目次

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の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、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」及び「Yahoo!ファイナンス掲示板」を対象に、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。

【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

https://news.yahoo.co.jp/info/comment-policy

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照合し、同一の携帯電話番号を利用して取得された Yahoo! JAPAN ID からのコメントの投稿を制限しています。

さらに、ヤフーでは、ユーザーの安全性・利便性向上のため、パスワード認証からパスワードレス認証（SMS 認証、生体認証）への移行を推進しており、現在は Yahoo! JAPAN ID の新規取得時に携帯電話番号の設定が必須になっています。Yahoo!ニュース コメント欄においても、ユーザーの安全性・利便性向上に加え、不適切なコメント投稿の抑止を強化する施策の一環として 2022 年 11 月よりコメント投稿における携帯電話番号の設定の必須化を導入しました。この取組については、後掲の「22 年度の新たな取組について」で詳細を紹介します。

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。

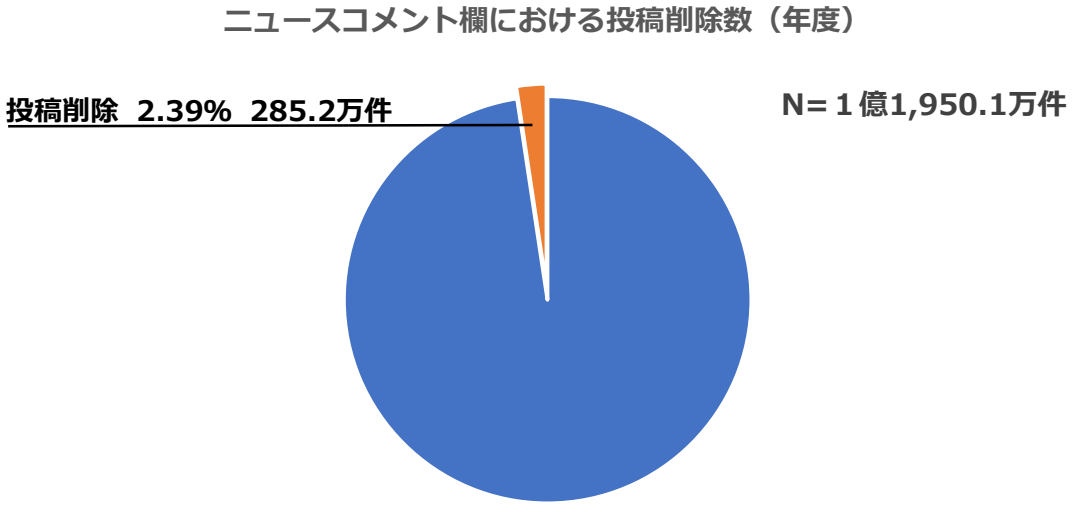
なお、非表示設定の内容は、設定したユーザー本人だけが確認できます。

22 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　コメント欄における 22 年度の投稿数は **1 億 1,950.1 万件（月平均 995.8 万件）**、投稿削除数は **285.2 万件（月平均約 23.8 万件）**でした。

投稿数のうち投稿削除数が占める割合についてみると、22 年度は **2.39%**でした。



投稿数・投稿削除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 年度は、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を背景として時間帯を問わずニュース全体の閲覧回数が増え、それに伴いコメントの投稿数及び削除数ともに 7-12 月期まで中長期的な増加傾向にありました。これに対し、22 年度は、パンデミックの落ち着きや大規模な国民的イベントが少なかったことから、投稿数及び投稿削除数が 21 年度に比べて減少しました。その中でも、安倍元首相銃撃事件とそれに関連する一連の報道等の影響を受け、投稿数及び投稿削除数ともに 7-9 月期に一時増加しています。

四半期ごとの数値をみると、22 年度下半期は投稿削除数及び削除割合が大きく減少しています。これには、従来からの AI やパトロールによる投稿削除等の取組に加えて、

・　「アカウント投稿停止措置」の対象拡大（21 年 10 月）や投稿時の携帯電話番号の設定必須化の導入（22 年 11 月）など、関連施策を実施してから、違反投稿の累積により投稿停止措置を受ける ID 数が約 4 割減少したこと

・　違反コメントを投稿したユーザーに対して掲出する「投稿時注意メッセージ」の施策の継続により、掲出ユーザー数が約３割減少したこと

等に見られるように、違反投稿を未然に防ぐ新たな取組の実施・強化の効果や、後述のAI のアップデート等の影響もあると考えられます（後述４、【共通編】「その他違反投稿を未然に防ぐ取組」）。

今後も、AI やパトロールによる違反投稿への対応などに加え、各種の違反投稿の未然防止の取組を講じるなど、投稿の事前・事後で複合的に施策を組み合わせながら、Yahoo!ニュースコメント欄を安全にご利用いただくための取組を推進してまいります。

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
22 年４－６月（月平均）	**3001.2 万件**（1000.4 万件）	**94.4 万件**（31.5 万件）	**3.2%**[1]
７－９月（月平均）	**3710.9 万件**（1237.0 万件）	**121.3 万件**（40.4 万件）	**3.3%**
10－12 月（月平均）	**2834.7 万件**（944.9 万件）	**37.5 万件**（12.5 万件）	**1.3%**
23 年１－３月（月平均）	**2403.4 万件**（801.1 万件）	**32.0 万件**（10.7 万件）	**1.3%**
年度合計（月平均）	**1 億 1950.1 万件**（995.8 万件）	**285.2 万件**（23.8 万件）	**2.4%**
参考 21 年度合計 （月平均）	1 億 5923.2 万件 （1326.9 万件）	658.4 万件[2] （54.9 万件）	4.1%

[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] 2021 年度版から数値の訂正を行っています。詳しくは、2021 年度版の該当箇所をご確認ください。https://about.yahoo.co.jp/common/transparencyreport_2021/

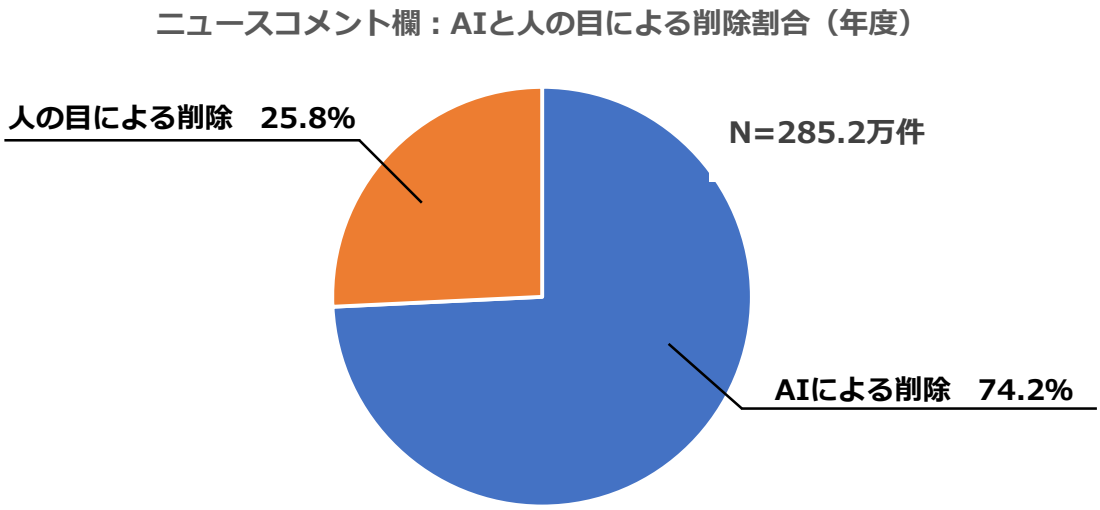
２．AI と人の目による削除の状況

（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。

22 年度における投稿削除件数のうち **74.2%**は、AI 等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 等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。22 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 **25.8%**でした。



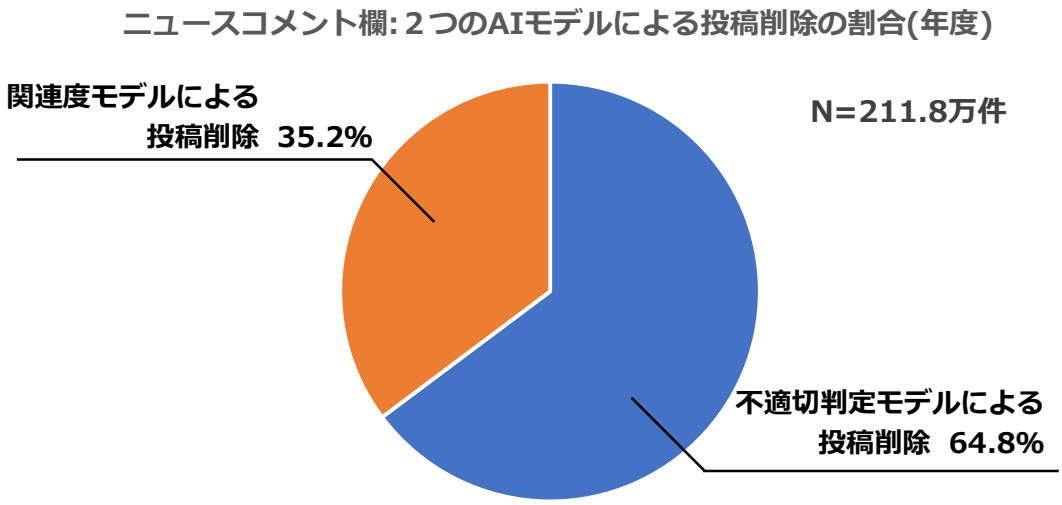
（２） AI による自動削除の状況

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判定を行っています[3]。

- **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、コメントと当該コメントに対する返信の関連度、それぞれを AI が判定

22 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 **64.8%**、関連度モデルによるもの[4]が **35.2%**でした。



（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、ユーザー等からの違反報告を契機とするものと、専門チームによる積極的な巡回・パトロールの優先順位付けのための AI 判定を補助とするものとに大別されます。

ユーザーからの違反報告については、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

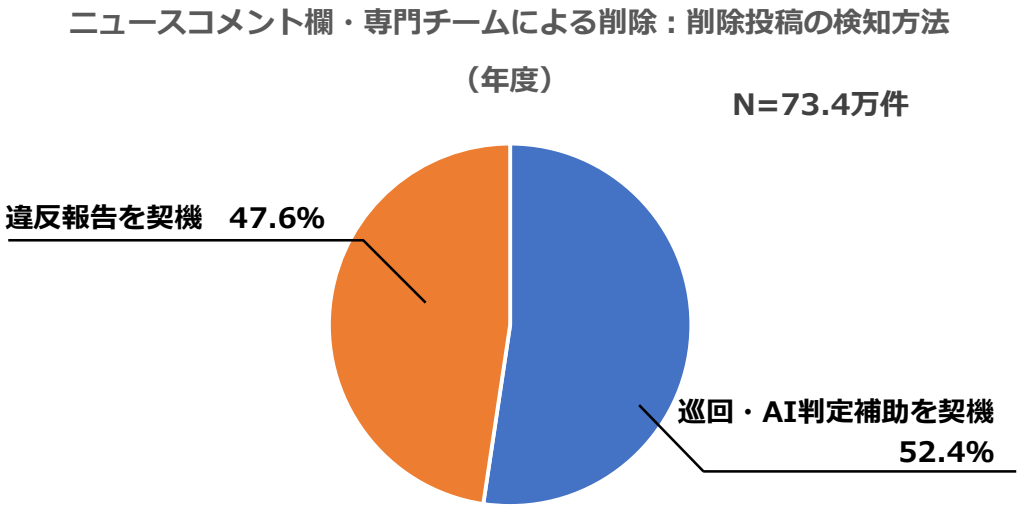
[3] 各モデルの具体的な仕組みについては、「ヤフーにおける投稿監視体制」の項目で詳しく紹介しています。
[4] 関連度モデルによる自動削除には、集計の仕組み上、特定の単語の使用を起因とする自動削除も含まれています。

け付けています。「関連度モデル」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た投稿についても、AI から違反報告がなされ、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、違反報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は、自動削除とならなかった全投稿を対象に、パトロールの優先順位付けのため AI の判定を補助として、コメント欄を積極的に巡回して目視確認しています。

22 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、違反報告を契機として削除されたものの割合は **47.6%**であり、**52.4%**と過半数が専門チームによる積極的な巡回・パトロールの優先順位付けのための AI 判定を補助として削除されたものでした。

なお、コメントパトロールで利用している AI は、精度の維持・向上のために不可欠なモデルのアップデート等を適宜行っており、その準備、進捗状況等によって判定数や削除率等の実績数に影響を与えることがあります。



（違反報告受付件数）

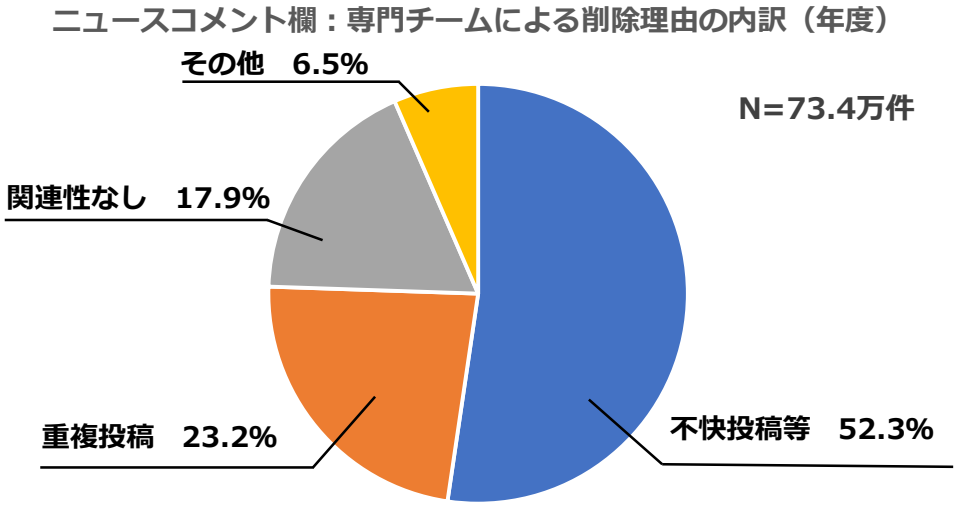
22 年度に受け付けたユーザーからの違反報告の件数は **320.2 万件（月平均 26.7 万件）**でした。1 件の投稿に対して複数の違反報告が寄せられることもあり単純比較は困難ですが、これを年間の投稿件数に対する割合としてみると **2.7%**に当たります。

Yahoo!ニュースでは、ユーザーからの違反報告を促進することによって、より迅速に違反投稿を検知することが可能となると考えています。こうした考えの下、21 年 12 月に違反報告の導線をユーザーによりわかりやすい形に改善したこと等の影響もあり、22 年度の違反報告受付件数は 21 年度の約 1.4 倍に増加しており、これらを違反投稿の検知に役立てています。

（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、「不快投稿等」と、それ以外の投稿の割合を見ていきます。なお、「不快投稿等」とは、ニュースコメント欄においては、コメントポリシーで禁止しているコメントのうち、誹謗中傷、ヘイトスピーチ、アダルト関係、不謹慎な投稿などユーザーが不快に感じると思われる投稿を指します。

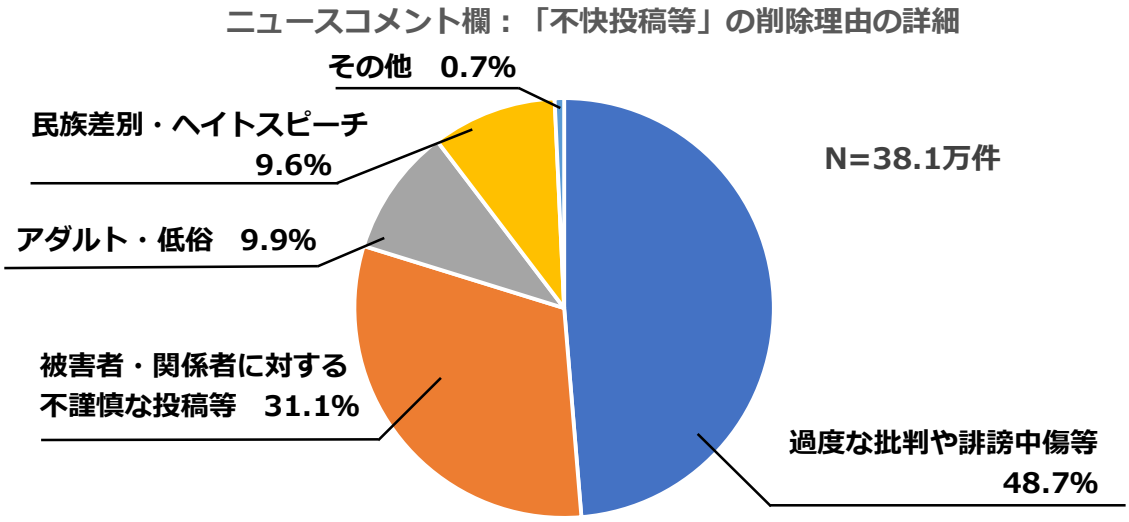
22 年度において「不快投稿等」として専門チームにより削除された投稿の割合は**52.3%**を占めており、続いて割合が高い順に「重複投稿」「関連性なし」「その他[5]」となっています。なお、21 年度までは複数の削除事由に該当する場合に「総合判断」と分類していましたが、2022 年 3 月に、禁止事項を運用実態に沿って細分化し、具体的な投稿例を追加することで、ユーザーにとってわかりやすくなるようコメントポリシーの改定を行ったため、内訳の「総合判断」の分類がなくなっています。



[5] 「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、**48.7%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[6]」「アダルト・低俗」「民族差別・ヘイトスピーチ」を理由としたものとなっています。



3．投稿削除請求及びプロバイダ責任制限法に基づく開示請求の状況

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、投稿の送信防止措置の依頼（以下「削除請求」といいます。）や同法に基づく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 22 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。

---

[6] 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）

（１） 裁判外の請求

削除請求件数：1 件（21 年度：3 件） （内訳：名誉・信用・プライバシー1 件）	
	うち削除（一部削除を含む）に至った件数：0 件（同：1 件）
開示請求件数：4 件（同：6 件） （内訳：名誉・信用・プライバシー3 件、著作権 1 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 22 年度は削除・開示請求件数ともに減少しました。
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）[7]

削除請求件数：0 件（21 年度： 2件）	
	うち削除（一部削除を含む）に至った件数：0 件（同：2 件）
開示請求件数：8 件（同：10 件） （内訳：名誉・信用・プライバシー 8件）	
	うち開示（一部開示を含む）に至った件数： 7件（同：8 件） （内訳：名誉・信用・プライバシー 7件）

22 年度の削除請求件数は0件で、開示請求件数はやや減少しました。なお、開示請求は全件が名誉・信用・プライバシーの侵害を理由とするものでした。

４．アカウントの投稿停止措置・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照合し、同一の携帯電話番号を利用して取得された Yahoo! JAPAN ID からのコメントの投稿を制限しています。

---

[7] 2023 年 8 月時点。なお、削除／開示に至った件数には、任意に削除／開示を行った件数も含みます。知恵袋、ファイナンス掲示板も同様。

22 年度中に投稿停止措置を受けた ID の数[8]は **8,774 件（月平均 731 件）**と 21 年度の月平均から約４割減少しました。

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 257 件**と 21 年度よりも約３割弱減少しました。

[8] 23 年 2 月のプレスリリース（Yahoo!ニュース、コメント投稿における携帯電話番号設定の必須化後の効果を公表 https://about.yahoo.co.jp/pr/release/2023/02/27a/）の「投稿停止措置を受ける ID 数」は各日の集計時点において投稿停止の対象となる ID の数（措置前の ID を含む）を指しますが、本レポートで紹介した「投稿停止措置を受けた ID の数」は実際に投稿停止を受けた ID の数を指します。

【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）
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たと判断したされた場合には、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。

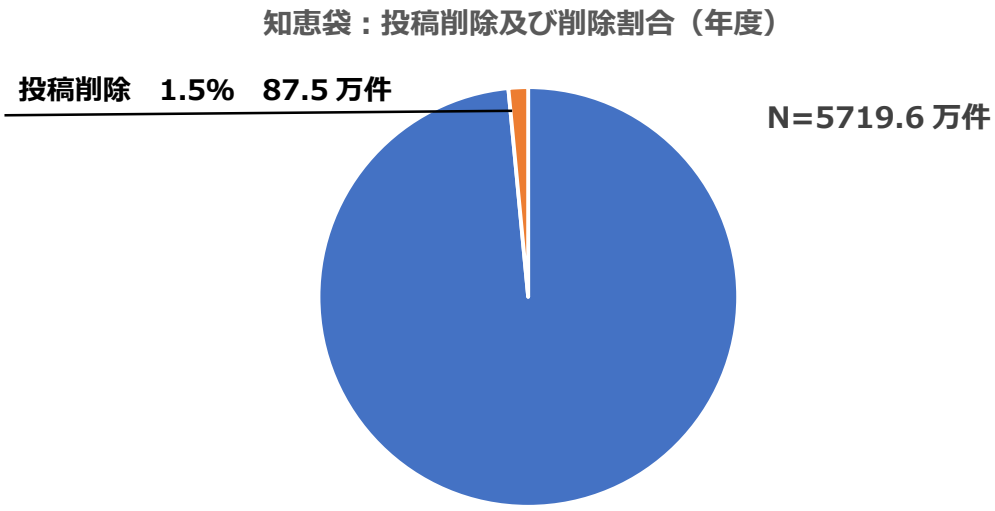
このほか、禁止事項に該当する蓋然性が高いと判断された投稿は、不適切な投稿として質問詳細ページにおいて非表示となることがあります。非表示状態となっている投稿は、ユーザーが「表示する」ボタンを押下すると、その内容を確認することができます。

22 年度における投稿削除等の状況

１．投稿削除の状況

22 年度の投稿件数は **5719.6 万件（月平均 476.6 万件）**であり、21 年度の投稿件数 5696.6 万件（月平均 474.7 万件）からやや増加しました。これに対し、投稿削除件数は **87.5 万件（月平均 7.3 万件）**と、21 年度の投稿削除件数 249.4 万件（月平均 20.8 万件）と比較して**約３分の１**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みると、22 年度は **1.5%**であり、こちらも 21 年度の 4.4%と比較して**約３分の１**となっています。



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[9]	削除割合
4－6月（月平均）	**1393.8 万件**（464.6 万件）	**24.8 万件**（8.3 万件）	**1.8%**[10]
7－9月（月平均）	**1490.2 万件**（496.7 万件）	**23.5 万件**（7.8 万件）	**1.6%**
10－12 月（月平均）	**1380.0 万件**（460.0 万件）	**19.0 万件**（6.3 万件）	**1.4%**
1－3月（月平均）	**1455.7 万件**（485.2 万件）	**20.2 万件**（6.7 万件）	**1.4%**
年度合計（月平均）	**5719.6 万件**（476.6 万件）	**87.5 万件**（7.3 万件）	**1.5%**
参考 21 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%

前述のとおり 22 年度は 21 年度と比較して、投稿削除数・削除割合ともに約３分の１となっています。

知恵袋では、18 年に利用のルールの大幅改定を行い、それ以降、改定後の利用のルールに照らし、過去のすべての投稿を対象に、AI 及び人の目を活用したパトロールを実施してきました。その結果、21 年度中に、利用のルールに違反する過去投稿のパトロールが一巡したこともあり、22 年度の投稿削除数に占める過去投稿の削除分は大きく減少し、同年度中に投稿された新規投稿の削除分が削除対応の中心となっています。

２．ＡＩと人の目による削除の状況

（１）違反投稿の検知状況

---

[9] 知恵袋では、投稿削除数は対象年度に削除された投稿の数（対象年度以前に投稿された過去投稿を対象年度中に削除した場合を含む）を示しています。

[10] 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

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。

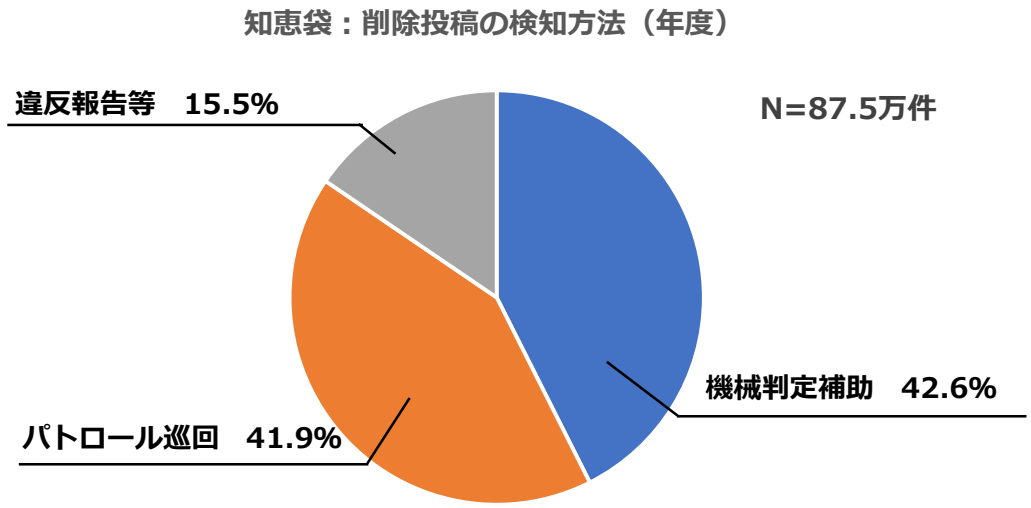
知恵袋で用いられている AI モデルは、専ら違反投稿の検知（機械判定補助）を目的としたもので、AI により違反投稿である蓋然性が高いと判断された投稿が「人の目」による審査フローに移される仕組みとなっています[11]。

また、ユーザーからの違反報告については、投稿ごとに違反報告フォームを設け、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、違反報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、22 年度において削除された投稿のうち **42.6%**は機械判定補助を受けて削除に至ったもので、**41.9%**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、違反報告を契機として削除された投稿の割合は**15.5%**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。

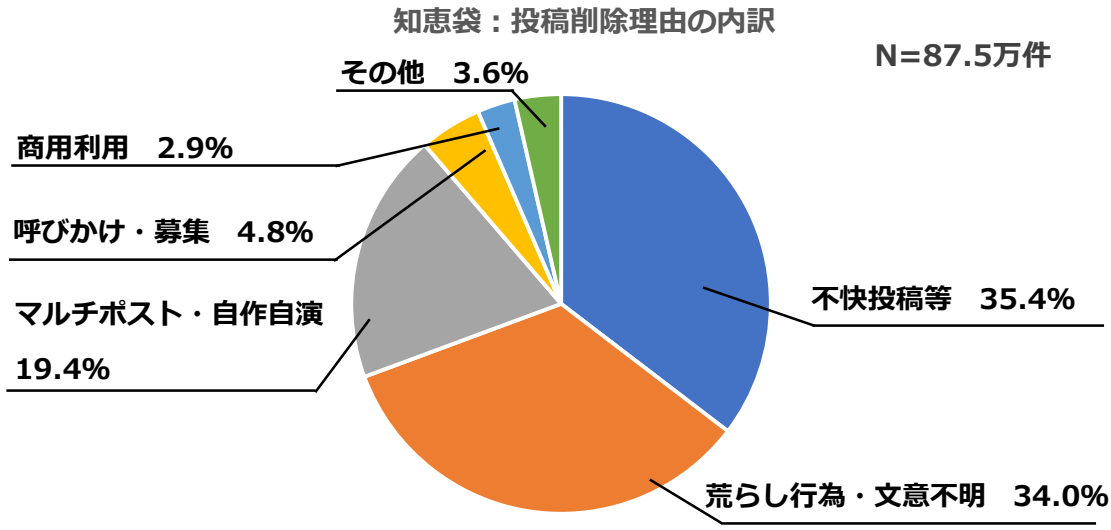
（違反報告受付件数）

22 年度に受け付けたユーザーからの違反報告の件数は、**117.3 万件（月平均 9.8 万件）**と 21 年度の 212.9 万件（月平均 17.7 万件）からおよそ半分に減少しました。1 件の投稿に対して複数の違反報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、**2.1%**に当たり、こちらも 21 年度の 3.7%から低下しています。

（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、「不快投稿等」と、それ以外の投稿の割合を見ていきます。なお、「不快投稿等」とは、知恵袋においては、利用のルールにおける禁止事項のうち、誹謗中傷、ヘイトスピーチ、アダルト関係、不謹慎な投稿などユーザーが不快に感じると思われる投稿を指します。

22 年度において「不快投稿等」として削除された投稿の割合は **35.4%**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」「その他[12]」「商用利用」となっています[13]。



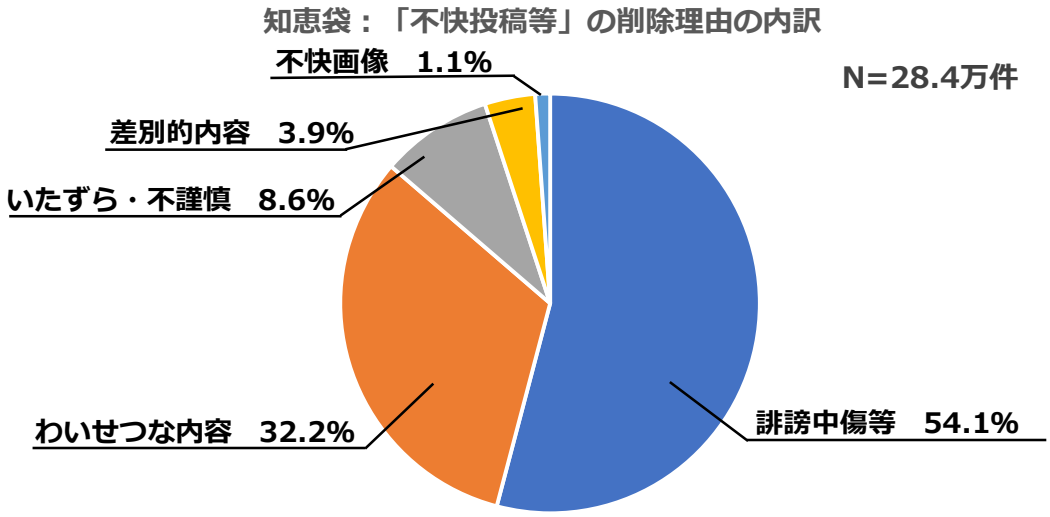
（不快投稿等の内訳）

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 **54.1%**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

[12] 「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。

[13] 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。

ずら・不謹慎」「差別的内容」「不快画像[14]」となっています。



（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。

22 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、**191.8 万件（月平均 16.0 万件）**であり、21 年度に同様に非表示措置が行われた投稿数 717.8 万件（月平均 59.8 万件）と比較して、約４分の１となっています。また、投稿件数に対する比率としては **3.4%**を占めていますが、こちらも 21 年度の 12.6%と比較して約４分の１となっています[15]。

３．投稿削除請求及びプロバイダ責任制限法に基づく開示請求の状況

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

[14] 「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。

[15] 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、現在残っている投稿のうち 3.4%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。

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、投稿の送信防止措置の依頼（以下「削除請求」といいます。）や同法に基づく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 22 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[16]からの請求

削除請求件数：8 件（21 年度：10 件） （内訳[17]：名誉・信用・プライバシー8 件）	
	うち削除（一部削除を含む）に至った件数：2 件（同： 5 件） （内訳：名誉・信用・プライバシー2 件）

② 行政機関以外からの請求

削除請求件数：80 件（21 年度：99 件） （内訳：名誉・信用・プライバシー75 件、著作権 4 件、その他 1 件）	
	うち削除（一部削除を含む）に至った件数：30 件（同：36 件） （内訳：名誉・信用・プライバシー26 件、著作権 4 件）
開示請求件数：8 件（同：15 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：0 件（同：2 件）

前年度と比較すると、行政機関からの（削除）請求件数、行政機関以外からの削除・開示請求件数ともに減少しています。

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。

---

16 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。
17 内訳には重複が含まれることがあります（以下同じ）

（２） 裁判上の請求（訴訟）

削除請求件数：6 件（21 年度：11 件） （内訳：名誉・信用・プライバシー6 件）	
	うち削除（一部削除を含む）に至った件数：3 件（同：6 件） （内訳：名誉・信用・プライバシー3 件）
開示請求件数：9 件（同：7 件） （内訳：名誉・信用・プライバシー9 件）	
	うち開示（一部開示を含む）に至った件数：4 件（同：4 件）・係属中 2 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減少した一方、開示請求件数はやや増加しています。

いずれの請求についても、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。

22 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 **2,065 件（月平均 172 件）と、**21 年度の 4,520 件（月平均 380 件）から半減しました。前述のとおり、利用のルール違反にあたる投稿削除件数が大きく減少していることに伴い、利用停止となる ID 数も減少しているものと考えられます。

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。22 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 27 件前後でした。**

【ファイナンス掲示板編】

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[18]。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[19]。

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。23 年４月には、マルチポストや個人情報の投稿等について、個人の顔写真の投稿など禁止される投稿の例を挙げ、より明確に禁止事項をお伝えできるよう規則の改定を行うなど、ユーザーが安心して利用できる環境の整備に努めています。

【掲示板】禁止行為、投稿に注意が必要な内容について
https://support.yahoo-net.jp/PccFinance/s/article/H000011273

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。

---

[18] 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。
[19] 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。

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
https://support.yahoo-net.jp/PccFinance/s/article/H000011280

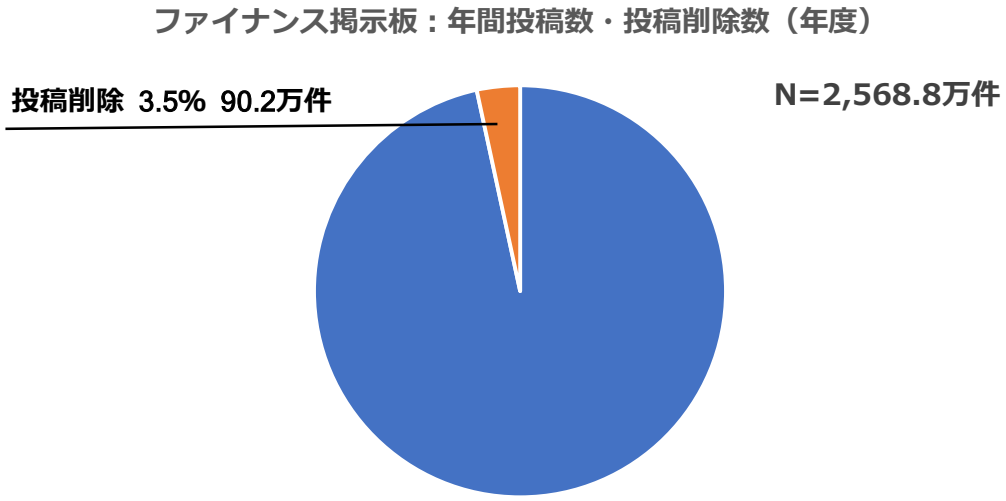
３．禁止行為に対する違反への対応について

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。

22 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 22 年度の投稿件数は **2,568.8 万件（月平均 214.1 万件）**で、そのうち投稿削除件数は **90.2 万件（月平均 7.5 万件）**でした。21 年度の投稿件数 2,772.9 万件（月平均 231.1 万件）及び投稿削除件数 111.5 万件（月平均 9.3 万件）と比較すると、いずれもやや減少しました。また、投稿件数のうち投稿削除件数が占める割合は **3.5%**となっており、21 年度の 4.0%からやや低下しました。



これを四半期別にみると次の表のとおりです。21 年度と比較すると、投稿数、投稿削除数及び削除割合いずれもやや低い数値となっています。

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
22 年 4 – 6 月（月平均）	**627.4 万件**（209.1 万件）	**25.2 万件**（8.4 万件）	**4.0%**[20]
7 – 9 月（月平均）	**680.8 万件**（226.9 万件）	**23.5 万件**（7.8 万件）	**3.5%**
10 – 12 月（月平均）	**635.8 万件**（211.9 万件）	**20.0 万件**（6.7 万件）	**3.1%**
23 年 1 – 3 月（月平均）	**624.8 万件**（208.3 万件）	**21.5 万件**（7.2 万件）	**3.4%**
年度合計（月平均）	**2568.8 万件**（214.1 万件）	**90.2 万件**（7.5 万件）	**3.5%**
参考 21 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%

2．AI と人の目による削除の状況

（1）違反投稿の検知状況

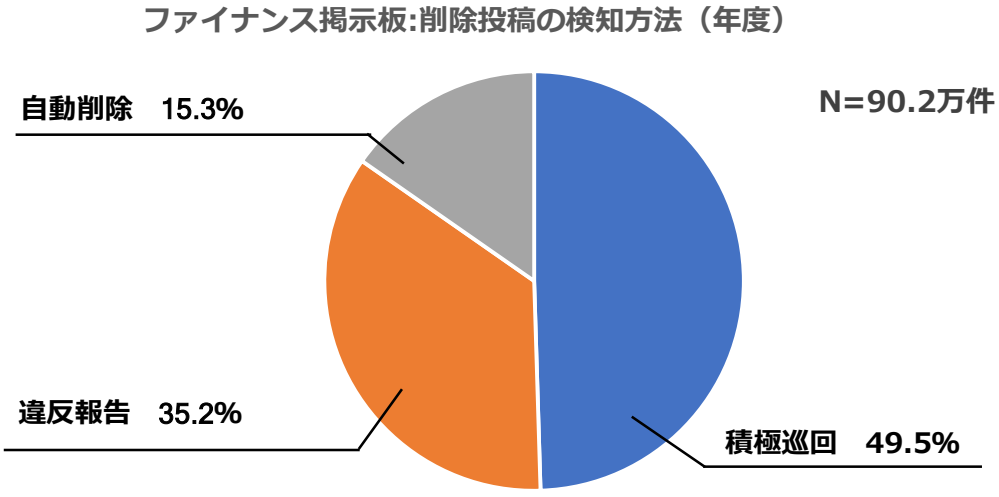
膨大な数の投稿について的確に違反投稿であるかどうかの判定を行っていくためには、「人の目」だけでなく、機械の力を活用することが欠かせません。そこで、ファイナンス掲示板では、AI と専門チームによる「人の目」の双方を組み合わせて投稿削除の対応を行っています。

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報告について、投稿ごとに違反報告フォームを設け、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、違反報告がない投稿についても、積極

[20] 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

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、22 年度において削除された投稿のうち **49.5%**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、**35.2%**は違反報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 **15.3%**を占めています。



（違反報告受付件数）

22 年度におけるユーザーからの違反報告件数は **309.5 万件（月平均 25.8 万件）**でした。１件の投稿に対して複数の違反報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 **12.0%**に当たります。

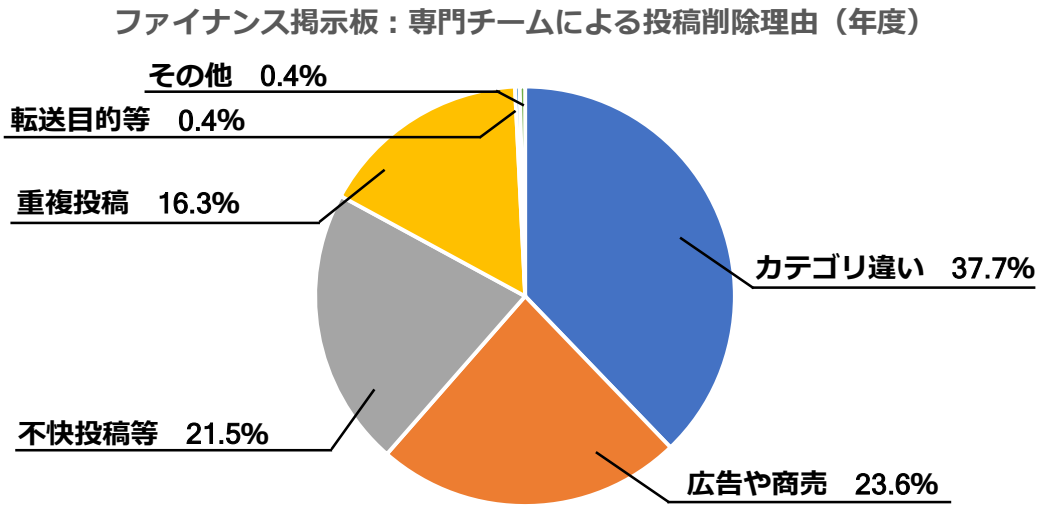
なお、このうち「不快投稿等[21]」に該当することを違反理由とした申告はおよそ **3 分の１程度**を占めています。

（２） 削除理由

22 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、「カテゴリ違い、スレッド違いの投稿」に該当するものが **37.7%**と最も多く、続

[21] 誹謗、中傷、わいせつな情報などが該当します。

いて割合が高い順に「広告や商売に関する情報」「不快投稿等」「重複投稿、マルチポスト」「転送を目的とした URL の投稿等」「その他[22]」となっています。



３．投稿削除請求及びプロバイダ責任制限法に基づく開示請求の状況

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、投稿の送信防止措置の依頼（以下「削除請求」といいます。）や同法に基づく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 22 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。

22　個人情報の投稿、投稿行為が法令違反を構成するもの等が該当します。

（１） 裁判外の請求

削除請求件数：67 件（21 年度：26 件） （内訳：名誉・信用・プライバシー65 件、著作権 2 件）	
	うち削除（一部削除を含む）に至った件数：50 件（同：12 件） （内訳：名誉・信用・プライバシー48 件、著作権 2 件）
開示請求件数：26 件（同：7 件） （内訳：名誉・信用・プライバシー26 件）	
	うち開示（一部開示を含む）に至った件数：1 件（同：0 件）

前年度と比較すると、裁判外の削除請求件数、開示請求件数ともに増加しています。
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：2 件（21 年度：9 件） （内訳：名誉・信用・プライバシー2 件）	
	うち削除（一部削除を含む）に至った件数： 2件（同：6 件） （内訳：名誉・信用・プライバシー 2件）
開示請求件数：14 件（同：16 件） （内訳：名誉・信用・プライバシー14 件）	
	うち開示（一部開示を含む）に至った件数： 5件（同：11 件）・係属中４件 （内訳：名誉・信用・プライバシー 5件）

前年度と比較すると、削除請求件数、開示請求件数ともに減少しています。
いずれの請求についても、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

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。22 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

4,034 件（月平均 336 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 0～2 件程度でした。**

【共通編】

22 年度の新たな取組について

１．コメント欄への投稿における携帯電話番号設定の必須化（ニュース）

前述のとおり、ヤフーでは、ユーザーの安全性・利便性向上のため、パスワード認証からパスワードレス認証（SMS 認証、生体認証）への移行を推進しており、現在は Yahoo! JAPAN ID の新規取得時に携帯電話番号の設定が必須となっています。その結果、ID のアクティブユーザーの約 7 割が SMS 認証をはじめとするパスワードレス認証に移行しています[23]。

一方で、コメントの「投稿停止措置」を受けたユーザーにおいては、携帯電話番号の設定がない ID の割合が 5 割以上と高水準になっており、その背景として、「投稿停止措置」を受けたユーザーが別の ID を利用して不適切な利用を繰り返すケースが一因となっていました。このため、Yahoo!ニュース コメント欄においても、ユーザーの安全性・利便性向上に加え、不適切なコメント投稿の抑止を強化する施策の一環として、携帯電話番号をもとにした「投稿停止措置」を確実に実施できるよう、2022 年 11 月より、コメント欄への投稿にあたり携帯電話番号の設定を必須化しました。

２．違反投稿判定モデルの判定精度向上のための取組（知恵袋）

AI を活用して効率的にパトロールを行い、ユーザーが安心して利用できる投稿型プラットフォームを維持するためには、データ学習等による AI の精度の維持・向上、モデルのアップデート等の対応が不可欠です。知恵袋では、機械判定に用いられるモデルの１つとして、知恵袋のすべての禁止行為への違反を対象として判定を行う違反投稿判定モデルを採用し、学習データを読み込ませることで判定精度の向上に努めています。

そこで、違反投稿判定モデルの精度検証のため、22 年度の全投稿の約 0.3%にあたる投稿を抽出したテストを行って 22 年度期初と期末の各モデルを比較し、同じ投稿の中から違反投稿と判定すべき投稿数のうち何割を正しく判定できたかの検証を行いま

---

[23] 2022 年 5 月時点

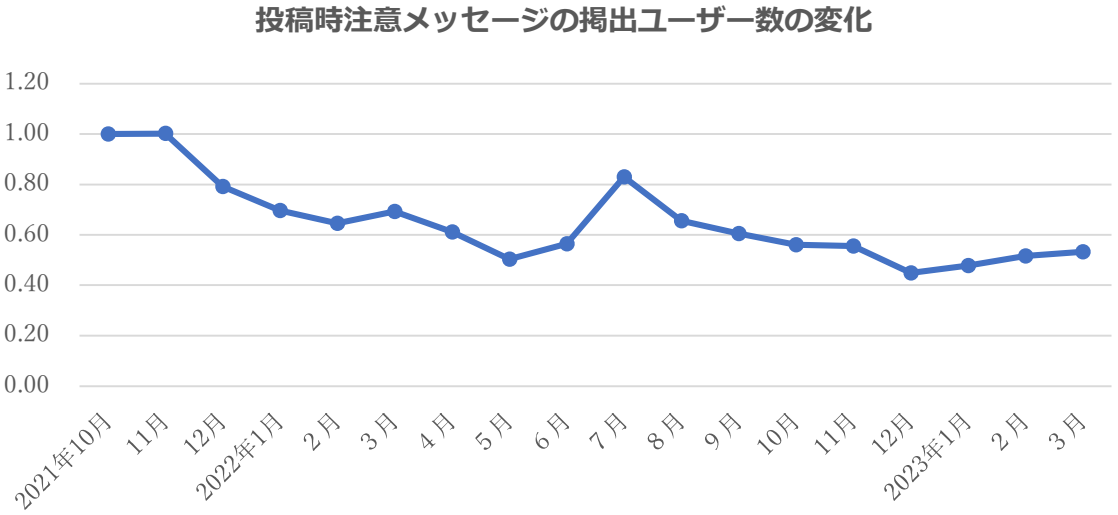
した。その結果、違反投稿を正しく判定した割合は、質問が 5 pt、回答が 6pt 上昇しており、判定精度の向上を示しています。

その他違反投稿を未然に防ぐ取組

投稿時注意メッセージの掲出（ニュース）

Yahoo!ニュースでは、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喚起し、投稿するコメント内容の再考を促すためのメッセージを掲出しています。この注意メッセージ掲出の取組自体は 2020 年 7 月から行っていますが、21 年 10 月に「コメントの投稿ができなくなる可能性がある」と警告する内容へ表現を強めています。なお、23 年 6 月には、「誹謗中傷等に関しては、法改正により投稿者の情報開示について簡易な裁判手続きが導入されています。」という内容を追記するなど、随時アップデートを行いながら取り組みを継続しています。

この取組の効果として、メッセージが掲出されたユーザー数について、掲出条件が同じ 21 年度下半期と 22 年度の月平均を比較すると、約 3 割減少しています。21 年 10 月の数値を 1 とすると、23 年に入ってからは概ね 0.5 前後の水準で推移しています。



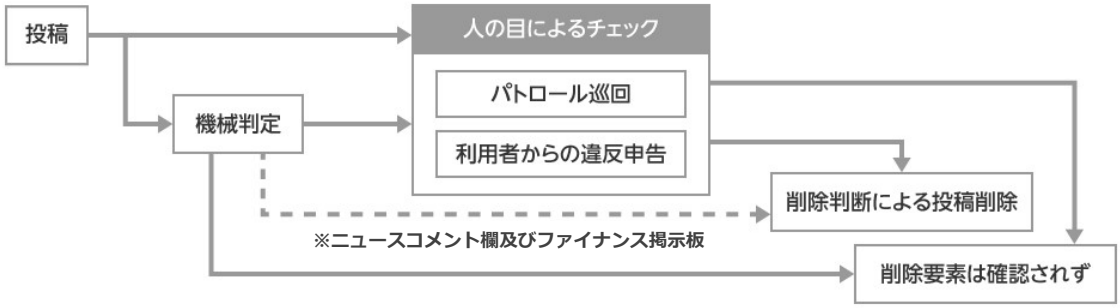
ヤフーにおける違反投稿の投稿監視体制

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種類のモデルをコメントポリシー違反による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、コメント多様化モデルを加えた４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コメントポリシーの各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。「関連性のない投稿」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。

（コメント多様化モデル）
コメント多様化モデルは、建設的かつ多くの人が共感している意見だけでなく、異なる視点の意見など、より多様なコメントを上位に表示させるためのモデルで、23 年 4 月より導入したものです。同モデルは、まず同一の記事に対して投稿された各コメントの意味・内容を把握したうえで、それぞれの内容の類似度によってグループ分けし、各グループから抽出した代表的なコメントをコメント欄の上位に表示します。
同モデルの導入により、ユーザーが多様な意見を知るきっかけの提供を目指しています。さらに、似た内容のコメントが並んだ結果、そのコメントの意見が多数派で正しい

意見であるかのような印象をもたらし、他のユーザーからも同様の投稿やさらに過激化した投稿が過度に集中してしまう、いわゆるエコーチェンバー現象の軽減効果も期待しています。

なお、本 AI モデルを含むコメントの並び順は、ユーザーからの評価などを踏まえて、引き続きさらなる改善に取り組んでいきます。

（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、利用のルールにおいて禁止事項として定められているわいせつ等の投稿や質問/回答になっていない・文意をなさない投稿、過度な批判や誹謗中傷・他人を不快にさせる内容の投稿を指し、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。

（違反投稿判定モデル）

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利用のルールのうち「誹謗中傷等」「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、違反投稿判定モデルは、知恵袋の投稿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、低品質投稿判定モデルにおける判定を補完して専門チームによる優先的な確認への移送等を行っており、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。

違反投稿判定モデルの精度向上に向けた取組については、該当箇所（【共通編】 2）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。

（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 217 万件）と、それら各質問に対するすべての回答約（約 559 万件

[24] 投稿の削除等の監視業務を専門とする人員数

[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

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、6 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。

なお、本 AI 技術については、複数の特許を取得済みです。

（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として、ヤフーでは、ニュースコメント欄と知恵袋について、21 年度から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。このようなサービス運用による対策に加え、ファクト情報の伝達・支援、啓蒙啓発・リテラシー向上施策にも取り組んでいます。

ファクト情報の伝達・支援としては、新型コロナウイルス関連情報や医療情報といった、特定の検索結果や知恵袋投稿の上部の目立つ場所に、公的機関等による正しい情報を掲載しているほか、偽情報等の打ち消し・注意喚起記事を、最も目立つ場所である Yahoo!ニューストピックスへ積極的に掲載したり、コメント欄や特設サイトなどにおいて、専門家による解説・フォローアップを行う等しています。

[25] 件数はいずれも 23 年 4 月 7 日現在

また、ファクトチェック・イニシアティブ（FIJ）や日本ファクトチェックセンター（JFC）の活動に賛同し、資金面での支援に加え、コンテンツ面での連携（特設サイトへのリンク、ファクトチェックコンテンツの Yahoo!ニュースへの配信など）を行っています。

啓蒙啓発・リテラシー向上施策としては、有識者やメディアと連携して、フェイクニュース対策に関する啓蒙啓発コンテンツ制作を行い、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～
https://news.yahoo.co.jp/special/fakenews/
Yahoo!ニュース健診
https://news.yahoo.co.jp/kenshin/

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。

（以　上）