Co-designing Metaverse Ethics: Perspectives of Jamaican Youth

Ethical Guidelines

An initiative hosted by

Project Amplify

YOUTH CAN DO IT

Jesus College OXFORD

Project Amplify

YOUTH CAN DO IT

Jesus College OXFORD

Ethical Guidelines

Project Amplify is a free initiative that aims to elevate the voices of young people in the design and use of new and emerging digital environments. The 2023 implementation of Project Amplify focused on the Metaverse. One of the cornerstones of Project Amplify is the development of a set of ethical guidelines by which other organisations, developers, government agencies and policymakers can inform their own design processes of the Metaverse and other digital spaces.

1 PRIVACY & DATA PROTECTION

Consider how personal data is collected, stored and shared and how that is communicated to the end user. Recommendations include: collect and store only necessary data, require multi-factor authentication, utilise encryption best practices, decentralise data storage, simplify Terms and Conditions (T&C's) and incorporate multimodal access to T&Cs.

2 AUTONOMY

Autonomy refers to the amount of control participants have as they interact with and within a digital environment. Here we prioritise creative expression, self-paced/self-guided onboarding and navigation, clear and enforceable reporting mechanisms, and explicit guidelines of what it means to be in community. Data protection is also a key feature of a space that fosters user autonomy.

3 HUMAN FLOURISHING

An ethically-designed space promotes the well-being of its users. Some ways to accomplish this are through the inclusion of mental health resources, implementation of parental controls particularly for youth participants, and the adoption of clear and enforceable community guidelines. Finally, there is power in simple positive nudges including allowing users to track their usage and reminding them to take breaks.

4 ACCESSIBILTY & INCLUSIVITY

Adopting Universal Design Principles for Learning, providing multi-language translation, having dynamic support channels, and reducing the connectivity requirements are all ways to make one's platform more accessible. Equally important is ensuring individuals feel valued and respected, and ultimately included. Diverse representation in the team of developers as well as a broad range of avatars for users are some considerations which engender inclusivity.

5 REDUCING DECEPTIVE DESIGN PATTERNS

Reducing deceptive or dark design patterns starts with employing an ethical code of conduct to which one's developers are beholden. Consider how the very design of the space and the devices used to access it could be adversely manipulated and safeguard against these possibilities. Furthermore, educate users and allow for an effective evaluation, feedback and refinement loop.

LEARN MORE

Table of Contents

SUMMARY

Co-designing Metaverse ethics: Perspectives of Jamaican youth Ethical Guidelines

Project Amplify: An initiative hosted by Youth Can Do. I.T. & Jesus College, Oxford

Project Amplify is a free initiative that aims to elevate the voices of young people in the design and use of new and emerging digital environments. The 2023 implementation of Project Amplify focused on the Metaverse. The project is a collaboration between Youth Can Do I.T., (YCDI) and Jesus College, Oxford. YCDI is a dynamic youth-focused organisation committed to enhancing technological literacy and creativity, empowering youth, and fostering ethical digital citizenship. Amplify was funded by Jesus College, University of Oxford. As a constituent college, Jesus College actively contributes to shaping intellectual discourse and advancements across various disciplines, including those related to science, technology and youth participation.

The deliverables of Project Amplify included: 1) A set of ethical guidelines by which other organisations, developers, government agencies and policymakers can inform their own design processes and 2) A final report documenting the research rationale, methodology, findings and recommendations of the project.

This document is the manifestation of the first deliverable – a set of ethical guidelines co-created with the young people, and informed by the workshops and guided sessions led by academics and industry experts and the university-aged facilitators. The document provides a set of guidelines that emerged around the following themes: 1) **Privacy and Data Protection**, 2) **Autonomy**, 3) **Human Flourishing**, 4) **Accessibility and Inclusivity** and 5) **Reducing Deceptive Design Patterns**. Each theme begins with a brief introduction or overview, provides the guidelines in a checklist format, and concludes with a rationale for the guidelines provided.

This document is the result of a co-design process involving both adults (facilitators and project coordinators) and young people working together to develop these guidelines. The participants, ranging in age from 10 to 17 years, cited their involvement in this process as critical, sharing the following perspectives:

- "*I think young people's voices are essential in the co-design process because they [adults] might not know what young people want.*"
- "*What older people think is not what young people consider fun and we need to make unique things with our imagination and not close it up [limit our imagination].*"
- "*Young people might come up with really good ideas.*"

For their part, the facilitators noted that including the youth perspective was a matter of respect and an acknowledgment of the participants' rights. Furthermore, by incorporating the youth voice at the design level, it ensured their perspectives would be reflected in the final product.

The deliberate inclusion of the youth voice in the co-design process offers valuable insight. Particularly in instances where the target audience, either partially or wholly, is composed of youth, we posit it is also an ethical responsibility to consider how and why youth will interact with the intended product from their vantage point. Our project found that the value lies in the reality that there may be a disconnect between an adult's and youth's perspective. Technical development as a co-design process also recognises the young person as an individual capable of shaping the world around them. Simultaneously, it acknowledges the need for adult guidance, supervision and direction.

To learn more about the research underpinning the aforementioned themes, and subsequent guidelines, you can read the full report here: *Project Amplify Final Report*

For an interactive checklist that can be used to either self-assess your organisation's current incorporation of the ethical guidelines, and/or to plan how best to incorporate them moving forward, you can check out our *Ethical Guidelines Tool: ycdi.online/ethical-assessment.*

However, these intricacies are thoroughly addressed and represented in our complementary ethical guidelines document.

We extend our sincere gratitude to all the young people and their parents/guardians, the University-aged facilitators, the expert presenters who conducted the capacity-building training sessions, and the YCDI staff for their time and dedication to Project Amplify. We also thank Jesus College, Oxford for their collaboration, generous funding and support of this initiative.

Suggested citation:

Johnston, S-K., Wilson, C., & Derrell, C. (2024). *Co-designing Metaverse ethics: Perspectives of Jamaican youth.* Youth Can Do I.T., Jesus College, Oxford. https://doi.org/10.5287/ora-bpdmxbwdo

License:

Additional Notes:

Please note that the views expressed in this document are those of the authors and do not represent the views or positions of Youth Can Do I.T., Jesus College, Oxford or The University of Oxford. Any errors or omissions are those of the authors.

Project Amplify

Jesus College OXFORD

PRIVACY AND DATA PROTECTION

PRIVACY AND DATA PROTECTION

Overview

Privacy and data protection refer to the principles, legal frameworks, and practices aimed at safeguarding individuals' personal information from unauthorised access, use, disclosure, alteration, or destruction. *How can we ethically design the Metaverse to ensure we promote privacy and data protection?*

Guidelines

Terms and Conditions (T&Cs): Identify multimodal options for presentation

- ☐ Use animation to communicate terms of use and privacy policies.
- ☐ Include a voice-activated reader feature.

Rationale: People access, process and retain information differently. A visual representation can enhance user comprehension of the data collection process. Similarly, using a voice-over allows the information to be assimilated in a different way, catering to different learning needs and abilities.

Terms and Conditions (T&Cs): Simplify the language used in accordance with legal requirements.

- ☐ Present a concise but comprehensible and informative list of privacy rules and settings.
- ☐ Have clear and succinct explanations of how data will be accessed, processed, shared and stored.

Rationale: Simplification of language makes the T&Cs more accessible. Importantly, it allows for **informed consent** as it increases the likelihood that the user will understand the terms to which they are agreeing. The key caveat here is "in accordance with legal requirements" as one may be legally bound to include codified terms and jargon. However, including a simplified version will allow for greater engagement and a better understanding of the terms.

Terms and Conditions: Demonstrate transparency in how data is being used in the Metaverse

- ☐ Present a visual representation of how data is being accessed and used.
- ☐ Include information about the trends in virtual reality (VR) data usage to promote more informed choice about an individual's data can be used in the Metaverse.

Rationale: Persons interacting with these environments may not be technologically aware or well versed in how their data is being used. By understanding not only how their individual data may be used but the trends in VR data usage, it again allows the user to provide informed consent and enhance the user experience.

Privacy Structures: Build the Metaverse on privacy structures such as SOLID

- ☐ A play on Solid Pods, *MetaPods* can be created for the Metaverse. Users can either be instructed on how to set up their own or the MetaPods could be created as a standard user feature. If the former,

instructions would be embedded within the design of the Metaverse spaces. At its core, this is also an autonomy-enhancing structure.

- ☐ Demonetise personal data and ban the sharing of user data to third parties without explicit user consent.
- ☐ In addition, regular [sanity] checks must be conducted to ensure that the users' data is secured.

Rationale: Solid pods allow for the decentralised storage of data and give the user control over how their data is accessed, shared and stored. By adopting a similar framework within the Metaverse, participants become actively involved in how their data is stored and shared, which also encourages more intentionality around how they engage in the space. Additionally, and perhaps controversially, banning the sale of data to third parties could likely increase user trust and engagement, which could lead to a higher volume of users on the platform. Here organisations will have to confront the tension between profitability and data privacy. At a minimum, user consent should be obtained for the sharing of data with third parties. Finally, sanity checks enable site administrators to demonstrate and confirm the structures are working as designed.

Data Security and Encryption

- ☐ Require two-step verification.
- ☐ Restrict the sharing of identifiable information.
- ☐ Have password-enabled features required to access privacy settings.
- ☐ Prevent the taking of screenshots for sensitive content or make screenshots unclear once a camera is detected, for example, in cases where people take a picture with an external (phone) device.
- ☐ Implement privacy considerations into the design process from the start.
- ☐ Require verification for syncing when allowing for sign-in using another account, such as an email account.
- ☐ Employ a design feature that randomises the type of authentication mechanisms required each time a user is accessing personal data. For example, across a three-day period, on Day 1, the user would authenticate with a password, Day 2 via FaceID and Day 3 using a fingerprint. This design feature would be supported by a mechanism that randomises the different authentication types of each time access is requested.
    - This design can be supported by a double layer of protection requiring two-step verification. In this scenario, a code is employed alongside a password, fingerprint or FaceID.
    - If incorrect information is given a predetermined number of times (e.g., three times), this prompts the account to be deactivated for 24 hours.
- ☐ Implement a child-friendly version of the aforementioned feature which allows children to give assent as well as authenticate access to their personal information using different mechanisms.

Rationale: Additional layers of data protection and encryption offer a safeguard against a user's data being stolen, manipulated or misused. Additionally, by varying and combining authentication methods across the different security types: physical (e.g., a key card or debit card), what you know (e.g. a Personal Identification Number [PIN]), and who you are (e.g. retina scan, fingerprint) it further reduces the likelihood of a hacker or nefarious party being able to access a user's data.

Furthermore, adopting a child-friendly version allows children to better understand the concept of informed consent and impresses on them the importance of safeguarding their personal data.

Minimise Data Collection

- ☐ Reduce the collection/saving of passwords and other personal information (e.g., credit card details).
- ☐ Retain data only as long as necessary, and prioritise user consent.
- ☐ Classify or tag apps that may utilise a user's data for purposes other than those to which the user consented.
- ☐ Passwords should not be held by any apps outside of a designated decentralised data store, such as a MetaPod, explicitly designed for that purpose and properly encrypted. Accessing the Metaverse or another digital space, or connected app, would initiate a verification process through the data store.
- ☐ It is important to consider that with all the collection of multiple data for authentication, there is a risk of over-collection of personal data. Therefore, there needs to be accountability frameworks that govern the management of this information for authenticators.

Rationale: Many apps collect unnecessary data that neither enhances their functionality nor improves the user experience. By reducing the amount of data collected from individual users and managing data access through a secure data store, trust and user engagement can be significantly improved.

Communication

- ☐ Allow people to communicate within a dedicated space shared with their trusted contacts, ensuring a sense of security and familiarity. However, it is important to balance this with opportunities for exposure to diverse perspectives, to prevent the formation of echo chambers where only similar viewpoints are reinforced.

Rationale: Dedicated spaces reduce the risk, particularly for younger users, of engaging with parties that may have malicious intent.

Personas

- ☐ Encourage creative use of personas to protect Metaverse users. Avatars and personas can serve a dual purpose of representing how a user wants to appear online while strategically protecting one's identity. The caution here is that people could abuse this feature by misrepresenting who they are to gain the trust of a vulnerable person or group or misrepresenting specific political, ethnic, cultural or religious groups. Therefore, an avatar or persona creation should still be guided by a set of enforced community guidelines.

Rationale: An avatar or other online persona helps us to curate or manage our digital identity by allowing for the freedom of self-expression, within (suggested) parameters designed to not only protect our privacy, but also our safety.

Project Amplify

AUTONOMY

AUTONOMY

Overview

Autonomy involves placing individuals' values at the core of their interactions in digital environments, including the collection, use, and storage of their data. It also extends to empowering people in how they create, share, and manage information, including having control over the information they receive. *How can we ethically design the Metaverse to ensure we promote autonomy?*

Guidelines

Creative Expression

- ☐ Allow people to have control over how they express themselves in digital spaces (e.g., creative expression through pictures, text, sound, etc.).
- ☐ Provide the ability to use fun filters. These filters must make a reasonable attempt to reflect a wide range of ethnicities, be inclusive of the disabled community, and generally allow for representation across as many communities as is feasible. This is also important for advancing inclusivity.

Rationale: Self-expression is one way to promote autonomy thus enhancing the user experience. This does not however mean the promotion nor acceptance of slander, inappropriate, or malicious content. Thus, autonomy is not the absence of regulation but instead an understanding of how the tools and features of the space can be used to enhance the individual and collective user experience.

Self-guided/Self-paced Onboarding

- ☐ Embed a design feature which offers self-paced experiences to support the onboarding process of engaging in the Metaverse. It should include key safety features that users should consider.
- ☐ The onboarding guide should be dynamic and iterative to account for any new considerations of how to be more autonomous in the Metaverse.
- ☐ Allow people to freely access materials online in different forms for both educational and entertainment or edutainment purposes (including but not limited to video, text and audio).
- ☐ Have clear written guidelines under a 'learn more' or similarly labeled link about what is available in the Metaverse. There should also be the option to activate automated fact checking of the information available to ensure it remains accurate, current and that there are no links to unsolicited advertisers.
- ☐ Put terms and conditions in a strategic position which is likely to enhance visibility, access and engagement. Strategic in this sense is both contextual and relative and may manifest differently across various platforms. Thus, platform owners should demonstrate the efficacy of their selected methods through published reports on user engagement and read times of their terms and conditions.

Rationale: Onboarding processes for new apps, websites and technology can often be tedious and not human-centred. Either the "rules of the space" are contained in lengthy text documents replete with jargon, or presented in a single modality that may only be seen or readily accessible on initial login or activation. The aforementioned guidelines suggest an onboarding process that is driven by the user and regularly updated to allow for a better user experience.

Reporting Mechanism

- ☐ Incorporate a reporting mechanism that has a built-in feedback loop which allows for timely access to an administrator, and a prompt and appropriate response when someone perceives that their rights have been violated.

Rationale: Allows users to feel empowered, safe and supported when using the Metaverse.

Community

- ☐ Provide the opportunity to make different communities for varied purposes and interests.
- ☐ Reduce instances of direct messages from advertisements: Design a function in the chat box to report unsolicited advertising to the administrators and block advertisers' messages.

Rationale: Users often interact in spaces such as the Metaverse with the hope of finding/building community. By allowing users to shape what their community looks like (and what it doesn't), this provides a sense of autonomy over their environment.

Data Protection

- ☐ Alert participants when a new user enters the space.
- ☐ Incorporate an AI-enabled feature that warns when a user is sharing information with an unverified source. The warning should include the possible risks such as vulnerability to hacking. This information should be presented in a clear manner so that people, without prior knowledge of what hacking involves, understand the risks of sharing their personal/sensitive information.
- ☐ Create an AI-enabled mechanism which selectively filters out sensitive information that users do not wish to share with third-party applications. The filtering of information would be based on preset controls by the user.

Rationale: These features ensure users are informed about potential risks in a manner which is accessible regardless of their level of familiarity with such spaces and terminology. It also provides active steps for them to protect themselves, thus increasing their autonomy.

Accessibility of Terms and Conditions (T&Cs)

- ☐ Create opportunities for T&Cs to be more understandable for the general population to allow people to feel more empowered about the decisions they make in the Metaverse.
- ☐ Include an app that can transform complex information into multiple, easily understandable representations. Supporting features of this app could include:
    - The ability to have a summarised and simplified version of the T&Cs by copying and pasting it into the app.
    - A design feature which allows users to create an AI-generated video of the T&Cs. This can also increase the accessibility of the information.
    - The opportunity to connect to more information and resources. Two possible formats are: 1) provision of automated responses for simple requests which could be enhanced by a dynamic, interactive and animated chatbot, and (2) access to a trained (human) T&Cs agent.
    - A feature in the form of a word highlighter that highlights the most important parts of the T&Cs document for users. The core applications of this feature include:
        - Highlighting complex words which users can select to determine a more simplified meaning of the word/concept.
        - Permitting users to continue to simplify the T&Cs as many times as they desire to ensure they understand the document (e.g., the use of a slider which reduces or increases the level of T&Cs complexity).
        - Highlight the keywords in the T&Cs.
    - Another enhancement is a design feature which converts the T&Cs into animations. The core aspects of this feature could include:
        - Highlighting the key aspects of the T&Cs: These key aspects could be tailored to users based on a mechanism that collects expressed preferences regarding what is most meaningful to them. For example, one user might prioritise understanding how their data is being used, while another might be more concerned with knowing with whom their data is shared.
        - A narrator feature that allows users to select a specific voice and language based on their preferences and needs. This option can also contribute to enhanced accessibility.
        - The incorporation of universal visualisations that would be understandable to people globally while allowing context and culturally specific options, as needed. This can be supported by generative AI and also is linked to enhancing accessibility and inclusivity.

Rationale: All the suggested design features, while neither exhaustive nor non-negotiable, approach autonomy through accessibility. If users do not fully understand the Terms and Conditions presented, nor the expectations surrounding how to use the space, it will ultimately hinder their ability to be autonomous, informed and engaged users.

Authentication

- ☐ Autonomy can be improved if people have *multiple* and *context-specific* authentication mechanisms. This is because certain contexts may be better suited to different types of authentications, such as fingerprints, facial ID, or passwords.

Rationale: This approach bridges data protection and autonomy by enabling users to choose the method of data sharing they find most comfortable, while also providing information on why one type of authentication may be more effective than another.

Design

- ☐ Implement a structured process (e.g., a life cycle) or mechanism that will help Metaverse users to maintain their autonomy in an ethical way. This should be executed from the moment that users enter the space to the point at which they disengage and possibly re-engage. The life cycle would:
  - Allow users to see in real-time what data are being collected and how they are being used.
  - Generate periodic reports summarising data usage, access, and sharing activities.
  - Offer a structured process for users to disengage from the Metaverse, ensuring they can easily export or delete their data.
  - Inform users about any significant changes in data policies or features that might impact their autonomy, encouraging them to re-engage with updated terms.
  - Allow users to provide feedback on their autonomy experience and suggest improvements.

Rationale: By developing an autonomy life cycle of a typical user at the design phase, developers can better ensure that an autonomous usage is integrated into each stage of the user experience.

General

- ☐ Autonomy should be used in a way that does not contradict social good. There should be an ethical aspect of using one's autonomy to benefit the Metaverse or digital space (e.g., sharing data if it will help to have a net positive impact). This means that mechanisms should be put in place for users to have meaningful reflections on their own decision making, and features should be implemented to ensure that people are thinking and acting critically about their own decisions. This prompting for reflection may also generate critical feedback which can also be channelled in a way that is beneficial for the community.

Rationale: This approach bridges data protection and autonomy by enabling users to choose the method of data sharing they find most comfortable, while also providing information on why one type of authentication may be more effective than another.

Project Amplify

Jesus College OXFORD

HUMAN FLOURISHING

HUMAN FLOURISHING

Overview

Human flourishing, as collectively understood by participants, involves the holistic development and fulfilment of individuals, encompassing physical, emotional, mental, and social well-being. Applied within the context of Amplify, *Human Flourishing* describes the idea that when people are online, they are safe, and they feel that their well-being is considered. *How can we ethically design the Metaverse to promote human flourishing?*

Guidelines

Mental Health Resources

- ☐ Wellness checks and breaks should be built into both the design and the practice (encouraged use) of the Metaverse and similar spaces. Interactions within the space should start with mental health prompts. This could be achieved using an AI-generated bot that provides a validated wellness/human flourishing scale and then alerts an appropriate administrator if there are concerning responses. Violations of the space guidelines and reports of same could be made using the same AI-generated bot mechanism.
- ☐ Mental Health resources can be made accessible in varied and engaging ways.

Rationale: The check-ins serve as a reminder that mental wellness is an important part of one's overall well-being. However, as neither the administrators nor the community may be mental health experts, it is important that persons are linked to resources that can provide support rather than leaving that onus on other users of the space.

Parental Controls

- ☐ Parental controls should include a ban on explicit content, use of derogatory language, and evidence-based limits on the screen time per week.
- ☐ Establish opportunities for meaningful parent collaboration with minors in the Metaverse.

Rationale: As young people/minors may be one of the primary users of the space, this allows parents to be involved in their child's Metaverse use while setting reasonable boundaries for safe exploration.

Community Guidelines

- ☐ Establish clear and enforceable community guidelines that outline what is permissible in the Metaverse and the consequences for violating these rules. Guidelines should explicitly define terms that may be open to interpretation and share how those definitions were reached. Furthermore, provide a rationale for why actions such as, but not limited to, cyberbullying, derogatory content and harassment are prohibited.

- ☐ For spaces that cater to children and more broadly minors, we recommend the prohibition of gambling. One way to enforce the same is banning crypto currency or any form of digital currency that could be converted to cash in Metaverse spaces designed for minors.
- ☐ Promote use of the Metaverse (and similar spaces) that encourages users to be mindful of the intent and possible impact of how they engage in the space, not simply out of fear of consequences from violating guidelines, but in awareness of the repercussions on the community.
- ☐ Enforce the community guidelines in a timely manner to encourage adherence to the same and to ensure that inappropriate behaviour is swiftly addressed.
- ☐ Allow the community to participate in the creation of the community guidelines: this could serve as an incentive to encourage co-designing and prevent future violations.
- ☐ Incorporate a feature that would detect any derogatory words which either triggers a warning or removal of the offender, depending on the severity of the offense.

Rationale: If the community guidelines have been clearly communicated, then it is easier for both the community and the administrators to hold violators accountable.

Monitoring and Reporting

- ☐ Prevent links promoting fake news, phishing sites, and harmful or malicious content from being shared.
- ☐ Incorporate strict monitoring of chats to prevent cyberbullying by enabling AI-powered detection of special characters, images, emojis and symbols that could be deemed harmful, including speech that actively advocates for violence.
- ☐ Provide users with a reporting system supported by a backend team. Where feasible, include trained professionals (volunteer or paid), for example psychologists, to help persons more responsibly and safely navigate the Metaverse.
- ☐ Include language translators that help to identify harmful speech across multiple languages.

Rationale: Any enforcement of community guidelines will require an effective monitoring and reporting system, as well as dedicated resources to address concerns that arise.

Positive Nudges

- ☐ Include a feature to track and monitor one's use of the Metaverse.
- ☐ Incorporate reminders for users to take breaks from using the Metaverse .

- ☐ Ensure users are truly retaining their autonomy by prompting them to disengage after a certain amount of time and provide encouragement and examples of activities they can partake in offline.

Rationale: Users can lose track of the time spent in the Metaverse. Positive nudges encourage a healthy and balanced user experience, again particularly important for a youth demographic.

Project Amplify

Jesus College OXFORD

ACCESSIBILITY AND INCLUSIVITY

ACCESSIBILITY AND INCLUSIVITY

Overview

Accessibility pertains to the design of products, services, and environments in a way that enables equitable access and participation for people with diverse abilities, including those with disabilities. Inclusivity, on the other hand, refers to the principle of ensuring that all individuals, regardless of their background, identity, beliefs or abilities, feel valued, respected, and represented within a given context. *How can we ethically design the Metaverse to promote accessibility and inclusivity?*

Accessibility Guidelines

Design Principles

- ☐ Design the user interface to be accessible to individuals with visible and non-visible disabilities, including features such as easy-to-read fonts, color contrast adjustments, captions, text-to-speech. This could be especially important for people with dyslexia.
- ☐ Implement a voice-activated control system to enable participation from persons who may not be able to physically move the different components of their device (e.g., people with cerebral palsy)
- ☐ Ensure that the Metaverse is accessible to individuals with disabilities by incorporating Universal Design for Learning Principles (UDL).

Rationale: People have a diverse range of physical and cognitive abilities. By incorporating design principles that consider this range, it allows the Metaverse to be accessible to a greater user base and demonstrates accessibility as a priority and not an afterthought.

Language

- ☐ Include language translation as a standard feature.
- ☐ Embed a tool that can translate audio to sign language.
- ☐ Provide sign language as a form of communication between/among people, similar to how one can express themselves using emojis.

Rationale: Users can optimise their experience by participating in their native or preferred language, thus making the platform more accessible. Incorporating sign language (instead of merely enabling captions) also allows for greater inclusion in the space. Generally, the integration of these language features recognises not all users have the same “default”.

Support

- ☐ Provide a help feature that can be activated through audio, text, or movement for people who do not understand a specific area/feature when they are in the Metaverse.

Rationale: Acknowledges different learning styles, needs and abilities, allowing users to more quickly access help/support.

Bandwidth/Connectivity Requirements

- ☐ Enable core features of the Metaverse to function offline, allowing users to save certain processes or features locally and sync them upon reconnection. This ensures that individuals without continuous internet access can still interact with the Metaverse and have their offline activities automatically synced when they reconnect.
- ☐ Design the Metaverse to allow certain spaces or features to be accessible using only a fraction of the usual bandwidth. This may involve reducing the display quality of some elements to prioritise functionality and accessibility.

Rationale: The Metaverse's high bandwidth and data demands present significant barriers to mass adoption. Addressing these challenges could enable greater participation, particularly from regions like Latin America and the Caribbean, as well as the wider Global Majority.

Inclusivity Guidelines

Design and Developer Teams

- ☐ Recognise that inclusion also means incorporating people's thoughts and ideas in the design of the Metaverse. Design and developer teams should be intentional about including different lived experiences (e.g., race, socio-economic status, culture, gender, ideological beliefs) and decision making should benefit from this diversity and inclusion of perspectives.

Rationale: Diversity policies, when implemented correctly and fairly, are positive and allow for blind spots and gaps to be uncovered and filled. Often the focus is on diversity of users or is incorporated at the final product stage instead of ensuring different experiences and perspectives are also reflected in the decision makers. It is crucial to implement genuine human-centered training focused on designing an ethical Metaverse. The discussion should expand beyond simply advocating for an ethical Metaverse to actively cultivating ethical developers.

Diversity of Thought

- ☐ Establish dedicated Metaverse forums and virtual spaces where users can engage in discussions on controversial topics, fostering respectful dialogue and exchange of ideas**.**
- ☐ Develop algorithms that prioritise diversity in content recommendations, ensuring users are exposed to a wide range of unbiased perspectives and information sources.
- ☐ Design recommendation systems that mitigate echo chambers by promoting content that challenges users' viewpoints and encourages exploration of diverse opinions.

- ☐ Incorporate tools within the Metaverse that support critical thinking, such as fact-checking resources, logical reasoning exercises, and simulations for decision-making in complex scenarios.

Rationale: Upholding diversity in thought advances enriching discussions, fosters innovation, and promotes understanding by exposing individuals to diverse perspectives, challenging assumptions, and facilitating informed decision-making, thereby contributing to a more inclusive space, and more broadly, society.

Themes and Motifs

- ☐ Use colours, motifs and visuals that can be easily changed to represent the design for various cultural, national or international events or holidays.

Rationale: This will serve both as a learning tool but also a way for users to see themselves reflected in the Metaverse.

Avatars and Representation

- ☐ Expand the range of choices for users when customising their avatars, including hair, skin complexion, hair type, body shape, etc.

Rationale: Often, the precise skin tone, hair type, or body shape one desires is unavailable, leading users to settle for something approximating their own features. During our development of the Metaverse spaces, users spent a significant amount of time selecting an avatar that they felt represented them or how they wanted to be perceived. By offering a wider selection of options, individuals will experience a greater sense of inclusivity and authenticity.

Language and Translation

- ☐ Provide multilingual support, including automatic, real-time translations, to accommodate a wide range of languages.

Rationale: This allows for greater participation from a range of users. This feature also allows for persons who speak different languages to interact with each other.

Project Amplify

Jesus College OXFORD

REDUCING DECEPTIVE DESIGN PATTERNS

REDUCING DECEPTIVE DESIGN PATTERNS

Overview

Deceptive design patterns encompass a range of strategies employed in digital interfaces to manipulate user behaviour, often leading to undesirable actions or outcomes from the user perspective. How can we ethically design the Metaverse to ensure we reduce deceptive design patterns (also called dark design patterns)?

Guidelines

Ethical Code of Conduct

- ☐ Provide a code of conduct to which the developers of the space are also beholden.

Rationale: If the core values, training (including continuous professional development) and practices of your team (designers, developers, administrators, etc.) are founded on ethical principles, and save for ethics washing or ethics lip service, then it will be more likely to be reflected in the design and final product.

Terms and Conditions

- ☐ Provide a design feature where the complexity of the terms and conditions can be toggled from beginner to expert language. This also reflects people's minute by minute changing values or preferences for complexity.

Rationale: Allowing users to engage based on their capacity, comfort level and experience is likely to increase the accessibility of the terms and conditions as well as this likelihood that users understand both the terms and the consequences of breaching the same. This allows for true informed consent.

Design Principles and End User Education

- ☐ Provide skill building opportunities for users so that they can better recognise dark patterns in design. This fosters important skills such as critical thinking skills as well as critical thinking disposition (i.e., the willingness to use one's critical thinking). Here are examples of dark patterns that users could become more aware of through such educational initiatives and developers should be mindful of in their design:
    - *Emotional Manipulation through Biosensors*: As wearable biosensors become more integrated with digital interfaces, dark design patterns may exploit these sensors to detect users' emotional states. Interfaces could then be adjusted to exploit vulnerable emotional states for persuasive or manipulative purposes. Intrusive monitoring of emotional states raises significant privacy concerns, as users may unknowingly share sensitive information about their mental and emotional well-being.
    - *Augmented Reality Distractions:* Dark patterns in augmented reality interfaces may deliberately introduce distracting elements to divert users' attention and encourage unintended interactions or purchases. User's autonomy is compromised as their attention is manipulated, leading to decisions that may not align with their original intentions.

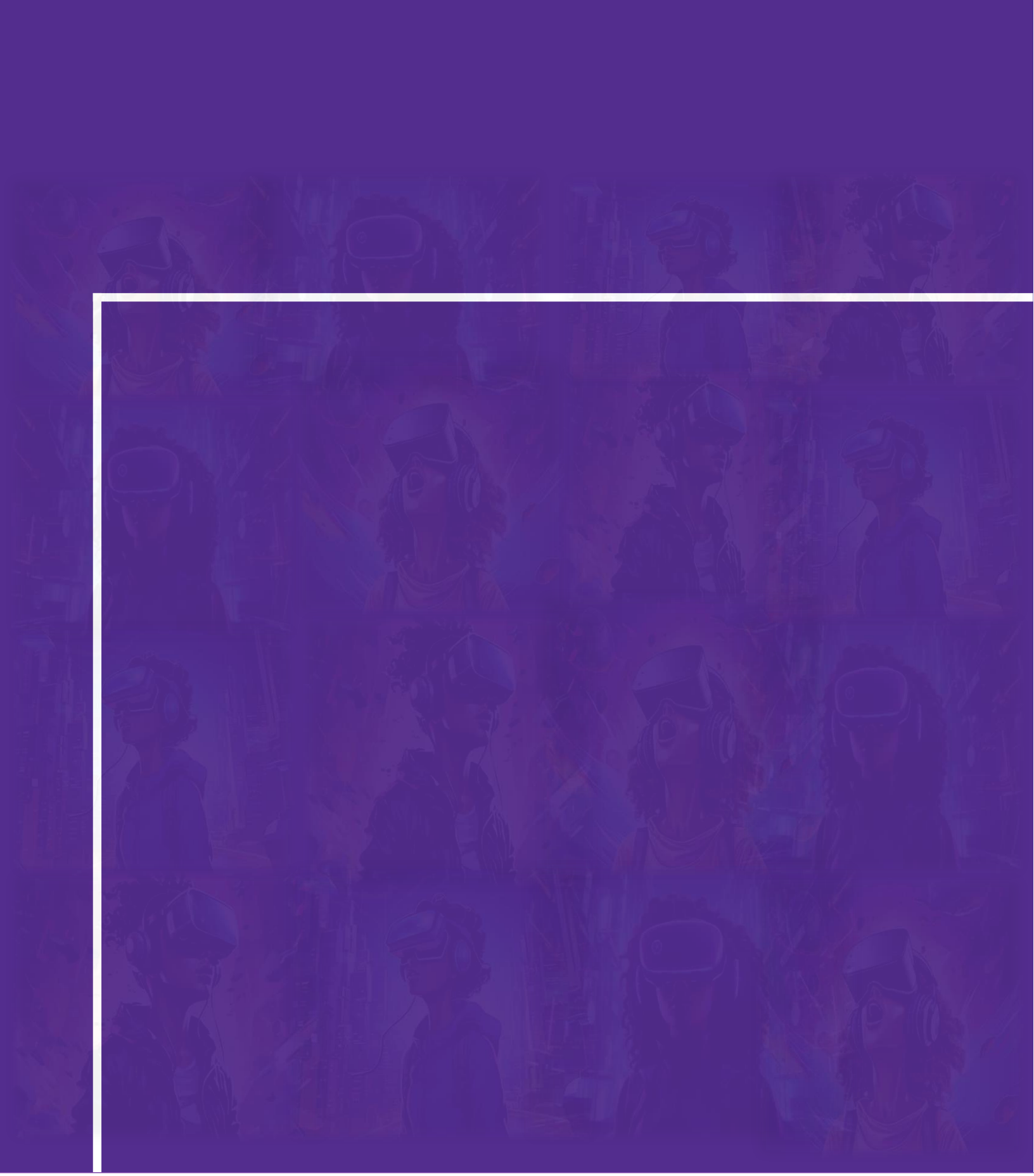
- *Behavioral Prediction Algorithms:* Advanced algorithms may predict user behavior with increasing accuracy, allowing platforms to strategically deploy dark patterns based on individual behavioral profiles to maximise engagement or sales. Constant monitoring and prediction of user behavior could erode privacy, as individuals may feel exposed and manipulated without their explicit consent.
- *Biometric Deception in Authentication:* Dark design patterns may exploit biometric authentication by employing deceptive interfaces that encourage users to unintentionally provide biometric data for unauthorised access. Such practices pose serious privacy and security risks, potentially leading to unauthorised access to sensitive information.
- *Algorithmic Gaslighting:* Algorithms may be designed to subtly manipulate users' perceptions of reality, making them doubt their own experiences or beliefs through tailored content and information. This may lead to algorithmic gaslighting. Users' autonomy is compromised as their understanding of reality is subtly distorted, influencing decision-making and beliefs.
- *Dynamic Social Proof Manipulation:* Dark patterns may dynamically manipulate social proof elements (e.g., user reviews, likes) in real-time to create a false sense of urgency or popularity for certain actions or products. Users may find it challenging to trust online information, as social proof loses its authenticity and becomes a manipulated tool.

Rationale: By educating users on how to identify deceptive patterns, it holds developers of the space to a higher standard as the community becomes the informal regulator. It also allows users to be aware of potential risks and dangers.

Evaluation, Feedback and Reporting

- ☐ Develop indicators such as Trust Indicators which users are first informed about and which can then be used to rate the Metaverse space and provide feedback.
- ☐ Identify or flag platforms which have been reported as being non-compliant as well as the number of times a platform has been reported.
- ☐ Create a feedback loop where users can rate and comment on the ethical design of the site. Feedback should also be provided to the users on the results.
- ☐ Incorporate the ability to report deceptive design patterns.
- ☐ Implement a feature that can automatically identify fake and malicious accounts for users.
- ☐ Implement stricter penalties for using deceptive patterns in design. Overall, greater proactive regulation is needed.

Rationale: By educating users on how to identify deceptive patterns, it holds developers of the space to a higher standard as the community becomes the informal regulator. It also allows users to be aware of potential risks and dangers.



Project Amplify: A collaboration between Youth Can Do I.T. and Jesus College, Oxford