



# 第 3 6 回 全 国 産 業 教 育 フ ェ ア 佐 賀 大 会

# さしフェア SAGA 2026

## 「全国高校生 DI 選手権大会 育てる!!方言 AI コンテスト」実施規則

### 1. 競技概要

競技は次の2つで構成される。

#### 構成①:T5 系自然言語処理(NLP)モデル実装・評価

参加者は AI 開発者が使用するプログラミング実験ノート「Jupyter Notebook」を用いて、事前に NLP モデルの機械学習およびパラメータ調整を行う。

競技当日は、大会で提示されるお題となる文章データを、事前に学習・調整したモデルに入力して、「標準語(以下「標準語 A」) → 方言 → 逆変換後標準語(以下「標準語 B」)」の二段階変換を行う。

#### 構成②:プレゼンテーション

参加者は、以下の内容を含むプレゼンテーションを行う。

- 方言の妥当性を示す根拠(別紙①～③参照)
- 方言変換 AI の出力結果および評価スコア  
競技当日に得られた方言変換結果および標準語 A と標準語 B の類似度(以下の 2 指標による評価)に基づき、方言変換 AI の性能を示すこと。
  - ・ BERTScore:文の意味の保持度を測定する指標(スコア範囲:0~1、高いほど良好)
  - ・ chrF++:文字・単語レベルの字面の一致度を測定する指標(スコア範囲:0~100、高いほど良好)※ 評価スコアは、事務局提供のツールにより自動算出する。
- 方言変換 AI のカスタマイズ内容  
事前学習および調整において、変更・工夫したパラメータや設定内容を説明し、その意図や効果について述べる。また、変換結果を踏まえた方言変換 AI の改善点について述べる。
- 方言変換 AI の社会での活用方法  
方言変換 AI の社会での活用方法について述べる。地域課題の解決という視点を踏まえ、具体的な提案を行う。

### 2. 規格・規定

#### (1) 使用機材

##### ア 対面参加

競技で使用するPC機材は大会事務局が準備する。

##### イ オンライン参加

競技で使用するPC機材は各参加者が準備するものとする。

周辺機器(外部ストレージ、外部ディスプレイ等)の使用可否については、対面・オンラインいずれの場合も大会事務局の指示に従うこと。

#### (2) OS・ハードウェア

##### ア 対面参加

使用するOSおよびソフトウェア環境は、大会事務局が準備した状態のものを使用する。CPU、RAM、GPU、ストレージ等のハードウェア仕様は大会指定とする。

##### イ オンライン参加

OSおよびハードウェア構成は各参加者が準備するものとする。競技に支障が出ないように、大会指定PCスペックを参考に、機械学習に適した性能の環境を用意すること。

ウ 大会指定PCスペック(対面参加用)

以下のスペックは、対面参加者向けに大会事務局が準備するPC環境である。

- CPU: Intel Core i7-13700H
- RAM:16GB
- GPU:GeForce RTX 4060 Laptop GPU 8GB
- OS:Windows11

(3) 開発環境の準備

ア 対面参加

競技に必要な開発環境(Python、Jupyter Notebook等)は、大会事務局が準備したものを使用する。

イ オンライン参加

競技に必要な開発環境(Python、Jupyter Notebook等)は、各参加者が事前に構築すること。また、大会当日に支障が出ないように、事前に動作確認を行うこと。

(4) ソフトウェア・ライブラリ

ア 言語・ツール

- 使用言語はPython 3系とし、Jupyter Notebookを用いる。
- 仮想環境はMiniforgelにより作成すること。

イ ライブラリ

競技の必須ライブラリはコードに組み込まれているため、各参加者が個別に準備する必要はない。  
【ライセンスについて】

本大会で配布するプログラムは、sonoisa氏による「t5-base-japanese-v1.1」(CC-BY-SA-4.0)を基に事務局が改変したものである。参加者が本プログラムや調整済みモデルを外部で公開・配布する際は、同ライセンスに従い、出典の表示および同一ライセンスでの提供が求められる。

(5) ネットワーク利用

ア 許可される利用範囲

- 大会事務局が設ける競技に関連する情報を事前に提供されたWebサイトの参照
- Python、Jupyter Notebookに関連する公式ドキュメントの参照

イ 禁止される行為

- インターネット上のクラウドサービスを利用して、参加者のPC以外で計算や処理を行うこと。
- インターネット上の外部サービスに接続し、データの取得や処理を行うこと。
- ソーシャルメディアやメッセージングサービスにアクセスすること。

3. データセットの仕様

競技では、各参加者自身で学習用のデータセットを準備し、それを使ってモデルを学習させることとする。以下を遵守すること。

(1) データセットの内容

エントリーシートで指定した方言データを収集し、主な学習素材とする。

(2) データ収集方法

利用するデータは、著作権を侵害しない範囲で、言語資料、インタビュー、アンケート、生成AIなどを使用して、方言データを合法的に収集すること。

ア データ量の目安

データは質の高い分析を行うために、最低2,000項目があることが望ましい。

イ データ形式の指定

データはCSV形式で学習すること。CSVファイルには見出し行および行番号の列は設けず、1行目から学習データを記載すること。各対は話し言葉で自然な文章を使用すること。1文あたり20文字以上が望ましい。(別紙参照)

ウ 個人情報の取り扱い注意事項

データ収集の際は、個人情報の保護に最大限の注意を払う必要がある。収集したデータに個人が特定されるような情報を含めないようにすること。

#### 4. 実施方法

##### (1) 一次審査

提出された資料を元に審査を行い、本選に出場する6チームを選出する。

なお、提出するプレゼンテーション動画には、以下の内容を含めること。

- 方言としての妥当性を裏付ける根拠(別紙参照)
- 「方言変換AIの出力結果および評価スコア」
  - ・ 作成した方言変換AIを使用して、「標準語A」→「方言」→「標準語B」の2段階変換を行った様子
  - ・ 「標準語A」→「方言」の変換結果およびその精度についての自己評価
  - ・ 標準語Aと標準語BのBERTScoreとchrF++
- 「方言変換AIのカスタマイズ内容」
  - ・ 事前学習および調整において、工夫したパラメータや設定内容およびその意図や効果
  - ・ 変換結果を踏まえた方言変換AIの改善点の考察
- 「方言変換AIの社会での活用方法」
  - ・ 地域課題の解決という視点を踏まえた方言変換AIの活用方法

##### (2) 本選

一次審査を通過した6チームで競技を行う。競技の流れは以下のとおりである。

- STEP1  
事前にカスタマイズした方言変換AIに、大会当日に提示されるお題の標準語Aを入力し、方言へ変換する。
- STEP2  
STEP1で出力された方言を、再度AIに入力して標準語Bへ逆変換する。
- STEP3  
標準語Aと標準語Bを比較し、BERTScoreとchrF++を算出することで変換精度を数値として確認する。
- STEP4  
STEP1～3で得られた変換結果およびBERTScoreとchrF++を、事前に準備した発表資料に組み込み、プレゼンテーションの最終準備を行う。  
※ STEP1～4の実施時間は45分とする。
- STEP5  
プレゼンテーションおよび審査員による質疑応答を行う。1チーム10分(発表7分/質疑応答3分)。  
※発表時間は、司会者による「発表を始めてください」の台詞終了時から7分以内とし、超過した場合は即座に発表を終了すること。

## 5. 審査基準

一次審査と本選は下記の観点で総合的に審査を行う。ただし、本選においては、質疑応答での対応力も評価に含める。

- 「方言としての妥当性を裏付ける根拠」
  - ・ 方言としての妥当性を裏付ける根拠が客観的に提示されているか
- 「方言変換AIの出力結果および評価スコア」
  - ・ 変換結果について、その妥当性を説明できているか
  - ・ 標準語Aと標準語BのBERTScore(0~1)が1に近い結果となっているかどうか
  - ・ 標準語Aと標準語BのchrF++(0~100)が100に近い結果となっているかどうか
- 「方言変換AIのカスタマイズ内容」
  - ・ カスタマイズ内容を明確に説明できているか
  - ・ 変換結果を踏まえた改善策の提案があるか
- 「方言変換AIの社会での活用方法」
  - ・ 独自性(工夫やオリジナリティがあるか)
  - ・ 実現可能性(アイデアが「実際に作れそう」「仕組みとして成り立ちそう」か)
  - ・ 社会的価値(この方言変換AIが、誰の役に立ち、どんな地域課題を解決できるか)
  - ・ プレゼンテーション技術(説明が分かりやすく、聞き手に伝わる話し方・資料になっているか)

## 6. 失格事項

### (1) 不正行為

- ア ネットワーク利用に関する違反  
「2 規格・規定 (5) ネットワーク利用 イ 禁止される行為」に該当する行為
- イ 不正なソフトウェアの使用  
大会で指定されていないソフトウェアやツールを使用すること。
- ウ 機材の不正改造  
大会事務局が提供する機材を改造したり、外部機器を不正に使用したりすること。
- エ データの不正変更  
競技に使用するデータを不正に変更すること。また、競技結果を意図的に誤認させること。

### (2) 規定違反

指定されたフローやルールに従わない行為

### (3) 安全および倫理上の違反

- ア 妨害行為  
他の参加者の競技結果や操作に影響を及ぼす行為、またはそれを試みること。他の参加者の競技を故意に妨げること。
- イ 安全基準不適合  
機材や環境に対する安全基準を無視する行為

## 7. 異議申し立て

審査結果に対し、異議の申し立てをすることはできない。

## 8. その他

本実施規則に定めのない事項については、大会事務局が協議のうえ決定する。また、競技の円滑な運営のために大会事務局が必要と判断した場合は、参加者に事前通知のうえ規則を変更することがある。

(別紙)

■ 方言の妥当性を示す根拠

プレゼンテーションでは、方言が正しいことを示すため、以下のいずれかを必ず示すこと。

- ① 方言辞典・文献からの引用
- ② 方言話者(地域住民、家族、教員等)による確認結果
- ③ 参加チームメンバーがその方言話者であること、およびその判断根拠

■ 学習させる CSV ファイル例(左側:標準語 / 右側:方言)

	A	B
1	その本どうだった？思ったより内容が深かったって聞いたけど。	その本どうやった？思うたより内容が深かって聞いたばってん。
2	確かに深かったけど、意外と読みやすくて一気に読み終わったよ。	確かに深かっばってん、意外と読みやすうて一気に読み終わったとよ。
3	それはいいなあ、最近は気軽に読める本を探してたところなんだ。	そいはよかねえ、最近は気軽に読める本ば探したところやったと。
4	気軽に読めるけど、読み終わったあとにちょっと考えさせられる感じだった。	気軽に読めるっちゃ読めるっちゃけど、読み終わったあとにちいっと考えさせられる感じやったばい。
5	そういうの好きかも、読みやすいのに余韻が残るやつ。	そういうと好いとるとよ、読みやすかっつとに余韻が残るやつ。
6	作者のインタビュー記事も読んだけど、書くときにすごく丁寧に世界観を作ってるらしいよ。	作者のインタビュー記事も読んだっちゃけど、書くときにすごう丁寧に世界観ば作りよるらしかとよ。
7	やっぱり丁寧な作品って読んでて伝わるよね、雑さがないというか。	やっぱ丁寧な作品で読んどったら伝わるよかねえ、雑さがなかつちゅうか。
8	その通り、気づいたら物語の中に引き込まれてた。	ほんにそいそい、気づいたら物語ん中に引き込まれとったもん。
9	最近そんな風に夢中になれるものがなくて羨ましいよ。	最近そがん風に夢中になれるもんがなかけん羨ましかねえ。
10	むしろ最近忙しすぎて、そういう時間を無理やり作ってる感じだよ。	むしろ最近忙ししかすぎて、そがん時間は無理やり作りよる感じたい。
11	忙しい中で本を読むって逆に集中できていいのかもしれないね。	忙しか中で本ば読むって逆に集中できてよかとかもしれんね。
12	読み始めたら止まらなくて、結局寝不足になっちゃったけどね。	読み始めたら止まらんごとになって、結局寝不足になつてもうたばってんね。
13	わかる、それやると次の日後悔するのにまた同じことしちゃうんだよね。	わかるわかる、それしたら次の日後悔するつとにまた同じことしてしまうっちゃんね。
14	でも寝不足でも満足感はあるから、まあいいかって思っちゃう。	ばってん寝不足でも満足感はあるけん、まあよかかって思うてしまうとよ。
15	気持ちはすごくわかるけど体調には気をつけてね。	気持ちはようわかるばってん、体調には気いつけんといかんよ。
16	最近はカフェインの力に頼りすぎてるから、そろそろ控えようと思ってるよ。	最近はカフェインの力に頼りすぎよるけん、そろそろ控えようて思いよると。
17	コーヒー飲む量が増えてくるとちょっと危険信号だよな。	コーヒー飲む量が増えてくると、ちょっと危険信号やんね。
18	しかも夜に飲んでしまって余計寝れなくなるという悪循環。	しかも夜に飲んでしもうて、余計寝れんごとなるっちゅう悪循環よ。
19	僕も前にそれやって、2日くらい眠気が変な感じになったことあるよ。	俺も前にそいばやって、2日くらい眠気が変な感じになったことあるとよ。
20	生活リズムって崩れると戻すの大変だよな。	生活リズムって、いっぺん崩れると戻すと大変よかねえ。