

# LLM が考える「役立つ不満」とは

平尾 七虹, 山西 良典 (関西大学)

## 「役立つ不満」の課題

### 【過去の研究】

「改善案が思いつくか」で  
人に役立つ不満を判断してもらった

- 基準を設けても**価値観の違い**が出る
- 「役立つ」基準は人によって違う

### 疑問点

- LLM は**何を基準に**「役立つ」とするか
- 人と LLM に違いは生まれるのか
- 発話の役割によって不満は分類できるか<sup>※</sup>

## LLM による不満の評価方法

日本語に優れた指示学習済み LLM を使用し  
LLM の判断基準を分析

データ：不満調査データのうち  
人がラベル付けした 1,601 行のコメント

### 分析

- 3 つの LLM を比較
- LLM 間での**共通点**を調べる
- 人と LLM を比較

Llama3-ELYZA-JP  
Llama3-Swallow  
Qwen2

### プロンプト

"あなたは情報学部に通う大学生です。以下のテキストを解析して、以下の5つの項目に従って教えてください。回答するときは、項目の数字を先頭に付けて回答してください。"

"1. ラベル（「役に立つ」または「役に立つとは言えない」のいずれかを選択）"

"2. ラベルを選択した理由"

"3. 発話の役割の中から最も適する1つを選択)"

"4. 不快に感じるかどうか（はい/いいえ）"

"5. その理由（もし「はい」の場合、不快に感じる理由）"

## LLM が考える「役立つ不満」

### Llama3-ELYZA-JP

役に立つ判定：917  
 ・不快に感じる：701 感じない：216  
 役に立つとは言えない判定：699  
 ・不快に感じる：666 感じない：3  
 最も多かった発話の役割：**知識獲得**

### 人との一致度

正解率：61.9%  
 精度：64.3%  
 再現率：68.1%

### Llama3-Swallow

役に立つ判定：926  
 ・不快に感じる：144 感じない：758  
 役に立つとは言えない判定：674  
 ・不快に感じる：490 感じない：177  
 最も多かった発話の役割：**知識提供**

正解率：60.6%  
 精度：63.3%  
 再現率：66.8%

### Qwen2

役に立つ判定：484  
 ・不快に感じる：12 感じない：472  
 役に立つとは言えない判定：993  
 ・不快に感じる：394 感じない：599  
 最も多かった発話の役割：**価値判断**

正解率：60.3%  
 精度：73.3%  
 再現率：43.7%

### 考察結果

役立つかの基準 { 情報提供 / 情報としての価値  
問題解決 ( 要望・具体案 )

不快に感じるかの基準 { 投稿者が不快に思っているか  
不満に思った出来事が不快か  
読み手が不快に感じるか

LLM に具体的な指示を設けず回答させることで  
判断基準となる軸が見えてきた

## 今後の可能性

考察結果で得られた基準を明確に指示することで  
人が考える「役立つ不満」が得られる可能性がある

### 人が考える「役立つ不満」

## 「役立つ不満」をカスタマイズできるアプリケーション

どんな不満をお探しですか？

基準を入力

search

( 基準の例 ) 事実・状況報告・解決案の提示