

## ナイスステップな研究者 2023 講演会

講演会「近未来への招待状～ナイスステップな研究者  
2023 からのメッセージ～」第3回

科学技術・学術政策研究所 企画課

## 【概要】

科学技術・学術政策研究所（NISTEP）では毎年、科学技術イノベーションの様々な分野において活躍され、日本に元気を与えてくれる方々を「ナイスステップな研究者」として選定している。選定された10名のうち、3名の方々に第3回講演会として2024年7月23日（火）に自らの研究活動やキャリアパスについて御講演いただいた。本記事では、それぞれの御講演内容に沿って要旨とポイントを記載している。

キーワード：ナイスステップな研究者

講演タイトル：視覚と言葉を繋ぐAIのこれまでと  
これから

オムロンサイネックエクス株式会社  
Vice President for Research 牛久 祥孝 氏  
(牛久氏提供)

2014年 東京大学 大学院情報理工学系研究科 知能機械情報学専攻博士課程 修了 博士（情報理工学）  
2014年 日本電信電話株式会社  
コミュニケーション科学基礎研究所 研究員  
2016年 東京大学 大学院情報理工学系研究科  
知能機械情報学専攻 講師  
2018年 オムロンサイネックエクス株式会社  
Principal Investigator  
2019年 株式会社 Ridge-i Chief Research Officer（兼務）  
2024年 オムロンサイネックエクス株式会社  
Vice President for Research

## 講演要旨：

2010年代から第3次AIブームが起き、人々の生活の裏でもいろいろなAIが活躍するようになって

きた。その中で、人の視覚（画像や動画）と言葉（テキスト）を結び付けて理解したり生成したりするAIの研究と応用も進んでいる。本講演では、こうした視覚と言葉をつなぐAIであるVision and Languageと呼ばれる研究分野について話す。AIとは何か、特に最近話題の生成AIとは何か、それまでのAIとは違うものなのかどうかということについて簡単に触れつつ、視覚と言葉をつなぐVision and Languageの代表例を紹介する。また最後に、こうしたVision and Languageの応用として、観測されるデータ（実験データ）とテキスト（論文などの科学的知識）を結び付けることによってAIロボット自体が研究を加速する、AIロボット駆動科学の取組についても言及する。

## 講演のポイント：

AI研究において3つの非常にセンセーショナルな事件が起きた年を紹介します。2011年には音声認識技術の精度が劇的に向上し、エラー率が10ポイントも改善しました。2012年には画像認識技術が飛躍し、認識精度が大幅に向上。2014年にはシンプルなモデルでも高精度な翻訳が可能となり、機械翻訳が進化しました。これらはすべて深層学習、いわゆるディープラーニング技術によるもので、第3次AIブームを引き起こしました。その後、2022年以降には生成AIが普及し、ChatGPTやStable Diffusionといった技術が社会的に認知を集めました。これにより、第4次AIブームになったと言う方もいます。

こういった音声・画像・テキストを理解する「理解の AI」というものが進化してきたのですが、単純化すれば 2 つのタスクで回っています。1 つは「識別」で、例えば脳の画像からどこが海馬かを判定するのであれば、画素ごとにこれは海馬だな、これは海馬でないかと色塗りをするように、離散値での推定を行うもの。もう 1 つは「回帰」で、例えば脳の画像から体積がどれくらいかを見積もるなど、連続的な値のどこら辺かを推定するものです。

深層学習が発見される前は、画像認識、機械翻訳などでいろいろ異なる技術を順次適用して結果を得ており、それぞれの研究者間で何をやっているのか分からないというような状況でした。しかし、深層学習というのは人間の脳細胞神経にインスパイアされた技術で、画像認識でも機械翻訳でも使う技術がほぼ同じになります。これが深層学習によって起きた一つの大きな変化です。そして同時期に起きたもう一つの変化が、大量の動画やテキストのペアが収集できるようになったことで、Vision and Language という画像とテキストを融合するような研究が増え始めたことでした。その例を 3 つ紹介します。

1 つ目の「ビジュアル質問応答」では、画像を入力し、質問を与えると AI が適切な回答を生成します。2 つ目は、画像が入ってきたときに、その内容をテキストで説明する「画像キャプション生成」です。3 つ目が、逆にテキストから画像が生成できるのでは、ということで、「キャプションからの画像生成」です。例えば「ハリネズミの魔法使いが車を運転している絵」といったユニークなリクエストに応じて AI が画像を生成することができます（図表 1）。

最近の ChatGPT といったアプリは、Vision and Language のいろいろなことができるようになったものだ、ということもできます。一見新しい技術のようですが、その根本は従来の AI と同じです。例えば画像キャプション生成では、数万のボキャブラリから画像にふさわしい 1 単語目を推定し、更に 2 単語目を推定していきます。これは正に「識別」です。逆に、キャプションからの画像生成では、各画素の RGB はどういう値かを「回帰」します。

最後に、AI ロボット駆動科学の話です。科学はいろいろな形態に進化してきており、今は第 4 世代と呼ばれています。この金属を燃やすとこの色になるぞといった経験的な科学、数式的にそういう現象を説明しようという方程式の科学、計算機を使ってのシミュレーションといった演繹的にデータを扱う科学、そして帰納的にデータを使っていくデータドリブン科学と進化していったのですが、このデータドリブン科学が更に AI によって加速し、AI ドリブン科学みたいなものが出てきています（図表 2）。

例えば、大量のデータから科学法則を発見するか、実験のような複雑な作業をロボットにやらせるとか。そういうところに需要が生まれてきていて、私自身も科学技術振興機構（JST）からの委託研究で、AI ロボットにノーベル賞を受賞させるような AI ロボットサイエンティストを作ろうという取組を行っています。

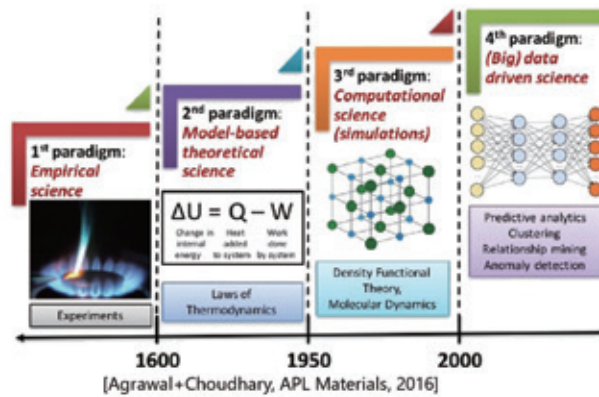
今後も、視覚や言葉、いろいろなデータを超えてつないでいながら、より人間の創造的な活動をサポートできるような研究を進めていきたいと考えております。

図表 1 キャプションからの画像生成



図表 2 深層学習がデータドリブン科学を加速している

## データドリブン科学の萌芽から20年



深層学習がデータドリブン科学を加速しているのが今！

## 講演タイトル：古代の智慧×最先端テクノロジーで イノベーションを創発



京都大学 人と社会の未来研究院 教授  
熊谷 誠慈 氏（熊谷氏提供）

2009 年 京都大学大学院文学研究科・博士課程修了 博士（文学）  
2009 年 日本学術振興会特別研究員 PD  
2011 年 京都大学 白眉センター 特定助教  
2012 年 京都大学 こころの未来研究センター 特任准教授  
2018 年 ウィーン大学 文献文化学部 特任教授（兼任）  
2021 年 内閣府ムーンショット型研究開発制度（目標 9）  
プログラムディレクター（PD）  
2022 年 株式会社テラバース 共同創業者  
2023 年 京都大学 人と社会の未来研究院 教授

### 講演要旨：

私は、インドやチベット等の古代文書の解読を通じて伝統知（伝統的な知識や知恵）を回収し、それらとテクノロジーを融合した「伝統知テック」の開発を進めるなど、「産・宗・学連携」による「総合知」を用いた研究開発を推進してきた。本講演では、その開発事例を紹介する。

私の研究開発グループは、2021 年 3 月、人工知能に最古の仏教経典『スッタニパータ』を機械学習させた仏教対話 AI「ブッダボット」を開発、公表した。その 2 年後の 2023 年 7 月には、上記対話 AI に ChatGPT の最新版（公開当初は ChatGPT4）を統合した「ブッダボットプラス」を開発し、より具体的かつ詳細な回答や解説を可能とした。同 9 月には、「親鸞ボット」と「世親ボット」を開発し、ボットの複数化・多様化を実現した。

また、テキスト対話だけにとどまらず、AR（拡張現実技術）を用いて、「テラ・プラットフォーム AR Ver1.0」を開発し、視覚・聴覚・触覚を用いた多感覚伝統知コミュニケーションを実現した。

これまで、AI に仏教のデータのみを学習させてきたが、今後、さらに仏教以外のデータを学習させていくことで、多数のサイバー・フィジカル融合世界の構築が可能となる。最終的には「一兆の宇宙」（テラバース）の構築を目指し、開発を推進している。

これらの総合知や技術を応用することで、新たな思想、文化、芸術などの創出可能性がある。また、これらを産業界にも応用し、実業家の古屋俊和氏と株式会社テラバースを共同創業するなど、「伝統知テック産業」の開拓を進めている。

### 講演のポイント：

私は仏教学を専門とする研究者で、古文書を解析し、サンスクリット語や古典チベット語などを用いて仏教の思想を解明するといったことを主な仕事としています。歴史書を基に遺跡を探すこともあり、例えばブータン仏教の開祖ツァンパギャレーの生誕地を世界で初めて発見しました。しかしながら、当時の私の研究は役に立たないと言われることも多く、また、文書に書かれている文字情報では机上の空論になってしまうのではないかと悩んでいました。

そんな中、2011 年のブータン国王の来日が大きな注目を集めました。すると、瞬く間に私の「役に立たない研究」が、「すばらしい研究」だと言われるようになりました。研究の社会的評価というのは絶対的でなく、時流に左右されやすいものだ気がしました。

そして、ブータンの仏教哲学を基に、新しい幸福哲学を展開していきました。行政の方々は興味を持ってくれるのですが、実際に政策に反映されて国民に利益として届くのは極めて難しく、行動するにしても何をやって良いか分からないといった悩みの中過ぎておりましたが、ある日転機が訪れます。

仏教界からの要請で、衰退気味の仏教を復興できないかということで、議論する機会があり、その中で AI と仏教を組み合わせるアイデアが生まれました。そしてできあがったのが、非生成系の AI 仏教チャットボット「ブッダボット」です。その後、高性能な生成系 AI である ChatGPT3.5 が登場したことを受けて、非生成系の旧式ブッダボットと生成系の ChatGPT4 を融合した生成系仏教対話 AI「ブッダボットプラス」も公表しました。

例えば、旧式ブッダボットが「SNS を始めるべきか」といった質問に対して、経典から近いと思われる情報を出力します。「何ものにも執着せず、余計なものを持たない人が、素晴らしい人である」と出ますが、これだけではピンと来ない。それに対して ChatGPT4 が回答を補足します。「SNS を始めるかどうかは個々人の自由ですが～」と（図表 3）。

こうした生成系 AI を統合したことで非常に性能は上がったのですが、むしろ不安を持ったという方々もいました。ゴードマ・シッダールタという一人の思想を AI でばらまいてしまうと、画一的な思想

統制がなされてしまうのではないかと。そこで、多様な思想を反映させるため、「親鸞ボット」や「世親ボット」といった複数の新しいチャットボットも開発しました。

また、テキスト対応だけでは物足りないということもあり、顔を見て声を聞いて意思疎通がしたいと思い、チャットボット×ARといったプロダクトも開発しました。あとは、こういった技術を用いて、新しいサイバー・フィジカル・システムを構築しようということで、「テラバース」（一兆の宇宙）を創るという構想を発表したのが2022年9月のことです。

背景としては、2040年には3割のお寺が廃寺になるとも言われていまして、お寺をインターネットでつないで、現実空間上の寺院の経営を安定化させるとか、VRを利用して仮想空間上にお寺や極楽浄土を作ってみようとか、そういったことにチャレンジしようと思いました。ですが、VRはいわゆる不気味の谷という現象もありましてまだ難しいということで、今はARに舵<sup>かじ</sup>を切っているところです。

社会や文化、産業への波及効果という点で言うと、伝統知とテクノロジーの融合により、物理空間の障害を越え、圧迫感や生きづらさを解消できる可能性があります。もちろん、悪用や誤用のリスクもありますが、新しいものを創出できる可能性は大いにあります（図表4）。

こういったコンセプトを作り、伝統知テック産業として産業界にも呼び掛けようとしたのですが、市場規模が0円のところにリスクを冒して参入するというのも企業側にとっては難しいということで、まず我々が株式会社テラバースというスタートアップを創業しまして、取りあえず市場を0円から1円以上にしたところですよ。

異分野融合で0を1に変える研究開発ができたのは、多様な考えを持った仲間たちとの協働のおかげでもあり、役に立たないと言われがちなものでも場合によっては新しい価値やプロダクトを生み出す可能性があります。本日の議論を通じて、新しい創発が出てくることを願っております。

図表3 新型「ブッダボットプラス」

**【例1】一般的なテーマも**

**旧式「ブッダボット」**

Q SNSを始めようと思いますが、どう思いますか。

A 自分の経験や考えに強くこだわって、その有益さを強調している。自分以外のものをすべて売っているに見下してしまうことになる。

Q SNSは必要ないですか。

A 正しくないかわかったことをしてはならない。

Q なぜ、SNSだけに頼ってはいけないのですか。

A 悪いことをすると、他人を傷つけるだけでなく、自分をも傷つけることになるからである。

**新型「ブッダボットプラス」**

Q SNSを始めようと思いますが、どう思いますか。

A **經典の文言**  
何ものにも執着せず、余計なものを持たない人が、幸福らしい人である。  
(SNSを始めるとどうかは個人の自由ですが、無駄なものに執着せず、必要なものだけを持つことが重要という意味です。)

**ChatGPT4の解釈**  
SNSは必要ないですか。

A 自分が得た知識だけに頼ってはならない。(SNSが必要かどうかは個々の状況によるので、SNSだけに頼らず、他の情報源や自身の判断も重視することが重要という意味です。)

Q なぜ、SNSだけに頼ってはいけないのですか。

A 学識や智慧を増やせない原因は、その人の生かさず消費な姿勢である。(SNSだけに頼ると、情報の多様性や深さが欠け、自己学習の機会を失う可能性があります。)

図表4 社会・文化・産業への波及効果

## ◎社会・文化・産業への波及効果

- ・伝統知とテクノロジーの融合により、物理空間の障害を越え、圧迫感や生き辛さを解消。
- ・仏教などの宗教や哲学、精神性等に、新たな活動の場や方法を提供。  
⇒新たな思想、文化、芸術などが創出可能に。
- ・「**伝統知テック産業**」の創出  
⇒「**株式会社テラバース**」を創立(熊谷&古屋)  
⇒伝統知コンサルティング&伝統知テクノロジー開発

## 講演タイトル：大規模言語モデルのこれまでとこれから



Sakana AI 株式会社 Research Scientist  
秋葉 拓哉 氏（秋葉氏提供）

2013 年～ 2015 年 日本学術振興会 特別研究員 (DC1)  
2015 年 東京大学 大学院 情報理工学系研究科  
コンピュータ科学専攻博士課程 修了  
2015 年 国立情報学研究所 特任助教  
2016 年 株式会社 Preferred Networks リサーチャー  
2018 年 株式会社 Preferred Networks 執行役員  
機械学習基盤担当 VP  
2023 年 Stability AI Japan 株式会社  
Senior Research Scientist

### 講演要旨：

進化的アルゴリズムを用いて多様なオープンソースモデルを融合することで新しい基盤モデルを効率的に開発する手法「進化的モデルマージ」を提案する。我々のアプローチは、既存のオープンモデルの膨大な集合知を活用するため、モデルを非常に効率的に作成できる。進化的モデルマージは、「非英語言語と数学的推論」「非英語言語と画像」といった、これまででは困難とされていた全く異なる領域のモデルをマージする方法すらも自動的に発見できることが分かった。

### 講演のポイント：

本日は大規模言語モデルのこれまでとこれからについて、特に「進化的モデルマージ」という新しい技術を中心にお話しさせていただきます。進化的モデルマージは、既存のオープンソースモデルを融合し、新しい基盤モデルを効率的に開発する手法です。

まず、Sakana AI という我々の会社について簡単に御紹介いたします。当社は 2023 年 7 月に設立され、日本を拠点に活動しています。共同創設者の David Ha は Google Brain Tokyo のリーダーを務めた機械学習・進化計算の専門家であり、Llion Jones はトランスフォーマー技術の開発に関わっ

た研究者です。我々は、ファンデーションモデルと呼ばれる最先端の AI 研究について、自然から着想を得たアイデアを使って、改善や新しいアプローチへの挑戦を目指しています。

現在の大規模言語モデルは、大規模なリソースを投入し、長期間にわたりモデルを学習させています。これにより便利な AI が開発されてきましたが、多大な計算資源と時間が必要で、様々な課題があります。一方、自然界に見られるアプローチは、柔軟性や適応性を持ちます。例えば、<sup>あり</sup>蟻は一つ一つは大した知能を持たなくても、協力して橋を作り、谷を越えることもある。これは、1 つ 1 つは非常に少ないリソースでも、大きな成果を達成できる可能性を示しています。

今回紹介するモデルマージは 2 つ以上の機械学習モデルをもとに 1 つの新しいモデルを作るという技術で、大きく 2 つのアプローチがあります。1 つ目は重みレベルのモデルマージです。これは 2 つのモデルの平均を取るような形で、新しいモデルを生成する方法です。簡単な手法ですが、結構うまくいきます。

二つ目はレイヤーレベルのモデルマージです。こちらはモデルの重みを変更せずに、異なるモデルのレイヤーを重ね合わせる手法です。このアプローチでは、パラメータの数が減ったり増えたりするという特徴があります。

これによって、1 つ 1 つのモデルの予測結果よりも性能が良くなったり、複数のモデルの持つ能力を統合したりすることができるようになると考えられています（図表 5）。

このモデルマージというのは英語圏ではかなり話題になっていたのですが、日本の AI コミュニティではそれに比べて話題になっていませんでした。これには日本語ベースモデルの特徴が関係していると考えていまして、まず日本にはプレイヤーが少なく、マージできるようなオープンソースのモデルというのが少ない点。もう 1 つは、技術的な理由により、日本語の LLM と英語の LLM ではマージがうまくいかないという点。しかし、むしろこれはチャンスだと私は捉え、進化的モデルマージという技術を開発しました。

進化的モデルマージは、先ほどの 2 つのアプローチを組み合わせ、進化計算と呼ばれるアプローチによってより複雑な新しいモデルを作るための技術です。アルゴリズムの詳細はスライドに記載のブログなどを見てもらえればと思いますが、かなり汎用性があり、いろいろなものを最適化することができます。

我々はこれを使って、非常に柔軟かつ適応的な新しいモデルの作り方を開発しました。その実証実験の結果の一つですが、日本語の言語モデルと英語の数学特化モデルを混ぜ合わせ、日本語で数学問題を解けるモデルを開発しました。元の日本語モデルは数学の能力が足りず、英語モデルは日本語が分からないのですが、2つを混ぜ合わせることで新しい能

力を発現することに成功したわけです（図表6）。

さらに、日本語の言語モデルと英語のビジョン言語モデルを組み合わせ、日本語対応のビジョン言語モデルを作成することにも成功しました。進化的モデルマージが単なる性能向上ではなく、今までできなかったことをできるようにする、非常に新しい技術であると考えております。

図表5 モデルマージとは？

## モデルマージとは？

$$\text{LLM}_{\text{new}} = \text{Merge}(\text{LLM}_1, \text{LLM}_2, \text{LLM}_3, \dots)$$

2つ以上のモデルを元に1つの新たなモデルを作る

**アプローチ2つ**

1. 重みレベルのモデルマージ
2. レイヤーレベルのモデルマージ

図表6 日本語数学 LLM

## 結果1: 日本語数学LLM

### 日本語LLM + 英語数学LLM → 日本語数学LLM

| Id. | Model                 | Type       | Size | MGSM-JA (acc ↑) |
|-----|-----------------------|------------|------|-----------------|
| 1   | Shisa Gamma 7B v1     | JA general | 7B   | 9.6             |
| 2   | WizardMath 7B v1.1    | EN math    | 7B   | 18.4            |
| 3   | Abel 7B 002           | EN math    | 7B   | 30.0            |
| 4   | <b>Ours (PS)</b>      | 1 + 2 + 3  | 7B   | <b>52.0</b>     |
| 5   | <b>Ours (DFS)</b>     | 3 + 1      | 10B  | <b>36.4</b>     |
| 6   | <b>Ours (PS+DFS)</b>  | 4 + 1      | 10B  | <b>55.2</b>     |
| 7   | Llama 2 70B           | EN general | 70B  | 18.0            |
| 8   | Japanese StableLM 70B | JA general | 70B  | 17.2            |
| 9   | Swallow 70B           | JA general | 70B  | 13.6            |
| 10  | GPT-3.5               | commercial | -    | 50.4            |

元の「日本語能力」と「英語数学能力」を維持出来ただけでなく、  
その両方を日本語数学能力として組合せて発揮出来る



<講演資料掲載ページ>

<https://www.nistep.go.jp/archives/57541>

※ 本記事内図表はすべて講演者の使用スライドより許可を得て掲載。