

AI・データサイエンス時代の 英語教育研究

千葉大学教育学部・大学院教育学研究科
理化学研究所革新知能統合研究センター

yishii@chiba-u.jp

<https://yutakaishii.com/>

自己紹介

- 専修大学附属高等学校
 - コミュニケーション英語，英語表現などを担当
- 早稲田大学大学総合研究センター
 - 高等教育開発（eラーニングの開発や教師の専門性開発など）
- 千葉大学教育学部・大学院教育学研究科
 - 英語教員養成科目を中心に担当
- 専門領域は言語テスト・ICTを活用した英語教育・英語ライティングの指導と評価

謝辞

- 本研究は以下のJSPS科研費の助成を受けています。
 - 混合研究法を用いた日本人英語学習者のライティングプロセスの解明（研究代表者：石井 雄隆）
 - 英語教育における高大接続に向けた発話自動採点システムの構築と運用（研究代表者：林 裕子）
 - ラーニングアナリティクスによる英語DDL（英語データ駆動型学習）の精緻化（研究代表者：西垣 知佳子）
 - 英語要約ライティングの形成的評価におけるフィードバック活用とその効果の検証（研究代表者：澤木 泰代）
 - 個別最適化されたe-Learningコースを活用したブレンド型授業－学習行動に着目して（研究代表者：小野 雄一）
 - 生成AIを組み込んだ日本語作文診断システムの開発と普及に関する研究（研究代表者：李 在鎬）
 - 英語教師の意思決定支援のための対話型ベイジアンデータ分析システム（研究代表者：草薙 邦広）
 - 第二言語における統語的複雑性に関する特徴量の検討（研究代表者：近藤 悠介）

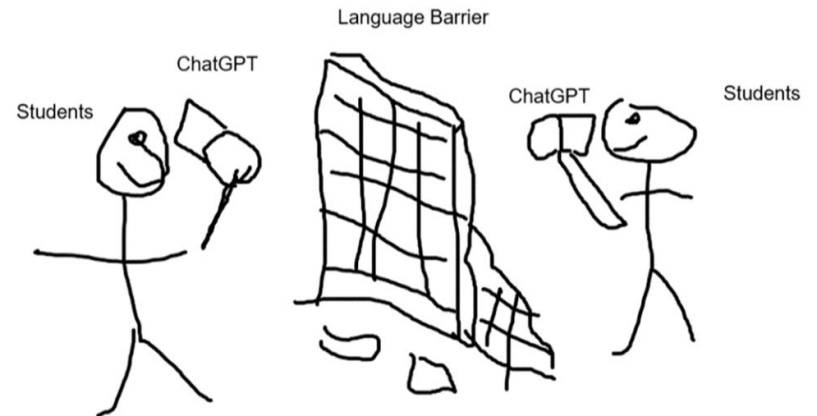
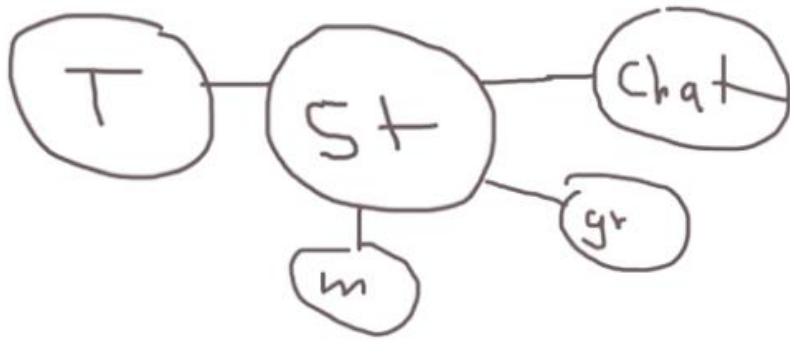
概要

- 教育現場では、アンケートやレポートなど、多くのデータが日々集められています。また、オンライン学習の普及により、学習履歴データが自動的に、そして大量に蓄積される時代となりました。しかし、その膨大なデータは本当に活用されているのでしょうか？今回は、このようなデータをどのように活用できるかに焦点をあて、特にAI・データサイエンスの手法を使った作文の自動採点についての研究をご紹介します。さらに、AI時代における英語教育のあり方について、参加者の皆さんと一緒に考えていきたいと思っています。皆様のご参加を心よりお待ちしております。

ChatGPTを説明するために 第二言語ライティング教師が選んだ 形容詞とフレーズ

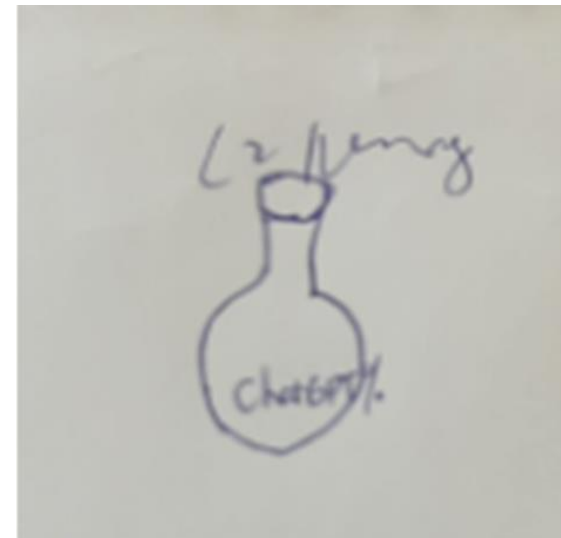
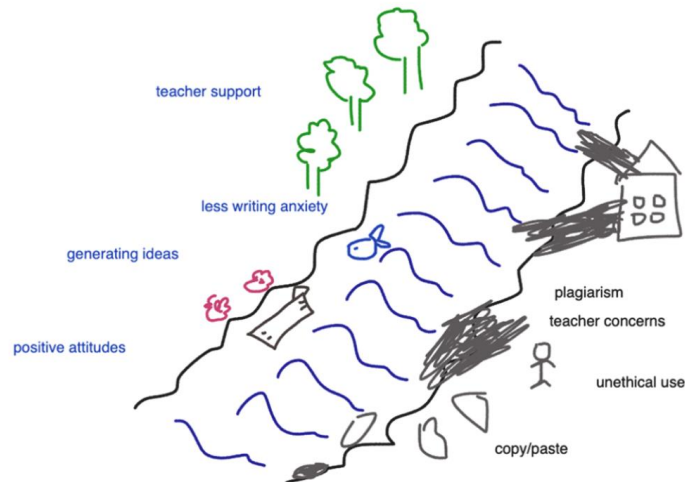
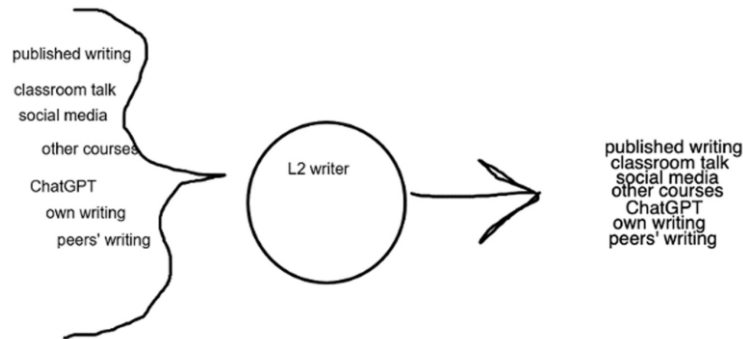
| Positive | Neutral | Negative |
|-----------------------------------|-------------------|-------------------------------------|
| Revolutionary (Norton) | Mysterious (Emma) | Troublesome (Norton) |
| Equalizing (Norton) | Evolving (Emma) | Formulaic (Cecile) |
| Helpful (Cecile, Katherine) | Hyped (Jack) | Time-consuming (Cecile) |
| Powerful (x2) (Emma, Kate) | | Disconcerting (John) |
| Promising (John) | | Ethically concerning (Alice) |
| Evolutionary (John) | | Scary (x3) (Kate, Katherine, Daisy) |
| Potentially revolutionary (Alice) | | Limited (x2) (Kate, Lily) |
| Exciting (Alice) | | Eerie (Lily) |
| Interesting (Jack) | | Threatening (Lily) |
| Useful (Jack) | | |
| Awesome (Katherine) | | |
| Innovative (Daisy) | | |
| Gainful (Daisy) | | |

ChatGPT as a tool



Xu, W., & Tan, X. (2024). Beyond words: L2 writing teachers' visual conceptualizations of ChatGPT in teaching and learning. *Journal of Second Language Writing*, 64, 101110.

ChatGPT as a resource



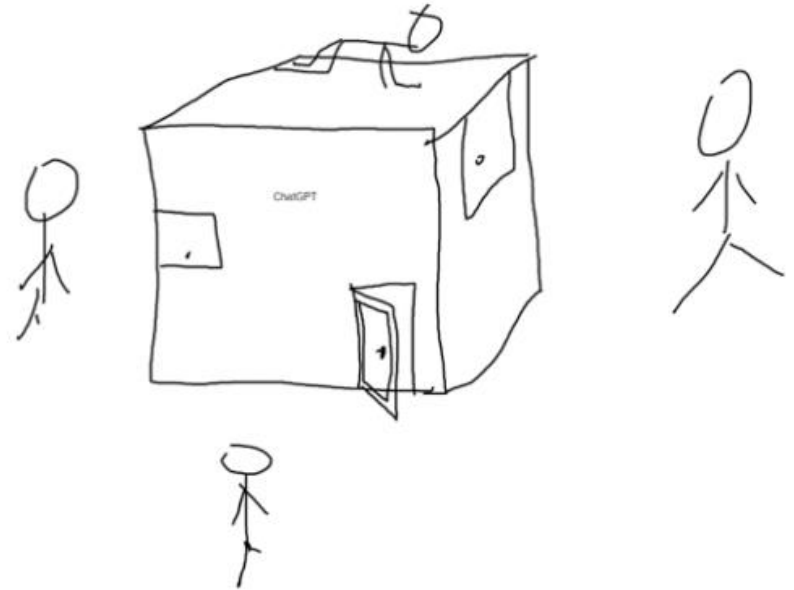
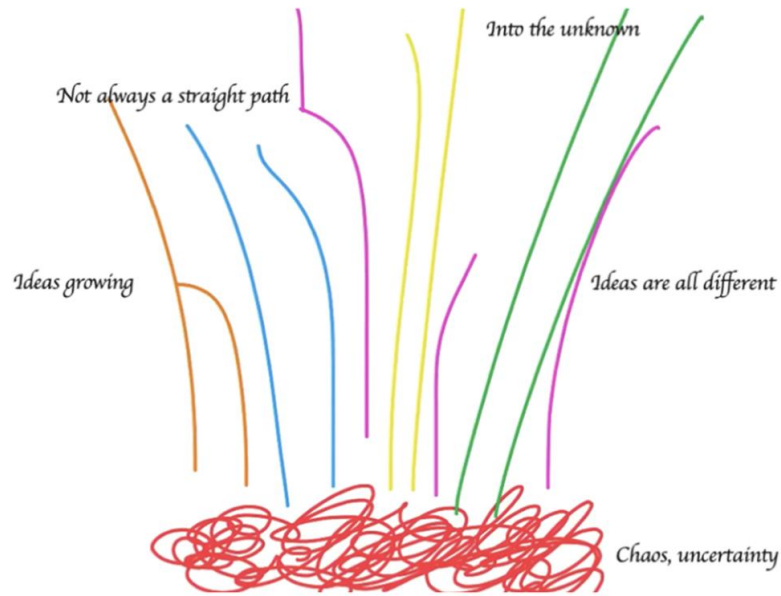
Xu, W., & Tan, X. (2024). Beyond words: L2 writing teachers' visual conceptualizations of ChatGPT in teaching and learning. *Journal of Second Language Writing*, 64, 101110.

ChatGPT as a threat



Xu, W., & Tan, X. (2024). Beyond words: L2 writing teachers' visual conceptualizations of ChatGPT in teaching and learning. *Journal of Second Language Writing*, 64, 101110.

ChatGPT as an unknown entity



本日の内容

- 背景
- 説明可能なAIと自動採点
- 生成AIと自動採点
- 人と共生するAIと自動採点
- 質疑応答

背景

社会的背景

- 社会の高度情報化
 - 大量のデータの生成
 - 記憶媒体の大容量化
- オンライン学習支援システムに蓄積される学習者の様々なデータを教育改善に利活用するラーニングアナリティクスなどが近年注目を集めている
- 生成AIを含む人工知能の発展は目覚ましく、それらをどのように社会応用するか、またデータ利活用のあり方についての議論が活発になってきています。

教育政策動向

- 教育データ標準（文部科学省, 2020）
 - 「データの種類や単位がサービス提供者や使用者ごとに異なるのではなく、相互に交換、蓄積、分析が可能となるように収集するデータの意味を揃える」（文部科学省）
- 「データ駆動型の教育への転換による学びの変革の推進」（教育再生実行会議, 2021）
 - 教育データの利活用推進の方向性
- 「教育データ利活用ロードマップ」（デジタル庁, 総務省, 文部科学省, 経済産業省, 2021）
 - 「教育のデジタル化のミッションを「誰もが、いつでもどこからでも、誰とでも、自分らしく学べる社会」と掲げ、そのためのデータの①スコープ（範囲）、②品質、③組み合わせ、の充実・拡大という「3つの軸」を設定

令和の日本型学校教育 (文部科学省, 2021)

- 個別最適な学び（「個に応じた指導」（指導の個別化と学習の個性化）を学習者の視点から整理した概念）
 - ICTの活用により，学習履歴（スタディ・ログ）や生徒指導上のデータ，健康診断情報等を活用することや，教師の負担を軽減することが重要
- 協働的な学び
 - ICTの活用による空間的・時間的制約を超えた他の学校の子供等との学び合いも大切

人文社会科学の学際化

- 計算社会科学
- デジタルヒューマニティーズ
- ELSI（倫理的・法的・社会的課題）
- 教育データサイエンス

データサイエンス

- データサイエンスは一般に「データを対象とした科学」あるいは「データから新たな価値を創出する科学」と定義される（柴田，2015）。
- データサイエンスに含まれる要素（竹村，2018）
 - データ処理
 - データ分析
 - 価値創造
- データ処理とデータ分析の重要性は広く知られるが、英語教育研究とデータサイエンスの統合において重要なことは、3つ目の価値創造という点
- 価値創造はたとえば、エビデンスに基づく政策形成や教育実践の場面への応用など（中西，2020）。
- このデータサイエンスによる価値創造を実現するには、英語教育の専門家だけでなく、他分野の専門家との協働が重要になる。

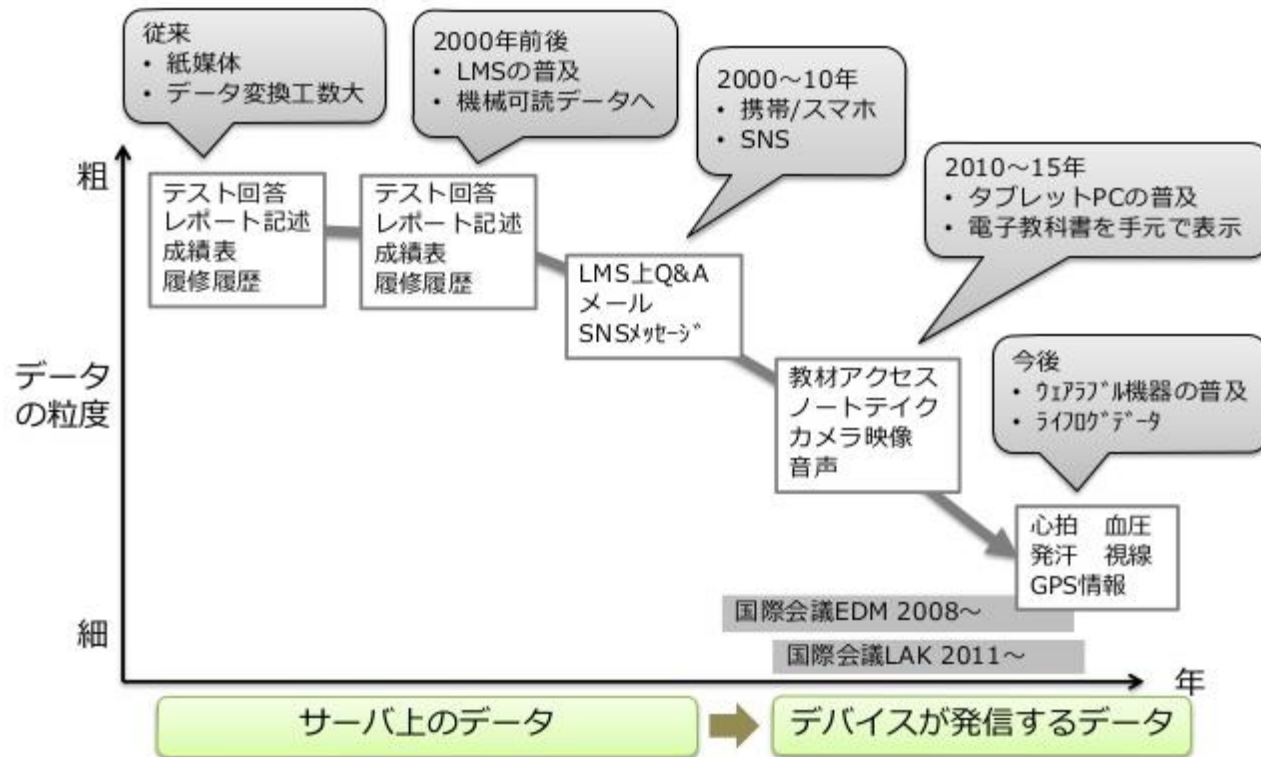
ラーニングアナリティクスとは？

- 「学習や学習が起こる環境を理解及び最適化するために学習者とそれにかかわるコンテキストについてのデータの測定・収集・分析・報告を行う」 (Ferguson, 2012)

ラーニングアナリティクスが近年注目される理由（山田, 2017）

1. 学習支援システムに蓄積されたデータ数は膨大であり、まさに教育「ビッグデータ」と言えること
2. ICT の発展により、端末の計算処理スピードが莫大に速くなったこと
3. 学習で使用する端末がモバイルにも広がり、アクセスログだけではなく、学習している場所など収集できるデータの量と質が大きく変化したこと

学習履歴データの電子化と細粒度化（田村, 2015）



ラーニングアナリティクスにより 可能になること (Bousbia & Belamri, 2013)

- 学習者モデリング
 - 知識、スキル、モチベーション、満足度、メタ認知、態度、学習の進捗状況などの詳細な情報が組み込まれたモデル
- 学習者の成績や学習成果の予測
- 適応的な情報推薦
- 学習者の行動解析
 - 個別指導、適応、パーソナライゼーションなど
- 学生に関するリアルタイムデータの視覚化

ラーニングアナリティクスの ユーザー別目的

<教育者>

- 学習者の学習プロセスを理解し、教育方法に反映
- 教育のパフォーマンスを向上
- 学習活動の認知的、行動的側面の理解



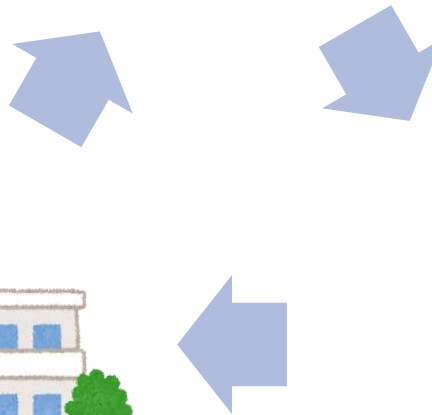
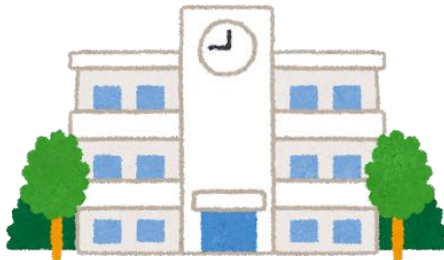
<学習者>

- 状況に応じた学習者支援
- 学習者に適したフィードバックの提供
- 学習効果の増大

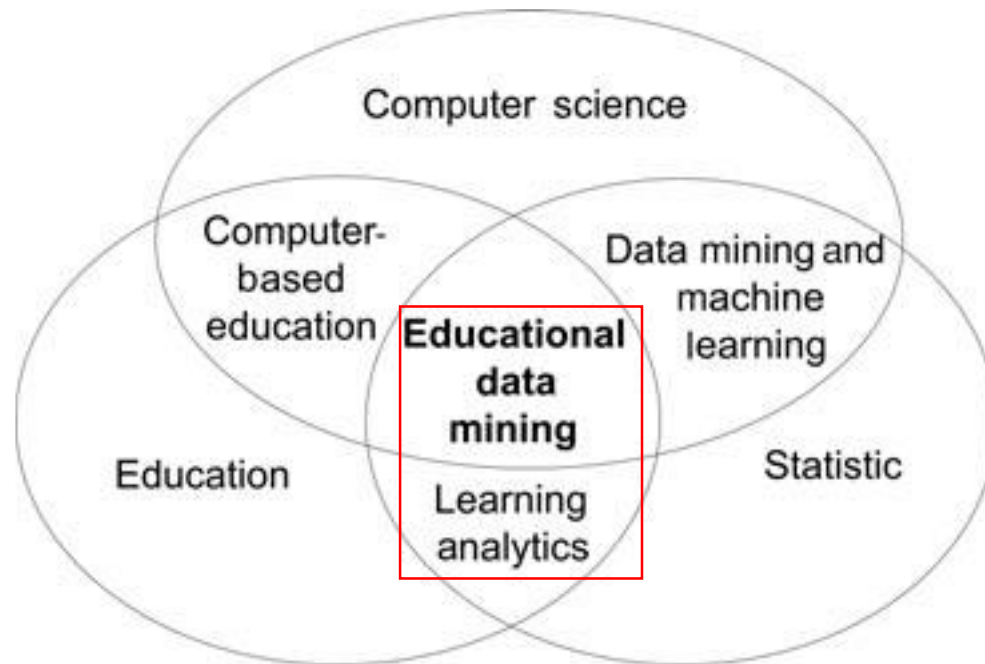


<管理者>

- 教育機関の資源を配分する最良の方法を検討



ラーニングアナリティクスと 関連する領域 (Romero & Ventura, 2013)

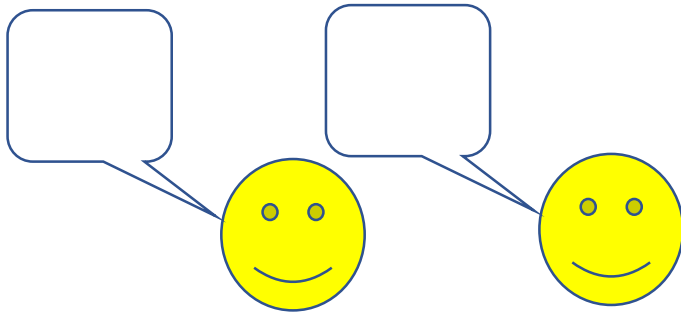


Data-centric AI

(データ中心のAI)

- AIシステムを構築するために使用されるデータを体系的にエンジニアリングする学問分野 (Ng, 2021)
- Data-centric AIは、学習データ開発、推論データ開発、データメンテナンスの3つの目標に分類される (Zha et al., 2023)

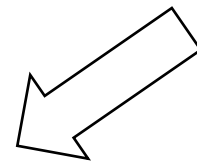
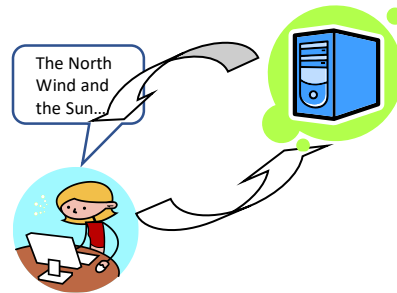
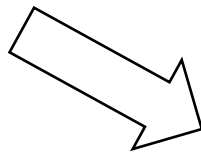
英語教育の現状と課題



日本人英語学習者の
英語を話す・書く能力の課題



話す・書く能力の評価は
労力とバイアスがかかる



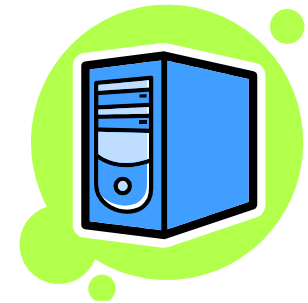
コンピュータを用いた英語を話す・書く能力の自動採点
AIによるフィードバックを用いた個別最適な学びの実現

自動採点システムの開発

1. 学習者がコンピュータ上でスピーキングやライティングを行う



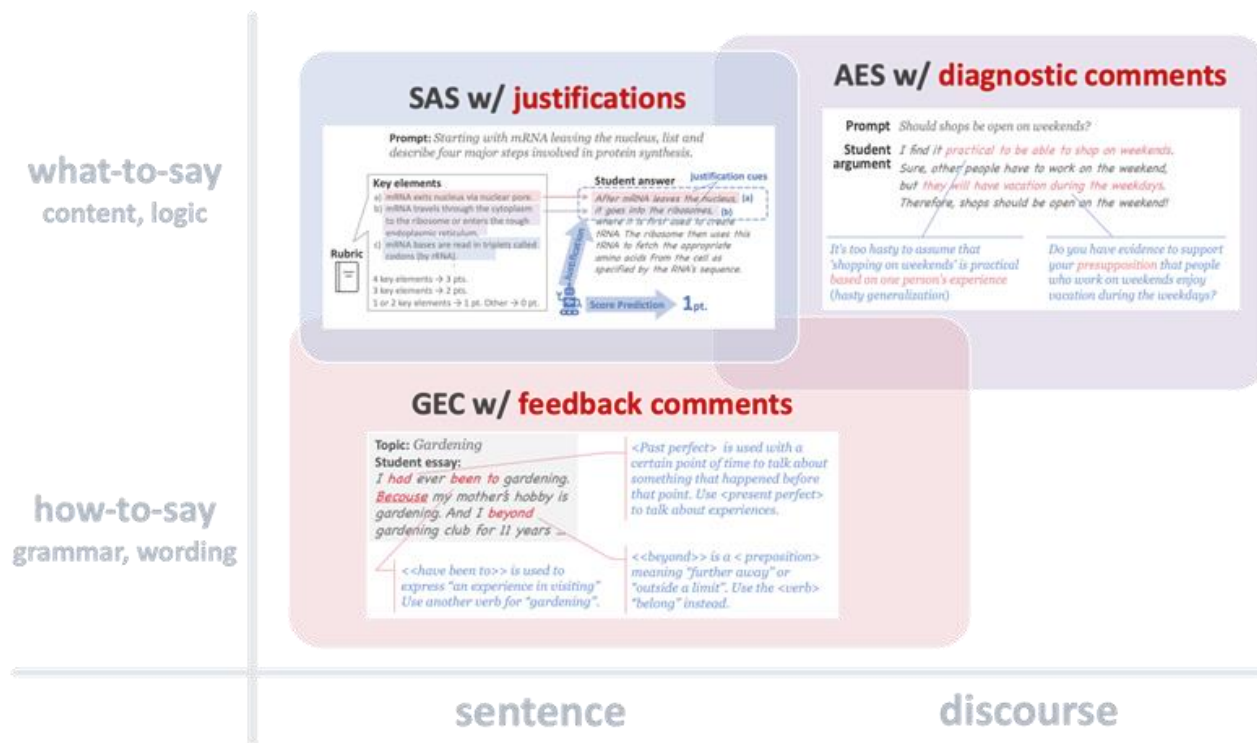
2. 開発したシステムが即時的にスピーキングやライティングを自動で採点する



3. 採点結果・個別最適なフィードバックを学習者に送信する

自動採点研究は英語教育学・統計学・情報学の知見を融合した学際領域

理化学研究所で取り組んでいる 自動ライティング評価タスクとその分類



説明可能なAIと自動採点

AI活用に求められる視点

- 倫理的な視点
- 社会的な視点
 - 公平性
 - 説明責任
 - 透明性
- システム的な視点
 - 可用性
 - 運用・保守性
 - 性能・拡張性
 - セキュリティ
- 機械学習的な視点
 - 精度

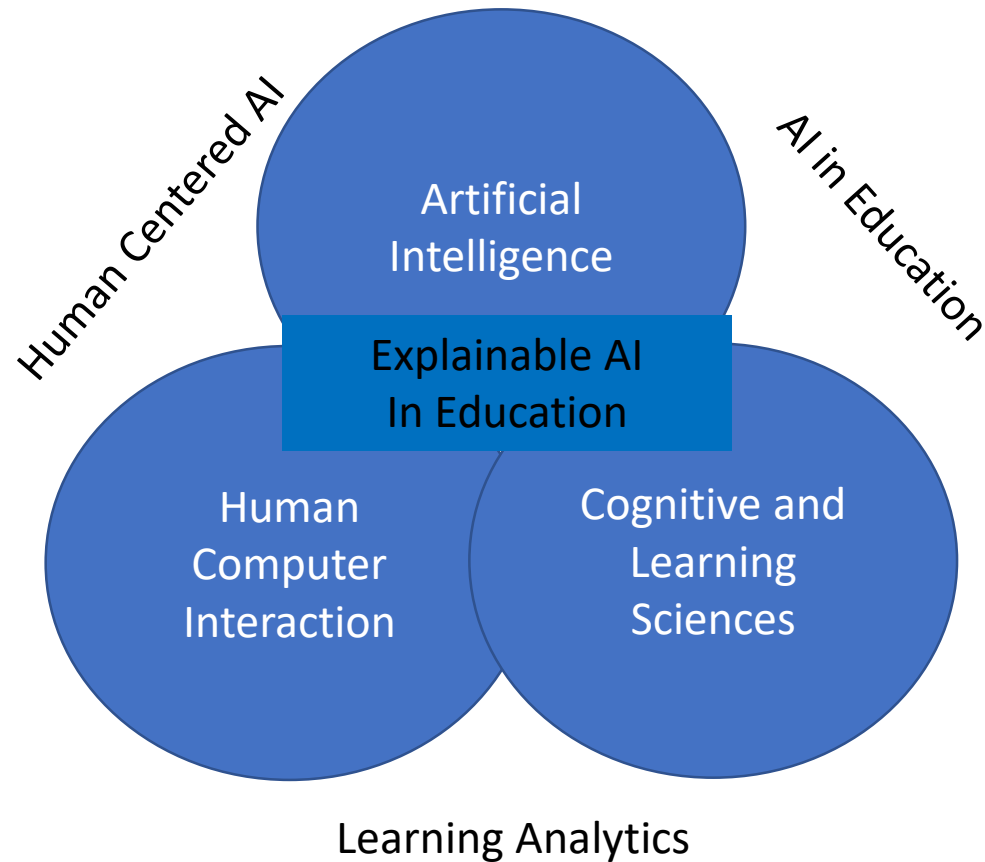
AIに求められる価値

- 精度
- 公平性
 - 学習時の入力データにバイアスが含まれると、バイアスに則った処理を獲得する
 - 歴史的バイアス
 - サンプルングバイアス
 - 加工バイアス
 - 学習バイアス
- 説明責任
 - 根拠の提示
 - 法律などの社会的な整備
- 透明性
 - 利用者が理解できる形で、AIシステムの情報を提示できること
 - どのようなデータで学習したか
 - どのような検証を行ったか
 - どのような基準や根拠に基づいて処理を行っているか

説明可能なAI（原, 2019）

- AIの出力に加えて、その出力を補助する追加の情報（モデルの解釈、判断根拠の説明、など）を出力する技術一般および研究分野全体を指す用語
- 大きな目的の一つは「ブラックボックスなAIは信頼できない」という不信の払拭
- AIの社会実装の際には、念頭に置くべき必要がある観点

説明可能なAIと教育に関連する分野 (Khorsavi et al., 2022)



インターフェース

アドバイス付き英作文自動採点デモ



次の日本語の下線部を英語に訳しなさい。ただし、回答に与えられた語句で文を始めること。

Q3 大阪からロンドンまで飛行機でどれくらい時間がかかりますか。

How is the time Osaka to London by airplane?

9 語

提出

How is the time Osaka to London by airplane ?

要チェック項目(4/6)

正解項目(2/6)

全ての項目(6)

採点基準 D

重要度
☆☆☆

語順が適切に書けている。

語順はfrom A to Bとなります。



採点基準 E

重要度
☆☆☆

主語が適切に書けている。

形式主語itで文を始めましょう。



採点基準 F

重要度
☆☆☆

「どのくらいの時間」が適切に書けている。

how longと表現しましょう。



採点基準 G

重要度
☆☆☆

「(時間が) かかる」が適切に書けている。

動詞takeを用いましょう。



フィードバックの方針

答案: "She spoke Japanese **fluent** because she **live** in Japan for 30 years."

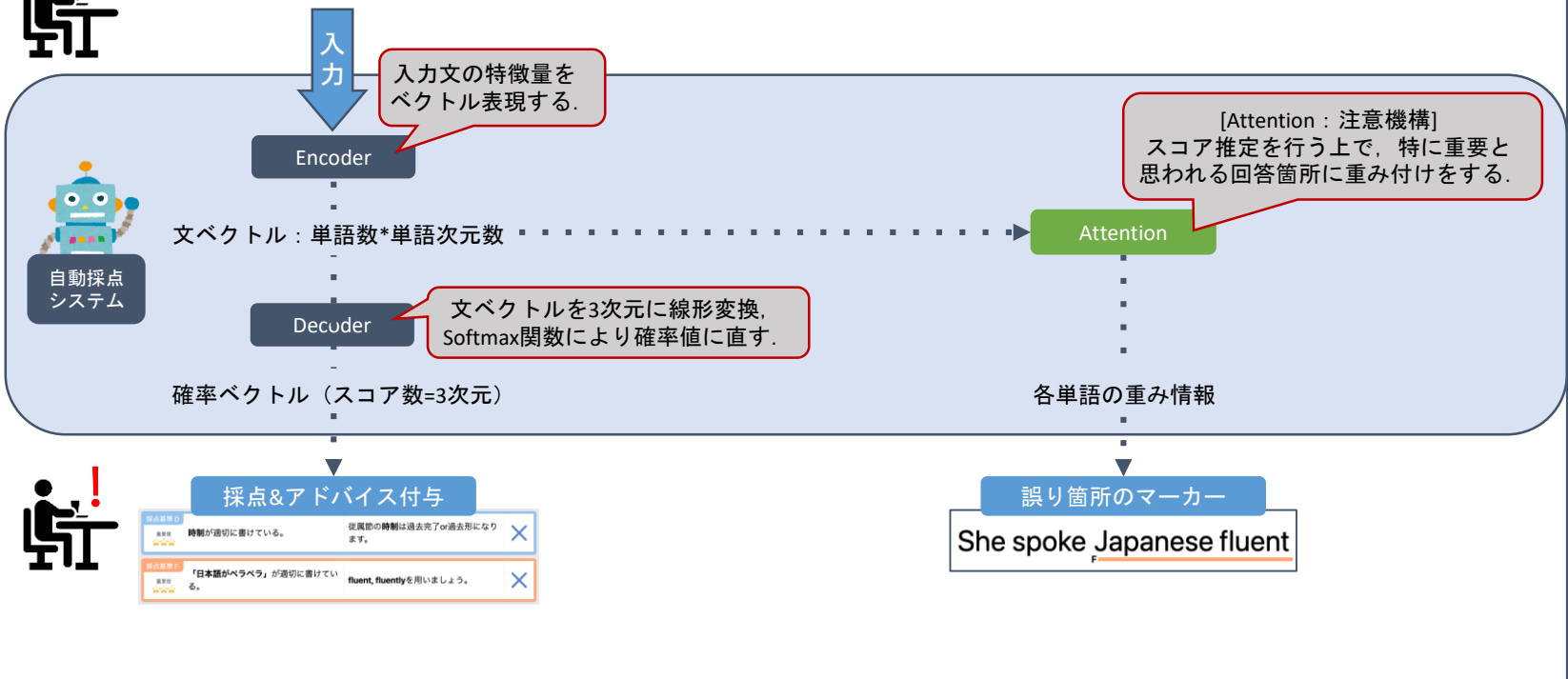


問題文: 次の日本語の下線部を英語に訳しなさい。

和文: 「彼女は日本で30年暮らしていて、日本語がペラペラでした。」



学習者はPC画面に表示される問題文を読んだ上で、回答をキーボード入力する。



採点モデル

出力

採点基準 F

重要度
☆☆☆

「日本語がペラペラ」が適切に書けている。

fluent, fluentlyを用いましょう。



She spoke Japanese fluent because she live in Japan for 30 years

根拠箇所

0 0 1 1 0 0 0 0 0 0 0 0

正規化+しきい値

根拠箇所重み

0.01 0.01 0.41 0.49 0.01 0.01 0.01 0.01 0.01 0.01 0.01 0.01

Attention Layer

文脈埋込表現

[[[0.1, 0.2, 0.3], [0.1, 0.2, 0.3], [0.1, 0.2, 0.3], [0.1, 0.2, 0.3], [0.1, 0.2, 0.3], [0.1, 0.2, 0.3], [0.1, 0.2, 0.3], [0.1, 0.2, 0.3], [0.1, 0.2, 0.3], [0.1, 0.2, 0.3], [0.1, 0.2, 0.3], [0.1, 0.2, 0.3]]]

Encoder

単語ベクトル

[[[0.1, 0.2, 0.3], [0.1, 0.2, 0.3], [0.1, 0.2, 0.3], [0.1, 0.2, 0.3], [0.1, 0.2, 0.3], [0.1, 0.2, 0.3], [0.1, 0.2, 0.3], [0.1, 0.2, 0.3], [0.1, 0.2, 0.3], [0.1, 0.2, 0.3], [0.1, 0.2, 0.3], [0.1, 0.2, 0.3]]]

入力答案

She spoke Japanese fluent because she live in Japan for 30 years

問題文

彼女は日本で30年暮らしていて、日本語がペラペラでした。

採点結果



Classifier

[[0.1, 0.2, 0.3]]

重み付き和
(文ベクトル)

AIを用いた学習における課題と 課題解決へのアプローチ

- **主体性をどのように動機づけるか**（吉田，2022）
- **エンゲージメントへの着目**
 - AIを用いた自動ライティングフィードバック研究において、近年盛んに研究がされている（Shi & Aryadoust, 2024）。
- **ライティング訂正フィードバック処理（WCF processing）**
 - 学習者が様々なタイプのフィードバックにどのように反応し、フィードバックからどのような恩恵を受けるかを明らかにしようとするプロセス重視の研究（Coyle et al., 2023）

エンゲージメントの分類 (Koltovskaia, 2020)

- **行動的エンゲージメント**

- 改訂作業，すなわち実際に行った改訂，原稿の精度を高めるために用いた改訂方略，改訂に費やした時間に関するもの

- **認知的エンゲージメント**

- 学生が自動文章評価システムをどれだけ深く処理したか（気づきや理解），およびメタ認知的操作と認知的操作に関するもの

- **感情的エンゲージメント**

- 自動文章評価システムに対する学生の即時的な感情反応と態度に関するもの

実験の様子



スクリーンキャスト



ビデオ録画

実験参加者



テスト冊子



研究者が観察シートにメモ

実験結果

- What the researchers did

- 複数の評価観点を採点項目として反映した項目採点モデルにより自動採点を行い、学習者に診断的なフィードバックをおこなうシステムを開発
- 日本人大学生を対象とした刺激再生法と半構造化インタビューを用いた実験を実施

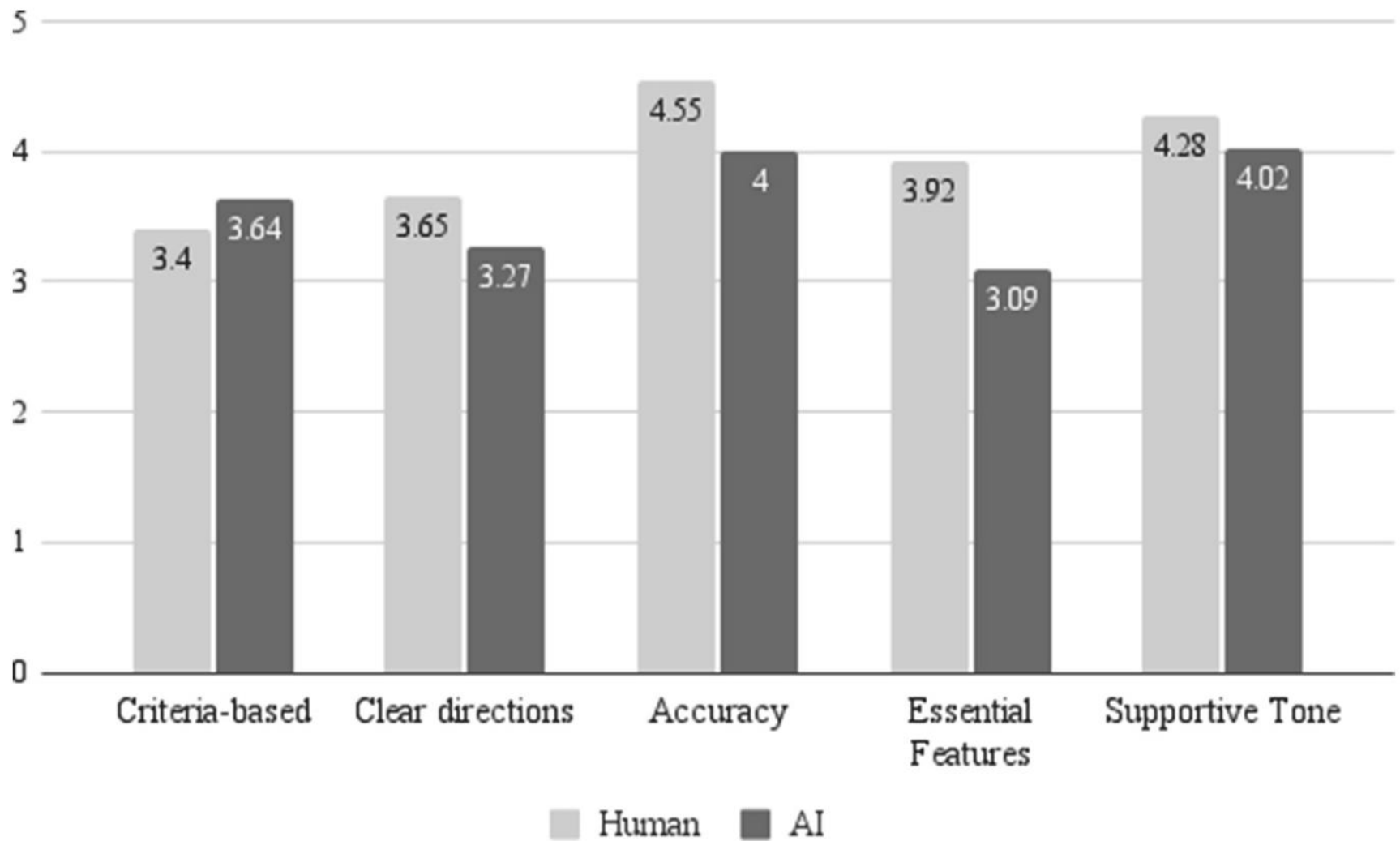
- What the researchers found

- 行動的エンゲージメントの観点からは、それぞれシステムの提案通りに修正しているが、認知及び感情的エンゲージメントの観点から考察すると、両者はやや異なるふるまいを見せた。ある実験参加者は、システムのフィードバックに納得して修正を行っているのに対し、別の実験参加者は、自身の回答も問題はないのではないかという意識を持ちながら、修正を行っていた。
- 作文自動評価システムのフィードバックへのエンゲージメントにおいては、**信頼**が中心的な役割を果たすことが、Ranalli (2021) より示唆されており、本研究の結果もそれを支持するような結果であった。

生成AIと自動採点

以下の英文はある学習者が書いたエッセイとフィードバックです。
フィードバックが2つありますが、どちらかは人間で、
どちらかはAIです。どちらがAIか考えてみましょう。

- Student Essay
- Why was the Montgomery Bus Boycott successful? Well the reason the boycott was successful was because Rosa Parks started to give flyers out to black people to stop using the buses. It did work because after a while the buses were not getting the money they used to get when the white and black people were there. Only the white were on the buses. The second way they got the boycott to be successful is, Rosa Parks was in the front of the bus but black people were in the back but she sat on the front which they were not supposed to be on. She...
- Feedback A
- Thank you for writing this essay on how the boycott succeeded! I think your claim, “Well the reason the boycott was successful was because Rosa Parks started to give flyers out to black people to stop using the buses” is interesting. But did Parks distribute the flyers? Better check that source again (I think it was Jo Ann Robinson and her students). Next time, you might use this sentence starter when adding evidence to support your claim, According to the “Bus Boycott Timeline,” _____. This will help you support your claim with clear evidence directly from the source. Your reader might also be convinced because you get your evidence from a source too. You are on the right track to becoming a great writer!
- Feedback B
- Great start! You have identified some key factors that contributed to the success of the Montgomery Bus Boycott. However, your essay could benefit from more detail and explanation. For example, you mention that Rosa Parks gave out flyers, but you don't explain what was on the flyers or how they helped the boycott. Try to provide more specific examples and evidence to support your argument. Additionally, it's important to note that Rosa Parks' refusal to give up her seat was not the only reason for the boycott's success. Can you think of any other factors that contributed to the boycott's success? Keep up the good work!



Steiss, J., Tate, T., Graham, S., Cruz, J., Hebert, M., Wang, J., ... & Olson, C. B. (2024). Comparing the quality of human and ChatGPT feedback of students' writing. *Learning and Instruction*, 91, 101894.

Target task

Feedback based on a checklist

- Automate checklist-based assessment by GPT-4
- In this experiment, we focus on checklist items related to topic sentences.

1 やってみよう2 学ぼう3 見直そう4 力試しをしよう

3. 要約解答の見直しと修正: Step 1で作成した要約解答を様々な観点から見直し、修正します。

Cycloclean

Can you imagine life without safe water for drinking, cooking, or washing? While lack of clean water may sound unfamiliar to many Japanese people, this is a real, life-threatening issue for people in other parts of the world. UNICEF (2016) reports that as many as 300,000 children under age 5 died from diarrhea worldwide due to the lack of access to safe water in 2015. Thus, we have a mission to find ways to supply water to areas where access to clean water is limited.

Now, a product developed by a Japanese company may offer a solution. In 2005 the company, Nippon Basic Co., Ltd. launched a product called Cycloclean, a bicycle loaded with a water purifier that only requires human power to produce drinking water. To use Cycloclean, you first pump water into a case attached to the rear seat from a variety of water sources such as rivers, ponds, and lakes. Then, by pedaling the bike, you can purify the water as it goes through filters inside the case. According to the company's website, one minute of pedaling the bike can produce as

アウトライン参照ハイライト参照メニューに戻る

② 要約ストラテジーが適切な箇所で作っているか確認します。チェックリストのうちでできている項目に☑マークを入れてください。できていない項目があれば、チェックリストの内容をヒントに修正してください。

People in some regions are suffering from lack of clean water for living. A product called Cycloclean may offer a solution. It is a special bicycle with a water purifier that can be operated without electricity.

Word count: 36 word(s)

★ チェックリスト

| | |
|-------------------------------------|--|
| <input checked="" type="checkbox"/> | P1の最後の文(“Thus, we have...is limited.”)はthesis statementなので、パラフレーズして要約に入れた。 |
| <input type="checkbox"/> | P2の最初の文(“Now, a product...a solution.”)はtopic sentenceなので、パラフレーズして要約に入れた。 |
| <input checked="" type="checkbox"/> | P3にはtopic sentenceがないので、このパラグラフの要旨を自分でまとめて要約に入れた。 |

Next

Topic sentences in the source text

Can you imagine life without safe water for drinking, cooking, or washing? While lack of clean water may sound unfamiliar to many Japanese people, this is a real, life-threatening issue for people in other parts of the world. UNICEF (2016) reports that as many as 300,000 children under age 5 died from diarrhea worldwide due to the lack of access to safe water in 2015. **Thus, we have a mission to find ways to supply water to areas where access to clean water is limited.**

Now, a product developed by a Japanese company may offer a solution. In 2005 the company, Nippon Basic Co., Ltd. launched a product called Cycloclean, a bicycle loaded with a water purifier that only requires human power to produce drinking water. To use Cycloclean, you first pump water into a case attached to the rear seat from a variety of water sources such as rivers, ponds, and lakes. Then, by pedaling the bike, you can purify the water as it goes through filters inside the case. According to the company's website, one minute of pedaling the bike can produce as much as 5 liters of drinking water, equivalent to the amount consumed per day by three adults. Thus, with Cycloclean people in remote villages and disaster zones can drink purified water as long as they have a water source nearby.

Despite its low-tech look, what this special bicycle can do is impressive. However, one problem is that only a small number of Cyclocleans were initially available. Another, related problem is that the product was quite expensive, at 550,000 yen per unit. For these reasons, many bicycles were bought by local governments in Japan, while only a small number was exported to foreign countries. More recently, however, with the increase of the demand of the product in Bangladesh, the company started to produce Cycloclean in the country by teaming up with a local bicycle company to offer a larger number at a more reasonable price. The product is in use in other surrounding Asian nations as well.

Because Cycloclean requires a water source to generate purified water, it cannot be used in dry areas. **However, this product may offer great hope to many people in areas worldwide where clean water is desperately needed, allowing us to take a step forward in improving public health in those areas.**

Checklist Items analyzed

1. I have paraphrased and included the topic sentence of paragraph 1, "Thus, we have a mission to find ways to supply water to areas where access to clean water is limited," in my summary.
2. I have paraphrased and included the topic sentence of paragraph 2, "Now, a product developed by a Japanese company may offer a solution," in my summary.
3. I have paraphrased and included the topic sentence of paragraph 4, "However, this product may offer great hope to many people in areas worldwide where clean water is desperately needed, allowing us to take a step forward in improving public health in those areas," in my summary..

Approach

- Describe the task to GPT-4, input a student summary, and let the GPT-4 predict the outcome.
- Input:
 - Task description
 - Source text
 - A student summary
- Output:
 - True or False for each checklist item
 - Additional natural language output explaining which part of the student's summary corresponds to a topic sentence.

Experimental settings

- Data:
 - Source text: “Cycloclean”
 - 15 student summaries composed by Japanese undergraduates
 - Score data:
 - Each summary is double-coded by two English teachers independently.
 - Evaluation results for the 3 checklist items are given as a score of either 0, 1, or 2.
 - 0 = false
 - 1 and 2 = True
- Settings
 - Evaluation metric: Accuracy (average across five runs)

Instructions for GPT-4 [1/2]

You are an experienced English teacher. You examine summaries written by non-native English speakers.

We tried to tell our task description as if we were talking to human colleagues.

Please check if the summary follows the rules below.

Rule_1: Whether a topic sentence of P4 is properly paraphrased and contained in the summary or not.

Response to Rule_1: (OK|NG). Reasons: ("{{text}}" in the summary corresponds to the topic sentence of P4. | There is no corresponding part to P4 in the summary).

We used a similar description to the one used in the feedback guideline written by an expert.

In the following [[[text]]] denotes topic sentence

Source:

[P1] Can you imagine life without safe water for drinking, cooking, or washing? While lack of clean water may seem like a simple problem, for many people, this is a real, life-threatening issue. In Japan, for example, access to clean water is taken for granted. However, in many parts of the world, this is not the case. UNICEF (2016) reports that as many as 2.2 billion people lack access to safe water. Diarrhea, a common disease caused by contaminated water, kills about 525,000 people worldwide each year. In 2015, 2.3 billion people did not have access to clean water. [Thus, we have a mission to find ways to supply clean water to areas where access to clean water is limited.]]]

We gave paragraph structure and topic sentences to the model.

Instructions for GPT-4 [2/2]

[P2] [[[Now, a product developed by a Japanese company may offer a solution.]]] In 2005 the company, Nippon Basic Co., Ltd. launched a product called Cycloclean, a bicycle loaded with a water purifier that only requires human power to produce drinking water. To use Cycloclean, you first pump water into a case attached to the rear seat from a variety of water sources such as rivers, ponds, and lakes. Then, by pedaling the bike, you can purify the water as it goes through filters inside the case. According to the company's website, one minute of pedaling the bike can produce as much as 5 liters of drinking water, equivalent to the amount consumed per day by three adults. Thus, with Cycloclean people in remote villages and disaster zones can drink purified water as long as they have a water source nearby.

[P3] Despite its low-tech look, what this special bicycle can do is impressive. However, one problem is that only a small number of Cyclocleans were initially available. Another, related problem is that the product was quite expensive, at 550,000 yen per unit. For these reasons, many bicycles were bought by local governments in Japan, while only a small number was exported to foreign countries. More recently, however, with the increase of the demand of the product in Bangladesh, the company started to produce Cycloclean in the country by teaming up with a local bicycle company to offer a larger number at a more reasonable price. The product is in use in other surrounding Asian nations as well.

[P4] Because Cycloclean requires a water source to generate purified water, it cannot be used in dry areas. [[[However, this product may offer great hope to many people in areas worldwide where clean water is desperately needed, allowing us to take a step forward in improving public health in those areas.]]]

Results and discussion

- Accuracy for Items 1 and 2 (84% and 87%, respectively) were relatively high compared to that for item 3 (67%).
- Sample outputs
- Limitations
 - Small test data
 - Single source text (lack of diversity in topics)
 - Simple instructions given to GPT-4
 - No exemplar given to GPT-4

| Checklist items | Accuracy |
|-----------------|---------------------------|
| 1 | 0.84 (.037) |
| 2 | 0.86 (.082) |
| 3 | 0.67 (.000) ⁴⁹ |

C2029 (successful)

Input Summary

A Japanese company produced Cycloclean, which is a bicycle that makes drinkable water by pedaling, and a one-minute pedaling equals to the amount needed for three adults. There were some problems at first, such as the amount of production and the cost. However, there was an increasing demand in Bangladesh, so the company worked with a local company to produce it more at an affordable price. It cannot be used in dry areas, but Cycloclean might help numerous people in the world.

Outputs

* Human coded labels and its original scores are shown in square

Topic sentence of P1

- False [False, 0]
- Reasons: There is no corresponding part to P1 in the summary.

Topic sentence of P2

- True [True, 1]
- Reasons: "A Japanese company produced Cycloclean" in the summary corresponds to the topic sentence of P2.

Topic sentence of P4

- True [True, 1]
- Reasons: "It cannot be used in dry areas, but Cycloclean might help numerous people in the world." in the summary corresponds to the topic sentence of P4.

C2060 (failed)

Input Summary

Cycloclean is a bicycle of that easily purifies much water by only pedaling. Although it was seemed that this invention helped people live in the country that is difficult to get safe water, it was introduced hardly because of the high price. To improve this situation, the company started local production of Cycloclean. This action led to reduce the price and Cycloclean was prevailed worldwide. It came to the key to solve the serious problem in the world.

Outputs

* Human coded labels and its original scores are shown in square

Topic sentence of P1

- **False** [True, 1]
- Reasons: **There is no corresponding part to P1 in the summary.**

Topic sentence of P2

- **False** [True, 1]
- Reasons: **There is no corresponding part to P1 in the summary.**

Topic sentence of P4

- **False** [True, 1]
- Reasons: **There is no corresponding part to P4 in the summary.**

C2050 (failed)

Input Summary

Clean water sources are needed in developing countries, so one of the solutions is suggested Japanese one product the name of “Cycloclean.” How to use of this item is really easy, so it should be the great way, but the cost prevents people needing water from getting this. It cannot generate water neither, so it is not perfect way.

Outputs

* Human coded labels and its original scores are shown in square

Topic sentence of P1

- False [False, 0]
- Reasons: There is no corresponding part to P1 in the summary.

Topic sentence of P2

- True [True, 1]
- Reasons: “Now, a product developed by a Japanese company may offer a solution.” in the summary corresponds to the topic sentence of P2.

Topic sentence of P4

- **True** [False, 0]
- Reasons: “**However, this product may offer great hope to many people in areas worldwide where clean water is desperately needed, allowing us to take a step forward in improving public health in those areas.**” in the summary corresponds to the topic sentence of P4.
- There is no corresponding part to P4 in the summary.

Copying the topic sentence of P4 here from the source text.

Conclusion

- Potential for delivering automatic feedback using large language models
 - Potential application of GPT-4 in content feedback generation for low-stakes purposes.
- The quality of generated feedback must be examined from multiple perspectives; Accuracy is not sufficient.
 - Meaningfulness and consistency of feedback (Bachman & Palmer, 2010)
- The usefulness of generated explanation for the feedback for stakeholders (students and instructors) should be examined.

人と共生するAIと自動採点

国語記述式読解問題

評論または小説を抜粋した問題文を読み、その内容について数十語程度で記述する

厳密に定められた採点基準に基づき、答案の内容の正しさを評価

問題

傍線部(1)「こうした～築いてきた」とあるが、それはどういうことか。70字以内で説明せよ。

採点基準（一部抜粋、要約）

- | | |
|----------|-----------------------------------|
| A | 言葉は他社との意思疎通を可能にする媒体であることが書けている…6点 |
| B | 言語には個別の経験を抽象化する作用があることが書けている…3点 |
| C | 世界との一体感が失われるということが書けている…6点 |

答案

言葉は様々な人とのやり取りを行う際の媒体となるが、その一方で

A 根拠箇所（6点）

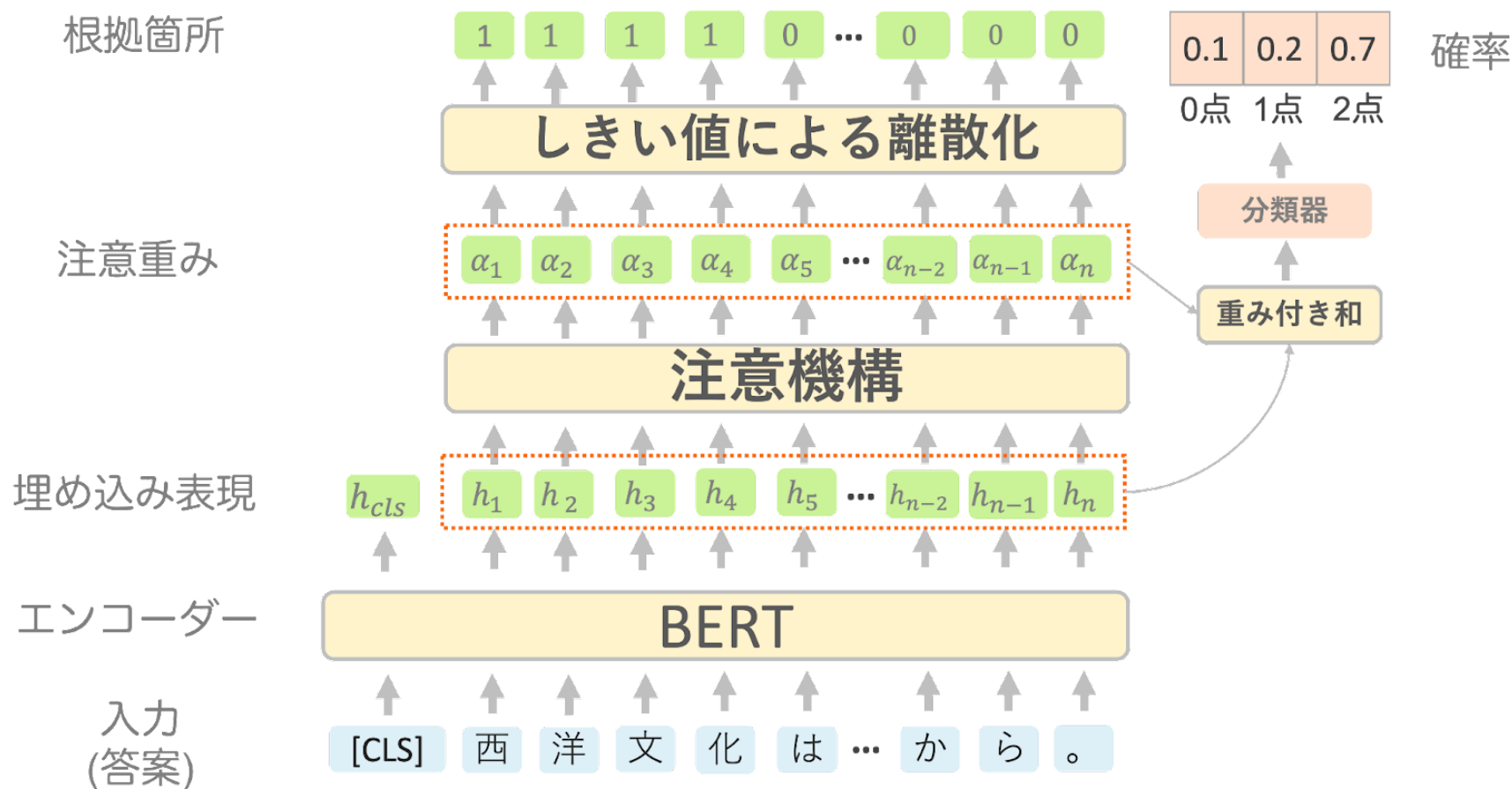
言葉は抽象的な記号でもあるので、世界との直接的な関わりを遮断してしまうこと。

B 根拠箇所（3点）

C 根拠箇所（6点）

MIZUMOTO et al. (2019) によって構築され、
FUNAYAMA et al. (2023) によって拡張された理研記
述式問題代ゼミデータより引用。
なお、採点基準は公開されていない。

自動採点システム (1/2)



自動採点システム（2/2）

合計点:
13 / 16

西洋文化の基底には「対決」のスタンスがあり、自分の考えを異質な他人を同意させるためには、言葉を尽くして説得する必要があるということ。
A B C D

| 採点項目 | 採点基準 | 採点結果 | 確信度 |
|--------|--|-------|-----|
| 採点項目 A | 採点基準 「西洋（では）」（＝話題の中心への言及）..... 2 点 | 2 / 2 | 99% |
| 採点項目 B | 採点基準 「他人を自分とは異質な考え方を持つ人間と見なす.....5点 内訳は、 ①<他人は自分とは違う>という説明..... 3 点、 ②異質な考え方を持つ..... 2 点 ※①の<違い>の中身の具体化 | 2 / 5 | 88% |
| 採点項目 C | 採点基準 「自分の意見に同意を得るために」という説明..... 3 点 | 3 / 3 | 99% |
| 採点項目 D | 採点基準 「言葉を尽くして他人を説得する」.....6点 内訳は、 ①「言葉を尽くして」..... 3 点、 ②「他人を説得する」..... 3 点 ※自分の考えを説明するなどは1点 | 6 / 6 | 99% |

AIの採点結果の信頼性を表す尺度である確信度を用いて
確信度が低い場合は人間の採点者が再採点を行う
→人間とAIの協調により記述式答案自動採点の品質を保証

メディア掲載：毎日新聞「現代文記述式問題をAIが採点 代ゼミが理研AIPと共同開発」

メディア掲載：日本経済新聞「記述式答案の採点AI、確信度を自己申告 人と協調」

実験参加者1（修正前）

西洋という項目が目に入ったので
主語が足りなかったという反省

異質な考え方を持つの方が足りなかったのかと思い、
異なる人間であるという考え方に含まれていると思って
いたので、なるほどなと思った

合計点:
9 / 16

基本的に他人は自分とは異なる人間であるという考え方があるため、自分の考え方を相手にきちんと説明して同意を取り付け説得しようとするということ。

採点項目 A

採点基準

「西洋（では）」（＝話題の中心への言及）..... 2点

0 / 2

確信度 86%

採点項目 B

採点基準

「他人を自分とは異質な考え方を持つ人間と見なす.....5点
内訳は、

3 / 5

確信度 99%

①<他人は自分とは違う>という説明..... 3点、

②異質な考え方を持つ..... 2点

※①の<違い>の中身の具体化

採点項目 C

採点基準

「自分の意見に同意を得るために」という説明..... 3点

3 / 3

確信度 99%

採点項目 D

採点基準

「言葉を尽くして他人を説得する」.....6点

内訳は、

3 / 6

確信度 99%

①「言葉を尽くして」..... 3点、

②「他人を説得する」..... 3点

※自分の考えを説明するなどは1点

言葉を尽くして自分の考えを伝えようとする説得であるって書いてあるから私は多分説得の話を書いたが、
言葉を尽くしてが必要なのかという気づきがあって、
そのまま直しました。

実験参加者1（修正後）

問: 下線部「こうした緊張のスタンスこそが饒舌な西洋文化を導いてきた」とあるが、それはどういうことか。句読点とも七十字以内で説明せよ。

西洋では他人は自分とは異なる考え方を持つ人間であるとみなすため、自分の意見に同意を得るために言葉を尽くして他人を説得しようとするということ。

70 字

提出

合計点:
16 / 16

西洋では他人は自分とは異なる考え方を持つ人間であるとみなすため、自分の意見に同意を得るために言葉を尽くして
A B C D
他人を説得しようとするということ。

半構造化インタビューの 結果と考察 (1/2)

- 「あなたはフィードバックに満足しましたか。また、それはどうしてですか。」
 - 「満足はしました。っていうのも特にまあ小説の方で項目CDの要素が確かになかったなって気付かされたりとかして。ただ問題1の評論の方で、採点項目Bが少し減点されてたんですけど、自分では表現したつもりだったので残念だった。(実験参加者2)」という回答が得られた。実験参加者2は、感情的エンゲージメントに関わる口頭産出を行った。具体的には、自分では表現したつもり項目が減点されたことに対して、納得できていない状況であった。このように、自信がある回答について減点などが生じると、学習者のエンゲージメントにおいて望ましくないことが生じる可能性についてはシステム開発において留意する点となる。
- 「根拠箇所が表示されることで、自動採点結果に対する信頼度は向上しましたか。また、根拠箇所の表示は回答を改善する上で役に立ちますか。その理由を教えてください。」
 - 「結構しました。役立ちました。ABCDのそれぞれの観点で何が不足していて何がちゃんと書かれているのかっていうのを根拠箇所を表示することで視覚的にわかりやすく理解できたので何が足りなくてどこまで削っていいのか、足りない要素を補うために文字数が必要になるんですけど、その文字数を確保するためにどこを削っていいのかっていうのが明示されていることで分かりやすかったので良かったです。(実験参加者4)」という回答が得られた。根拠箇所の表示は本システムの特徴の一つであり、学習者は採点結果に対してより深い理解を得ることが期待されている機能であるが、インタビュー調査の結果から、その効果が観察された。

半構造化インタビューの 結果と考察 (2/2)

- 「普段記述式問題の修正にはどのくらい時間をかけますか。今回その時間に変化はありましたか。その理由を教えてください。」
 - 「今回みたいにフィードバックが普通はないので、一旦書いたら要素、必要そうだなって思う要素を集めて文章にした後は文としておかしいところがないとか誤字脱字はないか文字数はちゃんとしているかっていうのを確認するのみなのであんまり直すのに時間は書けないです。今回はフィードバックでどこを直すべきかっていうのがきちんと示されていたのでそれに従って直していったのでいつもより記述問題の解答に時間がかかったかなと思います。(実験参加者4)」という回答が得られ、更に時間の変化については、「システムのほうが長いです。理由としては、詳しくというか項目とか基準とかを示してくれていたんで、直そうかなと思えるようになったというか。ここ直したらこんだけ上がるのかみたいなのがちゃんと分かるので意欲的に直そうとしました。(実験参加者5)」という回答も得られた。この質問項目については、行動的エンゲージメントと認知的エンゲージメントに関わる回答が得られた。普段記述式問題を修正する際には、一度読み直して、おかしいところがないか、誤字脱字はないかという点だけを確認するのに対し、フィードバックがあることで、学習者は意欲的に修正を試みた。
- 「このシステムの使いやすさについてどう思いますか。このシステムを使っているとき何か問題がありましたか。このシステムとそのフィードバックの強みと弱みはどのような点だと思いますか。」
 - 「自分ではこの基準を満たしていると思っていてもシステム上では違うって出た時に納得させられたらより良いのかなと思いました。(実験参加者1)」という回答が得られた。実験参加者1のインタビューデータからは感情的エンゲージメントに関わるコメントが得られた。学習者にとって自信がある回答に対して、システム上のフィードバックが納得いかない場合は学習意欲に影響が出る可能性がある。こうした問題に対し、本システムは、学習データから算出されたモデルの予測の信頼度合いを表す確信度を表示しているが、今回開発しているシステムでは少ない学習データを何度も学習しているため、やや高めの値が出る傾向があった。

事前に頂いた質問

テスト問題の作成（1/2）

- 本研究では、人間の専門家が作成したものと同等のL2評価教材を作成するためのリソースとしてChatGPTの可能性を検証することを目的とした。
- 韓国のCSAT（College Scholastic Ability Test）の英語セクションから、5つのリーディングパッセージと選択問題形式のテスト項目を抽出した。
- ChatGPTを使って、同じ形式のリーディングとテスト項目を作成した。次に、リッカート尺度の質問と自由回答形式の質問で構成される調査票を作成し、対象となるリーディングとテスト要素の多様な側面に対する参加者の認識を調査した。
- この調査の参加者は、50名の現職教師で構成され、対象教材の出典や研究の目的は知らされていなかった。
- 調査の結果、CSATとChatGPTが開発したリーディングは、対象パッセージの流れや表現の自然さにおいて類似していると認識された。しかし、CSATとChatGPTでは、CSATの方が魅力的な選択肢を含み、テスト項目の完成度が高いという結果が得られた。

テスト問題の作成 (2/2)

- ChatGPT (Chat Generative Pre-trained Transformer) の多肢選択式読解問題の生成能力を調査することを目的とした。
- 項目の品質を評価するために、心理測定モデルと人間によるレビューが採用され、人間が作成した項目のベンチマークが用いられた。
- その結果、ChatGPTで作成された多肢選択式読解力項目は、心理測定学的特性において、人間が作成した項目と同等であり、アンケート、専門家の判断、インタビューによる人間によるレビューにより、ChatGPTは読解教育・学習のためのテスト開発者・アシスタントとしての潜在的な能力があることがわかった。
- しかし、いくつかの欠点や潜在的な落とし穴も確認され、ChatGPTを教育目的の項目作成に適用する場合の改善の余地があった。

AIツールを活用した 文献検索やまとめなど

- Perplexity
- Consensus
- Elicit
- Connected Papers
- ResearchRabbit
- NotebookLM
- DeepResearch

質疑応答



本日の資料は以下のQRコードから
ダウンロード可能です

