

# 科学のための機械学習

瀧川 一学

<https://takigawa-lab.tokyo/>

東京大学 新領域創成科学研究科 メディカル情報生命専攻

北海道大学 化学反応創成研究拠点 (WPI-ICReDD)

理化学研究所 革新知能統合研究センター (AIP)

たきがわ いちがく

# 自己紹介：瀧川一学

関心

離散構造を伴う機械学習 + 自然科学における機械発見  
(機械学習の研究者) (機械学習のユーザ)

うどん県 出身

- 札幌 • 北大 工学研究科 (10年)
- 京都 • 京大 化学研究所 / 薬学研究科 (7年)
- 札幌 • 北大 情報科学研究科 (7年)
- JST さきがけ (3.5年)
- 京都 • 理研 AIPセンター 京大 iPS細胞研連携G (6年)
- 札幌 • 北大 化学反応創成研究拠点 (7年～現在)
- 京都 • 京大 国際高等教育院 (3年)
- 千葉 • 東大 新領域 **メディカル情報生命** 専攻 (現在)
- 理研 AIPセンター **化学反応情報学T** (現在)

機械学習×信号処理(信号源分離)

**バイオインフォマティクス・創薬化学**

機械学習×計算機科学(離散構造)

**材料インフォマティクス**

機械学習×幹細胞生物学

機械学習×化学

機械学習×データサイエンス教育

機械学習×分子生物学

機械学習×化学

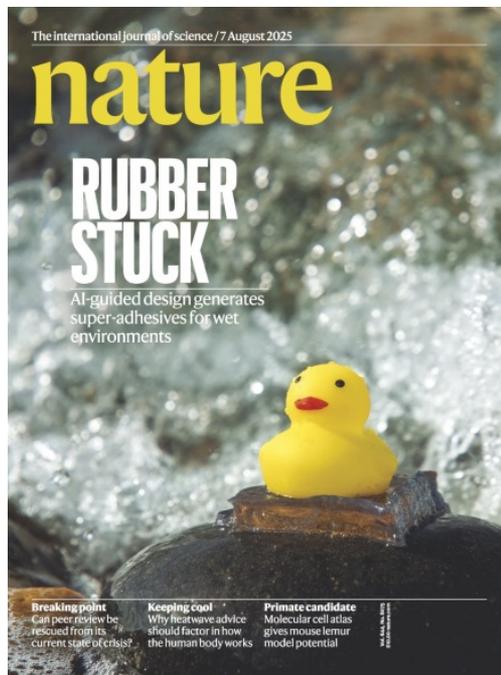
たきがわ いちがく

# 自己紹介：瀧川一学

機械学習屋(技術屋)であって  
科学者ではない

- Area Chair, NeurIPS 2025
- Area Chair, NeurIPS 2024
- Top reviewer, NeurIPS 2023.
- Top reviewer, NeurIPS 2022.
- Distinguished PC, IJCAI-ECAI 2022.
- Outstanding reviewer, ICML 2022.
- Highlighted reviewer, ICLR 2022.
- Expert reviewer, ICML 2021.
- Top reviewer, ICML 2020.
- Best reviewer, NeurIPS 2019.
- Active reviewer, IJCAI 2019.
- Top reviewer, NeurIPS 2018.

Bioinfoと機械学習で強力な  
水中接着剤を見つけた！



<https://www.nature.com/nature/volumes/644/issues/8075>

言葉に囚われた私たちの  
科学と人工無能について(?)



Can Artificial  
Intelligence  
Unravel the Faculty  
of Language?

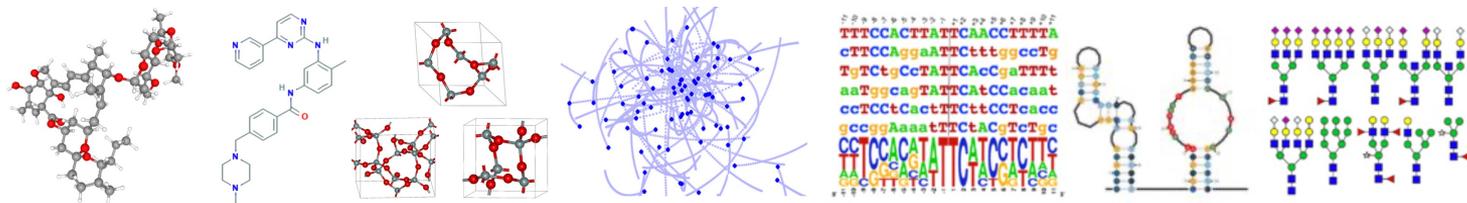
岩波書店

<https://www.iwanami.co.jp/book/b10134161.html>

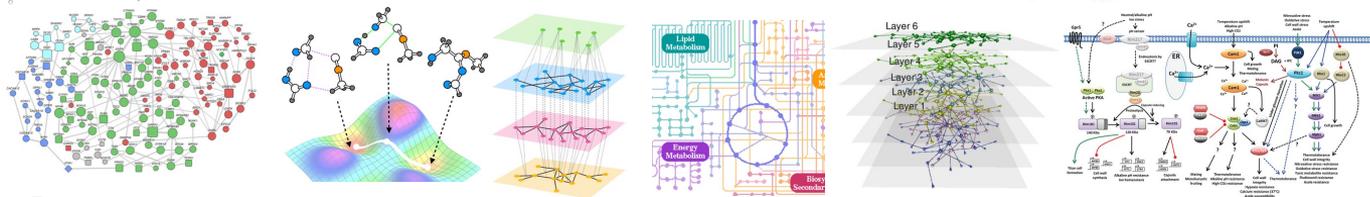
# 関心：離散構造(組合せ情報)を伴う機械学習

離散構造 = 有限個の離散要素の組合せによって生じる情報  
 → 集合、論理、系列、ツリー、グラフ、プログラム、言語など

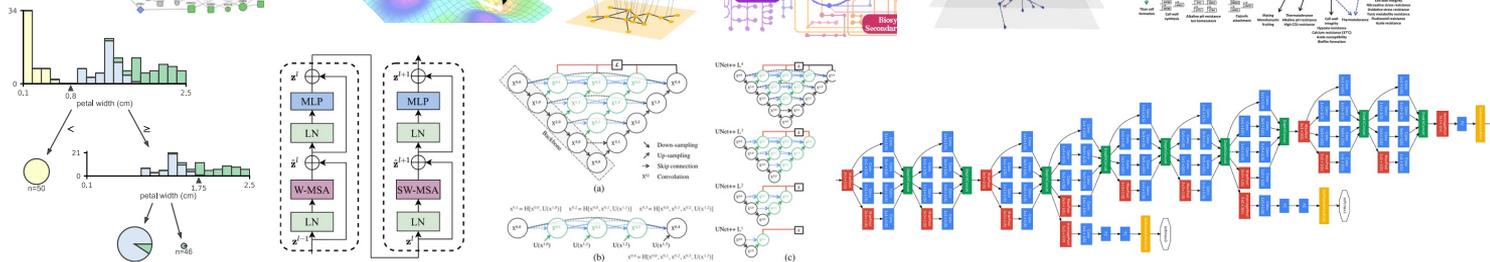
対象が  
離散構造



対象間の関係が  
離散構造



モデルが  
離散構造



# AI for Science: Goal Misspecificationを超えて

- 機械学習のゴールと科学者のゴールが根本的に異なり得る  
→ ニュースで目にする機械学習の成功例では**これが基本ズレてない**
- AI for Science  
**「データにないこと」を「データに基づいて」希求する方法の研究**  
→ もはや「機械学習」の問題ではなく、より難しい問題 (例: 機械発見)



今まだ分かっていない(=今データにない)  
ことを発見・理解したい



過去の延長線上で未来を統計的に  
予測する

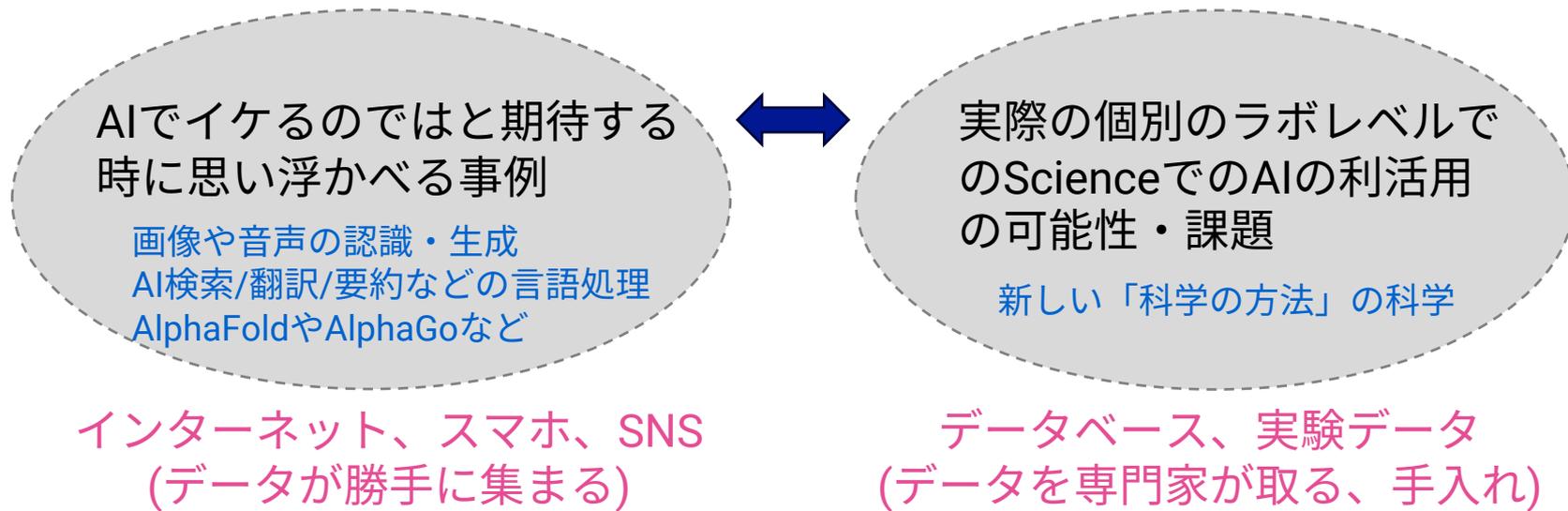
# 今日の話題提供 (機械学習屋の目線から)

お題：「マルチモーダルデータ×AI」と生命科学の未来について議論するパネルディスカッション

1. はじめにインターネットありき問題
2. マルチモーダルは統計学的には多難すぎ問題
3. 予測めっちゃ当たるのに一つも理解は進まない問題
4. AIがReproducibility Crisisの火に油注いでしまう問題
5. みんながAI使うと全員で平凡に向かってしまう問題

**Takeaway**：データ化する現代における「**新しい科学のやり方**」の**真剣な再考**と情報科学の基礎研究(機械学習は**発展途上技術**)が必要

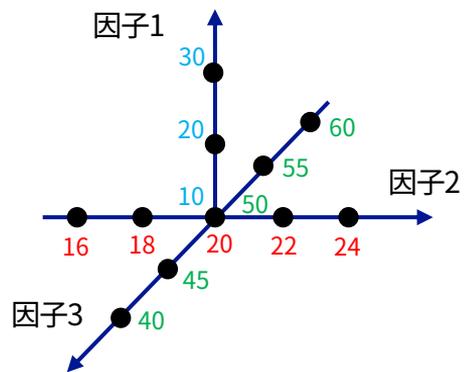
# 1.はじめにインターネットありき問題



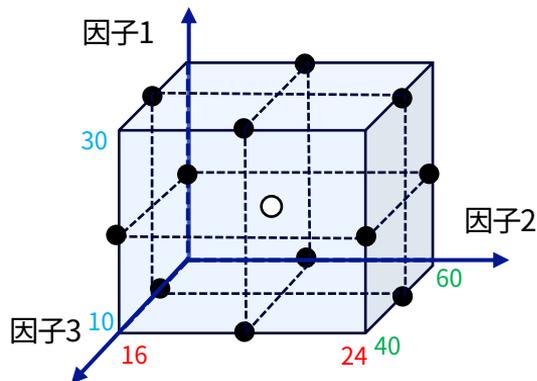
- 左と右には **データの規模・入手容易さ・内容** に大きな齟齬があるため、左の知見を右の状況に使うだけでは**Science利活用では全くworkしない!**
- その大前提の上で**機械学習技術・理論そのものの研究**がまだまだ必須

## 2. マルチモーダルは統計学的には多難すぎ問題

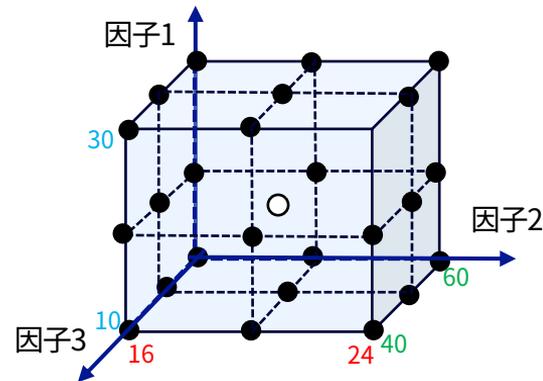
One Factor At a Time (OFAT)



Box Behnken



Full Factorial



- OFATでは交互作用が同定できない。同時に複数因子が変わる測定+複製が必要
- Cross-Modalな効果(交互作用)を考慮したい場合も同様。しかも、**同一対象、同一条件で同時に複数因子が変わる測定+複製が必要** or **RCT**を実施  
→ 実験数も指数的に増大, Bioの場合取れない点が多い (厳密な因果推論×), ...
- 機械学習は今あるデータの確率分布表現で実験計画と違う (相関 vs 因果)

### 3. 予測めっちゃ当たるのに一つも理解は進まない問題

教訓：予測ができることは理解・説明・解釈を直接は意味しない！！

機械学習で高精度に予測ができることは対象の仕組みの理解に直結しない



どうやって見えているものを認識？



どうやって音波から言葉を理解？



“ありがとう”

J'aime la  
musique



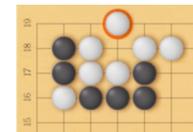
どうやって多言語を理解？



I love music



どうやったら囲碁勝てる？



# 科学的説明・理解とは何かの再考が迫られている

LLMの成功について機械学習屋がびっくりしたこと：

1. 意味など一つも理解せずとも驚愕の実用性があった
2. 「言語」についての理解は一つも深まらなかった

同じ現象を再現するものを「作って」理解しようという構成論的アプローチに疑問を投げかける...

“機械学習モデルを、そのまま科学的モデルとも見なすのは、飛行機を鳥の飛行の「説明」と見なすのと同種の概念的混同です。統計原理を使えば、モデルが対象の現象を捉えていなくても「予測」が可能な場合はあるのです”



Can Artificial Intelligence Unravel the Faculty of Language?

岩波書店

## 4. AIがReproducibility Crisisの火に油注いでしまう問題

機械学習予測の評価は本来激ムズなのに、あまりにお手軽に使えるので、チャラくて怪しい論文が量産され再現性の危機に火に油を注いでしまう...

250 | Nature | Vol 608 | 11 August 2022

News in focus

### IS AI FUELLING A REPRODUCIBILITY CRISIS IN SCIENCE?

'Data leakage' threatens the reliability of machine-learning use across disciplines, researchers warn.

By Elizabeth Gibney

From biomedicine to political sciences, researchers increasingly use machine learning as a tool to make predictions on the basis of patterns in their data. But the claims in many such studies

in Atlanta, whose paper<sup>2</sup> was examined by the pair, says that the field of conflict prediction has been unfairly maligned and that follow-up studies back up his work.

Still, the team's rallying cry has struck a chord. More than 1,200 people signed up to what was initially a planned as a small online

<https://doi.org/10.1038/d41586-022-02035-w>

22 | Nature | Vol 624 | 7 December 2023

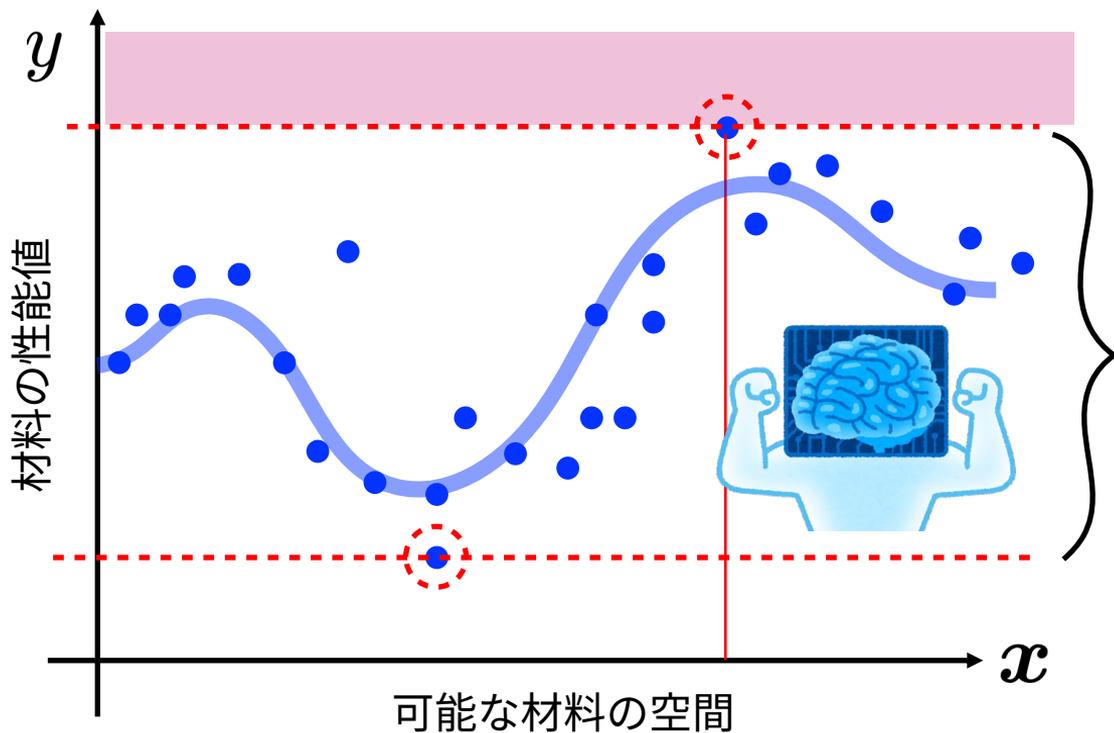
### IS AI LEADING TO A REPRODUCIBILITY CRISIS IN SCIENCE?

Scientists worry that ill-informed use of artificial intelligence is driving a deluge of unreliable or useless research. By Philip Ball

<https://doi.org/10.1038/d41586-023-03817-6>

## 5.みんながAI使うと全員で平凡に向かってしまう問題

機械学習は「平均(期待値)」で代表して予測する



機械学習予測  
(見本例との誤差最小化)

予測値は訓練データのベストより悪く、ワーストより良い  
凡庸な値しか取れない…

## 5.みんながAI使うと全員で平凡に向かってしまう問題

- 機械学習は「平均(期待値)」で代表して予測する  
→「平均」とは「平凡・凡庸」である
- 平均点取れたら悪くはないかもしれないが**良くも無い**
- AI予測は「無数の人たちの平均像」(平凡中の平凡)
- つまり、、、**「科学でAIを使う」と**
- **他人と同じようなことをすることによって、他人より抜きんでようとする勝っても負けても不幸せな競争を産みがちな**ので気をつけて！

# 今日の話題提供 (機械学習屋の目線から)

お題：「マルチモーダルデータ×AI」と生命科学の未来について議論するパネルディスカッション

1. はじめにインターネットありき問題
2. マルチモーダルは統計学的には多難すぎ問題
3. 予測めっちゃ当たるのに一つも理解は進まない問題
4. AIがReproducibility Crisisの火に油注いでしまう問題
5. みんながAI使うと全員で平凡に向かってしまう問題



May the **ML Force** be with you...

**Takeaway**：データ化する現代における「**新しい科学のやり方**」の**真剣な再考**と情報科学の基礎研究(機械学習は**発展途上技術**)が必要