測可能な形で予測性能や後続タスク性能が改善される

→　**LLMの計算資源の投入競争に発展。**

Preferred Networks

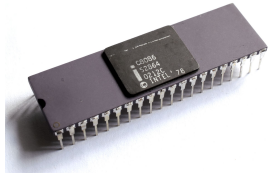
コンピュータの歴史

1950 1960 1980 2020

専用機の時代 | 汎用機の時代 | GPUの時代

- 科学技術計算
- 事務計算
- :

IBM S/360



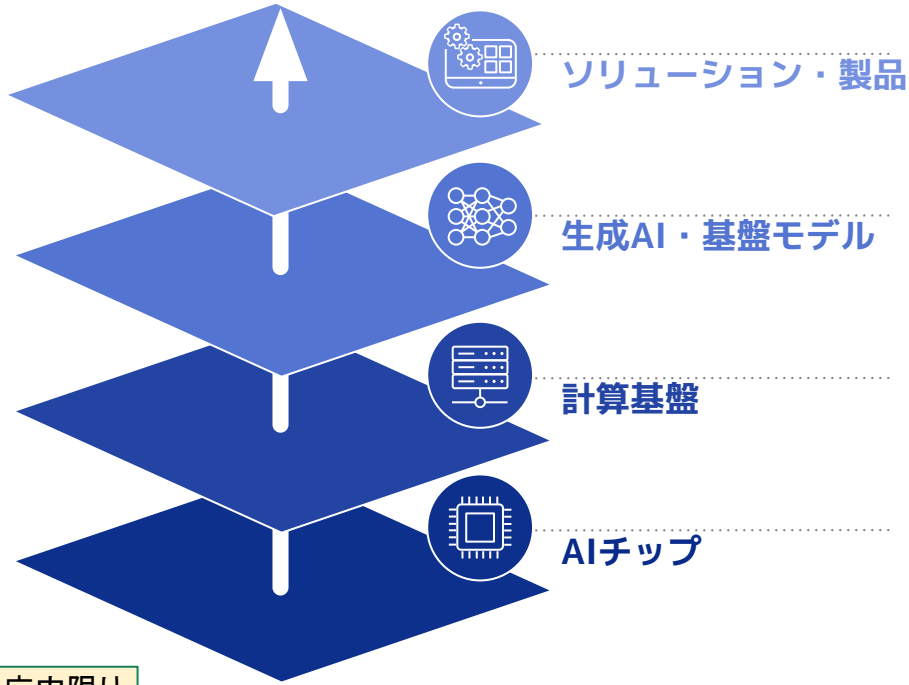
intelプロセッサ

エヌビディア 時価総額4兆ドル突破 “史上初” メディア伝える

2025年7月10日 4時09分 半導体

AIは計算機科学の総合格闘技： AI技術を垂直統合する必要性

Preferred Networks(PFN)は、生成AIを一気通貫で国産化



様々な産業・消費者向けのソリューション・製品群

MATLANTIS
計算化学シミュレータ

Preferred Networks Visual Inspection
外観検査ソフトウェア

PFN 3D Scan
3Dスキャンサービス

rypko
キャラクター生成プラットフォーム

kachaka
自律移動ロボット

ハイパーパラメータ自動最適化フレームワーク

PLaMo 13B
大規模言語モデル

PLaMo
マルチモーダル基盤モデル
(2024年リリース予定)

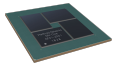
Preferred Potential (PFP)
物質の電子状態・エネルギー計算モデル

GPUクラスタ

MN-3
(MN-Core™ クラスタ)

MN-4
(MN-Core™2 クラスタ

MN-Core™ 2による計算能力のクラウド提供
(2024年開始予定)


MN-Core™

MN-Core™ 2

次世代機

貴省庁内限り

計算基盤の電力効率

計算基盤

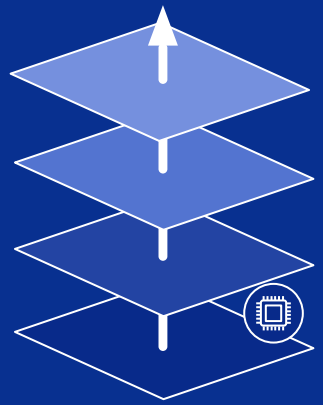
PFNは、電力効率の高い計算基盤の構築のため研究開発を続けています。独自のAIチップMN-Core™（第1世代）を搭載したMN-3は、スーパーコンピュータの電力効率ランキングGreen500で2020年6月、2021年6月、2021年11月に世界1位を3度獲得しました。

詳細: https://projects.preferred.jp/supercomputers/

貴省庁内限り

Preferred Networks

AIチップ

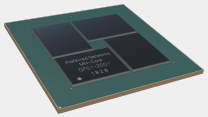


MN-Core™シリーズのロードマップ

PFNはAIの計算資源としての半導体の需要増をいち早く予測し、2016年にAIプロセッサーMN-Core™シリーズ第１世代の開発を開始。現在ではフラッグシップモデルと生成AI向けの2つのラインで開発・製品化を進めています。

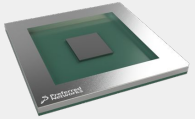
Flagship
AI学習・推論
高性能計算(HPC)

MN-Core
(TSMC 12nm)



2016年～開発
2020年～社内利用
2023年～計算力社外提供

MN-Core 2
(TSMC 7nm)



2023年～試験運用
2024年～社外販売・
PFCP™で計算力社外提供

超省電力AI アクセラレータ
(Samsung 2nm)

開発中

2024年～NEDO委託事業で
研究開発、成果を応用し
製品化予定

次世代

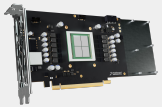
検討中

Rapidus、さくらインターネットと国産AIインフラの提供を目指し基本合意

2016 2020 2023 2026

生成AI向け
LLM推論など

MN-Core L1000



2024年～開発
2026年提供開始予定

MN-Core L2000

検討中

MN-Core L3000

検討中

詳細: https://projects.preferred.jp/mn-core/

Preferred Networks

アジェンダ

1. 人工知能研究の今まで
    - 最初の50年 – 賢い人の思考を対象とした人工知能
    - 次の15年 - 普通の人の思考を対象とした人工知能
2. **人工知能技術の課題とガバナンス**
    - 生成モデルの限界
    - 今のAIは「計算機科学の総合格闘技」
    - **人工知能のリスク**
3. 人工知能研究のこれから
    - 人にできない思考を対象とした人工知能
    - 人間との棲み分け - 超知性と共存する社会へ