資料15－1－3

Microsoft

デジタル空間における情報流通の健全性確保に関する Microsoft / LinkedIn の取り組み

デジタル空間における情報流通の健全性確保に関する　検討会

2024 年 3 月 28 日

目次

マイクロソフトの情報の完全性に関する原則

マイクロソフトは、信頼できる情報をサポートし、推進するために 4 つの原則を採用しています。

1

表現の自由

マイクロソフトは、表現の自由を尊重するとともに、マイクロソフトのプラットフォーム、製品、サービスでお客様が情報を作成、公開、検索できるようにします。

2

権威あるコンテンツ

マイクロソフトの製品では、内部データおよび信頼できるサードパーティのデータを活用することにより、サイバー影響工作に影響されないコンテンツの表示を優先します。

3

利益の不獲得

マイクロソフトは、サイバー影響工作に関連するコンテンツやそのアクターから意図的に利益を得ることがないようにします。

4

プロアクティブな取り組み

マイクロソフトのプラットフォームおよび製品が、サイバー影響工作に関連するサイトやコンテンツの拡散に使用されないようにプロアクティブに取り組みます。

有害なコンテンツに対するマイクロソフトのアプローチ

すべてのサービスに関して、マイクロソフトは 6 つの重点領域に基づいて、人々やコミュニティを保護する堅牢で包括的なアプローチを取ることに取り組んでいます。

1. 強力な安全性アーキテクチャ
2. 永続的なメディア来歴とウォーターマーク
3. 不適切なコンテンツや行為からの自社サービスの保護
4. 業界全体および政府や市民社会との強固な連携
5. 人々を技術の悪用から守る近代化された法律
6. 市民の意識向上と教育

Bing 検索: 調査とアクション用のツール

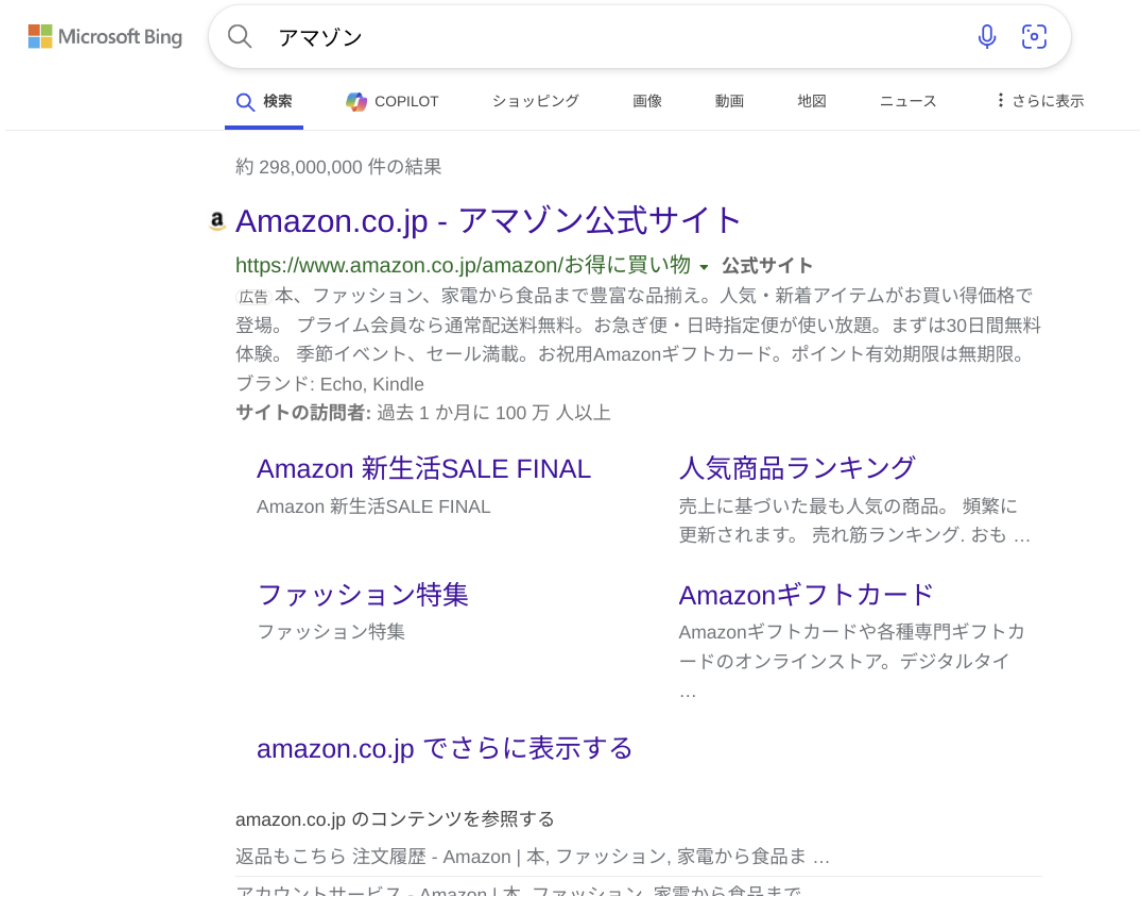
検索アルゴリズムが、検索インデックス内で**ユーザーのクエリ**に関連するクロールされたコンテンツを特定し、有用と思われる結果を優先的に表示します。有用性とは、バランス、または "関連性" と "品質" を意味します。"関連性" と "品質" は、**ユーザーの質問に対してどれだけ適切に回答しているか、専門知識、信頼性、および権威を評価するシグナル**により自動的に決定されます。

安全性と情報へのアクセスのバランスの維持

- ユーザーは World Wide Web 上の最も関連性が高く、質の高いコンテンツへの瞬時のアクセスを期待しています。
- 複雑で進化し続けるアルゴリズムが必要です。
- サードパーティによりホストされているコンテンツ。
- 情報へのアクセスの基本的権利。
- ユーザーに提供されるコンテンツは、一般的に "プッシュ" されるのではなく、"プル" されます。ユーザー クエリに固有です。
- "コミュニティ" なし – ユーザーはコンテンツを操作したり、バズらせたりすることはできません。

Bing のランキングの原則

- 信頼性が高い権威ある結果を優先します。
- 公平性とバランスを維持します。
  - ✓ 包括的なランキング
  - ✓ 複数の視点
- 予期しない攻撃的で有害なコンテンツから Bing ユーザーを保護します。
  - ✓ ただし、関連するすべての情報にユーザーがアクセスできるようにします。



Bing でのコンテンツのランキング方法

Bing の標準ランキング原則は、Bing の新機能においても、サードパーティのWeb コンテンツ提供時の指針となっています。

- 関連性
- 品質と信頼性
  - サイトの "権威" の分析を含む
- ユーザー エンゲージメント
- 鮮度
- 場所
- ページ読み込み時間

"権威" に影響する要素の例:

- 評判
- 議論のレベル
- 意見と歪曲のレベル
- 透明性のレベル
- 起源

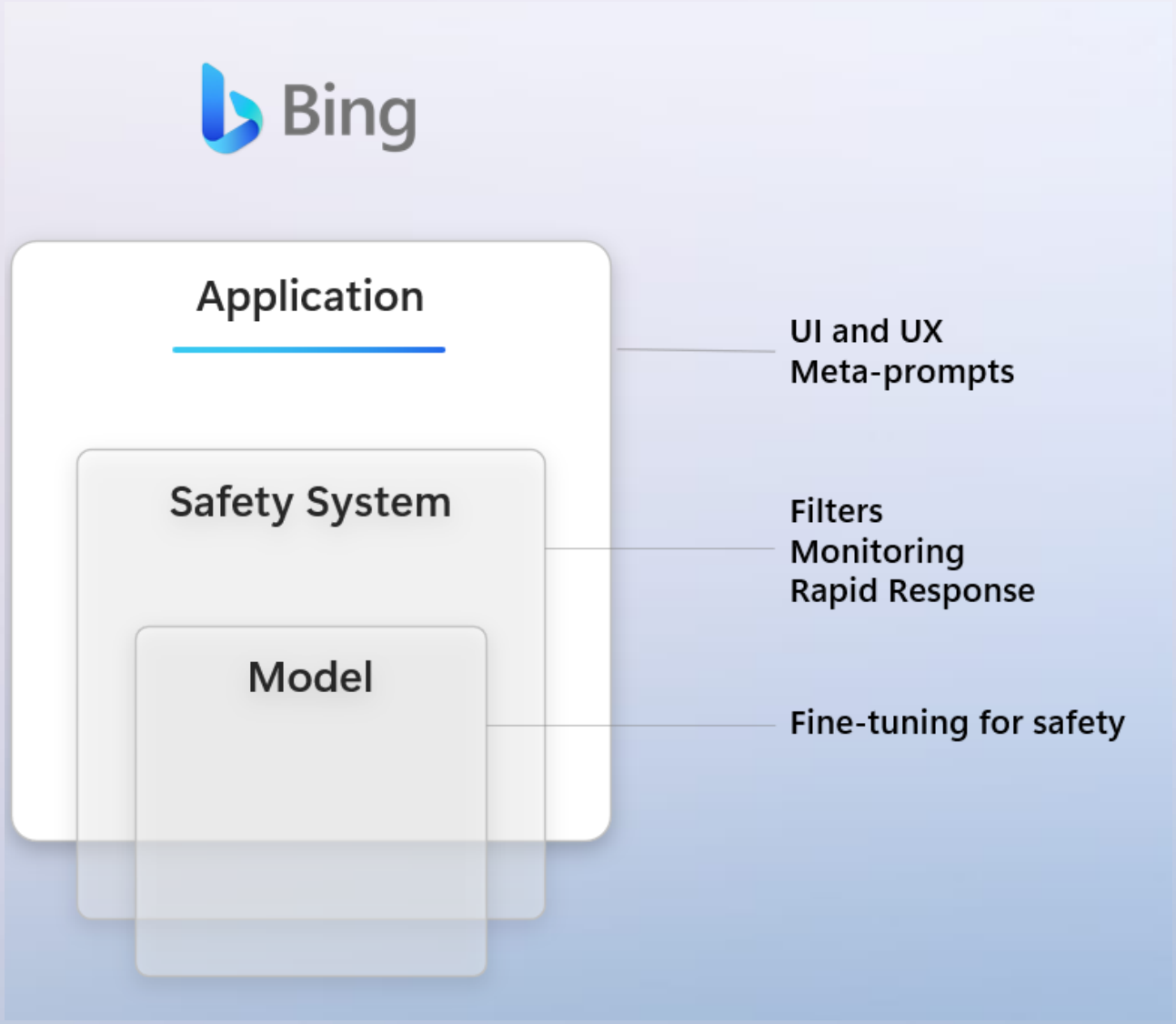
"関連性" の判断:

- 単語の一致
  - + 意味的同等性
- クエリの意図
  - 最良の意図を想定
  - 明示的な意図を尊重

Bing

責任ある開発に対する
マイクロソフトのアプローチ

弊害の軽減: 階層化アプローチ



セーフサーチ機能

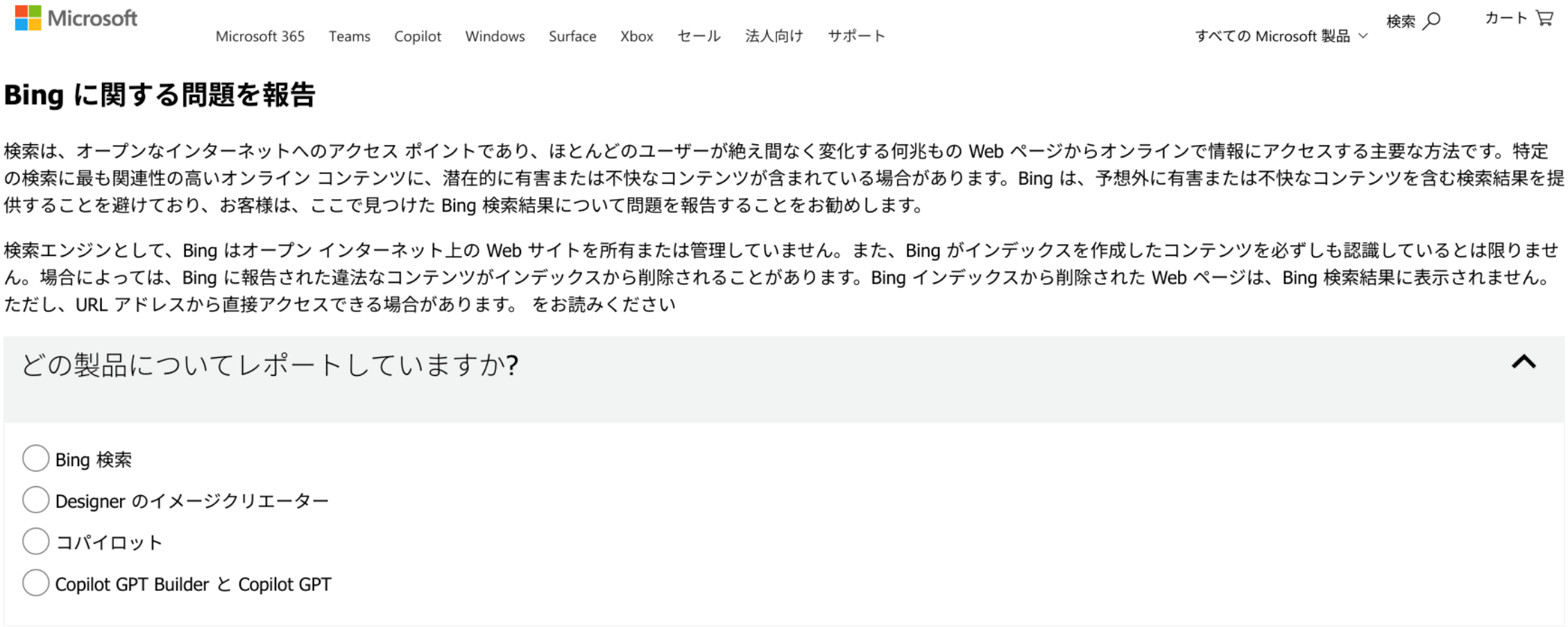

Microsoft Bing
ウェブ検索
English
設定
検索
検索
国/地域
言語
音声
ホームページ
個人用設定
セーフサーチ
高レベル
検索結果から成人向けのテキスト、画像、動画を除外する
標準
検索結果からテキスト以外の成人向けの画像および動画を除外する
オフ
検索結果から成人向けコンテンツを除外しない
不適切なコンテンツの表示を継続しますか?
場所
お住まいの市町村、都道府県、または郵便番号を入力してください。位置情報を使用することにより、より関連性の高い検索結果が示されます。
市区町村または郵便番号を入力
検索のキーワード候補
入力時に表示される検索候補を参照してください
検索結果
新しいタブまたはウィンドウに検索結果のリンク先を開きます
新しいタブまたはウィンドウにニュースの検索結果のリンク先を開きます

フィードバック ツール


Microsoft Bing
総務省
English
Switch to Bing in English
ログイン
30
モバイル
検索
COPILOT
画像
動画
地図
ニュース
ショッピング
さらに表示
ツール
約 63,900,000 件の結果
総務省
https://www.soumu.go.jp
総務省
ウェブ 総務省は、国の行政制度・運営、地方行財政、選挙・政治資金制度、情報通信などの政策を担当する省庁です。令和6年能登半島地震関連情報やマイナンバーカード、ふるさ ...
統計情報
統計情報 - 総務省
組織案内
総務省の組織を掲載しています。 自治行政局は、地方公共団体の円滑な行政運営 ...
改正法令
新規制定・改正法令・告示 法律 原則として、総務省が所管する主な法令・告示を ...
総務省｜サイトマップ
総務省ホームページのサイトマップを掲載しています。 法人番 ...
報道資料
総務省所管事業分野における障害を理由とする差別の解消の推進に関する対応指 ...
総務省の紹介
総務省では、行政機関が行う政策の評価に関する法律に基づき、平成14年度から ...
soumu.go.jp から結果を検索
検索
soumu.go.jp の他のコンテンツ
総務省｜行政相談｜災害時の行政相談活動
総務省 | 携帯電話ポータルサイト
総務省｜ふるさとワーキングホリデー ポータルサイト
総務省
日本の省庁
総務省は、日本の行政機関のひとつ。行政組織、地方自治、地方公務員制度、選挙、政治資計、消防など国家の基本的諸制度を所管してい
soumu.go.jp
ウィキペディア
YouTube
Facebook
Instagram
大臣 松本剛明
副大臣 渡辺孝一・馬場成志
大臣政務官 長谷川淳二・船橋利実・西田昭二
事務次官 内藤尚志
上部組織 内閣
内部部局 大臣官房・行政管理局・行政評価局
審議会等 地方財政審議会・行政不服審査会・
Bing についてフィードバックをお寄せください
フィードバックが関連するページの特定の領域をクリックしてください。
ご提案
良い点
改善が必要な点
フィードバックを入力 (必須)
リンクの欠落や誤りなど、詳細を入れてください。名前、メール、その他の個人情報は入れないでください。
スクリーンショットを含める
法律上またはポリシー上の問題ですか? プライバシーに関する問題を報告
送信
キャンセル

Bing

ユーザー レポート



Web 用の AI コパイロット (副操縦士)

適切な検索

Bing は安全で、より関連性が高くなった、使い慣れた検索機能であり、より優れたランキングと結果を返します。

Web のナビゲーション
天気の確認

完全な回答

Bing は Web 全体からの結果を検討して、ユーザーが求めている回答を見つけて要約します。

包括的な要約
比較的な洞察

チャット エクスペリエンス

チャットを使って質問し、提案を得られます。
複雑な調査を絞り込んで、より適切な提案を得ることができます。

旅行の計画
買い物の下調べ

創造性の刺激

既に存在するコンテンツのみの検索に限定されません。
簡単な記述を入力するだけで、新たなコンテンツを作成できます。

メールの作成
献立の作成

Microsoft Start

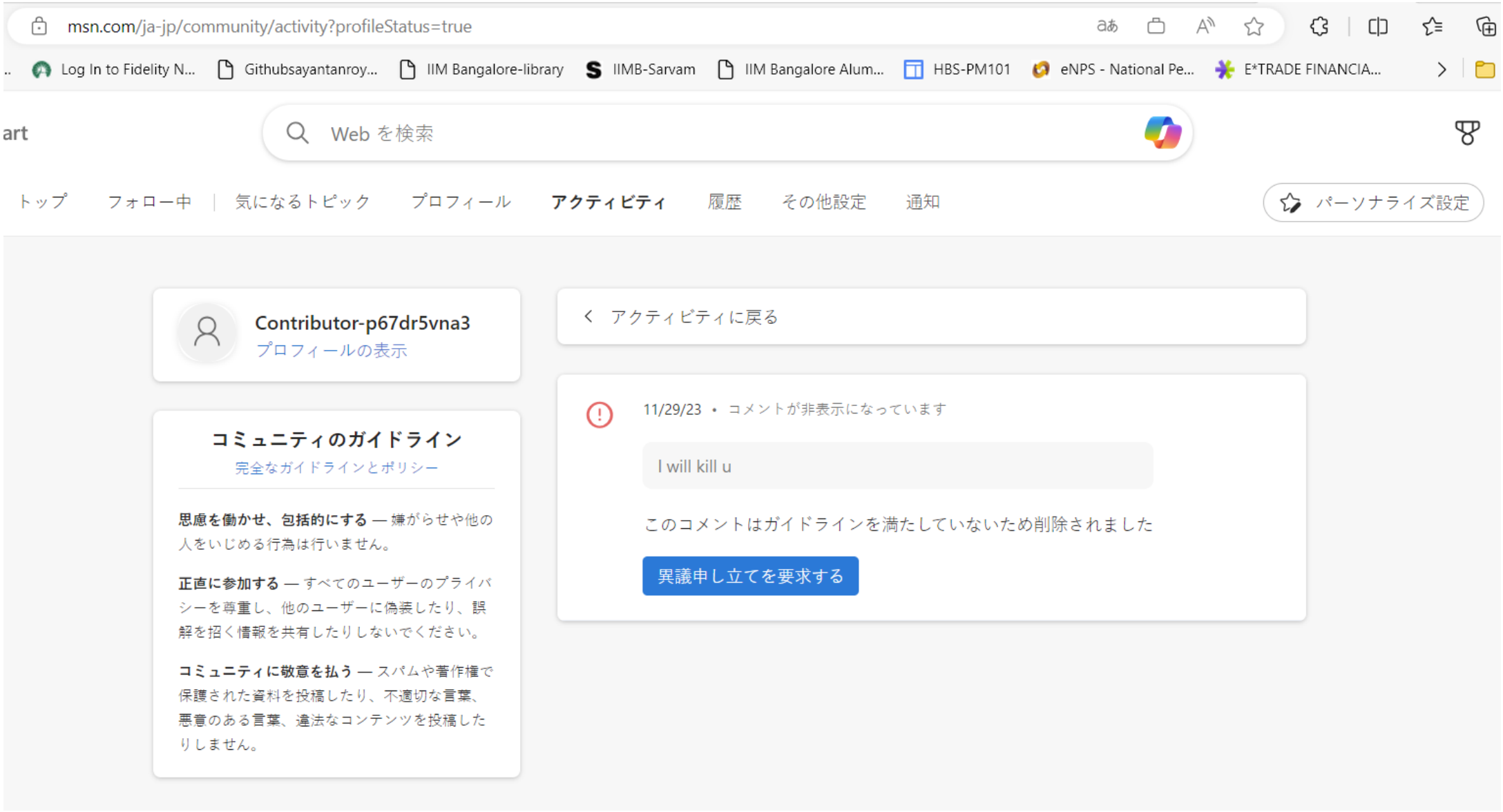
Microsoft Start は、

パーソナライズされたニュース フィードと情報コンテンツを提供します。プレミアム パブリッシャーからのニュースや、ユーザーの興味に合わせたタイムリーな更新情報が提供され、必要なときに必要な場所で利用できます。

Microsoft Start コミュニティのガイドラインは、
マイクロソフトのサービスで許可されているコメントの
種類を規定しています。

誤解を招く虚偽のコンテンツ、有害なコンテンツ、
その他の公衆または個人の安全、身体、精神、
財務上の健全性を損なうコンテンツ、論争を引き
起こすことを主な目的としたコンテンツを
禁止しています。

Start 上でのコメントは、自動化されたモデレーション
モデルの組み合わせと、ユーザーによる
"問題の報告" を通して管理されています。



必要に応じて、Microsoft Start は、特定のユーザーのコメント機能を停止する場合があります。
ユーザーは MS Start の自身のユーザー プロフィール ページで、コメントが削除された理由を確認できます。

LinkedIn

サービスの概要と
コンテンツ モデレーション ポリシー

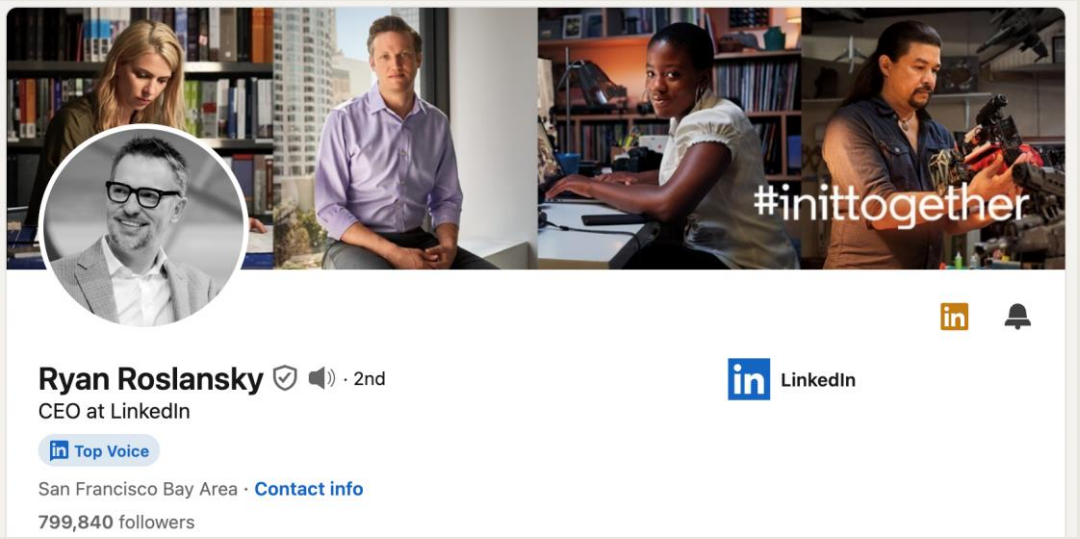
2024 年 3 月

10 億を超える
ユーザー
200 を超える国

- LinkedIn は実名制のプラットフォームです。
- LinkedIn は世界最大のプロフェッショナル ネットワークであり、200 を超える国で 10 億を超えるユーザーが利用しています。
- LinkedIn の**ビジョン**は、「世界で働くすべての人のために、経済的なチャンスを作り出す」ことです。
- LinkedIn の**ミッション**はシンプルです。「世界の人々をつなげることで個人と組織の生産性を高め、さらなる成功に結びつける。」

日本での LinkedIn

- 日本には 400 万人を超える LinkedIn メンバーがいます。
- 2023 年 12 月には 81 万 2,000 人を超える MAU を記録しました。
- 2024 年 1 月は、日本では、対象となる偽・誤情報に対するコンテンツ モデレーション アクションは実施されていません。



LinkedIn の基盤はメンバープロフィールです。

- メンバー プロフィールは、ユーザーの職務経歴を伝えるデジタル ポートフォリオです。
- プラットフォーム上でのメンバーの活動やメンバーが共有しているコンテンツはこのプロフィールに関連付けられ、他のユーザーが見ることができます。
- 多くのメンバーは自身の活動を関心のある専門分野に限定しています。
- メンバーは表示されるコンテンツがビジネスにふさわしい性質のものであることを期待しています。

LinkedInで上質なプロフェッショナルライフを経験していただくために

ここは、お互いを尊重し合い、お互いの成功を支援するコミュニティです。

安全性

LinkedInでは安全な話題のみで交流していただくようお願いします。

詳細はこちら

信頼性

身元を偽らず、真正かつ確実な情報を共有してください。

詳細はこちら

専門性

仕事に関するものであれば幅広い交流を許容しますが、プロフェッショナルであるよう心がけてください。

詳細はこちら

LinkedIn では、安全な環境を構築することを基本的な優先事項としており、これは私たちのビジネスにとって絶対不可欠なものです。

コンテンツモデレーションポリシー

偽アカウント

- 偽アカウントは実在する人に基づいていないため、誤・偽情報の拡散により使われうると考えられます。
- LinkedIn では、人工知能などのテクノロジと専門家チームを活用した偽アカウント対策に、積極的に多額の投資をしています。
- 昨年、LinkedIn は全世界で 8,000 万件超の偽アカウントをブロックしました。
- 2023 年には、検出されたスパムや詐欺の 99.6% を自動防御機能で削除し、検出された偽アカウントの 99.7% はメンバーから報告を受ける前にブロックしました。

コンテンツ モデレーション ポリシー

ユーザー生成コンテンツ

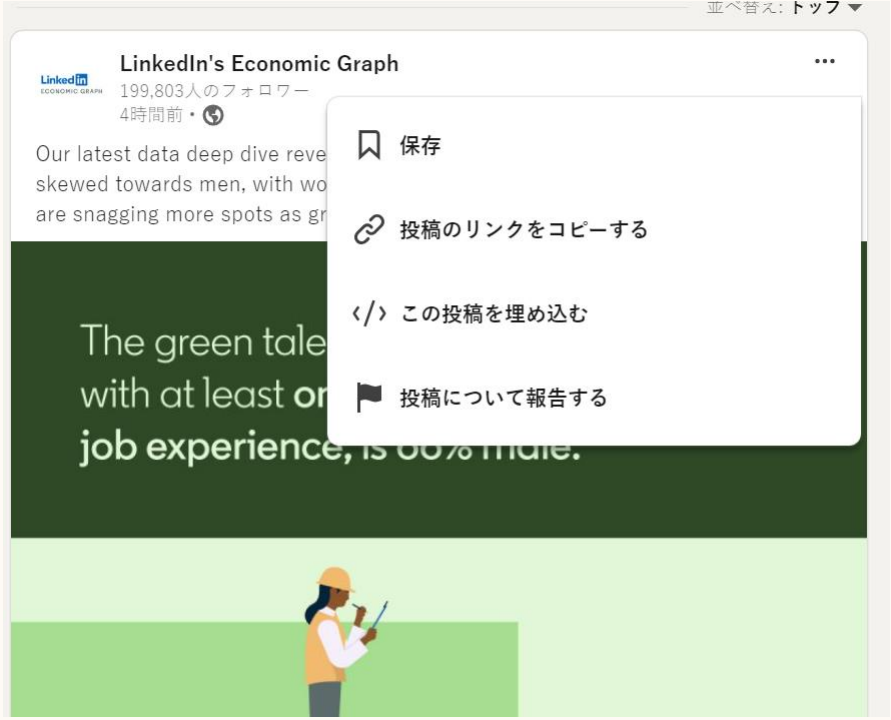
- メンバーが投稿するコンテンツについても、ビジネスに関連するものに限定しており、そうしたコンテンツを管理するための基準を策定しています。
- これらの基準を、将来を見据えた**プロフェッショナル コミュニティ ポリシー**で具体的に定めています。
- コミュニティ ポリシーでは、LinkedIn での職業倫理に反する行為は一切許されないと明言されています。
- たとえば、ハラスメント、誤情報、ヘイト スピーチ、またはいじめを決して許さないという方針が明確に定められています。
- これらのポリシーに繰り返し違反すると、アカウントが永久に削除される可能性があります。

LinkedIn の広告掲載ポリシー

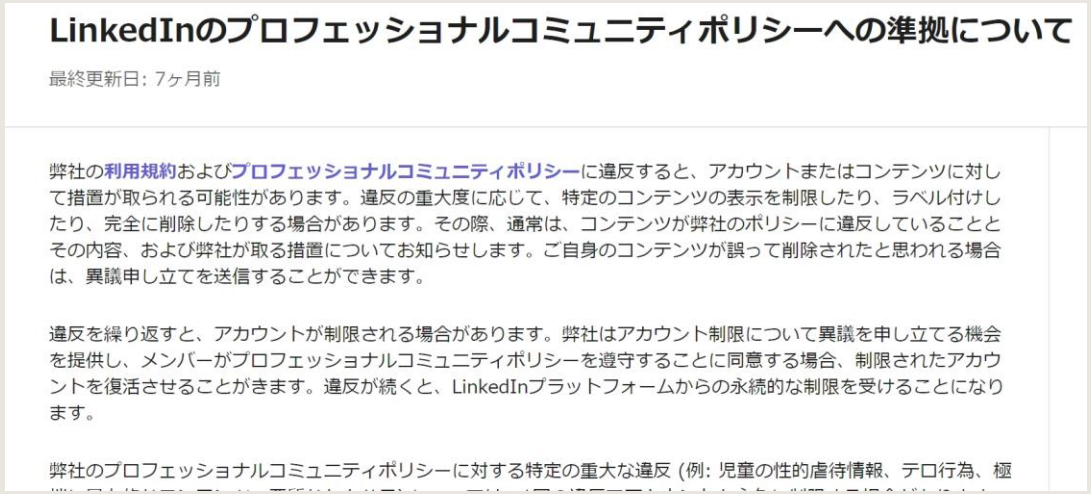
- LinkedIn の広告掲載ポリシーでは、誤情報、偽情報、詐欺および虚偽のコンテンツを禁止しています。
- LinkedIn はユーザーとの広告収益分配は行っていません。

違反の報告

- メンバーには、弊社のプロフェッショナル コミュニティ ポリシーに違反していると思われるコンテンツを、製品内報告メカニズムを使用して報告することを推奨しています。通常、報告されたコンテンツは訓練を受けたコンテンツ レビュアーが審査しています。
- さらに、個人情報の開示、スパムや悪意のあるページ、違法な素材のコンテンツなどの潜在的な違反を、自動的に社内のコンテンツ モデレーション チームに通報しています。
- 報告または通報されたコンテンツがプロフェッショナル コミュニティ ポリシーに違反していると判明した場合、そのコンテンツはプラットフォームから削除されます。



違反に対する措置



LinkedInのプロフェッショナルコミュニティポリシーへの準拠について

最終更新日: 7ヶ月前

弊社の利用規約およびプロフェッショナルコミュニティポリシーに違反すると、アカウントまたはコンテンツに対して措置が取られる可能性があります。違反の重大度に応じて、特定のコンテンツの表示を制限したり、ラベル付けしたり、完全に削除したりする場合があります。その際、通常は、コンテンツが弊社のポリシーに違反していることとその内容、および弊社が取る措置についてお知らせします。ご自身のコンテンツが誤って削除されたと思われる場合は、異議申し立てを送信することができます。

違反を繰り返すと、アカウントが制限される場合があります。弊社はアカウント制限について異議を申し立てる機会を提供し、メンバーがプロフェッショナルコミュニティポリシーを遵守することに同意する場合、制限されたアカウントを復活させることがきます。違反が続くと、LinkedInプラットフォームからの永続的な制限を受けることになります。

弊社のプロフェッショナルコミュニティポリシーに対する特定の重大な違反 (例: 児童の性的虐待情報、テロ行為、極

- 違反を繰り返すと、アカウントが削除される場合があります。
- アカウントの削除について異議を申し立てる機会を提供しています。
- 弊社のプロフェッショナル コミュニティ ポリシーに対する特定の重大な違反 (例: 児童の性的虐待コンテンツ、テロ行為、極めて暴力的なコンテンツ、悪質なセクハラ) については、1 回の違反でもアカウントが永久に削除される場合があります。
- 通常はポリシーの文言に違反するようなコンテンツでも、意識の向上や批判を目的としてシェアされている場合には、許容されることがあります。

透明性

LinkedIn の透明性レポートは、年に 2 回、次の URL で公開しています。
https://about.linkedin.com/transparency/community-report

Community Report

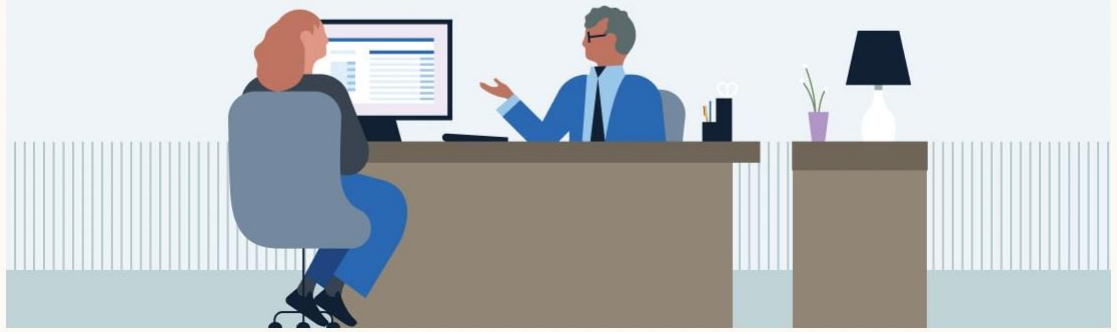
How we enforce our User Agreement and Professional Community Policies for our members globally.

Go to report

Government Requests Report

How we respond when governments ask for member data or for content to be removed.

Go to report



マイクロソフトの責任ある AI の基準

記録

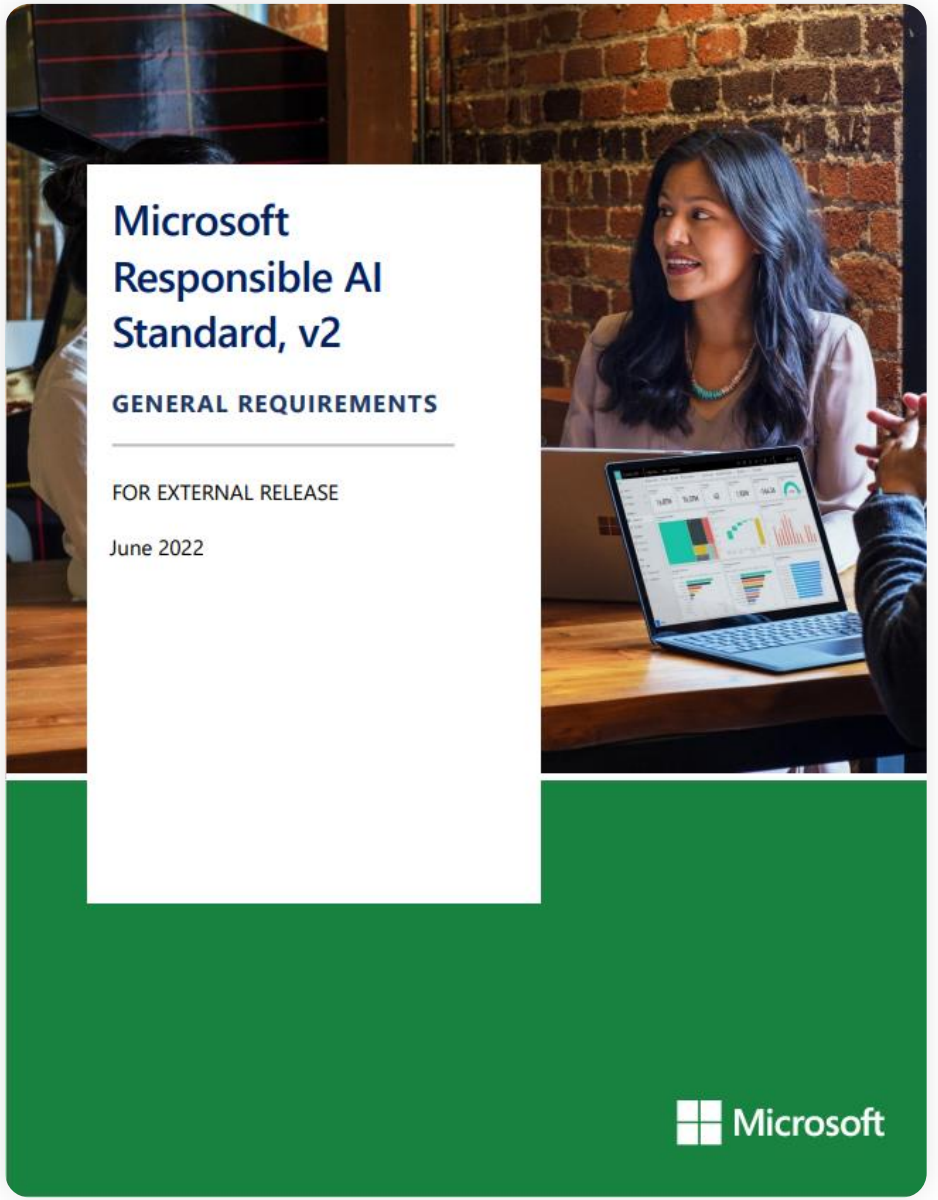
設計による責任ある AI のプラクティス、つまり AI システムの設計、構築、およびテストを手引きするプロアクティブな手段を記録します。

フレームワークの確立

責任ある AI のプラクティスの成熟化と規制要件の進化に対応する、**持続性のあるフレームワーク**を確立します。

反映

6 つの AI の原則の意味、およびそれらを遵守するのに必要な手順について、**より深く考察して**反映します。



マイクロソフトの AI の原則

公平性

信頼性と
安全性

プライバシーと
セキュリティ

包摂性

透明性

アカウンタビリティ

青少年の保護

継続的な調査と評価

セーフサーチ

ファミリー セーフティ設定

親と若者向けのリソース

4 つの重要な行動

ヒントとガイダンスを参考にして理解度を高め、インタラクティブ クイズで習熟度をチェックしましょう。

ID の保護

「hello1234」といったパスワードを使っていませんか? そのメールは本当にパスワード リセットのメールでしょうか、それともなりすましを狙った詐欺メールでしょうか? 被害にあうことが多いのは若い人たちです。[1]

自分を守る方法を学ぶ >

いじめ行為を 我慢しない

悪気のないコメント? それともネットいじめ? ソーシャル メディア上でのサイバー ストーカー行為、炎上、暴露、荒らし、なりすましを経験した生徒は、最大 34% に上ります。[2]

ネットいじめへの対処法 >

リスクのある状況を 見極める

自分の裸の画像を送信するのは、果たして賢明でしょうか? ネット上で 読む内容すべてを信じるべきでしょうか? 大人が 「グルーミング」しようとしている兆候を見極められますか?

リスクとその回避方法を知る >

自分の行動に 責任を持つ

自分と意見が合わないからといって誰かをはねつけるべきでしょうか? 気に障る投稿にはどのように反応すべきでしょうか? 「デジタル マナー運動」でネット上の ふるまいについて学んでみましょう

善良なデジタル市民になるには >

クラスルーム ツールキット

生成 AI を
安全に責任を持って
活用するために

Microsoft

Cyberlite

副操縦士としての AI

生成 AI ツールを協力的なアシスタントと考えましょう。生成 AI ツールはあなたの指示に従い、タスクをうまくこなしてくれますが、賢く責任を持ってそれらを利用する責任はあなたにあります。

AI は完璧ではない

AI ツールは多くのことをうまくこなせる一方で、常に回答を出すようにトレーニングされているため、間違うこともあります。したがって、注意を怠らないことが大切です。

必ずファクトチェックを行う

ファクトチェックを習慣にしましょう。AI によって生成された情報を盲目的に信じてはいけません。信頼できる情報源を使って必ず検証を行いましょう。

偏見に気を付ける

生成 AI モデルは時々、応答の中で偏見を示すことがあります。必ず、批判的な目で出力を吟味し、必要に応じてプロンプトを調整することで対応しましょう。

情報源の引用表記を必ず付ける

生成 AI の支援を得て完成した成果物には必ず引用表記を付けることで、功績を認めるようにしましょう。

自分の情報を保護する

信頼できない Web サイトやアプリに個人情報を渡してはいけません。また、プライバシー ポリシーを読んで、自分のデータがどのように使用されるのかを理解しましょう。忘れてはいけない点として、複雑な文書の要約に AI ツールを利用することはできますが、ファクトチェックと検証を必ず行うようにしましょう。

ウェルビーイングに注意する

LLM とのコミュニケーションは自然なように見えることがありますが、これには問題が生じる可能性があります。スクリーン タイムを制限し、実生活で大切な人と時間を共に過ごすことで、テクノロジとの健全な境界を設けましょう。


ありがとうございました