YOUTH, MENTAL HEALTH, AND THE METAVERSE: REVIEWING THE LITERATURE

Katie Salen Tekinbaş
Madison E. Taylor
Andre Adame
Stephen M. Schueller
F. Ria Khan

Produced by the Connected Learning Alliance
# CONTENTS

About Raising Good Gamers ................................................. 5  
Acknowledgments ................................................................. 5  
Youth, Mental Health, and the Metaverse Advisory Board ............. 6  
Executive Summary ............................................................... 7  
  Focus and Questions ......................................................... 7  
  Anticipating The Metaverse ................................................ 8  
  Snapshot of Our Findings .................................................. 8  
  Implications ................................................................. 9  
1. Introduction ...................................................................... 11  
  1.1 Questions and Focus ...................................................... 13  
     Focus on Teens and Young Adults ...................................... 13  
     Anticipating The Metaverse ............................................. 14  
     Attention to Equity and Inclusion .................................... 14  
  1.2 Understanding Adolescence ........................................... 14  
  1.3 Windows of Developmental Sensitivity and Age Appropriate Design .................................................. 15  
  1.4 Defining “Metaverse” ..................................................... 16  
     1.4.1 Building Blocks of the Metaverse .............................. 17  
     1.4.2 Enabling Technologies ........................................... 17  
     1.4.3 A Youth-Inclusive View ........................................... 17  
  1.5 Youth Engagement With Metaverse-like Spaces .................. 18  
  1.6 Methods ................................................................. 19  
2. Findings ........................................................................... 20  
  2.1 Common Concerns Associated With Youth Mental Health and Participation Online .......................... 20  
     2.1.1 Screen Time ......................................................... 20  
     2.1.2 Maladaptive Behaviors ........................................... 21  
     2.1.3 Exposure to Harmful Content .................................. 21  
     2.1.3.1 Eating Disorders ................................................ 22  
     2.1.3.2 Self-injurious Thoughts and Behaviors ................. 23  
     2.1.3.3 Substance Abuse ............................................... 24  
     2.1.3.4 Sexually Explicit Material ................................. 26  
  2.2 Social Risks .................................................................... 27  
     2.2.1 Distressing Encounters ............................................. 28  
     2.2.1.1 Cyberbullying ...................................................... 28  
     2.2.1.2 Sexual Harassment .............................................. 29  
     2.2.1.2 Hate Speech and Related Actions ......................... 30  
     2.2.1.3 Extremism ....................................................... 31  
     2.2.2 Social Comparison .................................................. 33  
     2.2.3 Fear of Missing Out ................................................. 34
ABOUT RAISING GOOD GAMERS

The promise and potential of games and new media as spaces of learning, social exchange, and civic engagement among youth will not be realized without a shared public agenda addressing the growing problem of toxicity online. While policies and infrastructures have long been in place to assure safe, public playgrounds for kids in the real world, no such approaches exist for the online playgrounds where young people are spending increasing amounts of time.

Raising Good Gamers (RGG) is an initiative focused on developing a response: designers, technologists, activists, researchers, funders, academics, parents, and educators. How might we develop and support gaming communities that cultivate empathetic, compassionate, and civically engaged kids? What might it look like to develop youth’s socio-emotional capacities to positively shape the climate of gaming clubs and communities? What role can the design of games, gaming communities, and associated technologies play in mitigating abuse? How do we build the foundations of a healthy community directly into the platforms and communities themselves?

The goal of RGG is to create a sustainable movement that can change the culture of online gaming for everyone. It was founded through an initial collaboration between the Connected Learning Lab at UC Irvine, Games for Change, and Take This. www.raisinggoodgamers.com

ACKNOWLEDGMENTS

This paper is part of the Youth, Mental Health and the Metaverse project, commissioned by The Jed Foundation in collaboration with Raising Good Gamers. It represents a substantial review of research related to youth, digital media, and mental health and was focused on understanding existing knowledge and predictions about the potential impact of the metaverse on youth mental health. The literature review was shared, in various stages, with an Advisory Board, convened by The Jed Foundation and facilitated by members of the Raising Good Gamers (RGG) initiative, as the grounds for identifying risks and opportunities, and recommended actions to address them. RGG team members included Katie Salen Tekinbaş, Susanna Pollack, Eve Crevoshay, Alana Selkowitz, and Rachel Kowert. Following the completion of the literature review The Jed Foundation hosted a broader convening of mental health, digital experts, and youth to discuss and refine the advisory board’s recommendations. This was followed by the publication of a report by The Jed Foundation—Can the Metaverse Be Good for Youth Mental Health? Youth-Centered Strategies for Ensuring and Enhancing the Mental Health and Safety of Young People in the Metaverse—describing the potential risks and opportunities of the metaverse on youth mental health, and specific recommendations for actions that metaverse platforms and developers, users, parents/caregivers, and regulators can take to help lower risks and help ensure the metaverse is responsive to the mental health needs of adolescents. Funding for this paper was made possible through the generous support of The Jed Foundation and the School of Computer and Information Sciences at UC Irvine. Jenna Abrams at the Connected Learning Lab offered invaluable editorial assistance. Thank you to Tiara Lynn Agresta for the layout and design of this report.
Youth, Mental Health, and the Metaverse Advisory Board

Grace Ahn
Associate Professor at Grady College of Journalism and Mass Communication, University of Georgia

Monica Arrambide
CEO and Founder, Maven Youth

Paula Cuneo
Co-Founder and Managing Partner, Fierce XR

Mark DeLoura
Games and Education Technology Consultant, Level Up Games

Ashley Elliott
Lead Consultant, Vivid Innovations Consulting; Resident Therapist, Psychological Wellness Services Arlington

Laura Erickson-Schroth
Chief Medical Officer, The Jed Foundation

Kishonna Gray
Associate Professor, Writing, Rhetoric, & Digital Studies and Africana Studies, University of Kentucky

Weszt Hart
Head of Player Dynamics, Riot Games

Celia Hodent
Game UX Strategy consultant / PhD in Psychology

Daniel Kelley
Director, Strategy and Operations, Center for Technology and Society, Anti-Defamation League

David Kleeman
Senior Vice President, Global Trends, Dubit Limited

Jennie Ito
Senior Product Policy Manager, Roblox

Amanda Lenhart
Head of Research, Common Sense Media

Michael Milham
International Neuroimaging Data-Sharing Initiative

Natasha Miller
Senior Research Scientist II (Social Design), Blizzard Entertainment

Alex Newhouse
Deputy Director at Center on Terrorism, Extremism, and Counterterrorism

Brian Nowak
Managing Director, Morgan Stanley

Mike Pappas
CEO and Co-Founder, Modulate.ai

Michael Preston
Executive Director, Joan Ganz Cooney Center at Sesame Workshop

Stephen Schueller
Associate Professor, Psychological Science, University of California-Irvine

Petr Slovák
UKRI Future Leaders Fellow & Associate Professor of Human-Computer Interaction

Jessica Stone
Chief Executive Officer, Chief Psychology Officer, Virtual Sandtray, King’s College London

Tiera Tanksley
Assistant Professor, School of Education, University of Colorado, Boulder

Rachelle Vallon
Guidance Counselor and Wellness Coordinator, Quest to Learn

Janis Whitlock
Senior Advisor, The Jed Foundation
EXECUTIVE SUMMARY

The impact of online social platforms on the mental health and wellbeing of adolescents has become a subject of concern as both positive and negative effects have been observed. Rising rates of anxiety, self-injury, and suicidal thoughts among adolescents have raised alarms and prompted researchers to explore potential links between technology use and wellbeing. A review of hundreds of studies presents a complex and uncertain relationship between youth mental health and social media use, with most studies reporting small associations in both negative and positive directions. The Surgeon General’s 2023 report emphasizes that the effects of social media vary among children due to cultural, historical, and socio-economic factors. While some children may experience excessive or problematic social media use, others benefit greatly from it, reporting feelings of acceptance, support, creativity, and connectedness.

This paper explores what we know about the risks and benefits of adolescent engagement with online social technologies, as a step toward anticipating and hopefully helping to shape the next iteration of the social internet: the metaverse. As the concept of the metaverse emerges, the potential benefits and harms on adolescent mental health and wellbeing come into question. Although research on metaverse environments is limited, insights from adolescents’ use of social media, online games, and augmented/virtual reality (AR/VR) provide a foundation for understanding the potential impact of metaverse spaces. The review focuses on youth ages 13-24, a significant transition period, while acknowledging that younger children are also active on social media. This is important to note as experiences with social and immersive technologies prior to the teenage years may affect wellbeing in later stages of adolescence and young adulthood.

Focus and Questions

Our review of the literature was focused on understanding and unpacking findings related to adolescent digital media use and impacts on adolescent mental wellbeing, both positive and negative. We sought to answer the following research questions:

**RQ1:** What are the risks and benefits to youth mental health associated with metaverse-like spaces?

**RQ2:** What are the protective factors associated with positive mental health outcomes among youth in metaverse-like spaces? How do these factors interact with the potential negative impacts on youth mental health related to participation in metaverse-like spaces?

We oriented our review around identifying both benefits and risks, with an eye toward informing a set of key recommendations. We were looking specifically at 1) adolescence as a developmentally significant period, 2) social media, games, and AR/VR technologies are precursors of the metaverse, and 3) issues of equity and inclusion. We reviewed available scholarly and gray literature, from fields
such as psychology, child development, learning sciences, game studies, and human-computer interaction (HCI), using a framework (see Appendix) organized around risks and protective factors.

**Anticipating The Metaverse**

The metaverse, a term coined by American writer Neal Stephenson in his 1992 sci-fi novel *Snow Crash*, has in recent years become shorthand within the tech industry and popular media more broadly for vast, persistent, and immersive simulated worlds where users can play, learn, work, shop, and socialize with others. The immersive worlds of the metaverse might be fully simulated, like virtual worlds, or might include layers of virtual content overlaid on the real world, like augmented worlds. What will likely differentiate metaverse experiences from current experiences with social media, games, or AR/VR is the feeling of mutual presence facilitated by environments with high visual and behavioral fidelity.

While significant uncertainty exists regarding what form the metaverse will ultimately take, we can look to the evidence-base around youth’s engagement with the social media, games, and AR/VR technologies now in order to anticipate and plan for the research, regulation, and design approaches needed to promote the positive effects of the metaverse while also protecting vulnerable youth.

**Snapshot of Our Findings**

- The current state of evidence does not support existing fears that technology is the driver of mental health problems among youth. Rather, the relationship between technology use and youth mental health must be understood as complex, multifaceted, and in need of further study.
- Screen time alone is not a reliable indicator of mental health outcomes, as other factors, such as sleep and bullying for example, are more powerful determinants of wellbeing in young people than digital screen use.
- There is some evidence that participation in multiplayer game environments has positive effects on youth wellbeing. These include outcomes like increased sense of agency and belonging, development of social emotional skills and habits, and activist stances.
- Online and offline vulnerabilities are closely connected, with negative experiences in online environments, such as social comparison or harassment, being influenced by factors like gender, socioeconomic status (SES), age, and race. These experiences are often linked to broader issues of poverty, instability, and marginalization. Moreover, the availability of different resources and support systems online can accentuate existing disparities in wellbeing, perpetuating inequalities that originate offline.
- One population that may be particularly vulnerable to negative mental health effects in the metaverse is adolescent girls. Generally, girls are at a greater risk of experiencing mental health disorders than boys starting in puberty and are also more relationally
inclined, which makes the social aspects of metaverse-like spaces more impactful. The metaverse needs to be designed with guardrails and scaffolds that support youth as they enter these spaces, especially girls.

- Online spaces are a beneficial resource for social support for many young people. Most teens and tweens say participation on social media platforms helps support their social-emotional wellbeing, boosting confidence and alleviating anxiety, loneliness, and depression.

- Peers and access to social support carry important protective effects for young people’s wellbeing. Metaverse-like spaces provide crucial social support to LGBTQ+ youth, for example, as well as youth from minoritized communities.

- Exposure to harmful content such as eating disorder content, content related to self-injurious thoughts and behaviors (SITB), and addictive substances content were found to be associated with negative mental health outcomes in youth, as is exposure to cyberbullying, hate speech, and sexual harassment. The role played by algorithms and moderation in exposing youth to harmful content is in need of further study.

- Social comparison is a common and robust factor through which social media is linked to negative outcomes on youth mental health.

- Online affinity networks play a crucial role in promoting youth wellbeing in the digital realm. These networks provide young people with genuine environments where they can develop important socio-emotional skills and behaviors, curate and share supportive media content, offer guidance to their peers on challenging topics, and foster a sense of community centered around marginalized interests and identities.

- Youth with high levels of self-esteem and resilience, along with strong self-regulating practices, are less susceptible to poor mental health outcomes associated with metaverse-like spaces.

- Family, friends, and educators can play a significant role in reducing risks and amplifying benefits by providing social support and guidance around digital technology use.

- Youth have shown that they are not powerless against exposure to risks online and use online platforms to combat harmful content, call attention to misinformation, and share mental health resources with others.

- Youth engage in a wide range of supportive behaviors online that could be amplified and supported to promote wellbeing.

Implications
Adolescence today is a life stage marked by active engagement with digital media, and youth will most certainly help drive what metaverse-like spaces can become, if given the opportunity. To achieve this goal, stakeholders must work together to steward forward a vision of the metaverse that takes the interests, needs, and vulnerabilities of youth into account, while also providing them with
the tools they need to support their mental health. Based on our findings, we propose the following guiding principles:

- **Equity matters**: Harms and benefits to youths’ mental health are not equitably distributed when it comes to participation in metaverse-like spaces. More precise and actionable knowledge requires studying the features of the platforms, how youth engage with metaverse-like spaces, and the experiences and backgrounds youth bring to their experiences there, in order to amplify or mitigate mental health risks for different adolescents.

- **There are no one-size-fits-all solutions**: Metaverse-like spaces should be designed in ways that support and acknowledge the great variation in youth across age, lived experiences, culture, identities, motivations, resources, and expertise. Doing so directs us to attend to research focused on identifying vulnerable populations, as well as groups, identities, and communities that are thriving online.

- **Recognize the importance of the pre-teen years**: Technologies developed with teens in mind will be used by pre-teens. We must consider how their pre-teen experiences with metaverse-like spaces may affect their mental health as teenagers. Partnerships between researchers and technology developers can support this work.

- **Center youth voices**: Engage youth as leaders in this work by centering youth experiences, needs, motivations, and assets in the design of codes of conduct, social dynamics, incentive systems, and moderation strategies. Shift the focus from solely protecting youth from online harm to creating opportunities for meaningful participation and empowerment.

- **Cross-sector support is needed**: There are important roles to be played by youth, platforms, guardians, policy makers, and professionals, in both counteracting metaverse-linked risks, and in amplifying its benefits. Supporting coordination and collaboration across sectors should become a core focus of philanthropic, business, research, and policy efforts.

This paper is a product of the Raising Good Gamers project at the Connected Learning Lab at the University of California, Irvine, an interdisciplinary research institute dedicated to studying, designing, and mobilizing digital technology in youth-centered and equitable ways. For recommendations emerging from the findings outlined here, please see [Can the Metaverse Be Good for Youth Mental Health? Youth-Centered Strategies for Ensuring and Enhancing the Mental Health and Safety of Young People in the Metaverse](#).
1. INTRODUCTION

Innovation and growth in the use of online social platforms has raised questions about the impact of such platforms, both positive and negative, on the mental health and wellbeing of adolescents. Increasing rates of anxiety, self-injury, and suicidal ideation among adolescents over the past decade (Mojtabai et al. 2016; Mercado et al. 2017; Duffy et al. 2019) have raised alarm bells about the state of adolescent mental health nationally, and have led researchers to look for possible connections between technology use and adolescent wellbeing outcomes. Some studies have found that high levels of engagement with digital media are responsible for a growth in mental health issues among adolescents, including higher levels of loneliness and depression, and lower overall wellbeing (Messias et al. 2011; Twenge and Campbell 2018; Rosenstein and Sheehan 2018; Kelly et al. 2018; Twenge et al. 2021). Other studies, however, including robust systematic reviews of evidence in the health sciences, show a more uncertain and complex relation between youth mental health and social media use (Baker and Algorta 2016; Dickson et al. 2018; Heffer et al. 2019; Keles, McCrae, and Grealish 2019; Orben and Przybylski 2019; Granic et al. 2020; Odgers and Jensen 2020; Orben 2020; Valkenburg et al. 2021a; Vogels and Gelles-Watnick 2023), highlighting small associations between digital media use and adolescents’ wellbeing in both negative and positive directions. A recent Surgeon General’s report emphasized that different children are affected by social media in different ways, including based on cultural, historical, and socio-economic factors (Office of the Surgeon General 2023). While social media use can be excessive and problematic for some children, for others social media use can be enormously beneficial. The Office of the Surgeon General (2023) reports the following:

Among the benefits, adolescents report that social media helps them feel more accepted (58%), like they have people who can support them through tough times (67%), like they have a place to show their creative side (71%), and more connected to what’s going on in their friends’ lives (80%). (p. 6).

A lack of evidence of a uniform effect of technology use among adolescents accentuates the importance of scrutinizing the features of the platforms, the ways in which youth engage with them, and the specific experiences and backgrounds youth bring to the platform, so that we may better understand not only the associated risks to mental health, but opportunities for positive development as well. Youth have shown that they are not powerless against exposure to risks online, and use online platforms to combat harmful content, call attention to misinformation, and share mental health resources with others (Oksanen et al. 2015; Cavazos-Rehg et al. 2017; Ito et al. 2020a).

This brings us to the concept of the metaverse, imagined as vivid networks of persistent virtual worlds that reach, interact with, and affect almost every aspect of human existence, from labor to leisure to art, health, and commerce (Ball 2022). The increasing popularity of gaming and online social interactions is considered a significant factor in the progression of the metaverse. This trend is also fueling the desire...
for more immersive experiences that connect individuals virtually. (UNICEF 2023). The arrival of a metaverse (whatever its eventual form), raises questions about the potential benefits and harms of such an environment on adolescent mental health and wellbeing. Because empirical research on metaverse environments is currently limited, we look to findings related to adolescents’ use of social media, online games, and augmented/virtual reality (AR and VR, respectively) upon which metaverse spaces are likely to be built. We focus our review on findings related to youth between the ages of 13-24, a key transition period from childhood into adulthood (Sawyer et al. 2018), but recognize that youth under the age of 13 are active on social media in many instances. This is important as experiences with social, immersive technologies prior to the teen years may affect experiences and mental health in the teen and young adult years to follow.

Our synthesis of existing research on youth, mental health, and metaverse-like spaces resulted in meaningful findings that countered, in most cases, the popular narrative that technology use was a driver of youth mental health problems. For example, we found that screen time was an unreliable indicator of youth mental wellbeing and that online spaces are a beneficial resource for social support for many youth, including those from vulnerable and minoritized communities (Nesi et al. 2023; Berger et al. 2022; Tanksley 2020a). Online affinity networks, or online spaces organized around shared interests, were shown to be an important part of youth wellbeing online. These networks provide young people with authentic environments in which they can develop crucial socio-emotional skills and behaviors. They can curate and share supportive media content, offer guidance to their peers on challenging subjects, and cultivate a sense of community around marginalized interests and identities (Ito et al. 2018; Ito et al. 2020b; Salen Tekinbaş et al. 2021). At the same time, exposure to harmful content such as eating disorder content, content related to self-injurious thoughts and behaviors (SITB), and addictive substances content were found to be associated with negative mental health outcomes in some populations of youth, as is exposure to cyberbullying, hate speech, and sexual harassment. Algorithms and moderation play a role in exposing youth to harmful content.

We found that negative social comparison is a common and consistent mechanism through which social media has been linked to negative mental health outcomes for youth. One population that may be particularly vulnerable to negative mental health effects is adolescent girls. Generally, girls are at a greater risk of experiencing mental health disorders than boys starting in puberty. Girls are also more relationally inclined, which makes the social aspects of metaverse-like spaces more impactful. Such findings suggest that the metaverse needs to be designed with guardrails and scaffolds that support youth as they enter these spaces, especially girls. Youth with high levels of self-esteem and resilience, along with strong self-regulating practices, are less susceptible to poor mental health outcomes associated with metaverse-like spaces. Family, friends, and educators can play a significant role in reducing risks and amplifying benefits by providing social support and guidance around digital technology use.
There is still much to be understood about how certain forms of engagement with online social technologies intersect with the varied identities, backgrounds, and motivations of a diverse population of youth (Ito et al. 2020a). We don’t yet understand, from an empirical perspective, how such patterns of engagement influence adolescent development and wellbeing. As a result, the relationship between problematic technology use and youth mental health must be understood as complex, multifaceted, and in need of further study. We start our review by identifying the core features and qualities associated with metaverse-like spaces and providing a framework for synthesizing our findings. We then compile a summary of our findings on the primary concerns associated with metaverse-like spaces, potential benefits of a metaverse-like space to youth mental health, and the risk factors and protective factors associated with adolescent mental health outcomes. We conclude by emphasizing the importance of scrutinizing the complex relationship between metaverse-like spaces and youth mental health outcomes.

1.1 Questions and Focus

Our review of the literature was focused on understanding and unpacking findings related to adolescent digital media use and impacts on their mental wellbeing, both positive and negative. We sought to answer the following research questions:

**RQ1:** What are the risks and benefits to youth mental health associated with metaverse-like spaces?

**RQ2:** What are the protective factors associated with positive mental health outcomes among youth in metaverse-like spaces? How do these factors interact with the potential negative impacts on youth mental health related to participation in metaverse-like spaces?

In reviewing the literature we were oriented around identifying both benefits and risks, with an eye toward informing a set of key recommendations for the Jed Foundation. We were looking specifically at 1) adolescence as a developmentally significant period, 2) social media, games, and AR/VR technologies are precursors of the metaverse, and 3) issues of equity and inclusion.

Focus on Teens and Young Adults

We focused on studies related to digital media use by youth between the ages of 13-24, in part because adolescents between the ages of 15-24 are at the highest risk of mental illness onset, with half of the mental health disorders with long-lasting effects starting in mid-adolescence (Kessler et al. 2005; Kessler et al. 2007; Merikangas et al. 2010). Teens and young adults have also been at the forefront of changes in how people engage with new technologies, information, social communication, activism, and learning. Given the importance of this developmental period (Yeager et al. 2018), it is essential to thoughtfully provide youth with tools and online social experiences that promote positive development while also safeguarding them from potentially harmful experiences. We look to the literature to tease out developmental needs of teens and young adults, as well as to understand the varying impacts youth may experience as they move from being young teenagers to becoming young adults. Throughout this report we refer to this age group as “youth.”
Anticipating The Metaverse

The metaverse, a term coined by American writer Neal Stephenson in his 1992 sci-fi novel *Snow Crash*, has in recent years become shorthand within the tech industry and popular media more broadly, as the next iteration of the social internet. Current virtual world platforms like *Second Life*, *Minecraft*, *Roblox*, and *Fortnite* allow youth to not only participate in gameplay together but also engage in social and distributed work, learning, play, and activism through a network of interconnected people (Gee 2004; Du et al. 2021). Youth are represented through avatars—digital representation of a user generated through computer technology (Holzwarth et al. 2006)—that are used to express social meaning and tie the user’s real world and virtual identities together (Park and Kim 2022). Game-adjacent technologies like Discord and Twitch extend the social spaces around communities and social media platforms like TikTok and YouTube and offer youth opportunities to create, share, curate, and connect. While significant uncertainty exists regarding what form the metaverse will ultimately take, we can look to the evidence-base around youth’s engagement with the social media, games, and AR/VR technologies now, in order to anticipate and plan for the research, regulation, and design approaches needed to promote the positive effects of the metaverse while also protecting vulnerable youth.

Attention to Equity and Inclusion

A final focus on our review was on issues of equity and inclusion. We sought to understand ways in which offline and online vulnerabilities are interrelated, for example, and how differential access and support available to youth might amplify offline inequalities in wellbeing (Ito et al. 2020a). We specifically focused on literature that addresses the structural inequities related to poverty, instability, and marginalization affecting young people. We also examined how online experiences for youth can be influenced by factors such as gender, socioeconomic status (SES), age, and race, potentially leading to stratification. In describing differences we use specific social and demographic terms as appropriate to the specific youth being described in a study. Like our colleagues at the Connected Learning Lab, we use the term “minoritized” to describe processes of marginalization from positions of institutional power and authority (Ito et al. 2020b). In place of the term “sexual minority”—an umbrella term for youth who identify as lesbian, gay, bisexual, or transgender, who engage in same-sex behavior, or who have same-sex attractions (Saewyc 2011)—we use LGBTQ+, except in instances where specific sexual identities are excluded from a study (e.g. trans youth).

1.2 Understanding Adolescence

Recent scientific progress has advanced understanding of adolescence as a dynamic developmental period—a formational social and emotional learning period for identity development. Adolescence is a period of great brain development, where brain systems that support control, reward, and social information processes undergo maturation and reconfiguration, starting from puberty and into young adulthood, that greatly influence their behavior (Andrews et al. 2021; Pfeifer and Allen 2021; Chiu and Chen 2022). It is also a period of heightened sensitivity and risk-taking, meaning that the brain is more susceptible to both physical and psychological harms (Steinberg 2014;
Adolescents are particularly susceptible to emotional manipulation, such as persuasive design and shady marketing practices, due to the incomplete development of their prefrontal cortex—the region responsible for managing emotions and impulses. Additionally, their vulnerability to stress and pressure, including peer pressure, is heightened compared to adults, as the prefrontal cortex does not fully mature until around the age of 25 (Konrad et al. 2013; Sawyer et al. 2018).

Further, adolescence is a crucial period for developing social and emotional habits important for mental wellbeing. These habits range from those focused on developing healthy sleep habits or exercising regularly, to learning to manage emotions, interpersonal relationships, and conflict (World Health Organization 2021). Metaverse-like spaces are increasingly important terrain within which this developmental work takes place—not just social media but also immersive online spaces such as social VR platforms and multiplayer gaming contexts (Magis-Weinberg et al. 2021). Expansion of metaverse-like platforms like AltspaceVR, Discord, Roblox, and Minecraft have offered adolescents the opportunity to experiment with and explore online spaces and places that reflect their interests and need for creativity, expression, and social connection (Reed and Joseff 2022). Recent hype about the future of a ‘metaverse’, defined by a commercially driven vision of an all encompassing future environment for social interaction online, adds to the urgency of our effort to understand what is known about the risks and opportunities to adolescent mental health (Rosenberg 2022c).

We need to better understand the risks and opportunities afforded by youth’s participation within these spaces to establish best practices from a design and research perspective—to reduce potential harm and promote healthy learning and development in this formative period. These represent compelling issues relevant to multiple outcomes, including education, mental health, wellbeing, and social inequities.

### 1.3 Windows of Developmental Sensitivity and Age Appropriate Design

Past research has established that each life stage exhibits its own unique trajectories, goals, and influences (Baltes et al. 1999; Arnett 2000; Lachman 2004; Blakemore and Mills 2014). As a result, activities taking place at different stages of life can “...differ from being of no impact, to being adaptive or maladaptive depending on which developmental period one examines” (Orben et al. 2022, P. 5).

Research into the possible impact of adolescent engagement with metaverse-like spaces should consider not only the life stage broadly—adolescence—but also the existence of windows of developmental sensitivity during the life stage. For instance, Orben and her colleagues (2022) conducted a longitudinal analysis involving 17,409 youth aged 10 to 21. Their research provides compelling evidence for specific periods during adolescence where the sensitivity to social media use varies based on age and sex. Interestingly, they discovered that higher estimated social media use during these sensitive periods predicts a decrease in life satisfaction ratings one year later, while lower estimated social media use predicts an increase in life satisfaction ratings. Moreover, they identified distinct age ranges for males (14-15 and 19 years old) and females (11-13 and 19 years old) when these sensitive windows occur.
Neuroimaging studies of social cognition have demonstrated similarly sensitive periods (Crone and Konijn 2018). Notably, research has found the change in social brain activity in the peer feedback condition is more pronounced for younger adolescents (ages 12–13-years) compared to mid-adolescents (15–16-years) (Van Hoorn et al. 2016; Crone and Konijn 2018). Collectively, these studies indicate that early adolescence may be a critical stage for social media’s impact on risk perception (Knoll 2015) and prosocial behavior (Van Hoorn et al. 2016). These findings align with research showing that adolescence is a period of social reorientation and social brain development (Blakemore and Mills 2014; Immordino-Yang and Gotlieb 2017). However, there is variability in the results regarding whether sensitive periods are more prominent in early or mid-adolescence. It is important to note that research has focused largely on usage of social media platforms like YouTube and Facebook, with a dearth of research in this area on games or AR/VR environments.

Recent work has been done on the policy front to define distinct developmental windows across adolescence. The Age-Appropriate Design Code, a policy framework implemented by the UK in 2020 that aims to enhance protections for children across all age groups who are likely to use online technologies, includes recommendations that point to distinct developmental windows in adolescence (Information Commissioner’s Office 2020). These windows include 10-12 years: The transition years; 13-15: Early teens; and 16-17: Approaching adulthood. The policy framework comes with information on key considerations (capacity, skills and behaviors a child might be expected to display) for each stage of development. For example, during a child’s early teen years (13-15) family remains a key influence, even as children seek greater levels of independence and autonomy. They may seek to actively flaunt parental or online rules and may emulate online influencers. Children in their early teens may exhibit tendencies towards idealized or polarized thinking, and they may be vulnerable to negative self-comparisons with others. They may also overestimate their capacity to handle the risks and challenges associated with online behavior and relationships. As a result, it is crucial to guide them towards various sources of support, including but not limited to parental assistance, to help address these issues effectively (Information Commissioner’s Office 2020). Policy frameworks like The Age Appropriate Design Code are intended to ensure that digital technologies are developed in a manner that considers not only the diverse age groups and developmental requirements of children and adolescents (Grace et al. 2023), but also specific developmentally sensitive windows within a life stage like childhood or adolescence. Gaining insight into the neurodevelopmental, pubertal, cognitive, and social changes that contribute to sensitive periods of influence (Orben et al. 2022) around engagement with metaverse-like spaces, and how these changes are influenced by individual differences, can guide the development of both policy and targeted interventions. Such interventions could aim to mitigate the adverse effects of a metaverse while simultaneously fostering its positive applications.

1.4 Defining “Metaverse”

No shared definition of the metaverse currently exists, either in academia or in commercial circles, although speculation abounds. For the purposes of this review, we draw on Rosenberg’s (2022a) formal definition of the metaverse: “The Metaverse refers to persistent and immersive simulated worlds (Orben et al. 2022).
that are experienced in the first person by groups of simultaneous users who share a strong sense of mutual presence” (Rosenberg 2022a, p.2). Rosenberg notes that immersive worlds can either be fully simulated, like virtual worlds, or can be layers of virtual content overlaid on the real world, like augmented worlds (Rosenberg, 2022b). From the perspective of youth and other users, the metaverse would be a place to engage in a set of daily activities that look a lot like activities conducted in non-digital and/or non-metaverse spaces: playing, learning, working, shopping, politically organizing, and socializing with others. What will likely differentiate metaverse experiences from current experiences with social media, games, or AR/VR is the feeling of mutual presence facilitated by environments with high visual and behavioral fidelity.

1.4.1 Building Blocks of the Metaverse

The internet, social networks, gaming, and virtual environments are current technologies upon which the metaverse will likely be built. These technologies include communication media, like text, graphics, and 3D virtual worlds that allow users to socialize and interact with others. AR and VR applications have introduced highly immersive digital overlays in the world, like those used in games like Pokémon GO and Super Mario AR, or social applications such as VR Chat. Blockchain technologies are showing up in virtual worlds like The Sandbox, which supports in-game trading, decentralized governance, and the ability to earn non-fungible tokens (NFT) that can be converted into currencies in the real world (Lee et al. 2021; Cerasa et al. 2022; Lovich 2022, Ravenscroft 2022).

1.4.2 Enabling Technologies

Technologies are the enablers that will drive the transition from the current internet to the metaverse (Lee et al. 2021). Examples of enabling technologies include AR/VR or extended reality (XR), artificial intelligence, blockchain, computer vision, emerging mobile computing technologies, the Internet of Things (IoT), and robotics, as well as high speed, edge and cloud computing, and microservices architecture (Lee et al. 2021; Cerasa et al. 2022; Howell 2022; Rosenberg 2022a, 2022b, 2022c). The metaverse would connect and interact with our physical world through human-computer interfaces like head mounted wearable displays or mobile headsets (e.g. Oculus Rift, Microsoft Hololens3), avatars, computer agents in the metaverse interacting with smart devices and robots, neural and voice processing devices, to name but a few (Lee et al. 2021; Benrimoh et al. 2022; Wallace et al. 2023).

1.4.3 A Youth-Inclusive View

Current definitions draw from commercial and technological perspectives and do very little to account for the potential needs and motivations of a diverse set of audiences, while also largely ignoring youth as a core audience (Lenhart and Owens 2021). In our synthesis of the literature we have chosen an alternate path. We have imagined the metaverse to be composed of many different kinds of experiences, for different audiences. As such, we will refer to “a metaverse” rather than “the metaverse” going forward, and “metaverse-like spaces,” more generally. Second, we have
sought to surface literature that counters a technologically deterministic view through a focus on the developmental and cultural needs of youth transitioning from childhood to adulthood. The transition from childhood into adolescence is a pivotal inflection point in the life course—this begins a formative period of learning and development that is not only shaping individual motivations, mindsets, and identities in new ways, but also contributing to shifts in life course trajectories that impact outcomes ranging from physical and mental health to education, social, and economic success (Immordino-Yang and Gotlieb 2017; Dahl et al. 2018). It follows that metaverse-like spaces should be designed in ways that support and acknowledge the great variation in youth across age, lived experiences, culture, identities, motivations, resources, and expertise. Doing so directs us to attend to research focused on identifying vulnerable populations, as well as groups, identities, and communities that are thriving online (Ito et al. 2020a).

1.5 Youth Engagement With Metaverse-like Spaces

Youth are engaging with social media and games in increasing numbers, in varying doses, and for a range of reasons. According to a 2022 study conducted by the Pew Research Center, “...there has been an uptick in daily teen internet users, from 92% in 2014-15 to 97% today. In addition, the share of teens who say they are online almost constantly has roughly doubled since 2014-15 (46% now and 24% then)” (Atske 2022). YouTube and Tiktok are the most popular social media platforms, followed by Instagram and Snapchat. The study pointed to demographic differences in choice of social media platform: teen girls are more likely than teen boys to use TikTok, Instagram and Snapchat, while teen boys are more likely than teen girls to say they use YouTube, Twitch and Reddit (Atske 2022). Far fewer youth are active users of VR devices, with a recent industry study reporting that “...although almost a third of all teens own VR devices (29% to be precise), ...only about 14% of youngsters used their VR devices on a weekly basis” (Piper Sandler 2023).

Games remain popular among youth: The Electronic Software Association 2022 report shows that 71% of youth under the age of 18 play video games (ESA 2022). The percentage of gamers that identify as female has consistently hovered around 40% since 2008. But over the past decade, women's share of the gaming population has grown to nearly half, with 48% identifying as female and 52% identifying as male.¹ This suggests that metaverse-like spaces will be for everyone. 83% of all video game players play with others online, or in person (up from 77% in 2021 and 65% in 2020) (ESA 2022), with an emphasis on play with friends, family, and partners. This increase is due, in part, to impacts of stay-at-home orders during the COVID 19 pandemic. One of the few bright spots for children and families during this time were online games like Minecraft and Roblox, and game adjacent platforms like Discord, which offered youth spaces and customizable virtual worlds where they could stay connected with friends (Roblox 2020; Navarro 2021; Barr and Copeland-Stewart 2022). ESA’s survey also showed that parents are active in supervising, and in many cases restricting, their children’s use of video games. For example, 79% of parents have screen time rules for playing video games such as specific time of day or how long their children can play (or both),

¹ No data was included about non-binary youth.
while 76% indicated their children must get a parent’s permission to communicate with others online (ESA 2022).

1.6 Methods

In order to effectively anticipate metaverse risks to youth mental health, we reviewed available scholarly and gray literature from fields such as psychology, child development, learning sciences, game studies, and human-computer interaction (HCI), using a framework (See Appendix) organized around risks and protective factors. We began scouring literature in the summer of 2022, incorporating search terms around the target demographic (e.g., “adolescent,” “teen,” “young adult,”), metaverse qualities (e.g., “social media,” “online gaming,” “virtual reality,”), and mental health outcomes (e.g., “wellbeing,” “depression,” “self-harm,” and “anxiety”). Searches were conducted first on academic databases including Academic Search Complete, APA PsycArticles, APA PsycINFO, Google Scholar, Jstor, ProQuest, Scopus, and Web of Science, and second using the standard Google search engine. As risks and opportunities associated with metaverse spaces were identified, we expanded our search to include more specific terms such as “harmful content,” “cyberbullying,” and “social comparison.” Finally, references of papers themselves were checked for additional citations. Considerations for incorporating literature into the review included empirical foundations, sampling methods, analysis, and length of the study. All papers selected for review were written in English.

Members of the research team were assigned literature to review by topic (e.g., maladaptive behaviors, identity, social isolation, etc.), meeting twice weekly for five months to discuss findings, identify areas of overlap, and construct a model for synthesis. Once an appropriate model was identified, the team worked collectively to draft the final text, which was reviewed by each member, at which point citations were finalized. A draft of the text was shared with members of a cross-sector advisory group for review and comment. The Jed Foundation convened the advisory group, with support from the Raising Good Gamers team. The review was then revised based on feedback from the advisory group, The Jed Foundation, and the Raising Good Gamers team.
2. FINDINGS

We begin our survey of the literature with a focus on findings related to screen time, maladaptive behaviors, and exposure to harmful content, including eating disorder content, self-injurious thoughts and behavior content, and addictive substances content. We then cover research related to social risks, including distressing encounters like cyberbullying, sexual harassment, hate speech, and extremism. Following this, we discuss findings related to social comparison, fear of missing out, isolation from friends and family, and privacy.

2.1 Common Concerns Associated With Youth Mental Health and Participation Online

There are several common concerns frequently affiliated with youths’ participation in metaverse-like spaces and their mental wellbeing. For example, previous research has emphasized links between the amount of time youth spend engaging with digital technology and mental health outcomes. However, such emphasis often overlooks important contexts of the content and activities that youth are engaging with. This section will examine the mental health implications associated with several common concerns around youth participation in metaverse-like spaces, including time spent with digital devices, the development of maladaptive behaviors, and exposures to harmful content.

2.1.1 Screen Time

A significant concern among various stakeholders when considering mental health outcomes in the metaverse is whether increased levels of digital media use may worsen the mental health of teens and young adults. Although this literature review will explore a number of risks, opportunities, and protective factors, specific comments on the relationship between screen time and mental health and wellbeing will be brief. On the whole, while some association may exist between total screen time and youth mental health, their relationship is weak at best. A review of the literature surrounding “screen time” and youth mental health shows mixed findings. Paulich and colleagues’ (2021) secondary analysis of 11,875 youth from the Adolescent Brain Cognitive Development (ABCD) study found that screen time predicted increased internalizing symptoms, behavioral problems, and sleep quality and quantity, and predicted decreased academic performance. However, under 2% of the variance in these ratings was explained by screen time, leading the investigators to conclude that the “negative impact on the subjects is likely not clinically harmful at this age” (Paulich et al. 2021). Oswald and colleagues’ (2020) systematic scoping review of studies on screen time and youth mental health found that while some studies showed higher levels of total screen time were associated with lower psychological wellbeing, lower emotional functioning, and lower self-esteem in adolescents from 12-14, and higher levels of anxiety and depression symptoms in adolescents aged 15-18, other studies they reviewed showed no relationship between screen time and mental health in adolescents. Other researchers have found that greater use of digital technology in adolescents is weakly associated with decreases in wellbeing, explaining 0.4% of the variation in adolescents’ wellbeing (Orben and Przybylski 2019).
One reason for these weak and inconsistent findings may be that screen time alone does not accurately capture adolescents’ activities on digital media. As such, prominent researchers have called for moving beyond evaluating the relationship between “screen time” and youth mental health (Orben 2020; Odgers et al. 2020). Rather, youth mental health in online spaces may be better understood through the types of online content consumed and the amount of each type, connections between different modes of digital technology use, and the individual differences among youth and specific populations of youth (Orben 2020; Granic et al. 2020). As such, this review adopts a similar framework, focusing on identifying both the mental health risks and benefits of youth engaging in specific behaviors and with specific types of content in metaverse-like spaces. Subsets of youth who may be most vulnerable to negative mental health effects are also highlighted.

### 2.1.2 Maladaptive Behaviors

Studies so far have largely failed to demonstrate a causal relationship between time spent engaged with digital media and worsened youth mental health outcomes (Granic et al. 2020). There is also some evidence to suggest that some mental health conditions, such as anxiety and depression, may exacerbate the use of digital media by youth (Ryding and Kaye 2017; Starcevic et al. 2018). Engaging in digital spaces like the metaverse is associated with worse youth mental health outcomes when teens and young adults display problematic behaviors akin to addiction, such as losing sleep due to their engagement (Glover et al. 2022; Royal Society for Public Health 2017), avoiding academic and social responsibilities (Glover et al. 2022), and interacting “passively” (e.g., browsing, scrolling, and reposting content) versus actively (e.g., chatting, sharing and posting original content) for a small subset of youth (Glover et al. 2022; Thorisdottir et al. 2019).

### 2.1.3 Exposure to Harmful Content

As a metaverse is an immersive digital world that allows for individuals to create content and quickly distribute such content to a variety of users, there is considerable potential for content that may harm youth mental health to be created and quickly distributed (Bell 2022). For this review, harmful content refers to forms of digital content (i.e., video, image, or text content from individual or corporate users) that are thought to be negatively associated with youth mental health upon exposure. Here are some examples of harmful content and the negative interactions youth may have with such content:

- A young person interested in fitness and healthy recipes frequents Instagram for inspiration. Over time, they get exposed to more and more content lauding individuals with thin, perfectly toned bodies, and they start to feel that they are not healthy enough if they don’t resemble the images they see online. Their new desire for thinness leads them to search for and discover content on these same social media platforms that promote disordered eating behaviors.

- A teen struggling with difficult life circumstances begins to follow social media accounts that post depression and other content related to mental health struggles that they feel resonates with them. Over time, one of these accounts begins to reblog posts that contain depictions of
self-harm in graphic detail. When the teen’s life circumstances become worse, they try one of the methods of self-harm described from the blog posts. When the method of self-harm reduces their emotional pain, the teen continues to use it when they get distressed.

In the following section we survey the literature surrounding four topics commonly thought to produce harmful content: eating disorders, self-injurious thoughts and behaviors, addictive substances, and sexually explicit material. Each subsection will examine what harms, if any, exposure and interactions with other users regarding content on this topic may pose for youth mental health, along with potential benefits, if any, that may come from exposure to this topic.

### 2.1.3.1 EATING DISORDERS

Metaverse-like spaces can be breeding grounds for content that promotes thoughts and behaviors associated with eating disorders in youth. Eating disorders are characterized by deep dissatisfaction with bodily appearance (normally a desire for thinness) and a variety of unhealthy behaviors individuals engage in to cope with bodily dissatisfaction, including food restriction, overexercising, and purging (Haines and Neurmark-Sztainer 2006). Many governing health agencies, such as the American Academy of Pediatrics (Collins et al. 2017), warn that social media networks can promote the unhealthy thoughts and behaviors involved in developing and maintaining eating disorders (Chassiakos et al. 2016). Likewise, there is a prominent concern in the clinical literature that unregulated and underregulated social media content may provoke disordered eating behaviors in youth, especially adolescent girls (Kim and Lennon 2007; Primack and Escobar-Viera 2017; Custers 2015).

Exposure to disordered eating in peers may lead youth to endorse beliefs and engage in behaviors related to eating disorders (e.g., drive for thinness, body dissatisfaction, fasting, purging, binge eating, etc.) (Quiles Marcos et al. 2013; Keel and Forney 2013). Eating disorders may partially develop and be maintained from social contagion, where youth unintentionally kickstart the development of eating disorders in their peers through modeling dieting practices, attitudes promoting thinness, teasing others about their weight, and rewarding weight loss (Allison et al. 2014). Given the easy access to peer content related to the above online, a metaverse may support the development of eating disorders in youth (Primack and Escobar-Viera 2017). Content that promotes disordered eating behaviors in the pursuit of thinness, including but not limited to other youth with almost superhumanly thin bodies, congratulatory weight loss videos, dangerous exercise routines, and videos that provide instructions on food restrictions, are spread via hashtags and easily accessible to youth (Sukunesan et al. 2021; Custers 2015). Similarly, content that promotes muscularity that is widely available online, such as fitspiration content and images of peers and celebrities with highly muscular bodies, may also encourage muscularity-oriented disordered eating, particularly among boys (Ho et al. 2016; Rodgers et al. 2020; Roberts et al. 2022).
While moderating eating disorder content on metaverse-like platforms through hashtags can be difficult (Chancellor et al. 2016; Gerrard 2018; Sharp and Gerard 2022), one design feature that could be targeted to reduce over-exposure to eating disorder content is the design of algorithms that manage search and content. Personalized content-curation algorithms on social media networks like TikTok (#foryou) could potentially lead to eating disordered behavior in adolescents (Logrieco et al. 2021). If a teen or young adult views a video on anorexia on their homepage, the algorithm will continue to suggest similar videos, contributing to obsessive behavior (Logrieco et al. 2021; Gerrard 2018). These kinds of algorithms are either not able to distinguish harmful content from harmless content, or can allow some videos of harmful content to slip through (Logrieco et al. 2021; Gerrard 2018). This can occur for several reasons, such as challenges of incorporating accurate measurement practices (e.g., differentiating self-starvation from fasting and portion control) into machine learning models (Wang 2021) and platform features promoting persistent “rabbit hole” use toward more extreme and less monitored content (Harriger et al. 2022). These programming oversights can even result in the continued exposure of eating disorder content to youth who are the most at risk; one study found that 55% of teens being treated for eating disorders continued to be proposed pro-ana content by TikTok, with their main way of searching for pro-ana videos being #foryou (Pruccoli et al. 2022).

Exposure to harmful eating disorder content also happens within eating disorder-centric communities online. Oksanen and colleagues (2015) mention that all social media sites have active pro-anorexia communities (pro-ana), where youth engaging in eating disorders share content related to disordered eating and form close social bonds with each other. These bonds allow for youth at risk for or currently experiencing eating disorders to continue to be exposed to thoughts and behaviors from other youth that encourage disordered eating, contributing to the maintenance of their own eating disorders (Custers and Van del Bulck 2009; Rodgers et al. 2016). These social bonds can alienate youth from adults and physicians who try to stop their disordered eating behaviors (Oksanen et al. 2015). As such, eating disorder communities online also contribute to the spreading of harmful eating disorder content within these communities, harming youth already struggling with these disorders.

2.1.3.2 SELF-INJURIOUS THOUGHTS AND BEHAVIORS

Metaverse-like spaces may also expose youth to content discussing and displaying self-injurious thoughts and behaviors. Self-injurious thoughts and behaviors (SITB) refers to the full spectrum of suicidal and non-suicidal thoughts and behaviors, including self-injurious thoughts (i.e., thoughts of harming oneself with no intent to die), self-injurious behaviors (i.e., harming oneself with no intent to die), suicide threats with no intent to die, suicidal ideation, suicide plans, and suicide attempts (Nock 2010). Some researchers have raised concerns that youth who are over-exposed to SITB content
online may be more at risk for engaging with SITB themselves (Biernesser et al. 2020). Systematic reviews on social media’s impact on SITB in youth have found that exposure to SITB content on social media was associated with youth glorifying and normalizing SITB, in addition to developing other maladaptive behaviors that may increase long-term risk for SITB (Biernesser et al. 2020; Dyson et al. 2016; Marchant et al. 2017; Nesi, Burke, Bettis et al., 2021). Notably, these relations appear to be the most relevant for youth who are already experiencing some level of SITB, as well as adolescent girls and LGBTQ+ youth (Biernesser et al. 2020; Dyson et al. 2016; Marchant et al. 2017; Nesi, Burke, Lawrence et al. 2021).

Much like with eating disorders, there is a robust network of users and SITB-centered blogs online that facilitate the sharing of SITB content, often through photos and videos (Rodham et al. 2013). Even blogs for other mental health conditions online, such as depression, may occasionally share or repost graphic images and video clips of SITB (Cavazos-Rehg et al. 2017). The easy sharing of SITB content in these communities allows for adolescents on social media to easily search for information on how to self-harm without having to have face-to-face conversations that may reveal their intentions to parties that can help them (Cash et al. 2013; Lewis et al. 2012; Lewis and Baker 2011; Lewis et al. 2011).

The same algorithmic programming insights that enable the continued curation of eating disordered content also apply to SITB content (Rodham et al., 2013). Algorithms that have challenges differentiating terms that generate harmful content from similar terms that do not generate harmful content, and that “rabbit hole” users into viewing more extreme and less monitored content, could lead to youth being unintentionally exposed to SITB content or being shown very graphic SITB content (Wang 2021; Harriger et al. 2022).

Metaverse-like spaces may also facilitate interactions between youth struggling with SITB that increase their exposure to SITB content and teach them thoughts and behaviors that facilitate and encourage engagement in SITB. Suicidal youth tend to be more closely connected with other suicidal youth on social media platforms, such as Twitter, than youth in general, giving them the capacity to share new SITB-related content amongst each other quickly (Colombo et al. 2016). Youth struggling with SITB can receive harmful advice from other users, such as on how to engage in self-harm covertly or new methods for self-harm (Cavazos-Rehg et al. 2017). Engagement with communities centered around SITB may also further normalize self-harm and encourage youth to continue engaging with SITB (Jacob et al. 2017).

2.1.3.3 SUBSTANCE ABUSE

Metaverse-like spaces also have the potential to expose youth to content that may encourage consuming addictive substances, most commonly alcohol, to the point of
substance abuse. Substance abuse occurs when a youth consumes addictive substances such as alcohol or other drugs to the point where it causes notable impairments in their daily functioning (Kim et al. 1998). About 5% of teens and 15% of young adults qualify for substance use disorder, with the use of most substances more broadly declining or remaining steady since the pandemic (Substance Abuse and Mental Health Services Administration 2017; National Institute on Drug Abuse 2022). Even for those who do not meet the qualifications for substance use disorder, teens and young adults may still consume substances such as alcohol to the point of impairment in many major areas of their lives, including academics, relationships with peers, and even additional adverse mental health outcomes (Popovici and French 2013; Wechsler et al. 2000). A growing number of researchers are expressing concerns that exposure to risky substance use behaviors, and advertisements for alcohol and other addictive substances in metaverse-like spaces on social media, may lead to increases in substance use-related problems in adolescents and young adults (Moreno and Whitehill 2014; Costello and Ramo 2017). While substance use-related problems can be considered a mental health concern in and of itself, excessive substance use in youth has also been associated with depression and anxiety (Rao 2006; Ohannessian 2014). While there are many types of substance abuse, this review will focus on patterns involved in alcohol abuse as a model for how metaverse-like spaces may impact substance abuse and youth mental health, given the high prevalence of studies on this subject.

Some evidence indicates that youth exposed to peers’ drinking behaviors on social media may be more at risk for alcohol abuse. Nesi and colleagues (2017) found that high schoolers who had seen photos of their peers with alcohol or had peers who made posts about drinking on social media were more likely to experience heavy episodic drinking one year later than their peers who had not seen such content online. A review of 19 studies on the relationship between alcohol-related problems and social media use in youth found that engagement with alcohol-related social media content in adolescent and young adult populations was modestly related to alcohol consumption and alcohol-related problems (Curtis et al. 2018). Notably, these studies did not provide concrete explanations for why this relationship exists. One proposed explanation for this relationship is that youth may become more motivated to consume alcohol after seeing posts on social media from friends and other peers glamorizing drinking (Tucker et al. 2013).

Another proposed explanation is that social media advertisements for alcohol may be targeting adolescents and young adults. McCreanor and colleagues’ (2013) review of youth exposure to alcohol on social media notes that social media companies may engage in many strategies online that appeal to adolescents and young adults and may encourage excessive drinking behaviors, including competitions, interactives games, celebratory posts on holidays and other special occasions that encourage drinking,
Advertisements may also be specifically targeted towards vulnerable youth, often through peer-to-peer marketing (WHO 2022). Indeed, there is some evidence that not only have young adults interacted with alcohol advertisements on multiple channels, but also those who interact more with alcohol advertisements through digital channels are more likely to have high episodic drinking (Critchlow et al. 2016). The practices are also found in other studies on drug use and social media in youth; advertisements for e-cigarettes on social media predicted lower risk perception of e-cigarettes, and subsequently more e-cigarette use (Zheng et al. 2021).

Other authors have noted that alcohol advertisements on social media and other digital platforms may not be adequately regulated by these platforms and have a considerable chance of reaching underage youth (Mosher 2012). While some social media platforms may restrict advertising alcohol to users who are underage in their country, many social media platforms lack adequate age verification practices or have practices that are easy to circumvent (Dunlop et al. 2016; Barry et al. 2015; Winpenny et al. 2014).

### 2.1.3.4 SEXUALLY EXPLICIT MATERIAL

Exposure to sexual content online is a concern for many caregivers, educators, and other authority figures (Dombrowski et al. 2007; Wolak et al. 2007; Gassó et al. 2019; Koletić et al. 2019). Studies suggest that somewhere between 19% to 32% of adolescents are unintentionally exposed to pornographic materials on the internet (Mitchell et al. 2003; Hardy et al. 2013; Hornor 2020; Ybarra et al. 2009). Even more youth intentionally engage in sexual content online (Ybarra et al. 2011). Although prevalence rates vary among studies, national and international studies reveal that online pornography use is common among boys and not uncommon among girls (Collins et al. 2017).

However, little work has examined the effects of exposure to sexual content online on the mental health of adolescents and young adults (Hornor 2020; Livingstone and Smith 2014). While some youth may be exposed to unwanted sexual content online (roughly 25-34%), studies have found only a small amount of those youth reported distress from the exposure (roughly 6-9%) (Mitchell et al. 2003; Wolak et al. 2006; Livingstone and Bober 2004). Most of those who were upset from the exposure appeared to be adolescent girls aged 9-12 (Livingstone and Bober 2004). As such, there is a lack of strong evidence that viewing unwanted pornography has a negative impact on the mental health of most adolescents. When relationships between sexually explicit materials and negative mental health outcomes are found, they often involve negative interactions with other users, such as non-consensual sharing of sexually explicit images, being sent unwanted nudes, solicitations for explicit images by other users, and other forms of cybersexual harassment and victimization, rather than the accidental or intentional viewing of sexually explicit materials themselves (e.g. Gassó et al. 2019; Razi et al. 2020; Patel et al. 2022).
Youth may also intentionally seek out sexual content online. While many adults are concerned about teens viewing pornographic content, youth report seeking sexually explicit materials online for relatively neutral or even positive reasons, including boredom, stress relief, curiosity, and sexual identity development (Attwood et al. 2018; Razi et al. 2020). For some youth, accessing sexually explicit materials online is the only way to receive relevant sexual education, especially for gender and sexual minority youth (Litsou et al. 2021). Similarly, youth are also aware of the shortcomings of pornography at depicting accurate sexual information, indicating that the depiction of youth as being blindly manipulated by pornographic materials they choose to view is unrealistic (Litsou et al. 2021). As such, watching pornographic materials online should not be seen as a wholly negative experience.

Research regarding negative mental health effects from intentionally seeking sexual content is equally as inconclusive as from unwanted exposure. Studies that aim to link pornography use to adverse mental health outcomes in youth have shown very mixed findings (Stulhofer et al. 2022). Some studies reported links between abstaining from pornography in adolescence and greater life satisfaction and less dysfunctional pornography use, and lower depression and anxiety in adulthood (Camilleri et al. 2021; Willoughby et al. 2018). However, these findings are complicated by mixed results for depression outcomes, a lack of investigating into moderating and mediating variables, and a lack of directionality (Willoughby et al. 2018).

Of the four categories of potential harmful content reviewed, exposure to content involving eating disorders, SITB, and addictive substances was found to be associated with negative mental health outcomes in youth. Content in these categories is often originally shared by peers and companies and further distributed through platform structures to youth to the extent where they may begin to emulate some of the thoughts and behaviors presented to them. In contrast, exposure to sexually explicit content does not appear to be a prominent risk for youth mental health unless if in the context of sexual harassment, coercion, and abuse.

2.2 Social Risks

One of the defining qualities of a metaverse is its function as a social space for users to connect with one another. However, participating in metaverse-like spaces can expose youth to risks of problematic encounters such as cyberbullying, sexual harassment, and hate speech. Other potential risks include social comparison and the fear of missing out. This section will explore how these social risks play a potential role in youth mental wellbeing. Here are some examples of how these risks might appear:

- A young person receives a string of unsolicited messages and pornographic images from an anonymous TikTok user. They feel a sense of helplessness, hopelessness, and fear despite having blocked the user, as new messages continue from a new profile. The young person feels that there
is no escape from the harassment without leaving the platform, but does not want to miss out on opportunities to continue connecting with friends online, so maintains their account on the platform.

- A teen makes a post on Reddit asking for advice for dating as a trans girl. While the comments on her post are mostly supportive, she is DM’d by strangers telling her that she is not a real girl so no one would want to date her and that she is just confused and needs God. The teen feels hurt after reading these comments, even though the moderators in the subreddit were supportive and banned the harassing accounts. She begins to feel like no one would really date her and closes herself off from developing a romantic relationship.

2.2.1 Distressing Encounters

When youth engage in online social spaces, they are risking exposure to distressing interpersonal encounters that could have detrimental impacts to their mental wellbeing. These encounters could be extensions of their offline social circles, contacts made online, or even from anonymous sources. This section will examine the body of research around interpersonal risks such as cyberbullying, sexual harassment, toxicity in online games, hate speech, and extremism that youth may encounter as they participate in metaverse-like spaces.

2.2.1.1 Cyberbullying

Research suggests strong links between cyberbullying and negative mental health outcomes for youth (Hinduja and Patchin 2017; Schneider et al. 2012; Bottino et al. 2015). Cyberbullying is defined as “an aggressive, intentional act carried out by a group or individual, using electronic forms of contact, repeatedly and over time against a victim who cannot easily defend himself or herself” (Smith et al. 2008). Although it is less prevalent than traditional bullying (Kowalski and Limber 2013), cyberbullying affects 10-20% of all adolescents (Bottino et al. 2015) and disproportionately impacts females and marginalized groups (Fryling et al. 2015; Ballard and Welch 2017; Marciano et al. 2020). For example, a study of massively multiplayer online games (MMOGs) found that males were far more likely to cyberbully other players, with females and LGBT+ players being far more likely to be victims of cyberbullying (Ballard and Welch 2017).

Cyberbullying can be more difficult to avoid and more likely to come from anonymous sources than other more traditional forms of bullying (Li 2007). A study from the American Public Health Association found that adolescents experiencing frequent cyberbullying (two or more times per month) are more likely to exhibit depressive symptoms than others (Schneider et al. 2012). Cyberbullying has also been linked to severe mental health outcomes such as substance abuse, suicidal ideation, and self-injury (Hinduja and Patchin 2017; Litwiller and Brausch 2013; Bottino et al. 2015; Biernesser et al. 2020; Nesi et al. 2021; Zhu et al. 2022). A meta-analysis of several longitudinal studies also found that experiencing depression and anxiety increases the
likelihood of becoming a victim of cyberbullying for youth (Marciano et al. 2020).

Litwiller and Brausch (2013) suggest that feelings such as low self-esteem, anxiety, and depression could cause youth to seek substances such as drugs or alcohol as a means of coping with negative feelings resulting from cyberbullying. Additionally, a large-scale study of over 1,900 middle-schoolers found that both cyberbullies and victims of cyberbullying were more likely to attempt suicide (Hinduja and Patchin 2017). Several studies also suggest an overlap between bullying, cyberbullying, and symptoms of depression (Schneider et al. 2012; Chang et al. 2013; Bottino et al. 2015), with one systematic review on the topic noting that “cyberbullies and cybervictims reported more experiences with school bullying, suggesting continuity and overlapping between traditional bullying and cyberbullying” (Bottino et al. 2015).

**SEXUAL HARASSMENT**

Barak (2005) identifies four types of online sexual harassment: active verbal (direct communication from harassers), passive verbal (sexually offensive user names or public posts), active graphic (harasser sends unsolicited pornographic image), and passive graphic (pornographic imagery posted in unexpected places). These forms of sexual harassment can take place in many metaverse-like spaces such as social media (Duncan et al. 2019), online games (Tang et al. 2020), and social VR spaces (Maloney et al. 2020). Harassment occurring in spaces that utilize VR can be especially traumatic due to the technologically enhanced sense of presence (Wiederhold 2022).

Studies report offline sexual harassment occurring more frequently than online (Hill and Kearl 2011; Mitchell et al. 2014; Ståhl and Dennhag 2021). Far greater sexual harassment victimization has been reported among girls than boys in both contexts (Zetterström Dahlqvist and Gillander Gådin 2018) with most frequent rates of reporting among the LGBTQ+ community (Mitchell et al. 2014). Findings of a longitudinal study supports the co-occurrence of offline and online sexual harassment, suggesting overlapping behaviors between the two (Leemis et al. 2019).

Sexual harassment is also a major concern associated with participation in a metaverse-like environment. Multiple studies have found significant links between online sexual harassment, anxiety, and depression among adolescents, with notably more severe outcomes associated with girls (Bucchianeri et al. 2014; Bendixen et al. 2018; Zetterström Dahlqvist and Gillander Gådin 2018; Duncan et al. 2019; Ståhl and Dennhag 2021). Online sexual harassment has also been associated with more severe outcomes such as substance abuse, self-harm, and suicidal ideation (Bucchianeri et al. 2014; Zhu et al. 2022). Other connections are also supported, as seen in a study of 2,869 high schoolers linking sexual harassment to lower self-esteem and body issues (Bendixen et al. 2018).
2.2.1.2 HATE SPEECH AND RELATED ACTIONS

Hate speech is another popular concern for youth mental health within metaverse-like spaces. Although there are some disagreements, many definitions of hate speech agree that it consists of publicly making statements that threaten, ridicule, or hold in contempt individuals because of some variety of demographic characteristics (Titley et al. 2014; Anti-Defamation League 2018). These demographic characteristics tend to include an individual’s race, skin color, nationality, ethnicity, faith, and sexual orientation (Titley et al. 2014). Many metaverse-like spaces that youth engage with, including online gaming communities, streaming platforms, and social media, have large numbers of users who engage in hate speech (Alava et al. 2017; Lakhani 2021; Gray 2012b). Here are some examples of how hate speech in online spaces might appear, in both overt and covert forms:

- A Muslim youth is streaming themself playing a popular video game on Twitch. While they are streaming, another individual enters the stream from browse mode and writes Islamophobic comments in the stream’s chat. The youth reads those comments, becomes distressed, and ends the stream.

- A queer youth of color is playing GTA V roleplay, a multiplayer mod for the PC version of the popular video game Grand Theft Auto V. Other players harass them in a variety of ways, without using conventional verbal slurs. For example, their character is killed, kidnapped, or surrounded and harassed physically in game, to the point that the youth becomes distressed and logs off. They begin to feel hopeless and helpless about being able to interact with others online and not encounter hate and harassment.

Preliminary evidence suggests that being a victim of hate speech online is associated with worsened mental health in youth. Online hate speech victimization has been found to be associated with increased depression symptoms in a large study of Spanish adolescents, almost as strong as the relationship between cyberbullying and depression found in other studies (Wachs et al. 2022). Exposure to hate speech in college subreddits was associated with greater expression of stress on these platforms in college students (Saha et al. 2019). Larger surveys on hate speech in online spaces have also found that 11% of users report experiencing depressive or suicidal thoughts as a result of being victims of hate speech (ADL 2021). As such, hate speech victimization may be related to increases in depression, stress, and suicidal thoughts for those victimized. However, more studies are needed that focus specifically on hate speech victimization in youth to further define its impact on their mental health.

It is very common for youth to witness or be victims of hate speech in metaverse-like spaces. 82% have witnessed hate speech on metaverse-like platforms and 64% have
reported being victims of hate speech themselves (Kowert and Cook 2022). One large survey found that 41% of youth reported watching or reading material that featured hate speech from extremist groups online (Grizzle and Perez Tornero 2016). Other studies focusing on hate speech victimization in young adults have found estimates as low as 20% and as high as 70% for young adults who have been victimized (ADL 2021; RESET and Pollytix 2021). In particular, youth who are LGBTQ+, Muslim, Jewish, racial/ethnic minorities, or female may be at an increased risk of experiencing hate speech victimization (ADL 2021; ADL 2022). Frequent experiences of hate speech online may lessen young people’s faith in individuals offline, realizing that some of them may want to call them slurs and engage in other hateful speech, but are unwilling to do so unless anonymous (Gray 2012b).

Non-verbal hate actions may also be a concern, particularly in online gaming. Youth in MMORPGs who have been suspected by other players as being of a non-White race/ethnicity through vocal cues or avatar appearance can be subject to griefing, or when other players go out of their way to ruin the game experience (Gray 2012a, Gray 2012b). The anonymity of players and a lack of effective moderation may lead to bolder acts of discrimination, slurs, and aggressive harassment than youth could be subjected to offline (Gray 2012a). Some metaverse-like platforms such as Roblox allow for users to create virtual worlds for the purpose of engaging in hate-based actions, such as Nazi roleplay games and death camps, or build spaces that perpetuate harmful stereotypes about marginalized groups (Kou and Gui 2023). Spaces may also be used to organize “raids” where users spam other servers (commonly pro-LGBTQ+) with offensive and discriminatory content (Gallagher et al. 2021; ADL 2022). While the specific mental health impacts of youth experiencing hate actions in metaverse-like spaces has not been as well researched as hate speech, witnessing or being the victim of hate crimes in offline spaces is related to worsened mental health for youth (Njoroge et al. 2021).

2.2.1.3 EXTREMISM

Extremism is another concern within the metaverse that may be related to youth mental health vulnerabilities. The radicalization of youth by far-right extremist groups in metaverse-like spaces, especially within gaming servers (e.g. Minecraft) and game-adjacent platforms (e.g. Discord), is becoming increasingly common (Alava et al. 2017; UNESCO 2017; Gallagher et al. 2021; Lakhani 2021; ADL 2022; Koehler et al. 2023; Olazola Rosenblat 2023). On these platforms, extremist groups strategically carve out their own spaces to increase familiarity and attractiveness of their propaganda to youth (Schlegel and Amarasingam 2022). While these groups normally target youth who already agree with many aspects of their ideology, other youth may be “rabbit hole-d” into extremist groups by being curious about their ideology (Schlegel and Amarasingam 2022). In particular, a growing number of parents, educators, and organizations are concerned that youth struggling with mental health issues may be more vulnerable to
being radicalized by extremist groups online (Vermeulen et al. 2022).

Despite this reasonable concern, the current empirical work seeking to determine if youth struggling with mental health issues are vulnerable to radicalization in metaverse-like spaces is very limited. No current studies have done longitudinal research examining if youth with mental health issues are more likely to be radicalized online than those who do not have mental health issues. Among literature examining the relationship between youth mental health issues and extremism, a recent review found mixed evidence that psychological vulnerabilities are related to youth mental illness (Harpviken et al. 2020). Notably, the studies included in this review were small in number (less than 6), focus on different forms of mental health concerns (e.g., ADHD, anxiety/depression, SITB, and self-disclosure of mental health concerns generally), and do not focus on radicalization in online spaces (Simi et al. 2016; Bjørgo 2005; Ellis et al. 2016; Coid et al. 2016; Bhui et al. 2014a). Additionally, some of these studies do not use gold standard clinical interviews to do so and attempt to make diagnoses “post hoc,” a major methodological concern common among studies attempting to link extremism and mental health (Simi et al. 2016; Bjørgo 2005; Trimbur et al. 2021). Other studies have expressed a small relationship between some symptoms of mental disorders and extremist ideologies/sympathy to political violence and terrorism (Rousseau et al. 2019; Ellis et al. 2016; Bhui et al. 2014b; Trimbur et al. 2021), but even these findings are mixed (Bhui et al. 2014a; Coid et al. 2016; Victoroff et al. 2010; Trimbur et al. 2021).

Despite this, there are some reasons to believe that some youth experiencing mental health issues might be vulnerable to radicalization within metaverse-like platforms. Some scholars have theorized that extremist conspiracy theories, such as QAnon, may provide believers with a sense of identity and moral purpose that helps them to cope with anxiety and uncertainty, followed by anxious uncertainty leading to defensive reactions that maintain ideological extremism when challenged (Fitzgerald 2022; Nash et al. 2011). Furthermore, loneliness, a risk factor and "symptom" of many mental disorders, may interact with identity fusion in gaming culture to increase support for extremist behavior (Richardson et al. 2017; Kowert et al. 2022). Relatedly, social isolation, a common experience for those suffering from mental disorders, has been stated to be a risk factor for being radicalized (McCauley and Moskalenko 2011; Loades et al. 2020).

As such, there are reasons to suspect that youth struggling with their mental health are more vulnerable to extremism online. However, stronger empirical work that directly connects mental health issues in youth with online extremism is called for. Future studies would benefit from stronger measures of symptoms of mental disorders, longitudinal methodologies, and a focus on youth in online spaces to determine how great the risk for radicalization is for youth with mental health issues. Such work should also attend to if sustained exposure to radicalized messages has an impact on
mental health. It is crucial that more work focus on the links between youth mental health and extremism, given the increased prevalence of extremism (both violent and non-violent) in online spaces occupied by youth, as well as the mental health risks hate speech and hate-related actions created by extremism pose for youth in these spaces. This will allow us to understand not only the role youth mental health plays towards vulnerability of being radicalized, but also how becoming radicalized by these groups may impact youth mental health.

2.2.2 Social Comparison

Negative social comparison is a common mechanism through which social media has been found to exert a negative effect on youth mental health, rightly a leading concern associated with metaverse-like spaces. Theory on social comparison has a long history, referring to one's tendency to look to other individuals as sources of information of how to act, think, and feel (Festinger 1954). As a result, one is able to assess the abilities, social standing, and performances of themselves and those around them (Festinger 1954; Verduyn et al. 2020). Social comparison takes place in online social spaces as well, most notably social media, and has been associated more strongly to passive use of social media (Appel et al. 2016). Within these contexts, users make upward comparisons (comparing oneself to those perceived as superior) and downward comparisons (comparing oneself to those perceived as inferior) to other users on the site (Wills 1981; Wood 1989; Vogel et al. 2014). Social comparisons, in particular involving the evaluation of “how popular” one is, may be of particular importance during adolescence when brain circuitry involving social interactions and rewards is particularly sensitive (Chiu and Chien 2022).

Evidence indicates that adolescents who are active on platforms such as Facebook are more likely to engage in “upward” social comparison (Vogel et al. 2014; Dodemaide et al. 2022). And while there are circumstances where upward comparison can be motivating for self-improvement (Verduyn et al. 2020), it is more commonly linked to negative comparisons and poor wellbeing (Dodemaide et al. 2022). This is reflected in an experimental study of 128 undergraduates that found that participants exposed to user content reflecting high activity on social media were more likely to report lower self-evaluations (Vogel et al. 2014).

Social comparison on social media has been linked to poor mental wellbeing, including lower self-esteem and depressive symptoms (Vogel et al. 2014; Nesi and Prinstein 2015; Appel et al. 2016). It has also been linked to other associated risks, such as poor body image and eating disorders (Ho et al. 2016; Lewallen and Behm-Morawitz 2016). A preliminary study of 619 eighth and ninth grade students found links between social comparison and feedback seeking (SCFS) on social media and depressive symptoms (Nesi and Prinstein 2015). A review of literature suggests that Facebook “encourages unfavorable social comparisons and envy, which may in turn lead to depressed mood” (Appel et al. 2016). Other research indicates a possible cyclical effect, where upward social comparison on social media leads to low mental wellbeing, which in turn results in further negative comparisons (Abi-Jaoude et al. 2020; Verduyn et al. 2020). However, despite providing
robust support for their correlation, the current literature still falls short of establishing that social comparison causes poorer mental wellbeing (Nesi and Prinstein 2015; Appel et al. 2016).

Social comparison takes place in other metaverse-like spaces as well. For example, one experimental study explored social comparison in VR, finding that lowering a user’s height in VR was associated with more negative self-evaluations (Freeman et al. 2014). Research has also suggested that members of virtual worlds like Second Life engage in comparing avatars (Kim et al. 2007). Social comparison has also been examined in design features of online social games (such as leaderboards to view the scores of others), which has been associated with increases in gaming continuance (Esteves et al. 2021). However, research linking social comparison in online games to mental wellbeing remains scarce, and is worth further exploration in the future.

2.2.3 Fear of Missing Out

When discussing social aspects in online spaces, it is worth including the fear of missing out (FoMO), a phenomenon described as “a pervasive apprehension that others might be having rewarding experiences from which one is absent” (Przybylski et al. 2013). For adolescents, who are particularly sensitive to social pressures due to developing reward and social circuitry within their brains, FoMO may be particularly detrimental to their mental health (Chiu and Chien 2022). Social media users have expressed habits reflecting FoMO ingrained within social media site designs and some games, such as prompting fears of missing timely interactions, fear of losing popularity, and fear of missing a valuable opportunity (Alutaybi et al. 2019).

Studies have linked FoMO to decreased mental wellbeing such as social anxiety and clinical depression (Reer et al. 2019; Gupta and Sharma 2021), hypothesizing that leaps in technology—namely social media platforms—have impacted social norms through an amplified sense of FoMO among adolescents. Results from a survey study of 402 adolescent users of Facebook indicate that an increased sense of FoMO is linked to a stronger desire for belonging and popularity, prompting an overall increase in Facebook use (Beyens et al. 2016). The study also found that users experienced increased perceptions of stress when they don’t feel a sense of belonging or popularity among their peers on the social media platform. A survey study of teens and young adults also linked high social comparison to feelings of FoMO (Reer et al. 2019).

Certain social media apps such as WhatsApp and Instagram have been identified to be more prone to prompting more severe “disorder” like behavior associated with FoMO in younger users (Rozgonjuk et al. 2020). It should be noted, however, that while a connection between FoMO related social media use and mental health outcomes has been identified, data remains inconclusive on a direct causal relationship. The relationship between FoMO and negative mental health outcomes may be better understood as one that mediates elements such as the user’s social needs and time spent on social media sites.
2.2.4 Isolation from Friends and Family

Several studies have linked participation in metaverse-like spaces to feelings of isolation and loneliness (O’Day and Heimberg 2021; MacDonald et al. 2022; Lawrence et al. 2022; Shannon et al. 2022). However, these linkages tend to be either negligible or have greater complexity to the relationship between variables (Reer et al. 2019; O’Day and Heimberg 2021). One review on the subject, for example, suggests that increased social media use does not necessarily indicate a direct causal relationship to social isolation, but rather that lonely individuals are more likely to use social media in ways harmful to themselves (e.g., problematic/addictive social media use) and harmful to others (e.g., endorsing extremist beliefs) (O’Day and Heimberg 2021; Kowert et al. 2022).

The exception appears to be in circumstances where users develop maladaptive patterns of behavior around their media use. For example, one survey study on Korean adolescents suggests a positive correlation between internet addiction and hours spent playing games, with high risk users being more likely to display characteristics such as asocial behavior, non-assertiveness, sexual avoidance, and a distrust of others (Seo et al. 2009). The category of high-risk internet users, however, represented only 3% of the study’s 676 sample size. Other studies have linked increased social and emotional loneliness among high schoolers and young adults to more severe cases of maladaptive internet and smartphone use (Yayan et al. 2019; Aslan 2022). These findings are reinforced in a meta-analysis that reports strong links between problematic social media use and outcomes of depression, anxiety, and stress among adolescents and young adults (Shannon et al. 2022). The findings of these studies suggest that engagement with social media and gaming alone does not have significant negative impacts on a user’s feelings of isolation, with the exception of extreme cases reflective of maladaptive behaviors that interrupt daily functioning.

2.2.5 Privacy

Connecting with others in metaverse-like spaces also opens certain risks to privacy. While these risks often lack direct links to mental health outcomes for adolescents, they tend to be associated with other social risks such as exposure to cyberbullying and sexual harassment. More open privacy measures in spaces such as social media run the risk of the wrong content reaching the wrong people. For example, poor practices around protecting one’s online privacy have been linked to greater likelihood of being cyberbullied (Mesch 2009). Youth may also miss out on opportunities to enhance their social mobility when content intended for close friend groups is openly displayed to potential college recruiters and employers (boyd 2014). Although drawing conclusive causal links between privacy and mental wellbeing is a challenge, current research does support an inverse relationship between the two. For example, depression among users has been associated with reduced privacy settings (Radovic et al. 2017). Research also suggests that individuals with symptoms of depression or anxiety are more likely to make poor decisions that put their privacy in jeopardy (McKenna et al. 2002).

The privacy of adolescents’ and young adults’ data within these platforms may also be a concern for their mental health. Current conceptions of the metaverse include the collection of an abundance
of user data, including information about youth’s identities entered when making their accounts, all communications and interactions with other users on the platforms, and transactions (Bibri and Allam 2022). The use of immersive technologies such as VR will also cause data to be collected and stored for youth’s facial and other bodily movements, and even the spatial data such as the layout of any rooms youth are in (Fernandez and Hui 2022; Bibri and Allam 2022; Wang et al. 2023). As such, data leakages in metaverse-like spaces have the potential to reveal large amounts of highly detailed and highly personalized information that may feel violating for youth.

Privacy concerns can arise in metaverse platforms that employ personalized advertising based on user data. The intrusive nature of targeted advertisements can lead to feelings of manipulation, invasion of privacy, and heightened materialism. Such experiences can negatively impact individuals’ emotional wellbeing and contribute to a sense of dissatisfaction and comparison with others (Grier and Kumanyika 2010; Alves de Castro et al. 2021). Personal data could also be used to create predictive profiles for youth, which while primarily used for commercial and advertising purposes, may also be used to create privileged classes of users based on predicted social and political characteristics while marginalizing others (Bibri and Allam 2022).

Additional privacy risks include the tracking of youths’ digital footprints and physical locations by predators, corporations, or even governments (Bibri and Allam 2022; Fernandez and Hui 2022; Wang et al. 2023). Privacy concerns may lead youth to develop a fear of surveillance and reputation damage. When individuals feel that their personal data within metaverse-like platforms is not secure, they may develop a fear of surveillance and a concern about their reputation. This fear can lead to self-censorship, withdrawal from online interactions, and a heightened sense of anxiety about their digital presence and its potential consequences on their offline life and relationships (Keith and Steinberg 2017; Wisniewski et al. 2022).

Beyond the potential mental health harms of a data leak or other surveillance practices, youth may be distressed by the idea that platforms are collecting intensive private information on them. Privacy is valued by adolescents (Davis and James 2012). Youth have reported discomfort on the idea that their profiles on social media websites may be used for research without their knowledge, even when conducted by academic institutions (Monks et al. 2015; Aggosto and Abbas 2016). However, privacy concerns may be more salient for young adults than adults and less salient for adolescents than young adults (Dhir et al. 2017).

Research has also suggested that adolescents do not have the same level of awareness as adults of what private information social networking sites are tracking until age 20 (Zarouali et al. 2020). Additionally, the emotional volatility, impulsivity, and desire for peer approval may make youth more vulnerable to social media marketing practices that use personal data (Montgomery 2015). However, current adolescents’ own privacy-protective practices surrounding data collection may be better on average than adults, starting from age 14 onwards, although these practices were low overall (Zarouali et al. 2020). Older adolescents also have a strong awareness that the information
they post online could be seen by anyone and may affect their reputation, and as such are mindful of what they post online (Moreno and Whitehill 2014; Aggosto and Abbas 2016; Adorjan and Ricciardelli 2019). These older adolescents also report feeling a strong sense of personal responsibility to manage their own privacy online (Adorjan and Ricciardelli 2019). Even younger adolescents, including those endorsing that privacy management is hard, report using privacy regulation strategies such as withholding sensitive information, setting profiles to “friends only”, embedding false information into their profiles, blocking threatening users, and consulting others if uncertain what to do (Davis and James 2012; Adorjan and Ricciardelli 2019; Madden et al. 2013).

To better protect youth privacy and minimize any mental health risks associated with the invasion of privacy, privacy enhancing technologies (PETs) could be used to make some data “unintelligible” before being shared with the broader network within the metaverse (Fernandez and Hui 2022). Machine learning algorithms could be taught to respect users’ privacy preferences across a variety of contexts within the metaverse (Lee et al. 2021). However, such privacy measures may conflict with the financial and political goals of these platforms during their development (Lee et al. 2021; Bibri and Allam 2022; Brown and Pecora 2014).

**A Spotlight on Video Games**

Adolescents spend a significant amount of time playing video games (Atske 2022), and questions regarding the risks and benefits of play on their mental health are important to address. Concerns over the potential effects of video game engagement have focused most recently on the mental health of players, with large variation in recommendations from policy organizations. For example, the World Health Organization included Gaming Disorder in the International Classification of Diseases (ICD-11), emphasizing excessive play time and “escalation of gaming despite the occurrence of negative consequences” as key components (WHO 2023, 6C51.0). In contrast, the American Psychiatric Association (APA 2023) did not identify any psychiatric conditions related to video games in the Diagnostic and Statistical Manual of Mental Disorders (DSM-5). The APA does recommend Internet Gaming Disorder as a topic for further research (Kardefelt-Winther 2015). The U.S. Food and Drug Administration took a different approach entirely by approving a video game designed to support treatment of children with attention deficit hyperactivity disorder (FDA 2020). The approval of the game was the first of its kind and signaled there was some evidence that video game play had some mental health benefits.

Despite this attention on video games as a public health issue and nearly three decades of research exploring the possible links between video games and negative health outcomes, no consensus has been reached.
This is due primarily to a lack of reliable, reproducible, and ecologically valid studies (Johannes et al. 2021). There are a few exceptions. Elson and Ferguson (2014) reviewed 25 years of experimental, cross-sectional, longitudinal, and meta-analytical research on violence in digital games and found that “empirical evidence regarding the impact of violent digital games on player aggression is, at best, mixed and cannot support unambiguous claims that such games are harmful or represent a public health crisis” (Elson and Ferguson 2014:1). A follow up study that re-evaluated experimental evidence on video games with aggressive content and aggressive behavior in youth reached a similar conclusion: “experimental studies were unable to demonstrate compelling short-term effects of aggressive game content on aggression” (Drummond et al. 2020:1). Johannes and colleagues conducted research on player wellbeing, motivations, and need satisfaction during play, combining survey data from players with telemetry data shared by two game companies (Electronic Arts and Nintendo of America). This approach was novel as most previous studies lacked the integration of actual player data, relying on self-reported player behavior instead. Johannes and colleagues found a small positive relation between game play and affective wellbeing (Johannes et al. 2021).

From a gender perspective, although boys tend to engage in gaming more frequently, the relationship between spending time gaming in metaverse-like spaces and depression has been observed in only a few studies, with most studies finding no association or only a small subset of boys being affected (Van Rooij et al. 2014; Padilla-Walker et al. 2020; Ohannessian and Vannucci 2020; Desai et al. 2010; Ferguson 2015; Kutner and Olsen 2008; Jones et al. 2014). Interestingly, moderate gaming levels have been associated with higher wellbeing and lower anxiety and depression among boys (Ohannessian 2009; Kutner and Olsen 2008; Jones et al. 2014; Ohannessian 2018). Boys seem only to be at risk for negative mental health effects from gaming when experiencing problematic/addictive internet or gaming use (which is a slightly more common and severe issue for boys) (Kim et al. 2018; Baloğlu et al. 2020; Müller et al. 2014). In the case of adolescent girls, increased gaming time has been linked to higher levels of internalizing symptoms like depression and anxiety (Ohannessian 2009; Ohannessian 2018; Desai et al. 2010).

**Multiplayer Games**

While many younger children play single player games, many of them designed with specific educational outcomes in mind (Gee 2004; Salen 2007; Butler 2017; Ibáñez and Delgado-Kloos 2018), children in early adolescence are ready for multiplayer experiences that allow them to play and socialize with friends online (Kafai et al. 2013; Du et al. 2021). Games like Roblox, Minecraft, and Fortnite are often referred to as precursors to the metaverse: persistent virtual worlds populated by groups of simultaneous users that include easy-to-use tools and give players the freedom to create things and set their own goals. There is some evidence that participation in multiplayer game environments has positive effects on youth wellbeing. These include outcomes like increased
sense of agency and belonging (Muriel and Crawford 2020; Vella et al. 2019), development of social emotional skills and habits (Schrier and Farber 2021; Mukund et al. 2022; Salen Tekinbaş et al. 2021), and activist stances (Cortez et al. 2022). However, a growing body of research also points to potential negative effects for some players, specifically around harmful conduct, such as harassment and bullying (Fox et al. 2018; Türkay et al. 2020). According to related research, harm in games follows a cyclical pattern in which the roles of victims, bystanders, and perpetrators are constantly changing. This means that individuals can switch between these roles, and both those who experience harm and those who witness it often become perpetrators themselves (Ballard and Welch 2017; Tang et al. 2020; Fryling et al. 2015). This cycle of harm leads to players engaging in harmful conduct more routinely (Hilvert-Bruce and Neill 2020; Beres et al. 2021). Adolescents, particularly those in their early teen years, may emulate the norms and behaviors of those individuals they see as successful or influential in the game, learning to replicate behaviors that can cause harm.

User Generated Virtual Worlds

Recent research on user generated virtual worlds (UGVW) identifies potential harms associated with metaverse-like spaces where users generate a large percentage of the content in the world. In platforms like The Sandbox or Roblox, developers provide players with easy-to-use tools and motivate them by making it possible to make money based on how much their creations are used. For example, using the tools in Roblox Studio, Roblox’s player-facing development environment, players can create, moderate, and monetize their own games or virtual worlds on the platform. Kou and Gui (2023) found that in addition to online harms related to user generated content, like UGVWs that recreate Nazi death camps (Pope 2022) and mass shootings (Brandom 2021), harm can also come from problematic incentive mechanisms, like ubiquitous microtransaction design, poor moderation, or unconstrained social design.

While increasing attention is being paid to adolescents’ engagement with video games, there is a need for more reliable, robust, and credible evidence that illuminates the influences video games have on youth mental health. Such research can be enhanced through collaboration with industry partners to obtain adequate data. In this section we look at research related to ways youth are engaging with metaverse-like spaces to give and receive social and emotional support, actively resist harmful content and behaviors, and seek refuge, connection, and care in identity-affirming online communities.
3. BENEFITS ASSOCIATED WITH YOUTHS’ SOCIAL PARTICIPATION ONLINE

Beyond a focus on risks, engagement with metaverse-like spaces also offers benefits to adolescent wellbeing. One way in which research has observed this is through youth expanding their social support within online spaces. One literature review on the topic reported strong similarities between friendships made online and in person based on qualities such as self-disclosure, companionship, and conflict resolution (Yau and Reich 2018). Another study involving 355 college students suggests that general social support and online social support are inherently linked, and using social media apps like Facebook can help users to maintain their online social support (Liu and Yu 2013). Recent work by Rideout and Fox (2018) shows that the majority of teens and tweens (87 percent) have gone online for mental health information, 64 percent have used a mobile health app, and 39 percent use the online space to seek out others with similar conditions.

While many younger children play single player games, many of them designed with specific educational outcomes in mind (Gee 2004; Salen 2007; Butler 2017; Ibáñez and Delgado-Kloos 2018), children in early adolescence are ready for multiplayer experiences that allow them to play and socialize with friends online (Kafai et al. 2013; Du et al. 2021). Games like Roblox, Minecraft, and Fortnite are often referred to as precursors to the metaverse: persistent virtual worlds populated by groups of simultaneous users that include easy-to-use tools and give players the freedom to create things and set their own goals. There is some evidence that participation in multiplayer game environments has positive effects on youth wellbeing. These include outcomes like increased sense of agency and belonging (Muriel and Crawford 2020; Vella et al. 2019), development of social emotional skills and habits (Schrier and Farber 2021; Mukund et al. 2022; Salen Tekinbaş et al. 2021), and activist stances (Cortez et al. 2022). However, a growing body of research also points to potential negative effects for some players, specifically around harmful conduct, such as harassment and bullying (Fox et al. 2018; Türkay et al. 2020). According to related research, harm in games follows a cyclical pattern in which the roles of victims, bystanders, and perpetrators are constantly changing. This means that individuals can switch between these roles, and both those who experience harm and those who witness it often become perpetrators themselves (Ballard and Welch 2017; Tang et al. 2020; Fryling et al. 2015). This cycle of harm leads to players engaging in harmful conduct more routinely (Hilvert-Bruce and Neill 2020; Beres et al. 2021). Adolescents, particularly those in their early teen years, may emulate the norms and behaviors of those individuals they see as successful or influential in the game, learning to replicate behaviors that can cause harm.

3.1 Social and Emotional Learning Support

Most teens and tweens say participation on social media platforms helps support their social-emotional wellbeing, boosting confidence and alleviating anxiety, loneliness, and depression (Rideout and Robb 2019). Further, a 2019 Canadian Paediatric Society report presents evidence
that media use appears to improve self-concept in teens by enhancing perceived friendship quality (Selhout et al. 2009; George and Odgers 2015; Underwood et al. 2012; Davis 2013), and fulfilling a developmental ‘need to belong’ (Veissière and Stendel 2018). The report also found that online communication has the potential to facilitate self-disclosure among isolated or socially anxious adolescents, resulting in increased social connectedness and decreased depressive symptoms (Canadian Paediatric Society 2019).

Participation in online communities around a shared interest have been shown to mediate social experiences for youth (Cole and Griffiths 2007; Ringland et al. 2016) and offer authentic contexts in which to develop important social and emotional skills and habits, including those related to conflict resolution and problem-solving (Slovák et al. 2018; Salen Tekinbaş et al. 2021). Such experiences are linked to apprenticeship within communities of practice (Gee 2005; Lave 1991; Wenger 2000) that emerge as part of the social interactions around participation on the platform—such as when new users are ‘shown the ropes’ by other community members. This points to the fact that peers and access to social support carry important protective effects for young people’s mental health, and increasing evidence suggests that online communication may be a critical avenue for peer-to-peer support for adolescents (Ito et al. 2020b).

Research also shows that adolescents with stronger offline networks report more robust online networks (George and Odgers 2015). With that said, we must also consider instances where access to strong offline social support may be limited or unpreferred. For example, social support from family and friends is vital to the wellbeing of LGBTQ+ youth (Snapp et al. 2015), however, youth of these communities are more likely to turn to online friends for emotional support (Ybarra et al. 2015). An interview study found social VR spaces to be a substantial resource for LGBTQ+ users to seek social support and build close relationships with others (Acena and Freeman 2021). Studies have shown that the increased social support that youth experience in many online settings may reduce their feelings of social isolation (McInroy 2020) and social anxiety (Best et al. 2014). Social support can also help youth become more socially adept, which can help them in their building of offline connections and friendships (McInroy and Craig 2018; McInroy et al. 2022). The findings of these studies indicate that online spaces remain a beneficial resource for social support for many adolescents.

### 3.2 Acts of Resistance

Youth are intentionally countering some of the risks to wellbeing described earlier in this paper. For example, social media and other metaverse-like spaces have very pronounced communities combating eating disorders. Oksanen and colleagues (2015) suggest that anti-eating disorder communities filled with their peers (anti-pro ana and pro-recovery) could be more fruitful avenues for combating eating disorders online due to youth suffering from eating disorders often feeling alienated from adults and physicians. Anti-pro ana videos also appear to be more liked and commented on positively on sites like YouTube than pro-ana videos, and pro-recovery content
creators will often use pro-ana hashtags to allow for their content to be seen by youth currently experiencing eating disorders (Oksanen et al. 2015; Yom-Tov et al. 2012). Social media content made by youth recovering from eating disorders can also increase awareness among peers about eating disorders and be therapeutic for those in recovery (Herrick et al. 2021). It should be noted, however, that some content that engages in “gallows humor” (e.g. “let’s confuse people who have good relationships with food”) could be harmful to other youth, especially their peers in eating disorder recovery.

We see evidence of similar resistance around self-harm related content. Despite the potential risks of the sharing of SITB content among youth in metaverse-like spaces, these spaces can also provide support to youth who are struggling with SITBs (Biernesser et al. 2020; Dyson et al. 2016). Youth can post general messages of comfort and also interact with each other directly to provide support, including emotional support, positive advice, and suggestions to seek professional help (Cavazos-Rehg et al. 2016). Given that many youth struggling with SITB hesitate to seek treatment, metaverse-like spaces could serve as crucial points of support for these youth, and even potential places of intervention (Rice et al. 2016).

Formal studies have also experimented within these virtual spaces as a means of counteracting the risks associated with them. For example, research has used Second Life (2003) for facilitating scenarios to educate youth about cyberbullying (Wright et al. 2009). Another study explored how VR might be used for men to take on the perspective of female sexual assault victimization, resulting in greater empathy (Ventura et al. 2021). Last, Black youth streamers and gamers have been shown to engage in what Cortez and colleagues (2022) call “speculative activism,” in video games like The Sims and Grand Theft Auto. Speculative activism is a “co-constructed repertoire of resistance practices, leveraged across digital and social media domains, oriented toward organizing new civic futures” (Cortez et al. 2022:56). In one study, Cortez and colleagues highlight the video game play of OTRGamerTV, a group of Black and Brown youth who regularly come together to engage in activist practices inside of multiplayer games like Grand Theft Auto V (GTA) (Cortez et al. 2022). Like the pro-recovery communities and practices highlighted above, BIPOC youth are “using socio-political action, critique, and other concrete actions” (Gutiérrez et al. 2019) to change the world around them, remaking and reordering the metaverse-like spaces in which they are engaging.

3.3 Spaces of Refuge, Connection, and Care

Scholars of race and technology have theorized and documented discriminatory practices shaping the current and metaverse-enabling technologies. These include data discrimination through biased algorithms (Noble 2018; Eubanks 2018), the racial violence of social media (Nakamura 2014), automation and discriminatory design (Benjamin 2019), as well as the surveillance (Browne 2015), suppression, and silencing of Black youth’s activism (Tanksley 2020a). At the same time there is ample evidence that minoritized youth are “retreating from open platforms like Twitter or YouTube, to more private and safe networked spaces to escape racist and hateful content” (Ito and Salen 2021), or to connect around shared interests and identities with relative anonymity. Byron
(2019) and Cho (2017) have studied Tumblr—a social media platform often associated with queer youth cultures. They find that Tumblr provides a space of refuge for many queer youth by offering an inclusive and affirming space in which to negotiate identity, find peers and friends, and locate communities that may be unavailable elsewhere (Byron 2019). Tanksley has closely studied Black girls’ use of private text messaging groups to support one another (2020b), creating spaces of personal healing in the face of the hyper-circulation of racially traumatizing imagery (2020b; 2022). Cavalcante’s (2016) digital ethnography of a transgender person transitioning online points to the possibility of “counterpublics” (Warner 2002) that serve as organized networks of care and concern, enabling transgender individuals to navigate their identity and cope with everyday challenges. Such counterpublics create feelings of belonging and ideological affiliation (Cavalcante 2016) that can contribute to adolescent mental wellbeing (Sawyer et al. 2018; Smith et al. 2021; Ortuño-Sierra et al. 2021).

Finally, youth are finding refuge in identity-affirming online communities created on customizable platforms like Minecraft and Discord. For example, the youth-led organization South Dakota Youth Activism is using the voice, video, and text chat app Discord to create safe spaces for BIPOC, LGBTQ+, and disabled youth. Minecraft servers like Autcraft and Therathrive create safe spaces for neurodivergent youth. Autistic youth are often the target of harassment and violence in online spaces (Cappadocia et al. 2012; Liu et al. 2022). Ringland (2019) found that the Autcraft community was actively engaged in making themselves a safe space for youth with autism. “Beyond simply keeping bullies out... the community has taken the label of ‘autism’ and turned it into something positive—a label worth identifying with” (p. 1).

A Spotlight on Virtual Reality

Some of the more unique qualities of some metaverse-like spaces are the incorporation of VR technology and embodied experiences through avatars. More specifically, it is anticipated that some versions of the metaverse will be based in VR. The proposed immersion afforded by VR is theorized to be accompanied by “greater embodiment over one’s avatar, greater sense of co-presence with internet-based friends, and generally more vivid experiences” (Paquin et al. 2023:3). This immersion may impact youth mental health in both positive and negative ways.

Avatars

There is some evidence that a youth’s use of an avatar may impact their mental health. When creating avatars, individuals tend to extend traits of themselves within the offline world into the avatar (Gilbert et al. 2014; Zimmermann et al. 2022). This process may result in individuals feeling more emotionally close and similar to their avatars, and eventually lead to the individual’s identity becoming “immersed” within this representation.

---

2 Queer, is an umbrella term inclusive of all non-heterosexual sexual orientations and non-cis-gender identities.
This immersion may result in the situations and experiences the avatar faces to feel more “real” to users and impact their emotions and behaviors (Yoon and Vargas 2014; Li et al. 2013). Wrzesien et al. (2015) found that teens were more likely to feel negatively when an avatar that was placed in a frustrating situation looked physically similar to them than when the avatar did not look similar to them. When those same teens were shown breathing techniques with the avatars, teens were more likely to feel relaxed when viewing the physically similar avatar than the avatar that did not look similar to them (Wrzesien et al. 2015). Additionally, over-identification with an avatar may also encourage the use of metaverse-like spaces in unhealthy ways. Individuals with higher social anxiety were more likely to identify with their avatars and also to participate in gaming in ways that impaired their functioning in offline worlds (Sioni et al. 2017).

Additionally, a lack of diverse avatar customization options provided within these spaces may also lead to experiences that negatively impact youth mental health. In particular, virtual worlds that make the default avatar White and limit access to non-White body parts can create feelings of frustration for some non-White youth and increase the amount of discriminatory interactions they experience (Kafai et al. 2010; Gray 2012a). The use of avatars in Social VR may also provide increased opportunities for youth to connect with each other, especially those from marginalized communities. Social VR offers opportunities for members of marginalized communities to make meaningful connections with others and bolster their social support (Acena and Freeman 2021). Embodied experiences in social VR also afford users a unique means of exploring gender identity (Freeman and Maloney 2021). Virtual avatars have also offered people with disabilities (both visible and invisible) opportunities to find meaningful work, escape visual biases, and reconnect with past qualities of themselves (Davis and Chansiri 2019). Perceptions of online social support and general wellbeing are also enhanced by presence-related affordances of VR technology (van Brakel et al. 2023; Gilbert et al. 2014).

However, researchers have also raised concerns that interacting with other users through avatars may allow for harassment and other negative interactions to be felt more intensely and vividly by youth (Maloney et al. 2020; Wiederhold 2022).

**Harmful Uses of VR**

Most of the work surrounding VR and other immersive technologies and mental health has been conducted in the context of delivering therapy to adults. In this work, VR is often used to simulate aversive situations feared by clients through visual and auditory stimuli to evoke an anxiety response that the client can learn to tolerate in a safe environment (Emmelkamp and Meyerbröker 2021). The effectiveness of VR therapy for many anxiety disorders suggests that encountering stressful or aversive situations in
VR can elicit feelings of fear and anxiety similar to the ones that would be felt from an equivalent offline encounter (Emmelkamp and Meyerbröker 2021; Parrish et al. 2016; Harris et al. 2002; Ridout et al. 2021). In other words, negative experiences within VR may cause emotions similar to if they had happened offline. News articles have highlighted that sexual harassment in some virtual reality platforms, such as Population One and VRChat, can feel particularly violating when combined with a haptic vest that relays sensations (Frenkel and Browning 2021).

Researchers have also identified potential side effects from the use of VR, such as dizziness, nausea, and headaches (Srivastava et al. 2014; Benrimoh et al. 2022; Biener et al. 2022). These physical side effects could negatively impact youth wellbeing. However, these immersive worlds could also provide experiences and social interactions that have a positive impact on youth mental health. For example, virtual reality nature environments could activate the parasympathetic nervous system and facilitate stress recovery (Annerstedt et al. 2013). Additionally, VR and other immersive technologies will likely continue to be useful in treating mental health issues in users (Cerasa et al. 2022).

Ultimately, evidence on the effects of virtual reality on youth mental health is still in its nascent phase. Based on the current evidence, it is likely that immersion within metaverse-like spaces through virtual reality may make interactions within the metaverse more impactful on youth mental health, in both positive and negative ways. However, most of these studies do not use youth in their samples or measure specific mental health outcomes. As such, more rigorous study of the effects of immersive technologies such as VR and the frequent use of avatars is needed for adolescents and young adults before stronger conclusions can be drawn.
4. INEQUITABLE BENEFITS AND HARMs

Harms and benefits to youths’ mental health are not equitably distributed when it comes to participation in metaverse-like spaces. Rideout and Fox (2018) found teens are more likely to rely on social media for connection and validation from their peers than adults, and are also most likely to benefit from these connections. In another study, they found that Black youth are most likely to encounter racist content online, an experience almost certainly also encountered offline (Rideout and Robb 2019). As noted earlier, online and offline vulnerabilities are interrelated. Negative online experiences, such as negative social comparison or harassment, are stratified by gender, sexual orientation, age, race, and class (Ito et al. 2020a). Recent research is shedding light on the factors that contribute to the connection between digital technology use and mental health. It emphasizes the interconnectedness of offline and online risks and raises doubts about making causal claims. In other words, the most influential factors in explaining this relationship are not solely the amount of social media usage but rather more familiar factors such as poverty, instability, social marginalization, and various forms of stress (Alegría et al. 2010; Benner 2017). In the following section we profile findings related to the intersection between mental health outcomes and factors such as age, class, gender identity, race/ethnicity, and sexual orientation.

4.1 Age

While the age range of 13-24 covers both adolescents and youth as defined by the World Health Organization (WHO 2021) and by Sawyer and colleagues (2018), younger adolescents may face different mental health risks in online spaces than older adolescents and young adults. During adolescence, adolescents go through a number of biological, cognitive, affective, and social changes that may impact their risk for developing mental health disorders (Pfeifer and Allen 2021). For younger adolescents, individual, familial, and school and community factors may all play equal roles in determining youth mental health (O’Connell et al. 2009; Triana et al. 2019). When adolescents lack proper development in a variety of key psychological factors (e.g., self-esteem and self-regulation), have issues with forming peer and community relationships, and lack structure, supportive relationships, or opportunities for belonging and mastery within families and school systems, they may be more at risk for developing mental health disorders (O’Connell et al. 2009). In contrast, familial and school factors, as well as the development of key psychological factors, play far less of a role in influencing the mental health of young adults (O’Connell et al. 2009). Issues with identity exploration, especially regarding one’s sense of adult status and self-sufficiency, perceived achievement in work and education, and failures to connect with individuals outside the immediate family may be more salient risk factors for developing mental health issues in young adults (O’Connell et al. 2009).

Younger teens may experience different kinds of impacts on their mental health, when compared to older adolescents and young adults. Due to ongoing neurological and psychosocial maturation,
young adolescents have underdeveloped critical thinking skills and may be more vulnerable to suggestions from others online regarding harmful content than older adolescents and young adults, particularly regarding disorder eating content (Fowler and Vinson 2020; Papathomas et al. 2018). Some interpersonal interactions that lead to negative mental health outcomes, such as cyberbullying, may also be more common and impactful in younger adolescents (Bottino et al. 2015; Hamm et al. 2015). At the same time, younger adolescents can receive protective benefits from family structure, such as parent supervision of online activities or lack of access to funds, as well as support through school-based interventions (O’Connell et al. 2009; Yang et al. 2022). Livingstone and Smith (2014) note that “Lack of parental involvement is associated with vulnerability to being groomed online for sexual abuse (Whittle et al. 2013), and with involvement in all kinds of bullying, including cyberbullying (Wang et al. 2009). Impacts on the mental health of older adolescents are tied to ongoing brain development in the area of critical thinking, as well as to a strong developmental need for social connection and identity development (Arain et al. 2013; O’Connell et al. 2009), and decreased parental supervision. The greater freedom that many older adolescents and young adults have with their online activities can lead them to experience negative online interactions and consume harmful content at much higher rates than more supervised younger adolescents.

4.2 Class

Youth from low socioeconomic status (SES) backgrounds may be more vulnerable to mental health concerns in a metaverse. SES is often defined in the literature as “the relative position an individual or family holds within a social structure based on their access to limited and valued resources” (Krieger et al. 1997). The most common measure of socioeconomic status is familial or household income, but education level and neighborhood disadvantage may also play a role in determining SES in a social context (Broomfield et al. 2022). Studies on SES and youth mental health have found that low SES youth were more likely to have a mental health problem than youth from higher SES (Reiss 2013; Yang et al. 2022). This disparity is often attributed to youth from low SES being exposed to a variety of destabilizing stressors while simultaneously being deprived of strong social support within communities (APA 2017). Low SES youth may also face difficulties accessing mental health care due to a lack of affordability (Castro-Ramirez et al. 2021). Hodgkinson et al. (2017) note that “Despite their high need for mental health services, children and families living in poverty are least likely to be connected with high-quality mental health care.” Few programs are tailored to the needs of children and families experiencing poverty, and research shows that offline vulnerabilities are mirrored online. A recent study found that youth who reported being from low SES were 1.25 times more likely to have experiences on social media that have been linked to negative mental health outcomes, such as exclusion from online groups, harassment from strangers, and negative comments and messages on their accounts, than youth from higher SES (Skogen et al. 2022). Other studies have suggested that low SES youth may be more vulnerable to cyberbullying victimization, which is linked to increased depression symptoms and suicidal ideation (Lee et al. 2023). Additionally, low SES youth may lack some of the social protective factors found to reduce risk of youth mental health problems within digital spaces. The families of low SES youth may have fewer social resources within
their families and communities to monitor their children, making youth in these families less likely to receive adult supervision than youth from more socioeconomically advantaged families (Walker et al. 2015; Horvat et al. 2003). As such, low SES youth may have fewer opportunities for parental or other adult mediation of digital content, placing them at a greater risk of encountering content online that may negatively impact their mental health (Cabello-Hutt et al. 2018). Likewise, youth from low SES families may also have less access to high-quality mental health care than youth from higher SES families, making them less likely to have external resources to help deal with any negative mental health effects that may arise when using metaverse-like spaces.

4.3 Gender Identity

Across the clinical literature, LGBTQ+ youth have been shown repeatedly to have increased odds of mental disorders over the general population. Studies have found between 12%-58% of transgender youth have depressive symptoms or depression, between 9-30% have attempted suicide, between 13-53% have harmed themselves, and between 2-16% have experienced eating disorders (Connolly et al. 2016). When compared to their cisgender peers, transgender youth have increased odds of depression, self-harm, and suicide attempts (Connolly et al. 2016; Reisner et al. 2015). One notable meta-analysis of over 35 studies on sexual minority youth and suicide found that transgender youth had the most increased odds, being 6 times more likely to make a suicide attempt than their cisgender peers (Di Giacomo et al. 2018).

Further, metaverse-like spaces can fail to support transgender and gender-nonconforming individuals. Selkie and colleagues (2020)’s interviews with trans high school students had students mention that social media was a site for harassment and exclusion. Other studies have found that transgender youth are distrustful of social media companies due to experiencing harassment themselves or perceiving harassment of their peers (Schoenebeck et al. 2021). Additionally, some platforms like Facebook only allow users to make verified and recoverable accounts if the name is listed on government and non-government IDs, serving as a barrier for transgender youth to use their chosen names on these platforms. Given the importance of gender affirmation for reducing symptoms of depression and anxiety in transgender youth, these platform features may negatively impact the mental health of transgender youth (Pollitt et al. 2021).

Conversely, providing transgender youth with space to affirm their identities has been associated in the literature with improved mental health. Russell and colleagues (2018) found that chosen name use was associated with lower depression, suicidal ideation and suicidal behavior in transgender and gender nonconforming youth, and effects were strongest if able to be used across more contexts (including social media). Fontanari et al. (2020) found that trans and other gender nonconforming youth having accessed more steps of gender affirmation (including social acknowledgement) was associated with fewer symptoms of depression and anxiety and improved feelings of social acceptance. Social media can be a good source for social support for transgender youth, with trans highschoolers reporting that social media provided them with peer support, validation of experiences, access to educational materials for friends and family, and more info about healthcare decisions (Selkie et al. 2020).
4.4 Race/Ethnicity

Racial and ethnic minority youth may also have unique mental health concerns within the metaverse. Nationally, a considerable amount of racial and ethnic minority youth experience mental health problems such as symptoms of depression, anxiety, eating disorders, and self-injury (Kann et al. 2018; Lipson et al. 2022). Rates of mental health disorders for racial/ethnic minority youth have been found to be roughly equivalent to rates for Non-Hispanic White youth, but mental health disorders for racial/ethnic minority youth may be more chronic (APA 2017; Kann et al. 2018; Lipson et al. 2022). However, racial and ethnic minority youth are significantly less likely to receive treatment for their mental health concerns than Non-Hispanic White youth (Alegría et al. 2010, 2012; Marrast et al. 2016). These disparities in treatment have been attributed to many factors, including structural barriers that result in the under-provision of treatment providers for racial and ethnic minority youth, stigma against receiving mental health treatment within these communities, and concerns over the availability of culturally-appropriate treatments and culturally sensitive providers (Hingwe 2021; Fisher et al. 2018; Lu et al. 2021; Marrast et al. 2016; Mennies et al. 2021).

In particular, both personal and vicarious discrimination against racial/ethnic minorities online may lead to negative mental health outcomes in racial/ethnic minority youth. Youth reported that online spaces, such as social media websites, are the most common place where they witness discrimination against racial and ethnic minorities (Tynes et al. 2008; Valkenburg et al. 2019). In studies that found links between digital media usage and negative mental health outcomes (such as symptoms of depression, symptoms of anxiety, and heightened negative affect) for racial and ethnic minority youth, being a victim of or witnessing discrimination against racial and ethnic minorities explained the relationship (Tao and Fisher 2022; English et al. 2020; Nereim et al. 2020). Being a target of discrimination against racial and ethnic minorities online has also been directly linked to increased levels of depression and anxiety for racial and ethnic minority youth (Tynes et al. 2008; Cano et al. 2020). Additionally, exposure to highly publicized incidents of racial violence online has been associated with symptoms of PTSD and depression in African American and Latinx adolescents (Tynes et al. 2019).

Metaverse-like spaces may also provide opportunities that bolster the mental health of racial and ethnic minority youth. Social media platforms can serve as places for racial and ethnic minority youth to further develop their racial and ethnic identities and form close-knit communities with other racial and ethnic minority youth (Lee 2012; Brock 2012). These spaces can help empower racial and ethnic minority youth by serving as spaces to mobilize community action for racial justice and other forms of social activism (Earl et al. 2017). Some metaverse-like spaces, such as remote learning environments, have been embraced by racial and ethnic minority families to lessen instances of racial discrimination experienced by their adolescents within school settings (Speak Up 2021). Digital spaces may also provide opportunities to improve access to mental health resources for racial and ethnic minority youth through digital interventions, although large-scale testing of such interventions is still needed (Schueller et al. 2019).
4.5 Sexual Orientation

Nationally, LGBTQ+ youth frequently experience serious mental health concerns. In 2022, 63-79% of youth who identify as LGBTQ+ reported experiencing symptoms of anxiety, 49-66% reported experiencing symptoms of depression, 35-53% reported seriously considering suicide, and 11-21% reported attempting suicide (The Trevor Project 2022). Additionally, rates of mental health problems for non heterosexual youth are often higher than those reported for heterosexual youth, including being 3.5 times more likely to attempt suicide (di Giacomo et al. 2018; Kann et al. 2016; Mustanski and Espelage 2020). These disparities are often attributed to minority stress theory: that LGBTQ+ youth must contend with chronic stressors from their stigmatized identity in addition to universal stressors that youth face (Meyer 2003). In particular, instances of anti-LGBTQ+ discrimination, particularly rejection and harassment from peers, are frequent stressors for LGBTQ+ youth (Katz-Wise and Hyde 2012; Goldbach and Gibbs 2017; Russell and Fish 2016; The Trevor Project 2022). These stressors are theorized to negatively affect the mental health of LGBTQ+ youth by causing emotional dysregulation, impoverishing social support, and contributing to thoughts of hopelessness and low self-esteem, all of which may leave LGBTQ+ youth more vulnerable to negative mental health outcomes (Hatzenbuehler 2009).

In particular, LGBTQ+ youth may be more vulnerable to negative interactions with their peers online than their heterosexual peers, contributing to worsened mental health. Youth who identify as LGBTQ+ were twice as likely to report being cyberbullied than heterosexual youth (Schneider et al. 2012; Kann et al. 2018). LGBTQ+ youth who experience this cyberbullying were more likely to report higher rates of depression (GLSEN 2013). Cyberbullying victimization for LGBTQ+ youth has also been associated with increased psychological distress (McConnell et al. 2017). LGBTQ+ youth also report some of the highest levels of hate speech online, with 64% of users reporting harassment and 52% of users reporting severe hate speech harassment (ADL 2021).

Metaverse-like spaces may also provide opportunities for LGBTQ+ youth above and beyond their heterosexual counterparts. Given a lack of education materials offline, digital spaces often provide key resources for LGBTQ+ youth learning about their identities (DeHaan et al. 2013; Attwood et al. 2018). Young adults reported using digital spaces to access educational resources regarding sexual identities and transitioning to a new sexual identity, find role models within LGBTQ+ communities, and develop their identities based on their sexual orientation (Fox and Ralston 2016). Metaverse-like platforms also provide crucial social support to LGBTQ+ youth. LGBTQ+ youth are more likely than LGBTQ+ youth to rate their online friendships as being more supportive than their friendships offline (Ybarra et al. 2015). While in-person support may be more effective at reducing negative mental health effects (such as depression), online social support can also be helpful (Ybarra et al. 2015; Cole et al. 2017). Digital spaces may also be fruitful for providing mental health resources to LGBTQ+ youth; these youth are more likely to use digital spaces for mental health-related support than heterosexual youth (Rideout and Fox 2018).
A Spotlight On Adolescent Girls

Adolescent girls are a potentially vulnerable population at risk of experiencing negative mental health effects within the metaverse. Generally, girls are at a greater risk of experiencing mental health disorders than boys starting in puberty. Some studies have found that the odds ratio of experiencing depression is significantly higher for girls than boys starting at age 15 onwards, with the odds of experiencing depression for the first time almost twice as high for adolescent girls than boys (Hankin et al. 1998; Pedersen et al. 2023). Over the past 10 years, the only significant increase in mental disorders amongst non-adults has been the increase in internalizing disorders in adolescent girls (Bor et al. 2014; Collishaw 2015). Many mechanisms for this gender difference have been suggested, including increases in pubertal hormones (Brooks-Gunn and Warren 1989) and a combination of heightened stress sensitivity and exposure to increased stressors (Thapar et al. 2012). Cyranowski and colleagues (2000) theorize that pubertal maturation sensitizes females to the depressogenic effects of negative life events. The authors note that girls are socialized to want to develop a sense of relationships with intimacy and base a sense of self off of connection with others. During adolescence, feelings of stress and anxiety arising from social and biological changes associated with puberty may cause adolescent girls to have more affiliative need as a way to cope with these stressors, and it is when these needs are not met that mental disorders may arise (Cyranowski et al. 2000).

Even beyond these general concerns, several studies comparing cisgender boys and cisgender girls have indicated that metaverse-like spaces may be more harmful for the mental health of adolescent girls than boys. For instance, a study utilizing data from the Millennium Cohort Study, involving 5,926 girls and 5,946 boys aged 13 to 15, found that 14-year-old girls who used social media more frequently exhibited higher depression scores and other clinically significant symptoms. Notably, this association was stronger for girls than for boys in the study (Kelly et al. 2018). Furthermore, there is evidence to suggest that chatting on social media platforms is associated with poorer mental health outcomes among adolescent girls. An analysis of a comprehensive dataset covering a wide population revealed that girls who engaged in more interactions with others on social media at the age of 10 experienced significantly lower wellbeing by the age of 15. Notably, this association was observed specifically among girls and not among boys (Booker et al. 2018). Another extensive dataset analysis conducted in the UK demonstrated that very frequent social media use among 13 to 14-year-old girls predicted lower wellbeing, including reduced life satisfaction and happiness, as well as increased anxiety, by the age of 16. In contrast, this relationship was not observed among boys in the sample (Viner et al. 2019).

\[3\] In our findings, demographics generally focused on and defined as girls were cis gender and assigned female at birth, while boys were defined as cis gender and assigned male at birth.
It is important to note that other studies and reviews have not found substantial evidence to support the idea that general social media usage leads to worsened mental health among adolescent girls (Orben and Przybylski 2019; Odgers and Jensen 2020). These studies argue that, after adjusting for confounders, many larger studies of social media use and mental health find that only adolescent girls with the lowest propensity for depression have slightly increased risk for depressive symptoms with daily social media use (Odgers and Jensen 2020).

Different mechanisms have been proposed for how increased social media use may be negatively impacting the mental health of adolescent girls, but not boys. Some studies have suggested that relational causes and body dissatisfaction may underlie these differences. Upward social comparison was suggested as a mechanism for how increased social engagement on social media platforms may lead to decreased wellbeing: as girls interact more with others online, they may start to compare themselves more to those they interact with; if they find themselves lacking this could lead to decreases in social wellbeing (Booker et al. 2018). Other studies have suggested that a variety of interpersonal factors (social comparison, relational aggression, desire for popularity, intimate yet fragile friendships) and greater self-objectification (body dissatisfaction) may all play a role in how increased time spent on social media leads to worsened mental health (Nesi and Prinstein 2015; Card et al. 2008; Twenge and Farley 2021; Flook 2011; Shih et al. 2006; Watkins et al. 2003; Yau and Reich 2019). Features of social media (e.g., idealized images of peers, quantifiable feedback) might intersect with adolescent developmental factors (e.g., salience of peer relationships) and sociocultural gender socialization processes (e.g., societal over-emphasis on girls’ and women’s physical appearance) to create body dissatisfaction in adolescent girls that translates into worsened mental health (Choukas-Bradley et al. 2022).

Several studies have examined key moderators and mediators that may explain the relationship between increased social media use and worsened mental health in adolescent girls. One study found that online harassment, sleep disruptions, poor self-esteem, and body dissatisfaction appear to be both more common and more clinically relevant for girls’ mental health and wellbeing than boys, especially at higher rates of social media use and online interaction (Kelly et al. 2018). The study also found support for one potential pathway: adolescent girls experiencing online harassment were more likely to have poor sleep, poor body image and low self-esteem; and that girls with poor body image were more likely to have low self-esteem. Another study found that after adjusting for cyberbullying, sleep, and physical activity, the relationship between time spent using social media and lessened life satisfaction was no longer significant in adolescent girls (Viner et al. 2019). Thus, the authors conclude that sleep and bullying are more powerful determinants of wellbeing in young people than is digital screen use.
5. CONNECTING RISK FACTORS AND PROTECTIVE FACTORS

This section seeks to identify both the risk factors and protective factors that augment the effects of participating in metaverse-like spaces to youth mental health. We then draw from the Strategic Prevention Framework (SPF), examining the mental health risk factors and protective factors associated with core Metaverse concerns (SAMHSA 2019). According to the SPF, risk factors are “characteristics at the biological, psychological, family, community, or cultural level that precede and are associated with a higher likelihood of negative outcomes” (SAMHSA 2019). Protective factors, on the other hand, are “characteristics associated with a lower likelihood of negative outcomes or that reduce a risk factor’s impact” (SAMHSA 2019). In this section, we focus on answering RQ2.

RQ2: What are the protective factors associated with positive mental health outcomes among youth in metaverse-like spaces? How do these factors interact with the potential negative impacts on youth mental health related to participation in metaverse-like spaces?

We begin by discussing individual risks and protective factors, before moving on to discuss interpersonal and community risk factors. For information on the research framework used for this section of the report, please refer to the Appendix.

5.1 Individual Risks and Protective Factors

Though this review has identified legitimate concerns regarding youth mental wellbeing in metaverse-like spaces, there are individual factors that serve to facilitate how youth engage with and are impacted by such concerns. For example, it may be difficult to imagine a highly confident adolescent regularly engaging in negative upward social comparison on social media, or an adolescent with strong self-regulation skills engaging in retaliatory bullying behavior. As such, this section will explore three key individual factors that have been found to play a significant role in augmenting mental health outcomes associated with metaverse-like spaces: self-esteem, self-regulation, and resilience.

5.1.1 Self-esteem

Self-esteem, defined as “an individual’s set of thoughts and feelings about his or her own worth and importance,” is a crucial part of the self-concept that goes through changes during adolescence (Rosenberg 1965; Moksnes and Reidunsdatter 2019). Self-esteem is regarded as being important for the mental health of youth, with more self-esteem associated with positive mental health and functioning for teens and young adults (Boden et al. 2008; Moksnes and Reidunsdatter 2019).

An adolescent or young adult’s pre-existing level of self-esteem may be particularly important in determining how their interactions with metaverse-like spaces will impact their mental health. Youth with low self-esteem have been shown in some studies to be more vulnerable than their peers.
to negative mental health outcomes from social media use, including depression and social media addiction (Wang et al. 2018; Banyai et al. 2017). Pre-existing low self-esteem may also encourage youth to engage in online spaces in ways that will worsen their mental health, such as excessively curating social media profiles to gain peer recognition and avoiding non-online interactions to the point of an internet addiction (Chua and Chang 2016; Fioravanti 2012). Conversely, high self-esteem has been shown to be a protective factor against interactions that could lead to negative mental health outcomes for youth in metaverse-like spaces, lessening the likelihood of maladaptive social media use and negative social comparisons (Martinez-Pecino and Garcia-Gavilan 2019; Niu et al. 2018).

Research also indicates metaverse-like spaces provide opportunities to both lessen and bolster the self-esteem of youth. Some evidence suggests that use of metaverse-like spaces is associated with lower self-esteem in youth. For example, increased social media use and investment in social networking sites has been found to be associated with lower self-esteem, possibly due to increases in social comparison (Blomfield and Barber 2014; Shah et al. 2019; Woods and Scott 2016; Vogel et al. 2014). Additionally, being at risk for currently experiencing social media addiction has been associated with low self-esteem in youth (Andreassen et al. 2017; Banyai et al. 2017). Notably, these studies lack designs that can show causality, leaving it uncertain if this relationship is due to social media use lowering self-esteem in youth, low self-esteem leading youth to engage more with social media, another factor leading to both low self-esteem and increased social media use in youth, or some combination of the above.

Other studies have found that use of metaverse-like spaces is associated with heightened self-esteem in youth, especially when these spaces lead to positive social interactions (Best et al. 2014). Spending time socializing online, having more friends on social media, and receiving positive feedback on online profiles has been associated with, and in some cases even found to improve, high self-esteem in youth (Apaolaza et al. 2013; Ma 2022; Valkenburg et al. 2021b).

There is also some evidence to suggest that only small subsets of adolescents are vulnerable to the effects of metaverse-like spaces on self-esteem. Valkenburg et al. (2021b) found that most adolescents experienced no or very small effects of social media use on self-esteem, but 4% received positive effects and 8% received negative effects. While more work needs to be done to identify these subgroups, some factors that may make youth more vulnerable to experiencing lower self-esteem from social media use are identifying as female (Steinsbekk et al. 2021; Fioravanti 2012; Ma 2022), receiving negative feedback online (Drogos 2015), and other-oriented social media use (i.e. spending more time focused on the content of others rather than focusing on self-generated content) (Steinsbekk et al. 2021).

Self-esteem has been associated with many of the mental health concerns for youth highlighted in this review, particularly within social risks and body image issues. In general, studies that find certain types of social media use to be associated with worse mental health also tend to find that use
associated with lower self-esteem (e.g. Stanton et al. 2017). Cyberbullying and sexual harassment has been linked to lower self-esteem in the victims, with more instances of bullying and harassment associated with lower self-esteem (Bendixen et al. 2018; Brighi et al. 2012). This may be especially true for adolescent girls, where girls experiencing online harassment were more likely to have low self-esteem (Kelly et al. 2018). Social comparisons on social media have also been linked to lower self-esteem in youth (Vogel et al. 2014; Lee 2014). Likewise, body image concerns have also been linked to low self-esteem, where girls with poor body image were more likely to have low self-esteem (Kelly et al. 2018).

Likewise, self-esteem appears to be relevant to body image concerns and disordered eating behaviors in youth in metaverse-like spaces. Exposure to manipulated appearances on social media and other platforms (particularly through image modification or modification of avatars) can lead to decreases in self-esteem and body satisfaction, particularly in adolescent girls (Shah et al. 2019). Low self-esteem has also been found to partially explain relationships between increased social media use and disordered eating behaviors in adolescents (Livet et al. 2022). As such, youth with lower self-esteem may be more vulnerable to the negative impacts on body image and disordered eating from exposure to manipulated appearances online.

5.1.2 Self-regulation

Self-regulation is considered a vital component to the development of youth individuals. Many brain regions associated with self-regulation, such as the regions of the frontal-parietal network, cingulo-opercular “salience” network, and the dorsal-attentional network, are in the middle of a maturation period during adolescence that lasts into the mid-twenties (Chiu and Chien 2022). As such, the ability to self-control in many youth may not be to its fullest potential until adulthood, although youth are still able to develop self-regulation skills during this developmental phase. In particular, self-regulation skills, or having the ability to understand and manage one’s emotions and behaviors in diverse contexts, are key to promoting healthy and positive habits that continue well into adulthood (Napolitano et al. 2011).

In terms of mental wellbeing for youth individuals, studies show that youth who have developed better self-regulatory skills, especially emotional self-regulation, have better ability to achieve personal or academic goals (Buckley and Saarni 2014), as well as higher aptitude of resilience and psychological wellbeing (Mertens et al. 2022). In the context of online spaces, research also suggests that stronger self-regulation skills can diminish the negative impacts on work-technology conflict and self-esteem among youth exhibiting maladaptive use of social media (Khan, Khan, and Moin 2021). Youth individuals from marginalized backgrounds and low-SES, however, struggle to gain self-regulatory skills as a product of their situation and environment (Caughy et al. 2018).

With these considerations, in the advent of the digital age, many studies have sought to examine how self-regulation presents itself in the context of social media and online activity. Strong self-regulation skills may yield positive consequences for youth who engage regularly on social media by enhancing
their social contact, independence, and communication (Gupta 2022). Furthermore, stronger self-regulation can facilitate more self-enhancing social comparison (Johnson and Knobloch-Westerwick 2014) and positive coping strategies (Reinecke and Rieger 2020) over social media.

Conversely, poor self-regulation can result in frequent social media self-control failure—or instances where social media use intervenes with the achievement of a user's other goals (Du, Kerkhof and Van Koningsbruggen 2021; Reinecke et al. 2022). This can occur due to the high gratifications obtained from social media use, which can displace other activities important to healthy wellbeing (i.e., sleep, offline socializing, and physical activity) and distract users from achieving goals that may contribute to reduced stress and anxiety (Reinecke et al. 2022).

Reward-sensitivity may also play a role in youth's self-regulation practices in metaverse-like spaces. Youth who are more reward-sensitive (meaning they experience rewards more saliently) may be more inclined to continue to engage with metaverse-like spaces to the detriment of their wellbeing. Youth who experience dysregulation in brain circuitry involved in reward-processing (i.e., the amygdala, ventral striatum, orbitofrontal cortex, etc.) were more likely to use digital platforms to the point of impairment (Kuss et al. 2018; Lin and Lei 2015; Turel et al. 2014; Chiu and Chen 2022). As such, youth who are more reward-sensitive may struggle with self-regulation in metaverse-like spaces without external guidance.

5.1.3 Resilience/Psychological Endurance

Resilience refers to the process and outcome of successfully adapting to demanding or difficult life experiences. It involves demonstrating mental, emotional, and behavioral flexibility in response to both external and internal demands (APA 2023). Resilience may be the product of a variety of factors, including the ways in which individuals view and engage with the world, the availability and quality of social resources, specific coping strategies, and interactions between the three (Herrman et al. 2011; APA 2023). Within the literature, trait resilience has been shown to be negatively correlated with negative indicators of mental health and positively correlated with positive indicators of mental health in youth, including satisfaction with life, positive and negative affect, depression, and anxiety (Hu et al. 2015).

One aspect of resilience, an individual's coping strategies, has been shown to play an important role in staving off negative mental health outcomes in metaverse spaces. A survey study of maladaptive gaming behaviors suggests that players who adopt more “approach” coping strategies—actively and adaptively managing stressors—are less prone to excessive game play and symptoms of depression and anxiety (Moge and Romano 2020). Conversely, avoidance coping strategies were positively correlated with game addiction, depression, and anxiety.

Resilience is a key factor in mitigating the negative mental health effects of cyberbullying and reducing further victimization (Hinduja and Patchin 2017; Navarro et al. 2018), even for youth currently experiencing distressing circumstances offline (McLaughlin et al. 2021). Low resilience
has been associated with greater risk to online grooming (Whittle et al. 2013). Resilience has also been shown to lessen the relationship between hate speech victimization and depression symptoms in Spanish adolescents (Wachs et al. 2022). Resilience may also lessen the likelihood of emotional outbursts online from young adults in response to hate speech victimization (Saha et al. 2019). Resilience has also been found to moderate more severe outcomes associated with social comparison, with high resilience being linked to lower suicide ideation and vice versa (Wetherall et al. 2019).

5.2 Interpersonal and Community Risk and Protective Factors

Family and friends can have an indirect impact on youth mental health outcomes associated with participation in a metaverse-like space, particularly in the likelihood of youth developing maladaptive behaviors. Multiple studies have found that family qualities such as low family cohesion, high parent-adolescent conflict, unhealthy family functions, and poor communication predicted maladaptive internet use and video game use among youth (Yen et al. 2007; Park et al. 2008; Lam 2014; Hyun et al. 2015; Lee et al. 2016; Adams et al. 2019; Yayman and Bilgin 2020). These findings emphasize the importance of family and friends to youths’ mental wellbeing as they engage with others online. The following section will take a closer look at the support offered by close relationships and the role of parent mediating strategies when dealing with metaverse-like spaces.

5.2.1 Social Support

Close interpersonal relationships serve an important role in providing instrumental, informational, and emotional assistance in times of need—otherwise known as social support (House et al. 1988). Cutrona and Russell (1990) identify five dimensions of social support: emotional, network, esteem, tangible, and informational. Emotional support refers to comfort and security received from others during times of increased stress. Network support is understood as an individual’s sense of group membership among those that share similar interests. Esteem support takes place when one’s sense of competence or self-esteem is bolstered by others. Concrete assistance, such as providing resources or physical assistance, is referred to as tangible support. And lastly, informational support refers to advice or guidance from others.

Social support plays a pivotal role in reducing negative mental health outcomes, with one survey study of 1,378 college students linking perceptions of high quality social support with lower chances of depression, anxiety, suicidal ideation, and eating disorders (Hefner and Eisenberg 2009). Furthermore, high perceptions of social support from parents has been linked to reducing risks of maladaptive uses of the internet (Esen and Gündoğdu 2010) and video games (Moge and Romano 2020). Strong peer relationships have also been linked to fewer reported instances of sexual harassment (Mitchell et al. 2014) and reduced symptoms of anxiety and depression (Ståhl and Dennhag 2021). Conversely, a longitudinal study of 3549 adolescents linked low social support to an increased likelihood of perpetrating acts cyberbullying and sexual harassment (Leemis et al. 2019). The consensus of these findings is clear: social support plays a substantial role in mediating the risks and amplifying benefits of metaverse-like spaces on youth mental wellbeing.
5.2.2 Family Practices

Youth mental wellbeing is also impacted by family practices around online activities, referred to as mediation. And while a majority of research on mediation practices centers on younger adolescents and early teens, such practices can influence healthy use patterns in the future (Yang et al. 2022). An extensive survey study of 6,400 parents revealed two primary forms of mediation categories: enabling mediation and restrictive mediation (Livingstone et al., 2017). Enabling mediation (also referred to as ‘active mediation’) refers to collaborative strategies such as discussing online behavior with youth and supporting their activities in online spaces. These strategies are linked to increases in online opportunities for youth as well as increases to their risk exposure. Restrictive mediation, on the other hand, refers to controlling strategies such as placing limits on child screen time and prohibiting certain activities or websites. A survey study of 1102 teen/parent dyads found that parents implement fewer mediation strategies in general as teens get older (Shin and Huh 2011).

According to Livingstone and colleague’s (2017) analysis, parents are just as likely to implement both enabling and restrictive mediation when they perceive increased levels of risk to youth online. Restrictive mediation becomes more prominent, however, when perceived risk associated with internet use reaches an even higher threshold. Studies have found that excessive restrictions on media use can backfire, linking it to youth developing problematic patterns of use that can disrupt daily functioning (Livingstone and Helsper 2008; Shin and Huh 2011; Fischer-Grote et al. 2019) and lead to symptoms such as depression, anxiety, and stress (Shannon et al. 2022) as well as mood dysregulation (Kwon et al. 2013). These findings coincide with those of Koronczai et al. (2020), who found that overprotective parenting in general can be a risk factor for problematic internet use.

Apart from such extremes, several mediating strategies have been found to be effective in reducing risks to youth online. For example, parental involvement in youth internet use has also been identified as a protective factor against the risk of grooming (Whittle et al. 2013). A survey study of 935 teen/parent dyads found that co-creating rules for web browsing with youth and discussing online safety can reduce the risk of cyberbullying perpetration and victimization (Mesch 2009). By giving advice on internet use, parents can also reduce the risk of their child being exposed to privacy risks and harmful content (O’Keeffe et al. 2011). A study of 3,000 adolescents found that active mediation of smartphone use among young girls can reduce the likelihood of maladaptive smartphone use (Lee et al. 2016). Another longitudinal study found that giving instructions and explanations on mobile phone use increased the likelihood of youth maintaining healthy use habits in the future (Yang et al. 2022). Additionally, setting technical limitations on youth media devices via hardware or software was reported to decrease the likelihood of problematic online gaming behaviors (Benrazavi et al. 2015). Overall, there appears to be a consensus that parent mediation—active mediation strategies especially—is effective in reducing online risks such as cyberbullying and harmful content, as well as reducing the likelihood of youth developing problematic patterns of use that could impact their wellbeing (Kirwil 2009; Elsaesser et al. 2017).
One other form of mediation discussed among scholars is co-use (also referred to as co-viewing or co-playing), where parents actively participate in the activity of their child (Livingstone and Helsper 2008; Nikken and Jansz 2014). This can look different depending on the activity, such as being present while the child surfs the internet (Livingstone and Helsper 2008), taking turns playing a single-player game (Musick et al. 2021), or simultaneously playing the same multiplayer game. While some research has found little evidence of the effectiveness of co-use practices in reducing online risks (Livingstone and Helsper 2008), an even closer look at specific practices and linked variables sheds a bit more light on the matter. For example, results from a survey study of 361 parents support a causal link between family video game co-playing and family closeness—a factor associated with the likelihood of maladaptive behavior—with even greater benefits to families with poor communication (Wang et al. 2018). This is supported by another interview study, finding that co-playing digital games could overcome physical and emotional barriers such as living at separate residences and disengagement (Musick et al. 2021). Strong family closeness is also associated with greater perceptions of social support, both of which are linked to lower levels of depression among adolescents (Cumsille and Epstein 1994).

It will be important to continue monitoring the role of parent mediation strategies within metaverse spaces. The current consensus of research indicates the effectiveness of parent mediating strategies—active mediation in particular—in reducing risks and amplifying opportunities associated with metaverse spaces (Mesch 2009; Livingstone et al. 2017; Yang 2022; and others), with co-playing having more indirect impacts through improving family cohesion and social support (Wang et al. 2018; Musick et al. 2021). Metaverse-like spaces have particular potential for co-playing opportunities between parents and their children, with examples seen through minigames in Roblox (Geffen 2021) and socially-driven AR games such as *Pokemon GO* (Koskinen and Meriläinen 2021). However, some research indicates that even within these opportunistic spaces, very few parents opt to co-play with their child when they don’t find the game to be interesting (Mavoa et al. 2018). These are important considerations for development of metaverse spaces as well as a point of focus for continued research.

### 5.2.3 Community Governance and Moderation

Concerns about the impact disruptive and harmful online behavior may have on adolescent mental health and wellbeing has encouraged research into the role community governance and moderation play in minimizing risk and promoting positive outcomes. Recent work by Matias (2019) focused on governance in online communities suggests the importance of social norms in regulating behavior. Many of these norms are expressed in written form by means of a platform’s terms of use, rules, and guidelines. An analysis of 2,190 online discussions found that posting community rules can increase compliance to norms among first time users. (Matias 2019). To enhance the effectiveness of these policies, Powell and Henry (2017) assert that “terms of use that ban hateful, offensive, and harassing content should be clearly defined, transparently communicated, and effectively enforced by service providers” (p. 215).
Unfortunately, user research of many metaverse-like spaces indicates that platforms struggle to deploy effective governance systems. For example, one study highlighted user distrust in corporate policies to punish perpetrators in social VR spaces (Zheng et al. 2022). Another study of queer female users on Tinder, Instagram, and Vine reported that governance measures such as Terms of Service and flagging mechanisms often fail to protect users from harassment and discrimination (Duguay et al. 2020). Failure to enforce posted community norms can be problematic, as witnessing ongoing behaviors such as hate speech can increase the likelihood of perpetration (Wachs et al. 2022). Grace and colleagues (2022) reviewed codes of conduct for 60 multiplayer game titles from several digital marketplaces, including Steam, PlayStation Network, Xbox Store, and the Nintendo Store. They found that documents meant to guide and inform players of expectations and rules are often inaccessible and filled with hard to understand legal and policy related language (Grace et al. 2022).

Many uncertainties exist around how to best manage the safety of metaverse-like spaces and govern online conduct. Research in online moderation has explored the role moderators play in guiding community behavior, working alongside—and sometimes at odds with—platform-driven algorithmic moderation (Grimmelmann 2015; Chandrasekharan et al. 2018; Salen Tekinbaş et al. 2021). Community managers or volunteer moderators are often overburdened and under-resourced (Bulut 2020; Sparrow et. al. 2021) and may struggle to both write and enforce codes of conduct or community guidelines (Grace et al. 2022). Recent research into alternative forms of online community governance has advocated for a shift in focus away from punitive moderation processes, which fail to effectively stop the perpetuation of harm, toward restorative measures (Salehi 2020; Xiao et al. 2022). Such an approach would focus on the victim’s immediate safety and health, and on addressing the harm based on the needs of that person, rather than on removal of the offending content or actor. No robust studies yet exist on the outcomes of a restorative approach to online moderation as it relates to adolescent mental health, although outcomes of restorative approaches on adolescent mental health in the context of schools are robust (Frias-Armentia 2018; Reimer 2020; Huguley et al. 2022).
A Spotlight on Persuasive Design: Dark Patterns

Persuasive design, which merges principles of behavioral design with digital technology, is frequently employed to manipulate human behavior, encouraging individuals to unknowingly act in favor of commercial interests rather than their own. The most commonly used persuasive design strategies are those deployed in service to the ‘attention economy’—a system where revenue is generated by monetizing engagement (Kidron et al. 2023). Persuasive design elements are sometimes referred to as “dark patterns.”

The exposure of youth to harmful content online may be partly influenced by the presence of dark patterns. Dark patterns are deliberate design techniques employed by designers who exploit their understanding of human psychology to implement “deceptive functionalities” within their products, leading users to engage in behaviors that are not in their best interests (Gray et al. 2018). In the context of virtual worlds, these manipulative user experience (UX) design practices often aim to increase user overengagement, enable undesired features, or facilitate the oversharing of personal and private information (Gray et al. 2018). Examples of such dark patterns found within platforms include denying users the option to permanently decline app notifications, restricting access to full functionality until users invite their friends to sign up, or coercing users to return to a platform within a specific timeframe to avoid losing a reward (Gray et al. 2018; Zagal et al. 2013). Additionally, other dark patterns entice users to spend more money than intended, diminish privacy settings, enable user tracking, and reveal more personal information to the platform than initially intended (Gray et al. 2018; Zagal et al. 2013). Dark patterns have a widespread presence across most platforms, with a staggering 95% of the 240 most widely used apps incorporating one or more dark patterns into their interface design (Di Geronimo et al. 2020).

The impact of dark patterns on youth mental health remains largely speculative, as empirical studies directly linking specific dark patterns to negative mental health outcomes in youth are lacking. However, there are indications that dark patterns could have adverse effects on youth. One study conducted on adults explored the effects of dark patterns and found that when aggressive dark patterns were employed, adults were nearly four times more likely to comply with the platform’s intended action (e.g., signing up for a paid data protection program). Interestingly, participants exposed to aggressive dark patterns reported worsened mood and expressed higher levels of anger in their qualitative feedback (Luguri and Strahilevitz 2021). Conversely, the group subjected to mild dark patterns showed no differences in mood or anger compared to the group exposed to no dark patterns, suggesting that the impact of dark patterns on mood is not universal and may vary across different scenarios (Luguri and Strahilevitz 2021).
Certain dark patterns have the potential to steer youth towards detrimental situations and behaviors that can harm their mental health. Various dark patterns actively encourage continuous platform usage, employing tactics such as persistent notification prompts, the threat of losing rewards for infrequent logins, or deliberately concealing the log-out or account deletion options within the platform interface (Gray et al. 2018; Zagal et al. 2013). Consequently, these dark patterns may lead youth to engage with these spaces in problematic ways, disregarding important mental health needs such as sleep (Bayer et al. 2022; Wong et al. 2020). This is especially challenging for adolescents who are still developing impulse control (Casey et al. 2011). Moreover, considering the significance of peer influence during adolescence, youth may be particularly susceptible to distress caused by dark patterns that jeopardize their social standing or create a sense of social obligation to engage with a platform (Konrad et al. 2013; Sawyer et al. 2018; Zagal et al. 2013). For instance, the dark pattern of impersonation, where platforms assume the identity of a user’s friends or other players to interact with them, can lead to conflicts arising from the platform’s actions as the impersonated player (Zagal et al. 2013). Additionally, certain online game designs may incentivize or reward users for engaging in antisocial behavior, causing distress among the victims of such actions (Zagal et al. 2013).
6. IMPLICATIONS

Adolescence today is a life stage marked by active engagement with digital media, and youth will most certainly help drive what metaverse-like spaces can become, if given the opportunity. Our findings suggest that there are important roles to be played by youth, platforms, guardians, policy makers, and professionals, in both counteracting metaverse-linked risks, and in amplifying its benefits. To achieve this goal stakeholders must work together to steward forward a vision of the metaverse that takes the interests, needs, and vulnerabilities of youth into account, while also providing them with the tools they need to support their mental health. Our findings point to the following guiding principles:

- **Equity matters**: Harms and benefits to youths’ mental health are not equitably distributed when it comes to participation in metaverse-like spaces. More precise and actionable knowledge requires studying the features of the platforms, how youth engage with metaverse-like spaces, and the experiences and backgrounds they bring to their experiences there, in order to amplify or mitigate mental health risks for different adolescents.

- **There are no one-size-fits-all solutions**: Metaverse-like spaces should be designed in ways that support and acknowledge the great variation in youth across age, lived experiences, culture, identities, motivations, resources, and expertise. Doing so directs us to attend to research focused on identifying vulnerable populations, as well as groups, identities, and communities that are thriving online.

- **Recognize the importance of the pre-teen years**: Technologies developed with teens in mind will be used by pre-teens. We must consider how their pre-teen experiences with metaverse-like spaces may affect their mental health as teenagers. Partnerships between researchers and technology developers can support this work.

- **Center youth voices**: Engage youth as leaders in this work by centering youth experiences, needs, motivations, and assets in the design of codes of conduct, social dynamics, incentive systems, and moderation strategies. Shift the focus from solely protecting youth from online harm to creating opportunities for meaningful participation and empowerment.

- **Cross-sector support is needed**: There are important roles to be played by youth, platforms, guardians, policy makers, and professionals, in both counteracting metaverse-linked risks, and in amplifying its benefits. Supporting coordination and collaboration across sectors should become a core focus of philanthropic, business, research, and policy efforts.
The implications of the research covered in this review are multifaceted. Firstly, as noted above, the research suggests the need for deliberate and mindful approaches to create metaverse-like online platforms and environments that foster positive development among adolescents. This calls for the involvement of various stakeholders, including educators, parents, policymakers, and technology developers, to collaborate and prioritize the wellbeing of young people in online spaces. Additionally, it emphasizes the importance of implementing robust guardrails and supports to protect youth from potential risks and harmful experiences online. This may involve implementing age-appropriate content filters, promoting digital literacy and responsible online behavior, and providing support mechanisms for youth who may encounter negative or harmful content or interactions. It is also crucial to provide scaffolds and supports for youth as they enter metaverse-like spaces, especially girls.

Furthermore, recognizing the potential of diverse youth as agents, experts, innovators, and communicators highlights the significance of empowering young individuals to actively contribute to the wellbeing of their peers in online spaces. Encouraging youth to support and uplift one another in digital communities can foster a sense of belonging, connection, and positive social support, ultimately enhancing their overall wellbeing.

Overall, the implications revolve around creating a balanced approach that harnesses the potential benefits of online social experiences while proactively addressing and mitigating the risks, ultimately aiming to promote the positive development and wellbeing of adolescents in the digital age.
7. CONCLUSION

This review emphasizes the great variation in mental health outcomes in adolescents across age, lived experiences, culture, identities, motivations, and expertise, and highlights specific vulnerabilities tied to factors like age, class, gender identity, race/ethnicity, and sexual orientation. It identifies several core concerns associated with youth participation in metaverse-like spaces and negative mental health outcomes such as depression, stress, and anxiety. Though screen time has often been a concern in discussions around digital technology, it was found to be an unreliable indicator of mental health outcomes. Rather, stronger connections were identified between youth mental wellbeing and maladaptive uses of technology that interrupt daily functioning. We also found evidence that content around eating disorders and self-injury may perpetuate within these online spaces and promote problematic behaviors. Furthermore, we identified various social risks such as cyberbullying, sexual harassment, hate speech, and upward negative social comparison that may result in negative mental health outcomes among youth. Intentional predatory design and marketing practices to engage, retain, and monetize youth is also a growing concern.

Despite these concerns, metaverse-like spaces also offer several opportunities to bolster youth mental wellbeing. We highlight the ways in which many adolescents are leveraging peer-to-peer support online, and the ways they are actively supporting each other online in times of need and crisis. Online social spaces provide additional avenues for youth to strengthen their social support by reinforcing offline relationships and establishing new connections. Youth have shown that they are not powerless against exposure to risks online, as they use the very same platforms that circulate harmful content around eating disorders and self-injury to promote positive messages combating that content.

This review also identifies various factors on the individual, interpersonal, and community level that may reduce risks to youth mental health. On the individual level, youth with high levels of self-esteem and resilience, along with strong self-regulating practices, are less susceptible to poor mental health outcomes associated with metaverse-like spaces. Family and friends also play a significant role in reducing risks by providing social support and guidance around digital technology use. Online communities can also do their part in reducing online risks to youth mental health by providing clear expectations for online behaviors in the form of community guidelines or terms of service.
8. REFERENCES


Rideout, Victoria, Susannah Fox, and Wellbeing Trust. 2018. Digital Health Practices, Social Media Use, and Mental Wellbeing Among Teens and Young Adults in the U.S. Wellbeing Trust.


Royal Society for Public Health. 2017. *#StatusofMind: Social Media and Young People’s Mental Health and Wellbeing*. RSPH.


Titley, Gavan, Ellie Keen, and László Földi. 2014. *Starting Points for Combating Hate Speech Online: Three Studies about Online Hate Speech and Ways to Address It*. Strasbourg, FR: Council of Europe, Youth Department.


Youth, Mental Health, and the Metaverse Literature Review

Research Framework

Risks and protective factors for youth mental health are displayed within three domains: individual, interpersonal, community.

**Risks**

<table>
<thead>
<tr>
<th>Individual Risks</th>
<th>Research Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identity</td>
<td>What does the research on youth mental health have to say about risks associated with the digital representation of self in metaverse-like environments? How might such risks be mitigated?</td>
</tr>
<tr>
<td>Maladaptive Use</td>
<td>What does the research on youth mental health have to say about maladaptive use associated with behaviors in metaverse-like environments? How might such risks be mitigated?</td>
</tr>
<tr>
<td>Age-appropriate content</td>
<td>What does the research on youth mental health have to say about risks associated with age-appropriate content in metaverse-like environments? How might such risks be mitigated?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Interpersonal Risks</th>
<th>Research Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social isolation/loneliness</td>
<td>What does the research on youth mental health have to say about risks associated with social isolation in metaverse-like environments? How might such risks be mitigated?</td>
</tr>
<tr>
<td>Predatory behavior</td>
<td>What does the research on youth mental health have to say about risks associated with predatory behavior in metaverse-like environments? How might such risks be mitigated?</td>
</tr>
<tr>
<td>Bullying/Harassment/Assault</td>
<td>What does the research on youth mental health have to say about risks associated with bullying and harassment in metaverse-like environments? How might such risks be mitigated?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Community Risks</th>
<th>Research Questions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exploitation</td>
<td>What does the research on youth mental health have to say about risks associated with exploitation in metaverse-like environments? How might such risks be mitigated?</td>
</tr>
<tr>
<td>Disparities/equity</td>
<td>What does the research on youth mental health have to say about risks associated with inequity and exclusion? What does the research say about ways in which metaverse-like environments might create disparities among young users or exacerbate inequities?</td>
</tr>
</tbody>
</table>
### PROTECTIVE FACTORS

<table>
<thead>
<tr>
<th>INDIVIDUAL PROTECTIVE FACTORS</th>
<th>RESEARCH QUESTIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive physical development</td>
<td>In what ways might participation in the metaverse affect young people’s physical development (positively or negatively)?</td>
</tr>
<tr>
<td>Academic achievement/intellectual development</td>
<td>In what ways might participation in the metaverse affect young people’s academic achievement/intellectual development (positively or negatively)?</td>
</tr>
<tr>
<td>High self-esteem</td>
<td>In what ways might participation in the metaverse support or hinder young people’s self-esteem (positively or negatively)?</td>
</tr>
<tr>
<td>Emotional self-regulation</td>
<td>In what ways might participation in the metaverse support or hinder young people’s emotional self-regulation?</td>
</tr>
<tr>
<td>Good coping skills and problem-solving skills</td>
<td>In what ways might participation in the metaverse help or hinder youth in developing coping and problem-solving skills?</td>
</tr>
<tr>
<td>Engagement and connections in two or more of the following contexts: school, with peers, in athletics, employment, religion, culture</td>
<td>In what ways might participation in the metaverse enable or hinder engagement and connection between two or more of the following contexts: school, with peers, in athletics, employment, religion, culture?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>INTERPERSONAL PROTECTIVE FACTORS</th>
<th>RESEARCH QUESTIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Family provides structure, limits, rules, monitoring, and predictability</td>
<td>What do we know about ways families can provide structure, limits, rules, monitoring and predictability around young people’s participation in the metaverse, or with digital media more broadly?</td>
</tr>
<tr>
<td>Family provides fact-based information and support</td>
<td>What do we know about the importance of families providing youth with fact-based information and support about the risks and opportunities associated with their children’s participation in the metaverse?</td>
</tr>
<tr>
<td>Supportive relationships with family members</td>
<td>What do we know about the kinds of supportive relationships young people need with family members around their use of digital media, and in particular participation in virtual worlds? What factors influence or hinder the development of such relationships?</td>
</tr>
<tr>
<td>Clear expectations for behavior and values</td>
<td>What do we know about ways families can set clear expectations for behavior and values around young people’s engagement with metaverse-like experiences?</td>
</tr>
<tr>
<td>COMMUNITY PROTECTIVE FACTORS</td>
<td>RESEARCH QUESTIONS</td>
</tr>
<tr>
<td>------------------------------------------------------------------</td>
<td>--------------------------------------------------------------------------------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Presence of mentors and support for development of skills and interests</td>
<td>What do we know about the role of mentors in metaverse-like experiences in supporting the development of young people’s skills and interests? What makes such relationships helpful and/or harmful?</td>
</tr>
<tr>
<td>Opportunities for engagement within the community</td>
<td>What kinds of opportunities for engagement within a community might young people have in the metaverse? What kinds of things facilitate or hinder such engagement?</td>
</tr>
<tr>
<td>Positive norms</td>
<td>What do positive norms in the metaverse look like? How are they created and reinforced?</td>
</tr>
<tr>
<td>Clear expectations for behavior</td>
<td>What do we know about how communities (online/offline) can set clear expectations for behavior by young people in metaverse-like environments and experiences?</td>
</tr>
<tr>
<td>Physical and psychological safety</td>
<td>What do you know about the potential risks and opportunities afforded by metaverse-like environments around young people’s physical and psychological safety?</td>
</tr>
</tbody>
</table>