Journal of Management and Business Education

JMBE

www.journaljmbe.com ISSN: 2605-1044 Published by Academia Europea de Dirección y Economía de la Empresa.
This is an open access article under the CC BY-NC license.

How helping students design ethical metaverse platforms can lead to safety and wellbeing for all

Cómo ayudar a los estudiantes a diseñar plataformas éticas en el metaverso puede contribuir a la seguridad y el bienestar de todos

Binod Sundararajan
https://orcid.org/0000-0002-4739-5842 (ORCID iD)
Dalhousie University (Canada)

Malavika Sundararajan
https://orcid.org/0000-0002-4003-9351 (ORCID iD)
Ramapo College of New Jersey (USA)

Sundararajan, B.; & Sundararajan,M. (2025) How helping students design ethical metaverse platforms can lead to safety and well being for all. *Journal of Management and Business Education*, 8(2), 313-328. https://doi.org/10.35564/jmbe.2025.0017

*Corresponding author: binod@dal.ca
Language: English
Received: 17 Feb 2025 / Accepted: 22 Jul 2025

Acknowledgments. We are grateful to our teachers for the inculcation of the ethical values.
Funding. The authors received no financial support for the research, authorship, and/or publication of this article.
Ethical Statement. The manuscript is fully based on secondary data available publicly and all original sources have been cited.
Declaration of conflicting interests. The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.
CRedit author statement. All authors have contributed equally to all phases of the research and manuscript preparation.

ABSTRACT

A lack of proximity and enhanced anonymity in virtual worlds seems to provide the license to Artificial Intelligence (AI) based metaverse users to misbehave. Bullying, abuse, spread of hatred and divisiveness and manipulations of minds in Metaverses are growing exponentially due to the speed and magnitude with which AI enabled bots in Metaverses can multiply and reproduce content. Online violence has begun spilling into the real world which is negatively impacting the psyche and wellbeing of children and young adults in society. Ethicists or well-meaning employees have spoken out against these violations in Metaverses. But we find many such ethics groups have been dissolved or silenced while employees who are whistleblowers are often fired, discredited, or dismissed. Business Ethics, Marketing, Management and Sustainability students are often asked to simply carry out an ethical analysis of cases and provide recommendations. While such processes have helped explore various ethical schools of thought, the application of these concepts to AI based metaverses seems less about what framework to apply and more about how to design

a fail-proof system to protect the safety and wellbeing of all. Such an approach will make students more aware of the consequences of their choices and develop a sense of responsibility towards the wellbeing of all. The paper proposes the use of a case study of a hypothetical company that has a metaverse platform and the challenges it faces in addressing abuse and scandals on its platform. The paper also offers a detailed review along with the pros and cons of several ethical frameworks and puts forth two key questions to students asking them to design a metaverse platform that can, a) ensure the wellbeing, security, and safety of the users who are not even aware that their minds may be swayed and manipulated; and b) find a way to convince companies that create AI-based metaverses to adopt ethical frameworks. Sample answers are provided to help faculty work with students understand the importance of always designing products and services with personal and others' well-being in mind rather than only making profits at the cost of people's safety and security.

Keywords. AI, Ethics, Metaverse, Wellbeing, Safety, Case Study

RESUMEN

La falta de proximidad y el aumento del anonimato en los mundos virtuales parecen otorgar una especie de permiso a los usuarios de metaversos basados en Inteligencia Artificial (IA) para comportarse de forma inapropiada. El acoso, el abuso, la difusión del odio y la división, así como la manipulación de mentes en los metaversos, están creciendo exponencialmente debido a la velocidad y magnitud con la que los bots habilitados por IA pueden multiplicarse y reproducir contenido. La violencia en línea ha comenzado a trasladarse al mundo real, lo que está afectando negativamente la psique y el bienestar de niños, niñas y jóvenes en la sociedad. Personas éticas o empleados con buenas intenciones han alzado la voz contra estas violaciones en los metaversos. Sin embargo, muchos de estos grupos éticos han sido disueltos o silenciados, y quienes actúan como denunciantes suelen ser despedidos, desacreditados o ignorados. A los estudiantes de Ética Empresarial, Marketing, Gestión y Sostenibilidad a menudo se les pide simplemente que realicen un análisis ético de casos y proporcionen recomendaciones. Si bien estos procesos han contribuido a explorar diversas corrientes del pensamiento ético, la aplicación de estos conceptos a los metaversos basados en IA parece depender menos del marco teórico que se aplique y más de cómo diseñar un sistema a prueba de fallos que proteja la seguridad y el bienestar de todos. Este enfoque puede hacer que los estudiantes sean más conscientes de las consecuencias de sus decisiones y desarrollen un sentido de responsabilidad hacia el bienestar colectivo. Este artículo propone el uso de un estudio de caso de una empresa hipotética que posee una plataforma de metaverso y enfrenta diversos desafíos relacionados con el abuso y los escándalos en su entorno virtual. Además, se ofrece una revisión detallada con los pros y contras de varios marcos éticos y se plantean dos preguntas clave a los estudiantes, instándolos a diseñar una plataforma de metaverso que pueda: a) garantizar el bienestar, la seguridad y la protección de los usuarios que ni siquiera son conscientes de que sus mentes pueden estar siendo influidas y manipuladas; y b) encontrar una forma de convencer a las empresas creadoras de metaversos basados en IA de adoptar marcos éticos. Se proporcionan respuestas de ejemplo para ayudar al profesorado a trabajar con los estudiantes en la comprensión de la importancia de diseñar siempre productos y servicios pensando en el bienestar propio y ajeno, en lugar de centrarse únicamente en obtener beneficios a costa de la seguridad y la integridad de las personas.

Palabras clave. IA, ética, metaverso, bienestar, seguridad, estudio de caso.

INTRODUCTION

What is the metaverse? Why has it become a problem? And why won't it be enough to have business students conduct regular analysis and application of ethical schools of thought to understand the workings of metaverses? - Metaverse is a shared virtual environment accessible via the internet (Muir and Hall, 2020). These virtual environments replicate physical worlds without the temporal and spatial boundaries (Zhuk, 2024). Although Metaverses offer real-time experiential learning and exciting creative opportunities to users, due to the actions of other users in the form of bullying, sexualization, and data and identity thefts among others they also, unfortunately, have highly detrimental psychological impacts on vulnerable populations who enter this universe (Sinha, 2023; Zhuk, 2024). Since the regulatory landscape for AI in virtual worlds remains nascent, a critical question raised by researchers is whether ethics is even computational to begin with (Johnson, and Verdicchio, 2023)? While it may appear efficient and attractive because of the ease with which one can formalize and compute rules, principles, theories, and frameworks, the lack of adherence to moral code by humans due to variations in perceived interpretations of unfairness or inequity can never be accounted for by technologists or ethicists. (Johnson and Verdicchio, 2023). Hence, a mere ethical analysis of business cases in the classroom may not suffice to instill moral intent and responsibility in our students.

In fact, one of the first things we teach business students is to assess the market share in an industry they want to enter or expand in. It is therefore natural for students or employees to design products and services that can cater to such large markets and make a profit on them. It should therefore not be surprising that when a market promises to be worth over US$13 trillion by 2030 (Zheng and Daugherty, 2023) a lot of people want a piece of that pie. If we are wondering why companies who create metaverse platforms allow hateful things to happen, it seems to point to the age-old choice of picking profits over people. Companies that create metaverse platforms allow people to choose whatever they want to pursue in the metaverse, but the problem arises when people's choices and actions lead to so much harm in the virtual world that its repercussions spill out into the real world.

Marketing students may be familiar with the term, caveat emptor' or 'buyer beware' (Hartman, et al, 2014), wherein the onus is on the customer or user to educate themselves about any product or service before they buy and accept consequences that come with the purchase and use of a product or service. While no product or service can be perfect, is it not the responsibility of the makers to at least provide safety instructions, safety nets, and support systems that will protect the users instead of harming them? What will convince marketing or other business or management students to look beyond the revenue streams?

In the real world, parents and teachers advise students to inculcate values and make informed choices about not talking to strangers, but in the virtual world with the advent of AI, the speed with which AI bots can spread information or reinforce thoughts that can be negative, is unimaginable. It spreads like uncontrollable wildfire. Children and young adults are also less likely to listen to any advice when they are experiencing heightened instant gratification in virtual worlds. The only other option is to ask the creators of these virtual platforms to be careful about what they allow online, and to program their AI bots to be vigilant about the safety of its users. Yet, the firing of Timnit Gebru at Google, or the changes to Meta's fact-checking program and the dissolution of several ethics groups across companies seem to indicate there is a concerted effort to devalue human dignity and human lives over profits. Thus, the primary purpose of this paper is to put forth the very question of how to design safe platforms and convince companies to do so, to the students themselves. Since the students are predominant users of AI and metaverses and will eventually also be part of the future workforce, having them literally design their world will help them explore the nuances and consequences of the choices, codes, and guidelines they opt to incorporate into their designed version of these metaverses.

The study begins with the review of the literature by focusing on the need for the design approach to addressing the ethical needs of AI and metaverses. It then discusses the value and harm of metaverses to users and finally highlights the ethical frameworks applied in AI and metaverses along with their pros and cons. Due to the extensive use of AI in metaverses and the nascent stage of ethics in virtual worlds, the ethical frameworks are borrowed from the literature on AI, metaverse, and social media or online technology platforms and summarized into three categories. Following this, a short case study on a hypothetical company that creates metaverses is presented, accompanied by a discussion of the teaching notes to assist faculty's use of the case to sensitize students to the various options they must make their future worlds better for themselves. The conclusion highlights the implications of the study and future research and applications of this approach.

LITERATURE REVIEW

Need for a business case with design oriented questions rather than ethical analysis alone

Thought exercises about the consequences of challenging, blowing the whistle, persuading, and reporting the big actors in the Big Techs and all their partners across industries, including higher ed, may help metaverse or AI ethicists understand the risks of extreme retaliations, active resistance, or mere inaction (Sætra, et al., 2023). But AI ethicists find it difficult to get sufficient support and space outside the system, even in academia (Sætra, et al., 2023). That does not mean we give up. No doubt finding support for encouraging and embedding ethical approaches may be difficult. But it is for this precise reason that as educators we continue to speak up and show our students through our lived experiences, how we can use reasoned, deliberate judgements to build ethical systems for ourselves that will ensure every individual's freedom and wellbeing.

The values of education have always been and will continue to be to equip students to take charge of their intellect and prepare students to take meaningful roles as active and informed citizens (Wines, 2008). Early studies in the late 80s revealed that business students would come to universities with the lowest level of moral reasoning skills than any other graduate student and were the only groups to experience a decline in their reasoning skills as they finished their studies (Wines, 2008). With efforts across colleges to enhance student preparedness through bridge programs, what we really need is a deliberate effort to help students go beyond ideologies and habitual orientations to engage in well-reasoned thoughts that convince the individual about their choice (Wines, 2008). Zimbardo's famous Stanford Prison experiment wherein college professors trained in reasoned judgements were so immersed in their role play enactment of board members that they didn't even realize they had turned aggressive, which resulted in monstrous outcomes, is an important example of how immersive experiences can overwhelm even the most balanced individuals (Wines, 2008).

One way to awaken students glued to their technology and raise their attention to considering the possible ramifications of harmful and untrammeled licensed tech on society is for faculty to use current case studies straight from the headlines and news articles along with small-group discussions on complex issues (Nourbakhsh, 2021). However, if these small group discussions debate whether the company or other stakeholders' situation is a virtue or sin, good or evil, or right or wrong (Russell, 1910) it will only result in airing of differing ideas. One recent study of small group discussions across multiple business ethics courses, found that students did not change their moral reasoning when exposed to the various ethical frameworks in the context of business case studies which had many ethical dilemmas (Ohreen, Sundararajan, Trifts, and Comber, 2021). The authors found that students retained their moral reasoning and ethical value systems learned from childhood and did not change their ethical stances, and only appreciated the different perspectives given by other student team members.

Further, determining if actions and therefore the insistence on recommended behaviors does not guarantee adherence to moral code by humans who design the metaverse or who use the metaverse. This is due to variations in perceived interpretations of unfairness or inequity which can never be accounted for by technologists or ethicists (Johnson, and Verdicchio, 2023). Therefore, to do good for the freedom and well-being of all, we need students to understand their responsibilities as part of the future workforce. Students can decide how they will design an ethical system that can present the options, the consequences, and offer opportunities for change, leaving the final decision to 'the humans behind the avatar in the metaverse to choose their path, because quite often, they themselves are the humans behind the various avatars.

It is essential that faculty provide scaffolding to make informed decisions because the metaverse and AI machine technology are quite alien to everyone (Nourbakhsh, 2021). If students are trained and conditioned in the safety of classrooms, it can carry over to other situations that arise in the workplace and in metaverses. Specifically for business students, in addition to teaching them the traditional areas of moral philosophy, ethical dilemmas, utilitarian ethics approach, and corporate social responsibility, faculty can offer other areas necessary to make sense of the dynamics of the business world. These include moral psychology, organizational design and behavior, motivational theories, and a session on how society, business, and law interact to illuminate the evolution of modern humanity, economics and societies (Spence, 2008).

Based on various news articles alerting the public to unethical incidences on metaverses, we present a short case study about a hypothetical company named, Metaversatality. This company intends to provide a metaverse platform for its users to build and play for commercial, educational, social, and entertainment purposes. We add teaching notes for the case in the discussion section. The crux of the exercise is to have students study the various ethical frameworks along with their pros and cons described in the literature review and then research all areas of the business to design a system they will themselves want to be a part of and one that will ensure the safety and wellbeing of all who join the metaverse. To this extent, we now discuss briefly the value of metaverses, the harms that can be caused by untrammeled use, and the views of ethicists, technologists, and researchers on the unethical practices in metaverses.

The value of metaverse to users

Metaverse users have the advantage of learning in 3-Dimensional digital spaces which are made more lifelike using virtual reality (VR) or augmented reality (AR). The gaming worlds, including e-sports, in which users have a character that can walk around and interact with other players (Knox, 2021) have led to many college scholarships for students' e-sports gaming abilities. The term "Metaverse" was first used in a 1992 cyberpunk novel by Neal Stephenson, titled, Snow Crash wherein characters were able to build things that in actual reality did not exist (Baloyan, 2022; Goyal, 2023). With gamifying being a common theme to increase engagement, business school students are often taught course content through simulation exercises. In fact, the metaverse's ability to simulate situations to help us learn and gain experiential knowledge are aplenty, for instance, several companies like Meta, Microsoft, and NVIDIA are developing applications for the metaverse that pertain to higher education, and vocational training among other industries (Torsen, 2022). Finally, another valuable tool for introverted individuals is the opportunity to speak up due to the confidence they get by being anonymous or by taking on an avatar (Yasuda, 2024).

The harm caused by metaverse to users

Sadly, the anonymity that offers confidence to some users to let their voices be heard is also one of the primary reasons for increased aggression in metaverses. The biggest concerns in the metaverse are in the form of harassment and bullying caused by aggression experienced by users due to the protection of anonymity of the abusers (Yasuda, 2024). In addition to the various

challenges such as loss of privacy, liberty, and dual-world issues, risks in metaverses can stem from deliberate malicious actions that can be forbidden or a result of the unintended consequences of those users who struggle with mental health challenges (Benjamins, et al., 2023). For example, due to lesser and lesser boundaries between real and virtual worlds, users carry their real world problems into digital environments which cause intense discomfort to others in the virtual spaces through their actions (Yasuda, 2024). Despite the rules of a game in the virtual world that may challenge one to duels and fights to death, willful vilification that result in shaming and undermining the dignity or self-respect of another, through denigration or degrading others, like vilification, and rape are unacceptable because of the psychological and emotional trauma that they will cause to the harmed person. Rape is one of the most serious moral wrongs that one can inflict on another (Spence, 2008).

Recent approaches by ethicists, technologists and researchers to counter unethical practices in ai and metaverses

Despite the advantages that disruptive innovations have brought into our lives, historically, critical elements like privacy, security, and other protection needs beyond the basic function have only been afterthoughts to its developers, as is the case for metaverses (Zheng and Daugherty, 2023). As a result, even though metaverses offer several advantages to their users, the harm caused by them are innumerable. The problem persists because of the delay in setting up a system of consequences that can deter harmful behaviors. Yet very weak voices seeking to embed ethical systems from within these large tech firms are few and far between and are often easily silenced or fired by the corporation. While strong and loud voices that do not have enough power from outside to make significant impact on the firm's choices are easily dismissed.

This conundrum forms the crux of this study's research question, "is there an ethical framework that companies, which design metaverses, can adopt that will allow them to sustain their profits and ensure the safety and wellbeing of users?" Before we explore various ethical frameworks, a review of a recent user experience survey helps us understand users' needs and expectations. Accenture's survey found the following needs among virtual users (Zheng and Daugherty, 2023):

1. 18–40-year-olds felt a) embedding mechanisms to combat misinformation, b) features that allow users to control their own safety, c) cross-industry standards for trust and safety, d) human moderators that proactively monitor user interactions and experiences, and e) built-in mechanisms to penalize users for bad behaviors, will provide the right ethical system in metaverses.
2. Users who are over 40 years prioritize built-in mechanisms to a) penalize users for bad behaviors, b) combat misinformation, c) allow users to control their own safety, d) enforce use of real names, d) vet users' identities, and e) provide clearly written and communicated community guidelines.

Overall security, privacy, trust and safety, were the common sentiments prevalent across all genders, ages, and geography, and therefore instead of playing catchup to counter cyberbullying and similar harmful behaviors, a proactive approach is necessary by metaverse builders (Zheng and Daugherty, 2023).

To address the above user needs, we undertake a review of the pros and cons of the various ethical approaches. The approaches are categorized into three broad groups. 1) human-centric approaches, b) system-centric approaches, and c) a combined human-system centric approach. We next discuss these approaches and mention the pros and cons of each.

The Human-Centric Approaches. While AI bots lack moral accountability as they do not possess sentiency and therefore do not have the capacity to experience suffering, avatars in metaverses are humans in the guise of virtual costumes who can make deliberate choices (Lin, 2023). Despite being in a virtual world, the actions are by humans, who are working with AI systems coded to make decisions as members of society (Kupiers, 2023). User experience surveys seem

to repeatedly call for a human-centric experience wherein people can feel free to express themselves authentically, in a safe and respectful manner (Zheng and Daugherty, 2023). Since the expectation comes from humans, AI ethicists can revert to understanding how ethics works for humans, a core function of which appears to be a way to balance individual self-interest with the well-being of society (Kupiers, 2023) and thus giving rise to the first category, the human-centric approaches. Here we need to state explicitly that "being able to express oneself authentically" does not give any user the license to say or do harmful things to others.

The first human-centric ethical approach is known as the Ethics, Trust, and Cooperation (ETC) framework (Kupiers, 2023). This approach describes how a society benefits functionally from the ethical beliefs of its individual members. The researcher explains that a society benefits from the positive- sum gains from cooperation between partners who are both trustworthy and can therefore trust each other to be vulnerable in that cooperative relationship. The principles and practices in this framework require a society that shows individuals how to be trustworthy and how to recognize whether others are trustworthy. The model sees trustworthiness and trust as central concepts on a causal chain that starts with ethical principles, which universally and across cultures, provide guidelines for people not to harm others physically, emotionally, or mentally. Ethical principles lead to trustworthiness which lead to trust, which in turn lead to cooperation on the one hand and social norms on the other. Cooperation is seen as the involvement of known and trusted partners collaborating in a positive-sum activity. While social norms are seen as the ability to count on others who may not be known as individuals avoiding costs for actively defending against or recovering from exploitation. Together, positive-sum activities and saving resources for defense and recovery lead to more resources for society. This approach asks companies to adopt normative frameworks such as virtue ethics, duties, contractual agreements, rights, and utility maximization (Kupiers, 2023), which requires the onus to be on the users' values and choices. Among these, the ease with which utility maximization can be programmed in a mathematically structured way seems to make it the most favored ethical system in AI (Kupiers, 2023). Utility maximization is a principle that describes how individuals choose an option that offers maximum benefit or value to themselves. This consequentialist approach is from the egoism school, where the maximization of utility is for the individual, as opposed to maximizing utility for everyone (Kupiers, 2023). **Pros:** This view can be helpful if an individual has had the opportunity to clearly discern what will provide the greatest benefit to their own and others' wellbeing in the short and long run. **Cons:** On the negative side, maximizing utility for just the individual without personal or reasoned ethics can lead to greed and an unwillingness to share, which may once again result in aggression in virtual worlds.

The second human-centric approach is from a Buddhist perspective, which shows how human beings are the ones that mainly suffer. But as indicated by Buddhist, Confucian, Aristotelian, and Bhartṛhari's teachings, humans are also the only ones who have the capacity to positively transform themselves by integrating the path of the middle way and by avoiding extremes, in their daily lives (Lin, 2023). Instead of fearing and dismissing metaverses altogether, the middle-way allows people to learn how to use it for beneficial rather than detrimental purposes (Lin, 2023). **Pros:** This is very helpful, for those who have internalized moderation and have the patience and willingness to avoid extremes. **Cons:** Sometimes desires and fears can overwhelm one's rationality to take the middle-path, thus taking one down a very dangerous and dark path before they can begin their ascension to a higher or middle-path. During that time, many others may be harmed by them.

The third human-centric approach is founded on Gewirth's arguments for principles of generic consistency[1] that build one's reasoning about why one develops virtues, and why one has duties towards others (Spence, 2008). Spence (2008) argues for the rights to freedom and well-being for

[1] The supreme principle of morality is the *Principle of Generic Consistency* (PGC), stating that every agent should act in accord with the generic rights of the recipients of her actions as well as of herself. https://iep.utm.edu/gewirth/

every being, real or virtual, as virtual beings are real beings inhabiting a virtual environment, based on the general principle of universalizability. He divides the reasoning into three stages.

Stage I-

1. I do X for purpose E.
2. (1) Entails E is good, and
3. My freedom and well-being are generically necessary conditions for my agency.
4. (2 and 3 entail): My freedom and well-being are necessary goods.

Stage II-

5. (4 entails). All other people ought, at least, to refrain from interfering with my freedom and well-being.
6. (5 entails) I have rights to my freedom and well-being.

Stage III-

7. (applying the argument from the sufficiency of agency to 6 entails) I have the right to freedom and wellbeing because I am a PPA (prospective purposive agent- a person who has the potential to be someone with a purpose).
8. Applying the principle of universalizability to (7) entails, all PPAs have rights to freedom and wellbeing.
9. (8) entails 9. I ought at least to refrain from interfering with the freedom and wellbeing of any and every PPA from which it follows that every agent is rationally committed to accepting the general moral principles of generic consistency (PGC) and
10. Act in accord with the generic rights of your recipients as well as of yourself.

This approach includes the development of Aristotelian virtues such as empathy, compassion, fairness, and ethical judgment that may be adopted by AI developers, users, and stakeholders to involve all of them to cooperate and work together for everyone's benefit (Giarmoleo, et al, 2024). Spence (2008) argues that the values of rights and role-driven morality manifested by the context and culture in organizations must match that of the universal morality principles, especially because avatars are real people who just happen to inhabit a virtual environment, and the right to freedom and wellbeing must take precedence in metaverses as well (Spence, 2008). **Pros:** It helps individuals develop character that creates a sense of conscientiousness due to awareness of the negative consequences to personal and societal wellbeing (Spence, 2008). It offers a way to reason about one's need for self-awareness and self-regulation that may help rational minds understand the value of having values that benefit oneself and others. **Cons:** If the mind is bound by biases, emotions, and clouded by desires and fears, it will not permit an individual to rationalize the importance of having values.

Systems-Centered Approaches. The system-centric approaches are founded on seeing technology as a human product that must inherently have moral intent and discernment.

The first system-centric approach is an extension of human-centric design but applicable to the company employees rather than the users and therefore falls under the system (organization) approach. It states that by design AI and any technology is related to ethics which requires a code of conduct for the engineers and technologists based on the ethical consequences it has on society (Giarmoleo, et al, 2024). The approach asks companies to adopt an act-centered approach or an agent-centered approach to ensure ethical behaviors in AI-related practices (Giarmoleo, et al, 2024). Act-centered approaches would include establishing and evaluating ethical procedures and guidelines along with impact assessment to shape behavior through established models or norms rather than giving importance to the innate ethical qualities or personal development of AI scientists or individual users. Like the human-centric approaches, agent-centered solutions will involve stakeholders and communities, improving cooperation in AI research fields, updating education programs, and behaving virtuously (Giarmoleo, et al, 2024). **Pros:** The system-centric places the

onus on the company to establish ethical procedures and monitor it as well as seeking its employees to develop virtues. If done with the right intent, it can create a heavenly place for users to grow through these metaverse applications. **Cons:** It assumes that the company will have leaders and employees who will always follow this despite pressures from outside or from within, which is not the case.

The second system-centric approach focuses on the organization's leader. It states that designating a leader to ensure the company designs and deploys the metaverse responsibly will empower employees to ensure that happens. The organization will need to establish principles and guidance for responsible innovation and use of the metaverse for the firm. It must decode challenges created by metaverses for the organization and apply a decision framework to help navigate the metaverse responsibly and strategically to meet users' ethical needs. To ensure safety and well-being for all, this approach recommends embedding responsibility into the very design of the metaverse. They suggest eight trust related dimensions like privacy, security, resilience and intellectual property rights and human dimensions like safety, sustainability, inclusion, and well-being (Zheng and Daugherty, 2023). **Pros:** From the organization's side, while the codes themselves cannot automatically make the AI system responsible, it will be able to guide the technologist towards making more responsible choices which can be translated into design features (Diakopoulos, et al., 2024) and offer those same reasoning and values to the users within the system. **Cons:** Despite metaverse spaces with codes and boundaries designed to care for all (Zheng and Daugherty, 2023), user identities can have limitless expressions and codes or boundaries cannot consider possible manifestations that may not be accounted for. It is also possible that the codes can be hacked, or too many codes and boundaries can stifle the value of the platform or creativity of the users. Additionally, the machine learning aspect of these systems can be affected or infected by actors with malicious intent or bad behavior and incorporate these as the norm.

Combined Human-Systems Approaches. These approaches combine purely user focused human-centric with the systems or organization embedded ethical frameworks. They can also be viewed as holistic approaches that require users and systems to be equally responsible.

The first combined approach adopts design principles which include simultaneously addressing Social, Economic, and Environmental needs (Yasuda, 2024). Social include - a) ensuring the safety and diversity of the metaverse, b) implementing unbiased data practices, c) improving accessibility, d) implementing automated moderation systems and reporting mechanisms, e) educating users with digital skills, and f) lowering the barriers to access the metaverse. Economic include - a) ensuring strong data and privacy protection, b) prohibiting oligopolistic control and promoting competition, c) protecting contributors, and d) obtaining users' consent for surveillance. Environmental include - a) introducing renewable energy, and sustainable hardware, and b) educating and enlightening users. All of these are sustained with the support of governance and legal frameworks. **Pros:** the holistic approach considers societal, environmental, and economic needs. **Cons:** It requires support of internal company governance and external legal frameworks. If any of the support systems collapse, it cannot be sustained.

The second combined approach entails the application of values identified in UNESCO's recommendation on the ethics of AI (Benjamins, et al., 2023) that a) respect, protect, and promote human rights, fundamental freedoms and human dignity, b) help environment and ecosystems to flourish, c) ensuring diversity and inclusiveness, and d) enable living in peaceful, just, and interconnected societies. Additionally, they seek that the specific ethical principles for the metaverse be as follows - fairness, equality, responsibility, accountability, privacy, security, safety, green, and sustainability, protection of children and other vulnerable groups, do-not-harm business models, transparency, inclusiveness, and liberty. The expectation is to have companies self-regulate the above ethical principles in all their actions. **Pros:** Once a company knows how to self-regulate, it will not need external motivators or incentives to make it adhere to ethical frameworks.

It will offer immense confidence to users that no matter what they can trust the system. **Cons:** The assumption that companies will not be tempted to compromise on the principles when they must choose profits over people is a tall order. Further, changing governmental laws and regulations as well as funding can trigger digression or suppression of preferred approaches. In a recent news article, many AI and other tech firms have urged the government to not pass strict regulations on AI systems for at least another ten years.

The third combined approach uses a utilitarian perspective (Anshari et al., 2022). It aims to ensure transparency and provide control over one's personal data by giving opt-out or opt-in options to help build trust in the businesses. To manage privacy damages, duties or rights-based strategy (deontology) is considered. It asks companies to understand they have a moral obligation to protect the privacy of customers. **Pros:** It invokes the moral obligation to safeguard customers' profile cards, which can in the long term benefit the company's reputation as a caring and transparent firm. **Cons:** It has a need to balance utilization and its restriction of user data to help with personalization but also alleviate data privacy concerns, which requires exceptionally conscientious individuals who handle user data.

The fourth combined approach asks the AI ethicist to consider whether their actions and communications will reduce chances of successfully reforming the system and offers a firm's employees two options (Sætra, et al., 2023).

Option 1: Be in the system, borrow the power of Big Tech and reform things for the better from the inside.

Option 2: Seek opportunities for change from the outside that actively avoids reliance on Big Tech. **Pros:** It provides options for those who are uncomfortable working for companies that may be engaging in unethical practices. **Cons:** The biggest risk is to the employee and that is too much pressure on one person to try and change the entire system.

DISCUSSION

Summary of ethical frameworks

The case of AI and metaverses seeks a scaffolded design methodology to instill in students a deep level of knowledge and awareness about how to conduct oneself and what to expect in the metaverse as AI and metaverses grow exponentially and take over future worlds. Variations of ethical principles that have been used to guide the complex relationships in the metaverse include business benefit evaluation, fairness, explainability, and reliability principles (Behera, et al., 2024). The nine ethical frameworks that prescribe either a human-centric, system or organization-centric, or a combination of human and system centric approaches provides a comprehensive view of all possible frameworks that can be adopted by students when designing it from the company's perspective.

The biggest point that students must consider is, whether it is fair for employees taking paychecks from companies to speak against unethical policies promoted by companies when designing metaverses or would the right thing be to independently speak from the outside of the system against unethical practices.

While working from within a system can be like a parasite, yet one could argue that for a greater good (consequentialism) it may be okay to violate virtue ethics. Unfortunately, such an approach may both protect and violate rights and duties, as the ends never justify the means. One may be a hypocrite who is complicit in ethics-washing (Sætra, et al., 2023). But as Sætra, et al., (2023) point out, "the master's tools will never dismantle the master's house, ~ Audre Lorde," because any change from within will be partly a result of the logic of the system ultimately making one beholden to the ideologies of the very system they are trying to reform. Therefore, working from within may only seemingly address minor ethical issues rather than truly changing the system to benefit

everyone's well-being (Sætra, et al., 2023). Should faculty then encourage students to go with the alternative, i.e., work for ethical changes from the outside? Unfortunately, there are downsides to working from outside. When one is insignificant and without any power, they will have no voice, or audience because those on the outside can be dismissed or discredited and can become lonely and frustrated enough to give up (Sætra, et al., 2023). The reality, however, is that it may not be possible to get true and complete independence from the system since Big Techs are ubiquitous in every industry and area of life (Sætra, et al., 2023). Yet if one had to choose and can get some support, even though working from the outside may be daunting, it may eventually have more lasting effects. It adheres to virtue and deontological ethics without contradictions and from a consequentialist angle, it avoids strengthening the system by avoiding becoming bound by incentive-based motivated actions (Sætra, et al., 2023).

With the above background, faculty can present the following case and its design questions.

The case[2]

The CEO of Metaversatality Inc, Channing Rose stated that although she started off "very pro-free speech, and free expression," she now felt that the programs on "fact checking and content moderation destroyed trust in the platform." Citing the various pressures from the huge institutions over the last decade being a test for the company, Metaversatality's chief global affairs officer, Jane Eyre, said the company is moving forward by getting rid of the number of restrictions on topics like immigration, gender identity, and gender that are subject to frequent political discourse and debate. These changes reflected in their revised hateful conduct policy now state that users can explicitly compare "gender or sexual orientation" to being mentally ill or abnormal. Further, by removing the 'restriction on referring to women as property or objects and removing restrictions on using derogatory terms, it creates exceptions that particularly target vulnerable groups, according to civil liberties groups. In response, Metaversatality's representatives said, 'it's important to note differences between offensive speech versus speech that can lead to violence, and that the company does not believe its role is to regulate what is offensive."

Reactions to channing rose's changes

A lawyer who represented the company said he would no longer do so. A charity group underlined how such changes will have negative consequences especially because "what needs to be, or what used to be, a safe space for sharing information, raising awareness, and building community is really going to move toward a downward spiral." The charity group representative felt that "codes of conduct are important ways to signal that an online community is for everyone, and that Rose's decision to remove restrictions that explicitly apply to gay, lesbian, bisexual, and transgender people could also send a message that perhaps those groups are less welcome. "Knowing that there are certain guidelines in place, and whether or not people choose to follow them, at least, creates that sense of a hopeful barrier or responsibility, or at least a moral ground that the platform that you are on wishes for you to abide by".

A review of several posts across various social media platforms indicates that users would delete their Metaversatality accounts, stop posting, or boycott the company in response to the changes. Since these changes were announced, Google searches for deleting Facescroll and Tenagram, two of Metaversatality's top social media platforms, have sharply increased, according to Google Trends.

[2] The case is based on a hypothetical company with characteristics drawn from multiple news articles about various AI and metaverse companies in the last decade.

Can we really trust Metaversatality and Channing Rose to shape the next era of humanity?

Can we really trust Metaversatality after witnessing several incidents in the past such as training advertisers to customize advertisements to users with broken heart to make them spend money on travel or the way the company's senior leaders dismiss allegations and scandals with condescension, claiming that "in a few weeks legislators will move onto something else, while the company prints money in the basement." The company is almost "central to the digital economy due to its control over critical infrastructure of human information and communications systems and access to the public sphere. Companies like Metaversatality that have provided social media platforms for colleges are now part of a group of companies that are the primary facilitators of deteriorating mental health, genocidal violence, and the rise of anti-democratic populism around the world.

Given that nearly 4 billion people regularly use Metaversatality's core products, which means the company not only "controls access to the internet in many countries, but it also determines viability and profitability for content creators, businesses, and politicians around the world." Such an influence over the public sphere has contributed to widespread democratic deterioration and the rise of digital authoritarianism. While we have seen large fines laid on the companies and some honest and sincere voices continuing to speak up and write articles despite brazen indifference, yet no one has really held Metaversatality responsible in any meaningful way. Metaversatality continues to grow having acquired over sixty companies and now creating its metaverse intended to be platforms for companies to run their online businesses. But, in the absence of federal privacy legislation that will enforce meaningful accountability, Radsch (2024) states that this may be the ascent of surveillance capitalism (Zuboff, 2018).

There does not seem to be a corner of the internet Metaversatality isn't tracking. Their trackers are embedded in millions of websites all over the internet, collecting data about where you go and what you do and sending it back to Metaversatality. An investigation showed that those trackers are on sites that even the most cynical among us might expect to be off-limits: those belonging to hospitals, including patient portals that are supposed to be protected by health privacy laws.

Even though a few years ago Meta, Microsoft, and other technology giants racing to build the emerging metaverse concept formed a group to foster development of industry standards that would make the companies' nascent digital worlds compatible with each other, many of these giants are yet to address the ethical needs of such virtual environments despite being repeatedly under scrutiny for multiple privacy law violations.

It is important for those entering metaverses to know that the whole ethos and business model of social media services is to regularly remind people of each other, ensuring maximum contact between everyone, as that is seemingly the optimum outcome for the company and its advertisers. While consumers have the option to block people who behave badly, relationship entanglements are often too complex to be dealt with by just blocking. Databases still hold users' shared history – a history that the service has encouraged people to assiduously compile but giving them almost no assistance to hide it when they no longer want to see it. Granted that the consumer must also be responsible for moderating their desires but for a naïve consumer with no way out once they get addicted, is an irresponsible act by companies. For instance, a game designed by Metaversatality has witnessed multiple virtual sexual assaults, the most violent one being the gang rape of a 16-year-old UK girl by adult men online.

Privacy violations aside, is it not time for the leaders of Metaversatality to question the ethicality of this type of advertising as well as non-monitored games with no consequences? Especially, behaviors which appear to lead to toxic rhetoric, psychological and sexual harassment, and addictive behaviors in the social media age. Would this not be a good moment for Metaversatality to step up and revamp its metaverse to support children and young adults to grow and flourish by establishing a transparent law of consequences as well as developing metaskills that will be available to all users to learn?

Question 1: If you were the CEO or part of the senior leadership team of Metaversatality, how would you design the metaverse to ensure safety and wellbeing for all, based on the nine ethical frameworks presented in class?

Question 2: If you are not the CEO, what design elements will help you convince your own company and other metaverse companies to adopt ethical designs to ensure safety and wellbeing for all?

Teaching notes

The case describes the questionable choices of a hypothetical company named Metaversatality which started out as a social networking platform for university students to connect with one another and has evolved into a global platform connecting billions of people worldwide. The reality is that people from various walks of life, have lived a fair bit of their adult lives on this platform, among other online platforms that have since emerged, with many having founded thriving businesses, established new and old connections, learned about multiple ways of being and doing things, and for the most part, it must be fair to say, have enjoyed their stay on the platform.

However, whenever one signs into or creates an account on this and other such social media outlets, while there is now a general awareness of giving up on some of their personal information, there is also an expectation that this information will be used wisely by the owners of such social media platforms. However naïve that may sound, users can at the least expect that their moments, emotions, lives, and feelings will not be exploited by the platforms. It is in this context that the case rests, where things done by these platforms, Metaversatality in particular, skirts on the fringes of moral and ethical lapses, and continuously so.

While rights groups and media watchers claim that this willful exploitation requires more serious checks, Metaversatality has chosen to be opportunistic, swaying to the vagaries of political demands of the day, or so it seems. In this context, it is for students and learners to ponder the design of AI fueled metaverse platforms to ensure overall wellbeing for all of humankind.

Teaching objectives

The target audiences for this case are senior undergraduate students or first year MBA students who have taken or are concurrently taking some business ethics courses.

Learning objectives

By analyzing and discussing the case, students should be able to:

1. Identify the ethical violations, the dangers present, and their consequences in metaverses, in this case Metaversatality being the point of confluence.
2. Select one or more ethical approaches to design an effective ethical system for metaverses that ensure wellbeing.

Recommendations for when and how to use this case

Educators can use this in a workshop style format or as a training and discussion session in their classrooms. For the workshop format, handouts of the nine ethical approaches can be given to all participants and a brief 20-minute review of the various models can be made. Additionally, participants can be provided with additional research related to how companies can balance profits and people's safety, sustainable marketing, sustainable strategic management, and latest legal guidelines for AI and metaverses. Each individual or group can give a presentation on the most effective design to be adopted by Metaversatality.

The case can also be assigned as group work with written submissions and oral presentations.

Discussion questions

The case ends with the following two questions posed.

Question 1: If you were the CEO or part of the senior leadership team of Metaversatality, how would you design the metaverse to ensure safety and wellbeing for all, based on the nine ethical frameworks presented in class?

Question 2: If you are not the CEO, what design elements will help you convince your own company and other metaverse companies to adopt ethical designs to ensure safety and wellbeing for all?

The answers for both the questions can vary as the design itself will be based on what each student believes is just and fair. Some sample answers could be as follows:

The design can be a combined human-system approach that requires both the user and the system creators to be equally responsible and prescribe signed agreements for holding each other responsible.

Another design could be adopting the law of karmic consequences, which is equally applicable to everyone, regardless of status, position, presence or lack of wealth or resources and this will actually be very easy to implement in the metaverse. For every good act done by a denizen of Metaversatality or other such platforms, there should be a karmic icon that glows some color (say blue) and for every vile act, the karmic icon can glow red. These being two extremes in the spectrum, every user, besides their status, can have this color bar, with no option to remove or hide. This will warn other users to either stay clear of such users or befriend them based on the color on their karmic color spectrum. Keeping with Metaversatality's new policy of allowing unfettered speech, Metaversatality's language engine can mine this and flag things accordingly. So, let the trolls hurl whatever abuses they can muster, all users will be able to see for themselves this wanton behavior and can unfriend, block, or avoid these users. Eventually, these users will find themselves only in self-created echo chambers or attract others who subscribe to this point of view. What then are the consequences of such behavior, students and others may ask? People who exhibit such vile behavior, hiding behind any real or perceived anonymity, will always find others to follow along, however, their impact on the larger Metaversatality community may be limited. Interestingly, unethical marketers will find ways to market their wares to even such depraved individuals, but eventually even companies that do this will realize that doing so will be going against any stated values they may have advertised and will likely abandon their efforts to market products or services to the red karmic icon accounts. The answer to this question poses both an existential aspect of social media platforms and how they operate their businesses and to users who populate these platforms.

CONCLUSION

The paper presents the need for educators to present newly defined ethical approaches to students, that are specific to the problems and requirements of AI based metaverses and ask them to design a metaverse that they will be a part of, which will ensure their own and others' wellbeing. The risks of metaverses are a problem that must be addressed urgently before it consumes a younger generation and aggravates toxic behaviors. Combining human and systems-centric ethical approaches, students can be encouraged to design systems that will reduce or remove these risks. A hypothetical case intended for business ethics students at the undergraduate or graduate levels, has been presented to facilitate discussions on how to build an ethical metaverse system. People in general are aware of ethical values, yet they do not follow them and will do so only when they begin valuing values (Dayananda, 2011). By helping students explore the pros and cons of each of the ethical approaches and designing an ethical system that would protect vulnerable populations and counter the risks in metaverses, educators can instill greater sensitivity, awareness and knowledge of ethos among students.

It provides practitioners with an opportunity to use this as a training case for their own companies in terms of thinking at a meta level and developing sensitivity to human desires, fears, and behaviors while designing ethically supportive environments as we go into new frontiers in metaverses. Trainers and educators must note that there can be many versions of each answer as the trainees/students choose from among the models presented to them. The idea is to engage their reasoning to elicit an ethical model that metaverse companies can adopt, which in turn will help consumers get out of addictive or toxic behaviors if they no longer want to behave that way. This is the key, i.e., providing metaverse users with the option to be safe, get out of the metaverse if they want to, and not feel forced, rushed, addicted, or pressured, and therefore put themselves in harm's way just to get some likes, awards, wins, recognition, or money.

It is important for those entering metaverses to know that the whole ethos and business model of social media services is to regularly remind us of each other with maximum contact between everyone being the optimum outcome. The user must also be responsible for choosing to moderate or refine their desires, but the continuous bombardment of attractive options with the promise of desire fulfillment seems like quicksand for a naïve consumer. However, Metaversatality is not solely to blame either for all the bad behaviors. It is the nature of the human mind to vacillate between being destructive, selfish, and selfless. Yet, Metaversatality can continuously provide an ethical system that reminds users of the consequences of these choices as well as opportunities to become valuable contributors in a social network rather than destructive consumers. The case offers students the opportunity to shape their own world through their actions and choices which will help them see how the right choices will lead to safety and wellbeing for all.

REFERENCES

Anshari, M., Syafrudin, M., Fitriyani, N.L. & Razzaq, A. (2022) Ethical Responsibility and Sustainability (ERS) Development in a Metaverse Business Model, Sustainability,14, 15805. https://doi.org/10.3390/su142315805

Baloyan, S. (2022). What You Need to Know About Metaverses Now, Feb 1, 2022, https://hackernoon.com/what-you-need-to-know-about-Metaverses-now https://doi.org/10.4324/9781315744841-3

Behera, R.K., Janssen, M., Rana, N.P., Bala, P.K., & Chakraborty, D. (2024) Responsible metaverse: Ethical metaverse principles for guiding decision-making and maintaining complex relationships for businesses in 3D virtual spaces, Decision Support Systems, 187(3):114337 https://doi.org/10.1016/j.dss.2024.114337

Benjamins, R., Viñuela, Y.R. & Alonso, C (2023) Social and ethical challenges of the metaverse, AI and Ethics 3:689–697 https://doi.org/10.1007/s43681-023-00278-5

Brom, M. (2023) Legal Representation and the Metaverse: The Ethics of Practicing in Multiple Realities, Depaul Business & commercial Law Journal, 21: 125-143

Dayananda. (2011). Bhagavad Gita Home Study Course, 9 Volumes Arsha Vidya Research and Publication Trust, Chennai, ISBN: 978-93-80049-39-7

Diakopoulos, N., Trattner, C., Jannach, D., Costera, I. M. & Motta, E. (2024) Leveraging Professional

Ethics for Responsible AI , Communications of the ACM, 67 (2): 19-21

Giarmoleo, F. V., Ferrero, I., Rocchi, M., & Pellegrini, M.M. (2024) What ethics can say on artificial intelligence: Insights from a systematic literature review, Business Society Review, 129:258–292. https://doi.org/10.1111/basr.12336

Goyal, R. (2023). 14 Best Metaverse Platforms You Can Already Enter in 2024, https://geekflare.com/best-Metaverse-platforms/

Hartman, L.P., Desjardins, J. & MacDonald, C. (2014) Business Ethics: Decision-Making for Personal
Integrity and Social Responsibility, 3rd Edition McGraw-Hill. ISBN 978-0-07-802945-5.
Johnson, D.G. & Verdicchio, M. (2023) Computing Ethics: Ethical AI is not about AI, Communications of the ACM, 66 (2): 32-34 https://doi.org/10.1145/3576932
Knox, P. (2021). What is the Metaverse? Meaning explained, New York Post, Oct. 21, 2021. https://nypost.com/2021/10/21/what-is-the-Metaverse-meaning-explained/
Kupiers, B. (2023) AI and Society: Ethics, Trust, and Cooperation, Communications of the ACM, 66(8): 39-42 https://doi.org/10.1145/3583134
Lin, C-T (2023) All about the human: A Buddhist take on AI ethics, Business Ethics, eth Environment and Responsibility, 32:1113–1122. https://doi.org/10.1111/beer.12547
Macfarlane, B., Desjardins, J., & Lowry, D. (2004). "The Ethics of Teaching Business Ethics- A Reflective Dialogue," Journal of Business Ethics Education, 1(1) pp: 43-54. https://doi.org/10.5840/jbee2004116
Nourbakhsh, I.R. (2021) AI Ethics: A Call to Faculty, Communications of the ACM, 64 (9): 44-45 https://doi.org/10.1145/3478516
Ohreen, D., Sundararajan, B., Trifts, V., & Comber, S. (2021) "Vygotskian Business Ethics: The Influence of Peers on Moral Reasoning in Business Ethics Education"., *Journal of Management Education.* Vol 46 (1), pp 1-36 - https://doi.org/10.1177/1052562921996019
Prasittichok P., & Klaykaew K.K. (2022). Meta-skills development needs assessment among undergraduate students. Heliyon, 8(1): e08787. https://doi.org/10.1016/j.heliyon.2022.e08787
Sætra, H.S., Coeckelbergh, M. & Danaher, J. (2023), The AI Ethicist's Dirty Hands Problem, Communications of the ACM, 66(1): 39-41 https://doi.org/10.1145/3529751
Sales, N.J. (2024). A girl was allegedly raped in the Metaverse. Is this the beginning of a dark new future? The Guardian, Jan 5, 2024, https://www.theguardian.com/commentisfree/2024/jan/05/Metaverse-sexual-assault-vr-game-online-safety-Meta
Sinha, E (2023) 'Co-creating' experiential learning in the metaverse- extending the Kolb's learning cycle and identifying potential challenges, The International Journal of Management Education 21(3):100875 https://doi.org/10.1016/j.ijme.2023.100875
Spence, E.H. (2008) Meta Ethics for the Metaverse: The Ethics of Virtual Worlds. In Proceedings of the 2008 conference on Current Issues in Computing and Philosophy. IOS Press, NLD, 3–12.
Stephen, K., Muir, L & Hall, H. (2020). Towards a definition of Metaskills, Information Research, 25(4), Proceedings of ISIC: the information behavior conference, Pretoria, South Africa, 28th September to 1st October 2020. https://doi.org/10.47989/irisic2010
Torsten, F. (2022) Meta Learning in Metaverse, IM + io, 2022, Issue 2, p82-85. 4p
Wines, W.A. (2008) Seven Pillars of Business Ethics: Toward a Comprehensive Framework, *Journal of Business Ethics*, 79:483–499 https://doi.org/10.1007/s10551-007-9411-7
Yadollahi A, and Yazdani S. (2020). Defining Cognitive, Higher Order Thinking, and Psychomotor Meta-Skills: Three-Curricular Content Concept Analysis. Journal of Medical Education. 2020;19(3): e110334. https://doi.org/10.5812/jme.110334.
Yasuda, A (2024) Metaverse ethics: exploring the social implications of the metaverse, AI and Ethics, https://doi.org/10.1007/s43681-024-00507-5
Zheng, D. and Daugherty, P (2023) Building a Responsible Metaverse, Accenture Report
Zhuk, A. (2024) Ethical implications of AI in the Metaverse, AI and Ethics, https://doi.org/10.1007/s43681-024-00450-5