

Yahoo!メールにおける Abuse対策の最新事情

LY Corporation / Communication Company

Masaharu Nakamura

LINEヤフー

自己紹介



中村 成陽（なかむら まさはる）

2013/04～ 経路探索サービスを提供する会社 エンジニア

2016/11～ ヤフー株式会社(現 LINEヤフー株式会社) エンジニア

Yahoo!メールの迷惑メール対策チームに配属

2020/04～ チームのマネージャーとなる

2021/10 “第11代黒帯～メッセージング技術～” を拝命

2022/10 “第12代黒帯～メッセージング技術～” を拝命

JPAAWG

3rd General Meeting

「**Yahoo!**メールにおけるなりすましメール対策 **DMARC**導入とブランドアイコン表示」

4th General Meeting

「現場発！メールサービスを支える運用者の集い **2021 秋**」

5th General Meeting

「大手メールサービスにおけるセキュリティ・なりすまし対策の最新の取り組み」

Agenda

01

直近のフィッシングメールの流行状況

02

日本におけるスパムメール対策の難しさ

03

Yahoo!メールでのフィッシングメール対策
～ スパムメールの対象データ蓄積

04

Yahoo!メールでのフィッシングメール対策
～ DisplayName / TLD 不整合検知

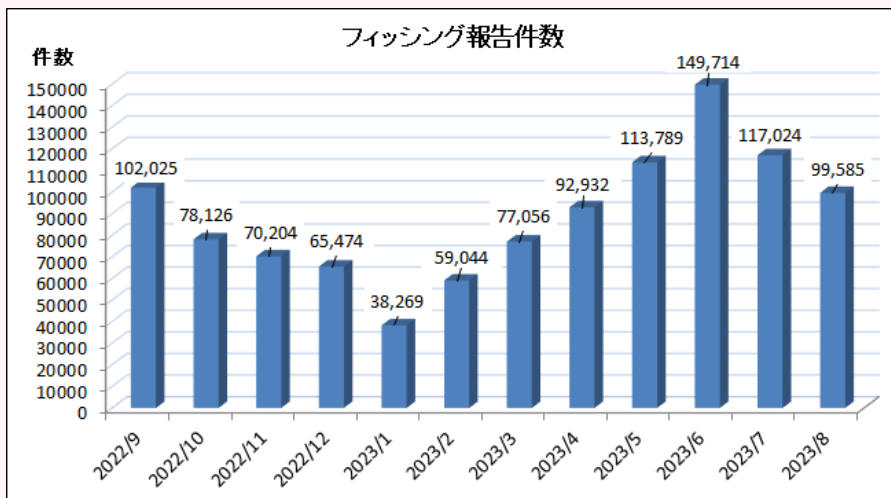
05

さいごに

直近のフィッシングメールの 流行状況

直近のフィッシングメールの流行状況

- フィッシング対策協議会が発表しているフィッシングの報告件数は増加傾向



※コロナ禍前の2019/06には6,218件
(出典：<https://www.antiphishing.jp/report/monthly/202308.html>)

- Yahoo!メールを利用しているユーザーからの関連問い合わせも直近増加している
- 大手ECや銀行などを騙るフィッシングメールが後を絶たない

直近のフィッシングメールの流行状況

- Cloudflare が8月に発表した「フィッシング脅威レポート2023」によると
フィッシング詐欺でなりすまされたブランドの
日本を含むアジア太平洋（APAC）地域1位は LINE であった

なりすまされた
APACのブランド上位:

1. LINE
2. JCBグローバル
3. インドステイト銀行
4. トヨタ
5. 東芝

(出典：https://release.nikkei.co.jp/attach/660827/03_202308171055.pdf)

- Yahoo!メールで受信したメールの DMARC の認証失敗通数は
2023年9月の一ヶ月分で昨年の同月比で約 69.2% の増加となっている

日本における スパムメール対策の難しさ

ユーザー嗜好

日本におけるスパムメール対策の難しさ

- （あくまで私見）日本のメールユーザーの嗜好
 - 届くはずのメールが届かない（誤検知）を非常に嫌う傾向にある
- 必要十分にスパムメールをブロックするのが理想だが、これは非常に難しい
 - フィッシングメールは日々巧妙化しており、対策をしてもいたちごっこ
 - そもそもそのメールが必要かどうかはユーザーによって異なる
- 厳しい条件によるメールブロックは技術的には難しくないが、現実的でない例）送信ドメイン認証に通過していないメールは全てブロックする、など…
- 可能な限り誤検知をなくす方針で対策を講じることになる
 - つまり検知漏れは増加する

法律

日本におけるスパムメール対策の難しさ

- 個人情報保護法や通信の秘密といった法律に準拠する必要がある
 - ユーザーのためのスパム対策であったとしても、個人情報を含む情報には最大限配慮する必要がある
 - 流行しているスパムメールの解析、分析をするとしても、通信の秘密（いつ誰とどんな通信をしているかといった情報を、第三者に知られないということ）に関する課題は常につきまとう
- 新しい対策を検討しても常に法務との確認が必要となり
対応ができない、あるいはできたとしても
当初よりも効果が薄いと見込まれてしまう方法での導入しかできないケースも多い
- DMARCの導入に関しては
これらについて法的な留意点を総務省が整理し、公開している

まとめ

日本におけるスパムメール対策の難しさ

- 法律を遵守した上で、約款の範囲内でルールベースのスパムメール対策を行うことになる
- ユーザー嗜好から誤検知によるスパム判定を可能な限りなくす方針とせざるを得ない
 - 一方で検知漏れによるすり抜けにも厳しい意見はいただく

この前提のもと、Yahoo!メールでは様々なフィッシングメール対策を行なっています

スパムメールの対象データ蓄積

Yahoo!メールでのフィッシングメール対策

スパムメールの対象データ蓄積

Yahoo!メールでのフィッシングメール対策

対象のスパムメール

- 根強く届き続けている **ウイルス添付ファイル付きメール**
 - 有償会員向けにウイルススキャンサービスを提供しているが全ユーザーに対策を提供する
- 性的脅迫を行いビットコインアドレスへの振り込みを要求する **セクストーションメール**
 - 5年ほど前から断続的に流行している
 - 内容がセンシティブであることから、受信時のユーザー問い合わせが多く発生してしまう

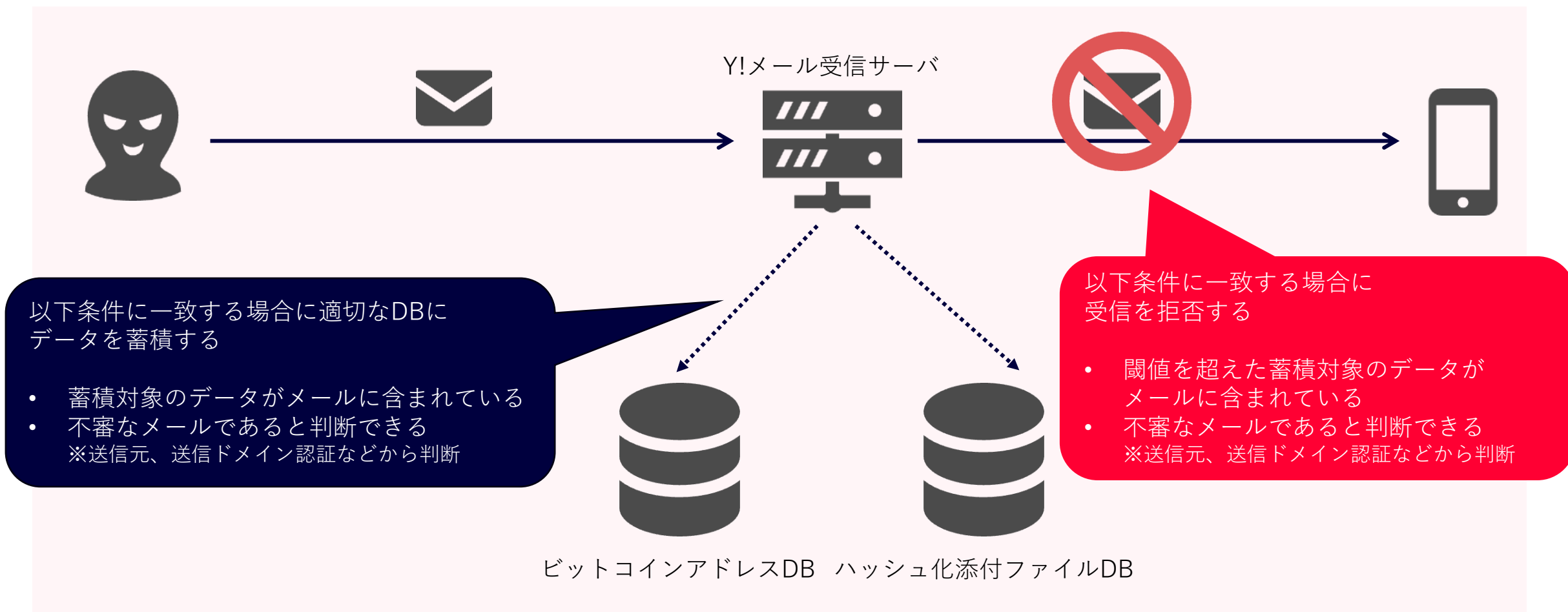
対策概要

- 不審な送信方法の対象データ※ を蓄積し、一定期間に閾値を超える通数を受信した場合にそのメールの受信を拒否する

※ハッシュ化した添付ファイルデータ、ビットコインアドレス

処理の流れ

スパムメールの対象データ蓄積



スパムメールの対象データ蓄積

Yahoo!メールでのフィッシングメール対策

Achievement

- 多い日には数100万通程の該当メールをブロックしている
- セクストーションメールの受信に関するユーザーからの問い合わせはほぼ0になった
- Y!メールからの送信の場合でも同様のDBを参照し不正利用防止に役立てている
- ハッシュ化添付ファイルについては特許を取得（特願2017-215181 /特開2019-087902）

Assignment

- 閾値に到達するまでには受信をしてしまうことになる
- データの蓄積、およびブロック対象とする「不審なメール」の判定条件はチューニングが必要

DisplayName/TLD 不整合検知

Yahoo!メールでのフィッシングメール対策

DisplayName / TLD 不整合検知

Yahoo!メールでのフィッシングメール対策

対象のスパムメール

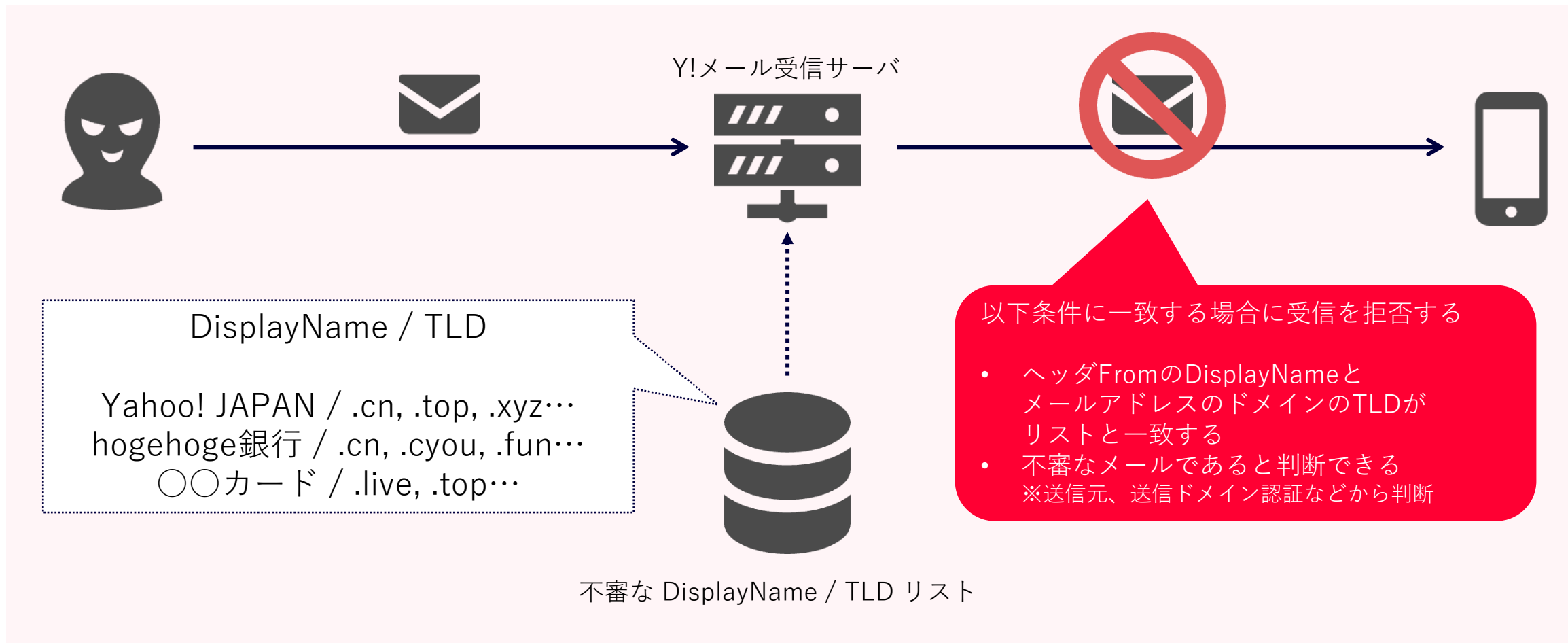
- フィッシングメールのうちDisplayNameで有名サービスを詐称するもの
 - 例) From: Yahoo! JAPAN <yahoojapan@example.xyz>
 - こういったメールはDMARCでのブロックができないため、課題となっている

対策概要

- 詐称されたサービスを表すDisplayNameと
そのサービスで利用されないトップレベルドメインの組み合わせの場合に受信を拒否する

処理の流れ

DisplayName / TLD 不整合検知



DisplayName / TLD 不整合検知

Yahoo!メールでのフィッシングメール対策

Achievement

- 多い日には数1500万通以上の該当メールをブロックしている
- DMARCでは捕捉しきれないDisplayNameを詐称したフィッシングメールをブロックできる

Assignment

- 自社以外、対象サービスが正規のメールで利用しているメールアドレスのドメインは自分達で収集せざるを得ない
- データの蓄積、およびブロック対象とする「不審なメール」の判定条件はチューニングが必要

さいごに

- 「ユーザー嗜好に合わせ可能な限り誤検知を減らす方針で、法律の範囲内で様々なスパムメール対策を実施する」というのがYahoo!メールでの方針
- 今回紹介した以外にも様々なロジック（ルールベース）や技術を利用したスパム対策を行なっている
- 受信側であるYahoo!メールのみでの対策では限界もあるためメールの送信を行う皆さんにも以下の対応を是非お願いしたい
 - DMARCへの対応（ポリシーのquarantine or reject化）
 - BIMlやYahoo!メールブランドアイコンへの対応
 - 正規メールで利用されるメールアドレス等の情報展開



LINEヤフー

LINEオープンチャットにおける Abuse対策の最新事情

LY Corporation / Communication Company

Toshihiro Harada

LINEヤフー

自己紹介

原田 利祐（はらだ としひろ）

～2004 個人事業主、会社経営（SEO/SEM、PCスクール/サポート）

2004～ ゲーム会社 プランナー/ディレクター

- ・ オンラインゲーム・サービス向けプラットフォーム
- ・ 会員コミュニティ、Eコマース等のWebサービス

2012～ ソーシャルゲームプラットフォーム会社 企画PM

- ・ ゲームプラットフォーム企画

2016～ LINE株式会社(現 LINEヤフー株式会社) 企画運営PM

- ・ LINE GAME PLATFORM プラットフォーム企画
- ・ LINEタイムライン（現LINE VOOM）サービス運営全般
- ・ LINEオープンチャット 日本サービスの立ち上げ
 - コミュニティ&リスクマネジメント等のサービス運営全般



01

LINEオープンチャットとは

02

オープンチャットのAbusing行為

03

モニタリングの流れ

04

対策事例の紹介① 青少年保護

05

対策事例紹介②

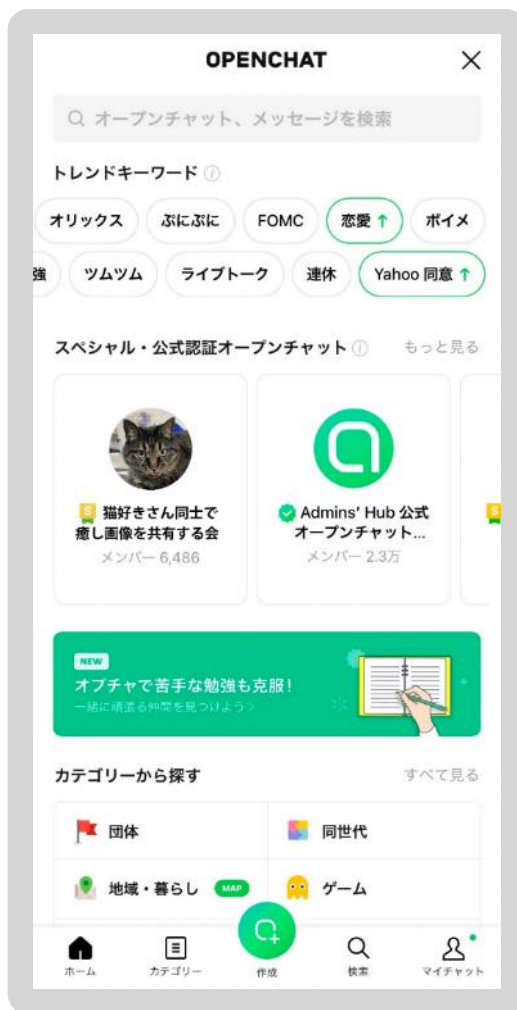
06

まとめ

LINEオープンチャットとは

興味関心ごとに対して複数人でトークができる

従来のLINEのプライベートのトークとは異なる別人格で、興味関心ごと別に複数人でトークが可能なオープンコミュニティ



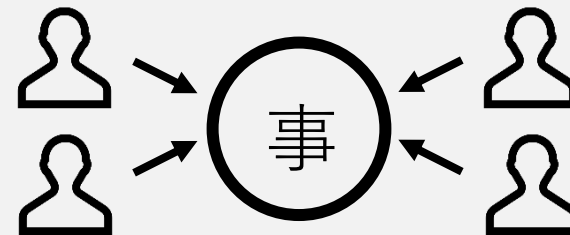
プライベートのトーク

家族・友だちなどリアルで繋がっている



オープンチャット

LINEアプリ上で見つけれられる



① LINEのグループトークとの主な違いは？

- ✓ 管理者の役割が存在しメンバーや投稿物を削除できる（勝手にメンバー削除されない）
- ✓ 個人のLINEのプロフィールとは別の人格で参加ができる
- ✓ すべての投稿物がモニタリングの対象（ガイドラインに基づき削除などの対応が行われる）
- ✓ 5000人以上参加可能（LINEのグループトークは最大500人）

オープンコミュニティは「運営の介入」がある

プラットフォームとして場所提供に加えてコミュニティの健全化への介入が必要となる

プライベートのトーク

完全なプライベート領域
リアル of 延長のつながり

運営介入

原則 なし

※規約違反等でのサービス利用停止はある

すべてが通信に秘密の領域

≠

オープンチャット

原則はオープンな領域
LINE上で新たなつながり

運営介入

あり

- ✓ コンテンツの削除
- ✓ ユーザーの利用停止
- ✓ オープンチャット露出制限 など

利用規約に基づき全てモニタリング対象

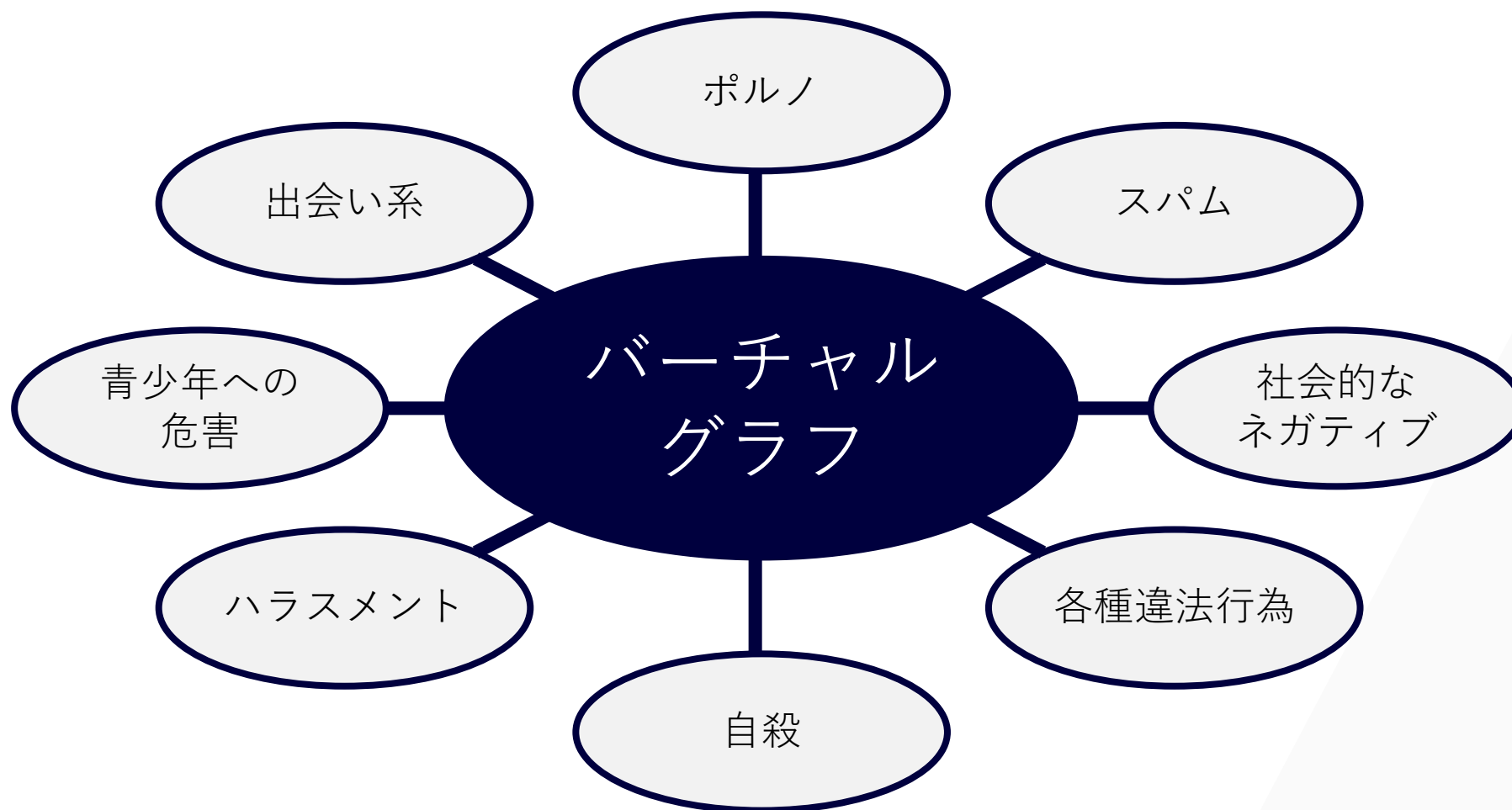
※公開設定で一部は通信の秘密に該当しますが利用規約に基づき全てがモニタリングの対象となります。

オープンチャットのAbusing行為

サービス開始から重点的に取り組んでいること

リアルグラフとは異なるコミュニティのリスク

LINEのプロフィールとは別の人格によってAbusing行為が発生しやすくなる状況が予測される



サービス開始直後の状況（2019年）

一斉に大量のユーザーが押し掛けたこともあり不適切な投稿が相次いだ



緊急

- ✓ 入口・検索の封鎖
- ✓ モニタリング強化
- ✓ 自動処理の強化

※当時の記事を再現したサンプル画像です

サービス開始から数か月後の状況（2019年）

徐々に社会的な存在意義や利便性の理解が深まってサービスが安定化してきた

運営

- ✓ 良事例の発信
- ✓ コミュニティ文化形成
- ✓ 管理者の教育



※当時の記事を再現したサンプル画像です

重点的に取り締まっているAbusing例

以下のカテゴリについてはサービスの存続にかかわるほどの最も重要なミッション

出会い禁止

LINE ID等の
連絡先交換の禁止

青少年保護

児童ポルノや
援助交際等の対応

違法行為禁止

宣伝スパム・詐欺行為、
誹謗中傷などの
社会問題

モニタリングの流れ

どのようにしてAbusing行為を検知・対応しているか

表現の自由と違反行為の境界線

表現の自由を前提としたうえでプラットフォームとして介入すべき範囲の明確化

安心・安全ガイドライン

知らなかったじゃすまされない! オープンチャットの禁止事項



出会い目的の行為・個人情報の投稿



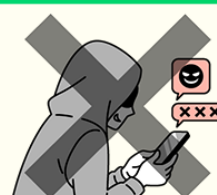
未成年の不健全な出会いや集まりの計画・勧誘



誹謗中傷・暴言・わいせつな投稿



著作権・肖像権・プライバシーを侵害する行為



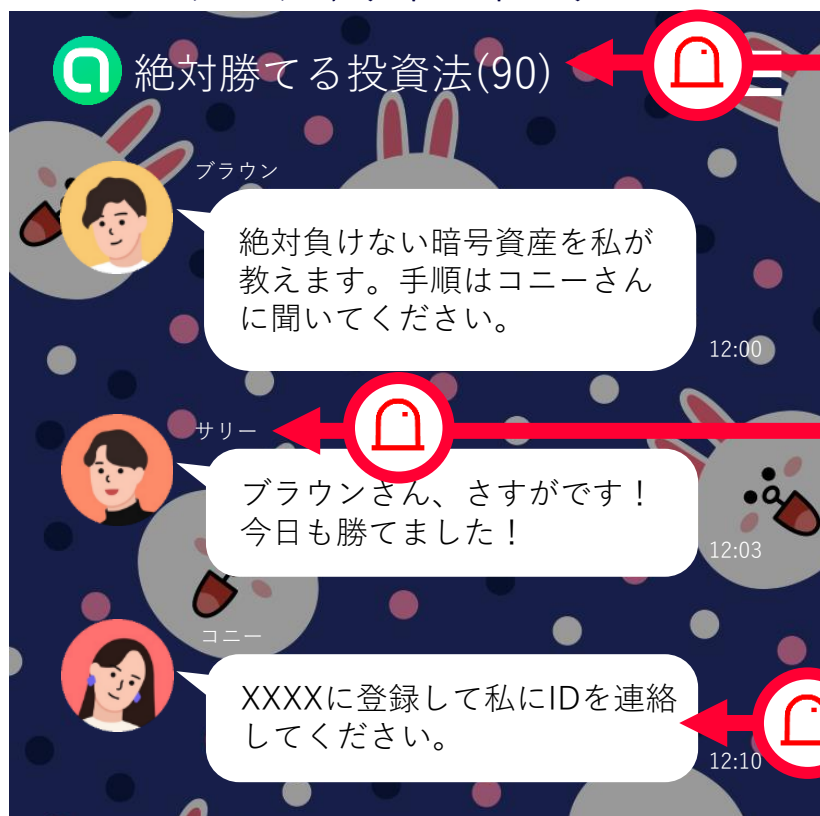
荒らし行為・違法行為・その他の迷惑行為

違反した場合はオープンチャット利用停止に加え、LINEアプリ本体の利用停止の措置を行うことがありますのでご注意ください。

モニタリングの対象になる単位

オープンチャット、投稿者、投稿した内容のそれぞれのを通報や自動検知によってモニタリングを実施

オープンチャットのトークルーム



オープンチャット
通報

個別ユーザー
通報

各投稿物を
通報

通報

通報する理由を以下から選んでください。

- ☒ スпам / 宣伝目的
- ☐ 性的いやがらせ / 出会い目的
- ☐ 迷惑行為
- ☐ トピックと無関係な内容
- ☐ その他

通報するとLINEヤフー株式会社に以下の情報が送信され、通報内容の確認・対応や不正利用防止ツールの開発を含む不正利用防止のために利用されます。また、上記目的の達成に必要な範囲で以下の情報を業務委託先に共有することがあります。

■送信される情報：
直近100件のメッセージ、通報相手の情報(ユーザー識別子/ニックネーム/プロフィール画像等)、オープンチャットの情報(オープンチャット名/メイン画像/メンバー数等)、トークルームの情報(トークルーム識別子等)、通報者の情報(ユーザー識別子/ニックネーム等)

同意して送信

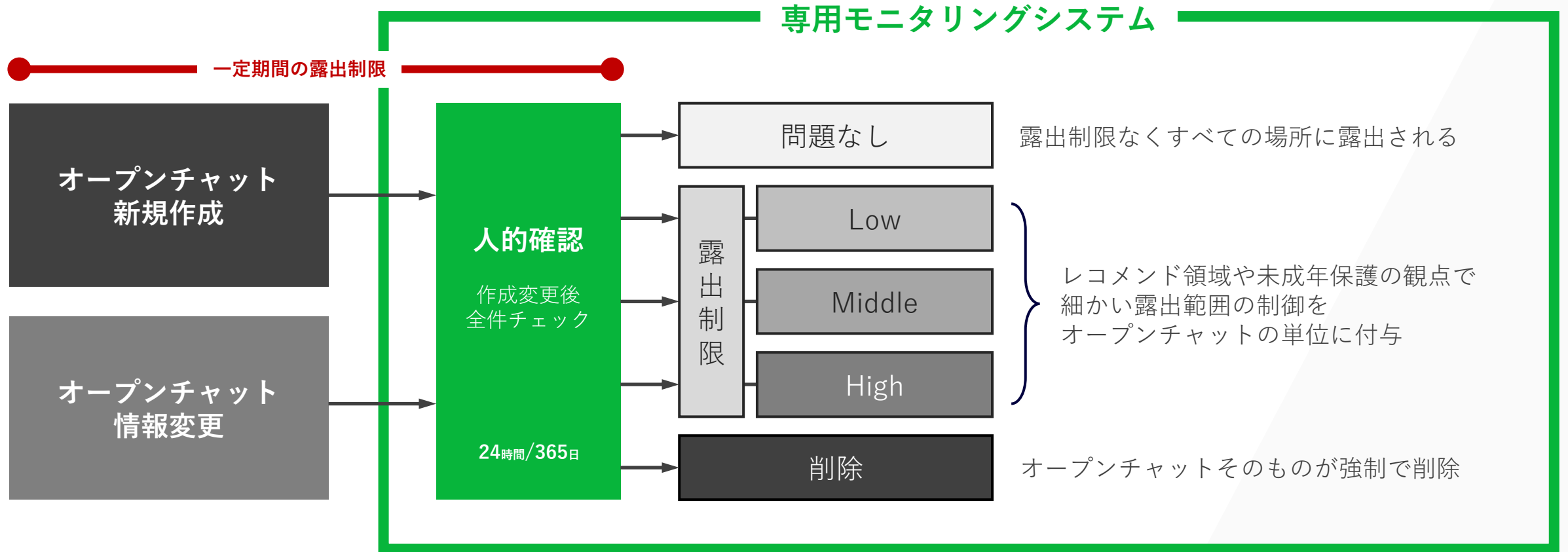
▶ オープンチャット
削除

▶ サービス自体の
利用停止

▶ 投稿削除

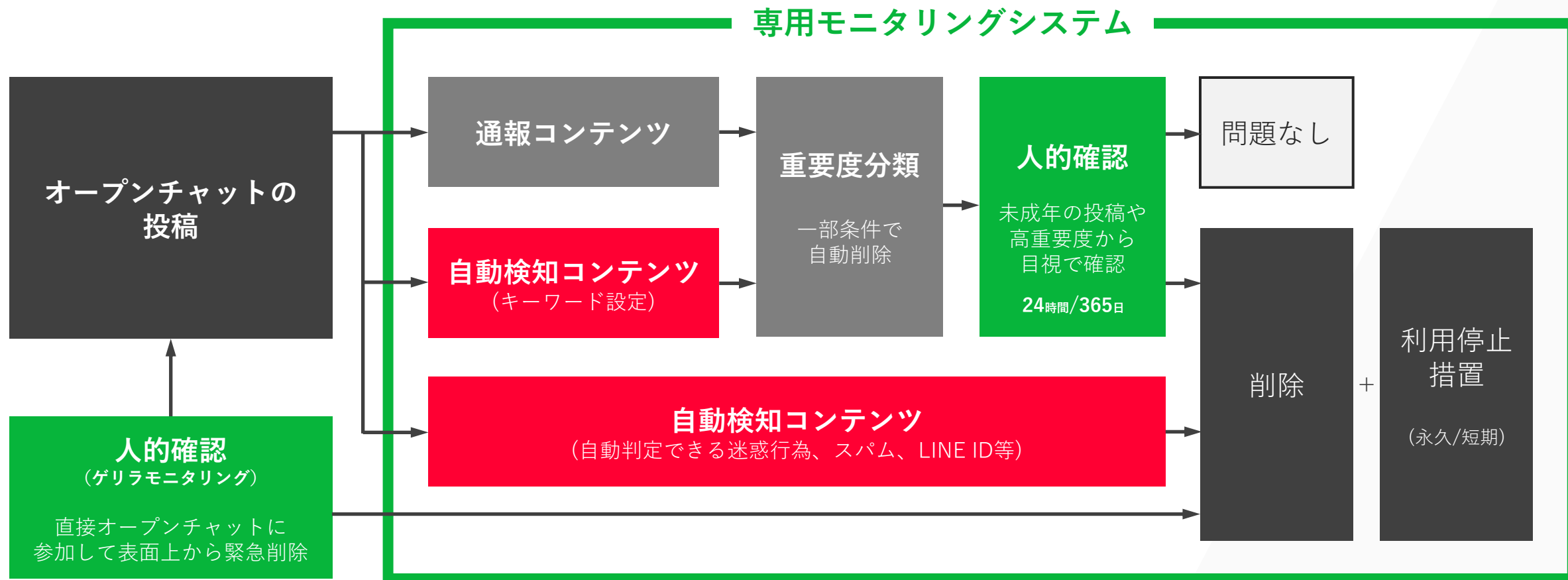
オープンチャット単位でのモニタリング

作成または内容変更されたオープンチャットは全て人的な確認をおこなってから露出



投稿単位でのモニタリング

自動化処理を積極的に活用しながら人的確認も併用し「できるだけ早く、正確に」システムを構成



自動化の現状と今後の取り組み

チャットならではの迅速な処理判断には自動化の強化が重要となる

現在

- ✓ 問題の可能性を含むキーワード単位の自動処理
- ✓ AIフィルターによる画像と動画の自動処理
- ✓ 自身のLINE IDや電話番号などLINEの友だち登録情報を自動処理
- ✓ スпам対応など数十種類のルールによる自動処理

自動検知したモニタリング対象のうち半数以上は自動で削除

今後

AI

AIの範囲をテキストにも拡大し、キーワード単位から文章を考慮
誤検知率を上げず自動検知と自動処理の割合を拡大させていく

対策事例紹介

青少年保護と投資詐欺対策

事例① 青少年保護

連絡先交換の全面的な禁止とユーザー啓発

事例① 青少年保護

幅広い年齢の利用者が集まるサービスでどのように青少年を保護するか

事例 大人との不健全な出会い

未成年の特徴

- ✓ 年齢と住まいを聞く
- ✓ 顔写真と自己紹介を登録
- ✓ カジュアルな会話
- ✓ 避難所を作る、他のSNS誘導

違反事案

- ✓ 年齢を偽り個人トークへ誘導する
- ✓ わいせつな写真を要求する

対策 モニタリング強化と啓発

- ✓ 未成年の検索制限やレコメンド制限
- ✓ LINE ID,QR,電話番号の自動削除 ※
- ✓ 未成年の優先モニタリング
- ✓ ゲリラモニタリング強化
- ✓ 未成年保護に特化した注意喚起
- ✓ 繰り返す場合は永久利用停止

出会いは禁止のサービス文化づくり

※未成年に限らず全年齢が削除の対象です

事例② 投資詐欺対策

著名人を名乗りエコーチェンバー効果による詐欺誘導

事例② 投資詐欺事案

複数の役割の人物を用いて劇場型で信用させる手法はオフラインでの詐欺手法と通じるものがある

事例 著名人の偽装とサクラで騙す

1. SNS等で著名人を装った広告出稿
2. 著名人をなりすまして
LINEの友だち追加をさせる
3. アシスタントを友だち追加させる
4. オープンチャットに招待する
5. サクラを使って信用を高める
6. LINEの1:1トークで投資指示
AI映像や著名人の合成ビデオ通話

対策 関連オープンチャット撲滅

- ✓ 海外LINEアカウントの大量保持
- ✓ 国外でオープンチャットを日本語作成
- ✓ 最初からサクラ要員の設定
- ✓ 同一会話を複数のトークルームで投稿
- ✓ 著名人や企業をなりすましている

などの特徴から

専用スクリプトや潜入モニタリングで
類似するオープンチャットごと一括削除

さいごに

オープンチャットAbusing対策のポイント

運営の介入を必要最低限に抑えながら日々変化するAbusingのトレンドに素早く正確に対応できるか

1

オープンコミュニティは場所の提供だけでは終わらない

2

表現の自由を尊重しつつどこまで介入すべきか明確な基準を設ける

3

日々増え続ける投稿と迅速な処理に対応するためAIの導入範囲を拡大

4

取り締まりを強化する一方で増加するフォルスポジティブを抑える

5

運営側の介入やAIだけに頼らず自浄作用が働く環境づくりも重要

ご清聴ありがとうございました



LINEヤフー