A Trust and Safety Analysis of the Metaverse

Aditya Saligrama

What is the metaverse?

What is the metaverse?

Commonalities:

- Virtual reality

- Digital economy

What is Trust and Safety (T&S)?

"The part of a digital service's operations that focuses on understanding and addressing the harmful content or conduct associated with that service."

(Digital Trust and Safety Partnership, 2021)

Traditional T&S challenges

For social media platforms:

- Spam

- Abuse

  - Harassment

  - Scams

- Advertising privacy issues

T&S challenges on the metaverse

- Content issues
  - Same as before

T&S challenges on the metaverse

- Content issues

  - Same as before

- User conduct in virtual reality

T&S challenges on the metaverse

- Content issues

  - Same as before

- User conduct in virtual reality

- Exploitation in the digital economy

T&S challenges on the metaverse

- Content issues

    - Same as before

- User conduct in virtual reality

- Exploitation in the digital economy

- ...and more

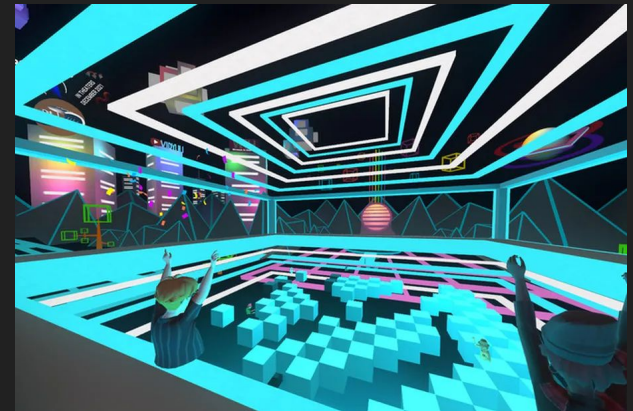# Content Trust and Safety Issues

Advertising on the Metaverse

Trust and Safety in Virtual Reality

Content Warning on following slides

Discussion of sexual harassment
Discussion of CSAM

Why does VR enable harassment?

- Physical interactions simulated online

- No verification of user identities

- Allows malicious actions without real-world repercussions

VR T&S back then: 2018

On VRChat:

- Frequent gendered harassment

- Racially problematic characters

VR T&S back then: 2018

VRChat's reaction:

- Awareness of problem

- Existing tools insufficient

  - Mute or block other users

Promise: build new tools to improve UX

VR T&S now: 2022

Still on VRChat:

- No restrictions on child accounts

- Children asked to perform sex acts in VR

- Insufficient moderation

VR T&S now: 2022

Microsoft shuts down AltspaceVR's social hubs to combat harassment

Igor Bonifacic   9:35 AM PST • February 16, 2022   Comment

THE INTERNET | FEB. 4, 2022

Mark Zuckerberg's Metaverse Has a Groping Problem

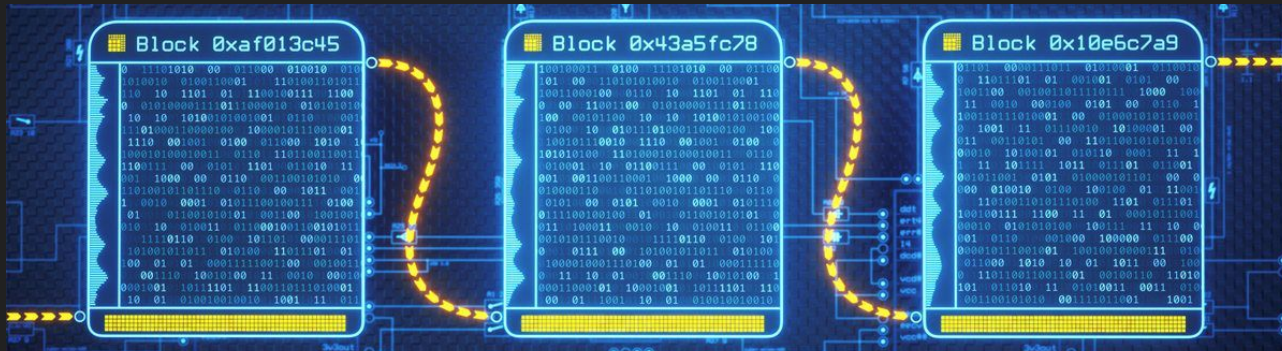By Matt Stieb

An Exploitative Digital Economy

A crash course on cryptocurrency and NFTs

Cryptocurrency:

- Decentralized currency

- No trust in central authority

Used for transactions:

- Secure

- Auditable

A crash course on cryptocurrency and NFTs

Non-Fungible Token (NFT): application of cryptocurrency

- For proving or transferring ownership of items

- "Receipt" associated with an item

Play-to-earn games: Axie Infinity

- Upfront investment in character tokens (NFTs)
  - $$$$!
- Win games => win cryptocurrency

Axie Infinity

Axie Infinity

MOTHERBOARD TECH BY VICE

The Metaverse Has Bosses Too. Meet the 'Managers' of Axie Infinity

MOTHERBOARD TECH BY VICE

The Digital Colonizers of 'Axie Infinity'

Axie Infinity

Axie Infinity Daily Active Users Drop 45% Since 2021 Peak

By David Thomas
5 April 2022, 04:09 GMT+0000

Axie Infinity players are leaving en masse after $600M crypto heist

Thoughts on fixing metaverse T&S

Rebuilding user trust in the metaverse

1. Restrictions on data advertisers can use

Rebuilding user trust in the metaverse

1. Restrictions on data advertisers can use

2. Employ more moderators

   a. Research on automatically detecting user harm in VR

   b. Informal moderation training for users?

Rebuilding user trust in the metaverse

1. Restrictions on data advertisers can use

2. Employ more moderators

    a. Research on automatically detecting user harm in VR

    b. Informal moderation training for users?

3. Reconsider necessity of cryptocurrency to application

    a. Improve crypto security

Questions?