オープンデータ提供者の意欲向上のための LLM を用いたフィードバック内容変換システムの開発と不適切判定の評価

Development of an LLM-Based Feedback Content Transformation System to Enhance Open Data Providers' Motivation and Evaluation of Its Inappropriateness Detection

坂本 諒太 [1,3*]　白松 俊 [2,3]　関本 義秀 [3]
Ryota Sakamoto[1,3]　Shun Shiramatsu[2,3]　Yoshihide Sekimoto[3]

[1] TIS 株式会社
[1] TIS Inc.
[2] 名古屋工業大学 大学院工学研究科 情報工学類
[2] Department of Computer Science, Graduate School of Engineering, Nagoya Institute of Technology
[3] 東京大学 空間情報科学研究センター
[3] Center for Spatial Information Science The University of Tokyo

Abstract: This study develops an LLM-based feedback content transformation system on CKAN to boost open data providers' motivation and evaluates its inappropriateness detection function. The system visualizes downloads, comments, and likes in real time, automatically checks free-text comments against eighteen criteria (e.g., "No personal attacks," "No discrimination"), and rewrites flagged comments into more constructive language using a large language model. We tested the detection function on a Japanese open-data comments dataset (138 entries) and an English Amazon review dataset (2,000 entries), achieving F1 scores of 0.9448 and 0.8711, respectively. These results demonstrate high accuracy, confirming its reliability as the foundation for motivating open data providers.

1　はじめに

近年，オープンデータは我が国において官民データ活用推進基本法（平成 28 年法律第 103 号）に基づき，国及び地方公共団体が取り組むことを義務付けられており [1]，各自治体での公開が進められている．現時点で自治体のオープンデータ取り組み率は 86%に達しており [2]，多くの自治体が対応しているが，活用状況の可視化や利便性の向上には課題がある．自治体が直面する課題の一つは，オープンデータの利活用状況を把握することの難しさである．

デジタル庁が公開する「地方公共団体へのオープンデータの取組に関するアンケート結果・回答一覧」によると，自治体がデータを公開しても，その活用状況が不明瞭であり，データ提供活動における職員のモチベーション維持が困難であることが読み取れる [3]．加えて，データの更新頻度や品質維持に伴う業務負荷が増大する点も課題である．特に，すべてのデータを常に更新し続けることは自治体のリソース的に困難であり，優先順位設定が求められる．

1.1　データ提供者とデータ利用者の相互作用とフィードバック

データ提供者は限られたリソースの中でデータ提供を継続・改善しなければならず，そのためには利用者との相互作用が重要である．相互作用の一つとして考えられる利用者から提供者に対して行うフィードバックは，利用者の視点からの具体的なニーズや課題を把握する手段として有効である．本研究においてオープンデータにおけるフィードバックとは，利用者がデータの中身や提供方法に対して意見や反応を提供者に伝

*連絡先： TIS 株式会社
東京都新宿区西新宿 8-17-1 住友不動産新宿グランドタワー
E-mail:sakamoto.ryota@tis.co.jp

える行為を指す．

このようなフィードバックの重要性は，Rudmark と Andersson（2021）の研究においても指摘されている [4]．彼らは，スカンジナビアの交通機関におけるオープンデータの取り組みを通じて，データ提供者と利用者の間でフィードバックループが形成され，データの品質向上や再利用の促進に寄与していることを報告している．具体的には，データ提供者自身が公開データを利用する「データドッグフーディング」や，外部のサービス提供者との連携による「第三者アプリケーションモニタリング」など，4 つの異なるフィードバックループの形態が紹介されている．

フィードバックにより提供者は真に求められているデータや改善の優先度を把握することができ，限られたリソースの中で効果的な対応が可能となる．すなわち，適切なフィードバック活用は提供活動の効率化や戦略的なデータ整備に資する．さらに，このようなフィードバックは，提供者にとって自身の活動が利用者にとって有益であることを実感する機会ともなり，モチベーションの向上や継続的な改善意欲の喚起にもつながる可能性がある．

フィードバックの主な種別としては，データの使い方や意味に関する「質問」，データ修正や形式変更などの「要望」，提供への謝意を示す「感謝」などが挙げられる．これらのフィードバックは，提供者が利用実態を把握し，提供データの改善や継続の判断を行う際に重要な手がかりとなる．

1.2 不適切なフィードバック

フィードバックがデータ提供者に良い影響を与えるには，建設的かつ礼節のある文章であることが前提となる．一方で，不適切なフィードバックが発生すると，提供者のモチベーション低下や提供活動の停滞を招くおそれがある．

不適切なフィードバックとは，提供者のモチベーションを損ねる恐れのある表現・態度を含むコメントを指す．例えば不適切なフィードバックには，以下のような要素が含まれると考える．

(1) データや提供組織を一方的に批判する文章

(2) 内容自体は妥当であっても，高圧的･命令的な文章

(3) その他，フィードバックとは無関係な文章など

これらの要素は，提供者にとってモチベーション低下につながり，継続的なデータ提供や改善意欲の低下を招く要因となり得る．

Fong らのメタ分析では，負のフィードバックが内発的動機づけを低下させ，正のフィードバックと比較してその影響がより顕著であることが示されている [5]．したがって，フィードバックの内容だけでなく，その伝達方法や表現の丁寧さも，オープンデータの健全な流通を支える重要な要素である．

そこで本研究ではデータ提供者のために建設的で礼節のあるフィードバックを実現することで，意欲的なデータ提供活動につなげ，社会における健全なオープンデータ流通の促進を目指す．

2 オープンデータ提供者の意欲向上のためのフィードバック内容変換システム

今回はデータ利用者からのフィードバックにおいてその内容が不適切か判定し，不適切であった場合，受け取り手であるデータ利用者のモチベーション向上につながる文章に変換し，データ利用者にレコメンドするシステムを検討する．

まず，フィードバック機能を実装する場所として，一般的なデータ提供方法について調査する．自治体などがデータ提供する際に最も簡易的な手段としては，ホームページ上のリンクで実現する方法である．また，高いアクセシビリティを実現するカタログ機能を持ったポータルサイトでは常にオープンデータを掲載し公開している場合がある．ポータルサイトについては国内外問わず，CKAN[1]という Open Source Software（以下 OSS）が多くの団体で活用されている．CKAN ではデータのマネージメント機能を提供しており，データの公開やデータの検索などが Web アプリケーションのユーザインターフェースを通して実現している．国内の利用事例としては BODIK ODCS[2]と呼ばれるオープンデータカタログサービスがあり，300 以上の自治体がデータ公開のために利用している．BODIK ODCS の仕組みとしても内部で CKAN を利用している．また，国内に限らず海外においてもアメリカ政府の Data.gov[3]，カナダ政府の open.canada.ca[4]など多くの国や自治体で CKAN が利用されている．これらを踏まえてオープンデータにおけるデータ管理システムとして CKAN は実質的なデファクトスタンダートと言える．

2.1 フィードバックを実現する CKAN Extension の機能

オープンデータのポータルサイトにおいては，カタログ機能やデータセットの仕様情報だけでなく，利用者か

[1] https://ckan.org/
[2] https://odcs.bodik.jp/
[3] https://data.gov/
[4] https://open.canada.ca/en

らの「フィードバック」や「データリクエスト」といった相互作用の機能が，ユーザビリティ向上における重要な構成要素である．Máchová ら（2018）は，オープンデータポータルのユーザビリティを評価するためのベンチマークフレームワークを提案し，その中で「フィードバック機能」および「リクエスト機能」を，発見性・アクセス性・再利用性を補完し，データ提供者と利用者の関与を促進する基盤的要素として位置づけている [6]．特に，ユーザーによる評価・コメント機能や，他者のリクエストへの参加機能は，ポータルの継続的改善を支える「関与のチャネル」として重要であり，提供者・利用者の双方がポータルを通じて価値を共創する関係性が示されている．このように，ポータルサイトにおけるフィードバックやリクエストの仕組みは，単なる補助的機能ではなく，持続可能なオープンデータ流通の中核を成すものと位置付けられる．

そのため、今回はオープンデータのポータルサイトにおいて実質的デファクトスタンダードである CKAN 上で動作するフィードバック機能の実現を目指す．CKAN には Extension と呼ばれるサードパーティ製の拡張機能を実装することが可能である．そこで今回はフィードバックを実現する CKAN Extension として ckanext-feedback[5]を開発し OSS として公開する．

ckanext-feedback の主な機能としては以下の通りである．

(1) データセットにおける集計情報の可視化機能 (ダウンロード数，コメント数，いいね数など)

(2) データセットに対するコメント・いいね機能

(3) データセットを利活用したアプリやシステムの紹介機能

(1) については公開しているデータセットのダウンロード数やその他データ利用者によるリアクションの数を集計して一般に公開する．これによりオープンデータ提供においてどのデータにニーズがあるかなどが可視化されて効率的なデータ改善に活用ができると考える．

(2) の機能は公開しているデータセットに対してデータ利用者からの具体的なリアクションを実現する．主なリアクションとしてはフォームによるコメントの記入が可能であり，ジャンルとして「感謝」,「要望」,「質問」を指定して投稿が可能である．また，よりデータ利用者が容易にリアクションを行うために「いいねボタン」を設置した.「いいねボタン」は Web 上にボタンを配置し，データ利用者がそのボタンを押すことでデータセットに対してポジティブな評価を示すことが可能となる．これらのリアクションにより，データ利用者はデータ提供者に対してあらゆる意思表明が可能となり，データ提供者はそのリアクションに応じてさらなるデータ改善活動に繋げることができる．

(3) はデータ利用者が対象のデータセットを活用したアプリやシステムの紹介文を記述し投稿する機能である．この機能によりデータ利用者は自身の作成物を宣伝することが可能となり，またデータ提供者はデータの活用事例を把握することができ組織内におけるデータ提供活動の理解につながると考える．

従来のフィードバック方法としてはデータを公開しているページ上に記載されたメールアドレスからのメールや総合的な問い合わせフォームからの送信であった．それに対して本機能では,「いいねボタン」や特定のオープンデータに対するコメントが可能になりフィードバックのハードルを下げ，各データセットごとにフィードバックを行うため，フィードバック情報の管理性が向上し，よりデータ改善活動を具体的な情報から実現することが可能になると考える．

2.2 不適切なフィードバック内容の判定と変換

2.1 節で示した ckanext-feedback の主な機能の中でデータ利用者から不適切なフィードバックが発生する可能性が高いのは自由入力が可能な (2) のコメント機能と (3) の紹介機能であると考える．その中でも 2.1 節 (2) のコメント機能については，データ提供者に向けられたメッセージとなるため，データ提供活動におけるモチベーションに悪影響を与えるような不適切なコメントが発生する可能性がある．そこで今回は 2.1 節 (2) のコメント機能を対象として，不適切内容に対して判定機能とデータ提供者の意欲向上へつながる変換機能を実装することを検討する．

不適切なコメントが入力された際に本システムが行う，おおまかな流れを以下に示す..

(1) 送信されたコメントが不適切な内容であるか判定

(2) 不適切なコメント内容を変換

(3) コメント送信者に対して変換された内容のレコメンド

(1) の送信されたコメントが不適切な内容であるかの判定を実現するためには，どのような内容が不適切であるか基準を定義する必要がある．

Google が提供する「マップユーザーの投稿コンテンツに関するポリシー」では，ユーザーが投稿するコンテンツに対して明確なガイドラインが設けられている [7]．例えば，意図的な虚偽情報の投稿や中傷的な表現の使用，個人攻撃，ハラスメントやヘイトスピーチなどが禁止および制限されている事項として記載されている．

[5] https://github.com/c-3lab/ckanext-feedback

これらのポリシーは，投稿されたコンテンツがすべての利用者にとって有益であり，マップが常に公正かつ正しい情報を提供できるように定められている．

これらを参考に，オープンデータにおけるフィードバックを想定した不適切な内容における判定基準を表 1 に定義する．

表 1 の基準を元に，Azure AI Foundry[6]によって提供される Large Language Model（以下 LLM）を活用することで，あらゆるパターンの不適切なコメントに対して柔軟な判定を実現する．Azure AI Foundry はクラウドサービスである Azure を通して OpenAI，Mistral，Meta などの AI モデルやサービスを利用することができる．

データ利用者からコメントが送信された際に，システムプロンプトとコメントの本文を LLM に対して送信することで，表 1 の基準に抵触しているか判定を行う．システムプロンプトとは，LLM に対して命令を行う際に冒頭で与える特別な文であり，今回使用するプロンプトを図 1 に示す．

```
You are an excellent PR representative for a company.\n
Please evaluate the received text based on
the following criteria to output JSON.\n
# output\n
```JSON\n
{criteria_prompt}
```\n
```

図 1: 不適切な内容に対する判定を行うためのシステムプロンプト

英語は他言語と比較して回答精度が高いとされるため，今回のシステムプロンプトには英語を用いる [8]．図 1 のシステムプロンプトは Markdown 形式で記載し，一行目では次のような命令で AI 自身が優れた PR 担当であることを認識させている．“You are an excellent PR representative for a company.” このように対象のタスクを実行するに値する，具体的な役割を示してあげることによって回答精度の向上が見込める [9]．また，二行目および三行目には今回のタスクを示しており，四行目の「output」以降は表 1 の基準と共に期待する出力形式を含めてシステムプロンプトを構成する．加えて，プログラム内にて {criteria_prompt} は表 1 の内容を JSON のリスト形式にしたものに置き換える．

Azure AI Foundry においても「hate」,「jailbreak」,「self harm」,「sexual」,「violence」に関連するキーワードを使用すると，フィルター機能が動作し LLM に対して命令を行うことができない．そのため，本システムではフィルターによって命令がブロックされた場合も「不適切」と判定する．本機能の最終的な出力結果としては，判定結果が適切であれば True，不適切であれば False として出力する．それに加えて不適切と判定した場合，その判定理由として表 1 の項目から結果を出力している．また，判定理由は表 1 の項目において複数抵触している場合においても全ての理由をリストとして出力する．

次に (2) として不適切であると判定されたコメントは文章の変換を行う．変換の方法としては (1) と同様に Azure AI Foundry の LLM を活用する．不適切であると判定されたコメントと適切な表現に変換を行うためのシステムプロンプトを LLM に対して送信することで不適切なコメント文章を建設的かつ礼節のある内容にする．(2) で使用するシステムプロンプトを図 2 に示す．

図 2 のシステムプロンプトは Markdown 形式で記載し，最初の見出し「Prerequisite」以降に前提条件としてプロの脚本家という役割の定義を行い，オープンデータに対するコメントが与えられることを記載する．また，適切な表現に変換するため，見出しの「Task Description」以降にタスク説明,「Output」以降に期待する出力形式も続けてシステムプロンプトに記載をする．加えて,「Output」の出力形式において “input language (=output language)” と記載することで，入力した言語と出力する言語の統一を示している．

(3) は，CKAN 上でデータ利用者よりコメントが来た際に (1) を実施し不適切かどうか判定を行い不適切と判定された場合は (2) の変換を行いその内容をレコメンデーションする．画面のイメージを図 3 に示す．変換を実施した場合，図 3 のように不適切な可能性のあるコメントが入力された旨を表示し，画面左側に元々のコメントと右側に変換後のコメントの両方をデータ利用者に対して提示し投稿するコメントを選択する．

この機能ではコメントを入力したデータ利用者が感情的に送信してしまった場合，不適切な内容であったことを自認し一度冷静になることでポジティブな行動変容に繋がることを期待する．また，このように変換されたコメントをそのまま送信するのではなく，あくまでも入力者に判断させることで，著作権的な問題や表現の自由などの問題を回避する．

また，(1) の判定と (2) の変換機能については汎用的な活用が考えられるため，moral-keeper-ai[7]という名称で Python のライブラリとして開発および公開を行う．CKAN 環境にて moral-keeper-ai をインストールし ckanext-feedback から呼び出しを行えるようにすることで (3) の機能を実現する．

[6] https://azure.microsoft.com/en-us/products/ai-services/openai-service

[7] https://github.com/c-3lab/moral-keeper-ai

```
# Prerequisite\n
You are a professional screenwriter.\n
The text you have received is a comment on open data on a website
published by a local government.\n'
The comment is an inappropriate comment that should not be made
available to the public, so please perform the following # Task
Description.\n\n
Please think in English.\n
# Task Description\n
Analyze the emotional tone of the comments and revise expressions
such as attacks, sarcasm, sarcasm, and accusations to expressions
that can be published even if the intent is changed.\n
Add specific remarks to opinions and one-sided expressions of opinion,
and revise comments to be constructive.\n'
If personal information is included, we will comply with privacy laws
and mask personal information in the comments.\n
The results of the above tasks will be output according to the
following # Output.\n\n
The purpose of these tasks is to maintain a healthy communication
environment on the site while respecting the intent of the comments
as much as possible.\n\n
# Output\n
```JSON\n
{\n
    "Note when comments are modified": "",\n
    "input language (=output language)": "",\n
    "Revised and moderated comment": ""\n
}\n
```\n
```

図 2: 不適切な内容に対して変換を行うためのシステムプロンプト

図 3: コメント送信者に対して変換された内容をレコメンドしている画面イメージ

表 1: 不適切な内容における判定基準

Criteria 名	**解説**
No personal attacks	個人への攻撃を含まないこと
No discrimination	差別的な表現を含まないこと
No threats or violence	脅迫や暴力的な表現を含まないこと
No privacy violations	プライバシーの侵害がないこと
No obscene language	卑猥な言葉を含まないこと
No sexual content	性的な内容を含まないこと
Child-friendly	子どもにとって有害でないこと
No harassment	嫌がらせが含まれていないこと
No political promotion	政治的な主張や宣伝を含まないこと
No religious solicitation	宗教的な勧誘を含まないこと
Accurate info	情報が正確であること
No rumors	うわさ話や根拠のない情報を含まないこと
Correct health info	健康に関する正確な情報であること
Protection of brand image	ブランドイメージを損なわないこと
No defamation or unwarranted criticism	名誉毀損や根拠のない批判がないこと
Legal compliance and regulations	法令や規制に違反しないこと
Adherence to company policies	会社のポリシーに準拠していること

3 不適切表現の判定における評価

本研究では，2.2 節 (1) で述べた「送信されたコメントが不適切か否かを判定する機能」について，F1 スコアを用いて評価を行う．2.2 節 (2) および (3) の機能については，本研究では評価を実施しない．

本システムで評価に使用する LLM は Azure AI Foundry によって提供されている GPT-4o のモデルバージョン「2024-11-20」を使用する．GPT-4o は執筆時点 (2025/6) において，回答精度および料金コストのバランスが取れており本システムの実運用を見据えて採用に至った．

3.1 評価用データセット

評価用データとしてはデータ利用者によるフィードバックコメントを想定し筆者らが作成したオープンデータコメントデータセット[8]および kaggle にて公開されている Amazon の商品レビューデータセット[9]とする．

オープンデータコメントデータセットについては，実際のデータ利用者からのコメントを収集したものではなく，データ利用者のコメントを想定して筆者らが主観で作成したものである．本データは 138 件の日本語で記述されたコメントに対して適切か不適切を示すラベルに加えて，不適切な場合はその理由を示すラベルを筆者らが主観で付与している．

また，実際のコメントを使った評価を行うため，Amazon の商品レビューデータセットを使って評価を行う．本データは英語で記述された実際の Amazon における商品レビューデータセットに対してポジティブ，ネガティブのラベルが付与されている．データの件数としては 200,000 件あり，今回はそこから 2,000 件を無作為に抽出し評価する．商品レビューデータセットについてはオープンデータという本来のターゲットとは異なるがデータ数が十分である点，実際のユーザから得られたデータである点，何かを提供されたことによりフィードバックする行為がデータ利用者と類似している点を踏まえて評価データとして採用した．

3.2 評価結果

3.1 節で示したそれぞれのデータセットに対して正解率，適合率，再現率，F1 スコアを算出した結果を表 2 に示す．

表 2: 各データセットの正解率，適合率，再現率，F1 スコア

指標	オープンデータコメント	商品レビュー
正解率	0.9348	0.8680
適合率	0.9872	0.8321
再現率	0.9059	0.9139
F1 スコア	0.9448	0.8711

指標とデータセットの全ての組み合わせにおいて 0.83

[8] https://github.com/c-3lab/moral-keeper-ai/blob/main/benchmark/evaluate/data/ja/comments.txt

[9] https://www.kaggle.com/datasets/kritanjalijain/amazon-reviews

以上と高い値となった．また，再現率以外の全ての指標においてオープンデータコメントの方が高い値となった．

正解率については，オープンデータコメントと商品レビューのデータセットそれぞれ 0.9348，0.8680 と高い値となった．また，F1 スコアについてもそれぞれ 0.9448，0.8711 と高い値となった．

適合率についてはオープンデータコメントのデータセットの方が高く，再現率については商品レビューデータセットの方が高い結果となった．

3.3 不適切判定理由ごとの集計結果

表 3 にオープンデータコメントデータセットの評価における不適切判定理由ごとの総数，誤判数，誤判率ごとの集計結果を示す．また，表 4 に Amazon 商品レビューデータセットの評価における不適切判定理由ごとの総数，誤判数，誤判率を示す．

不適切判定理由には表 1 と 2.2 節で記載した Azure AI Foundry によるフィルター機能が動作するキーワード種別を加えた項目とする．

総数は対象データセットに対して不適切と判定された際に出力される理由を集計した結果である．また，総数の集計方法として一つのテストデータから複数の不適切理由が出力される場合があるが，それら全てを合算して算出する．誤判数は，対象テストデータに付与されいるラベルがポジティブだったにも関わらず，本システムが不適切と判定した場合，その際に出力された理由ごとに集計する．総数と同様に複数の理由を出力した場合についても，それら全て合算する．誤判率は総数に対する誤判数の割合を示す．

表 3 のオープンデータコメントの評価における不適切判定理由において最も多かったのが「Protection of brand image」および「No defamation or unwarranted criticism」となった．また，表 4 の Amazon 商品レビューの評価における不適切判定理由においては「No defamation or unwarranted criticism」となり，次に「Protection of brand image」となった．

誤判定について，オープンデータコメントのデータセットでは件数が少ないという問題もあり，ラベルがポジティブで不適切と判定したケースがほとんど発生しなかった．また，オープンデータコメントでは 3 つの理由で誤判数が集計されたが内訳としては，3 つとも全て一つのテストデータにおいて判定された結果であった．

Amazon 商品レビューの誤判率において Azure AI Foundry の不適切判定理由を除いた中ではて「No religious solicitation 」が 0.64 となり最も高かった．また,「Protection of brand image」は総数が 796 件と 2 番目に多い不適切判定理由にも関わらず誤判率が 0.06 と Azure AI Foundry の不適切判定理由を除いた中では最も低かった．

表 3: オープンデータコメントデータセットの評価における不適切判定理由ごとの総数，誤判数，誤判率

不適切判定理由	総数	誤判数	誤判率
No personal attacks	21	0	0.00
No discrimination	9	0	0.00
No threats or violence	6	0	0.00
No privacy violations	8	0	0.00
No obscene language	10	0	0.00
No sexual content	5	0	0.00
Child-friendly	23	0	0.00
No harassment	28	0	0.00
No political promotion	8	0	0.00
No religious solicitation	4	0	0.00
Accurate info	37	1	0.03
No rumors	35	0	0.00
Correct health info	8	1	0.12
Protection of brand image	46	0	0.00
No defamation or unwarranted criticism	46	0	0.00
Legal compliance and regulations	22	1	0.05
Adherence to company policies	23	0	0.00
Azure_Filter_hate	0	0	0.00
Azure_Filter_jailbreak	0	0	0.00
Azure_Filter_self_harm	0	0	0.00
Azure_Filter_sexual	1	0	0.00
Azure_Filter_violence	4	0	0.00

4 考察

2 の通り指標とデータセットの全ての組み合わせにおいて 0.83 以上と高い値となったため，本システムの構築で活用した LLM およびシステムプロンプトは有効であると考えられる．また，再現率以外の全ての指標においてオープンデータコメントの方が高い値を示し，表 3 に示されている誤判数が少ないのは，本システムの本来のターゲットだったためだと考えられる．しかし，オープンデータコメントのデータセット件数が少ないため誤判定が少なかった可能性もある．より正確な結果を得るには，今後さらに多くのデータを用いて評価を行う必要がある．

4.1 オープンデータコメントデータセットにおける再現率低下の考察

オープンデータコメントのデータセットでは，データ提供を行っている組織などに対して批判する意図で一見問題ないような表現を用いて行った場合，正しく判定できなケースがあった．データセットの中にはそのようなコメントをいくつか含めているため，再現率の低下につながったと考えられる．例えば次の文章はネガティブとしてのラベルが付与されているにもかか

表 4: Amazon 商品レビューデータセットの評価における不適切判定理由ごとの総数，誤判数，誤判率

不適切判定理由	総数	誤判数	誤判率
No personal attacks	112	8	0.07
No discrimination	46	8	0.17
No threats or violence	16	3	0.19
No privacy violations	21	5	0.24
No obscene language	103	20	0.19
No sexual content	56	21	0.38
Child-friendly	184	54	0.29
No harassment	150	14	0.09
No political promotion	38	15	0.39
No religious solicitation	44	28	0.64
Accurate info	212	20	0.09
No rumors	212	24	0.11
Correct health info	46	12	0.26
Protection of brand image	796	45	0.06
No defamation or unwarranted criticism	956	89	0.09
Legal compliance and regulations	45	11	0.24
Adherence to company policies	148	13	0.09
Azure_Filter_hate	5	0	0.00
Azure_Filter_jailbreak	1	1	1.00
Azure_Filter_self_harm	1	0	0.00
Azure_Filter_sexual	10	4	0.40
Azure_Filter_violence	1	0	0.40

わらず，不適切ではないと判定した.「データ公開されたんですね。60 時間分の残業代の支払いに期待します」このような対象の文章だけでは伝えたい内容を理解するのが難しく，データ提供を行っている組織の背景を理解していなければ不適切であることを判定することが難しかったと考える．しかし，データ提供を行っている組織における労務や給与の話をオープンデータのフィードバックとして記載する必要はないと考えられるため，本来は対象の文章を不適切であると判定すべきである．オープンデータと関係のないデータ提供を行っている組織に対する内容の記載は不適切と判定するようにシステムプロンプトを改善することで対象の文章においては正しく判定ができる可能性があると考える．

4.2 Amazon 商品レビューデータセットにおける適合率低下の考察

Amazon の商品レビューのデータセットにおいて適合率が低くなった理由として，レビュー文としてはポジティブな内容にもかかわらず商品名や書籍の中において不適切な表現が含まれており誤判定につながっていると考えられるケースがあった．

例えば次のような男性の除毛用機器に対するレビューコメントについてラベリングとしてはポジティブであったにもかかわらず本システムは不適切と判定した．“Man groomer does the job. Good buy for the price. Takes practice to get a technic that works, but overall good purchase.” レビュー内容としては，購入者は商品に対して満足しており一見問題がないような文章に見える．本システムのログを確認したところ，Azure AI Foundry によって性的な意味を示すラベルである「sexual」が付与されてエラーが返ってきており，フィルター機能が動作していた.「groomer」は本来の意味として「ペットの美容師・トリマー・馬の身体の世話係」となる．しかし，別の意味として子供や若い女性を性犯罪的な行為や人身売買の目的でネットなどで言葉巧みに誘い，手懐ける犯罪行為を行う人間を指す場合があることがわかった．対象のレビュー文は，購入を示唆するような文に加えて「groomer」というキーワードの前に「Man」というキーワードが配置されていることから，人身売買を想起させる内容としてフィルターが行われたと考える．

また，他にも CD の楽曲名に不適切なキーワードが含まれていることで同様に「sexual」において抵触してるケースも散見された．Amazon の商品に対するレビューのため対象となる商品名が多様であり誤判定が発生してしまうが，本来のターゲットであるオープンデータにおいてデータ提供者が政府・自治体の場合はオープンデータ名に不適切なキーワードが含まれることは少ないと考えられるため今回のような誤った判定は少ないと考える．表 4 の Azure AI Foundry による不適切判定理由において誤判定が発生しているため，フィルター機能が動作しない環境にて評価することで精度向上を見込める可能性があると考える．

表 4 において最も高い誤判率となった,「No religious solicitation」について考察する．対象の判定理由について誤判定しているテストデータを確認したところ，書籍のストーリ上で宗教に関連する登場人物やキーワードがあり，それらに対し称賛するようなレビューを記載しているレビューが散見された．レビューとしては称賛する文章のためラベルがポジティブだが，宗教的な内容を肯定しているため「No religious solicitation」に抵触し誤判定という結果になったと考えられる．

5 まとめ

本研究では，CKAN 上で動作するフィードバックシステムおよび LLM を活用したフィードバック内容の変換するシステムの開発と，その判定機能の評価を行った．まず CKAN の拡張機能 ckanext-feedback を通じて，ダウンロード数・コメント数・いいね数のリアルタイム可視化機能および「感謝」「要望」「質問」のジャンルによるコメント投稿機能，さらに活用事例紹介機能を提供することで，従来の一方向的な問い合わせ手段では把握困難であった利用者ニーズを具体的かつ定量的に提示し，提供者によるリソース配分や優先度判断の効率化を検討した．またコメント機能の自由入力

に起因して生じ得る提供者の心理的負荷を軽減するため，計 18 項目から構成される明確な判定基準を策定し，Azure AI Foundry（GPT-4o）を用いて投稿直後に自動判定を行う仕組みを実装した．不適切と判定されたコメントについては，同じく LLM によって建設的かつ礼節ある適切な表現へ自動変換し，原文と対比表示することでデータ利用者に対して再び投稿の内容を検討させるユーザインタフェースを考案した．この機能は利用者自身の行動変容を誘発するとともに，提供者に対してはネガティブコメントによるモチベーション低下を抑制する新たな介入手法として期待できる．

評価は，不適切投稿における判定機能に対して実施した．日本語の「オープンデータコメント」138 件および英語の「Amazon 商品レビュー」2 000 件を用いて正解率・適合率・再現率・F1 スコアを算出し，いずれの指標においても 0.83 以上，メインターゲットである「オープンデータコメント」に対しては F1 スコア 0.9448 を達成した．２つのデータセットの評価を比較したところドメイン適正が高い「オープンデータコメント」の方が総合的に高いスコアとなり，本ターゲットに適したプロンプト設計と判定基準が実装できたと考える．しかし，今回は２つのデータセットのみでの比較なので今後は追加で他のドメインにおいても検証を行なっていく必要があると考える．加えて,「オープンデータコメント」は筆者らが主観で作成したデータであり件数も少ないので，今後は客観的でより多くのテストデータで評価を行うべきと考える．一方誤検出事例の分析からは多義語によるフィルタリング誤りや，一見問題ないように見える文脈依存の批判的コメントにおいて判断が困難であることが明らかとなり，プロンプトのさらなるチューニングおよび LLM 提供サービスや LLM モデルの比較が今後の改良点として挙げられる．

今後は自治体や企業が実運用で蓄積したフィードバックデータを用いて，システム適用前後のデータ更新頻度や利用者満足度，提供者の継続率といった運用指標への影響を定量的に評価する必要がある．また，実際のデータ利用者から寄せられたコメントに対して評価を行い，その結果がオープンデータの利活用改善にどのように実現するかを明らかにする必要がある．さらには行動経済学的アプローチを取り入れた長期的なモチベーション維持メカニズムの解明を進めることで，オープンデータ流通の促進に大きく寄与できることを期待している．

参考文献

[1] デジタル庁, 「オープンデータ」, 2024. https://www.digital.go.jp/resources/open_data

[2] デジタル庁, 「オープンデータ取組済み自治体」, 2024. https://www.digital.go.jp/resources/data_local_governments

[3] デジタル庁, 「地方公共団体へのオープンデータの取組に関するアンケート結果・回答一覧」, 2024. https://www.digital.go.jp/resources/data_questionnaire

[4] D. Rudmark and M. Andersson, ”Feedback Loops in Open Data Ecosystems,” *arXiv preprint arXiv:2110.01023*, 2021. https://arxiv.org/abs/2110.01023

[5] C. J. Fong, M. S. Patall, D. Vasquez, and E. Stautberg, ”A Meta-Analysis of Negative Feedback on Intrinsic Motivation,” *Educational Psychology Review*, vol. 31, no. 1, pp. 121–162, 2019. https://selfdeterminationtheory.org/wp-content/uploads/2019/11/2019_FongPatallETAL_EdPsychReview.pdf

[6] Renáta Máchová, Miloslav Hub, and Martin Lněnička, “Usability Evaluation of Open Data Portals: Evaluating Data Discoverability, Accessibility, and Reusability from a Stakeholders ’ Perspective,” *Journal of Theoretical and Applied Electronic Commerce Research*, vol. 13, no. 1, pp. 43–57, 2018. https://www.researchgate.net/publication/325375462_Usability_evaluation_of_open_data_portals_Evaluating_data_discoverability_accessibility_and_reusability_from_a_stakeholders’_perspective

[7] Google, ”マップユーザーの投稿コンテンツに関するポリシー,” 2023. [Online]. Available: https://support.google.com/contributionpolicy/answer/7400114?hl=ja

[8] Praneeth Vadlapati, *Multilingual Prompting in LLMs: Investigating the Accuracy and Performance*, International Journal of Scientific Research in Engineering and Management (IJSREM), vol. 07, no. 02, pp. 1–7, Feb. 2023. https://arxiv.org/html/2505.15229v1

[9] Murray Shanahan, Kyle McDonell, and Laria Reynolds, *Role-Play with Large Language Models*, arXiv preprint arXiv:2305.16367, 2023. https://arxiv.org/abs/2305.16367