

一人一人に寄り添う文章が 書けるAIを目指して

船井 正太郎

沖縄科学技術大学院大学 (OIST)

株式会社人々 / 慶應義塾大学SFC研究所

2021年12月18日 @ Tech BASE Okinawa

船井正太郎とは… 物理学者、だと思えます。

2005年 東京大学理学部物理学科 卒業

2010年 東京大学大学院で博士（理学）を取得

2015年 物理の研究を続けながら

AI（機械学習）の研究を開始

「グーグルの猫」に興味を持って…

2018年 沖縄科学技術大学院大学（OIST）に所属

画像の機械学習に本格的に取り組む

2020年 株式会社人々のAIシステム開発に参加

（初めは顧問、後に役員として）

言語の機械学習にも取り組むことに

2021年 慶応義塾大学SFC研究所にも籍を置き

AIを応用した社会システムの研究にも参加

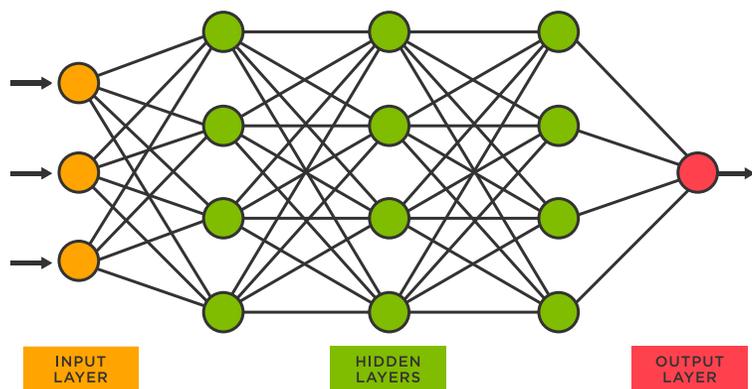
超弦理論という
高次元時空の物理学を
研究しています



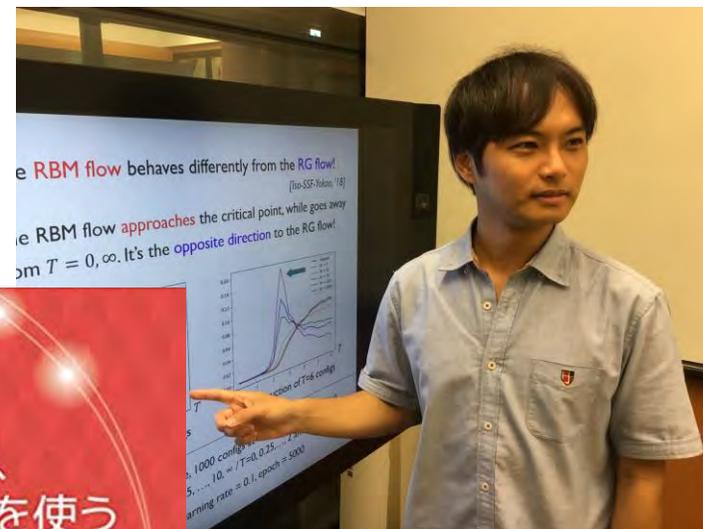
お金のあり方に
興味を持っています

船井正太郎の研究テーマを一言でいうと…

AI（機械学習）が数値化して捉えている、
データ（情報）の「特徴」とは何なのか、
物理学の概念を使って理解したい



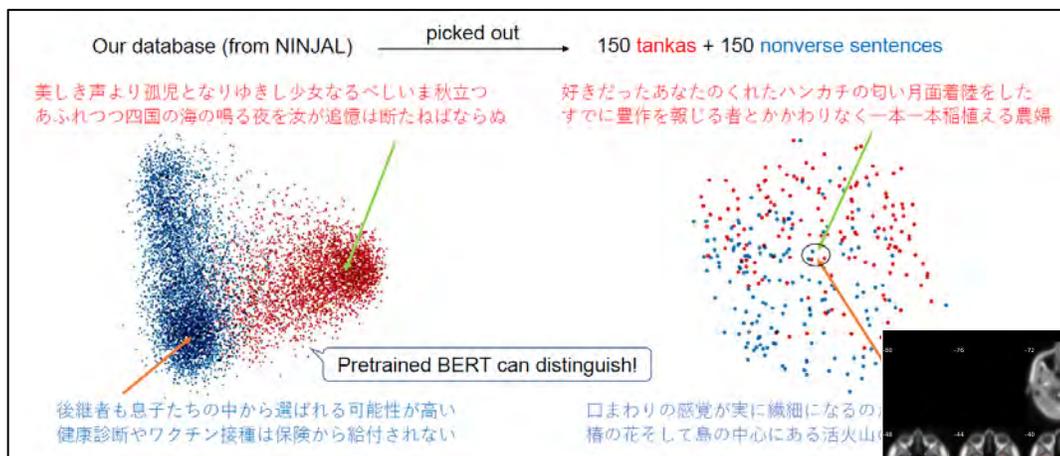
<https://www.tibco.com/>



AIを脳や意識のモデルと見做すことができれば、
「意識の物理学」が創れるかもしれない！

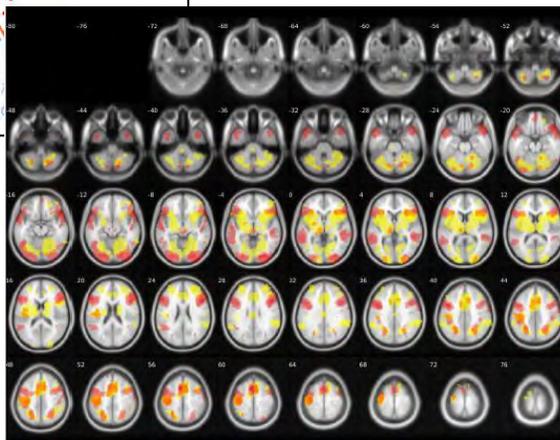
研究内容その1： AIは短歌を鑑賞できるか？

(機構間連携での研究)



短歌をAIに入力したときの
反応（捉えた特徴）を解析

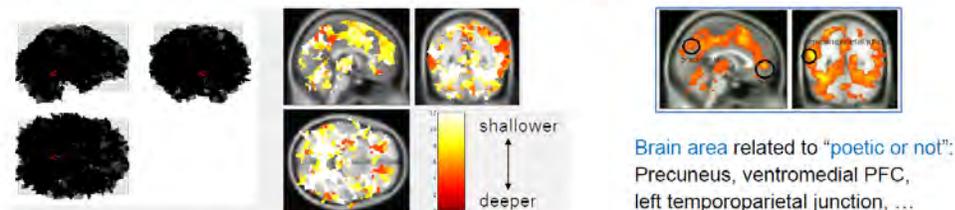
短歌を読む人間の脳活動を測定
AIの反応と比較すると類似点が見えてきた



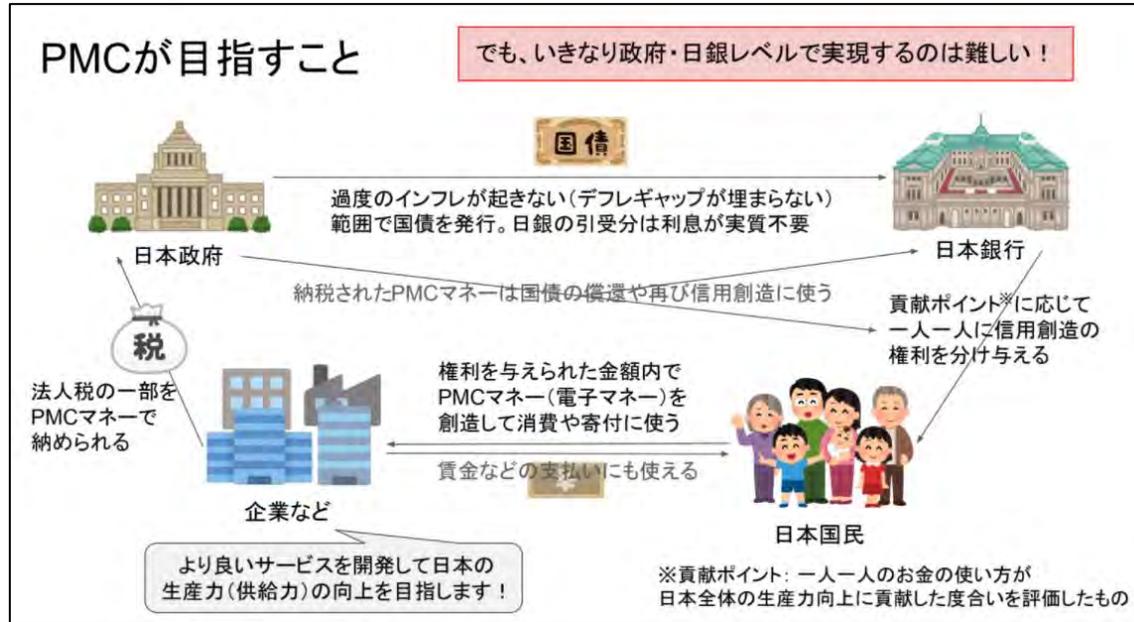
ご興味のある方はYouTubeで
見てみてください！

https://youtu.be/wP0RAWF6_h0

at deeper layers of BERT correspond to brain area correlated with
t (poetic or not) but found only weak correspondence.



研究内容その2： AIを用いた新しいお金の仕組み (慶應義塾大学SFC研究所での研究)



「失われた20年」の原因は行き過ぎた緊縮財政！
でも、単にお金をバラまけば解決するのか…？
新しいお金の仕組み (PMC、個人信用創造) で
皆で豊かになれる社会を目指そう！

松田政策研究所での対談・鼎談



作家さとうみつろうさんとの対談



実現に向けて一緒に動いてくださる
方々を募集しています！

台湾で活動中
2022年春に日本で活動開始！

研究内容その3：株式会社人々での研究開発（本題）

中国語チャット運用とAI開発の専門チーム

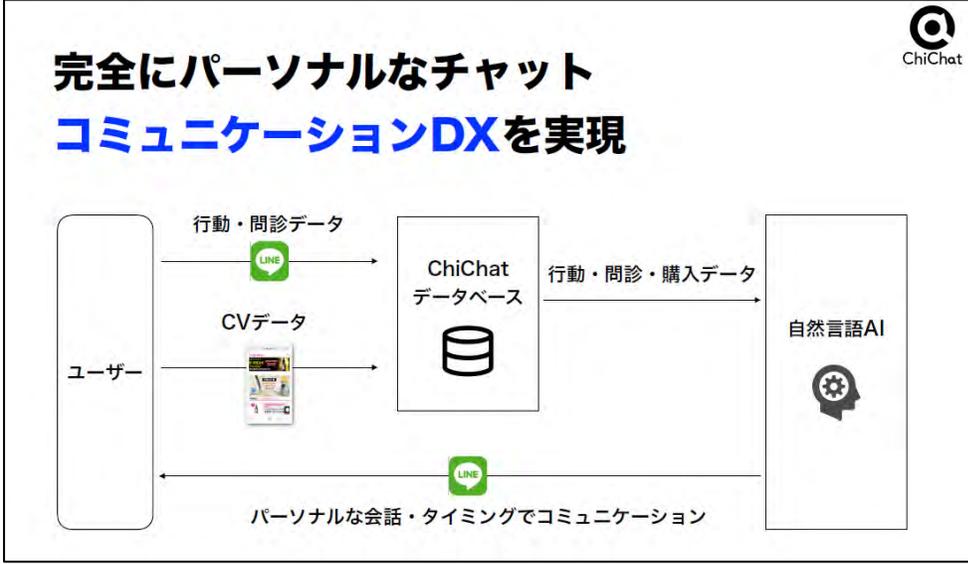


PM	中国語監修	AI設計	機械学習実装	データ整備
				
CEO 石川 真也	CCO 曾 雅淳	CDO 船井 正太郎	機械学習エンジニア 田村 健人	DBエンジニア 西川 雄三

CEO石川さんの目標は...

一人一人のユーザーに寄り添って、気持ちよく商品を買ってもらえるような、そんな文章が書けるAIが作りたい！

普通に考えると...



なぜならば、
自動文章生成の最先端とされるOpen AIのGPT-3でも…

Title: United Methodists Agree to Historic Split
Subtitle: Those who oppose gay marriage will form their own denomination
Article: After two days of intense debate, the United Methodist Church has agreed to a historic split - one that is expected to end in the creation of a new denomination, one that will be "theologically and socially conservative," according to The Washington Post. The majority of delegates attending the church's annual General Conference in May voted to strengthen a ban on the ordination of LGBTQ clergy and to write new rules that will "discipline" clergy who officiate at same-sex weddings. But those who opposed these measures have a new plan: They say they will form a separate denomination by 2020, calling their church the Christian Methodist denomination.
The Post notes that the denomination, which claims 12.5 million members, was in the early 20th century the "largest Protestant denomination in the U.S.," but that it has been shrinking in recent decades. The new split will be the second in the church's history. The first occurred in 1968, when roughly 10 percent of the denomination left to form the Evangelical United Brethren Church. The Post notes that the proposed split "comes at a critical time for the church, which has been losing members for years," which has been "pushed toward the brink of a schism over the role of LGBTQ people in the church." Gay marriage is not the only issue that has divided the church. In 2016, the denomination was split over ordination of transgender clergy, with the North Pacific regional conference voting to ban them from serving as clergy, and the South Pacific regional conference voting to allow them.

Figure 3.14: The GPT-3 generated news article that humans had the greatest difficulty distinguishing from a human written article (accuracy: 12%).

[\[2005.14165\] Language Models are Few-Shot Learners \(arxiv.org\)](https://arxiv.org/abs/2005.14165)

特定の目的に沿った文章を書くのは困難！



AIに大量の文章を入力すると、それらの特徴を掴んで、学習をしていく。

その学習結果に基づいて、与えられた文章の続きを「それらしく」書いていく。

人間が読んで自然と感じる文章が書けるかどうか議論の焦点。自然とは感じられない文章を書くこともまだまだ多い。

でも、解決策があるとすれば…

まず、文章の目的を限定すること（化粧品や健康食品の販促）

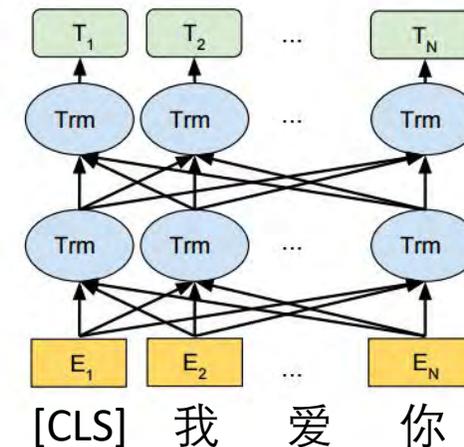
その上で、文章処理の最先端とされるGoogleのBERTと組み合わせて、新しいAIシステムを作ること！



<https://analyticsindiamag.com/>

1. 単語だけでなく「文全体」の特徴を捉えて、高次元ベクトルとして表現してくれる
2. 自然な文章は書けないが、単語の空欄補充は得意

文全体の特徴を表すベクトル



<https://medium.com/@cdathuraliya/>

```
input: ['[CLS]', '吾', '##輩', 'は', '[MASK]', 'で', 'ある', '。', '。', '名前', 'は', 'まだ', '無い', '。', '。', '[SEP]']
[CLS] 吾輩は猫である。名前はまだ無い。[SEP]
[CLS] 吾輩は犬である。名前はまだ無い。[SEP]
[CLS] 吾輩は人間である。名前はまだ無い。[SEP]
[CLS] 吾輩は狼である。名前はまだ無い。[SEP]
[CLS] 吾輩は私である。名前はまだ無い。[SEP]
```

日本と台湾で特許出願中

GPTとBERTをどう組み合わせたら良いか？

1. まずは、AIに書いてほしい文章のお手本（学習データ）を用意して、GPTとBERTに学習してもらいます。



弊社社員が作成した文章



影響力のあるブログ等の文章

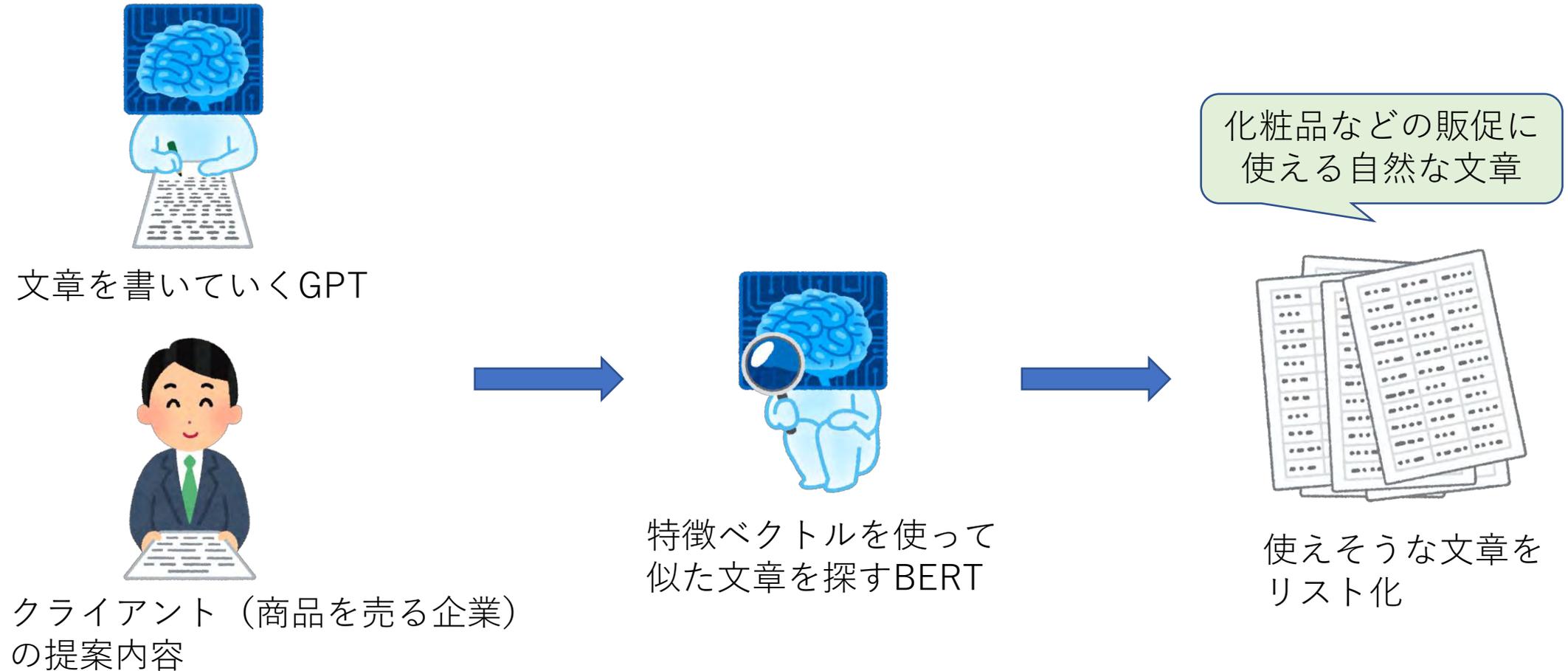


大量の学習データ

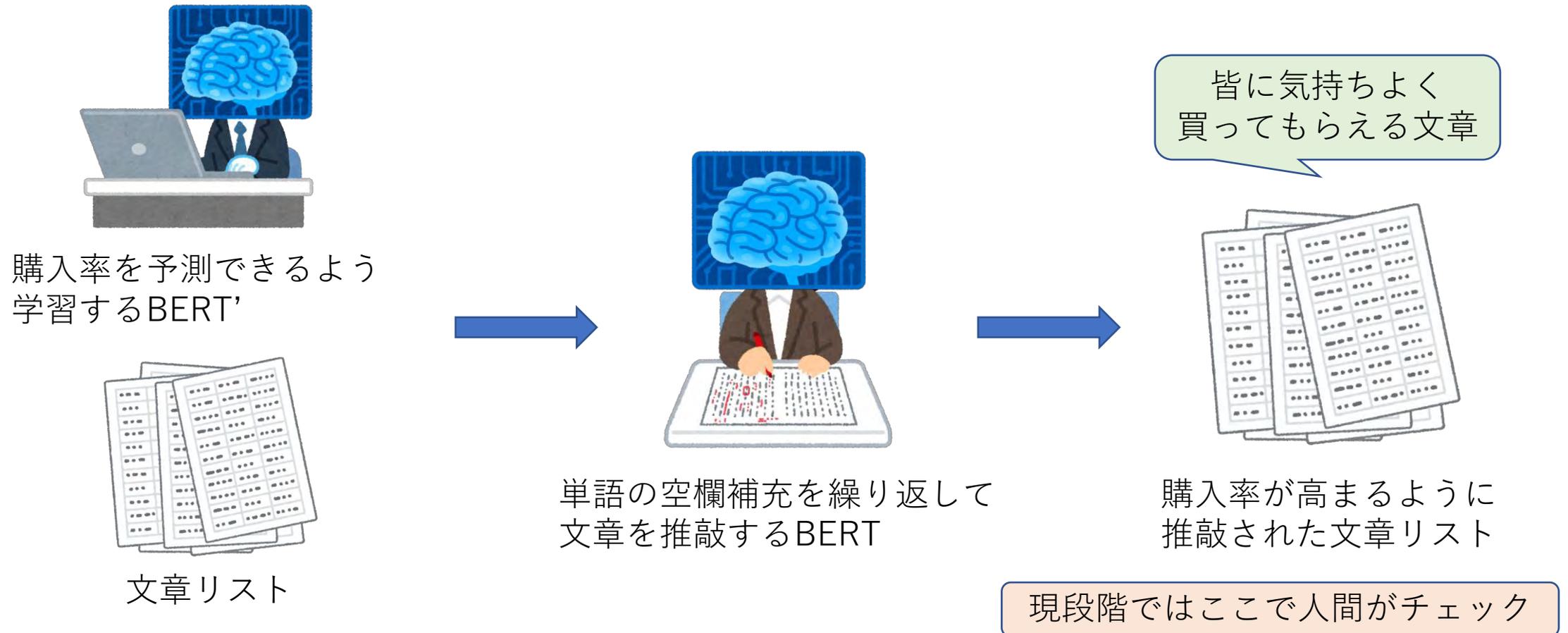


GPTとBERTが学習

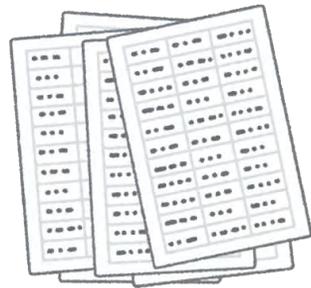
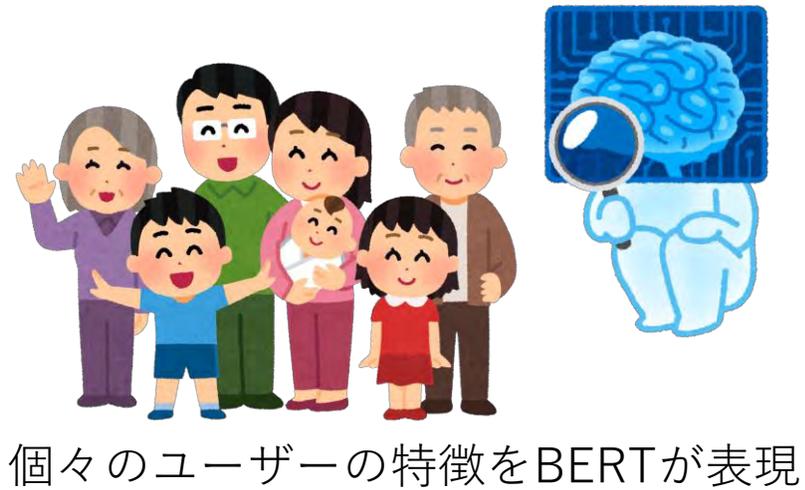
2. 学習済みのGPTに適切な書き出しを与えて、たくさん文章を書いてもらいます。
その中でクライアントの提案内容に近い文章をBERTに選んでもらい、リストにします。



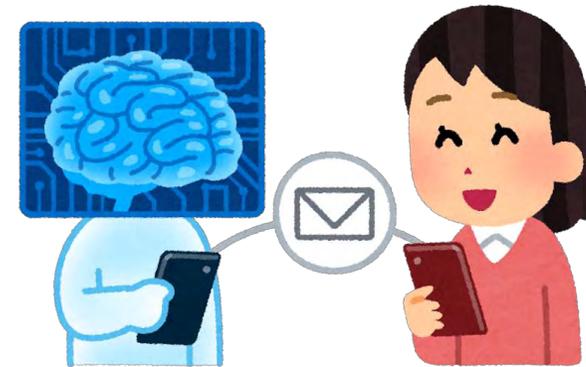
3. 文章を学習したBERTとは別に、弊社作成の文章を受け取ったユーザーの購入率(CVR)のデータを用意して、購入率が予測できるようBERT'に学習してもらいます（転移学習）。BERT'が予測する購入率が高まるように、BERTに文章を推敲してもらいます。



4. 個々のユーザーが、過去にどんな選択肢（悩みや習慣などのキーワード）を選んだか、どんな文章で購入したかを、BERTの特徴ベクトルを使って分析します。その結果に基づいて、リストの文章をそれぞれ最適なユーザーに届けます。



一人一人のユーザーに
寄り添った文章



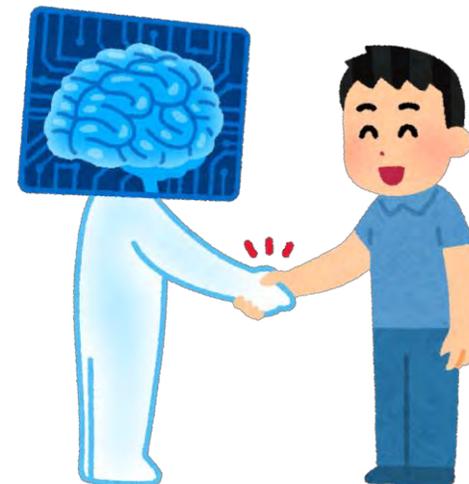
最適なユーザーに文章を届けます

最先端のAIたちを自在に組み合わせることで、独自のAI技術を開発しています！

残念ながらAI開発で遅れを取っている日本から、ユニークな技術を発信したいと思います。

一緒に開発してくれるエンジニアを随時募集しています。もしご興味があれば、お声がけください！

shotaro.funai@hitobito.co.jp



情報革命で誰もが国境を越えて

活躍できる世の中に

日本発のAI技術で世界を切り拓く