Understand User's Rights

Stanford CS224N Custom Project

Ray Hu
Department of Computer Science
Stanford University
rayhu@stanford.edu

Benjamin Ward
Institute for Computational and Mathematical Engineering
Stanford University
wardb@stanford.edu

Basant Khalil
Department of Computer Science
Stanford University
bkhalil@stanford.edu

1 Key Information to include

- Mentor (custom project only): We would like to request Jing Huang (hij@stanford.edu) as our mentor.
- Sharing project:

2 Research paper summary (max 2 pages)

Title	Know What You Don't Know: User's Rights in Online Terms and Conditions
Venue	Stanford CS-224N Winter 2025
Year	2025
URL	https://github.com/AI-knows-your-rights/t-c-ranker

Table 1: Sample table for bibliographical information

Background. Legal language is one of the most complex and sophisticated forms of natural language. A tragic 2024 case brought this issue into focus: a woman passed away at Disney, and the company argued that her family had waived their right to sue because she had accepted the terms and conditions (T&C) when she signed up for a trial of Disney+ in 2019. The legal clause, buried in fine print, favored the company, highlighting how consumers often unknowingly forfeit their rights due to legal documents' dense, technical nature.

Summary of contributions. Our project aims to develop an AI-powered system that can summarize and rate T&C, making them easier for consumers to understand. We have identified a community-annotated dataset containing 500 companies' T&C analyses by qualified attorneys, providing a strong foundation for classification and summarization. The dataset has been used to make a browser extension. Here is the screenshot. Google T&C are ranked as grade E as it includes many clauses that don't respect users' privacy. On the contrary, DuckDuckGo has an excellent rating. We extend the rating to more companies by leveraging AI.

Limitations and discussion. The project is based on a community-annotated dataset of 500 companies' websites. The company's T&C may have updated, but the community annotation probably cannot catch up with it, resulting the wrong annotation. Also, the annotations are all in English, potentially hurts the accuracy of other languages. Despite the limitations, we are still confident the

project is meaningful and will contribute to users' right awareness and push the companies to take more social responsibilities.

Why this paper? Our project aligns with ongoing research in legal text processing, document summarization, and explainable AI. Many legal contracts, T&C, and regulatory texts are dense and difficult to interpret, making it a rich area for NLP research.

Also, the legal language is highly context-dependent, just like culturally sensitive text. A clause might have different legal implications based on jurisdiction, industry norms, or company policies. For example, A "no-refund policy" may be standard in the U.S. but illegal in certain European jurisdictions. Similarly, arbitration clauses that waive consumer rights are controversial but often hidden in T&C. Our work can build on research in culturally aware NLP and legal interpretation by exploring how legal text is interpreted across different companies, industries, and jurisdictions.

Our work contributes to AI explainability (XAI) and transparency in corporate policies. AI is often seen as a "black box" in legal applications, and our work could help increase transparency in automated decision-making. Many T&C documents are deliberately obfuscated, so our model could provide consumer-friendly, interpretable AI outputs.

If successful, our project could pressure companies to write more consumer-friendly T&C by exposing unfair practices, incentivize companies to adopt better legal practices.

Our work could also contribute to an open legal dataset, potentially useful for future NLP research.

In summary:
• Legal NLP is an emerging, high-impact field with increasing research interest.
• Our project combines deep learning, legal text understanding, and social impact.
• It extends research on context-aware NLP and explainability in AI.
• It aligns with ongoing discussions on AI ethics, transparency, and regulatory compliance.
• It provides a practical use case that bridges academia and real-world legal challenges.

Wider research context. Our project advances legal NLP by tackling one of the hardest challenges in language processing—understanding and summarizing legal T&C for consumer awareness and corporate accountability. While our work has direct social impact, it also fits into the broader story of NLP research, addressing core challenges in representation, structure, deep learning application, and explainability.

For example, we will explore the following NLP research areas:

1. How do we adapt LLMs to specialized, high-precision domains? 2. How do we improve interpretability in AI-generated text? 3. How do we ensure fairness and bias mitigation in legal NLP? 4. How do we improve the dataset for future NLP research?

3 Project description (1-2 pages)

Goal. Our project investigates the effectiveness of transformer-based models in summarizing and rating legal T&C for consumer comprehension. Specifically, we aim to answer:
1. Can pre-trained language models effectively summarize legal T&C while preserving critical legal meanings?
2. How well do our models detect fairness and bias in T&C documents that matches the annotations by attorneys?
3. Can context-aware NLP techniques, inspired by culturally aware language models, improve legal text interpretation and fairness assessment?

We focus on fine-tuning transformer-based models to generate structured, explainable, and user-friendly summaries. Additionally, we evaluate whether contextual embeddings or retrieval-augmented generation (RAG) improves interpretability and bias detection.

Task. Our task is to automate the summarization and fairness evaluation of T&C documents using natural language processing (NLP) techniques. Specifically, we aim to:

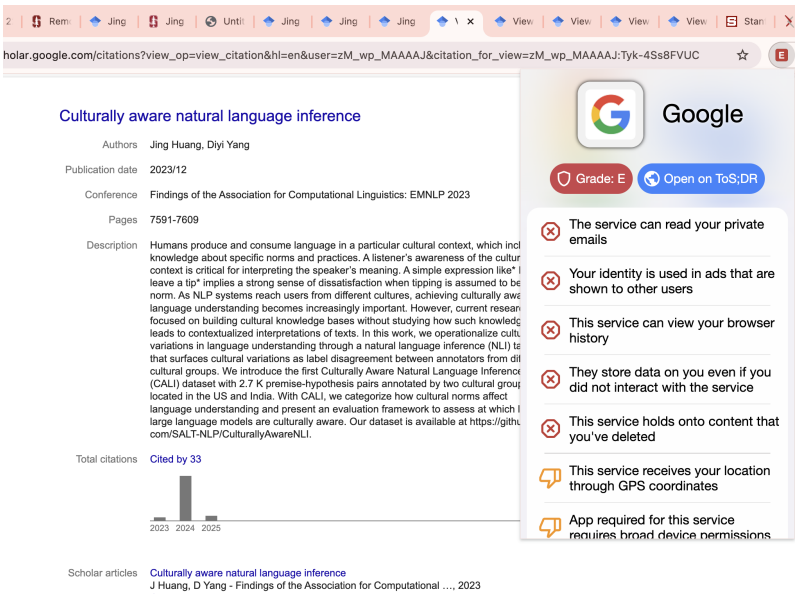

Figure 1: Community Annotated T&C rating and summaries

1. Summarize lengthy and complex T&C documents into concise, user-friendly explanations while preserving key legal implications.
2. Assign a fairness rating to T&C based on consumer rights protections, transparency, and legal risk, using a dataset of attorney-annotated contracts.

An illustration can be found in figure 1.

Data. We will use a community-annotated dataset of T&C from 500 companies, which qualified attorneys have analyzed and labeled. This dataset was previously used to develop a browser extension that show T&C fairness rating. It also includes Clause-level annotations.

To prepare the dataset for fine-tuning transformer models, we will:
1. Tokenization & Cleaning: Standardize legal text we acquired from the company. Remove the date/version mismatches.
2. Sentence Segmentation: Break long legal clauses into manageable units for NLP models.
3. Label Normalization: Align fairness ratings with a numerical scale for supervised learning.
4. Augmentation (if needed): Expand the dataset using semi-supervised techniques like zero-shot prompting from GPT models to generate additional annotations.

Methods. Our approach involves fine-tuning transformer-based models to perform legal text summarization and fairness evaluation.

We will explore and fine-tune pretrained sequence-to-sequence models to evaluate the best outcome.
We will use Hugging Face transformers to download models.
We will train on our annotated dataset, optimizing for text summarization quality and fairness classification accuracy.
We will implement our own legal text preprocessing pipeline, including clause segmentation, tokenization, and explainability analysis.
We will compare our model against existing summarization models (e.g., GPT-4, Claude) to assess performance gains.

Baselines. We will compare our T& summarization model with Pretrained Transformer Models such as OpenAI ChatGPT 4o, Claude, etc. We will compare our classifier with the pretrained classiers such as LegalBERT and the human expert annotations.

Evaluation. We will explore and use a bunch of evaluation methods such as n-gram-based, semantic similarity. For the classifier we plan to use ROC-AUC approaches.

Ethical Challenges Our project presents ethical challenges related to algorithmic bias in legal fairness assessments and misinterpreting AI-generated summaries. We will perform bias audits by cross-checking AI ratings with the community and asking for feedback. We will also put a legal disclaimer that AI summaries are informational, not legal advice, and recommend consulting a lawyer for critical decisions.

References