

Exhibit 41

FEF: Introduction

Page score: 76.58 · 📈 Manually synced: November 14, 2023, 11:33:45 AM · 10 views · Imported from · Managed by:

FEF: Introduction

(Work in progress)

What is FEF

FEF - Family Experiences Propagation (formerly SPACE = Signal Propagation and Consistent Enforcement)

FEF (Signal Propagation and Consistent Enforcement) is a project developed after Cambridge Analytica and other escalations to ensure consistent and timely enforcement against people who violate our policies. FEF propagates enforcement actions from one part of the Facebook family to the other platforms and products. Which policies and where these takedowns occur is based on Policy, Product and Legal guidance, and is only made in cases where the originating takedown and the connection between accounts are highly accurate.

FEF mission is to accelerate teams with a propagation platform that enables privacy-aware, cross-app and cross-surface enforcement across FB Inc.

How FEF works

FEF has been built as a configurable platform, which propagates enforcement actions across FB family of apps and surfaces, as defined by policy.

Policy can be configured in FEF in the following format:

When X happens, take:

1. enforcement action A, B, C.. in App 1
2. enforcement action B, C.. in App 2
3. enforcement action D, F.. in Surface A
4.

Where,

1. X could be any triggering event, such as deleting an account, accruing specified number of strikes etc.
2. Enforcement actions could be any action that runs on IAP (account deletion, page unpublish, feature limit etc..). Each of the actions can also be configured to run automatically or enqueued for additional human review.

FEF will handle the rest - executing actions, logging, reverting all actions if there is a successful appeal/reversal of the decision. It also comes with alerts, monitoring/dashboard and an engineer on-call.

Key Concepts & Terminologies

Match: This could be hard matching or soft matching. By hard matching for FB user and IG user, we mean the user explicitly tells Facebook both the FB and IG belong to them.

Hard Match/Connection: If a user explicitly tells Facebook both a FB and IG account belong to them, we say this FB account and IG account is a hard match/connection.

Soft Match/Connection: this means Facebook uses our models and take in features like phone numbers, email, etc to predict the probability of a FB user matches an IG user account. In FEP, we are currently taking FB-IG pair as a soft match if the probability is $\geq 99\%$.

Propagation: it's a cross surface enforcement, originating from a seed entity. For example, when a FB user is disabled, we disabled the hard connected IG account. This IG account disable enforcement is a propagation.

Seed Action/Seed Entity: The action on a seed entity. For example, when Facebook user is disabled, we disable the hard connected IG account. The Facebook user disable is the seed action and the Facebook user is the seed entity.

Propagated Action/Entity: The action on a propagated entity. For example, when Facebook user is disabled, we disable the hard connected IG account. The IG user disable is the propagated action and the IG user is the propagated entity.

Trigger: The violation category classified by FEP that has the same set of propagations. For example, CEI trigger consists of a number of CEI violation related codes, [code pointer](#).

FEP Value proposition

FEP team has built infrastructure that ensures quick actioning and provides transparency into our enforcement. We envision FEP as the key X-Surface propagation platform for the company and committed to its full support going forward, including:

1. Flexible enforcement propagation infrastructure that is easily adjustable based on a specific use case.
2. Reenable propagation infrastructure. In cases when initial (seed) enforcement action is reversed, we automatically reverse all follow-up actions.
3. Backfill and re-tires process that allow for any error mitigations.
4. Detailed [dashboard](#) that provides enforcement and reversal data on a disable code/label level (most granular level).
5. Performance tracking and thresholds that allow us control the quality of our propagations. Current thresholds:
6. Soft-match threshold: in instances when we use soft matches, we only employ ML with 99% precision rate
7. Reenable threshold: current acceptable re-enable rate threshold is 2.5% over long-term.
8. Alerts on system and harm levels.
9. On-call Eng and Ops personnel.
10. [Governance Model](#): we have developed clear set of rules and processes to ensure it is easy for partner teams to engage with FEP.

FEP current coverage

Apps and Surfaces coverage

When a Facebook user is disabled due to covered violation types and disable codes, we will do these auto enforcements:

1. **IG:** Disable the hard matched IG accounts
2. **Pages:** Force unpublish their solo-admined pages
3. **Groups:** Remove the user from the admin list for groups they are an admin of
4. **Business Managers:** Disable their solo-admined business managers
5. **Financial Entities:** Disable app linking for solo-admined financial entities that have no apps linked
6. **Ads:** Ads feature limit for the user
7. **Marketplace:** Remove their access for marketplace tab
8. **Ad Accounts:** Disable their solo-admined active ad accounts
9. **Whatsapp:** Disable the deterministic matched Whatsapp accounts (with the same phone number and device) after 14 days. For more information about FB-WA propagations, see [this doc](#).

Note: Items 2 - 8 is in the category of FB→Blue App.

When an IG user is disabled due to covered violation types and disable codes, we will do these auto enforcements:

1. **Facebook User:** Disable the hard matched Facebook accounts

Policies coverage

Currently, FEP is only propagating Red-level signals (user disables) for worst of the worst violations. We are working to expand to less severe violations.

Originating signal	Enforcement Propagation			
	FB->Blue App	FB->IG	FB->WA	IG->FB
Child Exploitation Imagery (CEI)	✓	✓	✓	✓
Counter-Terrorism (CT)	✓	✓	✓	✓
Grooming (IG)	✓	✓	✓	✓
Human Trafficking	✓	✓	✓	✓
Mass Murder	✓	✓	✓	✓
Sexortion	✓	✓	✓	✓
Intellectual Property	✓			

✓	Launched
✓	WIP

FEP Metrics

- Results highlight the pervasiveness of bad actors across the FB Family of Apps
- In H2 we enforced on millions of worst violators while keeping reenables below 2.5%

Surface	H2 2019	SPACE Takedown Volumes				
		December	November	October	September	August
FB Users disables	2.6MM	417K	478K	425K	551K	355K
Seed Takedowns						
Instagram users	852K	107K	140K	114K	762K	92K
WhatsApp users	345K	52K	65K	63K	90K	60K
Ad accounts	308K	43K	57K	48K	83K	37K
Business managers	62K	9K	13K	9K	17K	7K
Pages with copyright	311K	42K	56K	43K	84K	38K
Groups admin removal	228K	32K	39K	36K	55K	28K
Marketplace sellers active	208K	26K	36K	18K	80K	33K
Total Propagated Disables	2.3MM	311K	409K	336K	644K	298K
SPACE Counter Metrics (Reenables)						
Metric	H2 2019	December	November	October	September	August
Seed reenables	58K	3.4K	14K	5.7K	12.3K	14.8K
Seed reenable rate	2.2%	0.8%	2.9%	1.3%	2.22%	4.11%

What FEP is not

- Content - The system scope is actioning at actor level. Deletion of related/fan-out bad contents is not part of this system.
- Recidivism - This system does not address recidivism. Bad actors could attempt to create new accounts/profiles and try to get back on to the platform. We have a separate Integrity Actor Matching effort. In the future we may start to combine these efforts. FEP/SPACE is now the primary tool for prevention of recidivism. See here for more information: https://www.internalfb.com/intern/wiki/Actor_Accountability/Documentation/Space/SPACE_Arc
- Actor activity observation and classification - This system does not attempt to observe actor activities across FB Inc and identify/classify actors. The system is about defining "what to do" not "what is bad and how bad it is". (FUSS might be a better fit for that purpose).
- Investigation platform - FEP is not built to run investigations into fanned out entities. It propagated enforcement based on pre-defined set of policies and rules.



« Integrity Action Self Help Guide »

« Integrity Action Self Help Guide »