メディア透明性レポート（2021年度版）

メディア透明性レポートとは
【ニュース編】
【知恵袋編】
【ファイナンス掲示板】
【共通編】

メディア透明性レポートとは

ヤフーでは、ユーザーがそれぞれの関心分野において気軽に表現や意見表明・議論を行うことができる「場」として、複数の投稿型プラットフォームサービスを提供しています。

これらのサービスでは、それぞれ目的や特性に応じた利用のルールやガイドラインを定め、個人に対する誹謗中傷などの不適切な投稿を禁止し、違反行為に対して投稿の削除や投稿停止措置などの厳正な措置を行うなど、ユーザーが安心して利用できるよう様々な対策を実施しています。一方、現代において、投稿型プラットフォームは社会における情報流通の基盤として役割を有していることが指摘されています。ユーザーが生活の身近な問題から重要な社会問題に至るまで幅広い問題について、過度に委縮することなく自由に情報の発信を行うことができるようにしていくためには、関係法令も踏まえつつ、バランスの取れた対策を行っていくことが求められています。

そこで、ヤフーでは、各サービスにおける投稿削除の実績やその実施のための社内体制等について「メディア透明性レポート」としてまとめ、取組の内容について透明性を確保していくこととしています。
これによって、ユーザーや外部有識者にフィードバックをいただく機会を充実させることで、ヤフーにおける取組の検証・継続的な改善につなげていくというエコシステムを構築していきたいと考えています。

この「メディア透明性レポート」は、今回で２回目の公表となります。初回の公表は20年12月、社内に設置した「プラットフォームサービスの在り方に関する有識者会議」の提言を受けてのものでした。ヤフーでは、同提言を受け、今後とも、定期的にレポートを公表していくことと

しています。
今回のレポートは、21年度（21年4月1日～22年3月31日）の取組を対象としています。昨年度公表版と同様に、ヤフーにおける主要な投稿型プラットフォームサービスである「Yahoo!ニュースコメント欄」、「Yahoo!知恵袋」を対象とするとともに、新たに「Yahoo!ファイナンス掲示板」についても対象としています。

【ニュース編】

Yahoo!ニュース コメント欄について

1．Yahoo!ニュース コメント欄の提供目的

Yahoo!ニュースのコメント欄は、ニュースや世の中の出来事に関連する多様な意見や考え、感想が集まる場所です。Yahoo!ニュースでは、コメント欄で他のユーザー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います。また、インターネットの双方向性という特性を生かし、メディアによる情報発信に加えて、ユーザーが発信主体となる場を提供することで、さらなる情報の価値を創ることを目指しています。

2．禁止行為（コメントポリシー）について

Yahoo!ニュースのコメント欄が上記の目的に掲げたような「気づき」や「共感」を得られる場所であるためには、ユーザーに安心してご利用いただける環境が提供されていることが何よりも重要です。
Yahoo!ニュースでは、コメントポリシーを定め、投稿が禁止されているコメントや行為をわかりやすく示し、ユーザーに対し遵守をお願いしています。
なお、コメントポリシーにおいては、サービスを安全にご利用いただくため、不快な内容を含むコメントなどについて、必ずしも権利侵害や法令違反には至らない場合であっても禁止の対象としています。

・Yahoo!ニュース コメントポリシー

3．禁止行為（コメントポリシー）違反への対応について

コメントポリシーへの違反があった場合、対象投稿の削除を行うほか、違反の重大性や違反回数に応じ、投稿を行ったユーザーに対しコメントの投稿停止措置を行っています。また、投稿停止措置を受けたユーザーに対しては、Yahoo! JAPAN ID登録情報である携帯電話番号を削除の履

歴と照合し、同一の携帯電話番号を利用してYahoo! JAPAN IDの再取得等を行った場合、当該IDからのコメントの投稿を制限しています。

４．ユーザー自身によるコメントの非表示設定

記事を閲覧する際、コメント欄を表示させたくない場合には、ユーザー自身がコメント欄を非表示に設定することが可能です。また、特定のユーザーのコメントに限って表示させたくない場合には、ユーザー単位でコメントを非表示に設定することも可能です。
なお、非表示設定の内容は、設定したユーザー本人だけが確認できます。

21年度における投稿削除等の状況

１．投稿削除の状況

Yahoo!ニュースにおける21年度の投稿数は**1億5923.2万件（月平均1326.9万件）**で、投稿削除件数は**658.4万件（月平均約54.9万件 ）**でした。投稿削除件数は、年度を通じて投稿件数が増加したこともあり、前回公表時（21年3月）の月平均35万件から増加しています。
投稿件数のうち投稿削除件数が占める割合についてみると、21年度は**4.13%**となっており、昨年公表時の3.3%から増加しました。

（2023.9追記）
ニュースコメント欄における投稿削除件数について、AIによる自動削除件数の一部（関連度モデルによる投稿削除数）の計上が漏れていたことが判明したため、関連箇所の訂正を行っています（※）。
関係者の皆様には、深くお詫び申し上げます。
【訂正箇所】
投稿削除件数
（誤）513.1万件（月平均約42.8万件）
（正）658.4万件（月平均約54.9万件）

投稿削除割合（年度）※表中の四半期の数字も訂正しております。
（誤）3.22%
（正）4.13%

AIと人の目による削除割合（年度）
（誤）AIによる削除75.5%、人の目による削除24.5%
（正）AIによる削除80.9%、人の目による削除19.1%

２つのAIモデルによる投稿削除の割合（年度）

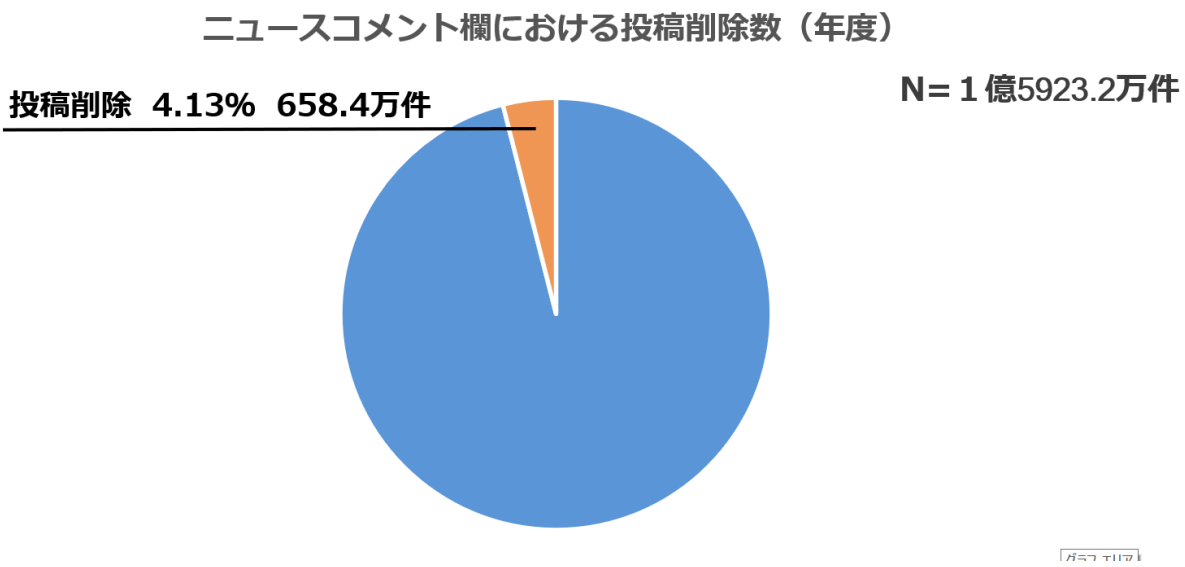
（誤）不適切判定モデルによる投稿削除数61.5%

関連度モデルによる投稿削除数38.5%

（正）不適切判定モデルによる投稿削除数72.7%

関連度モデルによる投稿削除数27.3%

（※）現時点において集計可能な期間の削除実績（月平均11.7万件）をもとに当該期間中の全投稿に占める当該削除実績の割合を算出し、これを残余の期間の全投稿数に乗じた数値（145.3万件）を用いています。



投稿削除件数及び削除割合の四半期ごとの推移は、次のとおりです。21年7-9月期には、緊急事態宣言の発出や東京オリンピック・パラリンピックの開催等を背景として投稿件数が大幅に増加し、これに伴って投稿削除件数も増加しました。

一方、10-12月期以降についてみると、投稿件数は依然として前回公表時を上回る水準で推移しているものの、ピークであった7-9月期と比較すると投稿件数・投稿削除件数ともに減少に向かっており、22年1-3月期の投稿削除件数は、7-9月期と比べて半数程度となっています。

投稿件数のうち投稿削除件数が占める割合についても同様の傾向を示しており、21年7-9月期にピークに達しましたが、10-12月期以降は減少傾向となっています。

ニュースコメント欄における投稿数・投稿削除数及び削除割合（四半期）

	投稿数	投稿削除数	削除割合
21年4－6月（月平均）	**3964.6万件**（1321.5万件）	**160.5万件**（53.5万件）	**4.1%**※1
7－9月（月平均）	**4847.8万件**（1651.9万件）	**229.2万件**（76.4万件）	**4.7%**
10－12月（月平均）	**3786.0万件**（1262.0万件）	**151.7万件**（50.6万件）	**4.0%**
22年1－3月（月平均）	**3324.8万件**（1108.3万件）	**116.9万件**（39.0万件）	**3.5%**
年度合計（月平均）	**1億5923.2万件**（1326.9万件）	**658.4万件**（54.9万件）	**4.1%**
参考（21年3月）	1,050万件	35万件	3.3%

※1　四捨五入の関係で、左記と割合が一致しないことがあります

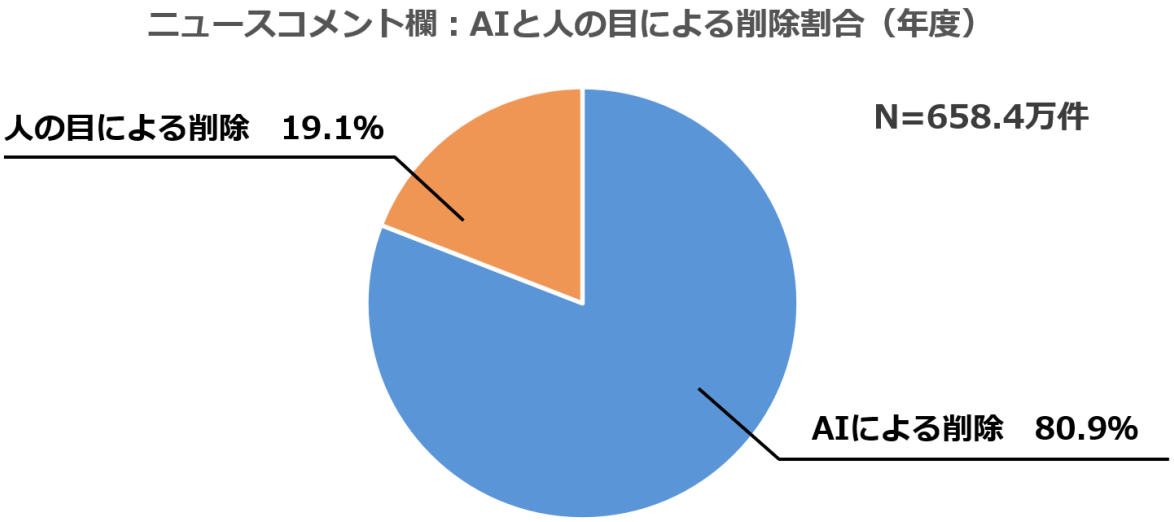
２．AIと人の目による削除の状況

（１）AIと人の目による削除の割合

膨大な数の投稿について的確に違反投稿であるかどうかの判定を行っていくためには、人の目だけでなく、機械の力を活用することが欠かせません。そこで、Yahoo!ニュースでは、AIと専門チームによる「人の目」の双方を組み合わせて投稿削除の対応を行っています。

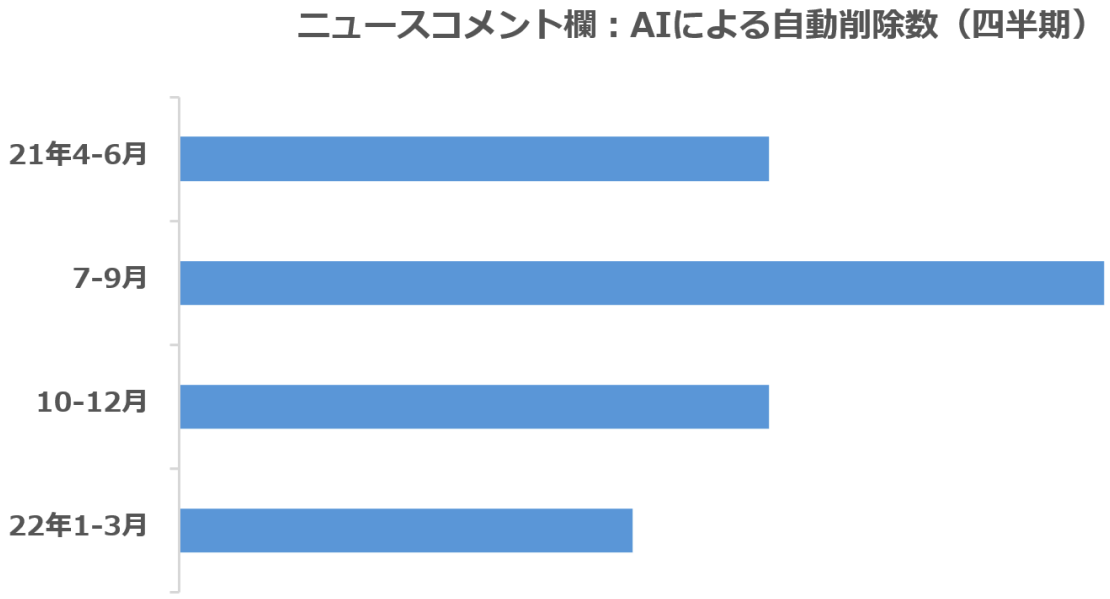
21年度における投稿削除件数のうち**80.9%**は、AIによりコメントポリシーに抵触していることが明白であると判断され、自動削除されたものです。AIによる自動削除は投稿から概ね数秒以内に行われており、多くの違反投稿がユーザーの目に触れる前に削除されているものと考えられます。

一方で、多岐にわたる内容の投稿について、ニュアンスを踏まえた丁寧な判定を行っていくためには、専門チームにおいて「人の目」による判定も欠かせません。21年度において「人の目」を経て削除が行われた投稿の割合は**19.1%**でした。



（２）AIによる自動削除の状況

Yahoo!ニュースでは、AIモデルを活用して違反投稿であるか否かの判定を行っています。以下では、その状況を示す例として、AIによる自動削除件数について四半期ごとの推移を示しています。先述のとおり、Yahoo!ニュースでは21年7-9月期において投稿件数の大幅な増加がみられましたが、同時に自動削除数も大きく増加しており、このことは、同時期においてAIが違反投稿の発見・削除に一定程度、有効に機能していたことを示唆しています。



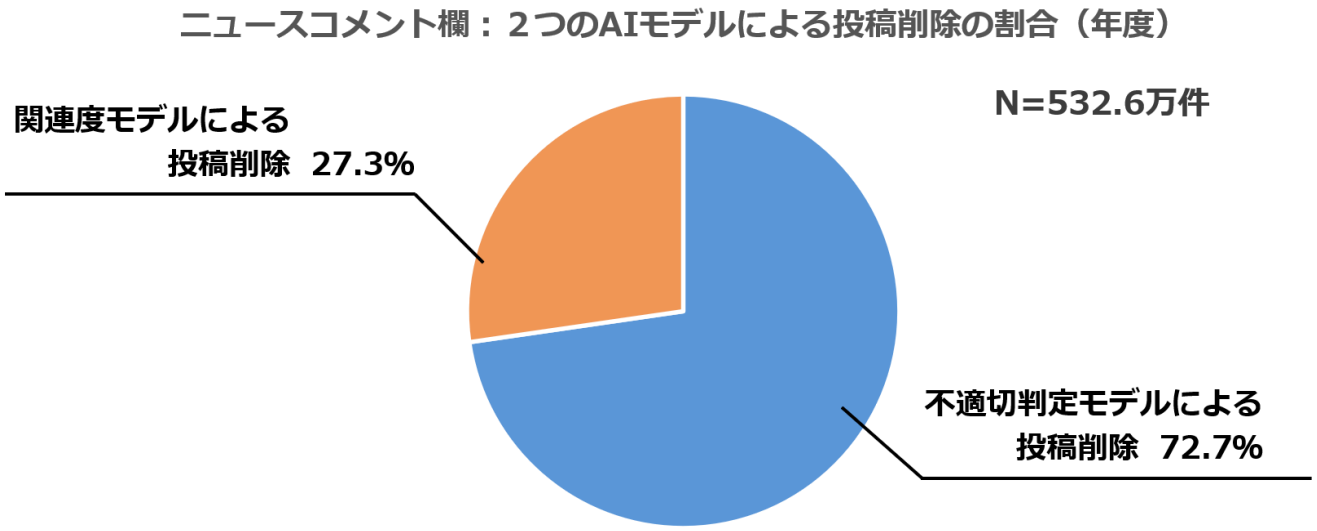
Yahoo!ニュースでは、違反投稿であるかどうかの判定や表示順の決定にあたり複数のAIモデルを使用していますが、違反投稿の自動削除には、このうち、以下の２種類のAIモデルを使用しており、それぞれ異なる観点から違反投稿の判定を行っています※2。

・**不適切投稿判定モデル**：過度な批判や誹謗中傷、差別、わいせつや暴力的等のガイドラインの項目に違反するおそれのあるコメントをAIが点数化

・**関連度モデル**：ニュース記事とコメントの関連度をAIが判定

21年度においてAIにより自動削除された投稿について、上記の２つのどちらのモデルにより判定されたかものであったのかについてみてみると、不適切判定モデルによるものが**72.7%**、関連度モデルによるものが**27.3%**でした。

※2　各モデルの具体的な仕組みについては、「ヤフーにおける投稿監視体制」の項目で詳しく紹介しています。



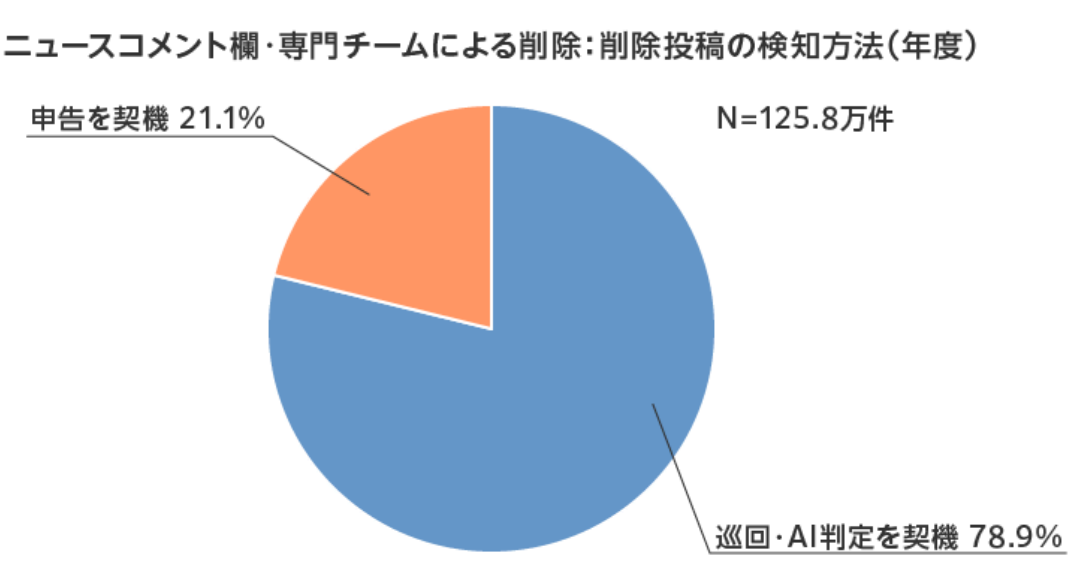
（３）専門チームによる削除の状況

専門チームによる投稿の削除は、ユーザーからの違反申告を契機とするものと、専門チームによる積極的な巡回・AIによる判定を契機とするものとに大別されます。

ユーザーからの違反申告については、投稿ごとに「非表示・申告ボタン」を設置し、コメントポリシーに違反すると思われる投稿について、ユーザーからの情報を幅広く受け付けています。申告が寄せられた投稿については、専門チームが１件ずつ投稿削除等の対応について検討を行っています。

一方で、違反申告がない投稿についても、積極的な検知に努めています。専門チームのスタッフがコメント欄を積極的に巡回して目視確認しているほか、「関連性モデル」により自動削除の対象とならないものの違反の可能性があると判断され、専門チームによる目視確認による検討フローに移されるものもあります。

21年度において専門チームにより「人の目」を経て削除された投稿のうち、申告を契機として削除されたものの割合は**21.1%**であり、**78.9%**と大半が専門チームによる積極的な巡回・AIによる判定を契機として削除されたものでした。



（違反申告受付件数）

21年度に受け付けたユーザーからの違反申告の件数は**230.0万件（月平均18.6万件）**でした。１件の投稿に対して複数の違反申告が寄せられることもあり単純比較は困難ですが、これを年間の投稿件数に対する割合としてみると**1.4%**に当たります。

Yahoo!ニュースでは、ユーザーからの違反申告を促進することによって、より迅速に違反投稿の検知が可能となると考えています。このため、後述の「2021年度の新たな取組」の項で紹介しているとおり、21年12月に違反申告の導線をユーザーによりわかりやすい形に改善しています。

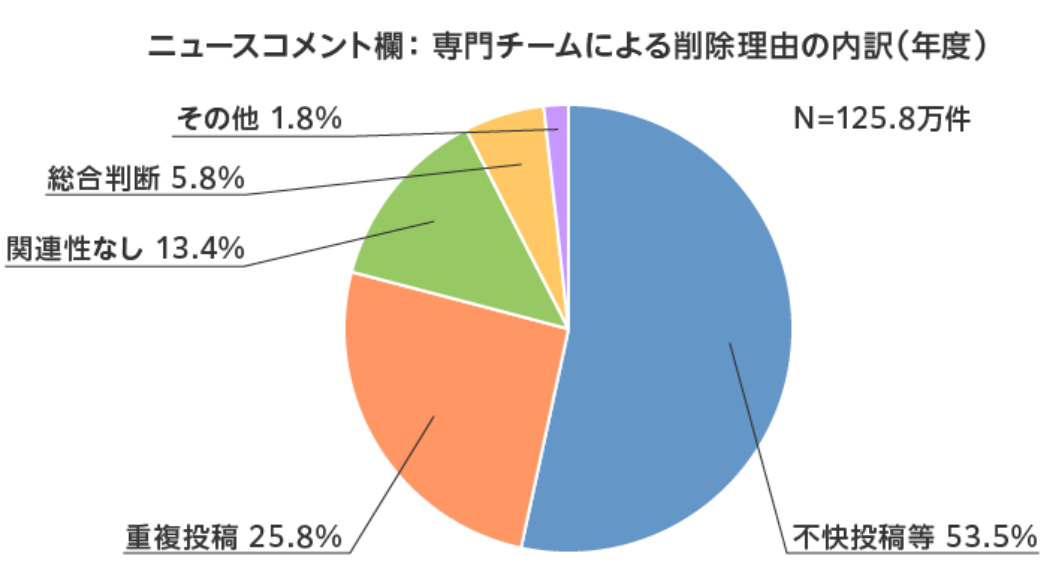
（削除理由の内訳）

専門チームにより削除された投稿について削除理由の内訳について、ここではまず、誹謗中傷やヘイトスピーチ、アダルト関係などユーザーが不快に感じると思われる「不快投稿等」と、それ以外の投稿の割合を見ていきます。

21年度において「不快投稿等」として専門チームにより削除された投稿の割合は**53.5%**を占めており、続いて割合が高い順に「重複投稿」「関連性なし」「総合判断※3」「その他※4」となっています。

※3　「総合判断」には、極端に短く否定的・攻撃的である投稿、特定のユーザーへの執拗な付きまといやあおり行為に該当する投稿、いわゆる「荒らし投稿」等が含まれます。

※4　「その他」にはいたずら投稿、宣伝・広告的な利用。個人情報の投稿、投稿行為が法令違反を構成しうるもの等が含まれます。

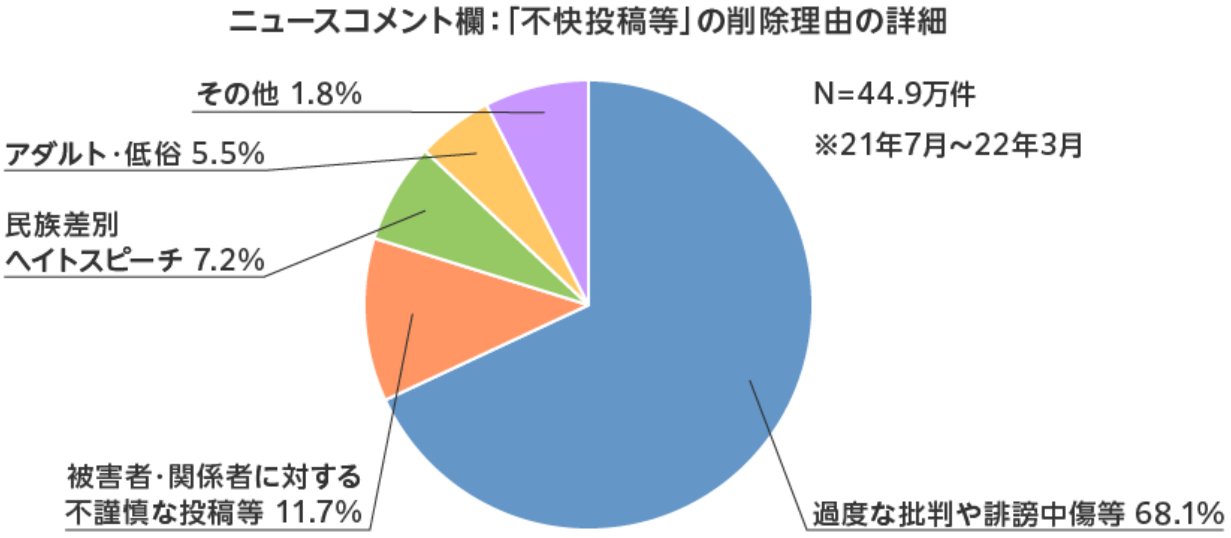


（不快投稿等の内訳）

次に、「不快投稿等」として削除されたものについて、更に理由の内訳をみると、**68．1%が「過度な批判や誹謗中傷等」**に該当するとして削除されたものであり、続いて割合が高い順に「被害者・関係者に対する不謹慎な投稿等※5」「民族差別・ヘイトスピーチ」「アダルト・低俗」を理由としたものとなっています※6。

※5　配慮に欠ける推測で、本人や関係者が目にしたら傷つくような投稿や、遺族感情を逆なでる投稿、亡くなった方をおとしめる投稿等が該当します（例：訃報や災害の発生に対して「おめでとうございます」と投稿）

※6　なお、本項目についてはデータの取得方法をサンプリング調査から全数調査に変更したため、対象時期は21年7月～22年3月となっています。



３．プロバイダ責任制限法に基づく削除の状況

特定電気通信役務提供者の損害賠償責任の制限及び発信者情報の開示に関する法律（以下「プロバイダ責任制限法」といいます。）は、プロバイダ等の損害賠償責任の制限、発信者情報の開示

請求等や発信者情報開示命令事件に関する裁判手続について定めた法律です。
ヤフーは、同法上のプロバイダ（「特定電気通信役務提供者」）として、提供している各投稿型プラットフォームサービスについて、利用規約に違反する投稿などによって権利を侵害された方から、同法に基づく投稿削除請求や投稿者に関する情報の開示請求を受け付けています。
Yahoo!ニュースにおける21年度の削除・開示請求の受付及び実施の実績を裁判外の請求及び裁判上の請求（訴訟）に分けて示すと次のとおりです。

（１）裁判外の請求

削除請求件数：3件（20年度：14件）	
	うち削除（一部削除を含む）に至った件数：1件（同：8件）
開示請求件数：6件（同：7件）	
	うち開示（一部開示を含む）に至った件数：0件（同：0件）

前年度と比較して21年度は削除請求件数が減少した一方で、開示請求件数は概ね同程度でした。削除・開示請求ともに、全件が名誉・信用・プライバシーの侵害を理由とするものでした。
なお、行政機関からの削除請求はありませんでした。

（２）裁判上の請求（訴訟）

削除請求件数：2件（20年度：2件）	
	うち削除（一部削除を含む）に至った件数：2件（同：1件）
開示請求件数：10件（同：11件）	
	うち開示（一部開示を含む）に至った件数：8件（同：6件）

21年度は削除請求件数・開示請求件数ともに前年度と同程度でした。なお、削除・開示請求ともに、全件が名誉・信用・プライバシーの侵害を理由とするものでした。

４．アカウント停止措置・問い合わせの状況

Yahoo!ニュースでは、コメントポリシーへの違反があった場合、対象投稿の削除を行うほか、違反の重大性や違反回数に応じ、投稿を行ったユーザーに対しコメントの投稿停止措置を行っています。

また、投稿停止措置を受けたユーザーに対しては、Yahoo! JAPAN ID登録情報である携帯電話番号を削除の履歴と照合し、同一の携帯電話番号を利用してYahoo! JAPAN IDの再取得等を行った場合、当該IDからのコメントの投稿を制限しています。
期間中に投稿停止措置を受けたIDの数は、**月平均で約1,140件**※7でした。

投稿停止措置に対して疑問・意見等があるユーザーからは、［お問い合わせ］フォームにより問い合わせを受け付けています。期間中に投稿停止措置を受けたユーザーから受けた問い合わせの件数は、**月平均で約340件**※8でした。

※7　21年6月～22年3月の平均
※8　21年11月～22年3月の平均

【知恵袋編】

Yahoo!知恵袋について

１．知恵袋の提供目的

Yahoo!知恵袋（以下「知恵袋」といいます）は日常のあらゆる疑問をほかのユーザーに向けて質問したり、それらの質問に回答することで、疑問を解決していく「知恵」共有サービスです。困っている人や課題を持っている人が、知恵袋を通じて「答え」や「気づき」を得られ、「答え」や「気づき」を提供した人は、その人の役に立ったという「助け合いの世界を作りたい」と考え、サービスを提供しています。

２．禁止行為（利用のルール）について

知恵袋では、すべてのユーザーにとって、安心・安全な環境でサービスを楽しんでいただくことを、何よりも重要だと考えています。
ユーザーの意見が誰かの「気づき」や「知恵」の種となるよう、質問や回答について自由な投稿を可能としていますが、同時にすべてのユーザーに対して、利用のルールを遵守して思いやりをもってご利用いただくとともに、誹謗中傷その他利用のルールに違反するご利用はご遠慮いただくよう呼び掛けています。

（利用のルールへのリンク）
https://chiebukuro.yahoo.co.jp/topic/guide/rule/

例えば、個人情報の書き込み、法令に違反するもの、悪質なリンク、不快に感じるもの、誰かを著しく傷つけたり、攻撃したりするような内容、などは投稿が制限されます。また、コミュニティサービスガイドライン、利用規約に抵触するような投稿もできません。

３．禁止行為（利用のルール）違反への対応について

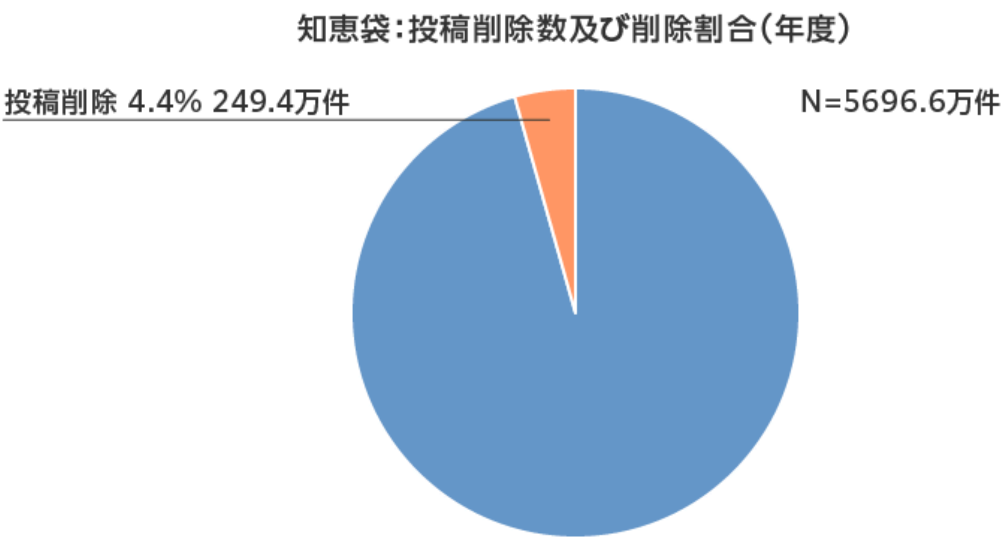
投稿が利用のルールに違反していると判断した場合には、投稿の削除を行います。また、一定期間内に利用ルール違反にあたる投稿を一定数行ったユーザーは、知恵袋を1週間、利用できなくなります（一時利用停止）。さらに、繰り返し利用ルール違反にあたる投稿を行い、複数回にわたって一時利用停止されたアカウントに対しては、知恵袋の利用停止を行うことがあります。
これらの措置に加え、明らかに悪意のある行為や不正利用には、知恵袋の利用停止およびYahoo! JAPAN IDの利用停止といった措置を予告なく行う場合があります。

このほか、禁止事項に該当する蓋然性が高いと判断された投稿は、不適切な投稿として質問詳細ページにおいて非表示となることがあります。非表示状態となっている投稿は、ユーザーが「表示する」ボタンを押下すると、その内容を確認することができます。

21年度における投稿削除等の状況

１．投稿削除の状況

21年度の投稿件数は**5696.6万件（月平均474.7万件）**であり、前回公表時（月平均450万件）からやや増加しました。これに対し、投稿削除件数は**249.4万件（月平均20.8万件）**と、前回公表時（21年3月）の月平均23万件と比較して**やや減少**しています。
投稿件数のうち投稿削除件数が占める割合についてみると、21年度は**4.4%**であり、昨年公表時の5.1%からやや低下しています。



投稿削除件数及び削除割合の四半期ごとの推移は、次のとおりです。

知恵袋における投稿数･投稿削除数及び削除割合（四半期）

	投稿数	投稿削除数	削除割合
21年4－6月（月平均）	1353.5万件（451.2万件）	64.5万件（21.5万件）	4.8%※9
7－9月（月平均）	1498.4万件（499.5万件）	65.9万件（22.0万件）	4.4%
10－12月（月平均）	1361.9万件（454.0万件）	53.6万件（17.9万件）	3.9%
22年1－3月（月平均）	1482.8万件（494.3万件）	65.4万件（21.8万件）	4.4%
年度合計（月平均）	5696.6万件（474.7万件）	249.4万件（20.8万件）	4.4%
参考（21年3月）	450万件	23万件※10	5.1%

※9　四捨五入の関係で、左記と割合が一致しないことがあります
※10 2020年度版から数値の訂正を行っています。詳しくは、2020年度版の該当箇所をご確認ください。
https://about.yahoo.co.jp/common/transparencyreport_2020/

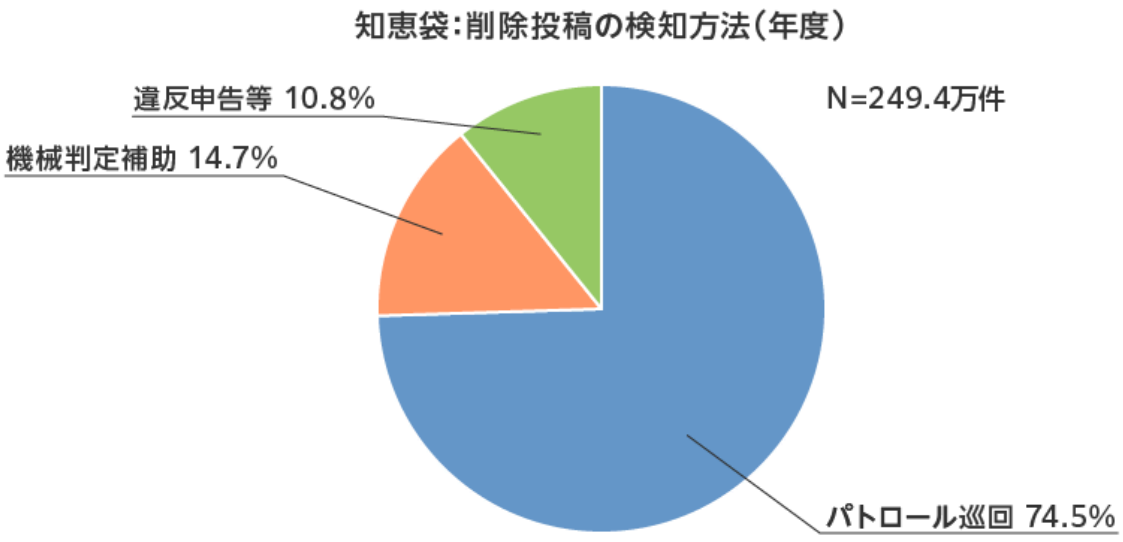
２．ＡＩと人の目による削除の状況

（１）違反投稿の検知状況

膨大な数の投稿について的確に違反投稿であるかどうかの判定を行っていくためには、人の目だけでなく、機械の力を活用することが欠かせません。そこで、知恵袋では、AIと専門チームによる「人の目」の双方を組み合わせて投稿削除の対応を行っています。
知恵袋で用いられているAIモデルは、専ら違反投稿の検知（機械判定補助）を目的としたもので、AIにより違反投稿である蓋然性が高いと判断された投稿が「人の目」による審査フローに移される仕組みとなっています※11。
また、ユーザーからの違反申告については、投稿ごとに違反申告フォームを設け、利用のルールに違反すると思われる投稿について、ユーザーからの情報を幅広く受け付けています。
一方で、違反申告がない投稿についても、積極的な検知に努めています。専門チームのスタッフがコメント欄を定期的に巡回して目視確認しています。
上記を通じて検知された投稿については、１件１件「人の目」により慎重な審査を経た上で、違反投稿であるかどうかの判定が行われています。
知恵袋では、21年度において削除された投稿のうち**74.5%**は専門チームによる積極的なパトロ

ール巡回により検知・削除されたものでした。一方で、機械判定補助を受けて削除に至った投稿の割合は**14.7%**、違反申告を契機として削除された投稿の割合は**10.8%**となっています。

※11 知恵袋では、Yahoo!ニュースと異なり知恵袋の投稿についてAIが自動削除を行うことはありません。



（違反申告受付件数）

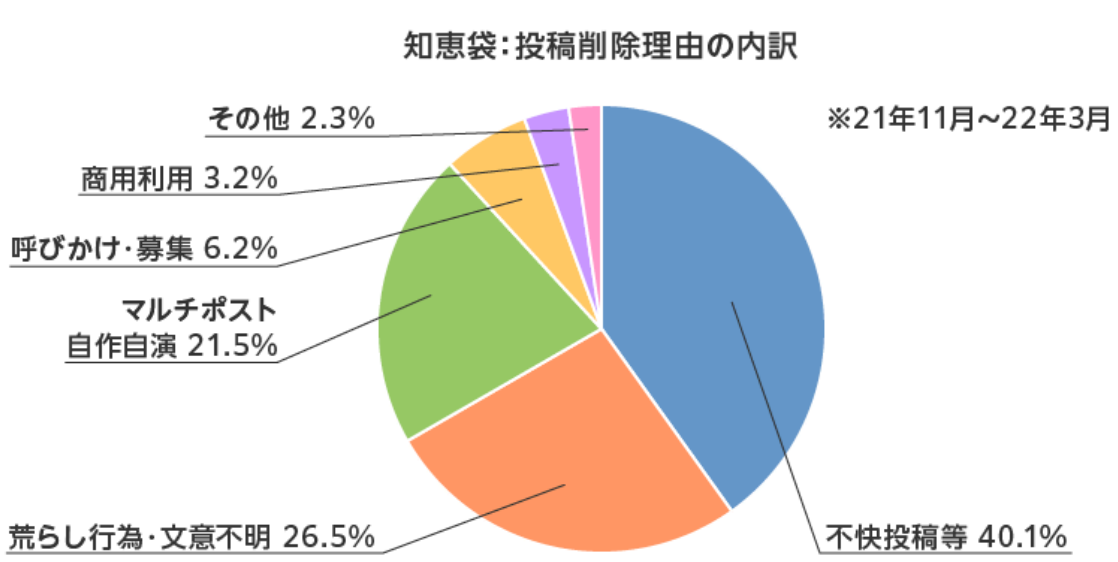
21年度に受け付けたユーザーからの違反申告の件数は、**212.9万件（月平均17.7万件）**でした。１件の投稿に対して複数の違反申告が寄せされることもあり単純比較は困難ですが、これを年間の投稿件数に対する割合でみると、**3.7%**に当たります。

（２）削除理由

利用のルール違反として削除された投稿について、ここではまず、誹謗中傷やヘイトスピーチ、アダルト関係などユーザーが不快に感じると思われる「不快投稿等」と、それ以外の投稿の割合を見ていきます。

21年11月～22年3月※12において「不快投稿等」として削除された投稿の割合は**40.1%**であり、続いて割合が高い順に「荒らし行為・文意不明」「マルチポスト・自作自演」「呼びかけ・募集」「商用利用」「その他※13」となっています※14。

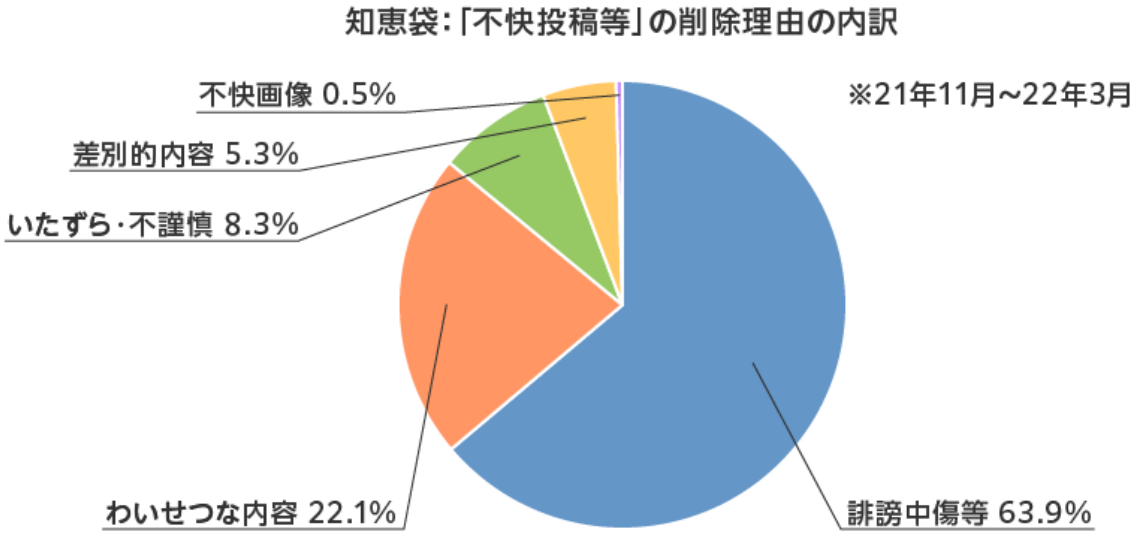
※12 集計の都合上、21年11月～22年３月のデータを掲載しています。
※13 「その他」には投稿行為が法令違反を構成するもの、個人情報の投稿、悪質なサイトへのリンクが含まれます。
※14 ここでは、投稿停止措置の対象となり得る違反投稿について分類しています。



（不快投稿等の内訳）

次に、「不快投稿等」として削除されたものについて、更に理由の内訳を見ると、「誹謗中傷等」が**63.9%**を占めており、続いて、割合が高い順に「わいせつな内容」「いたずら・不謹慎」「差別的内容」「不快画像※15」となっています。

※15 「いたずら・不謹慎」「不快画像」とは、それぞれ社会通念上、一般人が不愉快だと感じる可能性のある内容の投稿又は投稿を指します



（３）投稿の非表示措置

知恵袋では、利用のルールに違反する投稿は、不適切な投稿として、質問詳細ページにおいて非表示措置が行われることがあります。

これは、知恵袋が採用している低品質投稿判定モデルが個別の投稿についてスコアを付与し、これが一定の閾値以下となった投稿について、自動的に投稿の非公開措置を行っているものです。

21年度において、AIにより自動的に非表示措置が行われた投稿数は、**717.7万件（月平均59.8万件）**であり、投稿件数に対する比率としては**12.6%**を占めています※16。

※16 非表示となった後に投稿者自らにより削除されたものも含まれており、現に表示されている投稿のうち12.9%が非表示措置となっていることを示すものではありません。

３．プロバイダ責任制限法に基づく削除の状況

特定電気通信役務提供者の損害賠償責任の制限及び発信者情報の開示に関する法律（以下「プロバイダ責任制限法」といいます。）は、プロバイダ等の損害賠償責任の制限、発信者情報の開示請求等や発信者情報開示命令事件に関する裁判手続について定めた法律です。
ヤフーは、同法上のプロバイダ（「特定電気通信役務提供者」）として、提供している各投稿型プラットフォームサービスについて、利用規約に違反する投稿などによって権利を侵害された方から、同法に基づく投稿削除請求や投稿者に関する情報の開示請求を受け付けています。
知恵袋における21年度の請求の受付及び削除・開示等の実績について、裁判外での請求と裁判上の請求（訴訟）を分けて示すと次のとおりです。

（１）裁判外の請求

①行政機関※17からの請求

削除請求件数：10件（20年度：11件） （内訳※18：名誉・信用・プライバシー10件）	
	うち削除（一部削除を含む）に至った件数：5件（同：5件） （内訳：名誉・信用・プライバシー5件）

②行政機関以外からの請求

削除請求件数：99件（20年度：93件） （内訳：名誉・信用・プライバシー95件、著作権2件、肖像権1件）	
	うち削除（一部削除を含む）に至った件数：36件（同：43件） （内訳：名誉・信用・プライバシー35件、肖像権1件）
開示請求件数：15件（同：21件） （内訳：名誉・信用・プライバシー12件、著作権2件、肖像権1件）	
	うち開示（一部開示を含む）に至った件数：2件（同：1件） （内訳：名誉・信用・プライバシー2件）

※17 全件が法務省人権擁護局（地方支分部局（法務局）を含む）からの請求でした。
※18 内訳には重複が含まれることがあります（以下同じ）

前年度と比較すると、21年度における行政機関からの（削除）請求件数はほぼ同程度でした。
また、行政機関以外からの削除・開示請求件数をみると、削除請求件数は概ね同程度であった一方で、開示請求件数は減少しています。
いずれの請求についても、そのほとんどが名誉・信用・プライバシーの侵害を理由とするものでした。

（２）裁判上の請求（訴訟）

削除請求件数：11件（20年度：4件） （内訳：名誉・信用・プライバシー10件、著作権4件）	
	うち削除（一部削除を含む）に至った件数：6件（同：0件） （内訳：名誉・信用・プライバシー6件）
開示請求件数：7件（同：9件） （内訳：名誉・信用・プライバシー7件、著作権1件）	
	うち開示（一部開示を含む）に至った件数：4件（同：5件） （内訳：名誉・信用・プライバシー4件）

前年度と比較すると、21年度における削除請求件数は増加した一方、開示請求件数はやや減少しています。
いずれの請求についても、そのほとんどが名誉・信用・プライバシーの侵害を理由とするものでした。

４．アカウント停止措置・異議申し立ての状況

知恵袋では、一定期間内に利用ルール違反にあたる投稿を一定数行ったユーザーは、知恵袋を1週間利用できなくなる（一時利用停止）ほか、繰り返し利用ルール違反にあたる投稿を行い、複数回にわたって一時利用停止されたアカウントに対しては、知恵袋の利用停止を行うことがあります。
このほか、明らかに悪意のある行為や不正利用には、知恵袋の利用停止およびYahoo! JAPAN IDの利用停止といった措置を予告なく行う場合があります。
21年において、知恵袋において利用停止の対象となった**ID数は約4,520件（月平均約380件）でした。**

利用停止措置に対して疑問・意見等があるユーザーからは、［お問い合わせ］フォームにより問い合わせを受け付けています。21年度において、投稿停止措置を受けたユーザーから受けた問い合わせの件数は、**月平均で10件前後でした。**

【ファイナンス掲示板】

ファイナンス掲示板について

１．ファイナンス掲示板の提供目的と仕組み

Yahoo!ファイナンスの掲示板（以下「ファイナンス掲示板」といいます）は、株式、為替、FXなどの話題について、ユーザー相互が情報交換を行うことを目的として提供しています。
ファイナンス掲示板の仕組みは、銘柄別の掲示板とそれ以外の掲示板で異なっています。銘柄別の掲示板は、１銘柄１スレッドで構成されており、ユーザーがスレッドを作成することはできません※19。一方、「株式雑談」「FX、為替雑談」のカテゴリでは、ユーザーがスレッドを作成することができます※20。

※19 上場廃止になった企業は、「上場廃止・償還済み」カテゴリに移動します。ただし、一定期間後予告なく削除する場合があります。
※20 ただし、スレッドの作成はパソコン版又はアプリ版のみ可能であり、スマートフォンweb版からは行えません。

２．禁止行為について

ファイナンス掲示板では、禁止行為を定め、専用のページを設けてユーザーに対し周知を行っています。

【掲示板】禁止行為、投稿に注意が必要な内容について

株式、為替、ＦＸなど金融商品に関する投稿が行われることから、ユーザーに対し、掲示板で得られた情報のみを信頼したり過度に信用して投資決定を行うのではなく、信頼できる機関を通じて事実確認を行うことを推奨しています。
また、「風説の流布」や「相場操縦」といった金融商品取引法で禁止されている行為について注意を記載するとともに、証券取引等監視委員会への情報提供ページへのリンクを掲載し、法令に違反していると判断される投稿について同委員会への報告を呼び掛けています。

Yahoo!ファイナンスの掲示板について

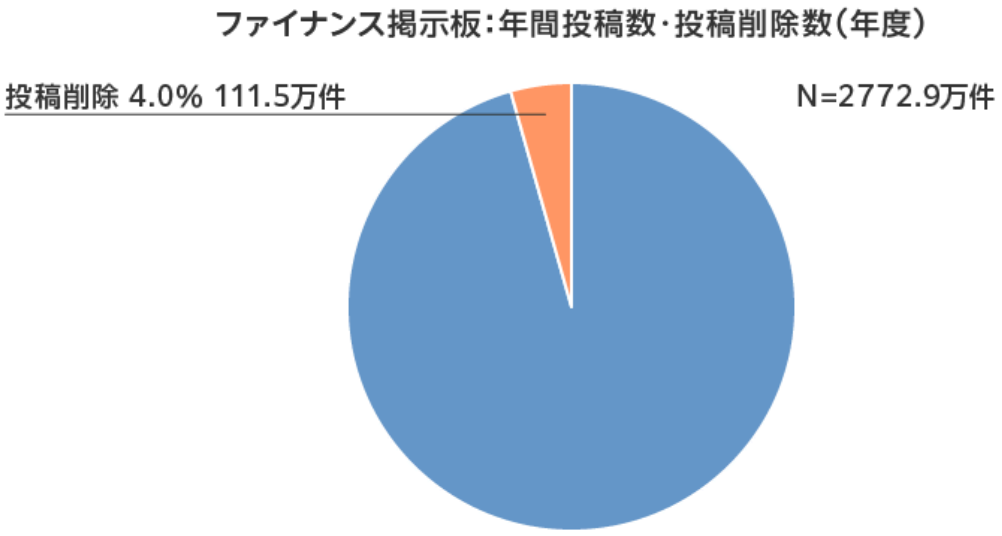
３．禁止行為違反への対応について

ファイナンス掲示板の禁止行為に該当すると判断された投稿については削除を行うとともに、一定の場合にはユーザーに対し投稿停止措置を行うことがあります。

21年度における投稿削除等の状況

１．投稿削除の状況

ファイナンス掲示板における21年度の投稿数は**2,772.9万件（月平均231.1万件）**で、そのうち投稿削除数は**111,5万件（月平均9.3万件）**でした。また、投稿件数のうち投稿削除件数が占める割合は**4.0%**となっています。



これを四半期別にみると次の表のとおりです。21年４－６月期においては、投稿削除数が低い水準となっています。ファイナンス掲示板では、後にみるように「広告・商売に関する情報」に該当することを理由とした投稿の削除が全体の投稿削除件数の過半を占めていますが、同時期においては、この項目に該当することを理由とする申告が他の時期と比較して約半数程度と少なく、このことが全体の投稿削除数にも影響したものと考えられます。

ファイナンス掲示板における投稿数･投稿削除数及び削除割合（四半期）

	投稿数	投稿削除数	削除割合
21年4－6月（月平均）	682.8万件（227.6万件）	17.8万件（5.9万件）	2.6%※21
7－9月（月平均）	700.7万件（233.6万件）	32.5万件（10.8万件）	4.6%
10－12月（月平均）	732.4万件（244.1万件）	31.0万件（10.3万件）	4.2%
22年1－3月（月平均）	657.0万件（219.0万件）	30.2万件（10.1万件）	4.6%
年度合計（月平均）	2772.9万件（231.1万件）	111.5万件（9.3万件）	4.0%

※21 四捨五入の関係で、左記と割合が一致しないことがあります

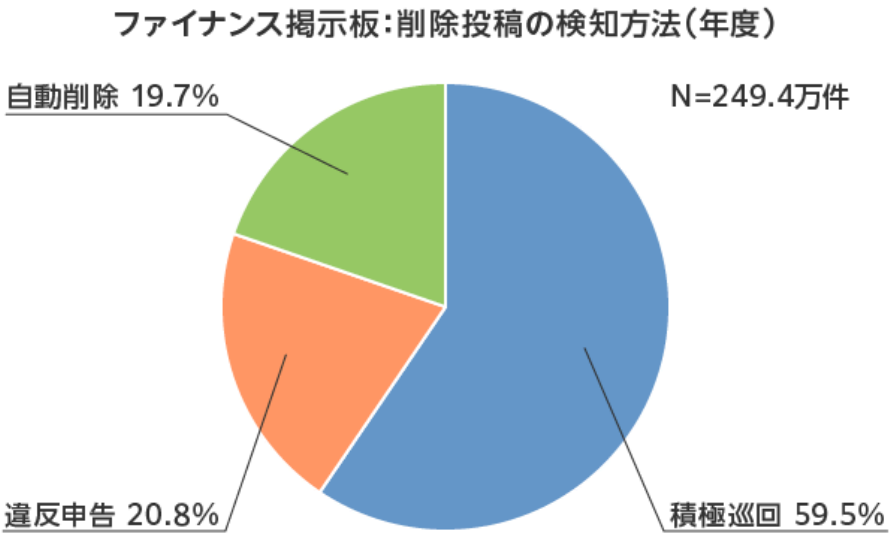
２．AIと人の目による削除の状況

（１）違反投稿の検知状況

膨大な数の投稿について的確に違反投稿であるかどうかの判定を行っていくためには、人の目だけでなく、機械の力を活用することが欠かせません。そこで、ファイナンス掲示板では、AIと専門チームによる「人の目」の双方を組み合わせて投稿削除の対応を行っています。

ファイナンス掲示板では、違反の可能性が明白である投稿については、AIにより自動削除されることがあります。また、ユーザーからの違反申告について、投稿ごとに違反申告フォームを設け、禁止事項に違反すると思われる投稿について、ユーザーからの情報を幅広く受け付けています。一方で、違反申告がない投稿についても、積極的な検知に努めており、専門チームのスタッフがコメント欄を定期的に巡回して目視確認しています。これによって検知された投稿については、１件１件「人の目」により慎重な審査を経た上で、違反投稿であるかどうかの判定が行われています。

ファイナンス掲示板では、21年度において削除された投稿のうち59.5%は専門チームによる積極的な巡回を契機として検知されたもので、20.8%は違反申告を契機として検知されたものでした。一方でAIにより自動削除された投稿は、投稿削除件数のうち19.6%を占めています。



（違反申告受付件数）

21年度におけるユーザーからの違反申告件数は**207.9万件（月平均17.3万件）**でした。１件の投稿に対して複数の違反申告が寄せられることもあり比較は困難ですが、これを年間の投稿件数に対する単純割合としてみると**7.5%**に当たります。

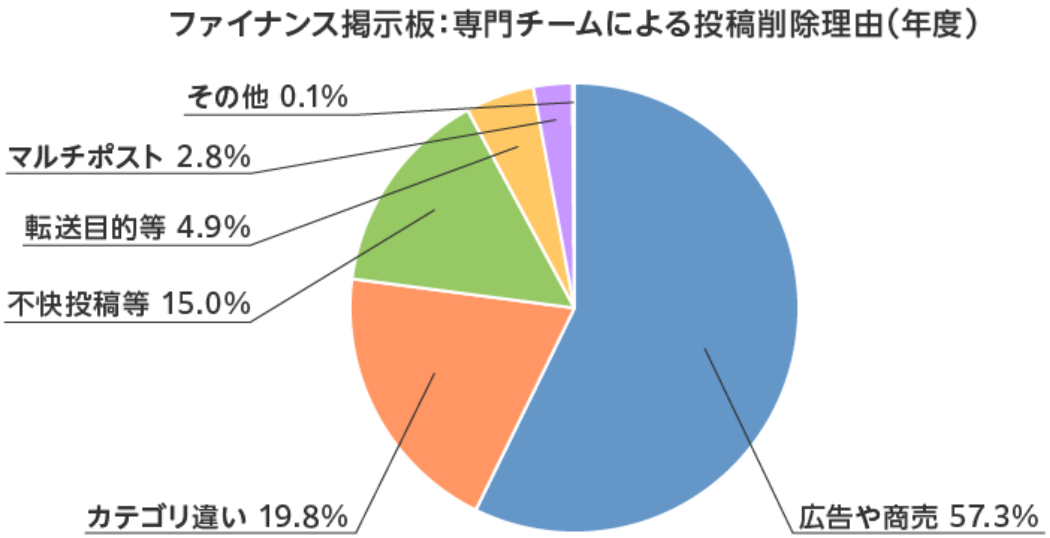
なお、このうち「広告・商売に関する情報」に該当することを違反理由とした申告はおよそ**３分の１程度**を占めています。

（２）削除理由

21年度において専門チームにより削除された投稿について、その削除理由を見ると、「広告や商売」に該当するものが**57.3%**と全体の過半を占めており、続いて割合が高い順に「カテゴリ違い」「不快投稿等※22」「転送目的等」「マルチポスト」「その他※23」となっています。

※22 誹謗や中傷、わいせつな情報、批判や論争に該当する投稿などが該当します。

※23 個人情報の投稿、投稿行為が法令違反を構成するものが該当します。



３．プロバイダ責任制限法に基づく削除の状況

特定電気通信役務提供者の損害賠償責任の制限及び発信者情報の開示に関する法律（以下「プロバイダ責任制限法」といいます。）は、プロバイダ等の損害賠償責任の制限、発信者情報の開示請求等や発信者情報開示命令事件に関する裁判手続について定めた法律です。
ヤフーは、同法上のプロバイダ（「特定電気通信役務提供者」）として、提供している各投稿型プラットフォームサービスについて、利用規約に違反する投稿などによって権利を侵害された方から、同法に基づく投稿削除請求や投稿者に関する情報の開示請求を受け付けています。
ファイナンス掲示板における21年度の請求の受付及び削除・開示等の実績について、裁判外での請求と裁判上の請求（訴訟）を分けて示すと次のとおりです。

（１）裁判外の請求

削除請求件数：26件 （内訳：名誉・信用・プライバシー23件、肖像権1件、その他2件）	
	うち削除（一部削除を含む）に至った件数：12件 （内訳：名誉・信用・プライバシー11件、その他1件）
開示請求件数：7件 （内訳：名誉・信用・プライバシー5件、肖像権1件、その他1件）	
	うち開示（一部開示を含む）に至った件数：0件

（２）裁判上の請求（訴訟）

削除請求件数：９件 （内訳：名誉・信用・プライバシー7件、人格権2件）	
	うち削除（一部削除を含む）に至った件数：6件 （内訳：名誉・信用・プライバシー4件、人格権2件）
開示請求件数：16件 （内訳：名誉・信用・プライバシー16件）	
	うち開示（一部開示を含む）に至った件数：11件 （内訳：名誉・信用・プライバシー11件）

４．アカウント停止措置・異議申し立ての状況

ファイナンス掲示板では、禁止行為に該当すると判断された投稿については削除を行うとともに、一定の場合にはユーザーに対し投稿停止措置を行うことがあります。21年度においてファイナンス掲示板においてアカウント停止の対象となったID数は**約7,900件（月平均660件）**で

した。

また、利用停止措置に対して疑問・意見等があるユーザーからは、［お問い合わせ］フォームにより問い合わせを受け付けています。21年度において、投稿停止措置を受けたユーザーから受けた問い合わせの件数は、**月あたり0～１件程度でした。**

【共通編】

21年度の新たな取組について

１．コメント欄非表示措置の導入（ニュース）

Yahoo!ニュースでは、21年10月19日より、一定以上の投稿数の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を導入しました。
導入（21年10月19日）から年度末（22年3月31日）までの期間にコメント欄が非表示となった記事数は**414件**で、1日あたりでは**2.5件**でした。集計期間中に媒体各社から「Yahoo!ニュース」へ配信された記事数は、1日あたり**7,450件**で、コメント欄が非表示になった記事の割合は全体の**約0.03%**でした。

同期間にコメント欄が非表示となった記事を媒体種別に検証したところ、件数が多い順にテレビが127件、一般紙・通信社が100件、週刊誌が61件、ネットメディアが52件、スポーツ紙・夕刊紙が46件、海外メディアが28件という結果でした。

非表示記事の件数や割合

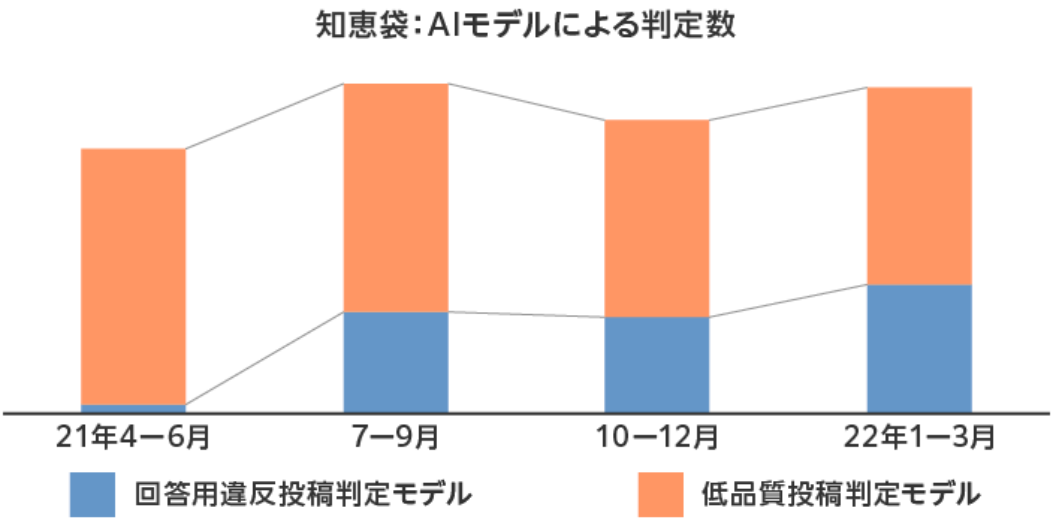
1日あたりの記事配信数(平均)	7,450件
1日あたりの非表示記事数(平均)	2.5件
コメント欄非表示記事数	414件
非表示になった記事の割合	約0.03%

非表示となった記事の媒体種別

テレビ	127件
一般紙･通信社	100件
週刊誌	61件
ネットメディア	52件
スポーツ紙･夕刊紙	46件
海外メディア	28件

２．新たな不適切投稿判定モデルの運用開始（知恵袋）

知恵袋では、機械判定に用いられるモデルとして、従前より利用されている低品質投稿判定モデルに加えて、21年6月より回答用違反投稿判定モデルを採用しています。
低品質投稿判定モデルは「不快投稿等」「文意不明」「わいせつな内容の投稿」の判定に特化しているのに対し、回答用違反投稿判定モデルは、知恵袋の回答欄におけるすべての禁止行為への違反を対象として判定を行っています。これにより、不快投稿等に関し低品質投稿判定モデルにおける判定を補完するほか、同モデルの判定対象とならなかった違反投稿についてもAIによる判定を行うことが可能となっています。
以下のとおり、21年7-9月以降、AIによる違反投稿の判定のうち約３～４割程度が回答用違反投稿判定モデルにより行われており、これに伴い２つのAIモデルによる判定数の合計も増加しています。



３．ユーザーへのメッセージ強化・違反申告の促進（ニュース）

Yahoo!ニュースでは、21年12月22日より、ユーザーによる違反コメントの報告を促進するために、投稿コメントごとに付されている違反申告アイコンについて「非表示・報告」というタイトルを明記し、違反報告の導線をわかりやすい形に変更しました。

また、違反コメントの投稿を事前抑止するために、AI判定モデルを用いて、違反と判定されたコメントを複数回投稿しているユーザーに対して注意メッセージの掲出を行っています。
事前抑止の効果を高めるため、21年10月にこの注意メッセージの表現を変更し、ユーザーがコメント欄を利用する際に必ず目に留まるよう、記事ページやコメント欄の上部など目立つ場所に注意書きを掲出しています。
さらに、同年12月にデザインとメッセージ内容をさらに強い表現へと変更しました。違反コメントを続けた場合に起こり得る法的リスクを明示するため、「発信者情報開示請求等を受けた場合、法令上の手続きにのっとり開示を行う場合があります」という文言を追加するなど、具体的かつわかりやすい形でユーザーへ警告しています。

また、通常のコメント投稿画面においても、Yahoo!ニュースとしてユーザーにどのようなコメントを投稿してもらいたいかを伝えるため、ユーザーの意見や質問、他のユーザーにも役立つ補足情報などを投稿していただきたい旨を明示することとしました。
あわせて、多くのユーザーがコメントポリシーを確認する機会を増やすため、投稿前にコメントポリシーを確認していただきたい旨を強調するよう、表示や文面を変更しました。

４．偽情報への対策（ニュース・知恵袋）

近年、米国や欧州においては、インターネット上での偽情報の流通が社会問題となっており、対策が進みつつあります。また、昨今の新型コロナウイルス感染症に関しても、事実に反する情報や根拠が不確かな情報の流通が大きな問題となりました。
偽情報に関しては、その分野・内容が多種多様であり判定や実態把握が困難であることや、これを信じることによる被害の抑止と表現の自由とのバランスなど各種の問題が存在しており、プラットフォーム事業者がいかにして過不足なく投稿削除等を行っていくべきか、関係各方面で様々な議論や検討が行われています。
ヤフーでは、ニュースコメント欄と知恵袋について、それぞれのサービスの性質に応じて21年度に以下のとおりポリシーの変更を行い、偽情報である旨が明らかである投稿を禁止しています。

知恵袋では、21年12月に利用のルールを変更し、「明らかな偽情報に関する投稿」として、明らかに事実と異なり、社会的に混乱を招く恐れのある投稿について削除の対象となる禁止事項として追加いたしました。
具体的には、「（そのような事実がないにもかかわらず）昨日、○○（地名）で大地震があった」などの投稿を対象としています。
22年3月を例にとると、当該禁止事項に違反して削除された投稿件数は、**約30件**でした。

また、Yahoo!ニュースにおいても、22年3月にコメントポリシーを変更し、「健康被害等をも

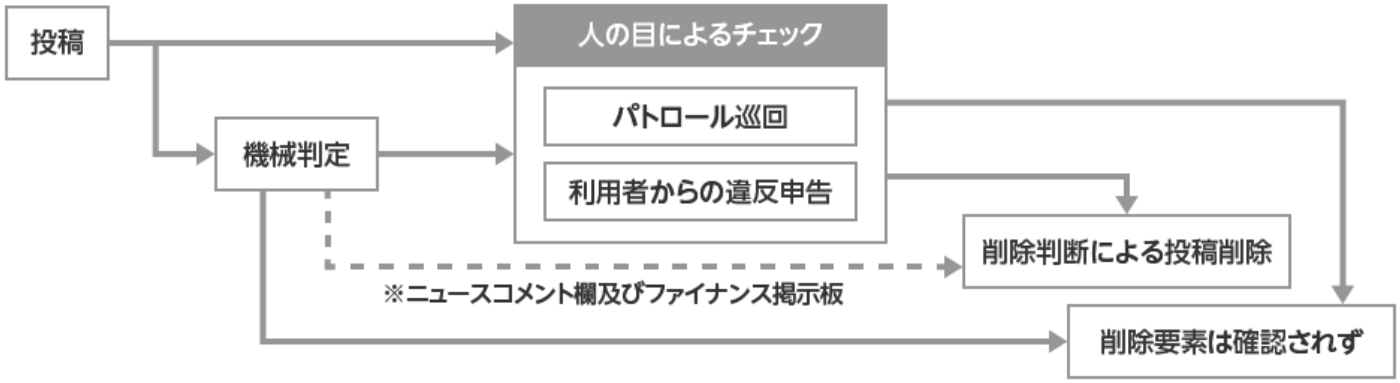
たらす可能性のある偽情報であって、ファクトチェックにより反真実であることが明らかになっている」投稿を、「明らかな偽情報」として投稿禁止対象に加えました。
具体的には新型コロナウイルスなど（新型コロナウイルスに限らない）のワクチンや治療薬などについての偽情報（ファクトチェック済みの情報に限る）を対象とするものであり、削除対象となる投稿内容は、関係当局のウェブサイトなどを参考として適宜見直しを行うこととしています。
導入月である22年3月に「明らかな偽情報」に該当するとして削除された投稿は、**約110件**でした。

なお、偽情報は、打ち消しの報道やファクトチェックがなされるまでヤフーとして真偽を確認することが困難であるため、ユーザーに対し、有害な偽情報と思われるコメントを見かけた場合、公的機関によるファクトチェック済みの情報を添えての違反申告を呼び掛けています。

ヤフーにおける違反投稿の投稿監視体制

１．AIと「人の目」の組み合わせによる違反投稿の監視

各サービスについてご紹介したとおり、ヤフーでは、AIと専門チームによる「人の目」を組み合わせて違反投稿の監視を行っています。
AIについては、当社が独自に開発したディープラーニングに特化したスパコン「kukai」等を活用した機械学習システムを導入し、迅速に禁止行為に抵触する投稿を検知できる仕組みを構築しています。具体的な活用方法はサービスによって異なりますが、機械学習システムにより投稿がコメントポリシーに抵触する内容である確率を判定し、コメントポリシーへの抵触の可能性がある場合は、自動削除を行ったり、専門チームによる確認の優先的なフローに移されることとなっています。
上記以外にも、専門チームによる「人の目」により、ユーザーからの違反投稿申告に基づく確認や自主的なパトロールも並行して行われています。



２．AIの仕組み

（１）Yahoo!ニュース

Yahoo!ニュースのコメント欄では、既にご紹介したとおり、関連度モデル、不適切投稿判定モデルの２種類のモデルを投稿の自動削除に利用しているほか、これに建設的モデルを加えた3 種類のモデルにより、「おすすめ順」における投稿の表示順序を決定しています。

（不適切投稿判定モデル・関連度モデル）

不適切投稿判定モデルは、過度な批判や誹謗中傷、差別、わいせつや暴力的などのガイドラインの項目に違反するおそれのあるコメントをAIが点数化するモデルです。同モデルで対象となるコメントは、「暴力的、差別的、過度に品位に欠けるコメント」等であり、同モデルは、算出したコメントの点数をもとに、コメントの掲載順位の変更や、自動的な削除などを行います。

関連度モデルは、ニュース記事とコメントの関連度をAIが判定するモデルです。関連度が低いと判定されたコメントは自動削除、あるいは専門チームによる目視による確認フローに移され、検知・判定の効率化に役立てられています。

これら２種類のモデルの生成方法については、専門チームが「人の目」により判定を行った結果を正解データとして学習し、その判断を再現します。学習の際には、ヤフーが持つ大規模テキストデータを言語理解力向上のためにあわせて活用しています。
性能とプロセスの検証を実施し、最新技術を取り入れて、継続的にモデルを改善しています。

（建設的モデル）

建設的モデルは、建設的であると考えられるコメントを上部の順位に表示させるためのモデルです。同モデルでは、「自分の意見を元に議論を喚起する発言」「客観的で、必要であれば根拠を提示している発言」といった評価軸でスコアを算出しています。

具体的には、一定数の一般ユーザーの集団に対し、「建設的コメントとなる条件」の定義を示し、「建設的コメント」を選択できなかったユーザーのみ対象から除外したうえで、記事とコメントのセット（1つの記事に対しコメントが複数個あり）を与え、建設的か否かの判断を行ってもらい、その正解データを利用して生成しています。
また、モデル生成後も、効果検証を実施し、導入前と比して、社内担当者の目から見ても適切な並び順となっているか、コメント欄の閲覧者が増えているのか等を参考に、定量的な指標をベースとして、効果の有無を測定しています。

（２）知恵袋

（低品質投稿判定モデル）

知恵袋において機械判定に用いられるモデルとしては、**低品質投稿判定モデル**があります。低品質投稿とは、アダルト関連や文意不明の（質問/回答になっていない）投稿、誹謗中傷を含む一般人が不快と感じる投稿を指し、専門チームにおいて過去に低品質投稿と判断した投稿を学習デ

ータとして利用し、機械判定モデルを生成しています。
同モデルでは、低品質投稿のスコアを算出し、スコアが一定の閾値以下の投稿については、自動的に回答の非公開措置（投稿をクリックしなければ全文を閲覧することができなくなる措置）を行うほか、専門チームによる優先的な確認への移送やランキング表示からの除外などの措置が行われています。

同モデルについては基準のアップデートを定期的に行っており、効果検証として、アップデートの前後における低品質投稿の対応件数を計測しています。当該データによれば、低品質判定がなされた投稿件数が増加しており、低品質判定能力の向上及びユーザーが目にする低品質投稿数の減少が確認できています。

（回答用違反投稿判定モデル）
「21年度の新たな取組について」「２．新たな不適切投稿判定モデルの運用開始（知恵袋）」において紹介させていただいたとおり、知恵袋では、21年6月から新たな機械判定モデルを採用しています。
具体的な内容については、該当箇所をご参照ください。

３．専門チームによる対応体制

（１）投稿監視体制

ヤフーでは、70人態勢※24の専門チームが24時間365日稼働で各サービスにおける違反投稿の検知・判定を行っています。
専門チームが検知・判定を行った結果は、AIモデルにおける正解データとして学習され、機械によってその判断が再現されることになります。

※24 投稿の削除等の監視業務を専門とする人員数

（２）教育体制

利用規約やサービスごとに定める禁止事項等の規定の他、内部の運用マニュアルを作成し、担当者ごとに判断にぶれが生じることのないように努めています。専門チームにおいては、配属時に1ヶ月以上の基礎研修（座学、モニタリング、OJT）を受けたスタッフが対応しています。
禁止事項等に抵触しているか否かについて疑義が生じた投稿については、専門チーム内での協議を経て複眼的な判断がなされるだけでなく、必要に応じて各サービスにおけるポリシーやルールの制定・改廃担当者や法的判断を行う専門部門へのエスカレーションを実施し、適切な判断がなされるように体制を構築しています。

インターネット上における言論空間の健全化を目指した取組

１．研究用データの提供（知恵袋）

ヤフーの開発部門では、情報科学、社会科学、学際領域など多岐にわたる分野において、大学、公的研究機関の研究者に広く利用していただくために一部ソフトウエアとデータを公開しています。
その一環として、知恵袋のデータベースからランダムサンプリングにより抽出した解決済みの質問（約247万件）と、それら各質問に対するすべての回答約（約649万件※25）について、国立情報学研究所（NII）を通じて研究者に対し提供を行っています。

（参考）
https://randd.yahoo.co.jp/jp/softwaredata
https://www.nii.ac.jp/dsc/idr/yahoo/chiebkr3/Y_chiebukuro.html

※25 件数はいずれも22年4月1日現在

２．建設的モデルのAPIの無償提供（ニュース）

Yahoo!ニュースでは、20年9月より建設的モデルの API を外部に無償提供しています。
同モデルを導入することにより、AI開発に必要となる大量の学習データの整備や計算コストなどの初期投資をかけずに、自社サービスにおけるコメントの健全化に向けた対策を実施することが可能となり、日本におけるインターネット上の言論空間の健全化に向けた業界全体の取組の向上につながることが期待されます。
現在、4社の導入実績があり、引き続き多くのお問い合わせをいただくなど導入企業は今後も増加していく見込みです。
（参考）
https://about.yahoo.co.jp/pr/release/2020/09/18a/
https://about.yahoo.co.jp/pr/release/2021/07/19a/

３．偽情報対策の取組

インターネットにおける偽情報の流通の対策について、ヤフーでは、一般社団法人セーファーインターネット協会主催の「Disinformation 対策フォーラム」へ参加し、有効な対策について多様なプレイヤーと議論を行っています。
(参考)
https://www.saferinternet.or.jp/anti-disinformation/

また、ファクトチェック・イニシアティブ（FIJ）の活動に賛同し、寄付の実施・法人会員として参加。偽情報・誤情報の検証結果を特設サイトからリンクするなどで連携しているほか、媒体社と連携し、ファクトチェック記事の配信を拡充する取り組みを行っています。

Yahoo!ニュースでは、有識者とフェイクニュース対策について議論を行い、対策コンテンツ制作の助言を受ける活動を実施し、その成果をユーザーに対する普及啓発に活用しています。
（参考）
「フェイクニュース」への備え～デマや不確かな情報に惑わされないために～
Yahoo!ニュース健診

その一環として、21 年 6 月よりフェイクニュース対策としてのリテラシー向上授業を実践しています。これまでに中学、高校、大学や NPO 団体イベントなどで出張事業等を開催しており、今後とも継続していく予定です。

４．関係省庁との連携

総務省が開催する有識者会議「プラットフォームサービスに関する研究会」では、大手プラットフォーム事業者各社に対し、各プラットフォームにおける誹謗中傷投稿の排除や偽情報の流通防止に向けた対策の自主的取組についてのモニタリングが行われています。
ヤフーは、21年、22年の２か年、計４回にわたって誹謗中傷・偽情報対策として行っている自主的取組について報告を行い、有識者の方々による評価を受けており、その内容は総務省が報告書として取りまとめて公表しています。
詳細については、下記をご覧ください。
（参考）
https://www.soumu.go.jp/main_content/000769270.pdf
https://www.soumu.go.jp/main_content/000831346.pdf