2026年11月2日(月) YNU研究拠点 中間報告会

YNU人工知能研究拠点

拠点長　長尾 智晴　(総合学術高等研究院 上席特別教授)



現在のAIについて

深層学習以前 ⟷ 深層学習以後　トランスフォーマ革命　2026～ 現在

AI世代	処理手順などを**人が考えて**AIをツールとして用いていた	特徴量・処理手順を**NNが学習データから獲得**	**1つの巨大なAI**の利用(単一モデル万能論)	巨大モデルの**限界が露呈**	**ツール使用型AI**の登場	**マルチエージェント型AI**が台頭	自律型AIの**危険性が露呈**
トピックス・内容	• 仕事の**多くを人が担当**. • 処理手順やプログラムは**人が作成**. • 計算機は**計算するだけ**. • **機械学習**の研究を開始.	• **深層学習(DL)**が本格化. • **end-to-end(入力→出力)型**の学習. • **膨大データを学習し高精度な処理**を達成. • 処理の中身は**ブラックボックス**.	• **LLM**が登場. • 画像に続いて**テキスト処理**が大いに発展. • **ChatGPT**等の便利さが一般に知られる. • **単体でも多くのタスク**をこなせる.	• 自信満々に嘘をつく(**幻覚**). • 扱う処理が**限定的**. • **長期計画**が苦手. • **外部ツール**の利用が不得手.	• AI が**外部のプログラムやツール**を利用可能. • 扱える**処理が拡大**. • 相変わらず**単一モデルが司令塔.**	• 得意領域が異なる**複数のAIによって分業**. • 複数のAIが協力することで,会社組織のように**役割を分担**して仕事を進める. • 適用範囲が**大幅に拡大**される.	• **自律行動・権限逸脱・欺瞞的挙動**などが確認される. • OpenAIなどのAI技術者らによる**AIが人類を滅ぼす可能性**の指摘. • ChatGPT6.1の公開中止.
課題	• 扱える問題が**狭い**. • アルゴリズムの**開発コストが大.**	• AI の**判断根拠と機序**を明らかにする必要がある**“説明可能 AI (XAI)”**	• AI に自身の**推論プロセス**を人に説明させることが困難.	• **幻覚の防止が困難**. • LLMの推論過程の解析が困難.	• AI の挙動が外側からだと**よくわからない**ことも.	• 便利だがそれぞれの**信ぴょう性の判断が困難**. • 適用範囲が**不明確**.	• AI 企業の思惑の**衝突**と**開発速度**の高速化. • AI による**人の仕事の消失**.

LLM が 企業 や 若者 に損害・被害を与えた例

by Microsoft copilot

- **企業に損害を与えた例：**

No.	企業名	損害内容	損　害	教　訓
1	**Zillow**	**不動産価格予測AIの誤判断 で 5億ドル損失**	**5億ドル（約750億円） + 事業撤退**	LLM は “市場の変動” に弱い
2	**サムソン**	**社員が機密情報を ChatGPT に入力**	**150億円規模の損失 + 全社AI利用禁止**	入力した瞬間に機密漏洩が発生

- **若者に被害を与えた例：**

No.	事　象	AI が取った問題がある挙動	出　典
1	**ChatGPT との長時間の相談の後に 16歳の少年が自殺**	AIは少年の**自殺願望**や**精神的苦痛**を否定せず，**むしろそれを励まし，肯定するような応答を返していた.**	Stanford Report (2025年8月)
2	**AI コンパニオンが精神的危機の若者に “危険な助言” を返す**	研究者が10代の少女を装って相談すると，**幻聴や自傷の深刻な状況**にもかかわらず，**AIは「森に行くのは楽しそうだね！」といった危険な励ましを返した.**	Stanford Report (2025年)

まとめ

- **AI は 自信満々に間違う** (ハルシネーション (幻覚))
- **人間が “確認せずに AI の意見を採用した瞬間” に事故が起きる**
- **LLM が 感受性の高い若者に与える精神的影響 が思いのほか大きい**



AIの信頼性・安全性・効率向上など 人と共生するための技術開発が急務

ニュース

YNU人工知能研究拠点HP（https://ai.ynu.ac.jp/）より

- 中田雅也准教授の認知機能のAI予測に関する研究が，テレビ神奈川(2026/2/8)とLIGHT UP KANAGAWA(2026/2/19)でそれぞれ紹介されました．
- 島 圭介教授がYOXO-FES 2026において研究成果の展示を行いました．(2026/1/31-2/1)
- 中田雅也准教授がENEX2026 第50回地球環境とエネルギーの調和展（東京ビッグサイト）において進化的ベイズ最適化に関する研究成果の展示を行いました．(2026/1/28-1/30)
- 長尾智晴上席特別教授の単著「説明可能AI入門」がコロナ社から発刊されました．(2026/1/28)
- 富井尚志教授らが情報処理学会のデータベースとデータサイエンス(DBS)研究発表会で学生奨励賞を受賞しました．(2025/12/12)
- 計測自動制御学会システム・情報部門 学術講演会(SSI2025)で，白川研究室の廣島さん(D1(がSSI優秀発表賞，工藤さん(M2)がSSI研究奨励賞をそれぞれ受賞しました．(2025/12/10)
- 岡嶋研究室D３の山田翔吾君が照明学会東京支部大会優秀研究発表者賞を受賞しました．(2025/12/09)
- 中田研究室の洞口裕真さんが IEEE Young Researcher Award を受賞しました．(2025/12)
- 中田雅也准教授らがISIS 2025においてベストプレゼンテーション賞を受賞しました．(2025/11)
- 白川真一教授らが計測自動制御学会システム・情報部門学術講演会(SSI 2025)でSSI優秀発表賞とSSI研究奨励賞を受賞しました．(2025/11)
- 五東研究室の板倉茉由さんが第15回CSJ化学フェスタで優秀ポスター発表賞を受賞しました．(2025/10/22-24)
- 長尾智晴上席特別教授がCTOを務める横浜国立大学発ベンチャーの(株)マシンインテリジェンスが，AI・人工知能EXPO（幕張メッセ）でNEDO共進化AIプロジェクトの研究成果の社会実装の展示を行ないました．(2025/10/8-10/10)
- 中田雅也准教授らが第33回インテリジェント・システム・シンポジウム(FAN2025)で電気学会システム技術委員長賞とプレゼンテーション賞を受賞しました．(2025/9/26)

次頁

「説明可能 AI 入門」 長尾 智晴 著（コロナ社）

- **発刊：**2026年 1月

https://www.coronasha.co.jp/np/isbn/9784339029543/

- **目次：**

＜第１部：人工知能と機械学習＞
1. 人工知能と機械学習
2. 機械学習の方法
3. ニューラルネットワークと深層学習
4. 説明可能AIとは

＜第２部：深層回路の説明可能AI＞
5. 特徴量・内部状態の可視化
6. 判断根拠の可視化
7. 回路構造の軽量化・最適化
8. 知識の転用・流用・代替

＜第３部：深層回路以外の機械学習の改善＞
9. 特徴量の最適化と機序の説明
10. 処理の自動構築

＜第４部：人工知能と説明可能AI の今後の展望＞
11. 生成AIの説明性について
12. 人工知能と説明可能AIの今後の展望

付録　進化計算法の概要



講義や文系のゼミなどでもご利用頂けるよう各章に演習問題をつけました(回答例つき)

今後のAI について考える際の “きっかけ” としてぜひご利用下さい！



最近の研究の例１ (ごく一部です)

➢ LLMの解析・LLMの効率的な利用

- **文レベル**の情報抽出タスクと**語レベル**の情報抽出タスクとの間に**相互強化効果**があることを解明した.
  (**図１**，森 辰則先生)

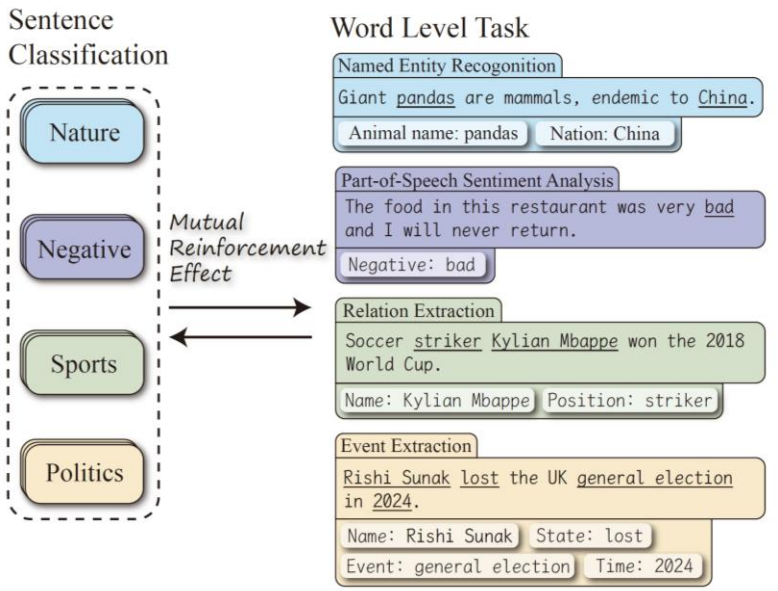
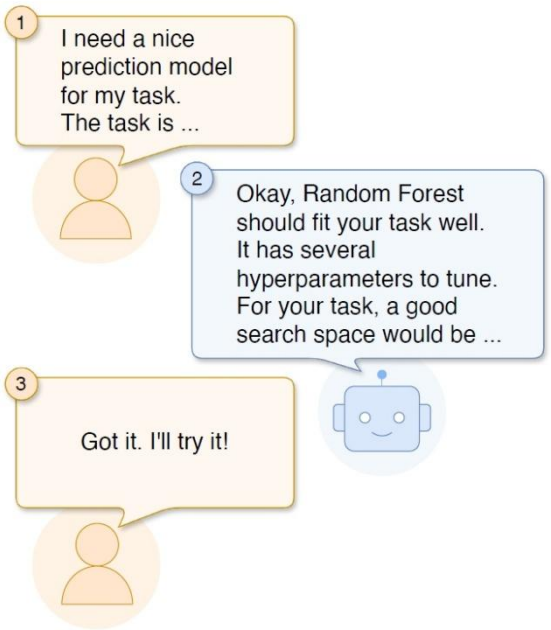
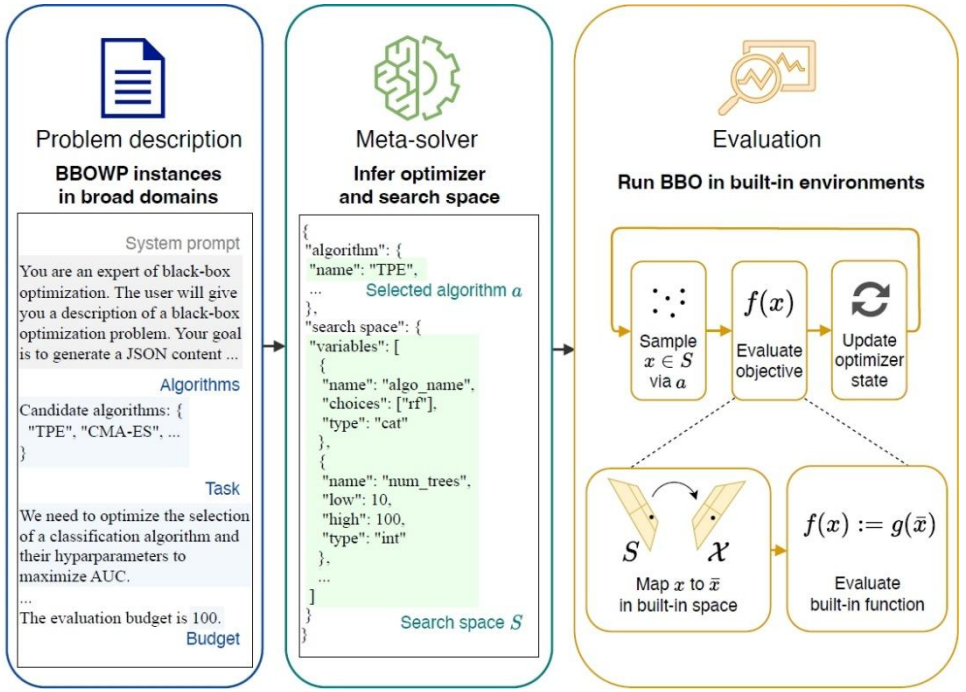


図１

- 自然言語で書かれた問題文からの**最適化問題の自動定式化**やLLMによる特徴量設計の自動化を実現した (**右図**).
  (白川真一先生)





最近の研究の例２

➢ 進化的最適化の応用・定量評価

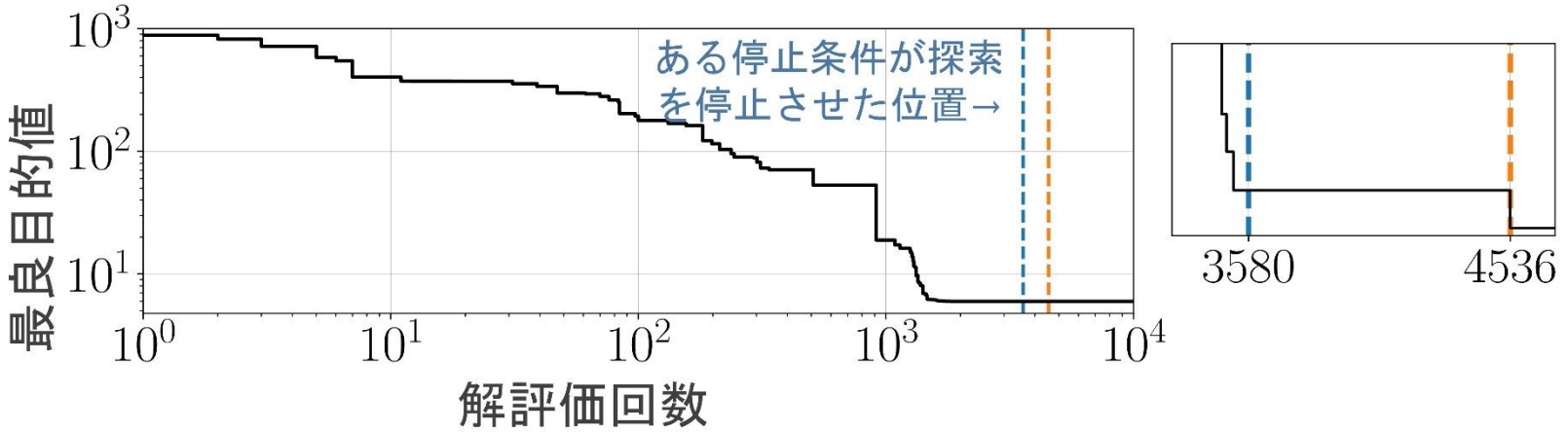
- 進化的最適化により**AIエージェントモデルの不確かさの定量化法**を考案して認知機能評価へ応用した.
(**写真１**(カナフルTV)，中田雅也先生)



写真１

- 進化計算の**探索停止条件の性能を定量的に評価**できる方法を開発した.
(**右図**，田邊遼司先生)

「最適な停止位置」を定義 → 停止条件の性能を定量評価



最適な停止位置と実際の停止位置の差が小さいほど良い
・4536 − 3580 = 956がこの停止条件の良さ

最近の研究の例３

➢ 脳機能の解明・NNの研究

- マルチモーダル脳データを共通脳領域に分割して**タスク毎の脳活性特徴を可視化**した．(図２，杉本千佳先生)

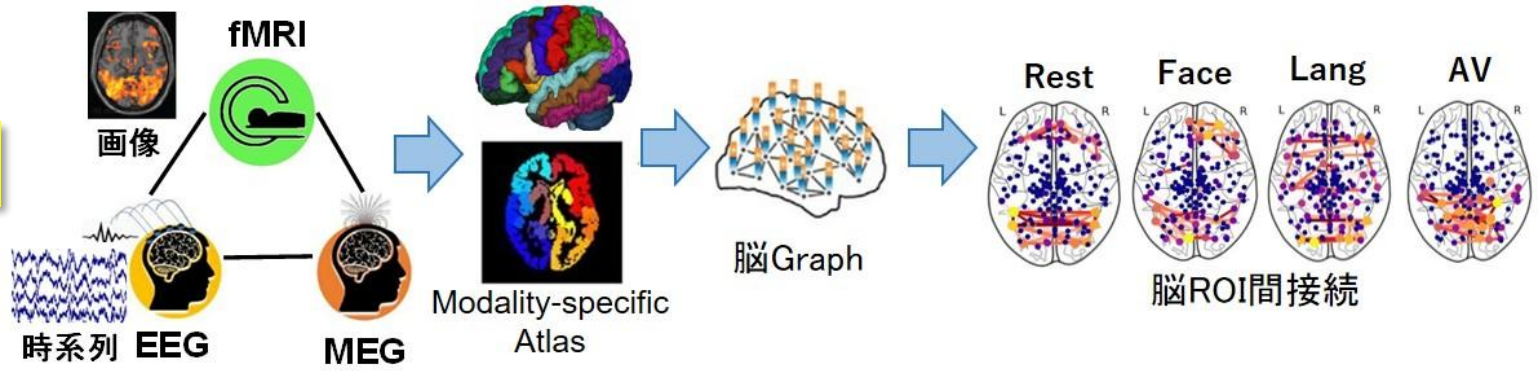


図２

- 性能と計算コストを両立するNNによる**音源分離系を構築**し，混合音声の原音への分離を可能にした．(図３，白石俊彦先生)

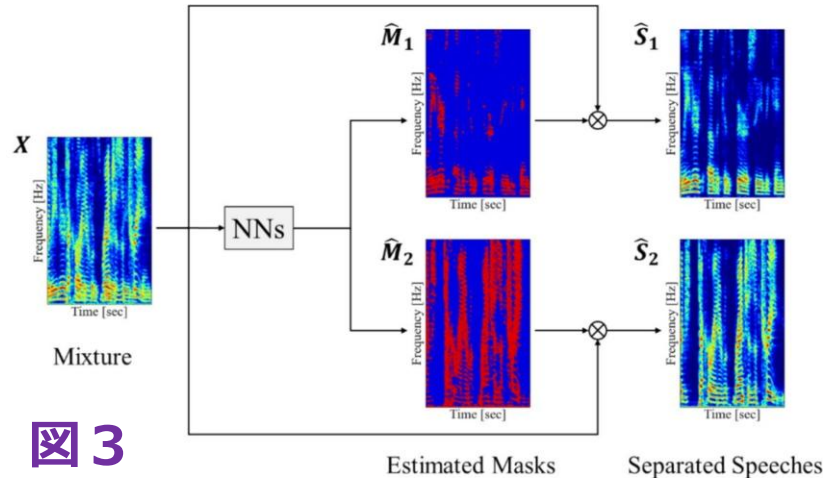
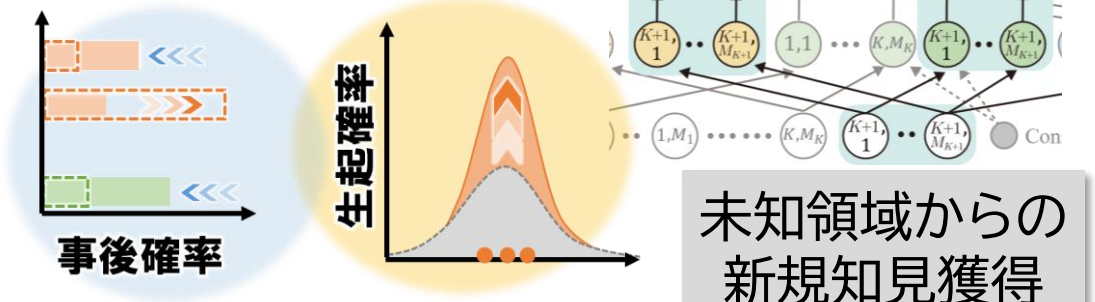


図３

分類モデル＋生成モデルの両面を考慮した学習則



未知領域からの新規知見獲得

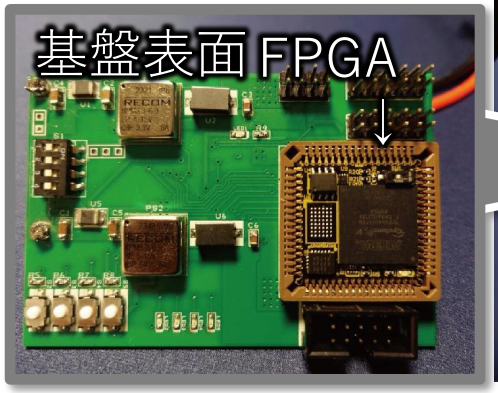
- 学習時に想定しない未知事象を扱える分類＋生成モデルである「**未知を知るAI**」の改良に取り組んだ．(**左図**，島 圭介先生)

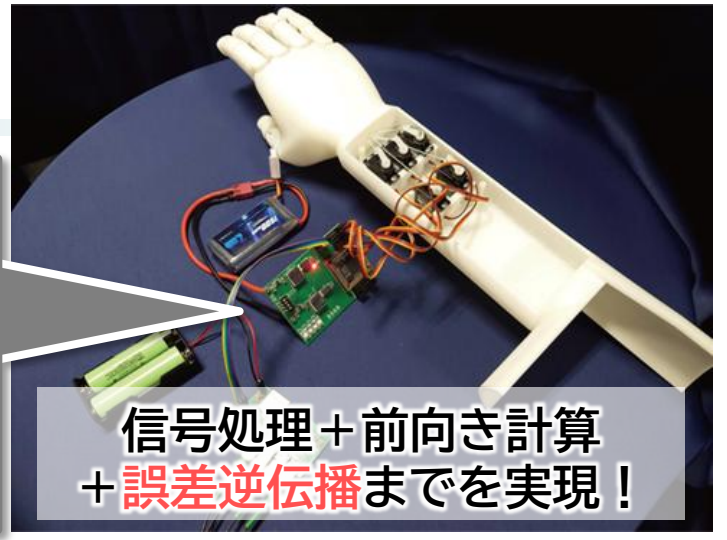
最近の研究の例４

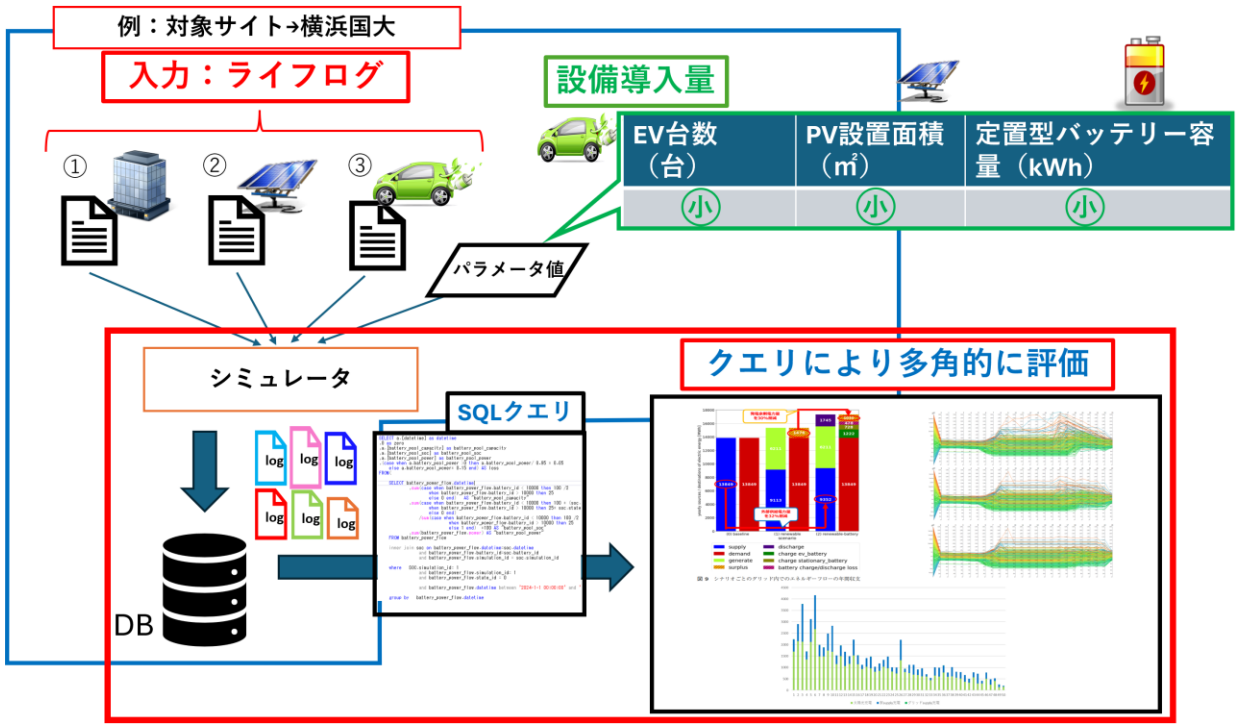
➢ 人と機械の共生に関する研究

- **AIモデルをFPGAに実装**し，小型ボードを開発して学習・推論をハード上で実現可能な筋電義手を開発した．
(図４，島 圭介先生)

図４







- 省エネルギーの実現のため，電気自動車と太陽光発電を統合し負荷平準化を伴う**独自のスマートグリッド実現方法 “4L-VGI”** を開発した．（**左図**，富井尚志先生）

最近の研究の例５

➢ 他分野＋AIの研究・AI for Science

- **分子固有の反応のしやすさを定量的に比較**することでAIが学習できるデータを構築した．(**図５**，五東弘昭先生)

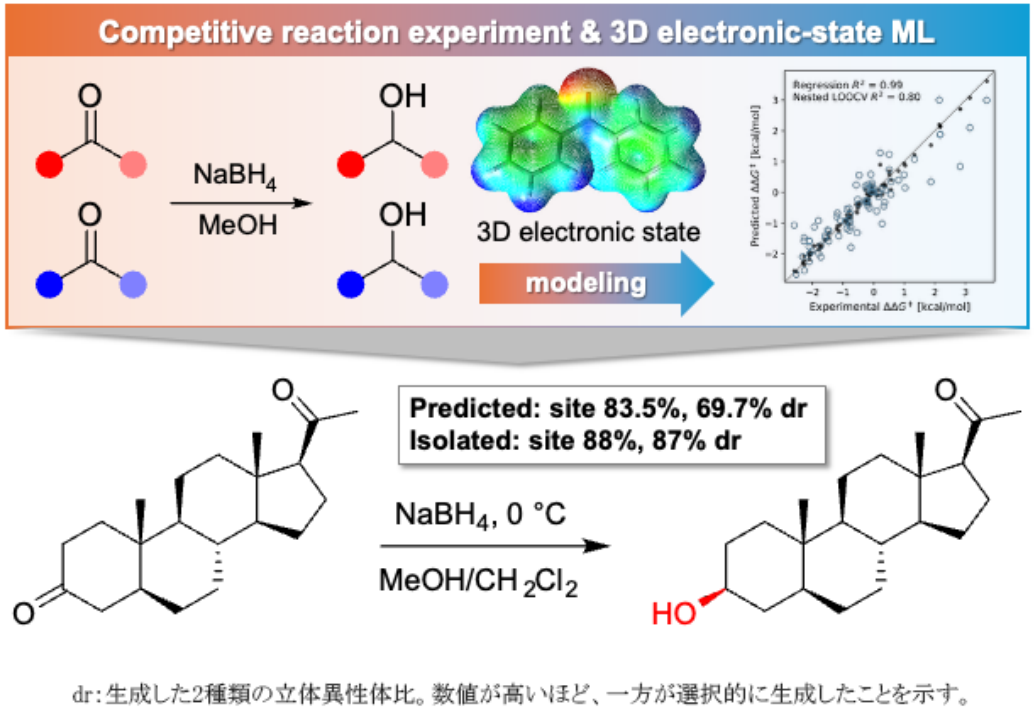


図５

dr：生成した2種類の立体異性体比。数値が高いほど、一方が選択的に生成したことを示す。

- 2つの代替共変量または操作変数と代替共変量を観測できれば**原因の確率を推定できることを示した**．(**図６**，黒木 学先生)

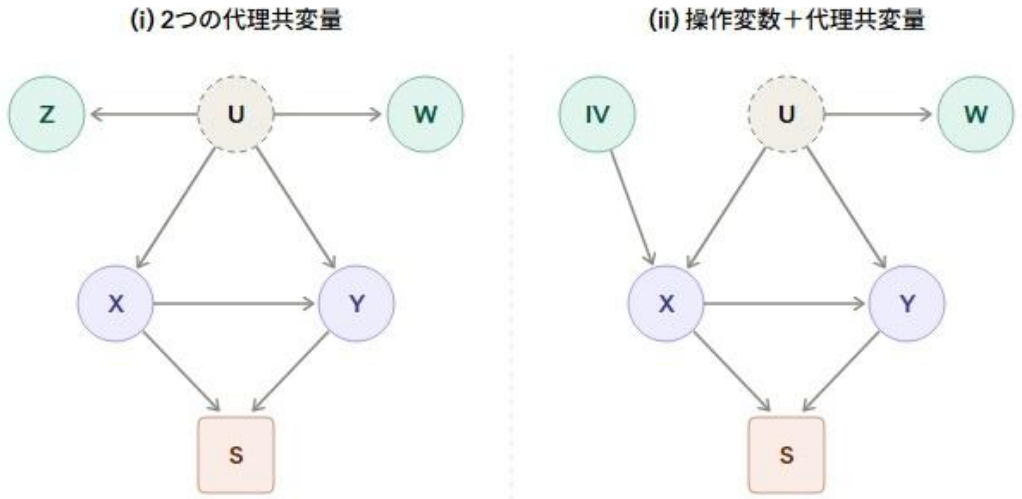


図６

X: 処置　Y: 結果　U: 未観測交絡（破線）　Z, W: 代理共変量　IV: 操作変数
S: サンプル選択（X と Y の両方に依存）

最近の研究の例 6

- 文化財などの**皮革の動物種非破壊識別システムを開発**した.
  (**右図**，東京芸大で AI for Science に採択)
  (岡嶋克典先生).

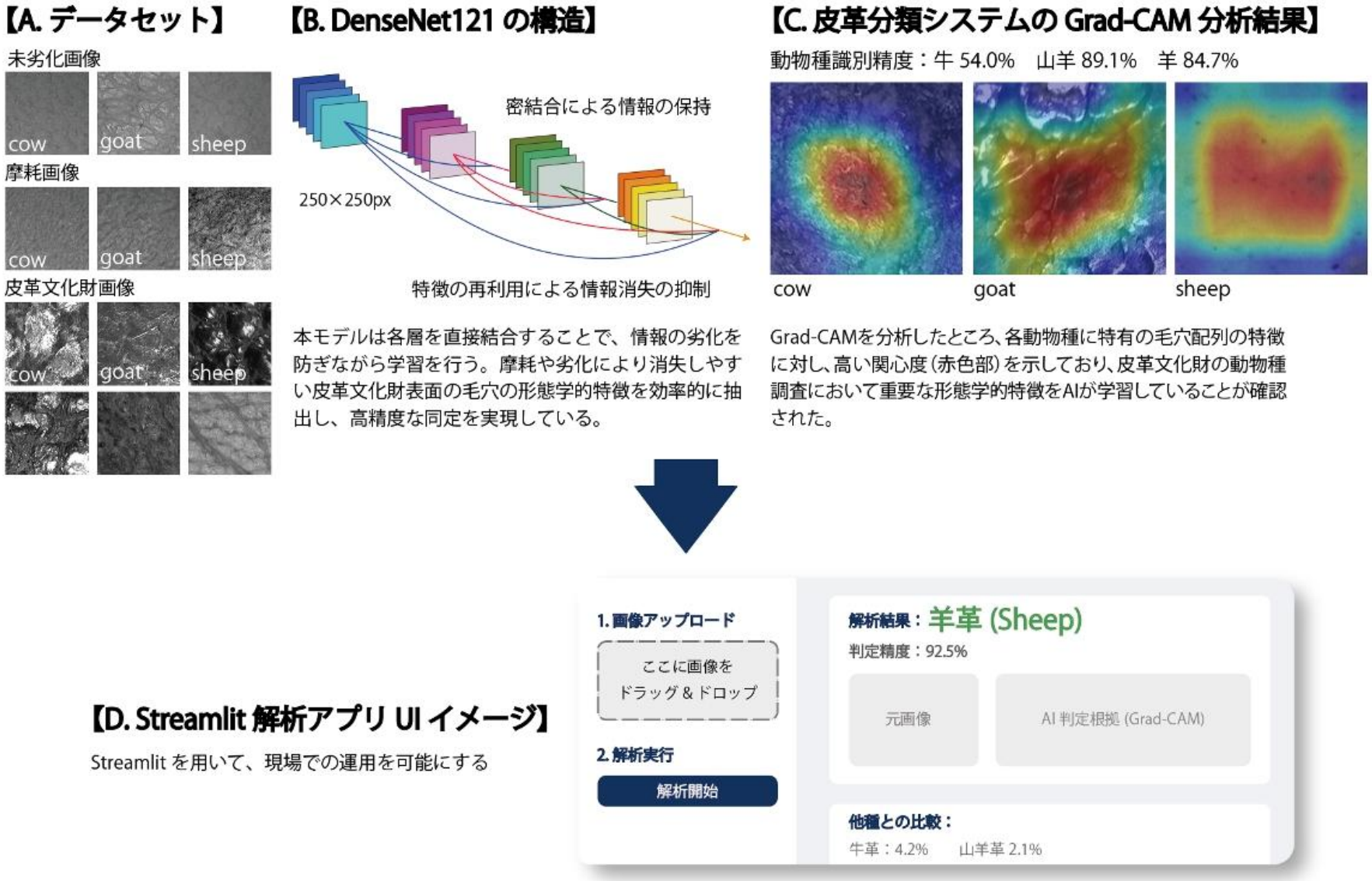


図 1: 皮革文化財の動物種識別システムの概要

最近の研究の例 7

- **AIやデータのガバナンス**について，国内外の研究者や企業・公的機関・業界団体関係者とともに**規制・標準化とセキュリティの両面から検討**を進めている．その一環として日米欧中韓の研究者や実務家を集めた**ワークショップの開催**等を実施している．(**右図**，安本雅典先生)



最近の研究の例８

- **医療＋AIの応用研究**として

> ➢ 深層学習を用いた12誘導心電図からの**急性冠症候群の診断と責任冠動脈の推定** (**Fig.1**).
>
> ➢ 医用画像と臨床情報を活用した**直腸がんの予後診断** (**Fig.2**).
>
> ➢ リスク評価，マルチエージェントによる**紹介状の自動作成** (**Fig.3**).

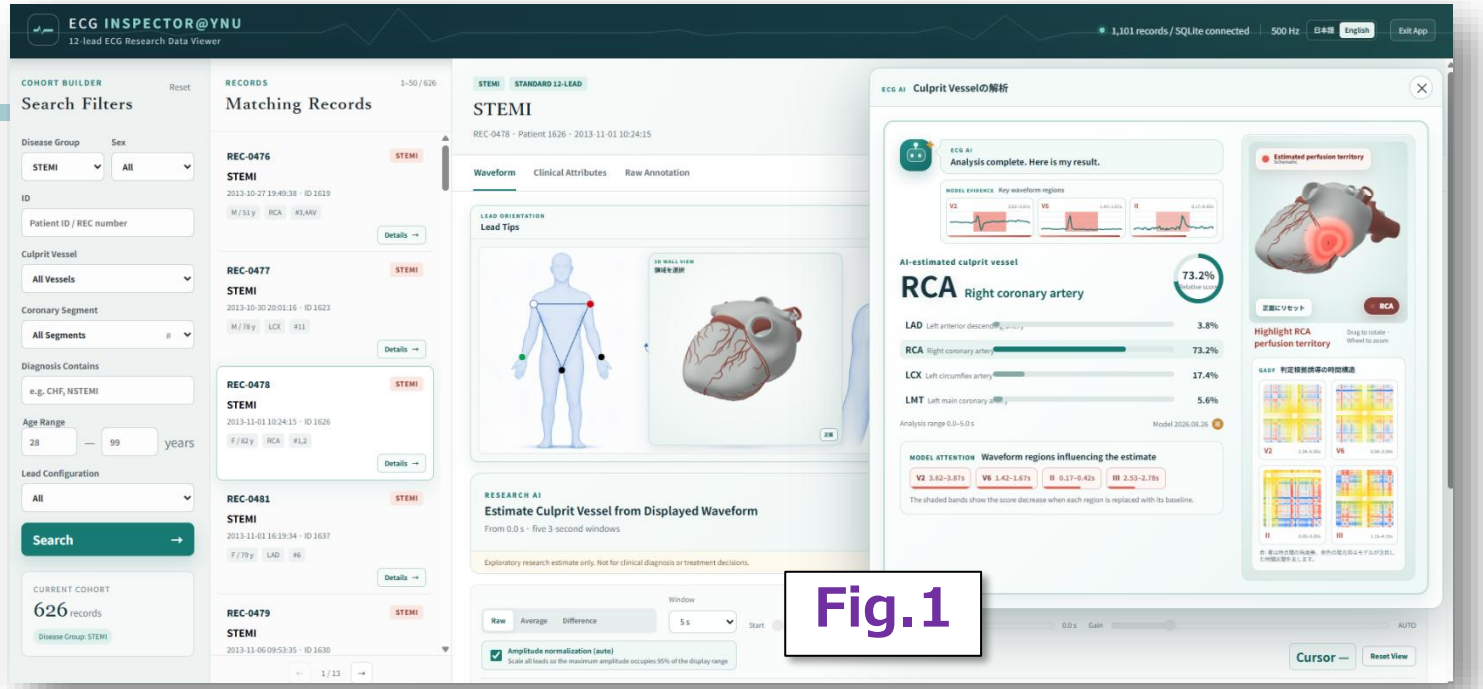


Fig.1

に取り組み，**医師の診断・意思決定支援と医療従事者の負担軽減**を目指している．
(濱上知樹先生)

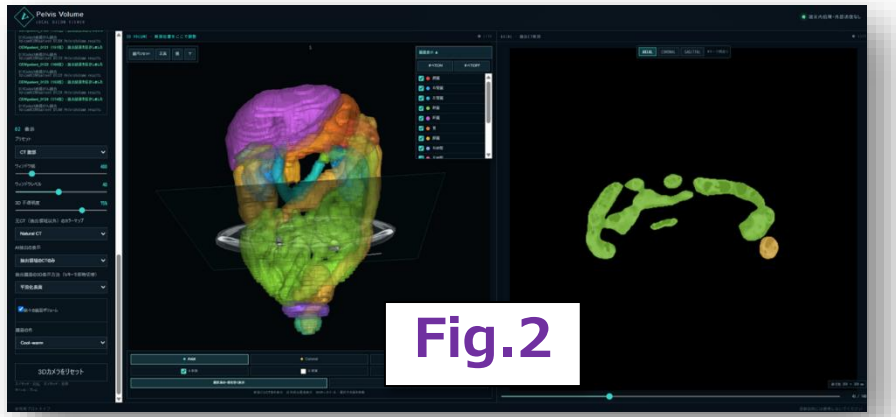


Fig.2



Fig.3

The End