Enabling Developers, Protecting Users:
Investigating Harassment and Safety in VR

Abhinaya S B, Aafaq Sabir, Anupam Das

North Carolina State University

What is Virtual Reality?

What is Virtual Reality?

360-degree virtual experiences

Synchronous voice chat

Social VR
Gaming VR
Streaming VR

What is Virtual Reality?

360-degree virtual experiences

Synchronous voice chat

Social VR
Gaming VR
Streaming VR

Haptic feedback

Full-body tracking

Immersive experiences

Harassment in VR

My First Virtual Reality Groping

Jordan Belamire · Follow

Published in Athena Talks · 5 min read · Oct 20, 2016

Listen

👏 2.5K   💬 257

Harassment in VR

Harassment in VR

My First Virtual Reality Groping

MIT Technology Review

Featured Topics Newsletters Events Podcasts

SIGN IN   SUBSCRIBE

HUMANS AN...   2.5K

The m...
alread...

A woman was s...
the first—and v...

By Tanya Basu

CNN BUSINESS   Markets Tech Media Calculators Videos

Harassment is a problem in VR, and it's likely to get worse

By Rachel Metz, CNN Business
Updated 10:01 PM EDT, Thu May 5, 2022

Harassment in VR

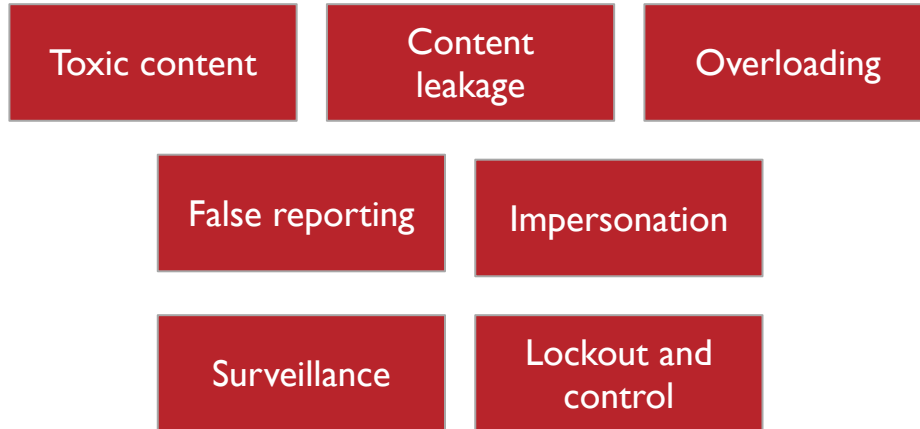## Virtual Reality Promised us a New World. Instead, It's Become a Breeding Ground for Harassment.

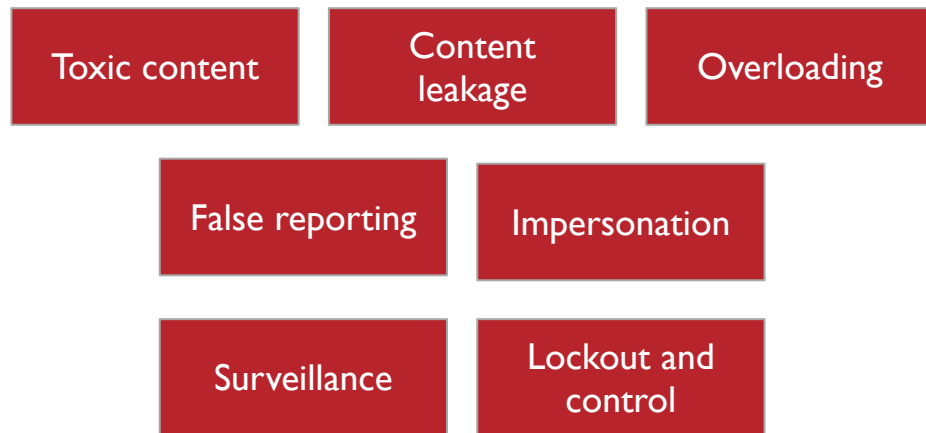*Misogyny is alive and well in the metaverse.*

BY MOIRA DONEGAN   **PUBLISHED: APR 13, 2023**

Harassment in VR: What is new?

Toxic content	Content leakage	Overloading

False reporting	Impersonation

Surveillance	Lockout and control

Taxonomy of online hate and harassment attacks [1]

[1] K. Thomas, D. Akhawe, M. Bailey, D. Boneh, E. Bursztein, S. Consolvo, N. Dell, Z. Durumeric, P. G. Kelley, D. Kumar et al., "SoK: Hate, harassment, and the changing landscape of online abuse," in 2021 IEEE Symposium on Security and Privacy (SP). IEEE, 2021, pp. 247–267.

Harassment in VR: What is new?

Toxic content

Content leakage

Overloading

False reporting

Impersonation

Surveillance

Lockout and control

Taxonomy of online hate and harassment attacks [1]

Embodiment, Immersion and Haptic feedback
Amplify
Virtual violence, Virtual sexual harassment

[1] K. Thomas, D. Akhawe, M. Bailey, D. Boneh, E. Bursztein, S. Consolvo, N. Dell, Z. Durumeric, P. G. Kelley, D. Kumar et al., "SoK: Hate, harassment, and the changing landscape of online abuse," in 2021 IEEE Symposium on Security and Privacy (SP). IEEE, 2021, pp. 247–267.

Safety Controls in VR

Safety Control	Function
Mute	Disable voice chat of self, or other users in a VR space
Block	Hide or change the appearance of user(s) in a VR space
Proximity setting	Control the distance at which other users can interact with a user in a VR space
Quick travel	Travel to a different location within a VR app
Safe zone	A user's private space accessible only to that user
Vote kick	Kick a user out of a VR space based on majority vote
Trust rank	Levels of trust assigned to a user

Table 1: Prominent safety controls available in VR apps (from our paper)

Multi-perspective study

Semi-structured interviews with:
- Targets of VR-based harassment

Multi-perspective study

Semi-structured interviews with:

- Targets of VR-based harassment

- VR developers

Study I

Consultation with psychology expert

Design of screening & interview protocol

- Targets of VR-based harassment
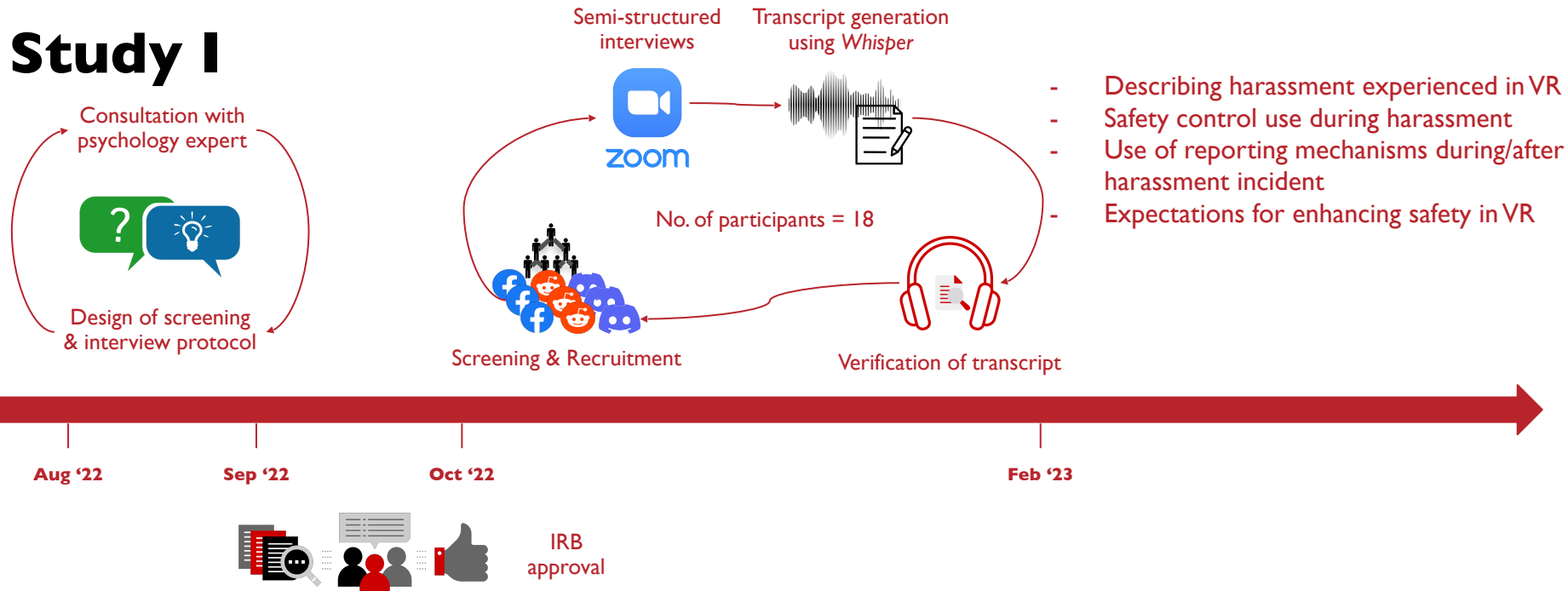- Excluded those with PTSD / emotional distress
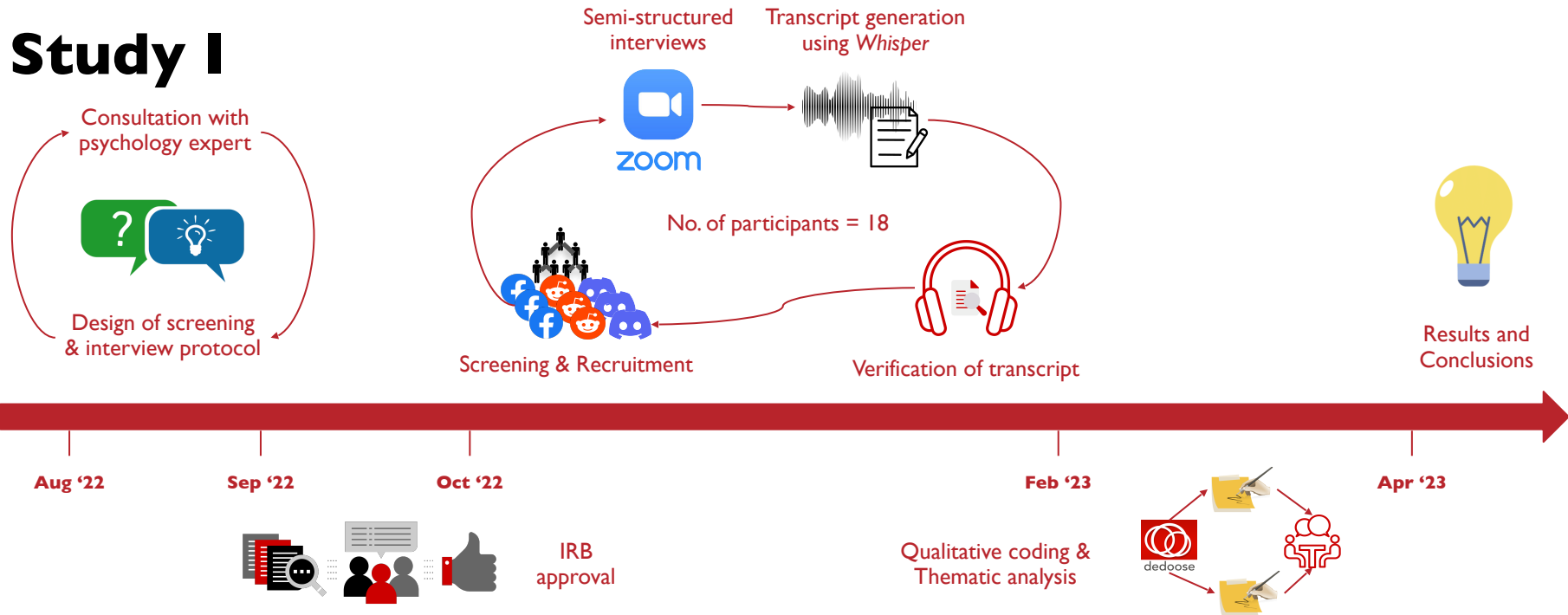- Rephrased sensitive questions

Aug '22

Population: Targets of VR-based harassment
RQ1: How do targets of VR-based harassment perceive the *usability and effectiveness* of existing safety controls and reporting mechanisms?
RQ2: What are the *expectations and recommendations* by targets of VR-based harassment for making VR safer?

Study I

Consultation with psychology expert

Design of screening & interview protocol

Semi-structured interviews

Transcript generation using *Whisper*

- Describing harassment experienced in VR
- Safety control use during harassment
- Use of reporting mechanisms during/after harassment incident
- Expectations for enhancing safety in VR

No. of participants = 18

Screening & Recruitment

Verification of transcript
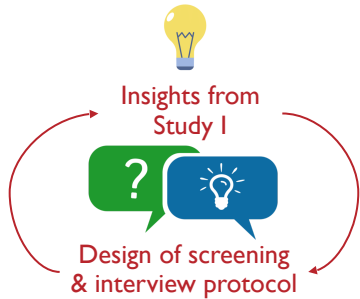
Aug '22   Sep '22   Oct '22   Feb '23

IRB approval

Population: Targets of VR-based harassment
RQ1: How do targets of VR-based harassment perceive the *usability and effectiveness* of existing safety controls and reporting mechanisms?
RQ2: What are the *expectations and recommendations* by targets of VR-based harassment for making VR safer?

7

Study I

Consultation with psychology expert

Design of screening & interview protocol

Semi-structured interviews

Transcript generation using *Whisper*

No. of participants = 18

Screening & Recruitment

Verification of transcript

Results and Conclusions

Aug '22   Sep '22   Oct '22   Feb '23   Apr '23

IRB approval

Qualitative coding & Thematic analysis

dedoose

Population: Targets of VR-based harassment

RQ1: How do targets of VR-based harassment perceive the *usability and effectiveness* of existing safety controls and reporting mechanisms?

RQ2: What are the *expectations and recommendations* by targets of VR-based harassment for making VR safer?

7

Study II

Insights from Study I

Design of screening & interview protocol
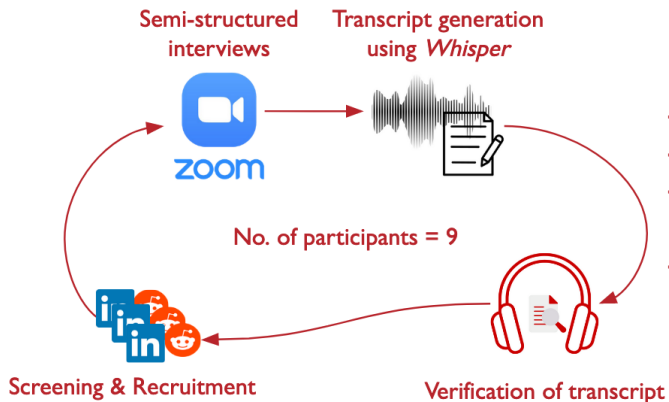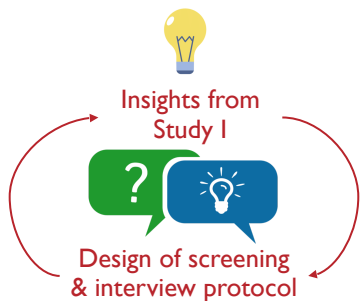
VR developers including
- UI/UX designer
- XR gameplay and tools engineer
- AR/VR maintenance and support

Apr '23

Jul '23

Population: VR Developers

RQ3: What are VR developers' perceptions of the *design and deployment of safety controls*?

8

Study II

Insights from Study I

Design of screening & interview protocol

Semi-structured interviews

Transcript generation using *Whisper*

No. of participants = 9

Screening & Recruitment

Verification of transcript

- VR development experience
- Solving usability challenges
- Challenges in implementing VR safety features
- Feasibility of user recommended features

Apr '23 Jul '23 Aug '23 Sep '23 Oct '23

IRB approval

Population: VR Developers
RQ3: What are VR developers' perceptions of the *design and deployment of safety controls*?

8

Study II

Insights from Study I

Design of screening & interview protocol

Semi-structured interviews

Transcript generation using *Whisper*

No. of participants = 9

Screening & Recruitment

Verification of transcript

Results and Conclusions

Apr '23   Jul '23   Aug '23   Sep '23   Oct '23

IRB approval

Qualitative coding & Thematic analysis

Population: VR Developers

RQ3: What are VR developers' perceptions of the *design and deployment of safety controls*?

8
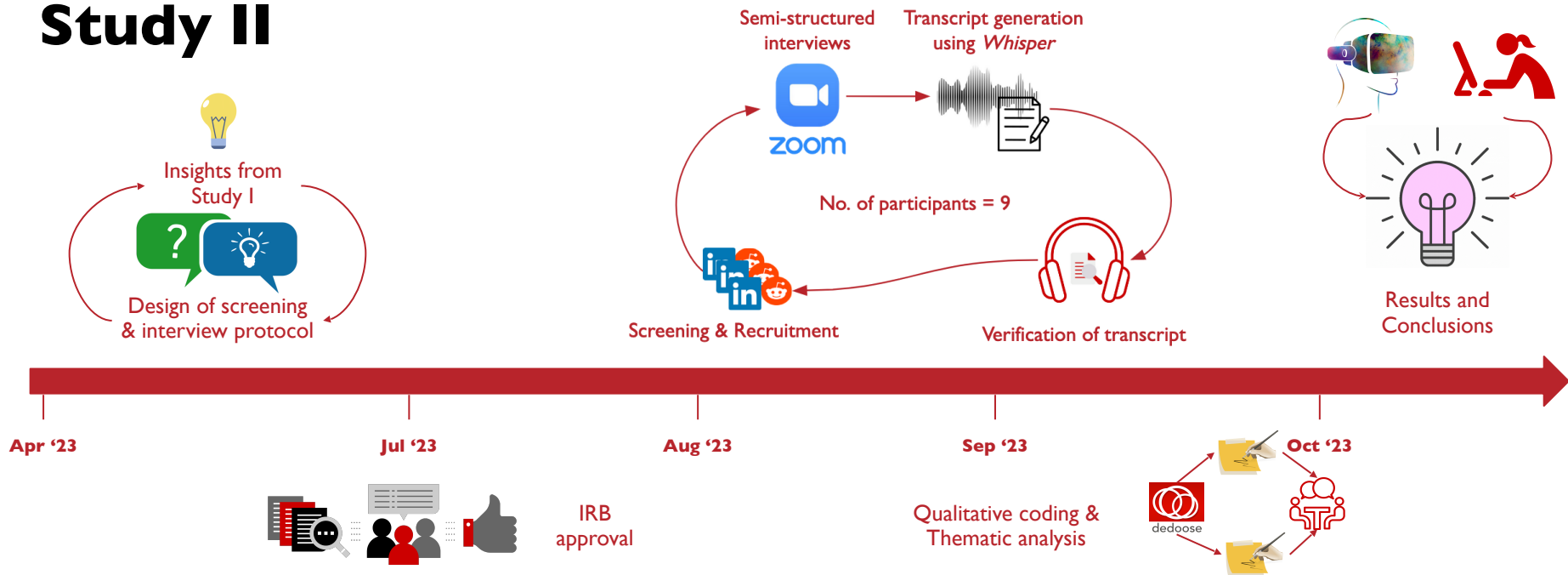
Results: Study 1

18 participants

RQ1: Perceptions on Safety Controls' Effectiveness

Safety controls are *ineffective* in many ways:
- Fail to provide feedback to harassers upon muting or blocking
- Do not remove the harasser from the game; merely change their appearance
- Affect communication with non-harassers
- Can be misused to cause further harassment
- Fail to stop further instances of harassment

RQ1: Perceptions on Ease of Reporting

Reporting is *time-consuming* and *cumbersome* if
- Typing through a keyboard in VR
- New user
- Video evidence is required

" *Each report takes 15 minutes if you're doing it properly. You have to get off the game, go through your footage, do a small edit of it, and write out the email. (Echo VR)* **"**

RQ2: **Expectations** for Safer VR

- In-app interventions to inform users of harassers in nearby VR spaces
- Age-based user segregation
- Live moderation
- Automatic detection of harassment
- Automatic detection of distress in users
- Tracking users' toxicity histories across VR apps

Results: Study II

9 participants

UI/UX designer
XR Gameplay & Tools Engineer
Researchers

RQ3: **Feasibility** of Users' Expectations

- Live moderation
    - Sustainable only through moderators embedded in the community
    - Augmented by automated abuse detection

# RQ3: **Feasibility** of Users' Expectations

- Live moderation
  - Sustainable only through moderators embedded in the community
  - Augmented by automated abuse detection
- Tracking users
  - Limited to a single platform (e.g., Meta, Steam)
  - Legal risks to VR companies due to infringement of users' privacy

RQ3: **Feasibility** of Users' Expectations

- Live moderation
    - Sustainable only through moderators embedded in the community
    - Augmented by automated abuse detection
- Tracking users
    - Limited to a single platform (e.g., Meta, Steam)
    - Legal risks to VR companies due to infringement of users' privacy
- Detecting distress in users
    - Will VR systems have sufficient input?
    - False positives may ruin user experience

RQ3: **Feasibility** of Users' Expectations

- Live moderation
    - Sustainable only through moderators embedded in the community
    - Augmented by automated abuse detection
- Tracking users
    - Limited to a single platform (e.g., Meta, Steam)
    - Legal risks to VR companies due to infringement of users' privacy
- Detecting distress in users
    - Will VR systems have sufficient input?
    - False positives may ruin user experience
- Identity checks
    - Non-invasive third-party verification

RQ3: **Challenges** in Designing Safety Controls

VR safety is not prioritized due to:

- Lack of *financial incentives* for VR companies
- High development costs; unstable economy
- Lack of *awareness* about safety risks in VR
- Lack of *legal or technical guidelines* for safety design
- Challenges in user testing of the safety controls
- Challenges in achieving a trade-off between *privacy* and *safety*
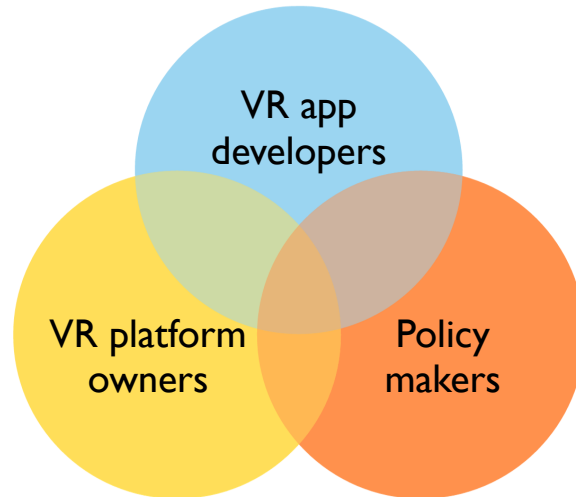
Recommendations
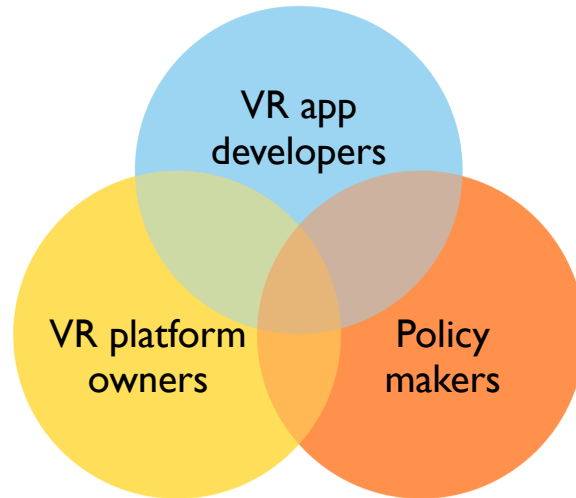
Recommendations
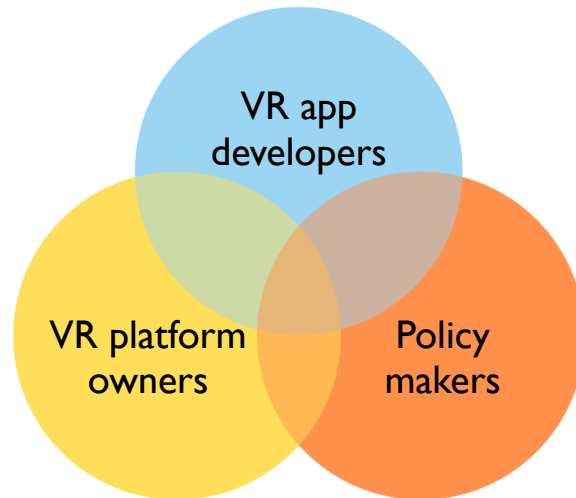
Develop open-source
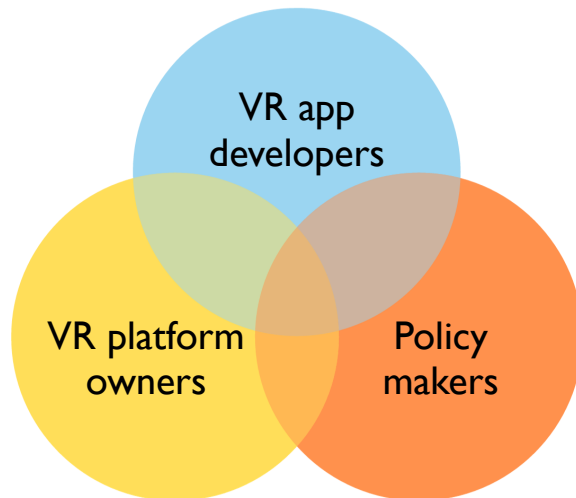safety libraries integrated
into game engines

Recommendations

Develop open-source safety libraries integrated into game engines

VR app developers

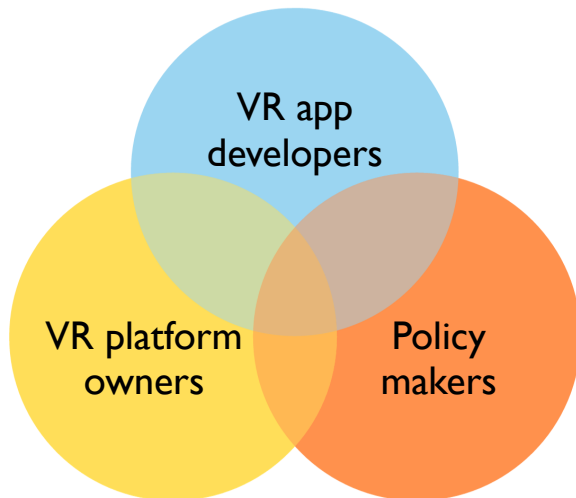Standardize baseline safety controls and reporting mechanisms across VR

VR platform owners

Policy makers

16

Recommendations

Develop open-source
safety libraries integrated
into game engines

VR app
developers

Standardize baseline safety
controls and reporting
mechanisms across VR

VR platform
owners

Policy
makers

Develop regulations for code
of conduct and moderation in
VR spaces

Recommendations

Develop open-source safety libraries integrated into game engines

Inform VR users about safety controls

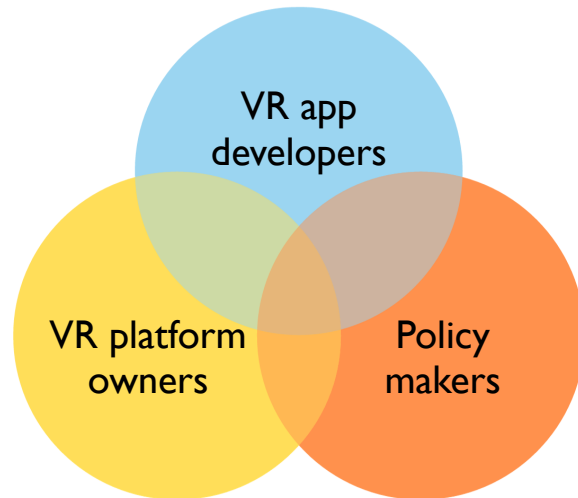Standardize baseline safety controls and reporting mechanisms across VR

Develop regulations for code of conduct and moderation in VR spaces

VR app developers

VR platform owners

Policy makers

Recommendations

Develop open-source safety libraries integrated into game engines

Inform VR users about safety controls

Standardize baseline safety controls and reporting mechanisms across VR

VR app developers

VR platform owners

Policy makers

Develop regulations for code of conduct and moderation in VR spaces

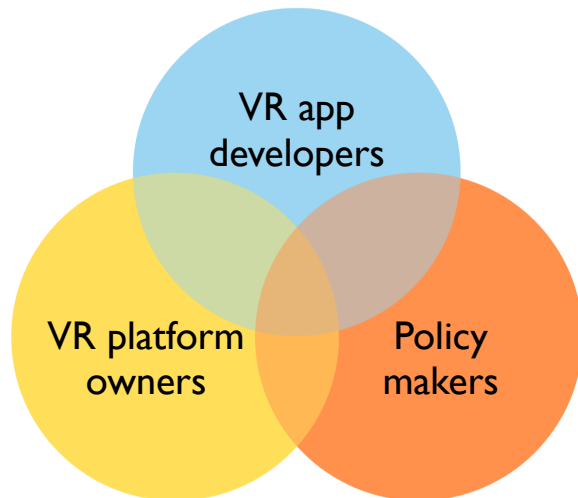Develop technical and legal guidelines mandating standards for VR safety

Recommendations

Develop open-source safety libraries integrated into game engines

Inform VR users about safety controls

Identify ethical ways for testing safety controls

VR app developers

Standardize baseline safety controls and reporting mechanisms across VR

VR platform owners

Policy makers

Develop regulations for code of conduct and moderation in VR spaces

Develop technical and legal guidelines mandating standards for VR safety

16

Thank You

Takeaways:

- Multi-perspective study on VR safety (n = 27) with:
  - *Targets* of VR-based harassment (n = 18)
  - VR *developers (n = 9)*
- Identified contexts where existing VR safety controls and moderation practices are *non-usable* and *ineffective*
- Contrasted VR users' *expectations* for safer VR with VR developers' perceived technical, legal, and financial *challenges*
- Made *recommendations* to VR platform owners, VR app developers, and policy makers for *improving safety in VR*

The Team
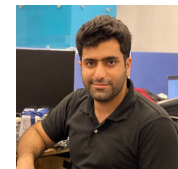
Abhinaya S B   Aafaq Sabir   Anupam Das

Email: asrivid@ncsu.edu

Read our paper   Access our study materials