



I. Collection, Use, and Disclosure of User Data (Topics 2(a), 2(b))

A. SME: Elizabeth Bugayong (Advertising)

- Privacy Policy: how data collected is used for advertising purposes
 - Privacy policy does not make a distinction between teens and adults on this point
 - Meta provides clarification on additional limitations for teen users in the privacy center that says that the only data we use for teens in ads are age and location. This is not included in the policy.
 - Categories of data that are addressed in the privacy policy as data we use to show ads
 - Profile information
 - Activity off our products
 - Content that you create across the app
 - Inferences
 - Meta uses stated age to target ads to users
 - Information we share with 3rd parties is aggregate, anonymized reporting data eg: total impressions an ad got from 13-17 year olds
- Data use and targeting restrictions on Youth Ads
 - Delivery means what we do to show an ad - most of the time this term refers to Ranking
 - Targeting refers to an advertiser selecting their audience
 - For BAY, options that advertisers have for targeting are minimal. In BAY, you can only choose age and city-level or higher location.
 - Pre-2021, Youth Ads used all data sources but had content restrictions for youth (eg: alcohol, gambling). All users including teens had the ability to opt out of the use of 3P data.
 - 2020-2021 time period, Meta began to look at a different approach for Youth Ads targeting. With GDPR, we introduced low-data use ads types in Europe for certain teens. In 2021, other laws passed that required that we did not create a "profile" of a teen (ie, the company may not use data to create a prediction of a teen's interests for advertising purposes) and Meta therefore decided to launch the low-data version globally to U18.
 - Slimmed Down Services - SDS. Teens were defaulted into the setting where 3P data was off for ranking, targeting and delivery use cases. Targeting was only on age, gender, city-level or higher location. Launched in Sep 2021.
 - Basic Ads Youth Tier: Announced in Jan 2023. Began to make advertiser side changes. Started rolling out in March 2023 to users. Launch

II. Identifying and Removing U13s; Data Deletion (Topics 3, 4(a), 4(b), 4(c), 4(d), 4(e), 4(f), 4(i))

A. SME: Ty Swarts (general background)

- User reporting on IG and FB
 - IG:
 - IG has 2 forms of user reporting: 1) contact form - reporting form at account level 2) in-app reporting / simplified user reporting flow
 - Contact form:
 - Links to help center article with link to a form that the user has to fill out. Information about the reporter and the account they are reporting.
 - In-app reporting is a button that allows a user to report another account as U13. Only available to 4% of the worldwide population. Testing since 2022. Has shown that it doubles the rate of false reporting so we have not rolled it out further.
 - Users do not have to have an Instagram account to report a U13 account. They can find the form in our help center.
 - FB:
 - Same as IG but only have contact form
 - We had a system for FB users to report IG users who x-posted content to FB but did not have FB profiles. This system generated a huge amount of false reports - 99% false. This is because the users were creators who are frequently targeted for false reports. We have deprecated this system because of the level of false reports. Users can still report these profiles on Instagram.
- Signals a user is U13
 - 4 ways users are detected:
 - Proactive: look at explicit or implicit admissions of age in bio or caption, FB and IG. Oct 2024
 - User reporting - details in other notes
 - Internal escalation - a different human review identifies an account that may be U13 and the account is escalated into the U13 queue
 - User self reports - edit DOB to U13. Does not go to human review, goes directly to UFAC aka checkpoint.
- Automation performed after accounts are flagged as potentially U13
 - Signals that these accounts are obvious adults, if yes, removed from the system
 - FB automation policies:
 - Pre-October 2024

- Adult classifier with predicted age of O18. Precision threshold of 90%
 - In the contact form, if the user report includes a birthday over 13
 - Account has existed for longer than 8 years
 - Previously manually reviewed
 - No media or no bio associated with the profile
 - Previously cleared appeals
 - Ignore if AP+ because it has multi-admin, does not represent an individual
 - If account is verified meta employee
 - If user self reported themself
- IG:
 - Pre Oct 2024
 - Adult classifier
 - If previously reviewed and marked as 18+
 - Ignore if no bio or media
 - Ignore if previously cleared appeal
 - Ignore if account age under 8 years
 - Ignore if user has successfully age verified
- Human review process
 - Half of the accounts are cleared by the automation policy
 - Accounts that are not cleared by automation go through human review
 - Human review queues are shared but underage reports are the top priority. Internal SLA of U24 hours but most finished in 4 hours.
 - Human reviewers follow a checklist:
 - Is the account displaying Child Sexual Abuse Material or Child Sexual Exploitation content? If so, account is escalated to a CSAM or CSE escalation
 - Does the account represent a non-human entity eg: pet or business?
 - Is there an explicit admission of age in the bio or in the “age assist widget” which will show historical elements of the profile. This was updated in Oct 2024.
 - Prior to Oct 2024, there was no non-human question
 - Prior to Oct 2024, could only see current not historical bio
 - Did 50% of the photographs in the account represent an underage user? In Oct, lowered the threshold to 25%
- Age checkpoints

- If yes to any questions except non human entity then account enters UFAC
- Meta's Oversight board maintains a list of people that are journalists, political figures, major celebrities. If an account is flagged that matches this list of people, it is intercepted on way to UFAC and reviewed by FTE against the policy
- Once in UFAC, you have 30 days to appeal
- If do not appeal in 30 day window, begins the clock for deletion
- If a user appeals and hard fails [explanation in other notes], clock for deletion begins immediately.
- Deletion process for online data (data in prod)
 - Deletion process begins on day 31 and takes 90 days to complete
- Abuse of the U13 reporting flow
 - One of most abused reporting flows at Meta
 - Well known way to get accounts taken down because it ends in account suspension / checkpoint
 - Top reported account in 2024 was the IDF by an extremely large margin
 - Other reported accounts, several times a day, Kim Kardashian, @creators handle, Logitech the mouse company was reported 200 times over the course of a year
 - **Stats:** in the US, ~14% of human reviewed reports are checkpointed
 - This varies by market. In Brazil, 60% of human reviewed reports are checkpointed. In India, only 1% are checkpointed. In India the abuse vector is primarily men harassing women. In the US the abuse vector is public figures or teen bullying.
- Cross-app enforcement
 - When an account has been triggered for a checkpoint, this checkpoint propagates to any hard-linked accounts, meaning accounts the user has explicitly linked
 - The user is only able to appeal from the original seed account that was checkpointed
 - If the user fails or does not complete appeal, all accounts are deleted

B. SME: Rohan Pritchard (cross-app enforcement)

- Underage enforcement in mid 2021
 - Using UFAC on Facebook and an old checkpoint experience on Instagram
 - At some point previously, cross-app propagation had been set up. This worked for hard linked accounts. It was set up at some point prior to 2020. At the point Rohan inherited it, it was broken.

- Late 2021, fixed cross-app propagation. At some point in the past, the way the triggers worked had broken.
- Cross-app propagation, 2021 to present
 - If a report comes in, we assess and then checkpoint them
 - They are put in a checkpoint or appeal experience for 30 days where they can submit an ID
 - At the point of checkpoint, we propagate the checkpoint to the hard linked account.
 - You can only create an appeal from the source account.
 - (Eg) If you are checkpointed on IG, your FB account is disabled. You will be asked to log in to Instagram to look at the enforcement. This works in both directions.
 - Cross-app propagation infrastructure supports multiple types of enforcements in addition to underage
 - Only works for hard link propagation. Looked at possibility of doing soft link propagation.
- Abuse of the reporting form
 - Most of our reporting processes for other issues are in app
 - In the case of underage, the reporting form was at the time an off platform / logged out contact form. We needed that because we wanted people without accounts to be able to report underage (eg: teachers).
 - As a result of it being off platform / logged out contact form, we experience a lot of abuse in the form.
 - Manual reviewer is told if unsure the account is underage, take it down. So it was therefore an attack vector where people would mass report accounts.
 - Introduced IP-based rate limiting and CAPTCHA. Also global-based rate limiting.
 - Never got to the bottom of the motivation of the attackers.
 - Scripting attack examples:
 - Delta checkpoint experience - multiple checks across devices that you have logged in. "Did you log in on this device?"
 - A bunch of malicious users got access to accounts via eg hacking passwords and they then got the Delta experience. They would get blocked by 2FA.
 - They figured out that if you reported the account as underage and it was a young-looking account, the underage checkpoint was higher priority than Delta checkpoint.

- Underage checkpoint was logged in, if they were logged in, they could submit an ID and when the appeal is successful they were given logged in Delta experience.
- This was just on IG.
- Reports of 600K users over the course of a few hours from this type of attack.
- There were 2 or 3 large scripting attacks like this that happened before we introduced friction.
- The scripting attacks were fixed in part by moving to IG UFAC so it was a logged in experience (2022).
- Scripting attacks were not the only type of form abuse:
 - Eg: celebrities / influencers frequent target of underage reporting
 - Legacy checkpoint experience on IG. In order to appeal, you were provided a contact form that you had to fill out. Before the migration to UFAC happened, that form was also subject to abuse as well. Attacker would submit an appeal with an ID that was underage on purpose to make sure the account was down forever.

C. SME: Kitty Lam (2021 Backlog)

- Background
 - Jan 2021: Increase of FB and IG jobs because of more accounts being created overall and raters being sent home due to Covid
 - Flows that go to human review:
 - Edit DOB: When a user requests to change Edit DOB.
 - There was discrepancy between IG and FB:
 - On IG, you could not select a birthdate under 13 because that does not meet TOS
 - On FB, when you edit DOB you can put your age as U13 so Edit DOB queue needs U13.
 - Experience was standardized to include ability to put a U13 birthdate
 - U13 reporting: User account is placed in underage human review queue. Account if believed to be U13 is checkpointed.
 - U13 checkpoint appeal queue is a separate queue
 - Volume of backlog:
 - 2.3M reports across all three queues Jan 2021. Peaked at 2.7M in April 2021.
 - Backlog was outpacing Meta's capacity to review because there continued to be incoming demand.
- Meta's efforts to reduce backlog

- Age verification taskforce kicked off in Feb 2021. Goal: address current backlog and prevent this problem going forward
- Review incoming jobs in 48 hrs
- Goal was to finish work in June 2021
- Levers:
 - Automations: Automated reviews of accounts. Prior to 2021 all review jobs were done by human review. May 17 2021 we had deployed 6 automations.
 - Reduced the backlog using these automations by 1.3M by May 17.
 - 6 automations are targeted toward different queues.
 - Auto closing jobs based on adult classifier in Edit DOB queue
 - Auto closed jobs based on adult classifier in U13 reporting queue
 - Account age automation for U13 reporting. When did the account get created, assuming they had created it X years before, they need to be O13.
 - Expirations of IDs for EditDOB in Appeals queue. When the ID is expired, reject.
 - Autoclear jobs for accounts with no photos in U13 reporting. Reviewer would have no signal to put them in a checkpoint.
 - Autocleared jobs that were previously cleared before.
 - From 0 automation, we saw a lift of 58% - 98% automation depending on the queue
 - Human Review improvements
 - Goal: increase efficiency of human review process
 - 5 changes:
 - Filtering photos that were pulled into the tooling for reviewer to see. Photos need to have humans in them to be useful to assessment. Filtered out photos without humans.
 - Added date information to photos more prominently
 - Photo zoom capability
 - Filtering profile content to only provide relevant information to make a decision on age determination.
 - Improved viewability of photos / overall UX
 - Increasing manual reviewer capacity

- With prior improvements, we can keep up with incoming demand but were not working on backlog at a fast enough pace
 - Requested additional capacity as well as a reprioritization of reviewers
 - \$2.5M increase in budget for these review flows to hire and train new reps
 - Brought in other reviewers from related queues at Meta to bring in to U13 review. You cannot always do this due to reviewer training, but if the reviewer has the right training.
- Results
 - May 17 2021, all of the levers were launched but there was still a 1.2M backlog
 - 453K in U13 enforcement queue
 - 63K in underage appeals queue
 - 733K in EditDOB queue
 - To address remaining backlog we implemented automations for the U13 backlog only as a one time measure and were not applied to new jobs coming in:
 - Auto ignore jobs with college or university degree signal. Account was not checkpointed. If account got another report, the account would be reviewed again and would not have used that automation.
 - Source of data: user-supplied. Or affiliation to a university or college group.
 - Auto ignore accounts with work experience signal.
 - Together these automations addressed 33% of the remaining backlog.
 - Edit DOB backlog:
 - 733K total - 718K from FB, 15K from IG
 - Auto reject entire remaining backlog of EditDOB requests. We believe it was less risky to require the users to not get age update granted. If they want to they needed to resubmit. This is lower risk than auto-accepting.
 - With these measures, by Aug 2021 the backlog was close to 0 on FB and IG. Existing supply of human reviewers then was able to manage incoming capacity.

D. SME: Mitali Paintal (soft matching background)

- Background

- Meta had a centralized team who owned soft matching
- Team used phone number and email to match across and within FB and IG
- Two kinds of emails and phone numbers:
 - Email and phone numbers that Meta collected directly - account creation, security (2FAC). Meta had phone number or email for all users because required during registration
 - Email and phone numbers where user did not provide directly but Meta inferred that it belonged to the user. For example, address book contacts could be imported by a user and Meta would then match a contact from her address book with her friend with the same name
 - FTC order 2019/2020: deprecated use of soft matching for ads purposes and more detailed disclosures. Highlighted uses like growth and integrity in the disclosures.
- Use cases
 - 1. Growth
 - Friend recommendations w/in and across apps
 - 2. Measurement
 - Meta began to report unique users in quarterly earnings. Using soft matching to dedupe users across apps.
 - 3. Ads
 - Matching contact points against advertiser consumer info - eg: Macy uploads a list of consumers and Meta targeted those ads. This is now deprecated.
 - 4. Integrity
 - Identifying people who were repeat offenders of Meta's TOS
- Accuracy
 - Meta measures accuracy of the inference. Different thresholds for different use cases - product team decides. If consequences were higher for inaccurate soft matching (eg: integrity use case) use higher precision. Not every use case used same accuracy.

E. SME: Laura Godlewski (soft matching background + analytics)

- Beginning of family metric reporting
 - Launched financial statements with family metric reporting in Q4 2019
 - In 2018, we were stating MAU and DAU metrics for FB alone in our financial reports. We wanted to be able to share a more expansive metric that demonstrated the whole family.