

研究報告

AI2L:学習支援ツールとしての AI 活用指針の定義と応用可能性

新谷正嶺^{1, 2, 3)}¹⁾中部大学生命健康科学部生命医科学科, ²⁾中部大学 AI 数理データサイエンスセンター, ³⁾名古屋大学高等研究院

要 旨

本稿は、2022 年 3 月 28 日に開催された第 3 回連携セミナー「AI と生命システム」で発表した経験を踏まえ、従来の応用指向型 AI (ブラックボックス的利用) に対抗する新たな概念「AI2L (AI to Learn)」を提唱する。AI2L は、AI を単なる効率化ツールとしてではなく、ユーザ自身の思考力や学習力を高めるための支援ツールとして活用するという哲学に基づく。具体的には、研磨不足と研磨十分のチタン板の画像分類実験において GradCAM を用い、酸化チタンの黒い微小面積領域の有無を可視化する手法や、テーブルデータに対してシンボリック回帰型 AI を適用し、温めた心筋細胞におけるサルコメア振動 (Hyperthermal Sarcomeric Oscillations / HSOs) の法則性を発見する事例を紹介する。これらの経験から、成果そのものに AI の痕跡を残さない、すなわち「学びを促進するための AI 利用法」の必要性和その応用可能性を議論する。

キーワード: AI2L、生成 AI、GradCAM、シンボリック回帰、HSOs、ホメオカオス、学習支援

1. はじめに

近年、生成AIの急速な進歩により、ChatGPTなどの大規模言語モデルが広く利用されるようになった。しかし、多くの現状の AI 活用は、望ましい出力を得ることに重点を置き、内部の判断根拠や解釈性が十分に考慮されていない(例えば、CNN (Convolutional Neural Network) を用いた画像分類など) [1-3]。こうした「応用指向」モデルは、効率性は高いがブラックボックス性が否めず、利用者がその結果を無批判に受け入れるリスクがある [4-11]。本稿では、AI を単なる効率化ツールとして受け入れるのではなく、「学習支援ツール」として主体的に活用することの重要性を説く [12]。

そこで本研究は、2022 年 3 月 28 日に開催された連携セミナー「AI と生命システム」で私が提唱した「AI to Learn (AIL)」という概念を出発点とし、後に AI2L と改称した経緯および具体的事例を通じ、AI2L の定義とその実践的応用可能性を議論する。

2. AI2L の概念とその定義

2. 1 応用指向 AI の問題点

従来の AI 利用は、入力に対して望ましい出力を得ることに注力するが、その内部プロセスは不透明であり、ブラックボックス性を放置している [4-8]。例えば、画像分類において、専門家は研磨状態の微妙な違いを判断できるが、一般利用者は AI の出力結果に依存しがちである。実際、私自身も研磨不足と十分な研磨状態のチタン板の違いを判別できなかったが、GradCAM (Gradient-weighted Class Activation Mapping)を活用することで、酸化チタンの黒い微小面積領域が判断の鍵であることが明確になった (図 1) [12]。

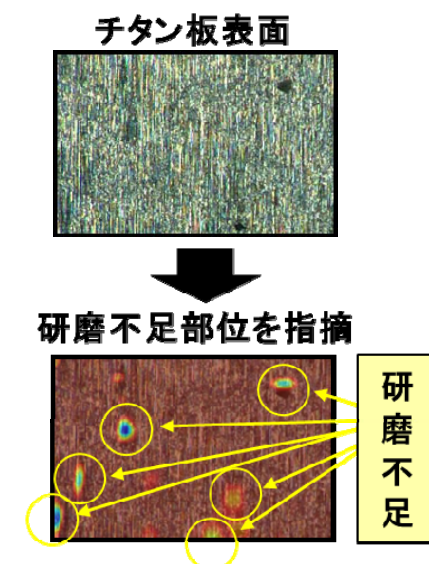


図1. 研磨不足部位の学習支援

2. AI2L の定義

AI2L (AI to Learn / AI2L) は、以下のように定義される。

1. **学習支援ツールとしての AI 利用:** 最終成果において AI の介在を排除し、あくまでユーザ自身の理解と成長を促すための補助ツールとする。
2. **透明性と説明可能性の追求:** AI の出力に対し、利用者がその根拠を検証可能なプロセスを必ず組み込む。
3. **主体的な活用による自己拡張:** AI の便利さに依存するのではなく、AI を用いて自らの思考や解析能力を向上させることを目指す。

この考え方は、単に「楽をするための AI 活用」ではなく、人間自身が AI との対話を通じて深い洞察とクリティカルシンキングを養うためのものである [12]。

2.3 命名経緯と改称

もともと「AI to Learn (AIL)」として提唱されたが、英語圏におけるネガティブなニュアンス (同スペルの略語が持つ別の意味など) を避けるため、「to」を数字の「2」に変更し、AI2L と略称することになった。この改称は、概念自体の内容を変えるものではなく、グローバルな展開を見据えたブランディング上の工夫である。

3. AI2L の実践事例

3.1 画像データによるチタン板の研磨状態判別

実験では、研磨不足と十分な研磨状態のチタン板の画像を撮影し、GradCAM という手法を用いて、元画像の一部を黒く塗りつぶすなどの工夫により、原因領域をシャープに可視化した [12, 13]。その結果、黒い微小面積領域の有無が研磨状態の判断ポイントであることが明確となり、専門の先生が直感的に判断できる基準を自分自身で得ることができた。これにより、精度を高めた画像分類 AI を利用しなくとも、状況判断が可能となった。

3.2 テーブルデータに対するシンボリック回帰

また、実験データ (テーブル形式) を入力に、シンボリック回帰型 AI を用いてデータ間の関係を適合率付きで数式として示す手法を構築した [14]。この数式がヒントと

なり、発表者が発見した、温めた心筋細胞におけるサルコメア振動 (HSOs) の解析において、カオス的な不安定性を示しながらも周期一定の「ホメオカオス」と呼べる性質を発見することができた (図 2) [15-19]。この経験は、AI の出力を単なる結果として受け入れるのではなく、そのプロセスから学び、自らの研究成果へと結実させる好例である。

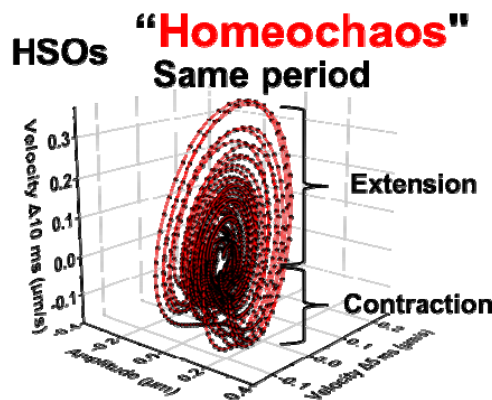


図2. 温めた心筋細胞のサルコメアで顕在化する熱筋節振動 (HSOs) の Homeochaos 性

4. AI2L の意義と応用可能性

4.1 教育分野での応用

AI2L の考え方は、学生が AI に依存せず自らの学びを深めるための支援ツールとして有用である。たとえば、授業で ChatGPT などの生成 AI を利用して解答例を得た後、必ずその理由や根拠を自分で検証するプロセスを組み込むことで、批判的思考や創造的問題解決力を養うことが可能となる。また、生成 AI の出力と自らの分析との乖離を議論することにより、AI の限界と可能性の双方を体感できる。

4.2 研究分野での応用

研究においては、膨大なデータ解析やシミュレーションに AI を活用するが、最終的な解釈や発見は研究者自身が行うべきである。AI2L の手法は、AI の解析結果を基にその根拠を解明し、再現性のある数理モデルを構築するための枠組みとして有用である。実際、シンボリック回帰による数式抽出の事例は、研究現場での新たな知見の発見に

大きく貢献する可能性を示している[18]。

4.3 医療分野での応用

医療の現場では、診断支援 AI が急速に普及しているが、AI の判断理由が不透明な場合、医師がその結果を説明できず倫理的問題が生じる。AI2L の理念に基づけば、AI はあくまで補助ツールとして利用され、最終判断は必ず医師が根拠を理解した上で行うべきである。このアプローチは、患者に対する説明責任を果たし、診断の安全性を高めるためにも極めて重要である。

4.4 他分野への波及効果

さらに、法律や経済分野においても、AI の判断を単純に受け入れるのではなく、その根拠を検証する姿勢は求められる。例えば、契約書ドラフトや判例検索においても、AI2L の考え方により、AI が提示する情報の裏付けを必ず自らの判断で確認することで、より精度の高い意思決定が可能となる。

5. 専門家の見解と先行研究

複数の専門家は、ブラックボックス AI の問題点と説明可能な AI の必要性を指摘している。Cynthia Rudin は高リスク分野では説明可能なモデルの使用を強く推奨しており[20]、また、Fei-Fei Li は AI を人間の能力拡張のためのツールとして活用すべきだと述べている[21]。さらに、Marcus は、AI を単に自動化ツールとして受け入れるのではなく、人間との協働によってその限界を補完すべきだと主張している[22]。これらの見解は、AI2L が掲げる「人間が AI の出力を解釈し学習する」という基本理念と整合性を持つ。また、GradCAM を用いた画像解析やシンボリック回帰を活用した数理モデル構築の先行事例は、すでに多くの分野で報告されており[12]、それらの技術を基盤とした AI2L の応用は十分に現実的である。

6. 結論

本稿では、AI2L (AI to Learn/AI2L) という概念を、従来の応用指向 AI の受容に対する反論として位置付け、学習支援ツールとしての AI 活用法の有効性を議論した。実践例として、研磨状態の判別やシンボリック回帰による数式抽出を通して、AI が単なる自動化ツールではなく、人間の思考や発見のプロセスを補助することを実証した。これにより、AI2L は教育、研究、医療、さらには法律・経済といった他分野においても、主体的かつ批判的な AI 利用のための有力な指針となると考えられる。今後、各現場での実践とフィードバックを通じて、AI2L の指針はさらに洗練され、未来においても人間の知性を拡張する強力なパートナーとなるであろう。

参考文献

- [1] Zednik, C. Solving the black box problem: a normative framework for explainable artificial intelligence. *arXiv preprint arXiv:1903.04361* (2019).
- [2] Paçacı, G., Johnson, D., McKeever, S. & Hamfelt, A. "Why did you do that?": explaining black box models with inductive synthesis. *arXiv preprint arXiv:1904.09273* (2019).
- [3] Kalmykov, V. L. & Kalmykov, L. V. XXAI: towards explicitly explainable artificial intelligence. *arXiv preprint arXiv:2401.03093* (2024).
- [4] Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D. & Batra, D. (2017). Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization. *Proceedings of the IEEE International Conference on Computer Vision (ICCV)*, 618–626.
- [5] Zhou, B., Khosla, A., Lapedriza, A., Oliva, A. & Torralba, A. (2016). Learning Deep Features for Discriminative Localization. *Proceedings of the*

IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2921–2929.

[6] Simonyan, K., Vedaldi, A. & Zisserman, A. (2014). Deep Inside Convolutional Networks: Visualising Image Classification Models and Saliency Maps. *arXiv preprint arXiv:1312.6034*.

[7] Sundararajan, M., Taly, A. & Yan, Q. (2017). Axiomatic Attribution for Deep Networks. *Proceedings of the 34th International Conference on Machine Learning (ICML)*, 3319–3328.

[8] Springenberg, J.T., Dosovitskiy, A., Brox, T. & Riedmiller, M. (2015). Striving for Simplicity: The All Convolutional Net. *arXiv preprint arXiv:1412.6806*.

[9] Doshi-Velez, F. & Kim, B. Towards a rigorous science of interpretable machine learning. *arXiv preprint arXiv:1702.08608* (2017).

[10] Lipton, Z. C. The mythos of model interpretability. *Communications of the ACM* **61**, 36–43 (2018).

[11] Rudin, C. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence* **1**, 206–215 (2019).

[12] Seine, A. Shintani. Artificial intelligence to learn (AIL): How to use AI without having to deal with AI-derived black boxes. *Chubu University Institute of Life and Health Sciences Bulletin* **19**, 18-22 (2022).

[13] elvaraju, R. R., Cogswell, M., Das, A. et al. Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization. *Proc. IEEE ICCV* 618–626 (2017).

[14] Schmidt, M. & Lipson, H. Distilling Free-Form Natural Laws from Experimental Data. *Science* **324**, 81–85 (2009).

[15] Kaneko, K. & Ikegami, T. Homeochaos: Dynamics of a Symbiotic Network with Population

Dynamics and Evolving Mutation Rates. *Physica D* **56**, 406–429 (1992).

[16] Shintani, S. A., Oyama, K., Fukuda, N. & Ishiwata, S. High-frequency sarcomeric auto-oscillations induced by heating in living neonatal cardiomyocytes of the rat. *Biochem. Biophys. Res. Commun.* **457**, 165–170 (2015).

[17] Shintani, S. A., Washio, T. & Higuchi, H. Mechanism of contraction rhythm homeostasis for hyperthermal sarcomeric oscillations of neonatal cardiomyocytes. *Sci. Rep.* **10**, 20468 (2020).

[18] Shintani, S. A. Hyperthermal sarcomeric oscillations generated in warmed cardiomyocytes control amplitudes with chaotic properties while keeping cycles constant. *Biochem. Biophys. Res. Commun.* **611**, 8–13 (2022).

[19] Shintani, S. A. Observation of sarcomere chaos induced by changes in calcium concentration in cardiomyocytes. *Biophys. Physicobiol.* **21**, e210006 (2024).

[20] Rudin, C. Stop explaining black box models for high stakes decisions and use interpretable models instead. *Nature Mach. Intell.* **1**, 206–215 (2019).

[21] Wang, D., Yang, Q., Abdul, A. & Lim, B. Y. Designing theory-driven user-centric explainable AI. *Proc. CHI Conference on Human Factors in Computing Systems*, 1–10 (2019).

[22] Dellermann, K., Calma, A., et al. Fostering critical thinking in the age of generative AI in higher education. *Proc. 52nd HICSS*, 1–10 (2019).

Title: AI2L: Definition and Application Potential of AI Utilization Guidelines as a Learning Support Tool

Author(s): Seine A. Shintani^{1, 2, 3)}

Address(es): 1) Department of Biomedical Sciences, College of Life and Health Sciences,

Chubu University, Kasugai, Aichi, 487-8501, Japan

2) Center for Mathematical Science and Artificial Intelligence, Chubu University, Kasugai, Aichi, 487-8501, Japan

3) Institute for Advanced Research, Nagoya University, Furo-cho, Chikusa-ku, Nagoya-shi, 464-8601, Japan

Keywords: AI2L, Generative AI, Grad-CAM, Symbolic Regression, HSOs, Homeochaos, Learning Support

Abstract: Based on experiences presented at the 3rd Joint Seminar "AI and Biological Systems," held on March 28, 2022, this paper proposes a novel concept termed "AI2L (AI to Learn)" as an alternative to conventional application-oriented (black-box) AI approaches. The AI2L philosophy emphasizes utilizing AI not merely as an efficiency-enhancing tool but rather as a support mechanism aimed at enhancing users' critical thinking and learning capabilities. Specifically, we introduce practical examples, including the use of Grad-CAM in an image classification experiment to visualize the presence or absence of microscopic black oxidized titanium regions in inadequately versus adequately polished titanium plates, as well as applying symbolic regression-based AI to tabular data for uncovering underlying principles of hyperthermal sarcomeric oscillations (HSOs) in heated cardiomyocytes. From these experiences, we discuss the necessity and applicability of using AI in ways that do not leave explicit traces of AI intervention in the final results—namely, employing AI methods designed to foster human learning.