

# 生成AIと人間が書いた文章の構造 とそれぞれから受ける印象の違い

神奈川県立厚木高等学校78期  
2年H組2班  $\alpha$

# 目次

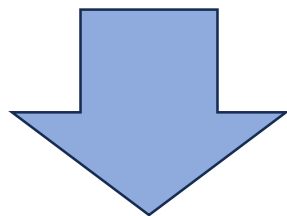
- 1 概要
- 2 実験1
- 3 実験2
- 4 考察
- 5 追加実験
- 6 結論・今後の展望
- 7 参考文献・文章引用元・使用ソフト

1

概要

# [1]背景

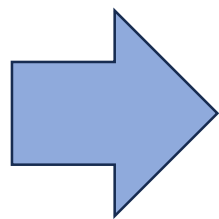
課題等でChatGPT等の生成AIを使用する  
学生が多く存在する



**人間が学習し文章等で表現する意味とは何か**

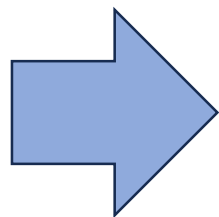
## [2]先行研究

生成AI



安全でバランスの取れた、  
中立的な文章を生成する  
傾向あり<sup>3</sup>

人間



具体的で、法律の規定や  
論文などソースの詳細な  
引用を含む傾向あり<sup>3</sup>

### [3]目的

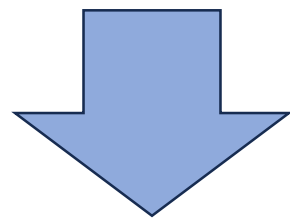
生成AIと人間の文章の相違点を見つけ  
「人間が学習し、表現する意味」を発見する

2

# 実験1

# [1]仮説

生成AIと人間が作る文章の「**品詞**」に注目する



文字数が近く、同じテーマについて述べた生成AIと人間の文章を比べた場合、使用されている**品詞の割合に有意差**があるのではないか



## [2]実験方法1

### ①人間の文章用意



3種類の書籍やサイトから、事象や用語等についての500文字程度の説明文を人間の文章として引用する

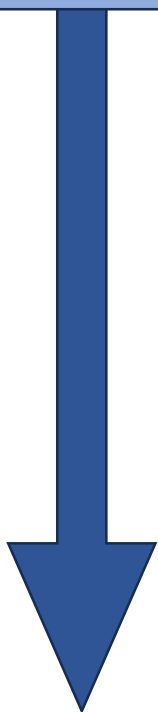
〈図1:人間の文章の例〉



ハンガリー生まれのフランツ・リスト(1811~1886)は、驚異的な技巧をもつピアニストであり、独創性に富んだ作曲家であった。彼こそ独奏ピアノリサイタルを一般的なものにした最初の人物であり、卓越した技巧を披露して、聴衆を感嘆させた・・・

## [2]実験方法2

### ②生成AIの 文章用意



ChatGPTに①と同じテーマについて同等の文字数で説明する文章を同じく3つ生成する

〈図2:ChatGPTの文章の例〉

フランツ・リスト（**Franz Liszt, 1811-1886**）は、**19世紀**ハンガリー出身の作曲家兼ピアニストであり、ロマン派音楽の重要な先駆者である。彼の音楽はその革新性と豊かな表現力で知られ、特にピアノ曲において革新的な作風を築いた。  
『ハンガリー狂詩曲』や・・・



## [2]実験方法3

### ③形態素解析

データ解析ソフトMeCabを用いて  
①と②で用意した文章の形態素解析を行い、品詞データを集める

### ④データ検証

品詞の出現頻度で有意差があるか  
 $\chi$  二乗検定で調べる

〈図3:  $\chi$  二乗検定の一部〉

結果

保存 コピー 消去 タブ変換

「カイ二乗検定の結果」  
(上段実測値, 下段期待値)

108	110
110.023	107.977
-----	
56	64
60.563	59.437
-----	
50	32
41.385	40.615
-----	
19	22

Rプログラム

保存 コピー 消去

#### i x j (カイ二乗検定)

```
## js-STARからの入力
levl = 6 # iの水準数
levJ = 2 # jの水準数

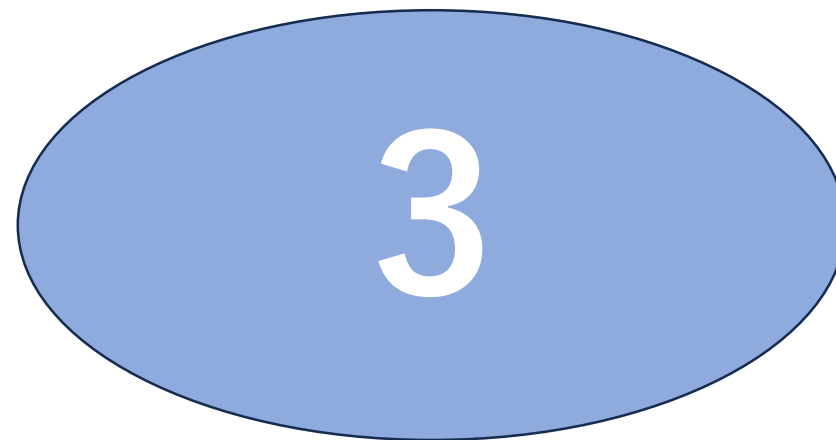
dosu <- c(
  108, 110,
  56, 64,
  50, 32,
  19, 22,
  30, 26,
  6, 10 ) # 度数
```

# [3]結果

帰無仮説「各文章の各品詞カテゴリの  
度数に偏りはない」が立証された

〈表1:文章に含まれていた品詞の割合〉

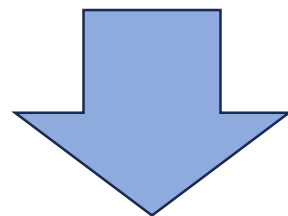
	人間	ChatGPT
名詞	44.8%	42.8%
助詞	21.7%	25.3%
記号	15.2%	11.6%
助動詞	5.3%	5.1%
動詞	10.6%	12.0%
その他合計	2.4%	3.0%



# 実験2

# [1]仮説

生成AIと人間が作る文章の  
「**わかりやすさ**」に注目する



アンケートを取った場合、それぞれの文章に対して**わかりやすいと答える人数**に有意差がある

## [2]実験方法

### ①アンケート作成

実験1で引用・作成した6つの文章を使用し、どちらの作成した文章がよりわかりやすいかを尋ねるアンケートを作成する

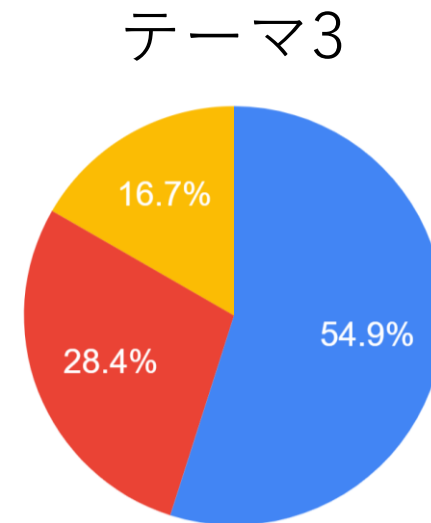
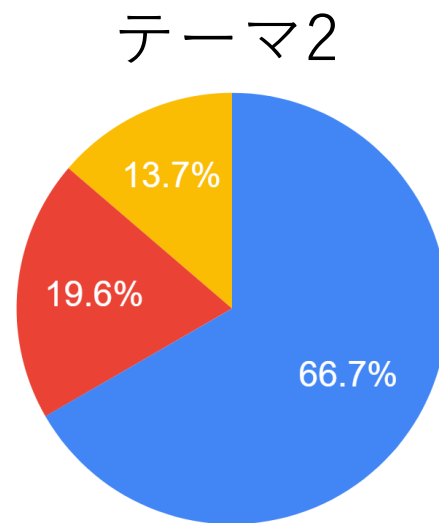
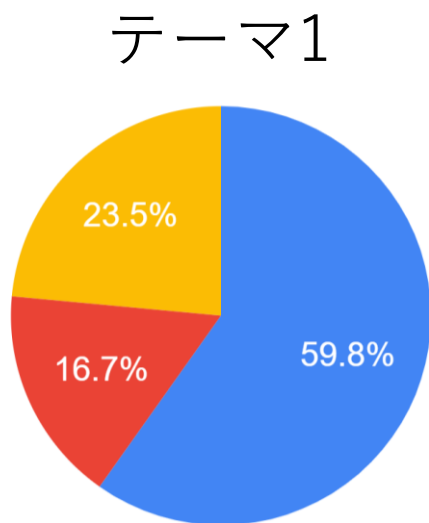
### ②データ検証

質問の結果から、それぞれの文章に対してわかりやすい、と答えた人数を二項検定にかけてわかりやすさに有意差があるか検証する

## [3]結果

3つすべてのテーマで**生成AIの文章の方がわかりやすい**  
と答える人が統計的に有位に多かった

〈図4-6:どちらの文章がより分かりやすいかという質問の回答(%)〉



赤→人間 青→ChatGPT 橙→どちらも変わらない



4

考察

# [1]実験1の考察

使用されている品詞の割合に  
大きな差はないと考えられる

## [2]実験2の考察

生成AIの文章の方が  
わかりやすいと考えられる

アンケート  
記述欄より

わかりやすいと感じた理由として「生成AIの文章は一般的な語彙が多い」、「人間は文章の構成や使用する語彙に特徴がある」と述べる人がいた

### [3]新たな仮説

人間の文章は生成AIの文章よりも  
構成や使用する語彙など表現の幅が広い

5

# 追加実験

(文章の難易度を測る)

# [1]実験方法

## ①難易度判定



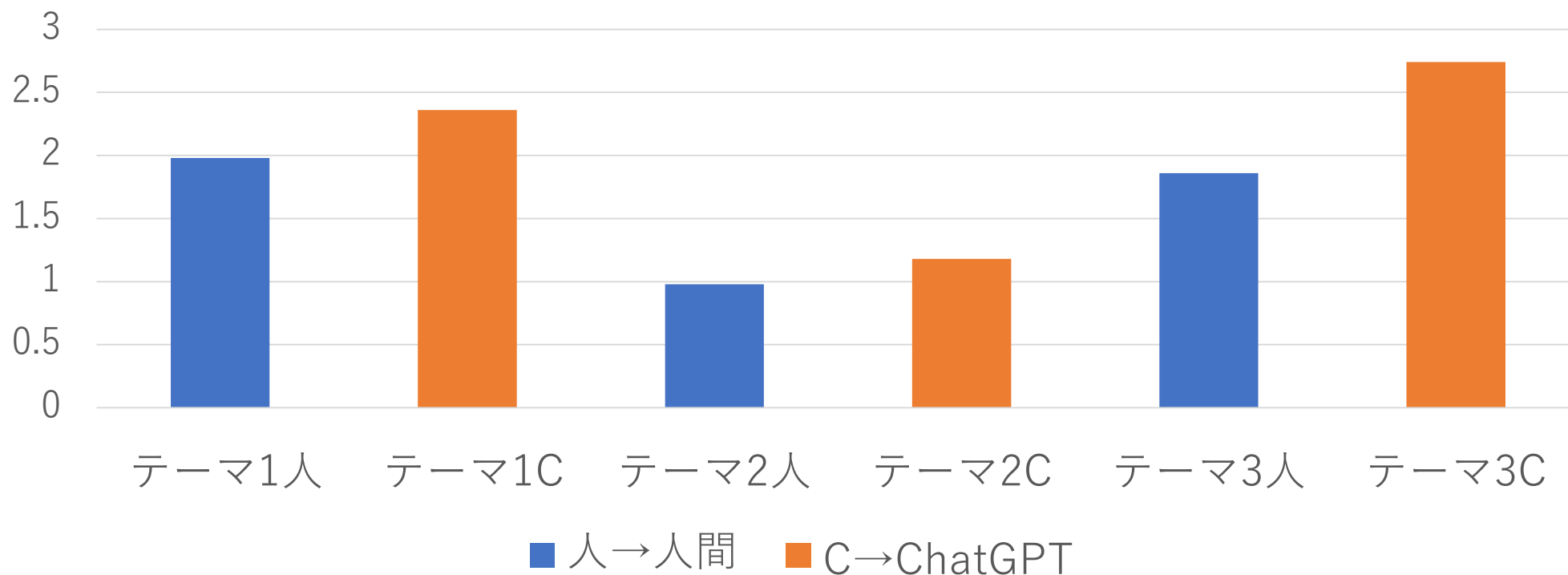
実験1と2で使用した文章を日本語文章難易度判定システム

「jReadability Portal」でリーダビリティスコア(高いほど読みやすい)として読みやすさを数値化し、テーマごとに比較する

## [2]結果

全てのテーマにおいて、AIの文章の方が人間の文章よりもリーダビリティスコアが高い結果となった

〈図7:リーダビリティスコア一覧(単位無し)〉



### [3]考察

AIの文章の方が様々な人が読みやすく、より一般的な語彙で作成されている可能性が高い



## [4]リーダビリティスコアの判定から見た考察の要因

リーダビリティスコアの判定要素である、総文数や1文あたりの平均語数に差あるため、これらがわかりやすさに影響する可能性があると考えた

〈表2:文章ごとの総文数と1文あたりの平均語数〉

	テーマ1 人間	テーマ1 ChatGPT	テーマ2 人間	テーマ2 ChatGPT	テーマ2 人間	テーマ2 ChatGPT
総文数	7	11	8	11	11	18
1文あたりの 平均語数	37.14	25.27	39.5	30.09	29.09	22.67



6

**結論・今後の展望**

# [1]結論

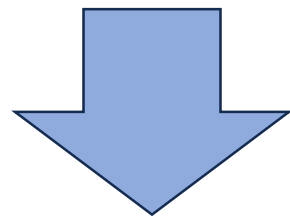
## ○生成AIの文章の方がわかりやすい

→各品詞の度数とは関連がなく、総分数や1文あたりの平均語数等が影響している可能性がある

○人間の文章の方がそれぞれ特徴的な構成や語彙を持ち、そのような**個性が見られることが人間の学習し、表現する意味の一つ**ではないだろうか

## [2]今後の展望

今回は難易度の高い説明文に注目した



小中学生向けの文章や日常会話においてのわかりやすさの違いを判定し、**人間の方が優れている場面や分野の特定**を目指す



7

**参考文献・文章引用元・使用ソフト**

# [1]参考文献

[1]いちばんやさしい、医療統計 2024年5月22日閲覧  
残差分析とは？使い方やエクセルでの計算方法をわかりやすく解説！

<https://best-biostatistics.com/contingency/zansa-bunseki.html#i-2>

[2]統計Web 2024年5月22日閲覧

27-4 予測値と残渣

<https://bellcurve.jp/statistics/course/9704.html>

[3]山下裕穀 ITmedia 2024年5月28日閲覧

ChatGPTと人間の違いは？約4万件の質疑応答から中国の研究者らが分析

<https://www.itmedia.co.jp/news/amp/2301/30/news058.html>

## [2]文章引用元

テーマ1「フランツ・リスト」

『楽器 歴史、形、奏法、構造』 監修 皆川達夫(1995) マール社

テーマ2「後期高齢者医療制度」

『用語集 医療』 WAM NET

<https://www2.wam.go.jp/content/wamnet/pcpub/iryo/handbook/dictionary/>

テーマ3「刺胞動物」

『動物学の百科事典』 日本動物学会(2018) 丸善出版

# [3]使用ソフト

[1]ChatGPT <https://openai.com/ja-JP/chatgpt/overview/>

[2]jReadability Portal <https://jreadability.net/>

[3]js-STAR\_XR+ <https://www.kisnet.or.jp/nappa/software/star/>

[4]MeCab <https://taku910.github.io/mecab/>

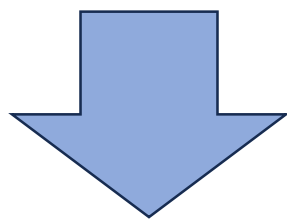


8

質問回答用スライド

# [1]品詞の数に着目した理由

初めてAIで作成した文章を見たとき、人間の文章と形容詞や指示語、形容動詞の数で差があるように感じた



品詞の数に有意差があると考えた

## [2]書籍の基準

### ①文字数が500文字程度

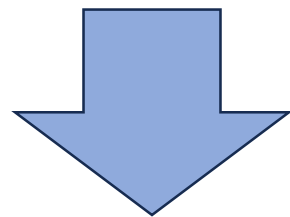
→実験1の形態素解析で、ある程度品詞の数を確保する必要があるため

### ②高校生が知らないレベルの単語の説明文

→実験2の対象が厚高生のため

### [3]文章のテーマは具体的に何か

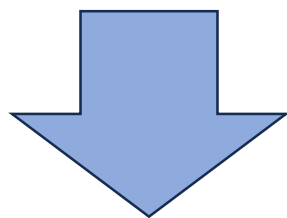
高校生が知らないような単語をテーマに設定



- ① フランツ・リスト(偉人・音楽家)
- ② 後期高齢者医療制度(福祉)
- ③ 刺胞動物(生物)

## [4]リーダビリティスコアとは

式：
$$\frac{\{\text{平均文長} - 0.056\} + \{\text{漢語率} - 0.126\} + \{\text{和語率} - 0.042\} + \{\text{動詞率} - 0.145\} + \{\text{助詞率} - 0.044\} + 11.724}{\text{単位のない値}}$$



文章の難易度を初級から上級に分類できる

## [5]実験で使った文章は何級か

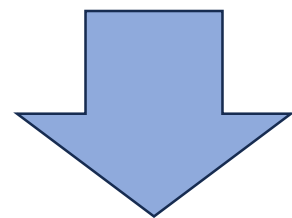
〈表3:リーダビリティスコアによる文章の階級〉

テーマ1 人間	テーマ1 ChatGPT	テーマ2 人間	テーマ2 ChatGPT	テーマ3 人間	テーマ3 ChatGPT
上級前半	上級前半	上級後半	上級後半	上級前半	中級後半

リーダビリティスコアは全てのテーマでChatGPTの方が高い(読みやすい)と判定されるが、階級で示すと同レベルと判定されるテーマも存在する

## [6]先行研究の結果はどのようにして出されたか

様々なテーマにおいて、ChatGPTに対し質問を行い、同じ質問の人間の回答を4万件収集



研究者200名超がそれぞれの特徴を比較した