The Ecology of Harmful Design: Risk and Safety of Game Making on a Metaverse Platform

Yubo Kou
College of Information Sciences and Technology,
Pennsylvania State University, USA
yubokou@psu.edu

Yingfan Zhou
College of Information Sciences and Technology,
Pennsylvania State University, USA
yxz5975@psu.edu

Zinan Zhang
College of Information Sciences and Technology,
Pennsylvania State University, USA
zzinan@psu.edu

Xinning Gui
College of Information Sciences and Technology,
Pennsylvania State University, USA
xinninggui@psu.edu

ABSTRACT

Metaverse platforms have been on the rise in recent years, offering three-dimensional (3D), immersive virtual worlds while encouraging user-generated content (UGC) in various forms. Roblox, a popular metaverse platform, enables its users to create a holistic virtual world (i.e., develop a 3D game) for other users to interact with. However, complex UGC is also challenging to moderate. Roblox has been notorious for its users' harmful designs, such as Nazi or terrorist role-playing mechanisms. In this study, we explore how harmful design takes place on Roblox. Through a grounded theory analysis of the 'r/Robloxgamedev' subreddit, we conceptualize an ecological view of harmful design, foregrounding three interconnected circumstances, namely sociotechnical risks, socioeconomic precarities, and normative (in)sensitivities, which work together to condition and give rise to harmful designs and bring about unique governance challenges to metaverse platforms. We conclude by laying out implications for design moderation.

CCS CONCEPTS

• **Human-centered computing** → Collaborative and social computing; Collaborative and social computing theory, concepts and paradigms; Computer supported cooperative work; • **Security and privacy** → Human and societal aspects of security and privacy; Social aspects of security and privacy.

KEYWORDS

Harm, Harmful Design, Game Making, Moderation, Platform Governance, Metaverse, Virtual World, Roblox

ACM Reference Format:
Yubo Kou, Yingfan Zhou, Zinan Zhang, and Xinning Gui. 2024. The Ecology of Harmful Design: Risk and Safety of Game Making on a Metaverse Platform. In *Designing Interactive Systems Conference (DIS '24), July 01–05, 2024, IT University of Copenhagen, Denmark.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3643834.3660678

1 INTRODUCTION

Persistent, three-dimensional (3D) virtual worlds, or metaverse, are increasingly popular today, facilitated by technological advances, increased bandwidth, as well as broader societal acceptance [25, 48]. The notion of metaverse came from the 1992 science fiction novel Snow Crash, and has attracted attention of HCI and virtual world researchers for many years [5, 19, 71, 99]. Contemporary instances of metaverse are characterized by heightened immersive experiences that can enhance how people communicate [70], work [73], learn [4, 43], and socialize [70]. The expanding scene and popularity of metaverse platforms naturally beg the question of who creates content for them. Tech companies such as Meta [10] and Microsoft [60] are eyeing user-generated content (UGC) as fuels for their future platform economy. Platforms like Roblox take a step even further employing a business model where they encourage users to create their own virtual worlds and share revenue from the monetization [11].

When UGC is incorporated into metaverse platforms, enormous governance challenges arise when it comes to harm and moderation [56]. Today, as the notion of metaverse entails significantly enriches user-generated activities and forms of UGC [19], some platforms like Roblox have already utilized scripting languages such as Lua to ease the entry to programming and empower a much wider and younger population to design and develop virtual worlds [76]. However, the empowered capabilities to create content in the metaverse come with unique risks, one of which is harmful design. In this paper, harmful design refers to user-designed virtual worlds that incur harm on players who interact with the virtual worlds. Such harm could be economic, emotional, or epistemic. For example, on Roblox, players have experienced ubiquitous microtransaction designs that sought to trick them to pay in order to advance in the games, sexually explicit interactions between avatars, as well as games that embed extremist ideologies [56]. In these examples and many others, harm does not just lie in static content in such forms as text or image that carries socially unacceptable meanings (e.g, text-based harassment [74] or image-based profanity [2]), but originates from design. However, despite the increasing amount of relevant news reports on harmful design on metaverse platforms (e.g., [50, 79]), the issue has received limited scholarly attention,

DIS '24, July 01–05, 2024, IT University of Copenhagen, Denmark

ACM ISBN 979-8-4007-0583-0/24/07
https://doi.org/10.1145/3643834.3660678

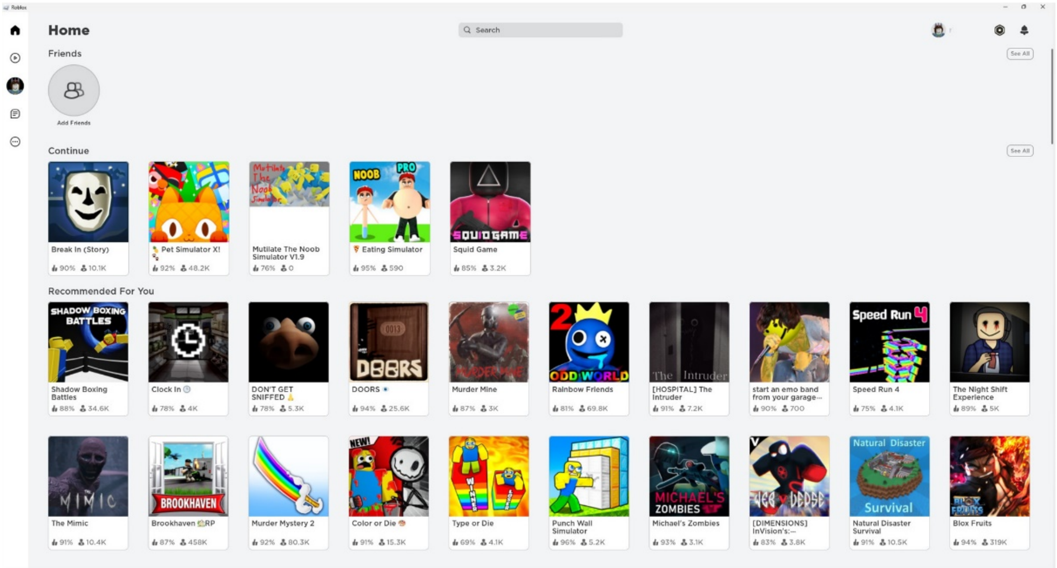


Figure 1: The Main Interface of the Roblox Client.

with one exception examining players experiences with harmful design in Roblox [56].

In this paper, we seek to understand how harmful design takes place on metaverse platforms. The metaverse platform we focus on is Roblox, which has nearly 50 million daily active users across 180 countries [86] and builds its business model on encouraging its users to create and interact with user-generated virtual worlds [83]. A virtual world on Roblox could also be considered as a game (or an "experience", in Roblox's own words). Using a grounded theory approach [39] to analyzing Roblox developers' discourses around harmful design on the 'r/Robloxgamedev' subreddit, we conceptualized an ecology of harmful design as an explanatory framework to contextualize and understand how harmful design takes place on Roblox. Specifically, we discuss three interrelated dimensions of the ecology which render Roblox developers as both perpetrators and victims: (1) They cope with sociotechnical risks embedded in the ecosystem of game making and unwittingly introduce harm into their design; (2) They negotiate with socioeconomic precarities, where they face exploitation from the Roblox platform but also seek to maximize profit from players of their creations; and (3) They normalize normative concerns about harmful design and focus on how to interact with moderation actions. Building on these insights, we further discuss governance challenges to metaverse platforms in light of the platformization of creative labor and harmful design, as well as implications for design moderation.

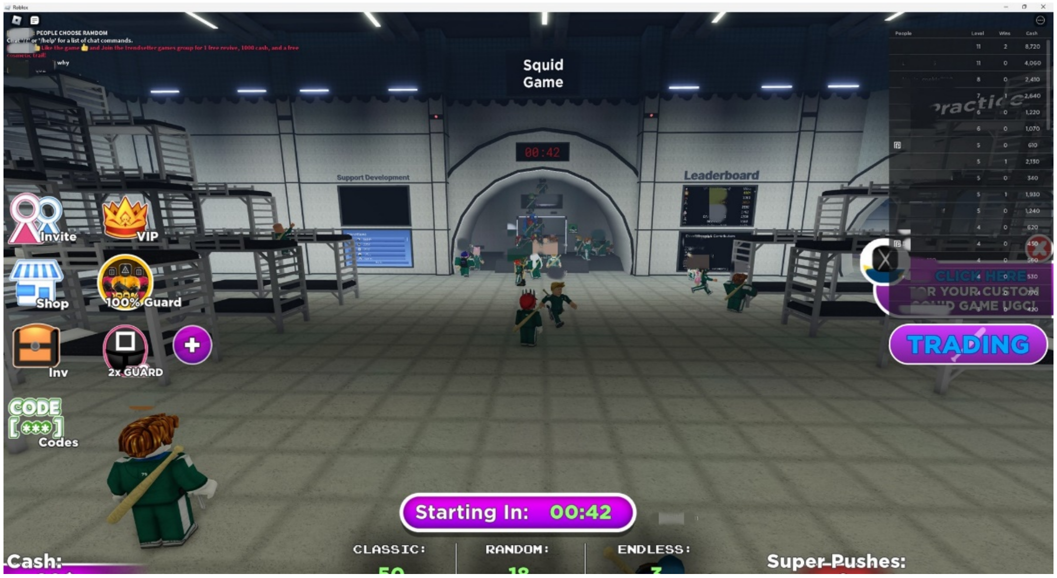
We contribute to the HCI and design literature in several ways. First, we provide an empirical account of harmful design on a metaverse platform, as well as conceptual insights into its origins and contexts; Second, we contribute to the moderation scholarship by elaborating on the emergent moderation issues from harmful designs produced by end users. Third, we reflect on governance challenges in light of the increasing platformization and monetization of creative labor.

2 BACKGROUND

Roblox is a metaverse platform with a massive player base, including a substantial number of younger children. It had nearly 50 million daily active users across 180 countries in 2021, with the US being the country with the most engagement time [86]. As a top online entertainment platform for kids and teens [85], the majority of its players are under the age of 13 [104]. For instance, in 2020, 54.86% of Roblox daily active users were under 13 years old, and 25% were under 9 years old [17]. More than two-thirds of kids aged 9 to 12 and a third of all Americans under the age of 16 in the US played it during the COVID-19 pandemic [51]. In the UK, there were around 1.5 million child players [20].

As a metaverse platform [104], Roblox features the core idea that a user can use one single avatar to access a network of connected virtual worlds (or games). The main interface of Roblox (see Figure 1) presents many games that are available for Roblox players. Those under the "Continue" category have recently been played, and those under the category of "Recommended For You" are curated by Roblox's recommendation algorithms. Players can find more games by searching or clicking the second button, "Discover," on the left sidebar.

Upon entering a game, players can operate their avatar within a 3D space in a third-person view. Each game has its own stories, setups, and gameplay mechanics. Figure 2 presents a screenshot of "Squid Game" on Roblox, a user-generated game that was created immediately following the success of "Squid Game," the most popular survival drama series on Netflix in 2021. In this particular game, its core gameplay mechanics simulate rules described in the TV show that players must follow to compete with others and only a few can win. In the screenshot, a group of avatars, each representing a unique Roblox player, were waiting in the lobby for a new competition to start (42 seconds left). While waiting, players



[redacted]: PEOPLE CHOOSE RANDOM
Chat '/?' or '/help' for a list of chat commands.
[redacted] 👍 Like the game 👍 and Join the trendsetter games group for 1 free revive, 1000 cash, and a free cosmetic trail!
[redacted] why

Figure 2: A Squid Game-Themed Game on Roblox.

typed to chat with each other and ran around to interact with other avatars as well as elements such as beds and walls in the lobby.

Roblox uses a business model where it encourages users to create games (called "experiences" by Roblox) and share revenue with them from the monetization of these games [83]. Roblox users can use Luau, a scripting language derived from Lua and used in Roblox Studio, to create games and other purchasable game content (e.g., virtual items for avatar decoration) for other players to purchase [91]. Roblox enables every end user to make and publish games, resulting in a community of 9.5 million developers and 40 million user-generated games [17]. About 5% of its tens of millions of child users have "published something of their own" [76]. When selling user-generated content and games, users will earn a share of Robux (Roblox currency) from each transaction. The rule is: The creator receives 30%, the seller or distributor gets 40%, and the rest 30% goes to the company. For instance, when a player creates a game and sells virtual items within the game, then the player gets 70%, since they are both the creator and the seller. However, if a player creates items and sells them in the Roblox marketplace, the player only gets 30% as the creator, while Roblox gets 70% as both the seller and the company [92].

Despite its success, Roblox has made numerous negative headlines related to harm and security issues, some of which stemmed from user-generated games. One of the main concerns is that Roblox users created harmful games that promote slavery, racial hate, and anti-Semitism [7, 50, 79]. In addition, Roblox has been criticized for making it very easy for children to spend large amount of money through microtransactions without their parents' knowledge [28, 75].

3 RELATED WORK

We ground our work in four strands of literature. First, we see clear continuity between the participatory culture of video game making and game making on Roblox. Second, we regard game making on Roblox as a form of creative labor which undergoes the process of platformization. Third, we draw from the value-sensitive design and values at play literature. Lastly, we also connect our work to the existing literature on online harm.

3.1 The Participatory Culture of Video Game Making

The dominant mode of video game making is producing games at professional studios or game companies [106]. Such mode is professional and publisher-centric [107]. However, in recent years, with more game engines, development tools, and distribution platforms becoming available, video game making has become democratized in a participatory culture that lowers the technical barrier so that amateurs could also produce game content [18, 23, 81, 106]. A participatory culture is "a culture with relatively low barriers to artistic expression and civic engagement, strong support for creating and sharing one's creations, and some type of informal mentorship whereby what is known by the most experienced is passed along to novices" [47].

Modding has been one of the primary forms of video game making in this participatory culture [81], where players and fans take an active role in "modification of a game through user-made additions of game content" [44]. Modders, the participants of modding, use their skills to undertake various types of modification of the game, such as graphic redesign, user interface (UI) customization (e.g., customizing the in-game information management dashboard UI design), game conversions (e.g., modifying in-game characters' capabilities), and hacking closed game systems (e.g., creating innovative modifications through reverse engineering) [96]. Modding demands a broad set of skills, such as programming, scripting, graphic design, and game design thinking [95]. It also requires modders to learn how to collaborate with others when it comes to large and complex games or mods [95]. Many modders acquire modding skills by participating in online modding communities [67, 94, 98].

Mods bring significant benefits to game companies, by increasing sales of the base game [80], providing innovative game designs and concepts to game companies [101], reducing marketing costs [97], and enhancing player experience [100]. However, modders are also exposed to potential exploitations and power imbalances. Modding can be viewed as immaterial labor and oftentimes, free labor which benefits the game company the most [45, 81]. Modders usually work long hours to complete a project, as modding is labor-intensive [45] Moreover, modders could be vulnerable. When viewed as derivative works, mods made without the copyright holder's consent can be seen as violations of copyright [105].

Independent (indie) game development, also represents the participatory and democratic culture of video game making [29, 30]. Indie game developers are "those who do not affiliate with large game companies or publishers but make and publish games in alternative ways such as self-funding/publishing, small teams/studios, and free labor" [61]. Resonant with themes in modding, indie game developers also face challenges such as learning [30], exploitation from big players [29], and teamwork [31].

Besides modding and indie game development, a few game companies have published game-making platforms, such as LittleBigPlanet, Minecraft, and Roblox, that afford players varying capabilities to create game content on their platforms. These platforms have been commonly celebrated for their blocky graphics, simple mechanics, and accessible programming languages to engage child players in learning and computational skill development [9, 41, 49, 68]. Prior literature on the platforms of LBP and Minecraft has focused on their educational and social potential. For instance, Ross et al. [93] reported how LBP 2 scaffolds the conventions and styles of game design for game creators to follow and build on. Ringland et al. discussed how Minecraft provides young players with autism a place to express them through their creations [84]. In comparison, Roblox's reputation of having rampant harmful designs and how harmful design becomes so on this game making platform is worth investigating.

Taken together, we see game making on Roblox has roots in the participatory culture of video game making which has already existed for a long time. Thus, there should be continuity in how amateurs take the initiative to learn and to make games. However, we also conjecture that the bottom-up game making practices led by amateurs could be complicated by the process of platformization, a process that we will discuss next.

3.2 The Platformization of Creative Labor and the Platform Ecology

While celebrating how the Internet has made online content creation a more open and egalitarian space, scholars have increasingly paid attention to the digital labor of content creators, against the backdrop that platforms exert centralized and nearly total control over content creators' creative labor, while deriving the majority of their profits from internet traffic and advertising income generated by content creators' work (e.g., [1, 6, 40]). In this context, creative labor refers to "commercializing and professionalizing native social media users who generate and circulate original content to incubate, promote, and monetize their own media brand" [14].

While platforms rely on the platformization of creative labor, content creators' labor conditions are also precarious. Because content with low visibility will have a limited audience reach and subsequently limited income, content creators must develop algorithmic knowledge about platforms' visibility algorithms in order to increase their visibility [52]. However, visibility is volatile by nature, due to the unpredictability of markets, industries, and platform features and algorithms [21]. In addition, even if creators strive to create content compliant with platform policies in order to make money, the ambiguity of platform policies also introduces much uncertainty and thus precarity [65]. Thus, platform studies have stressed such power imbalance between creators and platforms (e.g., [10, 54, 55]), observing how platforms may treat creators differently in terms of monetization [10] and career development [66].

While each platform seeks to better integrate their creators' labor into their platform logics, this platformization of creative labor also happens simultaneously with content creators who seek to diversify their labor and income stream across multiple platforms (e.g., YouTube, Instagram) [3, 64], emphasizing "Not putting all your eggs in one basket" [40]. Thus, content creators work not on one single platform but in an ecology of platforms. Here, platform ecology refers to "the multiple interrelations they create with their on/offline environments" [46], and "puts user practices and agency centre stage, accentuates the application of different platforms as an integral part of everyday life, and highlights the complexities of on/offline practices" [46]. Along this line, prior work has discussed how content creators' labor is supported or constrained by affordances of multiple platforms [22] or whether the effort to diversify platforms is effective [3].

Taken together, existing work on platformization and platform ecology lends us a perspective to explore the complex interaction between platforms and game making. Our work, in turn, enriches our existing understanding of creative labor in the platform era.

3.3 Game Design Ethics

Value sensitive design (VSD) is "a theoretically grounded approach to the design of technology that accounts for human values in a principled and comprehensive manner throughout the design process" [34]. Value could refer broadly to "what a person or group of people consider important in life" [35]. Batya Friedman and colleagues proposed VSD in the 1990s based on the observation that limited emphasis had been placed on values in the design of technology [33], but researchers started to raise concerns about human values such as privacy, autonomy, and informed consent [33, 34, 36]. Informed by VSD, Flanagan and Nissenbaum [27] proposed the values at play (VAP) framework to address the specific context of play. While the VSD focuses on analyzing systems already built, the VAP is suitable for interrogating the game design process [26].

Values that VSD and VAP research seek to cover tend to be abstract concepts, instead of specific design patterns. For instance, Flanagan and Nissenbaum provided a wide range of values such as diversity, security/safety, justice, and inclusion. Each of these values is fairly generic and abstract and can be applied to inform a broad range of design activities. Thus, these values need to be embedded in design practices in order to build the alignment between general values and particular design patterns [63]. For instance, the widely used usability heuristics [72] could be understood as a bridge, or intermediate form, between the general value of 'usability' and particular design patterns.

Methodology-wise, both VSD and VAP support empirical investigations which use empirical methods to understand and critique how values are embedded in technology in a real-world context [26, 34]. For example, drawing from VSD, Dadgar and Joshi conducted a field study to identify a set of values important to diabetic patients [15]. Informed by VSD and VAP, Kou and Gui conducted an online data analysis to surface four harmful design patterns in user-generated games on Roblox [56].

Taken together, VSD and VAP provide a productive perspective as we identify values in the game making practices on Roblox that are associated with harmful design. Starting from this view, this study then locates and analyzes Roblox developers' discourses around harmful design.

4 METHODS

The study was motivated by a general interest in understanding how harmful design takes place on contemporary metaverse platforms. This question was empirically motivated, and we sought to find an explanatory framework that can explain the circumstances that lead to harmful design. Here, circumstances correspond to the notion of value in VSD and VAP in capturing in broad terms what is deemed as important in the game making practices on Roblox. Thus, the scope of this inquiry aligns with a grounded theory (GT) methodology [13] which emphasizes theoretical construction. Next, we describe our methods in two parts, namely data collection and data analysis. The general workflow is illustrated in Figure 3. However, aligning with principles of GT [12, 13], these two processes were in fact highly interactive and mutually informed through the whole process of this research.

4.1 Data Collection

As we explained above, the general inquiry is well defined and aligns with Corbin and Strauss's view that a research question can "lead researchers into the data where they can explore the issues and problems..." [13]. Thus, we decided to turn to the 'r/Robloxgamedev' subreddit, where Roblox developers gather to discuss game making on Roblox, to answer this question. As of January 15, 2023, the subreddit had over 102 thousand members and is open to the general audience. The vibrant developer community and rich discussions on the subreddit provide a suitable site for our inquiry. In addition, Reddit's platform policy and API both allow data collection for research purposes. While Roblox has maintained an official developer forum, we did not find policy or API supporting such research purpose by the time of this research. Choosing this subreddit as a data source is in line with Corbin and Strauss's suggestion of being open to any source of data as fit [13]. Epistemology-wise, the study aligns with Charmaz's GT variant [11] that emphasizes a constructivist view and acknowledges researchers' subjectivity and pre-existing knowledge such as knowledge of related literature in the process of data collection and analysis. When we set out to explore the happening of harmful design on Roblox, we were already informed by the broad news coverage on harmful designs on Roblox (e.g., [7, 50, 79]).

Our data collection process lasted between the beginning of September 2022 and the end of November 2022. In the first week of September 2022, the research team, composed of four researchers, met to discuss the general areas of interest for this project. After that, we collected data for three years (from September 1, 2019 to August 31, 2022) from the subreddit, resulting in a dataset of 33,684 threads with 119,446 comments. Given the size of the dataset, it was infeasible for a qualitative analysis. Thus, each of us sampled and read the data to identify characteristics of data that pertained to harmful design. The purpose of this step was to familiarize ourselves with the data, through which each of us was able to develop an initial impression of the data in terms of how Roblox developers conversed, common topics they focused on, as well as how Roblox developers' conversations flew. For example, we noticed that Roblox developers would openly mention violent or sensitive content in their post titles and seek suggestions from their fellow developers. This type of thread would easily catch our attention in this initial read. Then, the research team met again to share their impressions of how Roblox developers engaged in meaningful conversations about harmful design in the collected data. Through this discussion, the research team agreed that the collected data had a substantial portion of threads involving harmful design, and thus would provide enough data for the GT approach that requires iterative data collection and analysis. Thus, we deemed that it was viable to employ this methodological approach to the dataset.

Then, two researchers first randomly sampled and identified 40 relevant threads with their associated 2758 comments and performed open coding on them. Here, the inclusion criteria we used for identifying a relevant thread were that it (1) was written in English, (2) was not deleted, and (3) mentioned harmful designs or harmful design-related experiences in either the post or the comments. The exclusion criteria were that the thread (1) was not written in English, (2) was deleted, or (3) was not about harmful design in any direct or indirect way. For example, a thread focused on explicit content was considered relevant, but a thread focused on programming for Roblox was considered irrelevant. An initial analysis of this small data set aligned with the GT methodology [12], resulting in basic codes that described harmful designs that Roblox developers talked about. This initial analysis, then, further informed our data collection effort, known as theoretical sampling in GT [13].

The GT methodology emphasizes an iterative process where researchers continuously collect and analyze new data through constant comparison, adding new codes, concepts, or categories to their existing set, until they have reached "theoretical saturation," meaning that researchers collectively decide that no new idea is found [13]. Thus, different from other qualitative methods such as content analysis that requires researchers to have or develop a codebook [58], the GT methodology does not generate or rely on a well-developed codebook through the analytic process. Following this principle, our analysis was performed after every week of data collection, where each coder had identified about 20 new relevant threads. Then, the four coders would hold a meeting to discuss the

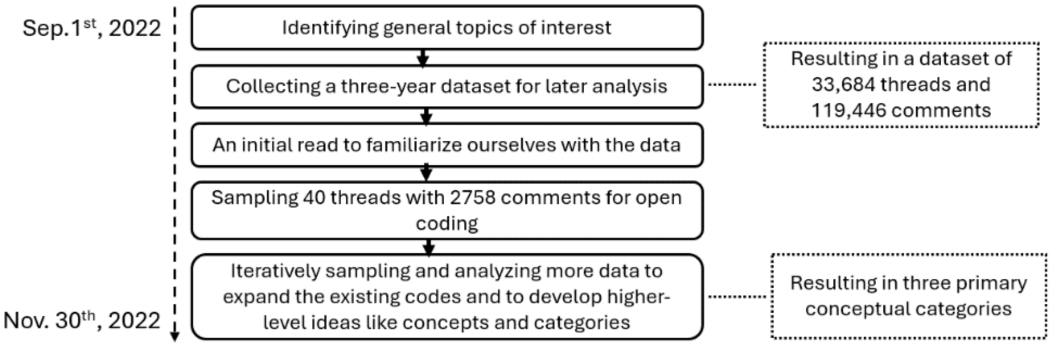

Sep.1st, 2022
Identifying general topics of interest
Collecting a three-year dataset for later analysis
Resulting in a dataset of 33,684 threads and 119,446 comments
An initial read to familiarize ourselves with the data
Sampling 40 threads with 2758 comments for open coding
Iteratively sampling and analyzing more data to expand the existing codes and to develop higher-level ideas like concepts and categories
Resulting in three primary conceptual categories
Nov. 30th, 2022

Figure 3: The workflow of our grounded theory approach.

new ideas within the newly identified threads, integrating the new ideas into our existing scheme of codes, concepts, and categories. In total, the research team identified 285 relevant threads with 11,378 comments, leading to 574 initial codes, 21 concepts, and three primary conceptual categories. The researchers continuously took memos throughout the data collection and analysis processes [13]. A memo is a written record of analysis [13]. For example, when analyzing the statement of "a window keeps popping up in my game asking me to buy something" from a Roblox developer, we wrote the memo as "a window that keeps popping up signals a potential risk. It is probably due to the game client having a viral free model that the developer does not understand. The risk happens when the developer lacks sufficient technical expertise and can incur financial harm if the developer ends up buying something through it. Such risk is also situated in a social context, where the developer likely have obtained this free model because of promotions from their friends or fellow developers. Thus, a viral free model can be viewed as a form of sociotechnical risk to developers who are new to game making on Roblox."

4.2 Data Analysis

Open coding is defined as "breaking data apart and delineating concepts to stand for interpreted meaning of raw data" [13]. A code describes what is in the raw data. Open coding was performed upon our initial dataset of 40 threads as well as after every new addition of data at a regular interval of one week. For example, a data record was coded as "community indicates the model could be used in child kidnap game" when Roblox developers talked about how a game scene mimics child abduction. The code was later grouped into a larger conceptual category called "normative (in)sensitivity in the moderation of game making."

To provide a bird's eye view of how the categories progressed through our GT analysis, when we completed the initial analysis of the 40 threads, none of the final categories appeared yet. Rather, we identified many conversations about how to make harmful designs and how to implement them to make money, leading to two primary categories concerning (1) how Roblox developers wanted to profit through game making, sometimes through harmful design, and (2) how Roblox developers knowingly created harmful design, disregarding player safety. However, through multiple iterations, we identified new concepts that added nuances to the first category, where developers shared their socioeconomic struggles as well as reliance on Roblox's infrastructure. Thus, the first category was refined to be "socioeconomic precarities in the monetization of game making" in the final theory, in order to capture their precarious conditions as they rely on Roblox to succeed in game making. The second category was further developed to be "normative (in)sensitivity in the moderation of game making" in the final theory, to account for how Roblox developers navigated ethical concerns and moderation practices on Roblox. Along with the refinements of these two categories, we additionally found that Roblox developers themselves were also vulnerable to exploitations and abuses, which further complicated the design processes that led to harmful design. We believed that such "sociotechnical risks in the ecosystem of game making" were significant enough to constitute a third reason behind how harmful design takes place on Roblox.

In sum, our GT analysis eventually led to three primary conceptual categories that form a holistic explanation of how harmful design takes place on Roblox. These three major categories are sociotechnical risks in the ecosystem of game making, socioeconomic precarities in the monetization of game making, and normative (in)sensitivity in the moderation of game making. We seek to construct a rich narrative for each conceptual category [11] to capture various elements and their interdependencies that are essential in the context of game making for Roblox.

4.3 Ethics Statement

The study was approved by the university's IRB office prior to the data collection and analysis efforts were made. The research team took caution in contemplating the potential risks and benefits of using publicly available data. The benefits of using data from the subreddit include the identification of harmful design in an emerging online context, and the potential of the research outcome to help promote online safety of metaverse platforms. We perceive no more than minimal risks associated with this research to people who made posts or comments in the subreddit. Following recommendations to disguise online data [8], we also paraphrased all the data to reduce its searchability. When reporting our findings, we use abbreviations (e.g., D1, D2, and D3) to represent interlocutors.

5 FINDINGS

Our grounded theory analysis pinpointed three primary categories (i.e., the sociotechnical, the socioeconomic, and the normative) of game making on Roblox where harmful design takes place. We do not stop at identifying how game developers introduce harm into their game design but explore the contextual factors behind such phenomenon and situate the practice of harmful game design in the broader culture of game making.

5.1 Sociotechnical Risks in the Ecosystem of Game Making

While Roblox hosts all the user-generated games and provides necessary development tools (e.g., Roblox studio) for a developer to make games, gaming making on Roblox is not necessarily a solitary practice. Instead, game making involves a large, complex sociotechnical ecosystem comprised of various user and developer groups as well as platforms and tools beyond those offered by Roblox. In this context, sociotechnical risk refers to how Roblox developers, especially new or unsuspecting ones lacking certain social or technical expertise, are exposed to risks inherent in the ecosystem of game making that they are unaware of.

One common risk for inexperienced developers comes from game design resources that they turn to. For example, both Roblox Studio and other Roblox developers provide free models (e.g., a 3D vehicle) for other developers to use in their own games. However, some free models contain backdoors or viruses and cause harm to unsuspecting developers. For example, a developer would seek opinions on a free model that they have adopted in their game:

> *D1: I have a major issue here. What are those things in the game I have been developing? I noticed a script and it is about duplicating something. Is it a virus? How should I deal with it?*
>
> *D2: Yes, they are a virus. They duplicate themselves every time you delete them in game. It could come from the free model which it was inserted in. When I was younger, I was scared by this type of stuff and didn't know what to do. I suggest that you use a trustworthy anti-virus plugin in the Toolbox.*

In the example above, D1 shared a screenshot of an anomaly in their game that was under development and sought to elicit feedback from other developers. D2 responded by identifying the cause as a virus and shared their experiences and suggestions dealing with viruses in free models. Across both developers' experiences, viruses in free models were designs intended to cause harm, or harmful designs. Such harmful designs posed a material risk when developers were new or inexperienced to know ways of risk mitigation.

When viral free models are used in a Roblox game, harm ensues after both the developers and the players who interact with the game. For example, viruses could harm the developer's game making environment. One developer observed that "*a window keeps popping up in my game asking me to buy something*" (D3). Another developer wrote that "*after my friend added a free model, now I got pop up ads in my Roblox studio*" (D4) A third developer described what a virus in a free model could do, writing:

> *If you get a pop up asking for your username, password, email, etc., it is a virus you got from a free model. It tries to make you turn on a http request and transmit the information about your game to an exploiting discord server through a webhook. (D5)*

In this quote, the developer detailed the steps a virus could take to engender harm on the unsuspecting developers, by leaking critical information to untrustworthy parties. The description also indicates a network of tools (e.g., viruses in Roblox game, discord server, and webhook) that are hidden and thus hardly knowable to inexperienced developers.

Viral free models are also harmful designs that can disrupt Roblox players' experiences. For example, a developer observed what a virus did to the players of their game:

> *My game has a free model with a hidden script that randomly moves players. This is why when they play my game, sometimes they are randomly teleported to a place. (D6)*

The quote above describes how Roblox developers could be harmed by viruses that perform actions against their original design goals. In addition, the viral free model could be harmful to Roblox players by moving their avatars against their will and thus violating their sense of autonomy.

In light of risks associated with viral free models, Roblox's moderation system enforces a strict policy against viral free models, oftentimes in a sweeping fashion. As a result, the anti-virus moderation decisions oftentimes hit inexperienced developers disproportionately. This is most evident in a thread which a father initiated on behalf of his daughter:

> *My daughter was banned for scamming. She recently began to build games and Roblox thought she made something illegal. My wife reached out explaining that there was no malicious intent, but Roblox refused to reinstate her account. . . My daughter had spent a lot of time and money on her only account, and is currently very upset. . . She was only trying things on Roblox, and wasn't sure what she made was considered a scam, but Roblox refused to provide a complete explanation. I wanted more information on what really happened.*

In the example above, a child developer who just started, presumably with little awareness of the risks associated with free models, received an account suspension for reasons she did not fully understand. The account suspension was costly to the young developer, because it was the only account in which she invested a significant amount of time and money. However, both the child developer and her parents were powerless in this situation. Thus, the father viewed such moderation as harsh and demanded more explanations about it. Clearly, in order to succeed in game making on Roblox, Roblox developers need to possess not only the technical expertise to make actual games, using free models or not, but also sufficient security knowledge about whether a free model contains harm. Otherwise, they can suffer losses incurred by viral free models.

Besides the technical risks, the ecosystem of Roblox game making also includes social risks that developers must navigate through. Social risks are the exposure to negative consequences stemming from collaboration with unfamiliar developers. Roblox developers can seek collaboration to develop large or complex games that require efforts from multiple developers. Roblox's platform supports developers to form groups and find collaboration. According to Roblox, a Roblox group allows "allows multiple creators to work on the same experience, use the same assets, share profits, and give credit to all contributors" [87]. However, the trustworthiness of such developer collaboration can be in question. For example, several developers complained about the loss or near loss of their games due to adverse behaviors by a former collaborator. One wrote:

> *A group member stole and deleted a game I have worked on for 2 months. Glad I was able to restore it. Otherwise, I would have lost years of work in another project. (D7)*

The quote above describes a scenario where developers within the same group had access to each other's games and corresponding codes, exposing developers to potential theft and unauthorized deletion. Adverse behaviors by collaborators also include the insertion of harmful designs in a game. Two developers conversed about this issue:

> *D8: A friend secretly added a UTG [short for ultimate trolling GUI] in my game, and I was banned by Roblox for seven days. What can I do about it?*
>
> *D9: The UTG has discriminatory commands. In the future, review your assets carefully before adding them to your game.*

UTG is a type of unacceptable script and thus harmful design on Roblox. D8 was punished for their collaborator's behavior of adding a UTG to their game, and D9 offered both an explanation for why it was a violation as well as a suggestion for future actions.

Developer groups and other mechanisms that facilitate collaboration between Roblox developers also expose them to possible scams, referring to malicious plans to obtain others' valuable assets or information. Here is an example where two developers talked about a common scam they encountered:

> *D10: Recently someone messaged me and wanted to use my avatar model in a thumbnail in their game. They also sent me an instructional video on how to make an avatar into a model. I don't know about this but wanted to make sure it's nothing sketchy.*
>
> *D11: This is a scam. They could get your avatar by simply searching it in Roblox studio. What they really wanted to do through the instructional video is to get your account information and steal your account. You should report them.*

The two developers above conversed about a possible scam that banked on the general norm of sharing. While the game making culture on Roblox celebrates amateur making and the spirit of sharing, much like what happened in modding [57, 67], it also exposes Roblox developers to possible scams, which could incur significant losses to them.

5.1.1 Summary. In this conceptual category, we discussed Roblox developers, especially new and inexperienced ones, are vulnerable to a variety of sociotechnical risks, such as using viral free models shared by others and collaborating with malicious actors. While the early days of game making practices saw enormous benefits of community-based practices such as sharing and learning [78, 101, 105], such practices on massive-scale metaverse platforms like Roblox no longer stay in accordance with the communal ethos. To successfully make games on Roblox, Roblox developers need to possess not just the technical expertise to develop games, but also sufficient security knowledge to mitigate these sociotechnical risks. However, these sociotechnical risks are not accounted for in Roblox's platform policy or moderation action. Rather, Roblox developers must bear the responsibilities for creating harmful designs and the losses incurred by harmful designs.

5.2 Socioeconomic Precarities in the Monetization of Game Making

Different from modding practices, in which modders are largely providing free labor to game companies [44, 59], the platformization of game making on Roblox comes with a set of socioeconomic arrangements that integrate Roblox developers' game making. The primary arrangement is a revenue sharing model, in which Roblox developers get to monetize their games and split Robux, Roblox's in-game currency, with the platform. But there are also other coexisting arrangements. For example, Roblox developers must pay an upload fee upfront to list their new avatar design for sale [88], or to promote their game in advertisement [89]. Such socioeconomic arrangements provide economic incentives for Roblox developers' game making, but also introduce socioeconomic precarities. Roblox developers find themselves facing uncertainties and insecurities with regard to their labor and financial stake. And such precarities, as mentioned frequently in the developers' discussions, become a condition for the emergence of harmful design.

Specifically, we found how the socioeconomic precarities are reflected in Roblox developers' acute awareness of unfairness in revenue sharing. One developer calculated that "*they first take 30% of the robux you have earned, and another 65% if you want to transfer Robux to real money. They take more than 80% in total*" (D12). Another wrote that "*Roblox is built on exploitation of people aged between 9 and 15. The players have to buy Robux from Roblox in order to purchase items in game. And when the fees are deducted, creators get less than 25% of what was spent*" (D13). A third even called for the unionization of developers, stating that "*I think Roblox developers should unionize and make Roblox to reduce their cut to a reasonable percentage*" (D14).

Despite the perception of unfair treatment and exploitation from Roblox, Roblox developers found it difficult to give up all their past investments in the platform (e.g., time, money, and learning how to make games in the particular development environment of Roblox Studio) and leave Roblox for another comparable game making platform. For example, one developer wrote:

> *If a development group wants to leave, they have to make sure that they have the financial support to develop a new game entirely different from the one they made on Roblox. Then, they have to worry about servers to keep the new game online, the player base, and then ways to monetize the game. (D15)*

As the quote above shows, the developer acknowledged Roblox developers' heavy reliance on Roblox's mature game making and publishing infrastructure. It would be overly costly for established developers to leave Roblox, for another game making platform such as Minecraft. In turn, they choose to stay in the unfavorable revenue sharing model, regardless of how unfair they may perceive the model as. While development groups face this dilemma, individual developers are also constrained by Roblox's socioeconomic arrangements. Two developers chatted about this issue:

> *D16: I have achieved the DevEx (short for Developer Exchange) goal. But I could not find a way to cash out my Robux.*
>
> *D17: There are a lot of factors to consider, including where your Robux come from, moderation history, whether you have premium, and whether your Robux is over 100k.*

DevEx is Roblox's official program that allows developers to exchange Robux for real money, but with multiple constraints. D17 shared their understanding of those constraints to D16, pointing to these constraints and highlighting that there is not a straightforward way to end the socioeconomic relationship with Roblox.

The socioeconomic stake of game making is substantial, but also requires sufficient knowledge about labor and economy to understand. This could prove challenging to children who "just want to make a game." Thus, Roblox developers could be quick to spot (or claim to have spotted) a child in forum discussions based on the implied socioeconomic knowledge, or lack thereof. Here is an excerpt:

> *D18: if you could make a map for my hangout game, I would pay you with 10% of Robux I made through the game.*
>
> *D19: This is not a job offer, but an empty promise. You don't even know whether your game could actually make money. A job offer would specify and guarantee when and how the payment would be done.*
>
> *D18: By this definition, every job offer is an empty promise. . . I already spent a lot of Robux on promotion. I don't want to do that again.*
>
> *D19: You must be a kid. Any type of job would need you to sign documentation to receive payment for work. . . This is why you don't worry when you grow up and work for an employer.*

In the conversation above, D18 promised a 'job offer,' seeking a developer to produce content (i.e., a map) for their game. D18's further explanation revealed their socioeconomic literacy of what a job offer entails, which, according to D19, matched neither what was prevailing on Roblox nor what was standard in the real world. Thus, D19 concluded that D18 must be a child. When Roblox runs its business model and encourages child developers to make games, their lack of sufficient socioeconomic knowledge is unattended to, putting them at a disadvantage.

The importance of possessing a certain amount of socioeconomic knowledge is also reflected in how Roblox developers could struggle to understand the mechanisms and formulas that determine whether and how much they get paid on Roblox. For example, a developer shared that "*I still haven't received my Robux yet, although I noticed premium players in my game many days ago. I made the game on my own. Is there anything that I missed?*" (D20)

As such, the socioeconomic conditions on Roblox have transformed game making into a commercial practice and predisposed Roblox developers to a set of socioeconomic concepts such as revenue share, profitability, monetization, and advertisements. Against this backdrop, design choices in game making can be motivated by socioeconomic incentives, at the expense of Roblox players, who are oftentimes children. Many developers openly discussed and admitted the economic risks associated with harmful design patterns. Cash grab game, in this context, refers specifically to low-quality games that are made for the sole purpose of generating revenue. One lamented:

> *Roblox has very good games, but they are less popular than the numerous popular low-quality games that target young children. In those games, you could sell children anything, and they would ask their parents for money to buy, for example, a funny pet in your game. (D21)*

The developer observed how certain designs could trick child players into purchasing behavior, thereby exposing them and their families to economic risks. They expressed concerns about the popularity of such harmful design patterns in Roblox games. Some other developers shared some other manipulative strategies that could accelerate children's spending in games. One wrote:

> *You could make the game addictive by adding pets. For example, players can buy eggs with Robux, and then the eggs could turn into pets. But you could set the probability of rare pets to be very low, like 1%. (D22)*

What the developer above suggested is loot box, a type of microtransaction, which is akin to gambling because of its chance-based nature [62]. The game design pattern they suggested is intended to target children at the expense of their and their families' economic interests. Regulations over such problematic microtransactions are unevenly distributed across the world. While countries like Belgium, Japan, and China have developed relevant laws to regulate such phenomenon, the U.S. legislature has not made much progress in this regulatory space [62, 69].

5.2.1 Summary. In this conceptual category, we described the precarious labor conditions, where Roblox developers perceive unfair revenue sharing but also significant barriers to leave. By building and owning a mature infrastructure that supports the entire cycle of game making and publishing, Roblox has acquired a dominant bargaining power over its developers when it comes to configuring the socioeconomic arrangements, such as revenue sharing, upload fee, and promotion fee. However, such unfair socioeconomic conditions are disguised under the utopian, inspirational tagline of "powering imagination" on Roblox's front page. Little do new developers know when they first sign up to become a creator. But as they become more experienced, some turn to harmful game designs to maximize their financial gains, even at the expense of players' economic interest, to meet their own financial goals.

5.3 Normative (In)Sensitivities in the Moderation of Game Making

Given the sociotechnical risks and socioeconomic precarities that Roblox developers face, the eagerness and aspiration to make a popular game and successfully monetize it becomes a highly valued, if not the foremost, goal of game making. The automated moderation on Roblox is conceived as more or less a barrier to monetization. As a result, player safety becomes a less regarded value in Roblox developers' game design practices. In this context, normative (in)sensitivity refers to the extent to which Roblox developers value normative goals such as compliance with community norms and platform policies when designing games.

Roblox developers have raised questions regarding the legitimacy of Roblox moderation, drawing from their previous negative experiences with moderation. A lot of developers expressed dissatisfaction regarding the standards that automated moderation enforced against them, stating that they were wrongly convicted by the moderation system. One wrote that "*Roblox's auto moderation is really bad. I was banned for 24 hours because of an image that I uploaded. I appealed the ban, and they got back to me four days later saying that my ban already ended so that I could log back in*" (D23) A second developer claimed that "*I just made a shirt with patterns near the top of the shoulder, and it was auto-moderated" (D24).* A third developer shared that *"I got a seven day ban for uploading crickets chirping because it was considered 'discriminatory and offensive*"' (D25)

When developers encounter problematic bans, the appeal process is their last and only resort. And developers talked about complexities in negotiating with the platform through the appeal process. For example, one developer stressed the importance of persistence, advising others to "*create multiple tickets and hopefully a good staff member will see it. In your ticket, explain why you think you are innocent, and try to get them to look further.*" Another developer also suggested a banned developer to "contact the platform as much as you can."

Roblox developers perceive issues with Roblox's automated moderation, such as receiving punishments they believe they do not deserve. In the meantime, they seek to better understand how the moderation system works and engage in ways to probe the boundaries of Roblox moderation. Here is a relevant conversation:

> *D26: I heard that Roblox doesn't like it if your game has too much blood too much. Do you think this is okay? [D26 shared a picture of the game they made]*
>
> *D27: As long as it's not too realistic, you are fine.*
>
> . . .
>
> *D28: You don't have a problem with blood. It is gore they are having a problem with. Also, you could make it an option so that your players can turn it on or off. This strategy helps here.*

In the conversation above, the developers conversed about Roblox moderation's boundary on the depiction of blood in game. D27 and D28 shared their respective opinions on where the line lied between acceptable and unacceptable depictions of gore and blood.

This type of normative insensitivity within some Roblox developers also manifested in conversations that focused on how to evade Roblox moderation, rather than whether a design was harmful or benign. For example, developers would discuss the appropriateness of designs that could be associated with fascism and terrorism. Below is an example:

> *D29: I notice red armbands in games and they are not banned immediately.*
>
> *D30: The developer probably didn't know red armbands are against Roblox policy.*
>
> *D31: He is making a history game and doesn't know the Nazi symbol?*
>
> *D30: But red armbands are used by other armies.*
>
> *D31: many cultures use swastikas, but this doesn't mean it is acceptable to use it in your game.*
>
> *D32: This is a bad idea. The account could be banned very soon.*

In this conversation, developers exchanged ideas about harmful design nuances related to Nazism. The focal point of the conversation, again, is not specifically about what is right or wrong, but about whether it complies with Roblox's moderation policy.

Lastly, with the accumulated knowledge about moderation as well as a certain degree of normative insensitivity, some Roblox developers might bypass Roblox moderation to embed harmful designs in their games. One developer mentioned a strategy where, if an image was deemed offensive by moderation, they could change a few pixels to try again. Another developer described in a conversation how they could find more harmful games on Roblox:

> *D33: if you want to have more gore, you have to go underground. For example, strip clubs and Nazi role-playing games are underground.*
>
> *D34: I visited strip club games before, but not Nazi role-playing games. Are they just WW2 role-playing games?*
>
> *D33: No, they are just Nazi games, sometimes with KKK, where you dress up like them. They exist, although not as common. . . For strip clubs you could use a few keywords, but for Nazi role-playing games, you have to have a Discord or come across them by chance.*

Here, D33 informed D34 about how certain harmful game designs existed on Roblox in secret, as well as possible ways to locate them. Both D33 and D34 were aware of how the games they were interested in violated Roblox's policy and normative values in society. Still, D33 and D34's conversation suggested their curiosity as well as their normative insensitivity regarding such games.

5.3.1 Summary. In this conceptual category, we observed how Roblox developers themselves disregarded ethical concerns about harmful designs in favor of design tactics and strategies to bypass Roblox moderation. Like many online platforms that prioritize scale and growth and only considers moderation in the second place [37], Roblox has focused on fostering a massive-scale, for-profit game making culture and operated a moderation system with many limits. Chiefly, its rigid enforcement of static standards, many of which are borrowed from content moderation, is perceived as lacking in both legitimacy and effectiveness. Roblox moderation has become learnable and gameable for Roblox developers. In this context, it is unsurprising that some Roblox developers choose to deprioritize normative values such as player safety and wellbeing, which are clearly stated in platform policies, in favor of practical ones such as monetization, success, popularity.

6 DISCUSSION

We reported on how harmful design takes place on a metaverse platform, by identifying three ecological dimensions, namely sociotechnical risk, socioeconomic precarity, and normative (in)sensitivity, which are interconnected and feed into each other. Within this ecology of harmful design, Roblox developers are both victims of risks and exploitations, as well as perpetrators of harm. What underpins this double role of Roblox developers is platforms' capitalist logics that operate to maximize profit from the platformization of user-generated games. Building upon these findings, we further discuss these governance challenges to metaverse platforms, and lay out implications for design moderation.

6.1 The Platformization of Game Making and the Emergence of Harmful Design

When metaverse platforms afford end users greater capabilities to create and interact in virtual worlds, moderation oftentimes falls short in identifying and punishing emergent forms of harm, such as avatar-based harassment in social virtual reality [32]. In our study, from the perspective of Roblox developers (whom Roblox calls creators), it is the harmful design that moderation struggles to address. Our findings showed that compared to conventional harms in modalities such as text and video, harmful designs can be particularly challenging to moderate because of its *perceptibility* and *complexity*. First, the perceptibility of harmful design means how players readily become aware of the design's associated risks. Roblox developers' conversations in our findings showed a general low level of normative sensitivity in injecting harmful design in their games. The harmfulness is compounded by the fact that the player population is largely composed of children. Second, the complexity of harmful design refers to the degree to which multiple parts are interrelated and work together to induce harm. Compared to harm embedded in conventional modalities (e.g., text-based harassment and image-based hate speech), harmful design is much more complex by several magnitudes. For example, while numerous algorithmic approaches have been developed to detect texts that endorse racism or Nazism (e.g., [16]), visual elements in Roblox, each of which could be acceptable on its own, could be combined in particular ways to render a scene that mimics a concentration camp, and Roblox's automated moderation falls shorts in detecting harm of such complexity. Thus, Roblox developers showed awareness of plenty of design strategies to bypass Roblox moderation.

Roblox developers are not only perpetrators but also victims. As shown in our findings, they are vulnerable to several interrelated risks, such as scams, frauds, viral free models, and sometimes incomprehensible account suspensions issued by Roblox moderation. When realized, these risks could lead to significant losses of time, effort, and money, which Roblox developers may have invested in their developer accounts for many years. This is not to justify their behavior of generating harmful designs, as social psychologists have long observed how people could act as both perpetrators and victims in their everyday life and derive various justifications for acting as such [77]. Instead, we seek to identify factors in the ecology of harmful design which can explain the propagation of harmful design.

The ecology of harmful design is mobilized and sustained by the platform's focus on the platformization and monetization of game making. It has been observed as a common issue that social media platforms have enjoyed a rapid growth and expanded to a sheer scale without proper attention to moderation, and only react when there is paramount societal attention to certain moderation-related issues [37]. Such observation also holds true for Roblox, as reflected in Roblox developers' conversations. They are primarily concerned with how to make popular games and monetize them, rather than what is safe design. Roblox developers' goals and desires align much with Roblox's business model of monetizing developers' game making labor.

From the participatory culture of game making to the platformization of game making, themes of exploitation [29] and of power imbalance [57] between companies and end users have resurfaced among Roblox developers. What's particularly alarming are the explicit socioeconomic arrangements that render other values less important. In the platform economy, Roblox's foremost priority—as a publicly-traded company—is to satisfy its shareholders' financial interests by boosting the base of its active user accounts, expanding its user demographics, and maximizing its gain from the revenue sharing model. Thus, aligning with the "growth over safety" doctrine of tech companies [24], harmful designs can exist within the purview of Roblox so long as they do not hurt the company's growth or trespass regulatory boundaries. In fact, certain types of harmful designs, such as low-quality cash grab games and gambling-like microtransactions, can directly bolster the company's revenue.

6.2 Governance Challenges to Metaverse Platforms and Implications for Policymaking

The ecological issues behind harmful design warrant a discussion of governance challenges that lie in how Roblox developers' game making practices are managed, structured, and incentivized. Our findings point to several recurring governance challenges to metaverse platforms and reveal several implications for policymaking.

First, how do metaverse platforms support the community of making? While the modding literature depicted a strong sense of community in modding practices where modders can learn from and collaborate with each other [67, 81], such communal spirit seems lost in game making on Roblox that is plagued by sociotechnical risks. Our findings reveal that Roblox developers struggled with learning how to make models and how to evaluate the safety of available resources and take advantage of safe ones. Without knowledge of risks associated with free models, new or inexperienced developers could fall prey to them. Furthermore, when Roblox developers form collaboration on complex projects, they lack a reliable collaboration infrastructure that could safeguard them against malicious collaborators who might steal their creations or insert harmful designs into their shared project.

Second, how do metaverse platforms foreground user safety as a key value? Informed by the VSD [33] and VAP frameworks [26], our empirical investigation has surfaced several problematic values that some Roblox developers uphold in their game making, such as profitability, popularity, and addiction. For example, D21 pointed to the existence of numerous popular, low-quality games that induce children into purchasing. D22 suggested a loot box design strategy that could profit off children's addiction. The disregard for the safety of children as players and developers is deeply concerning. While Roblox developers pursue economic interests over their players' safety, the platform also benefits financially through the revenue sharing model. In other words, metaverse platforms must center the safety of their users first, so that the creators will follow suit.

Third, how do metaverse platforms like Roblox respond to the ethical issues pertaining to a primarily child user group? What is distinctive about game making on Roblox is that child developers face ethical dilemmas in generating harmful designs that drive player engagement and spending. While much child research in HCI has discussed ethics of developing technology for children [102], where children are oftentimes placed in a passive role to be affected by technology, on Roblox, it is children who occupy an active role in developing technology that impacts other people and must wrestle with ethical situations in this process. Children are going through early, critical stages of moral development while learning how to justify and rationalize their choices [53]. Reading through Roblox developers' conversations, we noted concerning causes and rationales that they might utilize to justify the making of harmful designs, all of which converged at the goal of making popular, profitable games. However, this is not to put all the ethical burden on the shoulders of child developers. Ethical decision-making does not lie solely in the hands of designers, but a complex network of designers, users, and technologies that enable them [103]. In the context of game making on Roblox, this means that not just child developers, but also their parents, child players, and the parents of child players, as well as the platform, have a stake in ethical game creation. Child developers' parents are guardians and responsible for harm they cause (e.g., parental responsibility laws). And child players and their parents are also affected by unethical game design. Thus, a suitable governance framework should involve multiple stakeholder groups in ethical game making, including but not limited to child developers and their parents, child players and their parents, and platforms.

The host of governance challenges naturally lead us to reflect upon policymaking on metaverse platforms. We argue that policies on metaverse platforms should move beyond the content level and target the design level. Content-oriented policymaking tends to focus on creating an exhaustive list of content that is not allowed on the platform. But design-oriented policymaking should focus on governing the design processes. For example, to support the community of making, platform policies can be made to guide safe, trustworthy collaboration. To foreground user safety as a key value, platform policies should embed user safety as a key metric in evaluating and tracking the success of a game. Lastly, platform policies should promote ethical awareness of harmful effects of certain designs.

The governance challenges are severe enough to warrant attention beyond platform-level policymaking. For instance, recent years have witnessed emerging public events such as parents filing a class-action lawsuit against Roblox for its financial harms [82]. When metaverse platforms like Roblox profit from children creating games for other children to play, self-regulation is insufficient. Policymakers must closely scrutinize this underregulated phenomenon by examining how these platforms engage with their end users and collaborating with concerned parties, such as parents.

6.3 Implications for Design Moderation

Existing moderation approaches are efficient at addressing harms in traditional modalities as text, image, and video. However, given that harmful design exists not at the content level, but the design level, and that metaverse platforms' need for UGC grows in popularity and number at an exponential rate, it would be less effective if techniques to moderate content are adopted to moderate design. There is an imperative need for design moderation – a moderation approach that is suitable for moderating harmful designs. Design moderation refers to governance strategies to foster benign designs while discouraging harmful design patterns [56]. While design moderation can build upon existing understanding of content moderation, it has unique characteristics that demand focused research effort.

The chief distinction between content moderation and design moderation pertains to the modality of harm. Conventional modalities such as text and image enable harm to be easily produced, human-recognizable, and transferable across multiple platforms. Such harmful content is a 'static' object in the sense that it is out there for moderation techniques to act upon (e.g., detection, removal, or redaction). Harmful design, however, denotes a constant, dynamic process of making, through which the resulting product becomes harmful only through its interaction with the end users (i.e., Roblox players in this study). For instance, the simulation of Nazi camps is decidedly harmful even without user interaction, whereas for-profit microtransactions only become financially harmful when child users are manipulated into large spending.

Viewing harmful design as a process instead of static content holds important implications for design moderation. First, the common techniques of content moderation include excluding (e.g., banning unwanted users), pricing (e.g., raising the cost of participation through subscription or advertisements), organizing (e.g., deletion and filtering can alter the flow of content), and norm-setting (e.g., articulating policies and behavioral standards) [42], While content moderation is widely known for using 'excluding' and 'organizing,' we can immediately tell these two verbs might not do well on moderating harmful design, because they both hinge on clear identification of harm and its perpetrator. A harmful design is rarely clear at first sight. Thus, design moderation can emphasize the techniques of 'pricing' and 'norm-setting,' by enhancing the ethical awareness among designers and delineating the boundaries of harmful designs. At present, metaverse platforms' policies tend to take inspirations from content-oriented considerations. A pertinent example is how Roblox's policy discourages swearing in communication on their platform [90]. However, limited to no policies are directed at harmful designs. Thus, we see an urgent need to articulate harmful designs and make appropriate policies that can properly govern the design process on metaverse platforms.

Second, moderation also has descriptors, such as automatically/manually, transparently/secretly, ex ante/ex post, and centrally/distributedly [42]. In this study, Roblox developers' discourses problematized Roblox moderation that is performed automatically, secretively, ex post, and centrally, a common content moderation approach adopted by many online platforms. Such content moderation approach has already been criticized for its overstatement of accuracy, static policies and their enforcements, marginalization of the already marginalized, etc. [38]. This approach's limitations will only be enlarged when the target is harmful design, when automated approaches do not even have a comprehensive understanding of existent types of harmful design and new and inexperienced developers are being marginalized without efficient ways to appeal. Building upon this, we argue that design moderation should start from a community-based approach that emphasizes enhancing Roblox developers and players' ethical awareness of harmful designs through community building and collective sensemaking, rather than leaving it all to automated techniques.

Third, moderation approaches also differ in terms of how much control a platform has over its creators [42]. Although both content and design moderation are situated in platform economies that count on user-generated materials to drive up Internet traffic and a revenue sharing model to incentivize user creations, what is unique about design moderation is the vulnerability, powerlessness, and precarity of the end user designers. Content creators can diversify and publish content nearly simultaneously across platforms (e.g., Twitch, YouTube, and TikTok), but Roblox developers are tethered to this particular platform, because they have developed a peculiar set of platform-specific skills and assets and because there are limited alternatives to a whole game making and publishing infrastructure. This in turn underlies platforms' interest and willingness to improve their moderation in order to retain their creators and sustain their platform economy. Thus, moving forward, it is more plausible to address issues of harmful design if we can raise broader, societal awareness of this issue and form broader discourses and external pressure (the broad news coverage is already a good start), rather than counting on the platform/company to evolve a good system on its own.

6.4 Limitations

The work was focused on data from a single Roblox developer community. Thus, its findings may not generalize to other developer communities or other metaverse platforms. Much work can be done in the future to explore if similar findings can be identified in other developer communities or other metaverse platforms. In addition, the work was focused on data from a time period between September 1, 2019 and August 31, 2022. Thus, our findings may not reflect the latest dynamics within the Roblox developer community. However, the key dimensions and the ecological lens we identified in this work still hold theoretical implications for future research with similar interests.

7 CONCLUSION

In this study, we took a grounded theory approach to analyze how harmful design takes place on Roblox, a metaverse platform. We conceptualize the ecology of harmful design, where interrelated dimensions, including sociotechnical risks, socioeconomic precarities, and normative (in)sensitivities, work together to render Roblox developers as both victim and perpetrator. Thus, we see the governance challenges behind harmful design as existing at a grander scale, and demand attention beyond devising effective moderation techniques.

We call for more attention from both academia and industry to examining and mitigating harmful design on metaverse platforms. Specifically, academic researchers can identify and measure the forms and severities of harmful designs across different metaverse platforms such as Roblox, Fortnite, and Meta Horizon Workrooms. Developers for metaverse platforms can develop their ethical awareness of potential harms their designs can do to their users. Metaverse platforms devise and maintain appropriate content policies so as to prescribe safety design principles for their developers. Regulatory attention is needed to safeguard the work conditions and wellbeing of those who make content for metaverse platforms.

ACKNOWLEDGMENTS

We are grateful for the insightful and constructive feedback from the reviewers at DIS'2024. This work was partially supported by the NSF Grant #2326505.

REFERENCES

[1] Crystal Abidin. 2016. "Aren't These Just Young, Rich Women Doing Vain Things Online?": Influencer Selfies as Subversive Frivolity. *Social Media + Society* 2, 2 (April 2016). https://doi.org/10.1177/2056305116641342

[2] Wan Noor Hamiza Wan Ali, Masnizah Mohd, and Fariza Fauzi. 2019. Cyberbullying Detection: An Overview. *Proceedings of the 2018 Cyber Resilience Conference, CRC 2018* (January 2019). https://doi.org/10.1109/CR.2018.8626869

[3] Lauren Arnett, Robert Netzorg, Augustin Chaintreau, and Eugene Wu. Cross-platform Interactions and Popularity in the Live-streaming Community. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*, ACM. https://doi.org/10.1145/3290607.3312900

[4] Peter Ayton and Ilan Fischer. 2004. The hot hand fallacy and the gambler's fallacy: Two faces of subjective randomness? *Memory & Cognition* 32, 8 (December 2004), 1369–1378. https://doi.org/10.3758/BF03206327

[5] Shaowen Bardzell and Kalpana Shankar. 2007. Video game technologies and virtual design: A study of virtual design teams in a metaverse. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 4563 LNCS, (2007), 607–616. https://doi.org/10.1007/978-3-540-73335-5_65/COVER

[6] Ross Bonifacio, Lee Hair, and Donghee Yvette Wohn. 2021. Beyond fans: The relational labor and communication practices of creators on Patreon. *New Media & Society* (August 2021), 14614448211027961. https://doi.org/10.1177/14614448211027961

[7] Russell Brandom. 2021. Roblox is struggling to moderate re-creations of mass shootings. *The Verge*. Retrieved June 20, 2023 from https://www.theverge.com/2021/8/17/22628624/roblox-moderation-trust-and-safety-terrorist-content-christchurch

[8] Amy Bruckman. 2002. Studying the amateur artist: A perspective on disguising data collected in human subjects research on the Internet. *Ethics and Information Technology* 4, 3 (2002), 217–231.

[9] Noelene Callaghan. 2016. Investigating the role of Minecraft in educational learning environments. *Educational Media International* 53, 4 (October 2016), 244–260. https://doi.org/10.1080/09523987.2016.1254877

[10] Robyn Caplan and Tarleton Gillespie. 2020. Tiered Governance and Demonetization: The Shifting Terms of Labor and Compensation in the Platform Economy. *Social Media + Society* 6, 2 (April 2020), 205630512093663. https://doi.org/10.1177/2056305120936636

[11] Kathy Charmaz. 2006. *Constructing grounded theory: a practical guide through qualitative analysis*. Sage Publications.

[12] Tom Cole and Marco Gillies. 2022. More than a bit of coding: (un-)Grounded (non-)Theory in HCI. *Conference on Human Factors in Computing Systems - Proceedings* (April 2022). https://doi.org/10.1145/3491101.3516392

[13] Juliet M. Corbin and Anselm L. Strauss. 2015. *Basics of qualitative research: techniques and procedures for developing grounded theory* (4th. ed.). SAGE Publications, Inc.

[14] Stuart Cunningham and David Randolph Craig. 2019. *Social media entertainment: the new intersection of Hollywood and Silicon Valley*. NYU Press.

[15] Majid Dadgar and K. D. Joshi. 2018. The Role of Information and Communication Technology in Self-Management of Chronic Diseases: An Empirical Investigation through Value Sensitive Design. *Journal of the Association for Information Systems* 19, 2 (February 2018). Retrieved from https://aisel.aisnet.org/jais/vol19/iss2/2

[16] Thomas Davidson, Dana Warmsley, Michael Macy, and Ingmar Weber. 2017. Automated Hate Speech Detection and the Problem of Offensive Language. *Proceedings of the International AAAI Conference on Web and Social Media* 11, 1 (May 2017), 512–515. https://doi.org/10.1609/ICWSM.V11I1.14955

[17] Brian Dean. 2023. Roblox User and Growth Stats 2023. *Backlinko*. Retrieved June 20, 2023 from https://backlinko.com/roblox-users

[18] Christy Dena. 2008. Emerging Participatory Culture Practices: Player-Created Tiers in Alternate Reality Games - Christy Dena, 2008. *Convergence: The International Journal of Research into New Media Technologies* 14, 1 (2008). Retrieved June 20, 2023 from https://journals.sagepub.com/doi/pdf/10.1177/1354856507084418

[19] John David N. Dionisio, William G. Burns, and Richard Gilbert. 2013. 3D Virtual worlds and the metaverse: Current status and future possibilities. *ACM Computing Surveys (CSUR)* 45, 3 (July 2013). https://doi.org/10.1145/2480741.2480751

[20] Stuart Dredge. 2019. All you need to know about Roblox. *The Observer*. Retrieved June 20, 2023 from https://www.theguardian.com/games/2019/sep/28/roblox-guide-children-gaming-platform-developer-minecraft-fortnite

[21] Brooke Erin Duffy, Annika Pinch, Shruti Sannon, and Megan Sawey. 2021. The Nested Precarities of Creative Labor on Social Media. *Social Media + Society* 7, 2 (June 2021). https://doi.org/10.1177/20563051211021368

[22] Brooke Erin Duffy, Urszula Pruchniewska, and Leah Scolere. 2017. Platform-Specific Self-Branding: Imagined Affordances of the Social Media Ecology. In *Proceedings of the 8th International Conference on Social Media & Society* (*#SMSociety17*), July 28, 2017. Association for Computing Machinery, New York, NY, USA, 1–9. https://doi.org/10.1145/3097286.3097291

[23] Sean C. Duncan. 2010. Gamers as Designers: A Framework for Investigating Design in Gaming Affinity Spaces. *E-Learning and Digital Media* 7, 1 (January 2010), 21–34. https://doi.org/10.2304/ELEA.2010.7.1.21

[24] Elizabeth Dwoskin, Tory Newmyer, and Shibani Mahtani. 2021. The case against Mark Zuckerberg: Insiders say Facebook's CEO chose growth over safety. *Washington Post*. Retrieved June 22, 2023 from https://www.washingtonpost.com/technology/2021/10/25/mark-zuckerberg-facebook-whistleblower/

[25] E&T Editorial Staff. 2022. Children likely to spend 10 years of their lives in VR metaverse, study suggests. *Engineering and Technology*. Retrieved June 20, 2023 from https://eandt.theiet.org/content/articles/2022/04/children-likely-to-spend-10-years-of-their-lives-in-vr-metaverse-study-suggests/

[26] Mary Flanagan, Daniel C. Howe, and Helen Nissenbaum. 2005. Values at Play: Design Tradeoffs in Socially-Oriented Game Design. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (2005). https://doi.org/10.1145/1054972

[27] Mary Flanagan and Helen Nissenbaum. 2007. A game design methodology to incorporate social activist themes. In *Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '07*, April 2007. ACM Press, 181. https://doi.org/10.1145/1240624.1240654

[28] ParentsTogether Foundation. 2021. Warning for Parents: Kids spending thousands of dollars on "free" Roblox game. *ParentsTogether*. Retrieved June 20, 2023 from https://parents-together.org/warning-for-parents-kids-spending-thousands-of-dollars-on-free-roblox-game/

[29] Guo Freeman, Jeffrey Bardzell, Shaowen Bardzell, and Nathan McNeese. 2020. Mitigating Exploitation: Indie Game Developers' Reconfigurations of Labor in Technology. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW1 (May 2020), 23. https://doi.org/10.1145/3392864

[30] Guo Freeman, Nathan McNeese, Jeffrey Bardzell, and Shaowen Bardzell. 2020. "Pro-Amateur"-Driven Technological Innovation: Participation and Challenges in Indie Game Development. *Proceedings of the ACM on Human-Computer Interaction* 4, GROUP (January 2020). https://doi.org/10.1145/3375184

[31] Guo Freeman and Nathan J. McNeese. 2021. A Tale of Creativity and Struggles: Team Practices for Bottom-Up Innovation in Virtual Game Jams. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (April 2021). https://doi.org/10.1145/3449150

[32] Guo Freeman, Samaneh Zamanifard, Divine Maloney, and Dane Acena. 2022. Disturbing the Peace: Experiencing and Mitigating Emerging Harassment in Social Virtual Reality. *Proceedings of the ACM on Human-Computer Interaction* 6, CSCW1 (April 2022). https://doi.org/10.1145/3512932

[33] Batya Friedman. 1996. Value-sensitive design. *interactions* 3, 6 (December 1996), 16–23. https://doi.org/10.1145/242485.242493

[34] Batya Friedman, Peter H. Kahn Jr, and Alan Borning. 2002. *Value Sensitive Design: Theory and Methods*. University of Washington.

[35] Batya Friedman, Peter H. Kahn, Alan Borning, and Alina Huldtgren. 2013. Value Sensitive Design and Information Systems. *Philosophy of Engineering and Technology* 16, (2013), 55–95. https://doi.org/10.1007/978-94-007-7844-3_4/FIGURES/5

[36] Batya Friedman and Helen Nissenbaum. 1996. Bias in computer systems. *ACM Transactions on Information Systems* 14, 3 (July 1996), 330–347. https://doi.org/10.1145/230538.230561

[37] Tarleton Gillespie. 2018. Custodians of the internet: platforms, content moderation, and the hidden decisions that shape social media. Yale University Press.

[38] Tarleton Gillespie. 2020. Content moderation, AI, and the question of scale. *Big Data & Society* 7, 2 (August 2020). https://doi.org/10.1177/2053951720943234

[39] Barney G. Glaser and Anselm L. Strauss. 2009. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Transaction Publishers.

[40] Zoë Glatt. 2022. "We're All Told Not to Put Our Eggs in One Basket": Uncertainty, Precarity and Cross-Platform Labor in the Online Video Influencer Industry. *International Journal of Communication* 16, 0 (August 2022), 19.

[41] Sara M. Grimes. 2015. Little Big Scene. *Cultural Studies* 29, 3 (May 2015), 379–400. https://doi.org/10.1080/09502386.2014.937944

[42] James Grimmelmann. 2015. The Virtues of Moderation. *Yale Journal of Law and Technology* 17, (2015).

[43] Khe Foon Hew and Wing Sum Cheung. 2010. Use of three-dimensional (3-D) immersive virtual worlds in K-12 and higher education settings: A review of the research. *British Journal of Educational Technology* 41, 1 (January 2010), 33–55. https://doi.org/10.1111/J.1467-8535.2008.00900.X

[44] Renyi Hong. 2013. Game Modding, Prosumerism and Neoliberal Labor Practices. *International Journal of Communication* 7, 0 (April 2013), 19.

[45] Renyi Hong and Vivian Hsueh-Hua Chen. 2013. Becoming an ideal co-creator: Web materiality and intensive laboring practices in game modding. *New Media & Society* 16, 2 (2013). https://doi.org/10.1177/1461444813480095

[46] Oliver Ibert, Anna Oechslen, Alica Repenning, and Suntje Schmidt. 2022. Platform ecology: A user-centric and relational conceptualization of online platforms. *Global Networks* 22, 3 (July 2022), 564–579. https://doi.org/10.1111/GLOB.12355

[47] Henry Jenkins, Katie Clinton, Ravi Purushotma, Alice J Robison, and Margaret Weigel. 2009. *Confronting the Challenges of Participatory Culture: Media Education for the 21st Century*. The MacArthur Foundation.

[48] Trevor Laurence Jockims. 2022. Meta is opening its first store as VR headsets inch closer to mainstream reality. *CNBC*. Retrieved June 20, 2023 from https://www.cnbc.com/2022/05/08/meta-is-opening-a-store-as-vr-headset-sales-make-play-for-mainstream.html

[49] Thierry Karsenti and Julien Bugmann. 2018. The Educational Impacts of Minecraft on Elementary School Students. In *Research on e-Learning and ICT in Education: Technological, Pedagogical and Instructional Perspectives*, Tassos Anastasios Mikropoulos (ed.). Springer International Publishing, Cham, 197–212. https://doi.org/10.1007/978-3-319-95059-4_12

[50] Sean Keach. 2018. Roblox kids' game is a haven for twisted Jihadi, Nazi and KKK racist roleplay. *The Sun*. Retrieved June 20, 2023 from https://www.thesun.co.uk/tech/6710158/roblox-game-racist-jihad-nazi-kkk-racism-twin-towers-children/

[51] Olga Kharif. Kids Flock to Roblox for Parties and Playdates During Lockdown - Bloomberg. *Bloomberg*. Retrieved June 20, 2023 from https://www.bloomberg.com/news/articles/2020-04-15/kids-flock-to-roblox-for-parties-and-playdates-during-lockdown#xj4y7vzkg

[52] Erin Klawitter and Eszter Hargittai. 2018. "It's Like Learning a Whole Other Language": The Role of Algorithmic Skills in the Curation of Creative Goods. *International Journal of Communication* 12, 0 (September 2018), 21.

[53] Lawrence Kohlberg and Richard H. Hersh. 1977. Moral development: A review of the theory. *Theory Into Practice* 16, 2 (April 1977), 53–59. https://doi.org/10.1080/00405847709542675

[54] Susanne Kopf. 2020. "Rewarding Good Creators": Corporate Social Media Discourse on Monetization Schemes for Content Creators. *Social Media + Society* 6, 4 (October 2020), 2056305120969877. https://doi.org/10.1177/2056305120969877

[55] Susanne Kopf. 2022. Corporate censorship online: Vagueness and discursive imprecision in YouTube's advertiser-friendly content guidelines. *New Media & Society* (February 2022), 14614448221077354. https://doi.org/10.1177/14614448221077354

[56] Yubo Kou and Xinning Gui. 2023. Harmful Design in the Metaverse and How to Mitigate it: A Case Study of User-Generated Virtual Worlds on Roblox. In *Proceedings of the 2023 ACM Designing Interactive Systems Conference* (*DIS '23*), July 10, 2023. Association for Computing Machinery, New York, NY, USA, 175–188. https://doi.org/10.1145/3563657.3595960

[57] Yong Ming Kow and Bonnie Nardi. 2010. Who owns the Mods? *First Monday* 15, 5 (2010).

[58] Klaus Krippendorff. 1980. *Content analysis: an introduction to its methodology*. SAGE Publications.

[59] Julian Kücklich. 2005. Precarious Playbour: Modders and the Digital Games Industry. *The Fibreculture Journal* 05 (2005).

[60] Stuart Lauchlan. 2022. Game on! Microsoft’s near $70 billion gambit on the metaverse. *diginomica*. Retrieved June 20, 2023 from https://diginomica.com/game-microsofts-near-70-billion-gambit-metaverse

[61] Lingyuan Li, Guo Freeman, and Nathan J. McNeese. 2022. Channeling End-User Creativity: Leveraging Live Streaming for Distributed Collaboration in Indie Game Development. *Proc. ACM Hum.-Comput. Interact.* 6, CSCW2 (November 2022), 282:1-282:28. https://doi.org/10.1145/3555173

[62] Kevin Liu. 2019. A Global Analysis into Loot Boxes: Is It Virtually Gambling. *Washington International Law Journal* 28, 3 (2019), 763–800.

[63] Jonas Löwgren. 2013. Annotated portfolios and other forms of intermediate-level knowledge. *interactions* 20, 1 (January 2013), 30. https://doi.org/10.1145/2405716.2405725

[64] Renkai Ma, Xinning Gui, and Yubo Kou. 2023. Multi-Platform Content Creation: The Configuration of Creator Ecology through Platform Prioritization, Content Synchronization, and Audience Management. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, 2023. ACM Press. https://doi.org/10.1145/3544548.3581106

[65] Renkai Ma and Yubo Kou. 2021. “How advertiser-friendly is my video?”: YouTuber’s Socioeconomic Interactions with Algorithmic Content Moderation. *Proceedings of the ACM on Human-Computer Interaction* (2021). https://doi.org/10.1145/3479573

[66] Renkai Ma and Yubo Kou. 2022. “I’m not sure what difference is between their content and mine, other than the person itself”: A Study of Fairness Perception of Content Moderation on YouTube. *Proceedings of the ACM on Human-Computer Interaction* (2022). https://doi.org/10.1145/3555150

[67] Vittorio Marone. 2015. From Discussion Forum to Discursive Studio: Learning and Creativity in Design-Oriented Affinity Spaces. *Games and Culture* 10, 1 (January 2015), 81–105. https://doi.org/10.1177/1555412014557328

[68] Cecile Meier, Jose Saorín, Alejandro Bonnet de León, and Alberto Guerrero Cobos. 2020. Using the Roblox Video Game Engine for Creating Virtual tours and Learning about the Sculptural Heritage. *International Journal of Emerging Technologies in Learning (iJET)* 15, 20 (October 2020), 268–280.

[69] Kishan Mistry. 2018. P(L)aying to Win: Loot Boxes, Microtransaction Monetization, and a Proposal for Self-Regulation in the Video Game Industry. *Rutgers University Law Review* 71, (2018).

[70] Nicholas John Munn. 2011. The reality of friendship within immersive virtual worlds. *Ethics and Information Technology 2011 14:1* 14, 1 (May 2011), 1–10. https://doi.org/10.1007/S10676-011-9274-6

[71] Kizashi Nakano, Daichi Horita, Naoya Isoyama, Hideaki Uchiyama, and Kiyoshi Kiyokawa. 2022. Ukemochi: A Video See-through Food Overlay System for Eating Experience in the Metaverse. *Conference on Human Factors in Computing Systems - Proceedings* (April 2022). https://doi.org/10.1145/3491101.3519779

[72] Jakob Nielsen. 2005. Ten usability heuristics. Retrieved from https://www.nngroup.com/articles/ten-usability-heuristics/

[73] Mark E. Nissen and Richard D. Bergin. 2013. Knowledge Work Through Social Media Applications: Team Performance Implications of Immersive Virtual Worlds. *Journal of Organizational Computing and Electronic Commerce* 23, 1–2 (January 2013), 84–109. https://doi.org/10.1080/10919392.2013.748612

[74] Chikashi Nobata, Joel Tetreault, Achint Thomas, Yashar Mehdad, and Yi Chang. 2016. Abusive language detection in online user content. *25th International World Wide Web Conference, WWW 2016* (2016), 145–153. https://doi.org/10.1145/2872427.2883062

[75] Malcolm Owen. 2023. Family hit with $3,100 bill after kid goes on Roblox spending spree. *AppleInsider*. Retrieved June 20, 2023 from https://appleinsider.com/articles/23/05/22/family-hit-with-3100-app-store-bill-after-kid-goes-on-roblox-spending-spree

[76] Simon Parkin. 2022. The trouble with Roblox, the video game empire built on child labour. *The Guardian*. Retrieved June 20, 2023 from https://www.theguardian.com/games/2022/jan/09/the-trouble-with-roblox-the-video-game-empire-built-on-child-labour

[77] Monisha Pasupathi and Cecilia Wainryb. 2019. When I hurt others, and when I get hurt: Integrating victim and perpetrator experiences of harm into a sense of moral agency. *Social Development* 28, 4 (November 2019), 820–834. https://doi.org/10.1111/SODE.12334

[78] Nathaniel Poor. 2014. Computer game modders’ motivations and sense of community: A mixed-methods approach. *New Media & Society* 16, 8 (December 2014), 1249–1267. https://doi.org/10.1177/1461444813504266

[79] Felix Pope. 2022. Children’s game Roblox features Nazi death camps and Holocaust imagery. *The Jewish Chronicle*. Retrieved June 20, 2023 from https://www.thejc.com/news/news/the-nazi-death-camp-found-in-a-game-for-children-128DrQ3MoW1jzQa17oym2S

[80] Lev Poretski and Ofer Arazy. 2017. Placing Value on Community Co-creations: A Study of a Video Game “Modding” Community. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (*CSCW ’17*), February 25, 2017. Association for Computing Machinery, New York, NY, USA, 480–491. https://doi.org/10.1145/2998181.2998301

[81] Hector Postigo. 2010. Modding to the big leagues: Exploring the space between modders and the game industry. *First Monday* 15, 5 (2010). https://doi.org/10.5210/fm.v15i5.2972

[82] Emily Price. 2023. Parents File Another Class-Action Lawsuit Against Roblox. *PCMAG*. Retrieved May 4, 2024 from https://www.pcmag.com/news/parents-file-another-class-action-lawsuit-against-roblox

[83] Amanda Reaume. 2022. How Does Roblox Make Money? *Seeking Alpha*. Retrieved June 20, 2023 from https://seekingalpha.com/article/4486523-how-does-roblox-make-money, https://seekingalpha.com/article/4486523-how-does-roblox-make-money

[84] Kathryn E. Ringland, LouAnne Boyd, Heather Faucett, Amanda L.L. Cullen, and Gillian R. Hayes. 2017. Making in Minecraft: A Means of Self-Expression for Youth with Autism. In *Proceedings of the 2017 Conference on Interaction Design and Children* (*IDC ’17*), June 27, 2017. ACM Press, New York, NY, USA, 340–345. https://doi.org/10.1145/3078072.3079749

[85] Roblox. 2018. Roblox Emerges as a Top Online Entertainment Platform for Kids and Teens in 2017. *Roblox*. Retrieved June 20, 2023 from https://corporate.roblox.com/2018/03/roblox-emerges-top-online-entertainment-platform-kids-teens-2017/

[86] Roblox. 2022. A YEAR ON ROBLOX: 2021 IN DATA. *Roblox Blog*. Retrieved June 20, 2023 from http://https%253A%252F%252Fblog.roblox.com%252F2022%252F01%252Fyear-roblox-2021-data%252F

[87] Roblox. 2023. Group Collaboration. *Roblox Creator Hub*. Retrieved June 20, 2023 from https://create.roblox.com/docs

[88] Roblox. 2023. Marketplace Fees and Commissions. *Roblox Creator Hub*. Retrieved June 20, 2023 from https://create.roblox.com/docs

[89] Roblox. 2023. Promotion. *Roblox Creator Hub*. Retrieved June 20, 2023 from https://create.roblox.com/docs

[90] Roblox. 2023. Safety Features: Chat, Privacy & Filtering. *Roblox Support*. Retrieved June 20, 2023 from https://en.help.roblox.com/hc/en-us/articles/203313120-Safety-Features-Chat-Privacy-Filtering

[91] Roblox. Luau. *Roblox Creator Hub*. Retrieved June 20, 2023 from https://create.roblox.com/docs

[92] Roblox. Earning on Roblox. *Roblox Creator Hub*. Retrieved June 20, 2023 from https://create.roblox.com/docs

[93] Joel Ross, Oliver Holmes, and Bill Tomlinson. 2012. *Playing with Genre: User-Generated Game Design in LittleBigPlanet 2*. UC Irvine.

[94] Dongwan Ryu and Jiwon Jeong. 2019. Two Faces of Today’s Learners: Multiple Identity Formation. *Journal of Educational Computing Research* 57, 6 (October 2019), 1351–1375. https://doi.org/10.1177/0735633118791830

[95] Walt Scacchi. 2010. Computer game mods, modders, modding, and the mod scene. *First Monday* 15, 5 (2010). https://doi.org/10.5210/fm.v15i5.2965

[96] Walt Scacchi. 2011. Modding as an Open Source Approach to Extending Computer Game Systems. In *Open Source Systems: Grounding Research* (*IFIP Advances in Information and Communication Technology*), 2011. Springer, Berlin, Heidelberg, 62–74. https://doi.org/10.1007/978-3-642-24418-6_5

[97] Olli Sotamaa. 2007. On modder labour, commodification of play, and mod competitions. *First Monday* 12, 9 (September 2007). https://doi.org/10.5210/fm.v12i9.2006

[98] Shree Durga Subramanian. 2012. Moving past “hello world”: Learning to mod in an online affinity space. University of Wisconsin - Madison.

[99] Philipp Sykownik, Divine Maloney, Guo Freeman, and Maic Masuch. 2022. Something Personal from the Metaverse: Goals, Topics, and Contextual Factors of Self-Disclosure in Commercial Social VR. *Conference on Human Factors in Computing Systems - Proceedings* (April 2022), 17. https://doi.org/10.1145/3491102.3502008

[100] Sean Targett, Victoria Verlysdonk, Howard J. Hamilton, and Daryl Hepting. 2012. A Study of User Interface Modifications in World of Warcraft. *Game Studies* (2012).

[101] Sarah-Kristin Thiel and Peter Lyle. 2019. Malleable Games - A Literature Review on Communities of Game Modders. In *Proceedings of the 9th International Conference on Communities & Technologies - Transforming Communities* (*C&T ’19*), June 03, 2019. ACM Press, New York, NY, USA, 198–209. https://doi.org/10.1145/3328320.3328393

[102] Maarten Van Mechelen, Gökçe Elif Baykal, Christian Dindler, Eva Eriksson, and Ole Sejer Iversen. 2020. 18 Years of ethics in child-computer interaction research. In *Proceedings of the Interaction Design and Children Conference*, 2020. . https://doi.org/10.1145/3392063.3394407

[103] Maja van der Velden. 2009. Design for a common world: On ethical agency and cognitive justice. *Ethics and Information Technology* 11, 1 (December 2009), 37–47. https://doi.org/10.1007/S10676-008-9178-2/METRICS

[104] VentureBeat. 2020. Roblox believes user-generated content will bring us the Metaverse. *VentureBeat*. Retrieved June 20, 2023 from https://venturebeat.com/business/roblox-believes-user-generated-content-will-bring-us-the-metaverse/

[105] Ryan Wallace. 2014. Modding: Amateur Authorship and How the Video Game Industry is Actually Getting It Right. *BYU Law Review* 2014, 1 (January 2014), 219–255.

[106] Christopher James Young. 2018. Game Changers: Everyday Gamemakers and the Development of the Video Game Industry. University of Toronto.

[107] Peter Zackariasson and Timothy Wilson. The Video Game Industry: Formation, Present State, and Future. *Routledge & CRC Press*.