



Safety @ Meta

Dr Priyanka Bhalla, Head of Safety Policy APAC



Youth Safety

Our approach in action: Meta's multi-layered approach to youth

EXAMPLES

Active enforcement against nudity, Graphic violence, Self harm, Bullying or hate speech contents, No access to Dating for under 18 year-old users, etc.

EXAMPLES

Service Age requirements and collection, Content moderation public reports **EXAMPLES EXAMPLES** restrictions ('Community Standard Enforcement Underage reports, Report'), Youth Portal, Parents Guide, Age verification Menu of Options, Family Centre, Why am I Seeing This, Al age models, etc. Transparency **Understanding** Acces & Download your Information, etc. and Education user age Co-design and Meta's Best Interest of the Child Framework, Consultation **EXAMPLES** Design Guide for Youth, etc. for youth frameworks No alcohol, weight loss ads for under 18 year-old users, Safeguards, Ad restrictions No interest-based targeted ads Tools under 18 year-old users, No sensitive data from profile used **EXAMPLES** Privacy and safety by default settings, for targeted ads (for any user), etc. **EXAMPLES Parental** 'Take a break' wellbeing nudge, Making it supervision more difficult for adults to find and

Time setting, view accounts their teen is following/

blocking or teen's account privacy settings, etc.

with NCMEC, etc.

follow teens, 'Take it Down' partnership

Transparency Cente Бодлогууд Дүрэм хэрэгжүүлэх Аюулгүй байдал Онцлох зүйлс Засаглал

Эхлэл → Бодлогууд

Олон нийтийн дагаж мөрдөх хэм хэмжээ

Олон нийтийн дагаж мөрдөх хэм хэмжээнд Facebook, Instagram, Messenger, Threads дээр юуг зөвшөөрөх, үл зөвшөөрөхийг заасан

Оршил

Хүмүүс өдөр тутамдаа Facebook, Instagram, Messenger, Threads-ээр дамжуулан өөрт тохиолдсон үйл явдлаа бусдад түгээж, гэр бүлийнхэн, найз нөхөдтэйгөө холбоотой байж, олон нийтийн хүрээллийг бий болгож байдаг. Манай үйлчилгээ хэдэн тэрбум хүнд улс, соёлоор дамжин хэдэн арван хэлээр өөрийгөө чөлөөтэй илэрхийлэхэд тусалдаг.

Хүүхдийн бэлгийн

Transparency Cente Бодлогууд Дүрэм хэрэгжүүлэх Аюулгүй байдал Онцлох зүйлс Засаглал

мөлжлөг, хүчирхийлэл ба нүцгэн бие

Эхлэл 🗦 Бодлогууд 🤿 Олон нийтийн дагаж мөрдөх хэм хэмжээ

Бодлогын дэлгэрэнгүй хэрэглэгчийн интерфейс Мэдээлэл

Бодлогын дэлгэрэнгүй

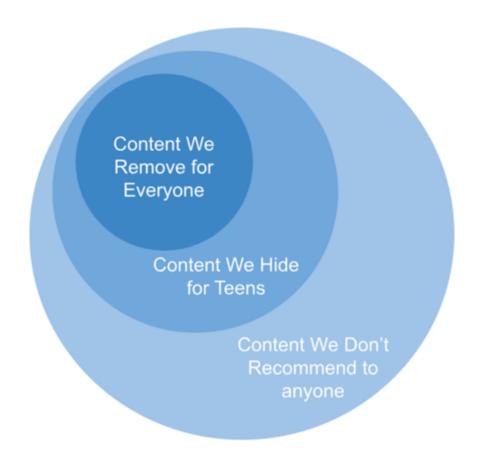
логийг өөрчлөх Өнөөдөр Одоогийн хувилбар 2025 оны 8-р сарын 1

Бодлогын үндэслэл

Бид хүүхдийн бэлгийн мөлжлөг эсвэл хүүхдэд аюул учруулахуйц контент эсвэл үйлдлийг зөвшөөрдөггүй. Бид хүүхдийн мөлжлөгийн талаар мэдмэгцээ холбогдох

Судалгааны хэрэгсэл

Хамгаалалтын олон үе давхарга



Х Үзүүлэхгүй устгах контент

Манай Олон нийтийн дагаж мөрдөх хэм хэмжээнд Facebook болон Instagram дээр үл зөвшөөрөх контентын төрлүүдийг заасан байдаг. Бид тэдгээр контентын талаар мэдсэн даруйд өсвөр үеийнхэн болон өөр ямар ч хүнд үзүүлэхгүйн тулд бүрэн устгадаг. Бид дүрэм зөрчсөн контентыг устгахдаа уунийг хуваалцсан бүртгэлд анхааруулга тавих бөгөөд манай бодлогыг давтан ноцтойгоор зөрчсөн бол уг бүртгэлийг идэвхгүй болгодог. Тэрхүү бодлого нь манай олон нийтийн сүлжээнд буй хүн бүрийг, үүнд өсвөр үеийнхнийг хамгаалах зорилготой.

Фер уеийнхэнд үзүүлэхгүй бусад контент

Зарим контент насанд хүрэгчдэд зохистой байж болох ч 18 нас хүрээгүй өсвөр үеийнхэнд хэт ахадсан байж болно. Бид мэргэжилтнүүдтэй хамтран өсвөр үеийнхэнд ямар төрлийн контент үзэхэд тохиромжгүй болох талаар улс орнуудад судалгаа явуулсан бөгөөд ийм төрлийн контентыг тэдэнд үзүүлдэггүй. Өөрөөр хэлбэл насанд хүрэгсэд тухайн контентыг үзэж болох ч 18 нас хүрээгүй өсвөр насны хүүхэд, түүний дагадаг бүртгэлээс хуваалцсан байсан ч, үзэх буюу харилцах боломжгүй.

Transparency Cente Бодлогууд Дүрэм хэрэгжүүлэх Аюулгүй байдал

Онцлох зүйлс Засаглал

Судалгааны хэрэгсэл

Эхлэл ightarrow Бодлогууд ightarrow Олон нийтийн дагаж мөрдөх хэм хэмжээ

Дээрэлхэх, дарамтлах

Бодлогын дэлгэрэнгүй хэрэглэгчийн интерфейс

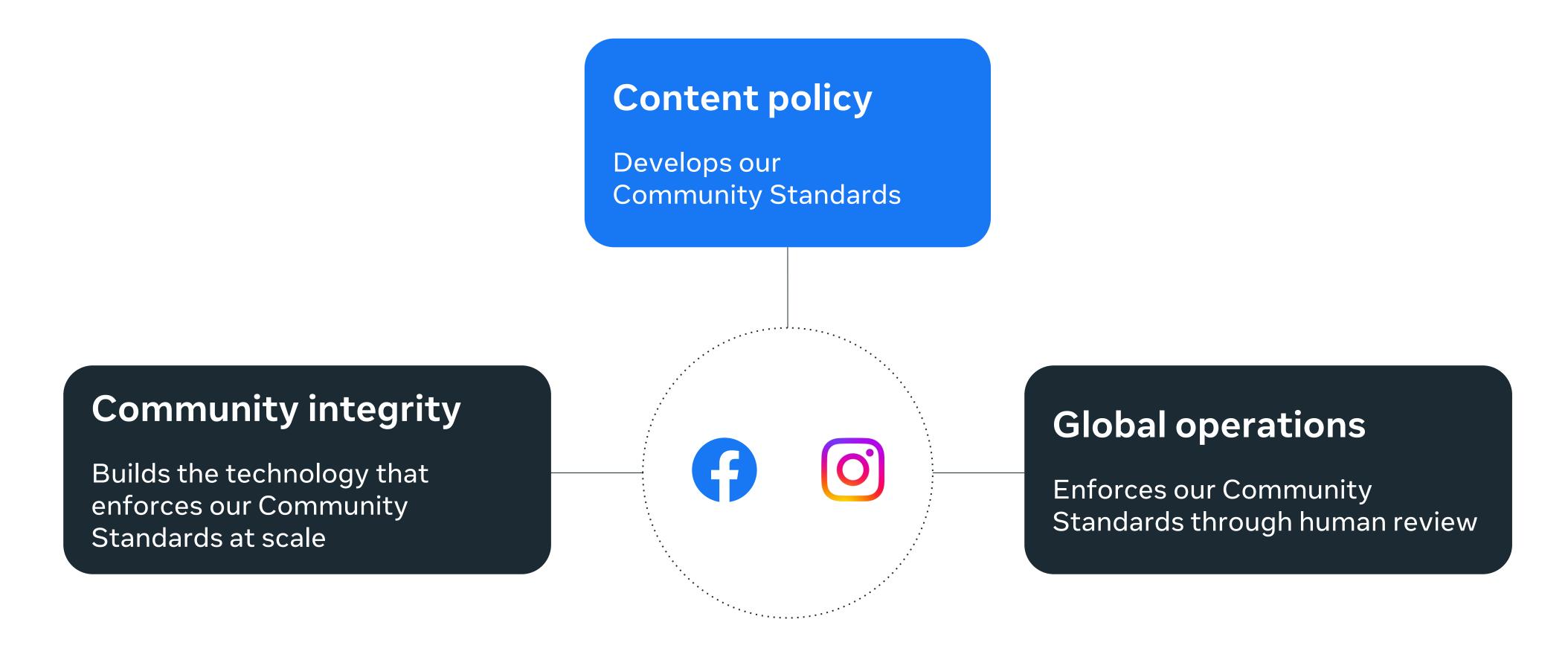
Бодлогын дэлгэрэнгүй

ЛОГИЙГ ӨӨРЧЛӨХ Өнөөдөр Одоогийн хувилбар 2025 оны 1-р сарын 23 2024 оны 7-р сарын 3 2023 оны 12-р сарын 14

Бодлогын үндэслэл

Дээрэлхэх, дарамтлах үзэгдэл нь олон газар тохиолддог бөгөөд заналхийлэх, хувь хүнийг таних боломжтой мэдээллийг задруулахаас эхлээд сүрдүүлсэн зурвас илгээх, хүсээгүй хорлонтой холбоо тогтоох гээд олон янзын хэлбэрээр илэрдэг. Ийм төрлийн зан төлөвөөс шалтгаалан хүмүүс Facebook, Instagram, Threads-д аюулгүй байх, хүндлүүлэх боломжгүй байх тул бид үүнийг хүлээн зөвшөөрдөггүй.

Content review: a partnership of multiple teams



Understanding age

Designing an ongoing, multi-layered approach to understanding user age



Age collected at registration

DOB requested at account registration through an age-neutral process with technical restrictions to make it harder for users to provide false information.



Content review

We also train content reviewers to remove accounts that appear to belong to U13s.



Community reporting

Anyone can report suspected underage accounts on IG, FB, and in Quest and we have dedicated channels to review these reports.



Machine learning to detect likely teens

We're investing in age prediction models to detect likely teens and ensure they receive ageappropriate experiences.



Educating parents

We remind parents of the minimum age in the Instagram Parents' Guide and our Parent Education Hub in VR and on IG.



Menu of options to verify age

We're testing additional options beyond ID upload for users to verify their age.

Why is age verification and parental approval at OS/App Stores level the best solution?

01

It's easier for parents. For example, in the US, the average teenager uses 44 applications on their phones. We believe that a significant step forward can be taken at a European level to ensure that parents only need to verify the age of their child once - and that their child will then be placed into an age-appropriate experience on every single app.

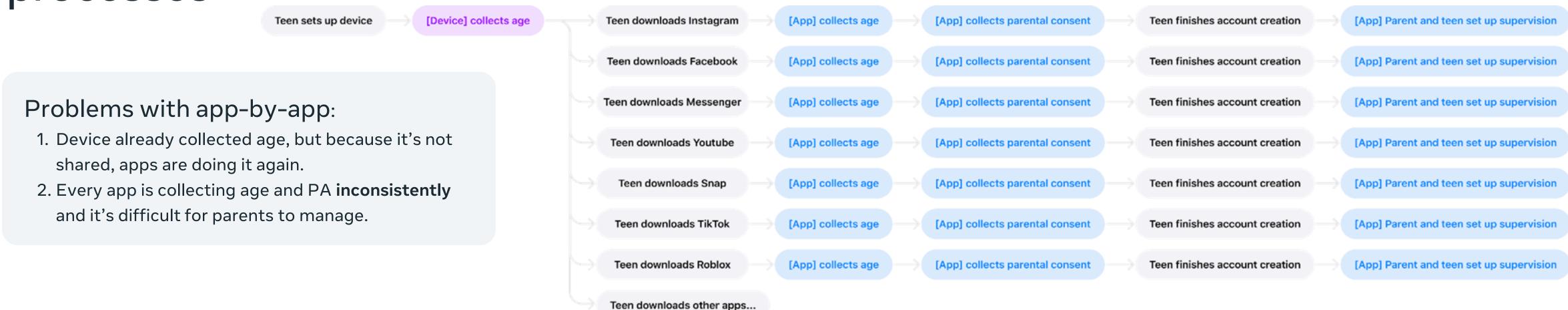
02

It's more privacy protective. Users can confirm their age, familial relationship and parental approval once, then decide from one place (at app store/OS level) which apps do and do not get this data. App-by-app parental approval and age verification requires parents and teens to provide every app with a parent and teen's personal information, increasing data privacy and security risks. Parents have concerns about this risk and three-in-four (76%) trust app stores over individual apps to securely handle the personal data needed to verify parental approval.

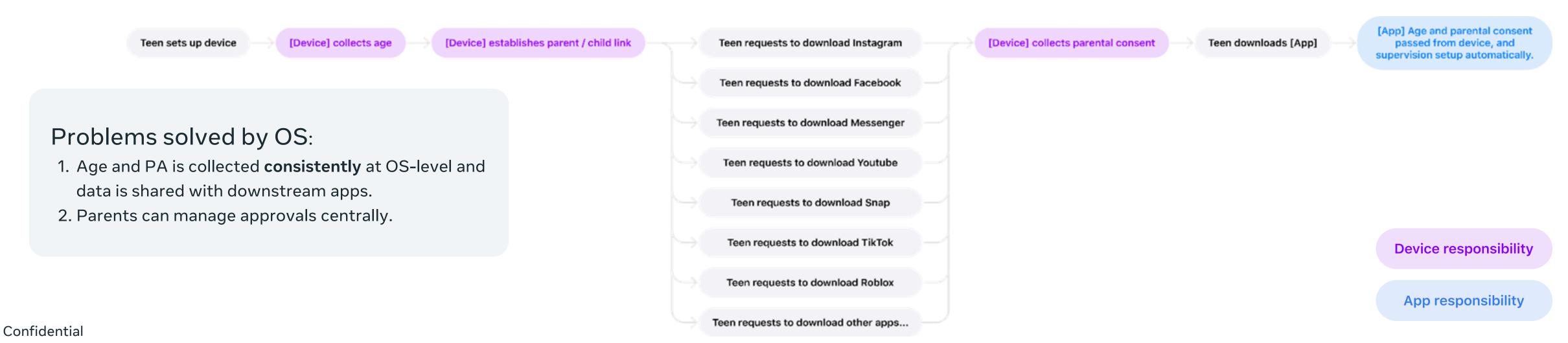
03

The capability already exists. Google/Apple already know the stated age of users and the apps they use, and have parental permission and approval built into the download step. Small or fast-growing apps popular with teens are unlikely to build out their own similar, reliable systems

Current app-by-app experience is high-friction for parental approval processes



OS/App store experience is much simpler for parental approval processes



Instagram Teen Accounts



Why Teen Accounts?

Addresses the largest concerns reported by parents - globally - around their teen's use of social media apps:

- Contact
- Content
- Screen time

Built-in protections:

- Limit who can contact teens and the content they see
- Teens under 16 won't be able to change any of these settings to be less strict without getting a parent's permission

5:26



Instagram













Your story

mwells77

aimi.allover



An experience made just for you

Soon, your settings will be updated automatically to help you explore what you're into and protect against unwanted contact.

Learn more















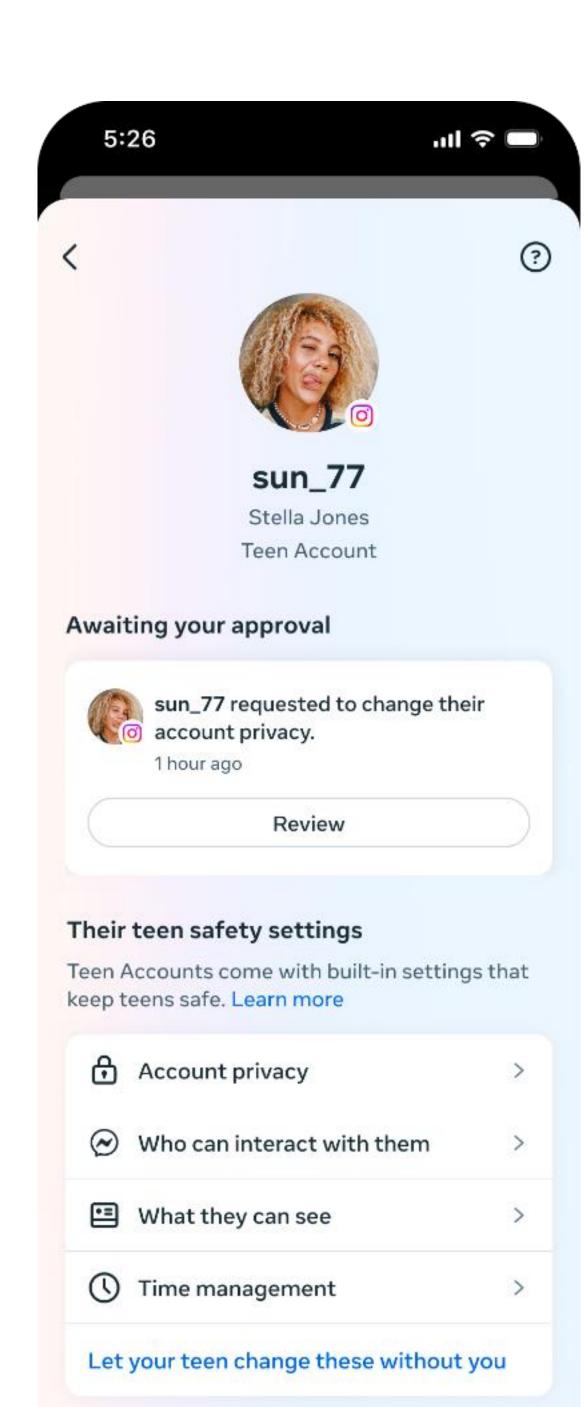
Built-in Protections

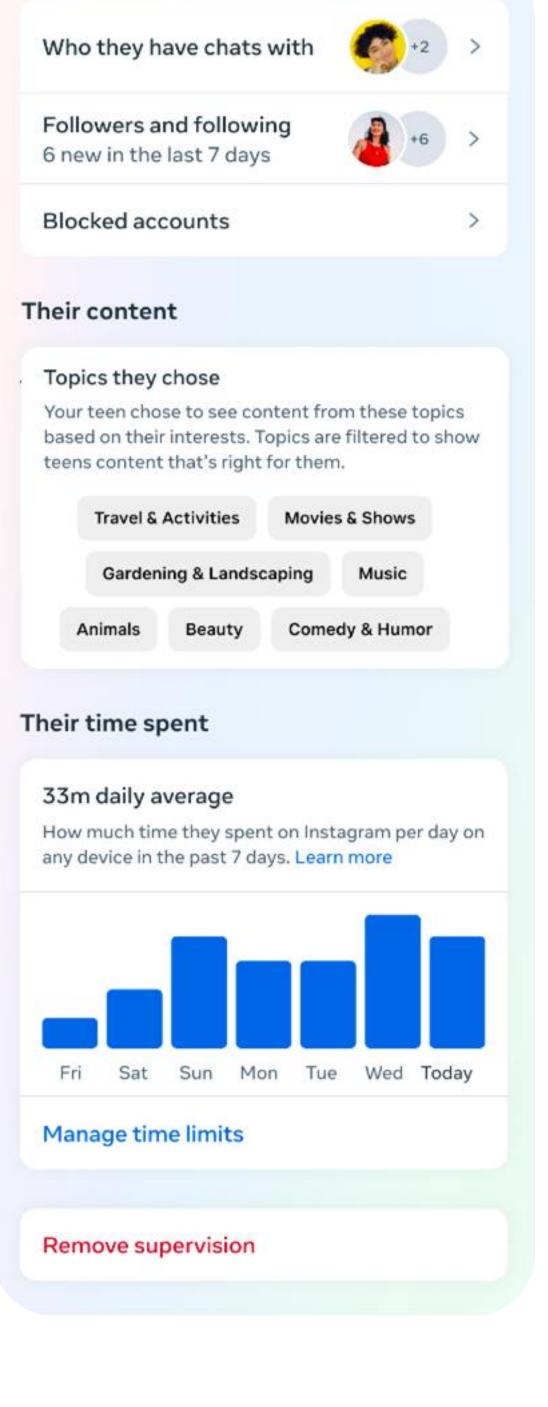
Control or Feature	New default state	Ability to edit
Privacy	Default to private upon sign-up	13-15: Parent approval needed to change 16-17: Teen can change on their own
Contact	Can only be messaged by people who they follow, or have messaged before	
Content	Sensitive Content Control set to Less	
Bullying & Harassment	Hidden words on Tags/Mentions/Reels Remixes restricted to people you follow	
Time	Sleep mode (10PM-7AM)	
	Daily limit reminder 60 min	

13-17 year olds, if supervised will need parental approval to change the above settings. Parents can place supervised teens into blocking hours.

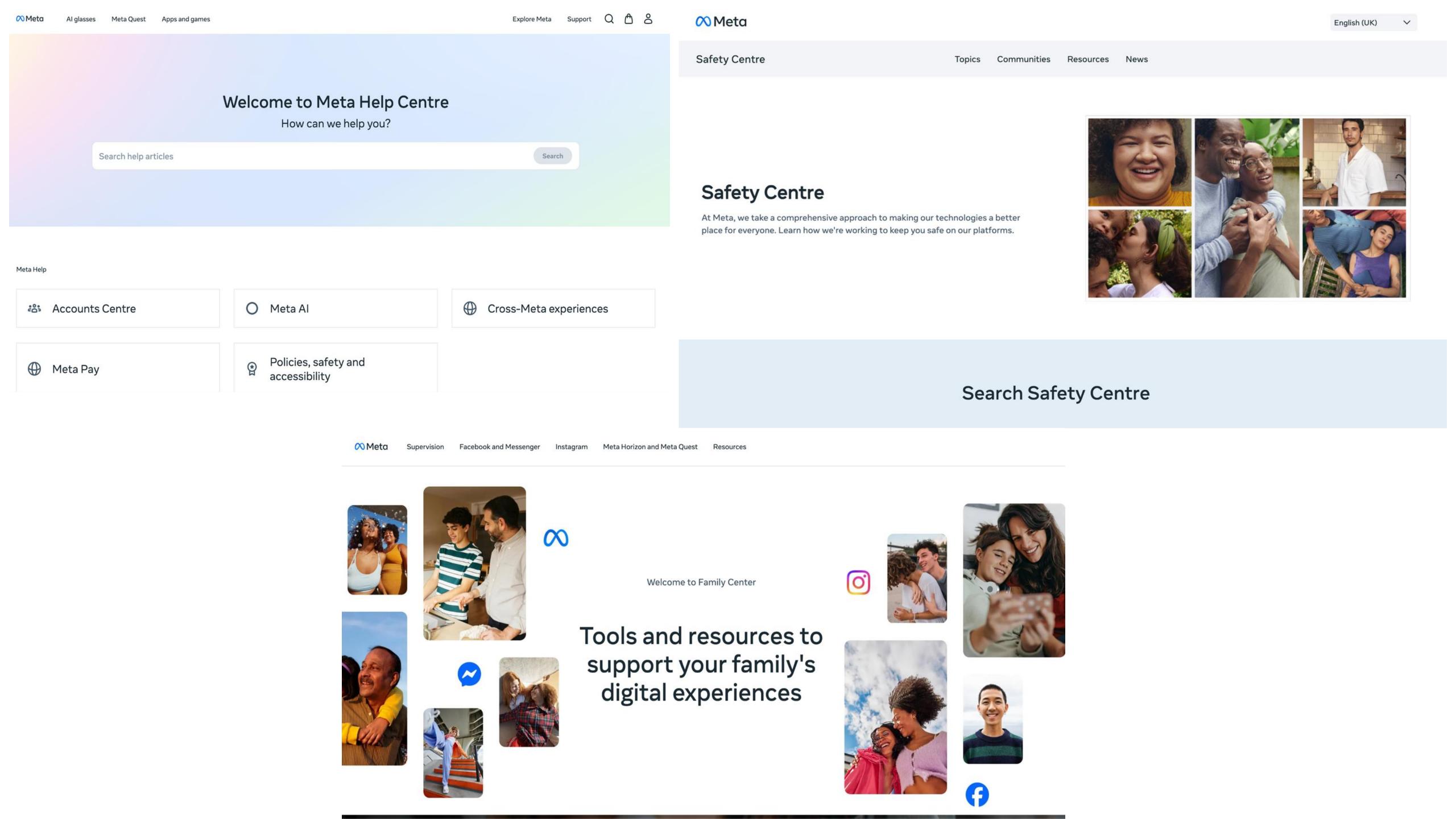
Supervision dashboard for parents

Once supervision is set up, parents can see their supervised teen's settings and insights such as time spent, followers and following, and blocked accounts in the supervision dashboard.





Resources



StopNCII.org

Owned & Operated by:



Industry Partners

Companies who will receive cases and hashes from StopNCII.org:













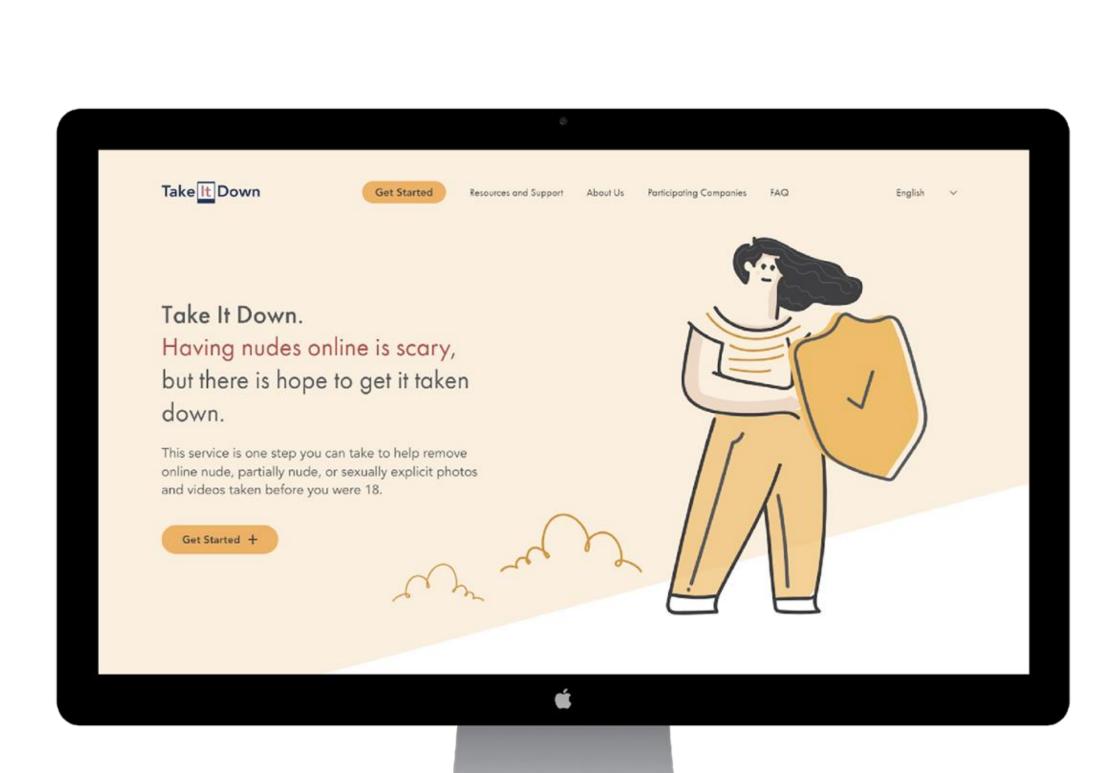


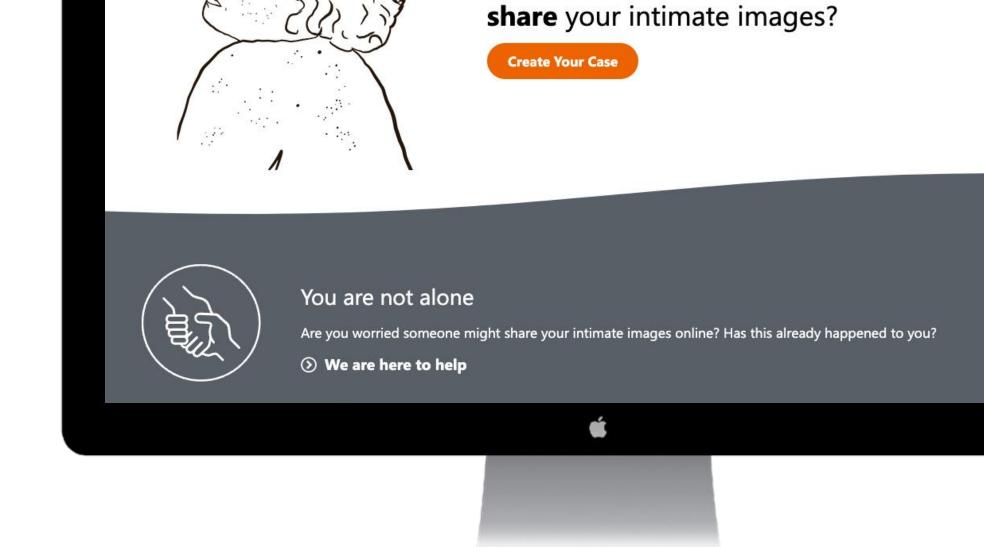
Snap Inc.



playhouse

REDGIFS





TakeItDown.ncmec.org

Owned & Operated by:









What do you do if someone is threatening to









Snap Inc.







Person creates a private case - images never leave their device

StopNCII.org
generates unique
hash ("digital
fingerprint")

StopNCII.org
shares the hash
with tech
companies

Tech companies find and remove matches of the hash on platform

Questions?