

総務省プラットフォームサービスに関する研究会 ヤフーの取り組みについて

2022年3月7日
ヤフー株式会社

プラットフォームサービスの運営の在り方検討会

● 会議体

プラットフォームサービスの運営の在り方検討会
(2020年6月設置、全4回開催)

● 目的

個人に対する誹謗中傷等を内容とする
投稿への対応を検討する

● 検討事項

- ・ 目指すサービスの方向性について
- ・ AIを用いた対策の効果と妥当性について
- ・ 措置基準や実績に関する透明性の確保について など

● 提言書の公表

- ・ 全4回の議論を経て、2020年12月23日公表
- ・ 誹謗中傷対策におけるエコシステムの構築と透明化を目指し、対応方針を明示

【構成員（敬称略）】

■ 座長

山本 龍彦（慶應義塾大学大学院法務研究科 教授）

■ 委員

小川 一（毎日新聞グループホールディングス 顧問）

沢田 登志子（一般社団法人ECネットワーク 理事）

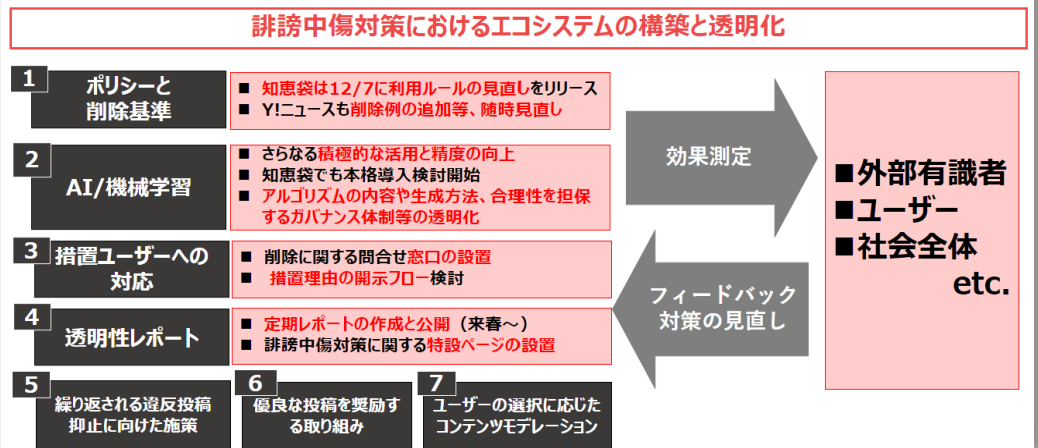
前嶋 和弘（上智大学 教授、総合グローバル学部長）

森 亮二（英知法律事務所 弁護士）

柳川 範之（東京大学大学院経済学研究科・経済学部 教授）

山口 真一（国際大学グローバル・コミュニケーション・センター 准教授）

提言に対するヤフーの対応方針



■ 提言書受領後の取組み

● ガイドラインにおける措置事例の明示

- 知恵袋の「利用ルール」やヤフーニュースの「コメントポリシー」をアップデート及び削除事例の追加。
- 必要に応じて適時のタイミングで継続的に見直しを行っている。

● 削除措置理由の明示

- ニュースコメントの投稿を削除されたユーザーより問い合わせがあった場合、CSより措置理由を開示する運用を開始。（2021年夏～）
※知恵袋は従前よりMy知恵袋ページで削除理由を確認可能（削除日から2週間）。

● 投稿停止措置に対する異議申し立て窓口を用意

- ニュースコメントにおいて投稿停止措置を受けたユーザーについては、異議申し立てのための問い合わせ窓口を用意し、措置の内容についてもあらためて確認を行っている。

● AIの積極的活用

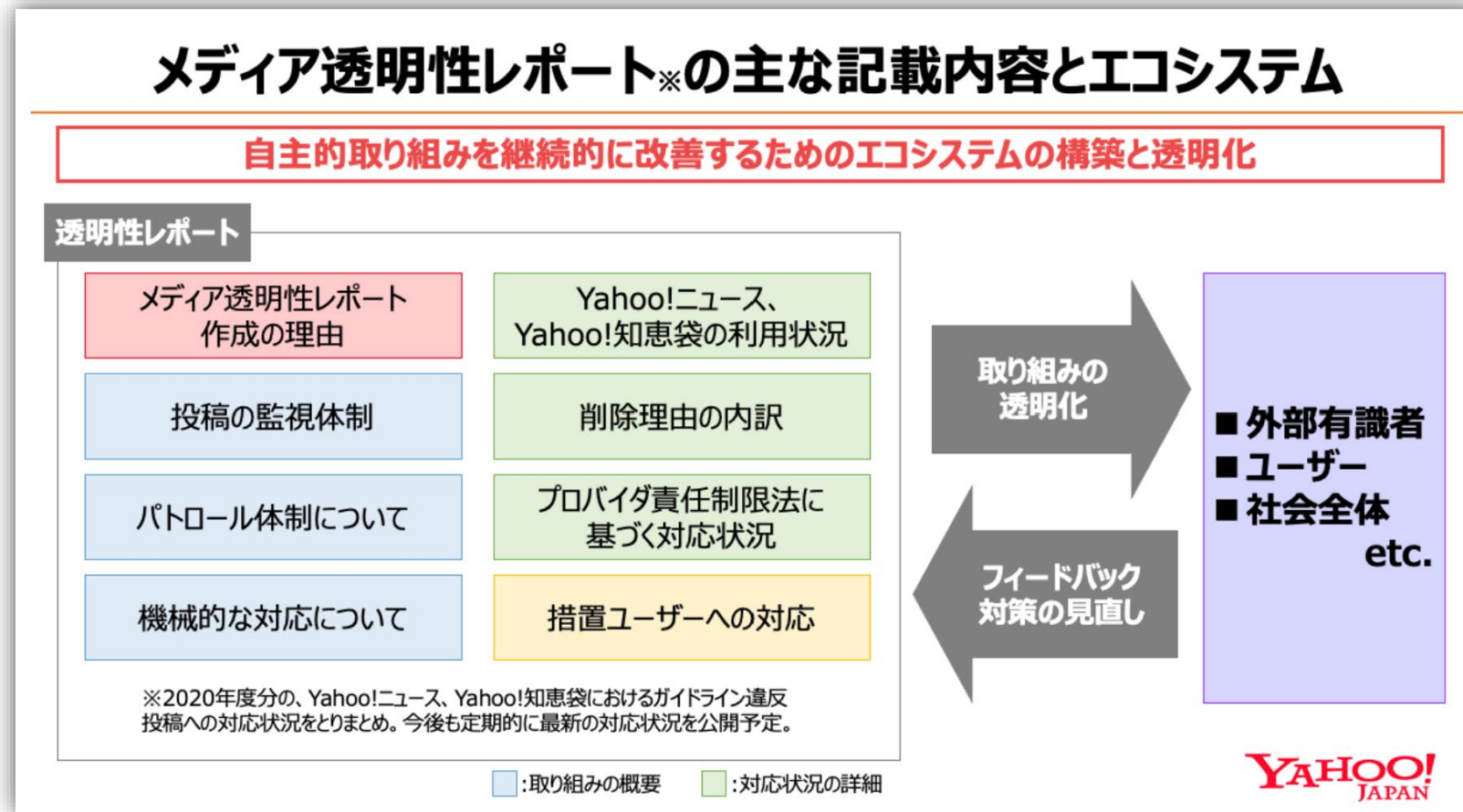
- ヤフーニュースコメントにおいて活用している建設的モデル（AI）について、外部への無償提供を開始（2021年5月～）、数社において導入が進んでいる。

プレスリリース：<https://about.yahoo.co.jp/pr/release/2020/09/18a/>
<https://about.yahoo.co.jp/pr/release/2021/05/19a/>
<https://about.yahoo.co.jp/pr/release/2021/07/19a/>

■ 透明性レポートの公表

誹謗中傷などのガイドライン違反投稿への対応状況をまとめた2020年度分のメディア透明性レポートを公表（2021年12月）

- 有識者会議提言書において透明性・アカウントビリティ確保の重要性についての記載があり、プラットフォーム事業者による自主的取り組みの内容について透明性レポートを作成。
- 主要な投稿型プラットフォームサービスである、「Yahoo!ニュース」のコメント欄および「Yahoo!知恵袋」を対象



透明性レポートのサマリ

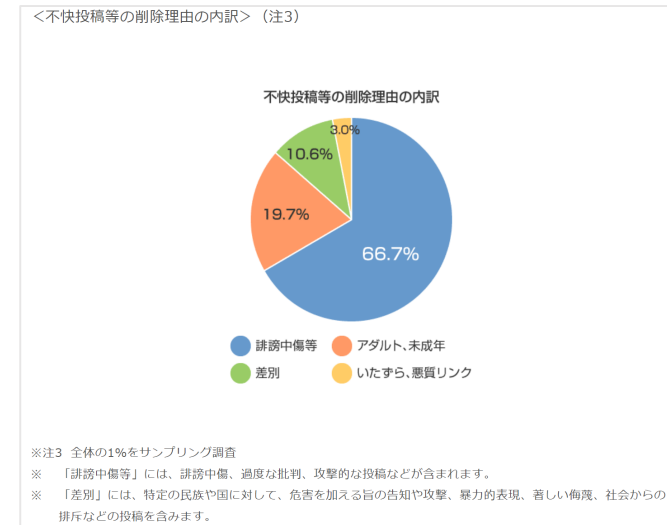
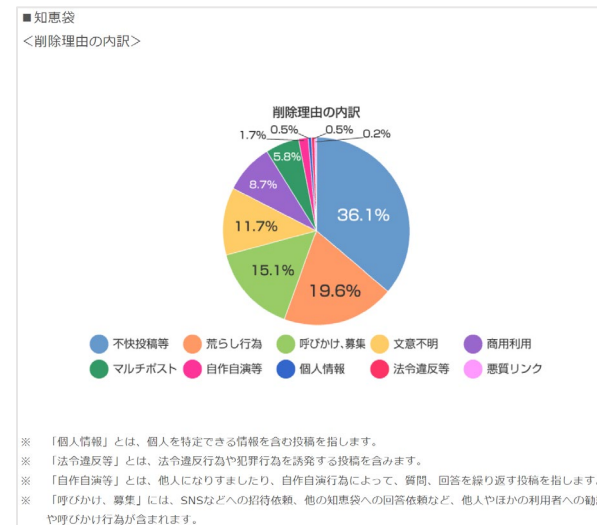
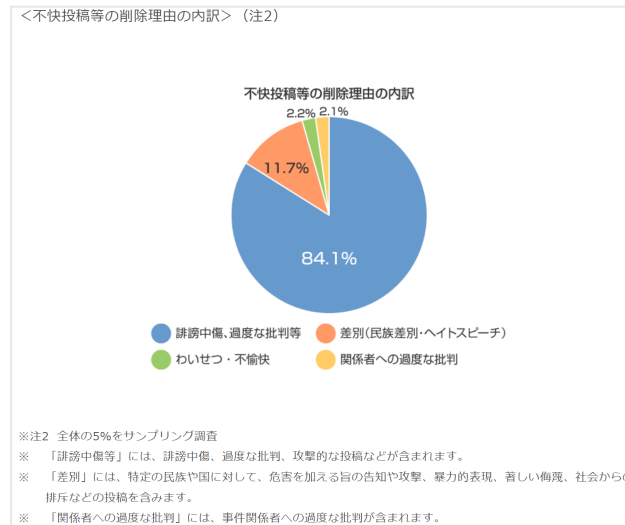
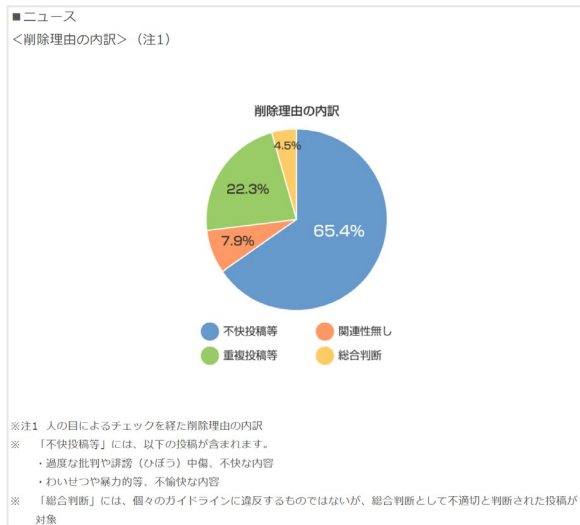
削除件数や削除理由内訳についても公表

ニュース

- 総投稿数に対するガイドライン違反を理由とした削除投稿数の割合は約3%
- 削除理由の内訳は、「不快投稿等」が最も多かった（約65%）
- 「不快投稿等」の詳細な内訳は、「誹謗中傷・過度な批判等」が最も多かった（約84%）

知恵袋

- 総投稿数に対するガイドライン違反を理由とした削除投稿数の割合は約7%
- 知恵袋における削除理由の内訳は、「不快投稿等」が最も多かった（約36%）
- 「不快投稿等」の詳細な内訳は、「誹謗中傷等」が最も多かった（約67%）



■ Yahoo!ニュースコメント欄における新たな施策

コメント欄の健全化に向けた3つの施策をリリース（2021年10月）

- 一部のコメント欄においては、「Yahoo!ニュース」が目指している目的を実現できていないという課題があった。
- 既存の取組みに加え、これまで以上に誹謗中傷への厳正な対策を講じるべく、3つの施策をリリースした。



1

**投稿時注意メッセージの対象者拡大と
投稿停止措置の強化**

2

記事単位のコメント欄の自動非表示

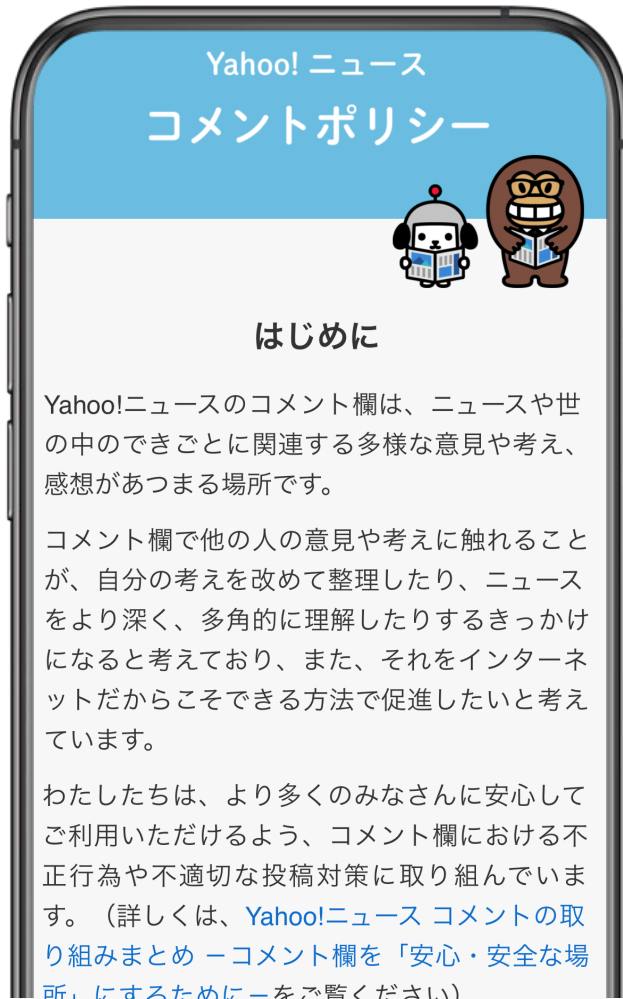
3

衆院選期間中の注意メッセージ掲出

プレスリリース : <https://about.yahoo.co.jp/pr/release/2021/10/19a/>

■ Yahoo!ニュースがコメント欄を提供する理由

- インターネットの双方向性という特徴を生かし、ユーザー個人にも発信の場を提供することで、さらなる情報の価値を創りたい
- その情報に触れた多くのユーザーの行動をサポートし、より良い社会の実現に貢献したい



Yahoo! ニュース
コメントポリシー

はじめに

Yahoo!ニュースのコメント欄は、ニュースや世の中のできごとに関連する多様な意見や考え、感想があつまる場所です。

コメント欄で他の人の意見や考えに触れることが、自分の考えを改めて整理したり、ニュースをより深く、多角的に理解したりするきっかけになると考えており、また、それをインターネットだからこそできる方法で促進したいと考えています。

わたしたちは、より多くの方々に安心してご利用いただけるよう、コメント欄における不正行為や不適切な投稿対策に取り組んでいます。(詳しくは、[Yahoo!ニュース コメントの取り組みまとめ - コメント欄を「安心・安全な場所」にするために](#)をご覧ください)



YAHOO! ニュース JAPAN

なぜ台風は怖い?被害を減らすために私たちができることは一専門家が解説

自治体の「警戒レベル」がわかりづらいと感じる

2年前、千葉で被害を受けました
まだ家の修理が完了していません

台風で避難した経験があり、
それから非常用持ち出し袋を常に用意しています

新しい発見があった

自分とは違う考え

とても共感できる

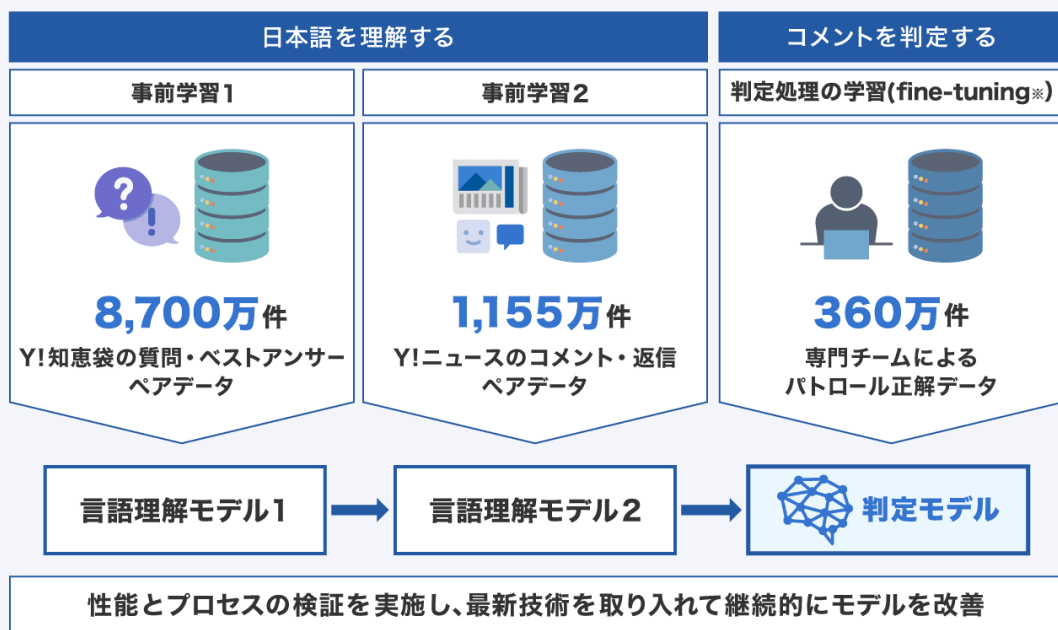
こういう考えもあるんだ

コメント欄非表示機能の概要

一定以上コメント数がある記事のコメント欄を対象に、AIが判定した違反コメント数などの基準に従い、コメント欄を自動的に非表示とする機能

- 恣意性を排除するため、AIが判定した違反コメント数など客観的な指標を用いた基準に従い、コメント欄を自動的に非表示。
- 活用するAIは、最先端の自然言語処理モデルとスパコン「kukai」を活用したAI判定モデルで、「Yahoo!ニュース コメントポリシー」のうち、過度な批判や誹謗中傷、差別、わいせつや暴力的などの項目に該当している度合いを総合的に点数化して自動削除するもの。

専門チームのパトロールによる削除判断を再現するモデル※を構築



※ 今回の施策で活用するモデルは、Yahoo! ニュース コメントポリシーのうち、過度な批判や誹謗中傷、差別、わいせつや暴力的などの項目に該当している度合いを総合的に点数化するもの

※ 膨大なデータで学習した既存モデルを用いfine-tuningを行うことで、少ないデータから性能の高いモデルを構築することが可能

AI活用イメージ



本機能の詳細 : <https://news.yahoo.co.jp/newshack/information/20211019.html>

コメント欄非表示機能導入後の状況

本機能による措置が適切に実行されているかについて有識者など外部の意見を伺って検証し、透明性をもって社会に示すため、導入開始後2ヵ月間の状況を公表（2022年1月）

＜本機能導入後の状況＞

- 導入後2ヵ月間において、非表示となった記事数は合計216件で、1日あたり平均3.5件（1日あたりの配信記事数に対して0.05%程度）
- 媒体の種類や記事の内容などを問わず、さまざまな記事のコメント欄が措置の対象となっている

非表示記事の件数や割合

| | |
|----------------------|--------|
| 1日あたりの 記事配信数(平均) | 7,511件 |
| 1日あたりの 非表示記事数(平均) | 3.5件 |
| コメント欄 非表示記事数 | 計216件 |
| 非表示になった 記事の割合 | 約0.05% |

集計期間：2021.10.19 - 2021.12.18

非表示となった記事の媒体種別

| | |
|-----------|-----|
| テレビ | 41件 |
| 一般紙・通信社 | 47件 |
| 海外メディア | 18件 |
| スポーツ紙・夕刊紙 | 31件 |
| 週刊誌 | 42件 |
| ネットメディア | 37件 |

集計期間：2021.10.19 - 2021.12.18

＜有識者からの意見＞

本件の検証にあたっては、「プラットフォームサービスの運営の在り方検討会」の有識者へ意見を伺った。

- コメント欄の非表示措置は、コメント欄に誹謗中傷などの投稿が集中しているような状況において、一定の有効な手段である
- 恣意的な運用とならないように非表示とする際の基準を明確にし、かつ、なぜこうした基準を採用するのかという理由を対外的に説明し尽していくことを前提に、本機能による施策を継続していく点についても支持する
- コメント欄の非表示措置は、建設的な意見も事実上削除してしまうことから慎重さが求められるため、違反者や違反投稿への対策をこれまで以上にきちんと両輪で進めるべきである

プレスリリース：<https://about.yahoo.co.jp/pr/release/2022/01/20a/>

■ 誹謗中傷投稿への対応について

● 不適切な投稿への直接的なアプローチ

→不適切な投稿の削除、反復継続される場合には、利用制限措置をする

● 不適切な投稿がなされないための環境整備

→利用者への注意メッセージや、分かりやすい禁止事項の規約、ポリシーの整備

→コメント欄非表示機能の導入

もともと、適切な投稿も巻き込んで非表示となるため、表現の自由への配慮についてのご意見もあり、その他の誹謗中傷投稿を減らしていくための施策も並行して検討をする必要がある。

今後も、外部の有識者の先生方のご意見などをいただきながら、コメント欄の非表示状況を振り返り、施策の見直しをするとともに、よりユーザーが安心して安全にサービスをご利用いただけるよう、サービスの品質向上に努めたい。

EOP