

# 教師ありインストラクション学習を中心とした 大規模言語モデルの作り方 ～国語教師や日本語教育者が大活躍～

【第73回】 大学等におけるオンライン教育とデジタル変革に関する  
サイバーシンポジウム「教育機関DXシンポ」



理研AIP 関根聡

# 本プロジェクトは下記のメンバーを中心に進めております

- 関根聡 (AIP)
- 乾健太郎 (東北大・AIP)
- 河原大輔 (早稲田大・AIP)
- 井之上直也 (北陸先端・AIP)
- 鈴木久美 (AIP)
- 安藤まや (AIP)
- 後藤美知子 (AIP)
- 15名程度のベテランアノテーター (AIP)
- 25名程度の新規アノテーター(AIP)

今年3月までMSのBing対話型検索の  
プロダクトマネージャー

日本語学、日本文学、英米文学  
出身の方が半数以上

バックグラウンドは様々

# 自己紹介

# 関根聡

[satoshi.sekine@riken.jp](mailto:satoshi.sekine@riken.jp)

- 1987 東工大理学部卒業、松下電器情報通信研究所入所
- 1992 マンチェスター工科大学修士号取得（計算言語学）
- 1994 ニューヨーク大学助手／博士課程入学
- 1998 ニューヨーク大学博士号取得、助教授／准教授
- 2000～ ランゲージクラフト設立（約40社との技術指導など）
- 2010～2014 楽天技術研究所ニューヨーク 所長
- 2017～ 理研AIPの言語情報アクセス技術チーム チームリーダー  
主に、構造化知識構築プロジェクト「森羅」を推進

その他、ソニーCSL、MSRでの訪問研究員

言語処理学会理事、情報処理学会NL研主査、その他多数の役職を歴任

現在、複数の企業において技術顧問



[https://aip.riken.jp/labs/goalorient\\_tech/lang\\_inf\\_access\\_tech/](https://aip.riken.jp/labs/goalorient_tech/lang_inf_access_tech/)



<https://www.criprof.com/magazine/2019/01/10/post-5551/>



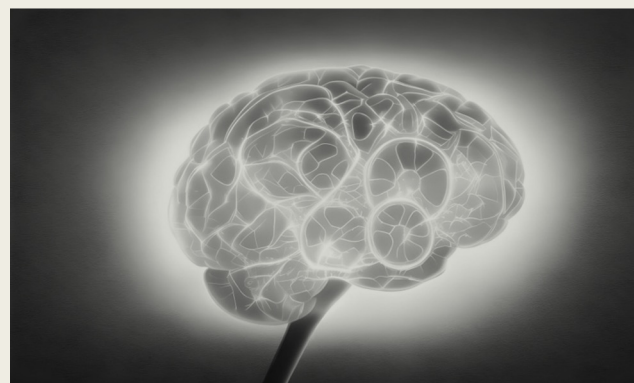
# 大規模言語モデルと インストラクションデータの重要性

# 大規模言語モデルの作り方

大量のテキスト  
(1T->13T token)  
日本語Wikipedia: 3.5B  
大規模パラメーター  
(1.7B->175B->1.8T)  
大規模GPU



RLHF  
(Reinforcement Learning  
from Human Feedback)  
Reward Model



Pre-training model



Fine Tuning model



Final model

# インストラクションの例



質問	回答
仕事に情熱を取り戻す 5つのアイデアを教えてください	<p>この質問からすると、質問者は仕事の情熱を失ってしまった状態のようですね。その理由によって取り戻す方法は変わるとは思いますが、一般的な仕事に情熱を取り戻す方法について5つばかり挙げてみます。</p> <ol style="list-style-type: none"><li>1) 初心を思い出す 誰でも仕事を始めた時には、その仕事に対して情熱を持っていたと思います。しかし時間が経ち、仕事の内容がわかってくると情熱を失うことはよくあります。そこで、現在の仕事への考え方を一旦リセットし、初心を思い出すことで情熱を取り戻すことができるかもしれません。</li><li>2) 仕事で喜ぶ人の顔を思い浮かべる どんな仕事でも、その仕事をする中で誰かの役に立ち誰かが喜んでくれることにつながっているはずですよ。その人の喜ぶ顔を思い浮かべると仕事に情熱を取り戻すことができるかもしれません。</li><li>3) 一緒に働く仲間のことを思う 多くの仕事は一人では成し遂げられません。様々な仲間と一緒に達成しているものだと思います。その仲間もその仕事が辛くとも一生懸命仕事に取り組んでいるものだと思います。そのような仲間のことを思い浮かべることで仕事に情熱を取り戻すことができるかもしれません。</li><li>4) 仕事は世界を変えること どんな仕事も何かの形で世界を変え、より良い世界を作っていくことにつながっているはずですよ。あなたのその働きが世界を変えていることを思うことで仕事に情熱を取り戻すことができるかもしれません。</li><li>5) 収入を考える 仕事をしている以上、その対価として収入を得ていると思います。その金額は少ないかもしれませんが、それでもあなたが自由に使えるお金です。そのお金を使って手に入れたり、成し遂げられることを思うことで、仕事に情熱を取り戻すことができるかもしれません。</li></ol>

# 2つの重要な論文 (その1)



68ページ

## InstructGPTの論文

Training language models to follow instructions with human feedback  
(L. Ouyang+ 20 authors, OpenAI)

arXiv:2203.02155v1 [cs.CL] 4 Mar 2022

arXiv:2203.02155v1 [cs.CL] 4 Mar 2022

---

### Training language models to follow instructions with human feedback

---

Long Ouyang\* Jeff Wu\* Xu Jang\* Diogo Almeida\* Carroll L. Wainwright\*  
Pamela Mishkin\* Chong Zhang Sandhini Agarwal Katarina Slama Alex Ray  
John Schulman Jacob Hilton Fraser Kelton Luke Miller Maddie Simens  
Amanda Askell<sup>1</sup> Peter Welinder Paul Christiano\*<sup>1</sup>  
Jan Leike\* Ryan Lowe\*  
OpenAI

**Abstract**

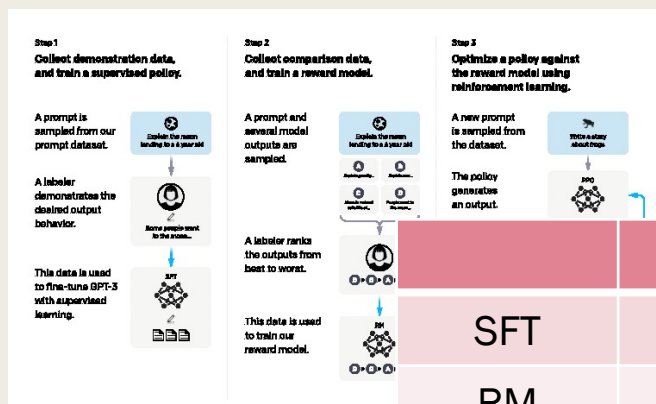
Making language models bigger does not inherently make them better at following a user's intent. For example, large language models can generate outputs that are untruthful, toxic, or simply not helpful to the user. In other words, these models are not *aligned* with their users. In this paper, we show an avenue for aligning language models with user intent on a wide range of tasks by fine-tuning with human feedback. Starting with a set of labeler-written prompts and prompts submitted through the OpenAI API, we collect a dataset of labeler demonstrations of the desired model behavior, which we use to fine-tune GPT-3 using supervised learning. We then collect a dataset of rankings of model outputs, which we use to further fine-tune this supervised model using reinforcement learning from human feedback. We call the resulting models *InstructGPT*. In human evaluations on our prompt distribution, outputs from the 1.3B parameter InstructGPT model are preferred to outputs from the 175B GPT-3, despite having 100x fewer parameters. Moreover, InstructGPT models show improvements in truthfulness and reductions in toxic output generation while having minimal performance regressions on public NLP datasets. Even though InstructGPT still makes simple mistakes, our results show that fine-tuning with human feedback is a promising direction for aligning language models with human intent.

**1 Introduction**

Large language models (LMs) can be "prompted" to perform a range of natural language processing (NLP) tasks, given some examples of the task as input. However, these models often express unintended behaviors such as making up facts, generating biased or toxic text, or simply not following user instructions (Bender et al. 2021; Bommasani et al. 2021; Keiron et al. 2021; Weidinger et al. 2021; Tamkin et al. 2021; Gehrmann et al. 2020). This is because the language modeling objective

\*Primary authors. This was a joint project of the OpenAI Alignment team. RL and JL are the team leads. Corresponding author: loo@openai.com.  
<sup>1</sup>Work done while at OpenAI. Current affiliations: AA: Anthropic; PC: Alignment Research Center.

# InstructGPTのデータとその作成



	Labeler	Customer	Total
SFT	12,895	1,533	14,428
RM	10,111	40,983	51,094
PPO		47,329	47,329

## <作業者の選別>

1. 有害情報に対する反応
2. 全体的なクオリティー
3. 有害質問に対して生成した回答の質
4. 特定の考え方に偏らない人であるか

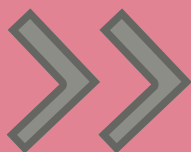
男女比均等、20代が多い、大学、修士卒がほとんど  
フィリピン人、バングラディッシュ人多い

## <作成方針>

正確性, 有害ではないこと > 役に立つこと (タスクにもよる)  
有害情報に対する注意が多い

## 結論

インストラクション付  
1.3B param. GPT-3



175B param. GPT-3



# 2つの重要な論文（その2）



18ページ

## Instruction tuningの評価論文

How Far Can Camels Go? Exploring the State of Instruction Tuning on Open Resources

(Yizhong Wang+ 11 authors, AI2, U. of Washington)

arXiv:2306.04751v1 [cs.CL] 7 Jul 2023

arXiv:2306.04751v1 [cs.CL] 7 Jun 2023

### How Far Can Camels Go? Exploring the State of Instruction Tuning on Open Resources

Yizhong Wang<sup>\*†</sup> Hamish Ivison<sup>\*†</sup> Pradeep Dasigi<sup>\*†</sup> Jack Hessel<sup>\*†</sup>  
Tushar Khot<sup>\*†</sup> Khyathi Raghavi Chandu<sup>\*†</sup> David Wadden<sup>\*†</sup> Kelsey MacMillan<sup>\*†</sup>  
Noah A. Smith<sup>\*†</sup> Iz Beltagy<sup>\*†</sup> Hannaneh Hajishirzi<sup>\*†</sup>

<sup>\*</sup>Allen Institute for AI <sup>†</sup>University of Washington  
{yizhongw, hamishi}@allenai.org

#### Abstract

In this work we explore recent advances in instruction-tuning language models on a range of open instruction-following datasets. Despite recent claims that open models can be on par with state-of-the-art proprietary models, these claims are often accompanied by limited evaluation, making it difficult to compare models across the board and determine the utility of various resources. We provide a large set of instruction-tuned models from 6.7B to 65B parameters in size, trained on 12 instruction datasets ranging from manually curated (e.g., OpenAssistant) to synthetic and distilled (e.g., Alpaca) and systematically evaluate them on their factual knowledge, reasoning, multilinguality, coding, and open-ended instruction following abilities through a collection of automatic, model-based, and human-based metrics. We further introduce TULU<sup>2</sup>, our best performing instruction-tuned model suite finetuned on a combination of high-quality open resources.

Our experiments show that different instruction-tuning datasets can uncover or enhance specific skills, while no single dataset (or combination) provides the best performance across all evaluations. Interestingly, we find that model and human preference-based evaluations fail to reflect differences in model capabilities exposed by benchmark-based evaluations, suggesting the need for the type of systemic evaluation performed in this work. Our evaluations show that the best model in any given evaluation reaches on average 83% of ChatGPT performance, and 68% of GPT-4 performance, suggesting that further investment in building better base models and instruction-tuning data is required to close the gap. We release our instruction-tuned models, including a fully finetuned 65B TULU<sup>2</sup>, along with our code, data, and evaluation framework to facilitate future research.

#### 1 Introduction

The latest generation of large language models has brought unprecedented attention to the potential of language technologies. To support imperative user requests and a chat interface, these models often undergo an *instruction-tuning* step which involves training on supervised input/output pairs. Instruction tuning corpora are often gathered via crowdsourcing (Dolly [12], Open Assistant [23]) or via distillation from another model (Alpaca [38], Vicuna [8]). However, while some public, instruction-tuned models are advertised as comparable to powerful closed-source proprietary models such as ChatGPT, most experiments that support such claims only cover a small set of tasks, and mostly rely on model-based evaluation metrics [6, 50]. We contend that the evaluation setup should

<sup>\*</sup>Equal contribution.  
<https://github.com/allenai/open-instruct>

# 結論

特定のジャンルや目的のインストラクション学習はその対象の性能を大幅に向上させる

- Instruction datasets targeted at specific domains and/or capabilities are extremely effective at improving model performance in those aspects.
- Larger or pretrained-for-longer base models consistently perform better after instruction tuning.
- Our model TŪLU 🦙 – fine-tuned LLaMa on a combination of existing instruction datasets – achieves the best average performance across benchmarks, although it is not the overall best when considering different evaluation settings independently.
- Even a very large (65B) model finetuned on a large mix of instruction datasets fails to outperform ChatGPT, although it does perform significantly better than similar smaller models.
- Model-based preference evaluation on open-ended instruction following correlates strongly with the average number of unique tokens generated by a model, suggesting that model-based preference evaluation has biases that may hide differences in model capabilities.

出力が長いほど、評価結果が良くなる傾向が見られた

# 評価結果

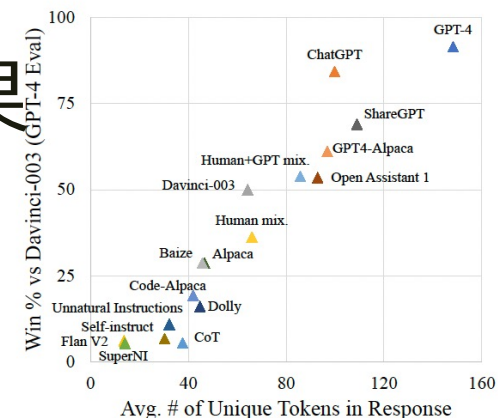


Figure 2: Win-rate scores of 13B models (trained on different datasets) given by GPT-4 strongly correlate with the average numbers of unique tokens in the model responses (Pearson  $r = 0.96$ ).

インストラクション学習は、どの実験においても性能を上げた

提案モデルは、平均的に最善だが全てのタスクで最善ではない

オープンなデータとモデルではChatGPTに（遙かに）及ばない

# まとめると

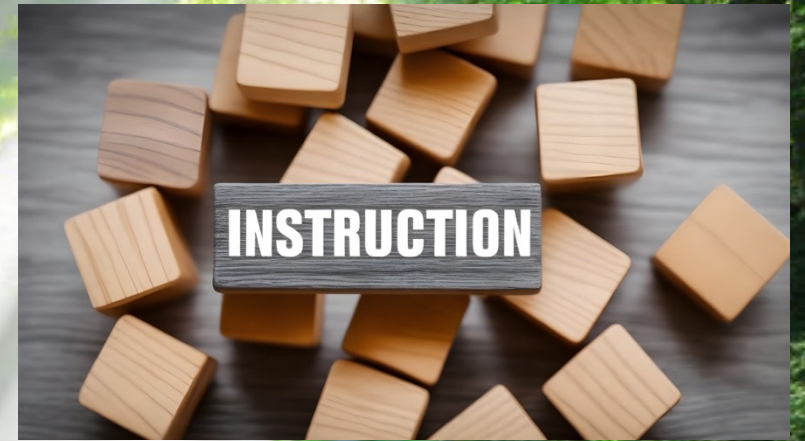
この2つの論文、読めば読むほど。。。  
GPTは何が凄い要素かって、LLMよりも  
Instruction部分なのではないかと思えてくる



もちろん

こんなInstructionを学習できるLLMは凄いのだが

ということで、絶対に両方が重要！



## 関連データとその問題点


# 重要な関連データ (Dolly by Databricks)

## Dolly-15k dataset Dolly-15k-ja dataset

<https://www.databricks.com/blog/2023/04/12/dolly-first-open-commercially-viable-instruction-tuned-llm>

<https://github.com/databrickslabs/dolly/tree/master/data>

<https://huggingface.co/datasets/kunishou/databricks-dolly-15k-ja>



databricks

- Platform
- Solutions
- Learn
- Customers
- Partners
- Company

Try Databricks

- Watch Demos
- Contact Us
- Login

Search

### databricks-dolly-15k dataset

**databricks-dolly-15k** contains 15,000 high-quality human-generated prompt / response pairs specifically designed for instruction tuning large language models. Under the licensing terms for **databricks-dolly-15k** ([Creative Commons Attribution-ShareAlike 3.0 Unported License](#)), anyone can use, modify, or extend this dataset for any purpose, including commercial applications.

To the best of our knowledge, this dataset is the first open source, human-generated instruction dataset specifically designed to make large language models exhibit the magical interactivity of ChatGPT. **databricks-dolly-15k** was authored by more than 5,000 Databricks employees during March and April of 2023. These training records are natural, expressive and designed to represent a wide range of the behaviors, from brainstorming and content generation to information extraction and summarization.

#### Dataset Viewer

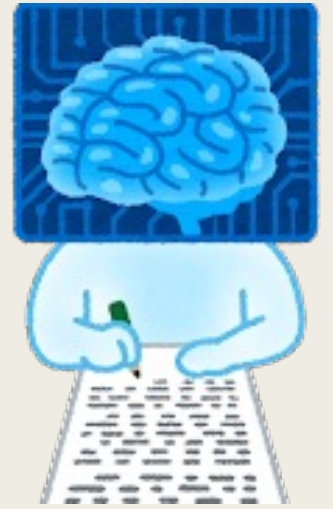
Auto-converted to Parquet </> API

Split

train (15k rows)

input (string)	category (string)	index (string)	output (string)	instruction (string)
"ワインは、一般的にブドウを発酵させて作るアルコール飲料です。酵母はブ..."	"closed_qa"	"200"	"ワインは、一般的にブドウを発酵させて作るアルコール飲料です。"	"ワインとは何か?"
" "	"classification"	"201"	"三銃士、Twix、snickersはキャンディバー、trident、hubba bubba、..."	"以下のものは、キャンディバーやガムですか？トライデント、トゥイック..."
"クシャーン帝国は、1世紀初頭にバクトリア地方でユエシ族によって形成された..."	"information_extraction"	"202"	"クシャーン・サザニア王国 (クシャーン・サザニアおうこく、Kushanshahs、また..."	"イランにおけるクシャーン帝国とササニアン帝国の関係は?"
" "	"brainstorming"	"203"	"1) 暗号は、まだ実生活での目的を見出すことができていない発明である。ピッ..."	"クリプトに課題を与えるもののリストを教えてください。"
" "	"general_qa"	"204"	"いつの時代も哲学者を悩ませてきた深い問いだが、ダグラス・アダムスは著書『..."	"生きる意味ってなんだろう?"

# 日本語Dollyデータセットの問題点



## <自動翻訳、表現の問題>

"オランダとオランダの違いは何ですか？",

"What is the difference between Holland and The Netherlands?"

"雪見だいふく"

"what are some things you can do on a snowy day",

"コードを学ぶにはどうしたらいいのでしょうか？"

"How can I learn to code?"

"どの楽器が弦楽器なのか、打楽器なのかを確認する：エクエ、バラライカ",

"Identify which instrument is string or percussive: Pahu, Morin khuur"

## <文化の問題>

"曜日を開始時のアルファベットで分類する。日曜日、月曜日、火曜日、水曜日、木曜日、金曜日、土曜日",

"人気ミュージシャンを5人挙げる", "テイラー・スウィフト、ヴァンス・ジョイ、ザ・スクリプト、コールドプレイ、レディー・ガガ"

"道路でアクセスできない州の州都はどこ？", "アラスカ州ジュノーは、道路ではアクセスできないが、車はフェリーで行くことができる。"

# 日本語Dollyデータセットの問題点

<自動翻訳、表現の問題>

"オランダの違いは何ですか？",

"What is the difference between Holland and The Netherlands?"

"What can you do on a snowy day",

"コードを学ぶ"

"How can I learn code"

"どの楽器が弦楽器なのか、打楽器か"

"Identify which instrument is stringed or percussive"

"Morin khuur"

<文化の問題>

"曜日を開始時のアルファベットで分類する。日曜日、月曜日、火曜日、水曜日、木曜日、金曜日、土曜日"

"人気ミュージシャンを5人挙げる", "テイラー・スウィフト、ヴァンス・ジョイ、ジョージ・ストラウエルド、ビートルズ、レディー・ガガ"

"道路でアクセスできない州の州都はどこ?", "アラスカ州ジュノーは、道路ではアクセスできないが、車はフェリーで行くことができる。"

日本語独自のデータが必要



プロジェクト



# プロジェクト



## <形態>

- 企業との共同研究として制作費（2000万円）を捻出。22社から申し込み

## <データサイズ>

- 3月末までに1万件のインストラクションデータ（質問回答ペア）を作成
- 12月末に5千件作成。3千件を研究目的で公開

## <課題>

- 質問作成の課題
  - ジャンル、目的、バランス設計、有害情報
- 解答作成の課題
  - 長さ、整形、立場と対象、書きっぷり、複数の回答作成、有害情報対応
- 質問作成者と回答作成者は異なった人、最終チェック担当者

# アノテーター

～元国語教師や日本語教育者が大活躍～

私は文章を書くのはそんなに苦ではないので結構好きでやっています😊

## ■ ベテランアノテーター（15名）

- 日本語学、日本文学、英米文学 出身の方が半数以上
- 元国語教師が創作、入試問題作成など重要なタスクをこなす

面白かったと言っていただけで嬉しいです！私も楽しんで書きました（笑）

## ■ 新規アノテーター（25名）

- 少なくとも2名が元国語教師。最初から重要なタスクをこなしている
- 様々なバックグラウンドを持っている人が集まっており、各人が得意とする分野／タスクの担当をお願いしている



# 元国語教師の創作

ヘビを主人公にして「もうこりごりだ」という内容の物語のいち場面を書いてください。

「主人公をヘビにして、「もうこりごりだ」という物語のいち場面を作成します。

「もうこりごりだ」は「もういやだ」という表現に置き換えられるので、物語の中に「ヘビが嫌だと思ふような出来事」を挿入するのがいいでしょう。

～物語のいち場面～

最近、近所で失踪事件が相次いでいる。ついこの前は1メートル先のミミズが。昨日はついにお隣のトカゲが連れ去られたと町内の掲示板に知らせがあった。<ふむ...。確かにあいつらを連れ去るのは容易だ。しかし、オレ様はどうか。オレ様が通るとみんな恐れ多くてサツと身をひるがえしていくんだ。今日だって...>

そう考えている最中に、しっぽに少しの圧迫感を感じた...と思ったら、次の瞬間宙に舞っていた。舞ったのも一瞬で、次に何が起こったのか分からないまま気を失った。

「...そう！.....で見つけ...！しっぽを.....で振り回すの！.....。」

グワングワン割れそうな頭を何とか持ち上げて、聞こえてきたのはウジャウジャ子供たちの途切れ途切れの声。うっすらと目を開ける。目ん玉をひん剥いた！！二十四の瞳がオレ様を見つめてる。

「これ、とぐろ巻いてるってやつじゃない？危なくない？」

子どもの一人がそう言った。<いや。オレ様怒ってない。ゆえにとぐろ巻かない。.....巻いてる！！！！！！>ギュッギュッギュッ。動けない。まったく状況がつかめない。

<中略>

...そう、無我夢中で遠くへ走ったつもりが、花壇の奥行きはそんなになかったらしい...もういやだ。もういやだ。ひっそりと静かなところで暮らしたい。



# LLMの性能が大幅に向上

## 人間による評価

rinna/youri-7b-instruction  
を  
Ichikara-instruction(4802)  
でSFTしたもの

By マネーフォワード

システム	勝率 (数)
houou-7b-instruction-v2	27.5 (11)
GPT-3.5-turbi-1106	55.0 (22)
引き分け	17.5 (7)

rakuda_human_evaluation_data													
prompt	A	Aコメント	B	Bコメント	Fluency		Accuracy		Detailed		Relevance		優劣
					A	B	A	B	A	B	A	B	
一票の格差問題とは何か説明し、これが日本の政治にどのような影響を及ぼしているかを論じてください。	<p>一票の格差とは、選挙において1人の有権者の投票価値が他の有権者よりも大きいことを指します。例えば、選挙区ごとの有権者数が大きく異なる場合、人口の多い選挙区の有権者の一票の価値が低くなってしまうことがあります。これは選挙民の平等な権利を侵害することになります。</p> <p>日本の場合、選挙区ごとの人口密度の差が大きく、それによって一票の格差が生じています。例えば、人口の少ない地方の選挙区では、一票の価値が人口の多い都市部の選挙区に比べて大きくなってしまいます。</p> <p>この一票の格差が日本の政治に与える影響としては、人口の多い都市部の選挙区の影響力が相対的に弱まることや、地方の利益が過大に反映されることが挙げられます。また、一票の格差が選挙制度の公平性を問題視されることで、政治の信頼性や選挙制度への不信感が生じる可能性もあります。</p> <p>このような影響を考慮し、一票の格差を解消するための選挙制度の改革や選挙区の再編成などが議論されています。</p>	問題点や改善案も書かれている点が良い	<p>一票の格差問題とは、選挙区ごとの人口や面積等に格差があるため、有権者ひとりあたりの投票価値が違ってしまふ問題のことです。</p> <p>この問題は、日本の国会である参議院において、2015年（平成27年）以降、抜本的な改革が行われていないために、常態化しています。</p> <p>一票の格差が最大となった2022年（令和4年）7月の参議院議員通常選挙については、最高裁判所大法廷において、投票価値の平等を定めた憲法第43条に違反する無効な選挙とする判決が出されています。これを受けて、参議院は、2022年（令和4年）9月27日に「参議院議員の選挙区選出議員の選挙に関する規定に係る憲法の規定に基づく参議院選挙制度の見直しに関する小委員会」（小委員会）を設置し、「1票の格差」の解消に向けた議論を行っています。</p> <p>日本の政治において、一票の格差問題は、選挙の度に論点となり、抜本的な改革が求められています。</p>	詳細な内容に踏み込んでいるが、その内容が正確でない	5	5	5	2	5	5	5	5	A



プロジェクト目標、形態と共同研究スキーム



# 情報

- データーや共同研究にご興味ある方は、まず、関根までご連絡ください

- [satoshi.sekine@riken.jp](mailto:satoshi.sekine@riken.jp)

- ホームページ

- <https://liat-aip.sakura.ne.jp/wp/llm%E3%81%AE%E3%81%9F%E3%82%81%E3%81%AE%E6%97%A5%E6%9C%AC%E8%AA%9E%E3%82%A4%E3%83%B3%E3%82%B9%E3%83%88%E3%83%A9%E3%82%AF%E3%82%B7%E3%83%A7%E3%83%B3%E3%83%87%E3%83%BC%E3%82%BF%E4%BD%9C%E6%88%90/?lang=ja>



質問、ディスカッションはslackに