

# 日本語不法行為事件データセット

Version: 2024-04-01

## 概要

日本語不法行為事件データセット (Japanese Tort-case Dataset, JTD)は、日本語・日本法における法的判断予測研究のためのデータセットである。JTD は不法行為判断予測タスク及びその根拠抽出タスク向けに設計されている。根拠抽出タスクは不法行為の成否判断に際して重要な根拠となった主張を、原告または被告の主張の中から抽出するタスクである。JTD は 41 人の法律専門家によって注釈付けされた 3,477 件の民事事件判決書に基づいて構築されており、7,978 事例（事例に内包される原告・被告らの主張は 59,697 事例）が収録されている。

本データセットは以下の2つのデータセットから構成されており、どちらも研究目的利用限定での提供となる。**それぞれ利用条件詳細が異なる**ため、利用申請書兼誓約書記載の条件を十分確認の上利用すること。

- 日本語不法行為事件データセット（元文書のみ）
- 日本語不法行為事件データセット（アノテーションのみ）

現在、dev・trainデータのみ公開中。testデータは後日公開予定。

なお、本言語資源は JST RISTEX JPMJRX19H3 並びに JST ACT-X JPMJAX20AM の助成を受けたものである。また、元データとなる判決書は株式会社LICより提供された。

## Citation

- データセット本体 (Dataset paper)
  - Hiroaki Yamada, Takenobu Tokunaga, Ryutaro Ohara, Akira Tokutsu, Keisuke Takeshita, Mihoko Sumida. *Japanese Tort-case Dataset for Rationale-supported Legal Judgment Prediction*. ArXiv Preprint, abs/2312.00480.
- アノテーションスキーム (Annotation scheme paper)
  - Hiroaki Yamada, Takenobu Tokunaga, Ryutaro Ohara, Keisuke Takeshita, and Mihoko Sumida. 2022. *Annotation Study of Japanese Judgments on Tort for Legal Judgment Prediction with Rationales*. In Proceedings of the Thirteenth Language Resources and Evaluation Conference, pages 779–790, Marseille, France. European Language Resources Association.

## ディレクトリ構造

- README.md (this file)
- README.pdf (pdf version of this file)
- annotations (日本語不法行為事件データセット（アノテーションのみ）)
  - annotation\_guideline (アノテーション時の作業向けガイドラインを含むフォルダ)
  - test.jsonl (テストセット, 後日公開)
  - train.jsonl (訓練セット)
  - valid.jsonl (開発セット)
- LIC\_documents (日本語不法行為事件データセット（元文書のみ）)
  - doc\_id\_list (収録している判決書のID。LIC社データベースに基づく。)

- test (テストセットに用いられた元文書, 後日公開)
- train (訓練セットに用いられた元文書)
- valid (開発セットに用いられた元文書)

## ファイル仕様

annotations (日本語不法行為事件データセット (アノテーションのみ) )

- このファイルはアノテーションデータを整形し、huggingface datasetとの互換性を持たせたもの。
- ファイル形式: UTF-8, LF, jsonl形式
- jsonlファイル各行の仕様
  - instnace\_ID: 事例ID
  - CD\_text: 裁判所による不法行為成否の判断 (タスクでは未使用)
  - CD\_decision:  $T_g$ , 不法行為の成否
  - UF\_text:  $U$ , 争いのない事実等
  - Date\_AD: 日付 (西暦)
  - Category: 事件分類 (タスクでは未使用)
  - Court: 担当裁判所
  - FC\_text\_plaintiff:  $P$ , 原告側主張 (具体的な事実の主張, Factual Claims)
  - FC\_labels\_plaintiff:  $R_g^P$ , FC\_text\_plaintiffに対する裁判所の採否
  - FC\_text\_defendant:  $D$ , 被告側主張 (具体的な事実の主張, Factual Claims)
  - FC\_labels\_defendant:  $R_g^D$ , FC\_text\_defendantに対する裁判所の採否
  - FC\_text\_other: 未使用
  - FC\_labels\_other: 未使用
  - NC\_text\_plaintiff:  $P$ , 原告側主張 (抽象的規範部分, Claims of Norms)
  - NC\_labels\_plaintiff:  $R_g^P$ , NC\_text\_plaintiffに対する裁判所の採否
  - NC\_text\_defendant:  $D$ , 被告側主張 (抽象的規範部分, Claims of Norms)
  - NC\_labels\_defendant:  $R_g^D$ , NC\_text\_defendantに対する裁判所の採否
  - NC\_text\_other: 未使用
  - NC\_labels\_other: 未使用

$P$ は, タスク内ではFC\_text\_plaintiffとNC\_text\_plaintiffを結合したリストとして与えられる。  $R_g^P$  も FC\_labels\_plaintiffとNC\_labels\_plaintiffが結合されたもの。  $D$ ,  $R_g^D$ も同様。

LIC\_documents (日本語不法行為事件データセット (元文書のみ) )

- ファイル形式: UTF-8, LF, プレーンテキスト
- 一般的なテキストファイルです。

## 追加資料

- <https://github.com/nightowl-jp/JTD-public>