

# ChatGPT と教育における信頼性



専門は人工知能、  
自然言語処理です

松林 優一郎  
東北大学 教育学研究科



教育評価測定論領域  
<https://www.eam.sed.tohoku.ac.jp/>

グループの方々にご意見頂きました。  
有難うございます。



自然言語理解チーム

[https://www.riken.jp/research/labs/aip/goalorient\\_tech/nat\\_lang\\_understand/index.html](https://www.riken.jp/research/labs/aip/goalorient_tech/nat_lang_understand/index.html)

TOHOKU  
NLP  
東北大学  
自然言語処理  
研究グループ  
GROUP

<https://www.nlp.ecei.tohoku.ac.jp/>

# ChatGPTとは

- 大規模言語モデル
- 大量の「人間の言語運用例」から「文章や対話の続き」を予測するように学習されたモデル
- 様々なジャンルの知識を備えた汎用モデル

# 何がすごいのか

- 文章の続きを予測する学習と人間にとって望ましい返答の学習を超大規模に行うと、
- 直近まで困難とされていた、言語に関わる様々な知的処理を高度に達成できている（ようにみえる）という事実
  - できてもおかしくはない予感があったが、ここまでできるようになるのかという驚き
- しかも、一つのモデルで様々なタスクを実現できる
  - 特定のタスク専用追加チューンされたモデルのほうが優れることも多いが、チューンのコストは大きい
  - 何故、どこまで出来るのかの分析は現在進行中
  - 社会適用が先行している状態
- 技術的解説
  - 黒橋 禎夫. ChatGPTの仕組みと社会へのインパクト.  
<https://edx.nii.ac.jp/lecture/20230303-04>
  - 岡崎直観. 大規模言語モデルの驚異と脅威.  
[https://speakerdeck.com/chokkan/20230327\\_riken\\_llm](https://speakerdeck.com/chokkan/20230327_riken_llm)
  - 松尾研究室. AIの進化と日本の戦略.  
[https://note.com/akihisa\\_shiozaki/n/n4c126c27fd3d](https://note.com/akihisa_shiozaki/n/n4c126c27fd3d)
  - JDLA緊急企画！「生成AIの衝撃」～ ChatGPTで世界はどう変わるのか？ ～  
<https://www.youtube.com/watch?v=TVaB5R4-uOE>

# 大学教育に与えたインパクト

- サービス開始直後から、大学における生成系AIの利用に関して様々な議論が起こり、各大学が利用に関する指針を策定
- 国内の動き
  - 東京外国語大学 3/22 [大学教育における AI について](#)
  - 上智大学 3/27 [ChatGPT等のAIチャットボット（生成AI）への対応について](#)
  - 東北大学 3/31 [ChatGPT等の生成系AI利用に関する留意事項（教員向け）](#)
  - 東京大学 4/3 [生成系AI\(ChatGPT, BingAI, Bard, Midjourney, Stable Diffusion等\)について](#)
  - （報道）4/6 文部科学省 「ChatGPT」など 学校現場での取り扱い示す資料作成へ
  - 神戸市外国語大学、群馬大学、島根大学、山形大学、東洋大学、…………
- より詳細な資料
  - 吉田 壘. ChatGPTを教育に取り入れる：アイデアと留意点.  
<https://youtu.be/slxkNglagHQ>
  - 金丸 敏幸. ChatGPTが語学（英語）教育に与えるインパクト.  
<https://youtu.be/hzsMs0h8OM8>
  - 吉田 壘 研究室. ChatGPT・AI の教育関連情報まとめ.  
<https://edulab.t.u-tokyo.ac.jp/chatgpt-ai-resources/>
  - 石井雄隆. ChatGPT関係のweb記事まとめ.  
[https://docs.google.com/document/d/1cqRwcNbCSHWIvpL3Nr-R0wqSUDHuV\\_AXWfzcmeSD3SU/edit](https://docs.google.com/document/d/1cqRwcNbCSHWIvpL3Nr-R0wqSUDHuV_AXWfzcmeSD3SU/edit)

# 議論の根幹にあるもの

- 利活用の方法：ポジティブな面は取り入れれば良い
  - 今後間違いなく社会で使われていくであろう技術に触れる機会を高等教育の現場から締め出すという選択肢は考えにくい。むしろ取り入れるべき。
- 火急の問題：従前の学習・評価体系の一部を**深刻に破壊する**
  - レポート課題に対するチーティング（？）
    - Stanford 大学の17%の学生が秋季の最終課題等にChatGPTを利用した <https://stanforddaily.com/2023/01/22/scores-of-stanford-students-used-chatgpt-on-final-exams-survey-suggests/>
  - プログラミング等における基礎演習課題が成立しない
- 利益のあるものは必ず使われる
  - 学生が語学学習支援にMTを使用する頻度の推移 [Jolley 2022]

今回はこれより遥かに波及が速い

[Pena 2011]	オーストラリア中等教育のスペイン語学習者の <b>33%</b> がMT利用
[Clifford+ 2013]	アメリカの大学のスペイン語学習者の <b>71%</b> はMTを時々または頻繁に使用
[Farzi 2016]	調査対象の大学ESL学習者全員が少なくとも何度かMTを使用し、 <b>84%</b> が頻繁に使用していた
[Valijärvi and Tarsoly 2019] [Enríquez Raído and Sánchez Torrón 2020]	<b>事実上すべての語学学習者</b> が学習活動にMTを使用しており、そのほとんどが頻繁に、特にライティングにMTを使用している

# 学習機会と評価の正確性・公平性の喪失

- 様々な知識や概念を調べ、習得し、集約・整理し、要約、一般化、価値生産する過程が重要
- 答えはあるが反復訓練が必要な技能の学習機会の喪失
  - 基礎力を身につける際の障害にならないか
  - Ex) 算数 vs 電卓、漢字学習 vs かな漢字変換、語学基礎 vs 機械翻訳、…調べ学習 (情報の整理・体系化) vs 生成系AI?
- 教師側も、学生にどこまで理解・能力があるかの評価が難しくなる
- 評価の公平性も失われる
- 対応の動向 (前回の吉田先生資料) [https://www.nii.ac.jp/event/upload/20230329-06\\_Yoshida.pdf](https://www.nii.ac.jp/event/upload/20230329-06_Yoshida.pdf)
  - 課題の出し方について様々な代替案、影響の軽減案が提言される
  - 一時的な保護措置として使用禁止とする場合も
  - どのような影響が及ぶかは分野・科目、学校毎に違うのでコース各自のルールがありうる
  - OpenAIも自ら教育者向けに留意点を公表 <https://platform.openai.com/docs/chatgpt-education>

# アカデミック・インテグリティの問題

- 学術上の不正行為は極めて厳格に取り扱われてきた
  - 試験中のカンニング、剽窃等 → 全単位無効化等の厳罰
  - 剽窃は学術上の重大な倫理違反
- 生成系AIの文章を取り入れる（or 参考にする）ことは剽窃なのか
- 生成系AIが書いた文章は本当にどこかの誰かが書いた文章ではないのか → （剽窃チェッカー）
- 各国際学会・出版局がポリシーを策定
  - Ex) ケンブリッジ大学出版局 <https://www.cambridge.org/core/services/authors/publishing-ethics/research-publishing-ethics-guidelines-for-journals/authorship-and-contributorship#ai-contributions-to-research-content>
    - AIは説明責任を果たせないので**著者資格を持たない**
    - AIの使用によって剽窃ポリシーに違反してはならない
    - **著者は、AIの利用を含め、研究論文の正確性、完全性、独創性について責任を負う**

# 「真実のような嘘」(Hallucination) が及ぼす影響

- 学術知識を体系的に学ぶ際は、受け取る情報が真実であることを前提としている
- Web検索は出典を当たれるが、生成系AIは必ずしもそうではない
  - 一部サービスは部分的に根拠文章を提供 (Microsoft Bing)
- 高等教育
  - 批判的思考力、ファクトチェックの訓練という考え方もあるが、これらは本来非常に高度な能力
    - Ex) 無数に蔓延るデマ、フェイクニュース、プロパガンダ
      - 言語処理技術でこれらを抑止する研究領域もある
  - 学術的知識の嘘は専門家でないで見抜けないという指摘も
- 初等・中等教育を部分的に保護すべきか
  - ※ ChatGPT は現在18歳以上のユーザー（もしくは親権者の同意を得た13歳以上）を対象に提供されるサービス



# 学習のための効果的な利用法の指南や 教師の正確な情報・適切な論点整理によるガイドの効果

- 機械翻訳の教育応用時の知見
  - [Thue Vold 2018] 学生に複数の機械翻訳結果を比較させ、議論させる
    - 学習者のメタ言語的認識を高めるためにMTテキストを使用することは「かなりの可能性」を持っているが、**使いこなす訓練、教師からの指導が最も重要**
    - **学習者のみでは有益な論点にたどり着けない**
    - 解釈の間違いで言語に関する**誤解を深める可能性さえある**
  - [Resende & Way 2021] 英作文に活用させる
    - 学習者は**適切なトレーニング**を受けることでより有効に機械翻訳システムを語学学習に活用できる。また、**その外国語に対し十分な習熟度がある場合に、より多くの恩恵**を受けられることができる
- 適切な活用方法について指導をすることが有効
- 学習者が自ら翻訳結果を評価できるような一定以上の基礎スキルを身に付けていることが重要
- 人間の教師が与えるような解説の必要性

# 生成系AI は良き教師になれるか (AIの教育応用に向けて)

- 安全性と信頼性は重要な観点 (教育における説明可能なAI)
  - 提供する情報の信頼性
    - 調べたことの全てに裏を取る、は通常の学習時には非現実的
  - 判断の自信度、自己の専門性把握
    - AIが自らの信頼性を自己評価する
  - 学習者の能力評価に対する信頼性
    - 学習者について何が不十分か、人間の教師と同じように説明・評価できる
- 教育応用に特化した生成系AI
  - 特化型はより信頼性を担保したシステムを実現できる
  - 目的がずれていると教育効果は希薄化する
    - Ex) Grammarly を用いた語学学習におけるエンゲージメント [Koltovskaia 2020; Ranalli 2021]
  - 知識提供者としての教師： 構造的な知の教授 vs 検索による断片的な理解
  - 助言者としての教師： 個別の状態に応じた適切なアドバイス

# 記述式答案自動採点における判断根拠の説明

[Mizumoto+ 2019, Funayama+ 2020]

- 国語の読解問題における記述式答案採点
- AIが得点の根拠とした箇所をあわせて出力する
  - AIが何を見ているか分かる (挙動の解釈性・人から見た納得感)
  - 人間が定めるループリックと同様の観点を見ているか (判断の妥当性)

答案 2点 3点 3点 3点  
 西洋人は自分と異なる人間を異人と見なして考えを同意させるために、言葉を尽くして自分の考えを伝えようとする説得を行うこと。

モデルが採点時に着目している箇所 (根拠箇所)



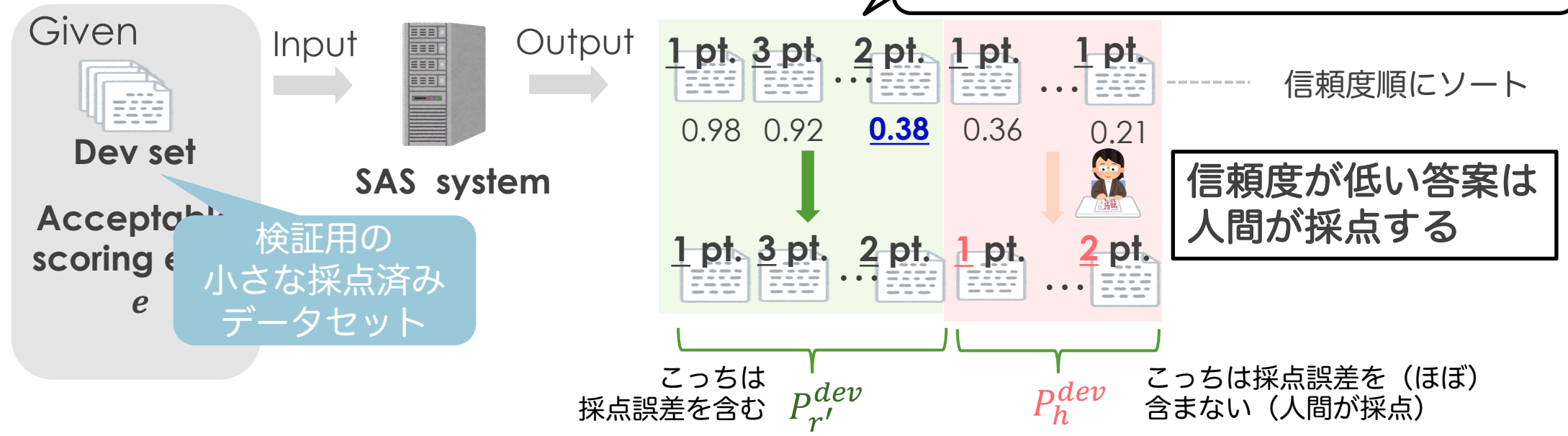
採点の根拠が示されるので、結果を解釈できる・結果に納得できる



# AIが回答に対する自らの信頼度を示し、 人とAIの分担で**目標の正確さ**を実現する [Funayama+ 2022]

① 目標の採点誤差を定める

② 検証用データで目標の採点誤差を超えないギリギリの信頼度まで採点させる



$$\tau = \operatorname{argmax} \left\{ \left| P_{r'}^{dev} \right| \right\}$$

$$\text{Subject to } \operatorname{Err} \left( P_{r'}^{dev} \cup P_h^{dev} \right) \leq e$$

人間の採点と機械の採点の採点誤差の合計が目標値を超えないように緑の量（AIの担当分量）を決める

# 学習者への有益な フィードバックコメントの生成

- ショートアンサーに対するフィードバックのデータセット構築

**Your Answer is Incorrect... Would you like to know why?  
Introducing a Bilingual Short Answer Feedback Dataset**

Anna Filighera, Siddharth Singh Parihar, Sebastian Ochs,  
Tim Steuer and Tobias Meuser  
Multimedia Communications Lab  
Technical University of Darmstadt, Germany

{anna.filighera, tim.steuer, tobias.meuser}@kom.tu-darmstadt.de

誤答に対し、何故間違っているのかを説明する

<b>Question:</b>	What are the challenges of Mobile Routing compared to routing in fixed and wired networks? Please name and describe two challenges.
<b>Answer:</b>	1) Due to hardware constraints, some nodes may be out of the range of others. 2) Mobile routing requires more flexibility. The environment is very dynamic and the routing mechanism has to adapt to that.
<b>Verification:</b>	0.5 out of 1.0 points (Partially Correct)
<b>Elaborated Feedback:</b>	While the second challenge of needing to be able to adapt to a dynamically changing environment is correct, the first challenge stated is not a challenge specific to mobile routing. In a wired network, nodes typically don't have a direct connection to each other node as well.

Table 1: An example answer with annotated feedback contained in SAF.

paper discusses the need for enhanced feedback models in real-world pedagogical ios, describes the dataset annotation p

<http://dx.doi.org/10.18653/v1/2022.acl-long.587>

**Towards Generating Counterfactual  
Examples as Automatic Short Answer  
Feedback**

Anna Filighera<sup>(✉)</sup>, Joel Tschesche, Tim Steuer, Thomas Tregel,  
and Lisa Wernet

Multimedia Communications Lab, Technical University of Darmstadt,  
Darmstadt, Germany  
anna.filighera@kom.tu-darmstadt.de

部分的誤答に対する反証的事実をフィードバックする

Table 1. Example student answer, common feedback and generated counterfactual.

<b>Question:</b>	What happens to the volume of the sound if you pluck a rubber band harder?
<b>Reference:</b>	The volume increases. The sound is louder
<b>Response:</b>	It vibrates more and it gets lower. → <i>Incorrect</i>
<b>Counterfactual:</b>	It vibrates more and it makes louder sound. → <i>Correct</i>

responses can be inclined to be classified as correct by an automatic grading model without straying too far from the initial response. However, an expert reevaluation of the modified responses shows that nearly all of them remain incorrect, only fooling the grading model into thinking them correct. While one of the counterfactual generation approaches improved student response, the general weakness of neural networks recommend further research

[https://doi.org/10.1007/978-3-031-11644-5\\_17](https://doi.org/10.1007/978-3-031-11644-5_17)



# 教師らしい応答の生成

<https://sig-edu.org/sharedtask/2023>

**SIG EDU**      Constitution   Officers   Members   News

---

Teacher Yes, good! And to charge it up, you need to \_\_ it \_\_

Student ...

Teacher connect to the source of electricity

Student i understand

Teacher plug it \_\_?

Student in

Teacher yes, good. And when the battery is full, you need to \_\_\_\_ (disconnect it)

## BEA 2023 Shared Task

### Generating AI Teacher Responses in Educational Dialogues

#### Motivation, Goal, and Purpose

Conversational agents offer promising opportunities for education. They can fulfill various roles (e.g., intelligent tutors and service-oriented assistants) and pursue different objectives (e.g.,

AI は

(a) 教師のように生徒に語りかけ、  
 (b) 生徒を理解し、  
 (c) 生徒の理解度を高めることができるか？

[Bommasani+2021]

to the development of conversational agents that are founded on more powerful generative language models.

- 最先端の生成モデルが、教育対話で生徒に返答できる優れたAI教師であるかどうか
- 教師と生徒の対話の実世界サンプルによるテスト

### [ DIALOGUE CONTEXT ]

Teacher Yes, good! And to charge it up, you need to \_\_ it \_\_

Student ...

Teacher connect to the source of electricity

Student i understand

Teacher plug it \_\_?

Student in

### [ REFERENCE RESPONSE ]

Teacher yes, good. And when the battery is full, you need to \_\_\_\_ (disconnect it)

# 終わりに

- 現行の学習・評価体系からのシフトは少なからず必要
  - ルール作り、評価方法の工夫
  - 検索ネイティブ → QAシステム・生成AIネイティブ時代の学習
- 生成系AIが教育において信頼できるAIとなるには
  - 挙動の説明・妥当性の評価、自己評価するAI（確信度）、人間の能力に対する理解と評価、構造的・段階的な知の教授
- 使う → 作るへ
  - 教育向けにチューンされたモデルの開発
  - 教育効果等の評価を開発のサイクルに組み込む
  - 技術の透明化・生成系AIのための研究基盤の構築