AIの現状と残された課題

Current State of AI and Remaining Challenges

山田 誠二

Seiji Yamada

国立情報学研究所／総研大

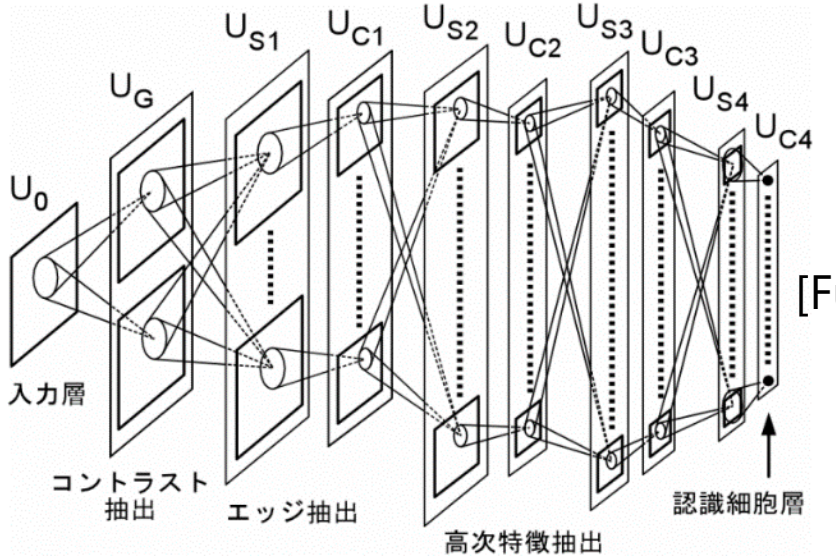
NII/Sokendai

人工知能学会 元会長

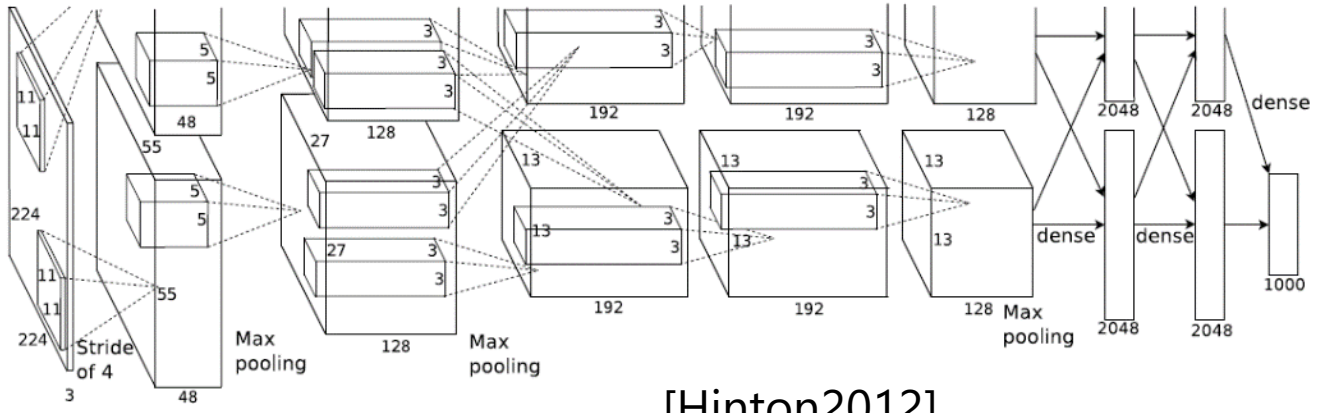
NNの復権：ディープラーニング

Deep learning: reinstatement of NN

- 最も成功した CNN
  - CNN 畳み込みニューラルネットワーク+プーリング層 pooling で多層 NN (4～100層)
  - 脳の視覚野 V1～V5
  - Neo-cognitron [Fukushima1979] の枠組みそのまま



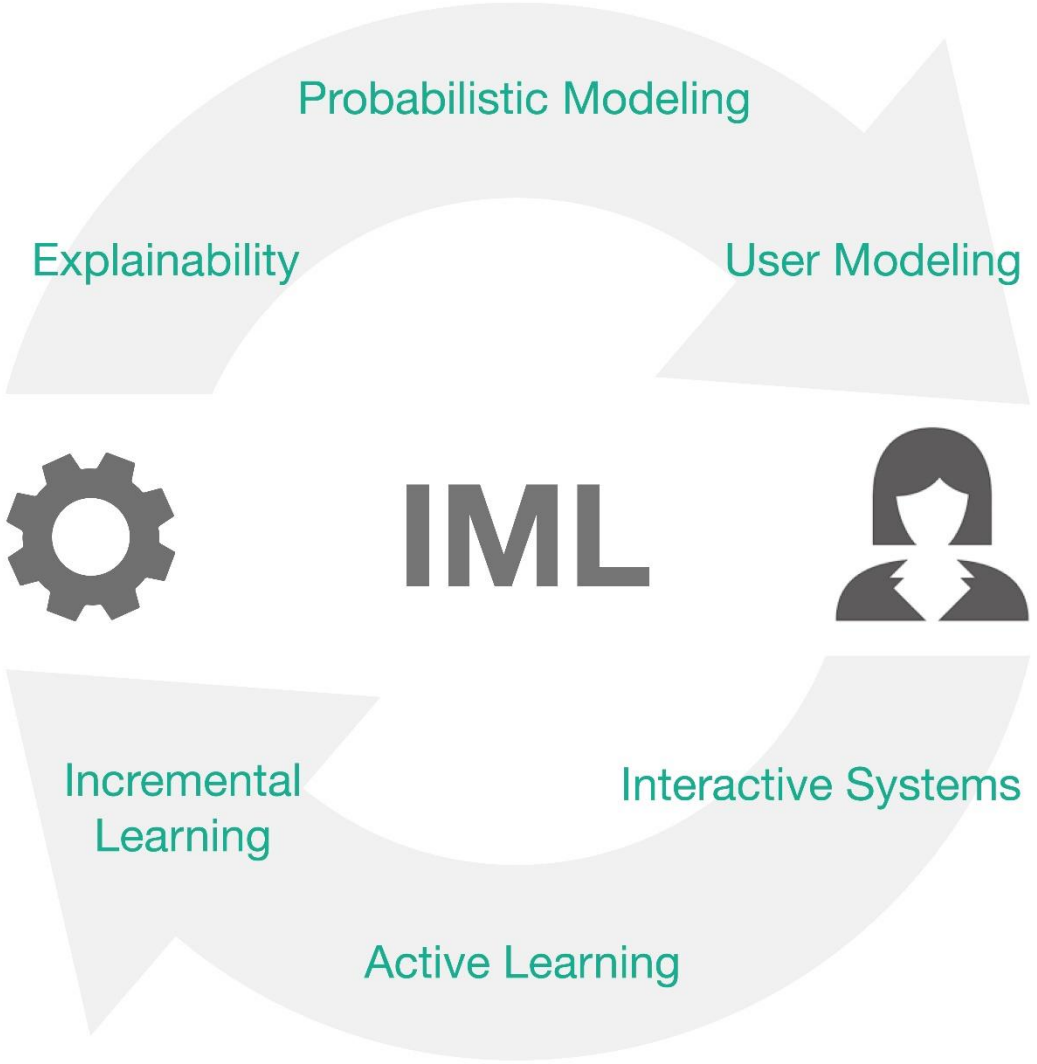
[Fukushima1979]



[Hinton2012]

AI研究 (ML) のトレンド ～NeurIPS, ICML, ICRLからの主観的解釈～

- 人間-AI協調 (ML・decision making) Human-AI cooperation
  - 人間の知識を導入して学習を高速化 ⇐ ある意味，80年代回帰
    Accelerating ML with human knowledge
- 人間の知識の利用方法
  - インタラクティブ機械学習・
    Human-in-the-loop, Interactive Machine Learning



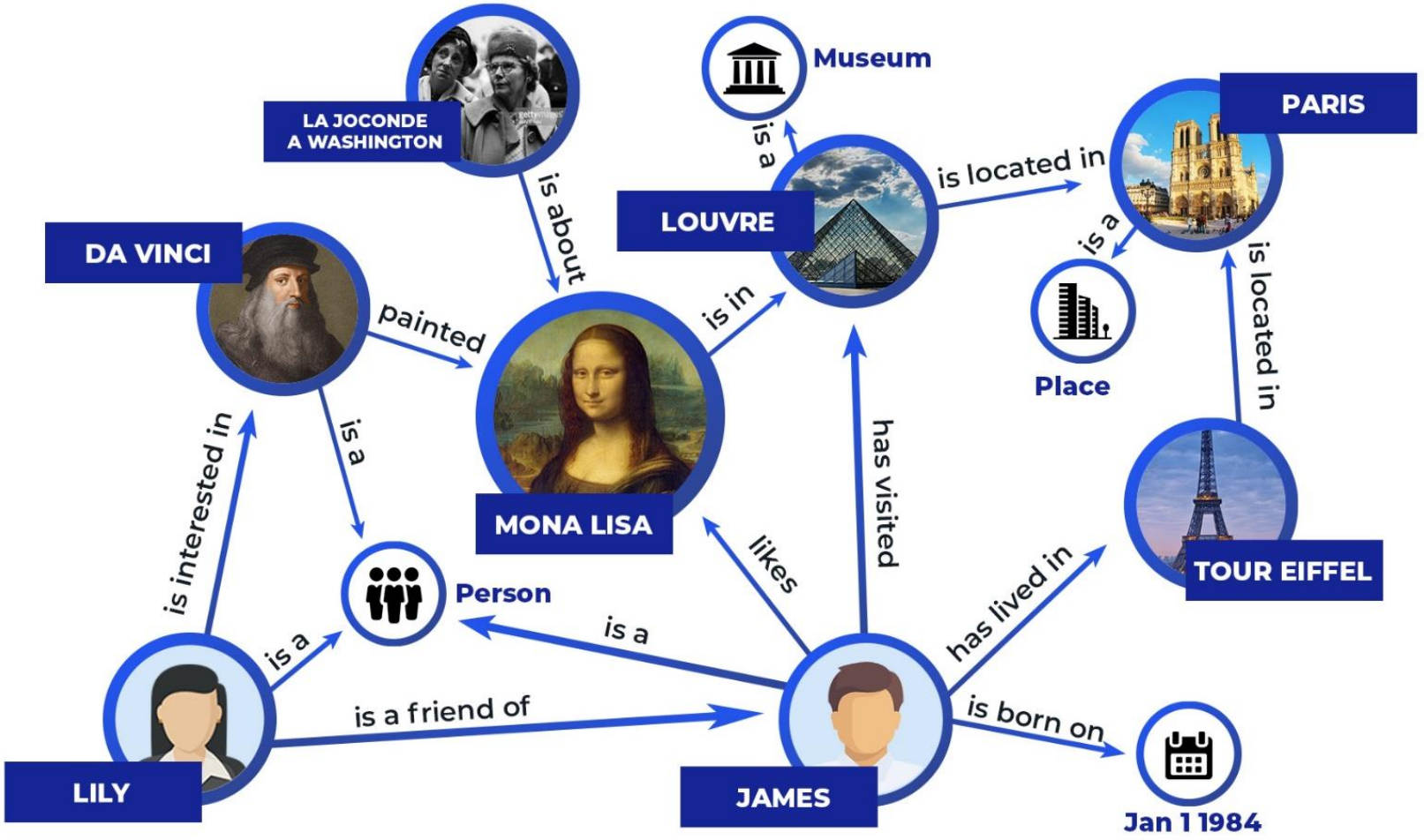
ユーザフィードバック User Feedback

- 知識表現 for ユーザーフィードバック Knowledge rep.
  - 明示的 explicit UF：知識グラフ knowledge graph
  - 非明示的 implicit UF：視点追跡 eye-tracking

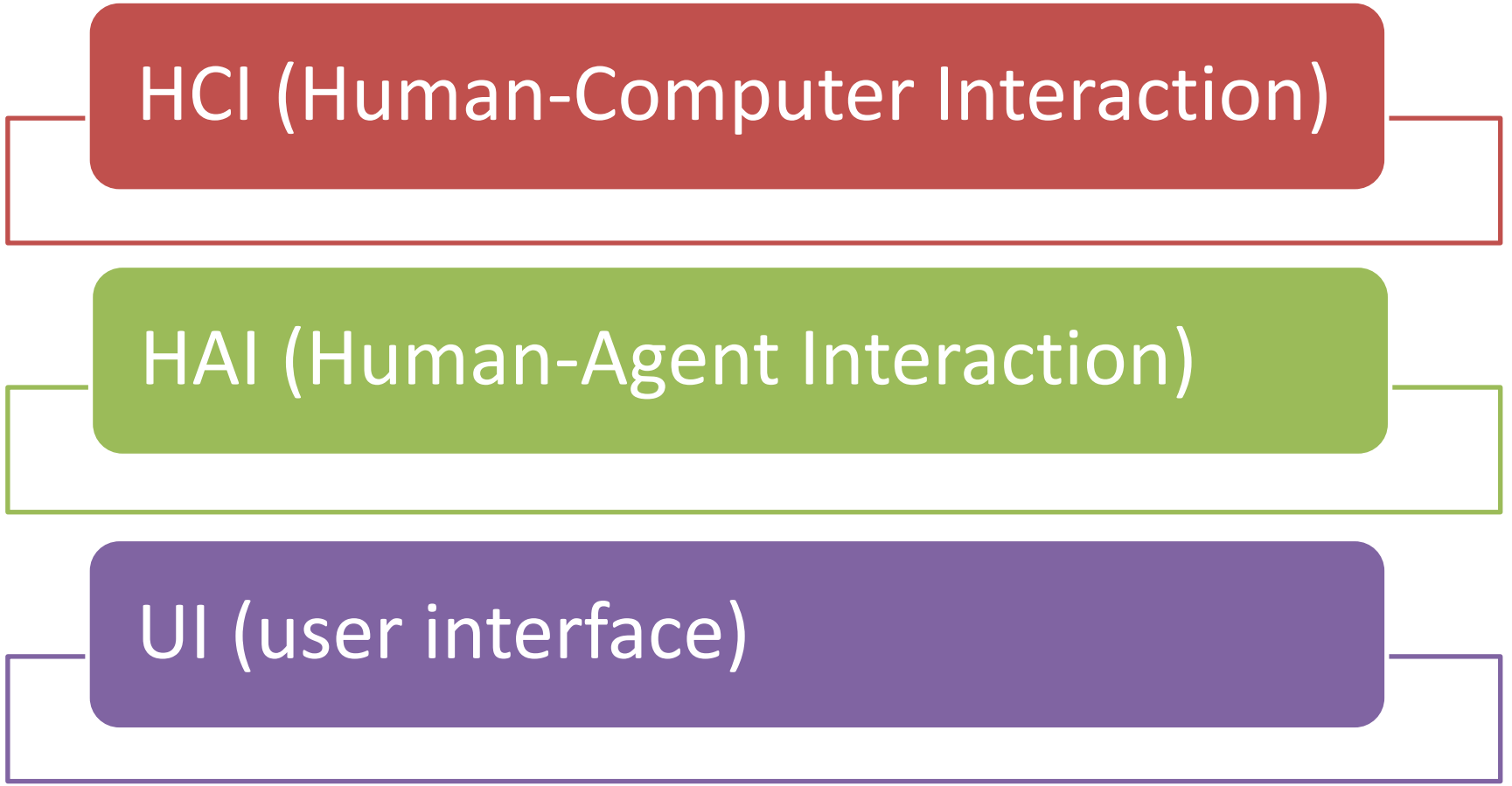
- 研究例
  - 知識グラフでLLMを強化！
  - 知識グラフで分類学習・強化学習を改善！

知識グラフ

- Googleの提唱する知識表現（semantic webの **RDF**, 意味ネットワークと類似，古くは意味ネットワークと類似）
- AIの世界で標準化されつつある



要素技術：人間とAIのインタラクションデザイン



少数訓練データからの学習

- Zero/one/few-shot learning：一つ／少数の訓練データからの学習 ⇐ 事前学習のあとのファインチューニング
  - ChatGTPの in-context learning
- 自己学習 self-learning
  - データ拡張 (data argumentation)により訓練データを水増し
- 半教師あり学習 semi-supervised learning
  - 同じクラスター内のデータは同じラベル
- 共通の仮定：類似データは，類似（正解）ラベルをもつ

AIの得意・不得意

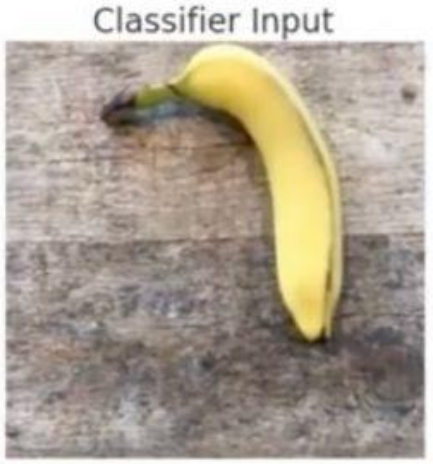
AI's strong and week points

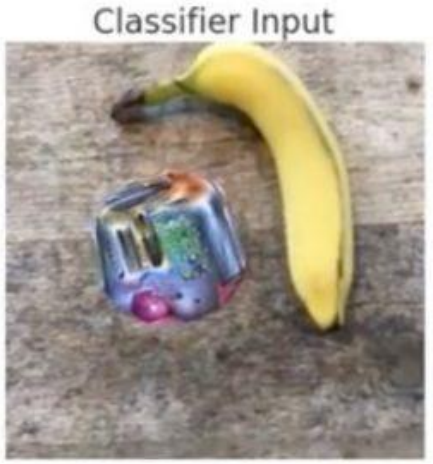
- AIの得意分野 strong points
  - （複雑だが）静的 static で閉じた世界 closed world
  - 例：(完全情報)ゲーム game，屋内環境 in-door environment

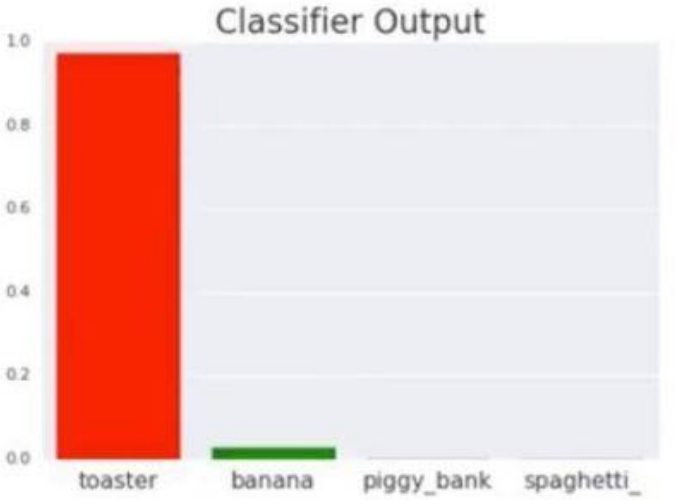
- AIの不得意分野 week points
  - 動的 dynamic で開いた世界 open world，**常識** common sense
  - 例：人間の行動を含む系 Human-AI systems, 屋外環境 out-door environment

（実は ....）だまされやすいAI Gullible AI in fact

常識：背景とは何か







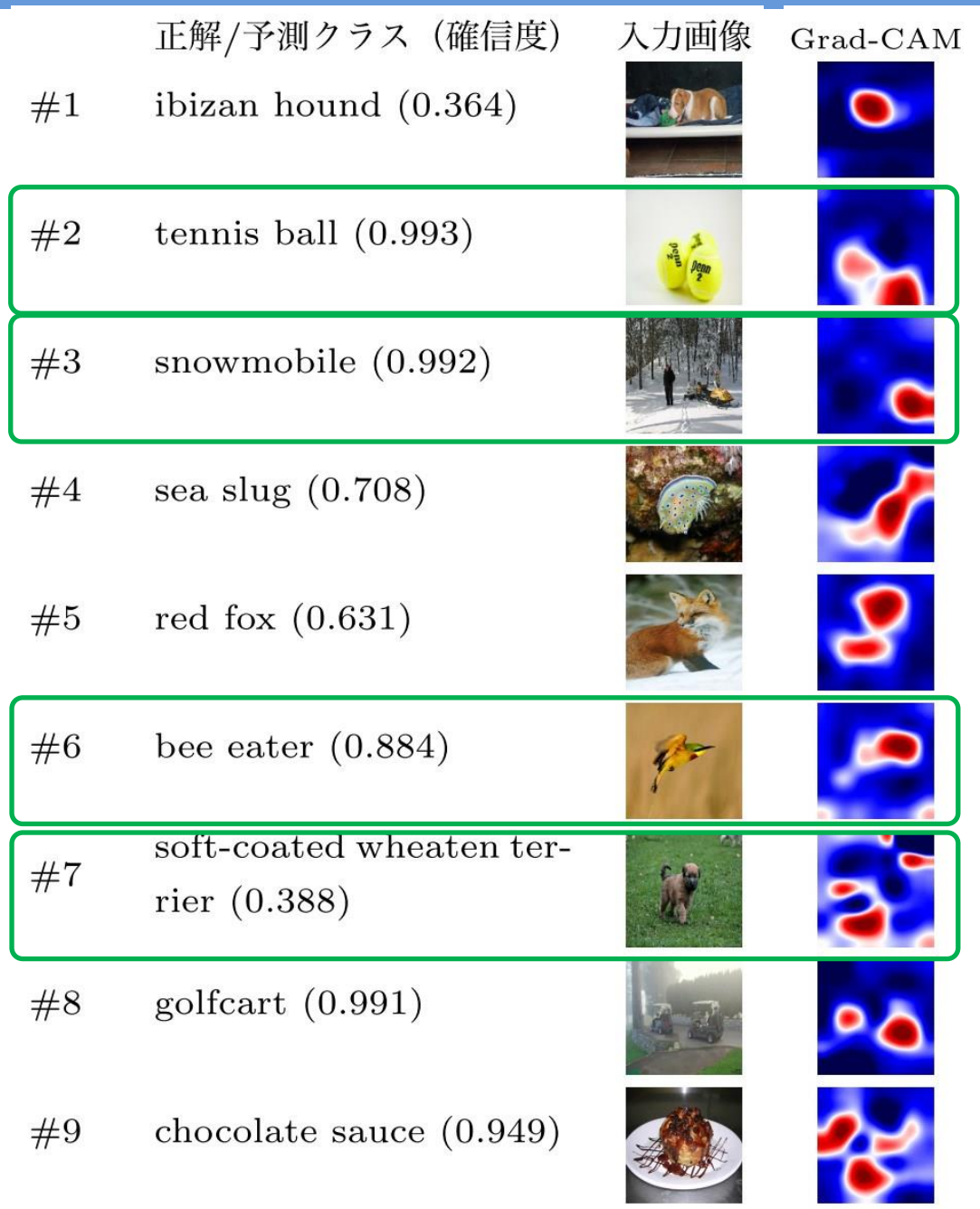
T. B. Brown, D. Mané, A. Roy, M. Abadi and J. Gilmer/Neural Information Processing Systems 2017

疑似相関

pseudo-correlation

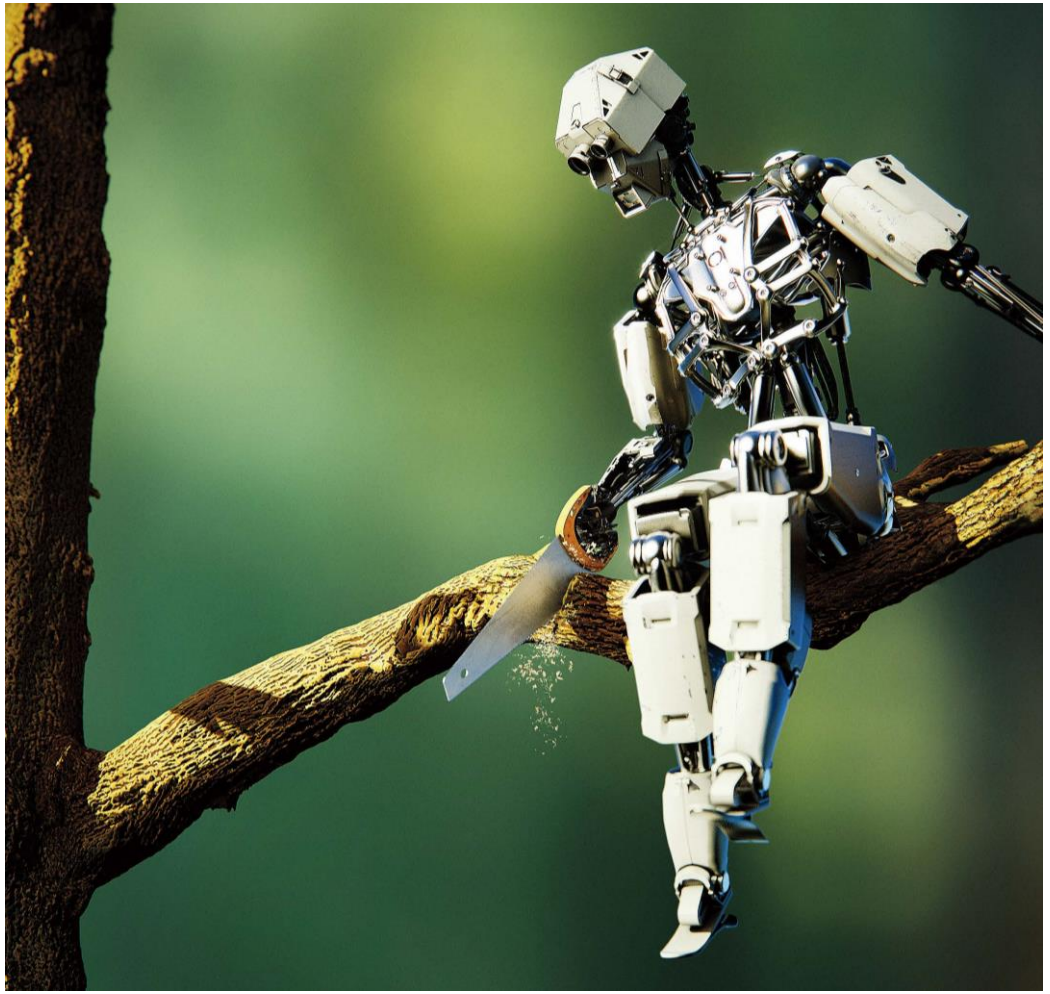
- Grad-CAMでも，かなりの疑似相関が見られる.

  Much pseudo-correlation in Grad-CAM



「常識」とは？ What's common sense?

- 物理的常識（右図）Physical CS (right figure)
  - 物理現象，自然現象
- 社会的常識 Social CS
  - 社会通念，モラル
- 膨大な量の知識
  - 書き尽くせない
  - 機械学習も非現実的



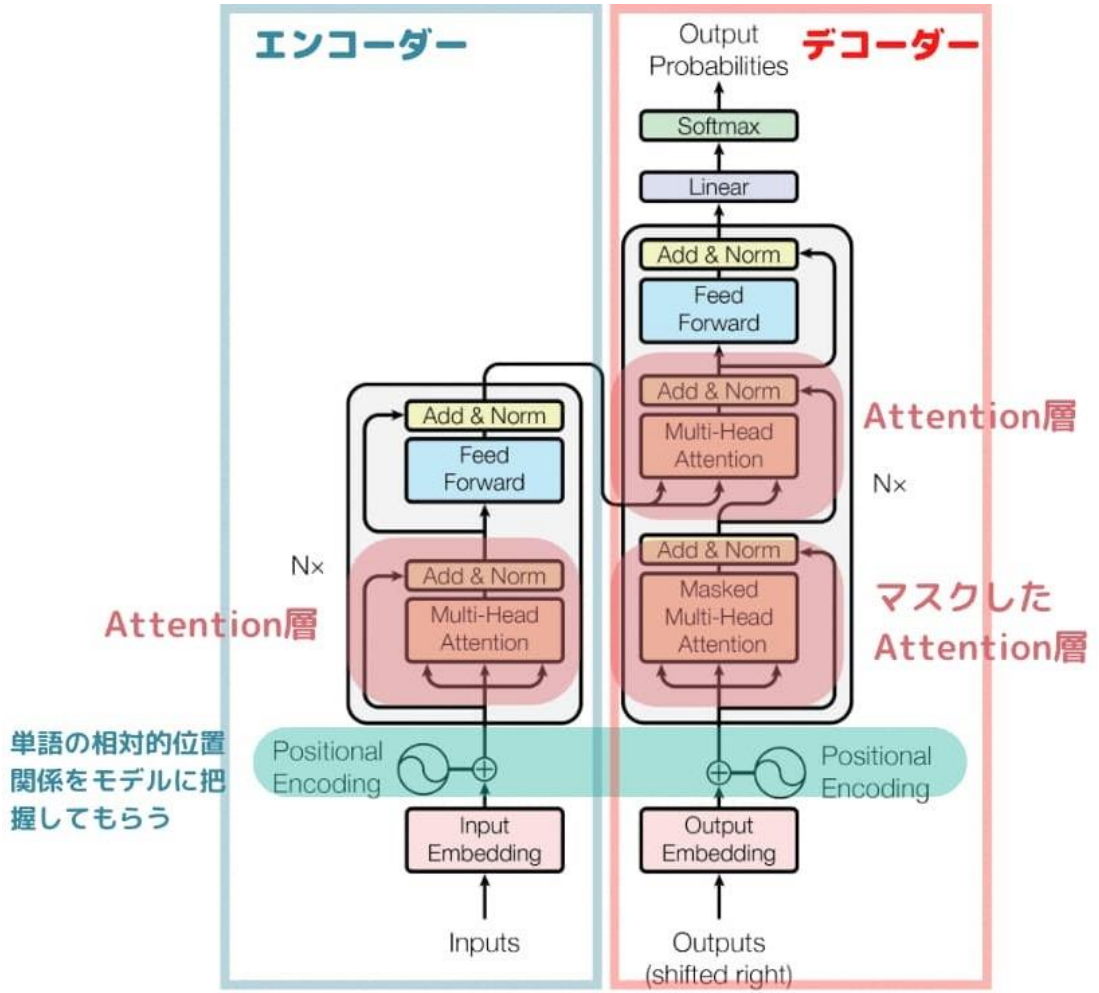
Communications of the ACM, 58(9) 2015

ChatGPTとは？

- OpenAIが開発，ローンチしたLLMのエンドユーザ向けフロントエンド.
- 基本，無料で使える.
- 使われているLLM
  - 現在GPT3からGPT4（有料）
  - GPT3
    - webデータは，**2021年9月**まで.
    - 例：Q：「日本の総理は誰？」，A：「菅義偉」

ChatGPTの本質とは何か？

- **LLM (大規模言語モデル)とは？**
  - **膨大なN-gramを使った次単語予測システム** N-gram：N個の単語系列から次の単語を予測する確率モデル
  - **GPT4**
    - **32,767**-gram！
    - パラメータ数 ＞ 数兆個
    - 20兆～30兆個の単語による訓練データ
      - この数は，人類がこれまで書いたすべての本にある単語の数に相当か越える．
  - 入力：単語の埋め込み表現
  - アテンションにより非連続・長期コンテクストを利用可能



Transformerの構造

ChatGTPで『常識』は？

Common sense in ChatGTP

- 本質的な問い：ChatGTPは常識を獲得したのか？ Did ChatGTP get common sense?

No!

– **口からでまかせ** hallucination

身近な『口からでまかせ』Familiar hallucination

- Q: こんな論文，教えて！ Tell me papers like this.

- A: 文献リストに『口からでまかせ』が！ Much hallucination in retrieved papers.
  - 複数論文の For several papers,
    - **筆者名を平気でマージ Merging author's names**
    - **論文タイトルを平気でマージ Merging titles**
    - **DOIを平気でマージ Merging DOIs**
    - **URLを平気でマージ Merging URLs**
    - . . . .

まとめ Conclusion

- AI(ML)のトレンド Trend of AI(ML)
  - 人間の知識を利用 Utilizing human knowledge.
- AIの得手・不得手 AI’s strong and weak points.
  - 疑似相関と常識のなさ Pseud correlation and few common sense.
- ChatGPTにも残る不得手＝口からでまかせ＝常識のなさ
  Weak points of ChatGPT = hallucination
  = few common sense.