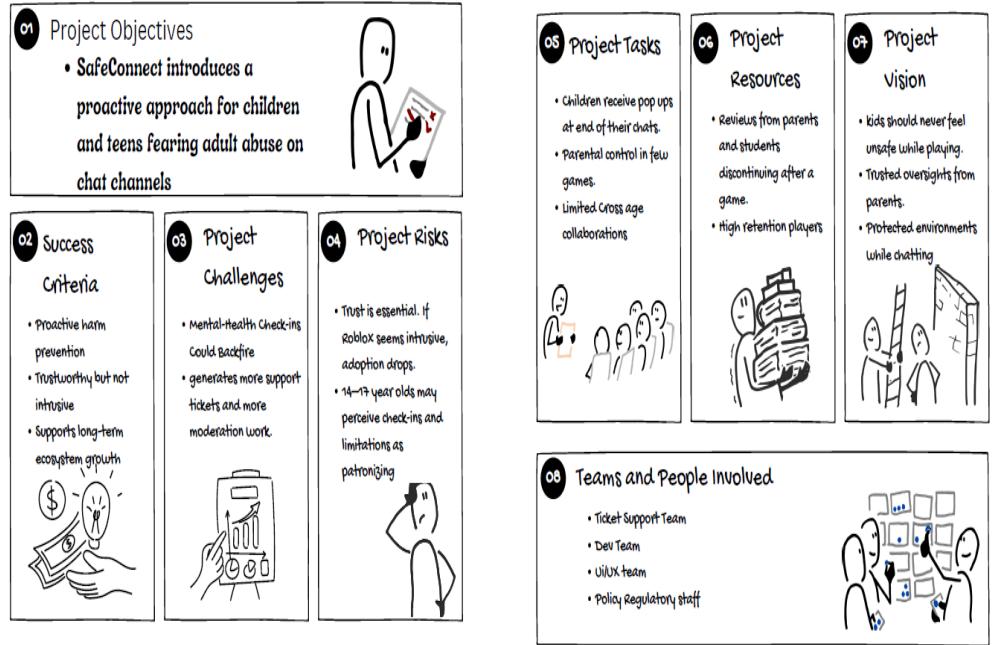


Project Plan

SafeConnect: Designing Safe, Confident, and Empowered Play on Roblox



Media Excerpts!!

A Growing Disconnect Between Messaging and Reality

If the AI places someone into the wrong age group, users over 13 can upload an ID to correct their age. Parents whose accounts are linked to their child's via Roblox's parental control tools can also correct their child's age. (ID checks are available only to users older than 13, likely because online privacy regulations are more stringent for younger children.)

Roblox says the images of users' faces will be used only for age estimation, and they'll be deleted after users are placed in an age bucket.

When asked why Roblox prioritizes open-ended communication among users, Baszucki emphasized the platform's social benefits. "There's a lot of kids who find their peeps on Roblox. There are a lot of kids [who play] when they're lonely or isolated." He claimed that famous parents have told him, "My kid would be dead if it were not for finding a community on a platform like Roblox to stay connected with."

But the mother of **Ethan Dallas** would strongly disagree. Ethan, who also felt lonely and isolated, looked for friends on Roblox and instead met a dangerous predator. Ethan tragically died by suicide after being groomed and sexually exploited by that predator.

Case Study on Roblox's Child Safety regulations and enhancements

A product initiative that protects children, reassures parents, and strengthens teen creativity — by reimagining how people connect on Roblox.

Most Roblox safety systems are **reactive**: age estimations, chat filters, post-incident moderation. But grooming and unsafe contact occur **before** detection; in the first DMs, the first friend request, the first casual conversation.

SafeConnect introduces a *proactive* approach focused on **three users**:

1 Children (8–13)

They need frictionless fun *without* unsafe contacts.

2 Teens & Young Creators (14–22)

They want a professional, collaborative, low-risk environment that helps them grow.

3 Parents

They want clarity, trust, and control — without needing to monitor every second.

SafeConnect improves life for all three groups.

✓ Overall Positive Impact

1. Increased emotional safety

Kids feel safer joining games or chats.

Teens feel less watched, more protected.

Parents feel less anxious.

2. Higher retention, healthier play

When users feel *safe*, they stay longer.

More positive interactions → stronger community.

3. Confidence for young creators

Teens & young adults build without fear of harassment.

This grows Roblox's professional creator ecosystem.

4. Reduced predatory risk

By cutting off cross-age funnels and adding check-ins, grooming risk is dramatically lowered.

5. Trust restored in Roblox

SafeConnect makes Roblox one of the most parent-trusted platforms in gaming.

Risks & Challenges of SafeConnect System

❖ Over-Restriction → Harm to User Experience

Risk:

Age-dynamic restrictions could block genuine friendships between kids of different ages (e.g., siblings, cousins, school seniors helping juniors).

Why it matters:

If it feels too strict, kids may quit or work around the system.

Mitigation:

- Allow *trusted contacts* approved by parents

❖ Privacy Concerns Around Age Detection

Risk:

Users may worry about face scanning, age verification, or data tracking.

Why it matters:

Trust is essential. If Roblox seems intrusive, adoption drops.

Mitigation:

- Use *behavioral age signals*, not biometrics
- Make parental settings optional, not forced
- Clear communication on how data is used

❖ Teens Feeling “Controlled” or “Babied”

Risk:

14–17 year olds may perceive check-ins and limitations as patronizing.

Why it matters:

Teen creators generate a big part of Roblox content; alienating them hurts growth.

Mitigation:

- Gradual reduction of warnings as age increases

❖ False Positives in Grooming Detection

Risk:

AI might flag normal conversation as “risky,” causing frustration and mistrust.

Why it matters:

High false positives → lower trust in the platform → backlash from community.

Mitigation:

- AI tuned with human review loops
- Clear “Undo / Dismiss” options
- Explainable reasons (“This message looked unusual because...”)

❖ Moderation Load Increases

Risk:

More nudges, alerts, and parental approvals generate *more support tickets* and *more moderation work*.

Why it matters:

Roblox's moderation team is already overloaded.

Mitigation:

- Automate routing and triage
- Prioritize high-risk cases

❖ Mental-Health Check-ins Could Backfire

Risk:

Some kids may get confused or scared by questions (“Is someone bothering you?”). Some teens may dismiss them instantly or find them annoying.

Why it matters:

This feature needs careful, gentle UX.

Mitigation:

- Use soft, child-friendly language
- Limit frequency to avoid spam
- Don’t mention “danger” directly — focus on *empowerment* (“You’re in control”)

✓ SafeConnect — Project Criteria

- User-Centered Criteria
 - Age-appropriate interactions
 - Emotional Comfort and confidence
- Product and Design criteria
 - Seamless integration
 - Customizable by creators
- Brand Trust
 - Long term ecosystem growth
 - Stronger youth community

