IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 5 September 2022, accepted 21 September 2022, date of publication 19 October 2022, date of current version 25 October 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3215736

RESEARCH ARTICLE

The Metaverse and Beyond: Implementing Advanced Multiverse Realms With Smart Wearables

SAJJAD ROSTAMI AND MARTIN MAIER
Optical Zeitgeist Laboratory, INRS, Montréal, QC H5A 1K6, Canada
Corresponding author: Martin Maier (martin.maier@inrs.ca)

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC) Discovery under Grant 2021-03224.

ABSTRACT With the online-everything transformation accelerated by a global Covid-19 pandemic, we may finally find ourselves on the verge of the next potentially paradigm-shifting step after the mobile Internet: *The Metaverse*. Among others, the Metaverse will utilize head-mounted devices (HMDs) and extended reality (XR), including but not limited to virtual and augmented reality (VR/AR), as the medium to connect avatars and users in the real world. In addition, the Metaverse is supposed to provide *gamified experiences* around emerging Web 3.0 technologies and is anticipated to be the precursor of the so-called *Multiverse*, which will serve as an architecture of advanced XR experience realms. In this paper, we focus on the anticipated 6G post-smartphone era, where smart wearables such as VR/AR HMDs are increasingly replacing the functionalities of smartphones. Our contributions are threefold: (*i*) we first extend Metaverse's primary focus on VR/AR to Multiverse's advanced XR realms of experience. Next, we gamify and implement all eight Multiverse realms of experience using Oculus Quest 2 and Microsoft HoloLens 2 as state-of-the-art VR/AR HMDs, experimentally investigating and comparing the performance of a (*ii*) single-player origami game and (*iii*) multi-player maze game across our proposed integrated VR/AR HMD and Amazon Mechanical Turk crowd-of-Oz (CoZ) platform.

INDEX TERMS 6G post-smartphone era, crowd-of-Oz (CoZ), extended reality (XR), gamified experiences, head-mounted devices (HMDs), Metaverse, multiverse, web 3.0.

I. INTRODUCTION

Global crises such as the Covid-19 pandemic highlighted the fragility of our current approach to globalized production, especially where value chains serve basic human needs. On the flip side, however, virtual experiences such as Zoom's cloud-based video platform for online meetings and events have skyrocketed in popularity as the Covid-19 pandemic's online-everything transformation took place. With the mass digital adoption of remote work and online social activities accelerated by a global pandemic, we may finally find ourselves on the verge of something big and potentially paradigm-shifting: *The Metaverse*—the next step after the Internet, similar to how the mobile Internet expanded and enhanced the early Internet in the 1990s and 2000s [1].

Unlike earlier Metaverse approaches, e.g., Linden Lab's Second Life, the current Metaverse, most notably Facebook's recent announcement on October 28 2021, is based on the social value of Generation Z that online and offline selves are not different [2]. The younger generation considers the social meaning of the virtual world as important as that of the real world, since they think that their identity in virtual space and reality is the same. Metaverse applications are mostly aimed at *games* as well as research and marketing simulations, and some office applications. At present, games (e.g., Fortnite and Roblox) are the most common platforms in the popularization of the Metaverse, including so-called serious games. *Extended reality (XR)* is the medium that connects avatars in the Metaverse and users in the real world. According to [2], the Metaverse will utilize the following hardware and software components: Hand- and non-hand-based input devices, scene/object/sound/speech recognition

The associate editor coordinating the review of this manuscript and approving it for publication was Nurul I. Sarkar.

TABLE 1. List of abbreviations and acronyms.

Abbreviation	Meaning
2D	Two Dimensional
3D	Three Dimensional
6G	Sixth Generation of Wireless Mobile Networks
AI	Artificial Intelligence
AR	Augmented Reality
ARG	Alternate Reality Game
AV	Augmented Virtuality
CoZ	Crowd-of-Oz
FG-NET-2030	Focus Group Network 2030
GFT	Google Flu Trends
HIT	Human Intelligence Task
HMD	Head-Mounted Device
HTML	Hyper Text Markup Language
IP	Internet Protocol
ITU	International Telecommunication Union
MRTK	Mixed Reality Toolkit
MTurk	Mechanical Turk
NFT	Non-Fungible Token
OpenCV	Open Source Computer Vision Library
SLAM	Simultaneous Localization And Mapping
STEM	Science, Technology, Engineering, and Mathematics
TCP	Transmission Control Protocol
TCT	Task Completion Time
TV	Television
V3	Version 3
VR	Virtual Reality
WMR	Windows Mixed Reality
XML	Extensible Markup Language
XR	Extended Reality
YOLO	You Only Look Once

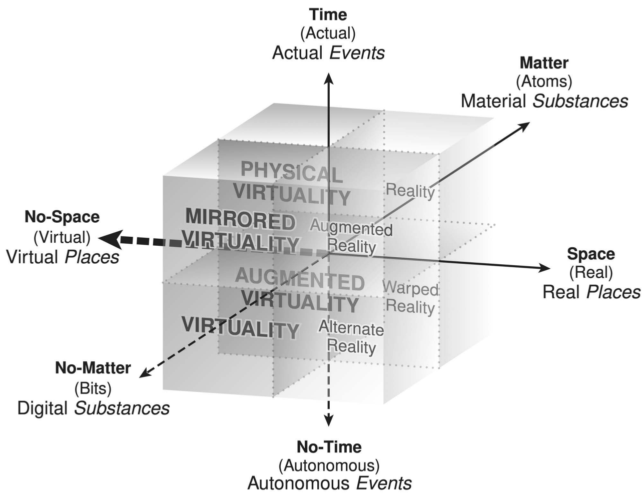


FIGURE 1. The Multiverse as an architecture of advanced XR experiences: Three dimensions, six variables, and eight realms [7].

and synthesis, motion rendering, and last but not least *head-mounted devices (HMDs)*.

The authors of [3], one of the most recent books on the Metaverse, consolidated various perspectives to give the following central definition of the Metaverse: "The Metaverse represents the top-level hierarchy of persistent virtual spaces that may also interpolate in real life, so that social, commercial, and personal experiences emerge through Web 3.0 technologies." According to [3], the Metaverse isn't meant to exist in a vacuum. Whether through virtual reality (VR), augmented reality (AR), or a smartphone, the Metaverse and the experiences inside it can connect to the real world. Importantly, in addition to incentivizing transactions, the Metaverse should provide what the authors call *gamified experiences*, given that gamification [4] will encompass the activity and story around emerging Web 3.0 non-fungible tokens (NFTs).

According to [5], the Metaverse will be the precursor of the so-called *Multiverse*. The Metaverse and Multiverse have in common that they aim at realizing the fusion of digital and real worlds, which is the driving theme for the Network 2030 vision of ITU-T Focus Group Network 2030 (FG-NET-2030) [6]. However, while the Metaverse primarily focuses on VR and AR, the Multiverse offers eight different types of advanced XR experience realms, including but not limited to VR and AR, as illustrated in Fig. 1 and explained in more detail shortly. With the rise of the Metaverse, the Internet will no longer be at arm's length. Instead, it will surround us and will be a network of interconnected experiences and devices far beyond mere VR. The Metaverse will be about being inside the Internet rather than simply looking at it from a phone or computer screen.

This paper builds on our recent work on the so-called Internet of No Things for the *6G post-smartphone era*, where smart wearables such as VR/AR HMDs are increasingly replacing the functionalities of smartphones [8], [9]. Our novel contributions are threefold: (*i*) we first extend Metaverse's primary focus on VR/AR to Multiverse's advanced XR realms of experience. Next, we gamify and implement all eight Multiverse realms of experience using Oculus Quest 2 and Microsoft HoloLens 2 as state-of-the-art VR/AR HMDs. More specifically, we experimentally investigate the (*ii*) single-player offline Multiverse implementation playing an origami game and (*iii*) multi-player online Multiverse implementation playing a networked maze game across our proposed integrated VR/AR HMD and Amazon Mechanical Turk *crowd-of-Oz (CoZ)* platform [10], which enables the crowdsourcing based remote control of SoftBank Mobile's *Pepper*, a state-of-the-art social humanoid robot, as well as its virtual avatar, and compare their performance in terms of task completion time and failure/success rate as our metrics of interest.

The remainder of the paper is structured as follows. In Section II, we first provide a comprehensive description of the current Metaverse vision(s). In Section III, we describe the eight different Multiverse realms of experience in technically greater detail. In Section IV, we describe the single-player offline implementation of the Multiverse realms of experience for playing an origami game with Quest 2 and HoloLens 2, followed by the description of the multiplayer online implementation for playing a networked maze game (beside the origami game) across our proposed integrated VR/AR HMD and Amazon Mechanical Turk (MTurk) CoZ platform in Section V. Section VI presents the comparison of our obtained experimental results. Finally, Section VII concludes the paper. For convenience, Table 1

TABLE 2. Current metaverse vision(s).

Company	Vision
Microsoft	Microsoft's vision goes beyond simply recreating physical space in digital space: It applies AI services to utilize boundaryless mixed-reality capabilities to allow people to share holographic experiences with Microsoft's key technologies by bringing their humanity with others and choosing how they want to experience or interact with the Metaverse.
Meta	Meta's vision focuses on augmenting users by moving beyond 2D screens toward immersive experiences to make the world closer, keeping people safely connected with communities and growing businesses using the next evolution of social technology.
Epic	Epic's vision is to foster an online playground by breaking down barriers between closed systems where users can join to play a multiplayer game, watch a movie, and freely try crafted the same in the real world as in this virtual playground.
Roblox	Roblox's vision fosters a rich platform for envisioning immersive co-experiences, where people can come together beyond digital boundaries with millions of 3D experiences to learn, work, play, design, and socialize.
Disney	Disney's vision for the Metaverse is next-generation storytelling by using data gleaned from theme park visits and consumers' streaming habits to deliver personalized entertainment experiences, including those from the company's Marvel and Lucasfilm studios.
Decentraland	Decentraland's vision is to provide opportunities for people to experience a new form of trading and sell/lease out land on an enterprise/peer-to-peer level at any time in a virtual world based on an Ethereum smart contract.
Siemens AG & NVIDIA	Envision an industrial Metaverse to create a photorealistic digital twin world, where teams worldwide analyze, identify, and optimize solutions/issues across the entire product lifecycle in an open ecosystem.

lists all the abbreviations and acronyms used throughout the paper.

II. CURRENT METAVERSE VISION(S)

In his long anticipated and recently published book "The Metaverse: And How It Will Revolutionize Everything," Canadian writer Matthew Ball argues that one of the most exciting aspects of the Metaverse is how poorly understood it is today [11]. Ball observes that for all the fascination with the Metaverse, the term has no consensus definition or consistent description. Most industry leaders define it in the manner that fits their own worldviews and/or the capabilities of their companies. For instance, Microsoft's CEO Satya Nadella has described the Metaverse as a platform that turns the entire world into an app canvas which could be augmented by cloud software and machine learning. No surprise, Microsoft already had a technology stack which was a natural fit for the not-quite-here Metaverse and spanned the company's operating system Windows, cloud computing offering Azure, communications platform Microsoft Teams, augmented reality headset HoloLens, gaming platform Xbox, professional network LinkedIn, and Microsoft's own Metaverses including Minecraft, Microsoft Flight Simulator, and even the space-faring first-person shooter Halo. Conversely, Mark Zuckerberg's articulation focused on immersive virtual reality as well as social experiences that connect individuals who live far apart. Notably, Facebook's Oculus division is the market leader in VR in both unit sales and investment, while its social network is the largest and most used globally. Furthermore, the Washington Post characterized Epic's vision of the Metaverse as an expansive, digitized communal space where users can mingle freely with brands and one another in ways that permit self-expression and spark joy, a kind of online playground where users could join friends to play a multiplayer game like Epic's Fortnite one moment, watch a movie via Netflix the next and bring their friends to test drive a new car that's crafted exactly the same in the real world as it would be in this virtual one. It would not be the manicured, ad-laden news feed presented by platforms like Facebook. Facebook hasn't said whether or not the Metaverse can be privately operated, but the company does say that there can be only one Metaverse - just as there is "the Internet," not "an Internet" or "the Internets." In contrast, Microsoft and Roblox talk about "Metaverses." For convenience, Table 2 list the current Metaverse vision(s) of the aforementioned companies along with others.

According to Ball, the Metaverse is still only a theory. It is an intangible idea, not a touchable product. As a result, it's difficult to falsify any specific claim, and inevitable that the Metaverse is understood within the context of a given company's own capabilities and preferences. However, he notes that the sheer number of companies which see potential value in the Metaverse speaks to the size and diversity of the opportunity for widespread disruption. Far from disproving it, confusion and uncertainty are features of disruption. While there are competing definitions and a great deal of confusion, Ball believes that it is possible to offer a clear, comprehensive, and useful definition of the term, even at this early point in the history of the Metaverse. Ball provides the following useful working definition of the Metaverse:

> A massively scaled and interoperable network of real-time rendered 3D virtual worlds that can be experienced synchronously and persistently by an effectively unlimited number of users with an individual sense of presence, and with continuity of data, such as identity, history, entitlements, objects, communications, and payments.

Simply put, according to [11], the Metaverse is envisioned as a parallel plane for human leisure, labor, and existence more broadly. So it should come as no surprise that the extent to which the Metaverse succeeds will depend, in part, on whether it has a thriving economy. Towards this end, payment rails are an important requirement to achieve a flourishing and fully realized Metaverse. Ball argues that's why so many Metaverse-focused founders, investors, and analysts see blockchains and cryptocurrencies as the first digitally native payment rail and the solution to the problems plaguing the current virtual economy. Ball argues that what matters is that blockchains are programmable payment rails. That is why many position them as the first digitally native payment rails, while contending PayPal, Venmo, WeChat, and others are little more than facsimiles of legacy ones.

According to Ball, ultimately, the Metaverse will be ushered in through *experiences*. Millions if not billions of users and dollars will be drawn to the new experiences and transformations that result. He argues that the very idea of the Metaverse means that more of our lives, labor, leisure, time, spending, wealth, happiness, and relationships will go online. Actually, they will exist online, rather than just be put online like a Facebook post or Instagram upload. Ball advocates that shifting any of the time passively consumed thru today's lean-back entertainment to *social, interactive, and more engaged entertainment* is likely a positive outcome, not a negative one. Hence, the Multiverse with its eight different types of advanced XR experience realms, including but not limited to conventional VR and AR, appears to be predestined to help usher in the Metaverse via novel experiences and the advanced social and interactive user engagement that results from them. Given its apparent importance to the emerging Metaverse, we provide a comprehensive investigation of all the Multiverse realms of experience in the remainder of this paper. To our best knowledge, this is the first time such a comprehensive study of the Multiverse realms of experience is conducted.

III. THE MULTIVERSE REALMS OF EXPERIENCE

The term Multiverse was introduced by B. Joseph Pine II and Kim C. Korn in their book titled ''Infinite Possibility: Creating Customer Value on the Digital Frontier'' as an architecture for the design of sophisticated XR experiences [7]. As shown in Fig. 1, the Multiverse consists of three pairs of variables, each with two opposite physical/digital dimensions Space/No-Space, Time/No-Time, and Matter/No-Matter, which give rise to a total of eight realms, each offering a different type of reality. It encompasses the multiple ways for *when* [Time $\leftrightarrow$ No-Time] experiences happen, *where* [Space $\leftrightarrow$ No-Space] they occur, and *what* [Matter $\leftrightarrow$ No-Matter] they act on. Each combination of the six variables yields a distinct realm, spanning the entire reality-virtuality continuum. In the following, we describe each realm in greater detail:

- **Reality:** The realm Reality consists of the variables Time/Space/Matter. An equivalent way of looking at it is Actual/Real/Atoms, see Fig. 1. We experience Reality as the realm of physical experiences through the *age-old medium of real life*. It is of course the realm with which we are most familiar.
- **Virtuality:** Virtuality lies exactly opposite Reality in the realm of No-Time/No-Space/No-Matter, consisting of Autonomous/Virtual/Bits. Virtuality is not subject to the physical laws of the real world. Virtuality together with Reality anchor the Multiverse. The name of each of the other realms relates directly to the two anchors in that the names of each realm on the right half of Fig. 1 denote their Reality-based nature, whereas the names of each realm on the left half of the Multiverse denote their Virtuality-based nature. These six other realms enhance, extend, or amend either our Reality- or Virtuality-based experiences. Therefore, they hold out greater possibility for value creation. A prime example of VR HMD is *Oculus Quest 2*.
- **Augmented Reality:** Of these, surely the most familiar is AR, characterized by the variables Time/Space/No-Matter. In the realm AR, digital technology is employed to enhance our experience of the physical world. The flagship AR HMD is *Microsoft HoloLens 2*.
- **Augmented Virtuality:** If bits can augment Reality, then logically atoms should be able to augment Virtuality. This is exactly what happens in the opposite realm of Augmented Virtuality (AV) characterized by the variables No-Time/No-Space/Matter, which is widely used for realizing digital twins. AV effectively flips a Virtuality experience from No-Matter to Matter, from Bits to Atoms. That means we're taking something material and tactile and using it to augment an otherwise virtual offering, resulting in an Autonomous/Virtual/Atoms experience. A popular example is *Nintendo's Wii*, where, for the first time, players at home could get physically, materially engaged in computer games, removing the experience from one residing primarily between the fingers and the brain to one involving the whole body.
- **Alternate Reality:** The shift from actual to autonomous events distinguishes Alternate Reality from the adjacent realm of the aforementioned AR, which both share the variables of digital substance and physical place. That means that if you can take the technology used to augment reality and then add a dimension of playing with time in some way, you can use that very same technology to alter people's view of the reality before them. Alternate Reality comprises the variables No-Time/Space/No-Matter. Its essence lies in constructing a digital experience and superimposing it onto a real place to create an alternate view of the physical reality. Alternate Reality derives its name from *alternate reality games (ARGs)* [12]. Such games have become increasingly popular in marketing circles as platforms for reaching the online gaming crowd.
- **Physical Virtuality:** Where Alternate Reality takes an otherwise virtual experience and plays it out in the real world, its opposite, Physical Virtuality, takes real-world objects (Atoms residing in Actual Time) and designs them virtually. Such a Time/No-Space/Matter experience occurs when virtually designed artifacts take

material shape. *3D printing* perhaps best captures the Actual/Virtual/Atoms nature of Physical Virtuality.
- **Warped Reality:** The last realm on the Real side of the Space dimension, Warped Reality, consists of the variables No-Time/Space/Matter. As opposed to Augmented and Alternate Reality, this realm isn't about embracing digital technology or bringing virtual places into the real world. Rather, it takes an experience firmly grounded in Reality and shifts only one variable, moving the event from Actual to Autonomous Time. This realm of Autonomous/Real/Atoms is not infused with the digital technology of No-Matter, nor does it reside in the virtual arena of No-Space. It just requires the offering to play with or manipulate time in some way that makes it clearly distinct and different from normal experience. Such reality-based time travel happens whenever experiences simulate another time in the past or future (albeit a fictional future), e.g., *living history museums* or *Star Trek conventions*.
- **Mirrored Virtuality:** Finally, Mirrored Virtuality, characterized by the variables Time/No-Space/No-Matter, is the exact opposite to Warped Reality, where Virtuality is tied to Real Time. Inside Virtuality, operating via some sort of avatar, you generally remain free to do whatever you wish to do, whereas inside Mirrored Virtuality you inexorably remain tied into what is happening in the real world, in real time, moment by moment. This realm derives its name from the term Mirror Worlds coined by Yale computer scientist David Gelernter in 1991 [13]. The use of any sort of online tracking tool, e.g., *Google Flu Trends (GFT)*, qualifies as Mirrored Virtuality since this realm offers a real-time view, a mirrored perspective, of what is going on out there, in the world.

IV. SINGLE-PLAYER OFFLINE IMPLEMENTATION: PLAYING ORIGAMI GAME WITH STATE-OF-THE-ART VR/AR HMDs

A. ORIGAMI GAME DESIGN

The authors of [14] argue that the need for engineering based careers has been steadily rising. To prepare our future work force, they developed a novel curriculum that integrates arts into STEM (Science, Technology, Engineering, and Mathematics) education, which gets students to think outside the box through interactive games. More specifically, to provide students with the addition of art, the authors applied the ancient Japanese craft of *origami*, which folds paper into geometric structures. According to [14], through the art of origami, students can be introduced to engineering concepts. Origami was shown to improve their spatial thinking and cognitive skills.

To provide an apt environment for children to learn, origami was deployed as a tangible interaction tool in a HoloLens based mixed-reality system [15]. The proposed tangible user interface system uses hand tracking with Kinect to overcome the limitations of HoloLens gestures and make the interaction more natural. The reported user study showed that mixed reality makes the origami process clearer and easier to understand, resulting in higher usability scores.

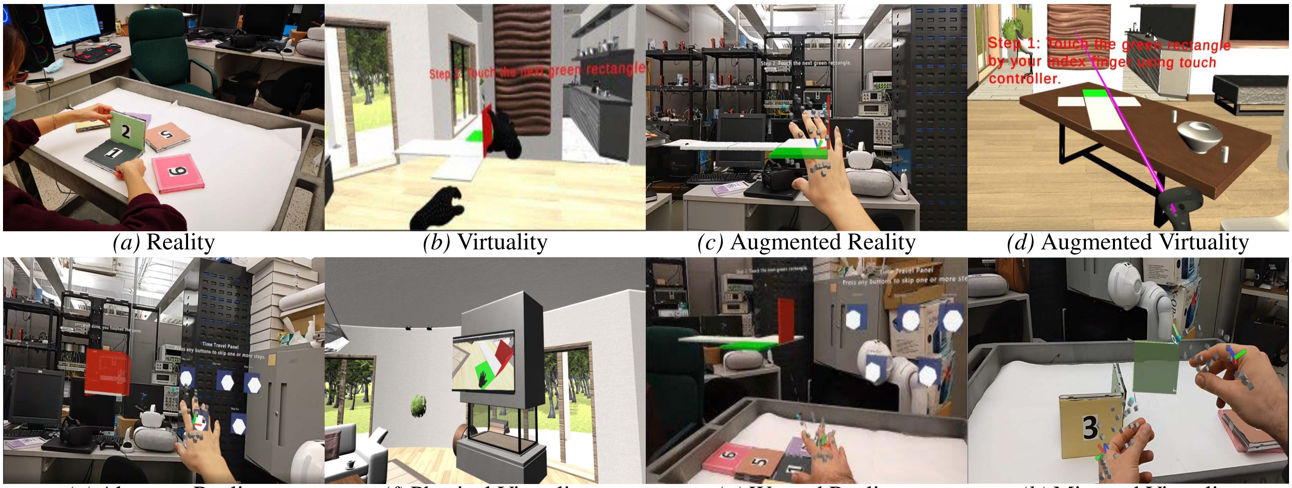
In this paper, we design a simple *cube origami game* for all eight realms of the Multiverse using Oculus Quest 2 and Microsoft HoloLens 2, two state-of-the-art VR/AR HMDs. As a cube has six sides, our origami game involves six steps to create the folded cube. Players are asked to successfully finish the cube origami game in each Multiverse realm within the shortest possible time. Depending on the specific variables of the Multiverse realm, players wear a different HMD and may use their actual hands, virtual hands, haptic controllers, and/or cognitive assistance, as explained next.

B. IMPLEMENTATION USING QUEST 2 AND HOLOLENS 2

Quest 2 and HoloLens 2 are the two flagship HMDs providing industry-leading solutions to deliver immersive single- and multi-user VR and AR experiences, respectively. The VR/AR experiences may be enhanced by the reliability, security, and scalability of Microsoft Azure and IBM Watson cloud AI services. Furthermore, both Quest 2 and HoloLens 2 provide several key design upgrades to their previous version, including an increased field of view, improved ergonomics, and more importantly, novel interaction modes that allow users to touch, grasp, and move virtual/holographic objects in ways that feel natural and intuitive.

While implementing the two Multiverse realms Virtuality and AR is straightforward with Quest 2 and HoloLens 2, the realization of the remaining Multiverse realms, apart from Reality grounded in real life, is more challenging since they require the use of virtual hands, haptic controllers, and additional support mechanisms, as we will see shortly. For illustration, Fig. 2 provides an overview of our implementation of all eight different Multiverse realms, which we describe in technically greater detail in the following.

To begin with, Fig. 2(a) shows how our cube origami game is played in the realm Reality by using six physical squares, each numbered and colored differently. Note that in this realm, the player is not given any cognitive assistance since it is rather obvious how to build a physical origami cube. Next, Fig. 2(b) shows the realm Virtuality using the aforementioned Quest 2 as VR goggles. Unlike in Reality, the player interacts with virtual squares to fold a virtual origami cube. Given that a player wearing VR goggles cannot see his own physical hands, we designed virtual hands that follow the player's physical hand movements in real time. Further, we embedded a text board in the virtual scene to give written instructions for each step. In addition, we give visual cues by coloring the next virtual square to be touched by the player in green. Once the player folded the square, its color changes to red and the subsequent virtual square is highlighted in green. By replacing Quest 2 with HoloLens 2, the realm Augmented Reality is realized, where the player interacts with holographic squares instead of virtual ones. Note that, unlike in Reality, here the player uses his own physical hands to interact with the holographic squares, as illustrated in Fig. 2(c). Similar to Virtuality, we embedded a holographic



(a) Reality *(b)* Virtuality *(c)* Augmented Reality *(d)* Augmented Virtuality

(e) Alternate Reality *(f)* Physical Virtuality *(g)* Warped Reality *(h)* Mirrored Virtuality

FIGURE 2. Single-player offline implementation: Creating an origami cube in the eight different Multiverse realms.

text board to give written instructions for each step and also used the same green/red coloring of (holographic) squares. After augmenting Reality, let us now augment Virtuality. Fig. 2(d) depicts the realm Augmented Virtuality. The player wears Quest 2, but has to tap into Matter instead of No-Matter. To do so, the player is provided with haptic feedback via a hand-held Oculus touch controller to update her about the right or wrong performed action while folding the virtual origami cube. Similar to the realm Virtuality, a virtual text board is embedded in the scene for nudging the player to perform the right actions during the game. In addition, a laser beam (see pink line in Fig. 2(d)) guides the player to select the folding a virtual object by pressing a button on the touch controller. For each successful step, the player feels a short haptic vibration in her hands. Next, to implement the realm Alternate Reality in Fig. 2(e), the player wears HoloLens 2 to access the holographic objects. Unlike in Augmented Reality, however, we need to realize the No-Time variable by embedding a holographic time travel panel to let the player jump forward to a specific step, backward, or replay it. Hence, the player is able to skip one or more intermediate steps by using the time travel panel. The player creates the physical origami cube immediately afterwards. The player can take advantage of an additional text board embedded on top of the holographic objects, which autonomously updates the player about the right decision. Fig. 2(f) depicts the implementation of the realm Physical Virtuality, where the player first wears Quest 2 to watch a short step-by-step tutorial on a virtual TV screen. At the end of the tutorial, the player is asked to take off Quest 2 and build the cube physically. To experience No-Time while creating a physical cube in the realm Warped Reality, the player wears HoloLens 2 to gain access to the aforementioned holographic time travel panel. As shown in Fig. 2(g), the player is able to virtually move back and forth in time to see the folding of the holographic cube and then creates one physically. Finally, Fig. 2(h) illustrates our implementation of the realm Mirrored Virtuality, where the player wears HoloLens 2. While building the physical origami cube, HoloLens 2 updates the status of the holographic cube according to the player's recent progress. Moreover, the player may perform additional actions on the holographic cube, e.g., rotating, relocating, and resizing.

V. MULTI-PLAYER ONLINE IMPLEMENTATION: PLAYING MAZE GAME WITH SOCIAL HUMANOID ROBOT PEPPER VIA INTEGRATED STATE-OF-THE-ART VR/AR HMDs AND AMAZON MTurk CoZ PLATFORM

In this section, we describe the multiplayer online implementation for playing a networked maze game across our proposed integrated VR/AR HMD and Amazon MTurk CoZ platform. Note that our proposed CoZ platform may be also used to play the aforementioned cube origami game, provided that a human is involved in lieu of the social humanoid robot Pepper in order to compensate for Pepper's limited dexterity, as explained in more detail next.

A. MAZE GAME DESIGN

Instead of the previous origami cube game, we select a maze game since Pepper's hands were not designed for folding objects. In our maze game, a group of online MTurk workers try to remotely steer Pepper to a hidden object (a book in our case) placed in an array of cubicles. To do so, Pepper first performs simultaneous localization and mapping (SLAM) to explore the physical environment autonomously by leveraging its built-in navigation sensors. Pepper merges all the data obtained from its navigation sensors to create a 2D map of the physical environment. Upon the successful completion of SLAM, Pepper uses the 2D map to find the approximately shortest path to the hidden object's location using the well-known *A* search algorithm*, one of the best techniques used in path-finding and graph traversals [17], [18]. For real-time object detection, we use Pepper's native NAOqi modules and

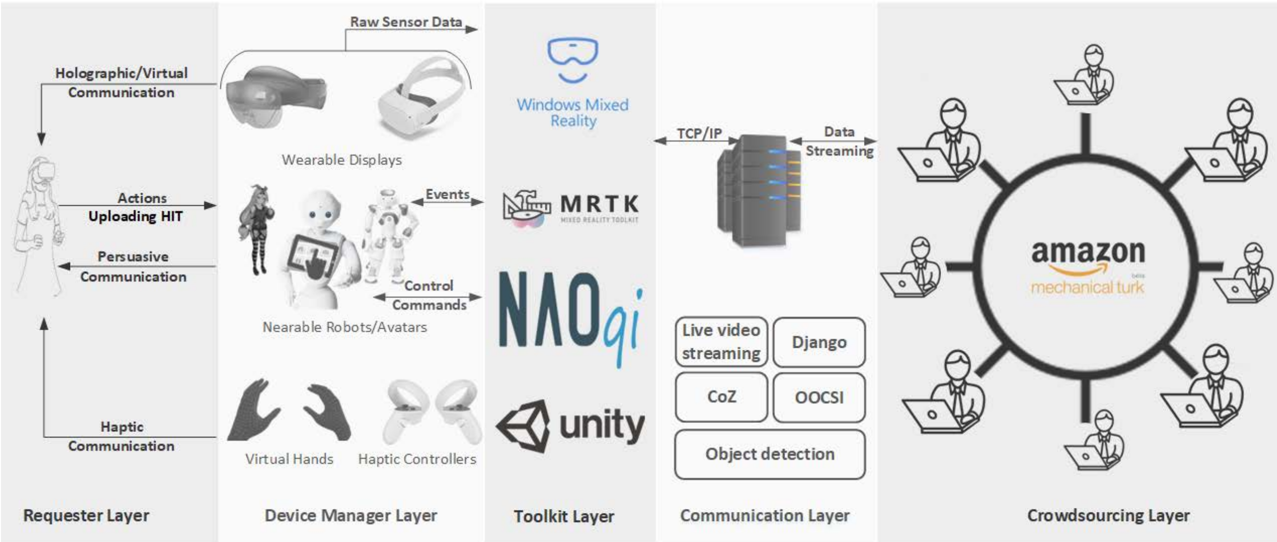


FIGURE 3. **High-level architecture of our developed CoZ platform integrating Quest 2, HoloLens 2, social humanoid robot Pepper, and crowdsourced Amazon MTurk workers.**

adopt a supervised learning approach based on a convolutional neural network to train our light dataset using ImageAI and YOLO V3. Using these techniques to let Pepper find the hidden object autonomously (i.e., without any support from local or remote human players) defines our *baseline* experiment.

Unlike in the baseline experiment, the maze game environment is dynamically changed by adding an obstacle in Pepper's path. Due to the fact that the physical environment is not the same as explored in the SLAM phase, Pepper's built-in sensors immediately abort its operation. As a result, Pepper freezes and cannot continue the assigned task. To help Pepper bypass the obstacle, workers may use our developed CoZ platform for navigating Pepper in the eight different Multiverse realms. Specifically, we realize the realm Virtuality using a local human player wearing Quest 2 to visually enable MTurk workers to remotely steer Pepper's avatar in a virtual version of the maze game environment. Note that the virtual avatar could jump over the obstacle quickly without passing around it. Further, we define a holographic border based on the output of the A* search algorithm for realizing the realm Augmented Reality. The border augments MTurk workers to steer Pepper without crossing the border. Next, we integrate haptic controllers with Quest 2 to realize the tangible aspects of Augmented Virtuality. If the avatar crosses the virtual border, the requester feels a haptic vibration in her hands and moves the avatar backward within the border. To tap into No-Time, we use HoloLens 2 and break down the maze game into four distinct steps for realizing Alternate Reality. In Physical Virtuality, the workers first watch a video on a virtual TV via Quest 2 to find out about the essential clues of the game in advance. In the realm Warped Reality, the requester wears HoloLens 2 to access a holographic time travel panel and an avatar that appears one step ahead in the path where Pepper has to move next. Finally, we realize the realm Mirrored Virtuality using Quest 2, where we synchronize Pepper with the avatar by sending a copy of every command, issued by the MTurk workers to Pepper, to the virtual avatar, thus mimicking Pepper's movements in the virtual world.

B. IMPLEMENTATION INTEGRATING QUEST 2 AND HoloLens 2 WITH SOCIAL HUMANOID ROBOT PEPPER AND AMAZON MECHANICAL TURK

Figure 3 depicts the high-level architecture of our developed CoZ platform, which integrates Quest 2, HoloLens 2, and Pepper with Amazon MTurk for crowdsourcing online workers/players. We apply a layered approach to design our platform, which provides improved scalability and simplified administration, troubleshooting, and maintenance. The developed services, interfaces, and protocols are structured into the following five layers:

- **Requester Layer:** A requester defines a human intelligence task (HIT) to recruit several online MTurk workers to accomplish a designated task. According to Amazon MTurk, a HIT represents a virtual task created by a requester and one or more workers can submit a proper answer. If the requester accepts the submitted answer, the involved worker(s) will receive a predefined reward. We applied CoZ to assist Pepper during the game to enable a real-time dialogue between the requester and worker(s). In addition, the requester uses Pepper to access the CoZ interface for defining and uploading a HIT via Pepper's in-built tablet.
- **Device Manager Layer:** We use a pointer mediator library to be able to interact with holographic objects, including 3D rendering and collision detection

in this layer. This library provides two types of interactions: (*i*) touching/grabbing holographic objects and (*ii*) detecting hand events. We also equip our virtual game objects with a so-called Collider and ImixedRealityPointerHandeler as intractable script to receive pointer laser beam events created by the haptic controllers.
- **Toolkit Layer:** We use several tools and libraries to handle all the events, motion commands, and processing of raw sensor data. Specifically, we used the Windows mixed reality (WMR) library to blend AR and MR experiences with compatible HMDs. In addition, we used Microsoft's mixed reality toolkit (MRTK) to accelerate cross-platform MR application development. Furthermore, we apply Pepper's NAOqi to execute the received commands on the robot. We use the Unity 3D game engine to fine-tune the relationship between head tracking, virtual avatar, and virtual hands.
- **Communication Layer:** The CoZ server is built using a Django web server and OOCSI Python middleware to create connections among the components and web clients. First, to join the toolkit layer with the CoZ platform, we implement a local TCP/IP socket. Next, we use Python libraries, including OpenCV, Flask, and CV2, to provide live video streams accessible over the Internet via Pepper's built-in cameras. Finally, to upload a HIT onto MTurk, we use Amazon Boto3.5 to integrate it with the CoZ server by converting a live HTML into XML format, which is a valid input format for the MTurk platform.
- **Crowdsourcing Layer:** The crowd workers use the CoZ interface to be able to have a real-time conversation with the requester through Pepper or HMDs, using a text-to-speech AI service to achieve more precise results. In addition, the CoZ has several social cue buttons to perform physical gestures and vocal cues on Pepper via embodied communication. Further, our CoZ user interface provides a section where workers can watch the physical environment through Pepper/HoloLens 2/Quest 2 cameras. Finally, we embed four directional buttons in the CoZ interface where the MTurk worker(s) can turn Pepper's head and move it in four different directions.

VI. EXPERIMENTAL RESULTS

Similar to [16], we use *task completion time (TCT)* as our primary metric of interest for the comparison of different XR environments, including but not limited to VR and AR. Figure 4 compares the experimental average TCT results of playing the cube origami game with five online (MTurk workers) and five offline (local human) players in each Multiverse realm. All the recruited players, aged 25 to 32, had no experience with HoloLens and Quest. The recruited players entered our lab freely and willingly, unaware of the study being conducted. Before starting the experiments, we trained all players in the same manner how to use the hand-held controllers, HoloLens, and Quest. For fair comparison, we did not recruit the same player to build the origami cube in two different realms in order to avoid biased players. Furthermore, we isolated the players from each other to create a distraction-free area during each experiment. Likewise, we avoided methods like standing next to the players while measuring TCT. Instead, in our experiment set-up, we used the Unity game engine that calculated the TCT score of each player automatically. This way, the players felt comfortable while playing the game and overcoming the task's challenges without feeling any external influence.

Let us first consider the Reality realm, where both online and offline players had no access to cognitive assistance. The players were supposed to build an origami cube based on their previous knowledge in this realm. The difference between achieved TCTs for offline and online players is 10 seconds. Conversely, in the Virtuality realm, the players had access to a virtual text board. Hence, they did not need to figure out how to build an origami cube based on their own knowledge. Instead, they could accomplish the task only by following the instructions on the virtual board and visual cues.

We observe from Fig. 4 that the offered virtual assistance is effective in that the players achieve a lower TCT than in the Reality realm. In the Augmented Reality realm, the players had the same visual cues and cognitive assistance, though a holographic version of them. Hence, players could use their real hands to manipulate the holographic game components. In doing so, they achieve an even lower TCT than in the Virtuality realm, where they had to use a virtual version of their hands. However, unlike the Virtuality realm, in the Physical Virtuality realm the players could take advantage of the provided haptic feedback via the hand-held Oculus touch controller to give players real-time feedback about their performed actions while folding the virtual origami cube. We observe that the achieved TCT in this realm is lower than the TCT score in the Virtuality realm. This result shows that haptic feedback is effective to enable the players to perform the right action on the virtual cube.

In the Alternate Reality realm, we observed that the players adopted an interesting approach, even though the players could use the time travel panel to build the origami cube by skipping one or more steps. More specifically, we observed that most of the offline players preferred to build the origami cube on their own rather than skipping several steps and finishing the game as soon as possible. Only one of the offline players skipped several steps to jump to step five of the game using the panel and perform the last step on her own. This is the reason for the considerable difference between the achieved TCT of offline and online players in this realm.

Next, the Physical Virtuality realm applies the same scenario as the Reality realm, except that the players in the Physical Virtuality realm could watch a video on a virtual TV via Quest 2 to find out about the essential clues of the game in advance. This is the main reason that both online and offline players could finish the game almost with the same TCT. Without watching the video, the achieved TCT is lower

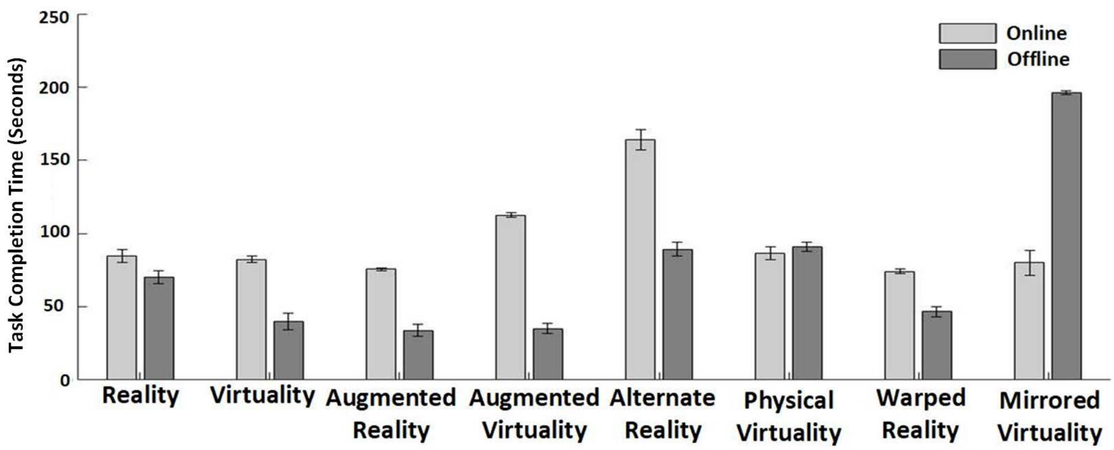


FIGURE 4. Experimental average task completion time (TCT) of cube origami game with and without using crowdsourced Amazon MTurk workers in different realms of the Multiverse (each shown with minimum-to-maximum TCT intervals).

TABLE 3. Experimental results on TCT and number of issued commands for playing maze game with online MTurk workers.

Multiverse Name	TCT (seconds)	Issued Commands
Alternate Reality	9	3
Augmented Virtuality	23	9
Virtuality	27	13
Physical Virtuality	68	26
Mirrored Virtuality	128	15
Warped Reality	134	17
Augmented Reality	171	21
Reality	237	31

than in the Reality realm, because players benefitted from the essential cues in advance to build the origami cube.

In the Warped Reality realm, the players had access to the time travel panel while building the origami cube. In this realm, the online players achieved the lowest TCT among all online scenarios. Conversely, the offline players achieved a lower TCT compared to the Reality realm. Moreover, although the offline players experienced the highest TCT in the Mirrored Virtuality realm, we witnessed that it was the most exciting realm for the offline players. This is due to the fact that they could see a synchronized holographic cube in real-time that allow them to resize, relocate, or turn the holographic components. In contrast, the online players preferred to accomplish the game as soon as possible to gain the offered reward.

Next, let us consider the maze game. Table 3 shows that the first three realms achieve a lower TCT than in the baseline experiment, which achieved a TCT of 34 seconds requiring 0 commands. This is due to the following reasons. First, unlike with the Pepper, physical limitations don't limit its avatar from passing a dynamic obstacle by simply jumping over it. Second, the virtual/holographic border helps workers issue fewer duplicated yet more effective commands. And third, the time travel panel provides players with the opportunity to finish the maze game with only three commands

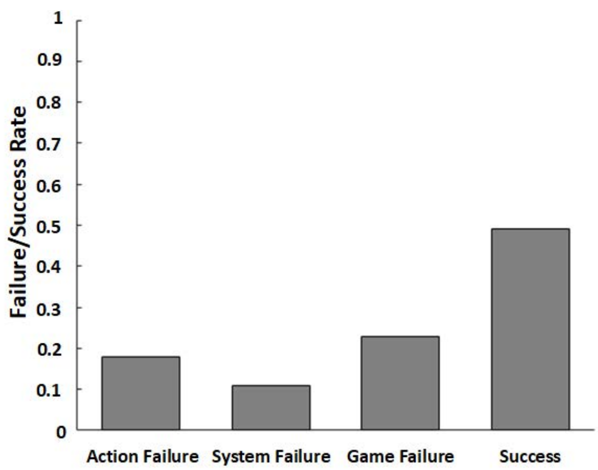


FIGURE 5. Experimental failure and success rates of maze game.

in the realm Alternate Reality. Conversely, the realm Reality achieves the highest TCT since Pepper can not bypass the dynamic obstacle autonomously by only relying on the previously acquired SLAM information. Further, in this realm, workers can not access any cognitive assistance such as our virtual/holographic border.

In total, we have submitted 128 HITs to Amazon MTurk, resulting in 61 recruited online workers to conduct our experiments. At least two MTurk workers were involved in each Multiverse realm during the maze game. As shown in Fig. 5, we had an action failure rate of 0.18 (or 18 percent), where the players did not make the right decision during the maze game. Furthermore, in 11 percent of the maze game instances, we encountered system failures, which in most cases were due to the lack of Internet connectivity and HoloLens 2 interruption. Game failure rate accounted for 23 percent of our encountered failures, where the inability to find a worker to continue the game or lack of practical knowledge of

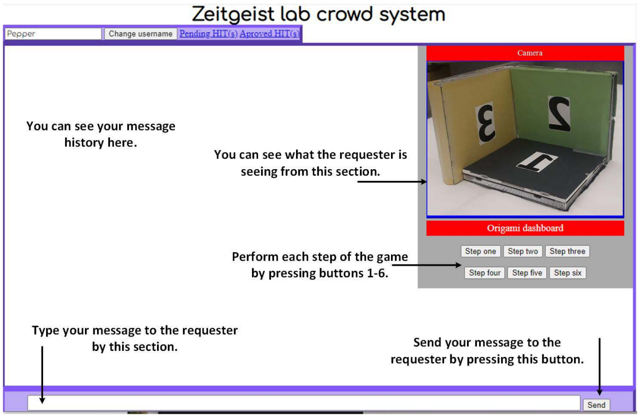


FIGURE 6. CoZ user interface providing workers with real-time video streaming, origami steps, definition of new HIT, view of submitted HITs, and message box.

workers were typical game failures. Finally, we observed that the remaining 48 percent of the submitted HITs were successfully completed in the sense that the workers performed all steps without facing any of the aforementioned failures.

To reduce the number of submitted HITs, we were inspired by several previous studies done by researchers, as discussed in more detail in the following. The design of tasks, workflows, motivation, and incentives, as well as the selection of suitable workers and their training are possible solutions that task owners can adopt to make their submission of HITs more effective [19], [20], [21]. Moreover, gamification is another technique used in crowdsourcing systems that may be adopted by task owners to increase the workers' intrinsic motivation to accept a task and participate actively [22], [23], [24]. By choosing gamification-based titles helped us get our HITs accepted sooner. Furthermore, we created a short user manual that explains the game rules along with a CoZ interface tutorial and an illustrative image of the CoZ platform to train the workers before starting the game, as shown in Fig. 6. Our CoZ interface design focused on training workers by offering them a gratifying task rather than simply filtering out unprofessional workers by submitting several back-to-back HITs.

Note that generally the lack of direct communication between workers and requesters is one of the main issues in crowd systems to keep the workers motivated [25], [26]. We tackled this shortcoming by establishing a peer communication link through our designed CoZ interface that makes it more manageable and feasible for workers to accomplish tasks. The CoZ platform provided this facility in that workers and requesters can communicate directly via text during the game. Hence, the hired workers were not supposed to accomplish a task independently. Instead, the requester could support the workers by providing immediate feedback during the decision-making process. The aforementioned approaches allowed us to recruit suitable workers and keep them interested till the end of the game. Keeping the workers motivated prevents them from aborting a given task unpredictably, resulting in improved HIT success rates.

VII. CONCLUSION AND FUTURE WORK

At present, games such as Fortnite and Roblox are the most common platforms in the popularization of the Metaverse. The emerging Metaverse should provide gamified experiences, including serious games, and is anticipated to be the precursor of the Multiverse's advanced XR realms of experiences going beyond conventional VR and AR. In this paper, we presented our integrated VR/AR HMD and Amazon MTurk CoZ platform, which uses Quest 2 and HoloLens 2 to enable the crowdsourcing based remote control of social humanoid robot Pepper and its virtual avatar. We've extended the Metaverse's primary focus on VR/AR to Multiverse's eight distinct XR realms of experience and experimentally compared their performance in terms of task completion time (TCT) and failure/success rate for a single-player origami game and a multiplayer maze game, played in both offline and online (i.e., networked) mode. Our obtained experimental results aim at providing insights into understanding the pros and cons of different types of reality and selecting the most suitable one to realize gamified experiences with the lowest TCT at acceptable failure/success rates. Taking the specific TCT requirements and failure tolerance of a given gamified experience into account, game designers may select their preferred single type of reality that satisfies the performance requirements. Alternatively, game designers may combine two or more types of reality in order to create cross-reality environments, in which players can cross or transverse multiple Multiverse realms within a given gamified experience, resulting in integrated physical-virtual environments whose respective weaknesses are mitigated and respective strengths are multiplied.

Leveraging on our obtained results, future research efforts may adapt collective intelligence mechanisms for nudging players to accomplish their assigned tasks faster and more successfully. Towards this end, the full potential of blockchains should be explored by exploiting the crowd intelligence in a cooperative game world. Blockchains allow players to earn tokens and thus help realize play-to-earn games, where players can earn or exchange cryptocurrencies or non-fungible tokens (NFTs). Using emerging Web 3.0 blockchain technologies, remote expert players can cooperate with local novice players across the Internet in different Multiverse realms to jointly perform complicated tasks faster and more accurately.

REFERENCES

[1] T. Winters, *The Metaverse: Prepare Now for the Next Big Thing!* Chicago, IL, USA: Independently Published, Aug. 2021.

[2] S.-M. Park and Y.-G. Kim, "A metaverse: Taxonomy, components, applications, and open challenges," *IEEE Access*, vol. 10, pp. 4209–4251, 2022.

[3] C. Hackl, D. Lueth, T. D. Bartolo, and J. Arkontaky, *Navigating the Metaverse: A Guide to Limitless Possibilities in a Web 3.0 World*. Hoboken, NJ, USA: Wiley, May 2022.

[4] D. Basten, "Gamification," *IEEE Softw.*, vol. 34, no. 5, pp. 76–81, Sep./Oct. 2017.

[5] R. Higgins, *Metaverse: A Definitive Beginners Guide to Metaverse Technology and How You Can Invest in Related Cryptocurrencies, NFTs, top Metaverse Tokens, Games, and Digital Real Estate*. Chicago, IL, USA: Independently Published, Nov. 2021.

[6] *Network 2030—A Blueprint of Technology, Applications and Market Drivers Towards the Year 2030 and Beyond*, Standard ITU-T FG-NET-2030, May 2019, pp. 1–9.
[7] B. J. Pine, II and K. C. Korn, *Infinite Possibility: Creating Customer Value on the Digital Frontier*. Oakland, CA, USA: Berrett-Koehler, Aug. 2011.
[8] M. Maier, A. Ebrahimzadeh, S. Rostami, and A. Beniiche, "The internet of no things: Making the internet disappear and 'see the invisible,'" *IEEE Commun. Mag.*, vol. 58, no. 11, pp. 76–82, Nov. 2020.
[9] M. Maier, A. Ebrahimzadeh, A. Beniiche, and S. Rostami, "The art of 6G (TAO 6G): How to wire society 5.0 [invited]," *J. Opt. Commun. Netw.*, vol. 14, no. 2, p. A101, Feb. 2022.
[10] T. Abbas, V.-J. Khan, and P. Markopoulos, "CoZ: A crowd-powered system for social robotics," *SoftwareX*, vol. 11, pp. 1–7, Jan./Jun. 2020, Art. no. 100421.
[11] M. Ball, *The Metaverse: And How It Will Revolutionize Everything*. New York, NY, USA: Liveright, Jul. 2022.
[12] M. Lupano, L. Farinetti, and D. Morreale, "OPUS: An alternate reality game to learn SQL at university," in *Proc. IEEE 45th Annu. Comput., Softw., Appl. Conf. (COMPSAC)*, Jul. 2021, pp. 121–126.
[13] D. Gelernter, *Mirror Worlds or: The Day Software Puts the Universe in a Shoebox. How It Will Happen and What It Will Mean*. New York, NY, USA: Oxford Univ. Press, Nov. 1991.
[14] J. Kennedy, E. Lee, and A. Fontecchio, "STEAM approach by integrating the arts and STEM through origami in K-12," in *Proc. IEEE Frontiers Educ. Conf. (FIE)*, Oct. 2016, pp. 1–5.
[15] Y. Song, N. Zhou, Q. Sun, W. Gai, J. Liu, Y. Bian, S. Liu, L. Cui, and C. Yang, "Mixed reality storytelling environments based on tangible user interface: Take origami as an example," in *Proc. IEEE Conf. Virtual Reality 3D User Interfaces (VR)*, Mar. 2019, pp. 1167–1168.
[16] N. T. Banerjee, A. J. Baughman, S.-Y. Lin, Z. A. Witte, D. M. Klaus, and A. P. Anderson, "Side-by-side comparison of human perception and performance using augmented, hybrid, and virtual reality," *IEEE Trans. Vis. Comput. Graphics*, early access, Aug. 18, 2021, doi: 10.1109/TVCG.2021.3105606.
[17] S. Russell and P. Norvig, *Artificial Intelligence: A Modern Approach*, 2nd ed. London, U.K.: Pearson, Dec. 2002.
[18] Y. Song and P. Ma, "Research on mobile robot path planning based on improved A-star algorithm," in *Proc. Int. Conf. Electron. Inf. Eng. Comput. Sci. (EIECS)*, Sep. 2021, pp. 683–687.
[19] C. Yin, Y. Li, and X. Liao, "Task decomposition and restructuring method for collaborative design of crowdsourcing platform," in *Proc. IEEE 4th Int. Conf. Nanosci. Technol. (ICNST)*, Jun. 2021, pp. 40–45.
[20] M. Shergadwala, H. Forbes, D. Schaefer, and J. H. Panchal, "Challenges and research directions in crowdsourcing for engineering design: An interview study with industry professionals," *IEEE Trans. Eng. Manag.*, vol. 69, no. 4, pp. 1592–1604, Aug. 2022.
[21] Q. Hu, S. Wang, P. Ma, X. Cheng, W. Lv, and R. Bie, "Quality control in crowdsourcing using sequential zero-determinant strategies," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 5, pp. 998–1009, May 2020.
[22] T. Ranas, Y. G. Sucahyo, and A. Gandhi, "Evaluating the gamification in tripadvisor: Is it effective for crowdsourcing platform?" in *Proc. 6th Int. Conf. Sci. Inf. Technol. (ICSITech)*, Oct. 2020, pp. 74–79.
[23] Y. Lu, X. Mao, M. Zhou, Y. Zhang, Z. Li, T. Wang, G. Yin, and H. Wang, "Motivation under gamification: An empirical study of developers' motivations and contributions in stack overflow," *IEEE Trans. Softw. Eng.*, early access, Nov. 23, 2021, doi: 10.1109/TSE.2021.3130088.
[24] B. Haworth, M. Usman, D. Schaumann, N. Chakraborty, G. Berseth, P. Faloutsos, and M. Kapadia, "Gamification of crowd-driven environment design," *IEEE Comput. Graph. Appl.*, vol. 41, no. 4, pp. 107–117, Jul. 2021.
[25] S. S. Bhatti, X. Gao, and G. Chen, "General framework, opportunities and challenges for crowdsourcing techniques: A comprehensive survey," *J. Syst. Softw.*, vol. 167, Sep. 2020, Art. no. 110611.
[26] W. Tang, M. Yin, and C.-J. Ho, "Leveraging peer communication to enhance crowdsourcing," in *Proc. World Wide Web Conf.*, May 2019, pp. 1794–1805.

SAJJAD ROSTAMI received the B.Sc. degree in information technology engineering from Sheikh Bahaei University, Iran, in 2013, and the M.Sc. degree in computer network engineering from Qazvin Azad University, Iran, in 2018. He is currently pursuing the Ph.D. degree in telecommunications with INRS. His current research interests include human–robot interaction, XR, and the Internet of Things.

MARTIN MAIER received the M.Sc. and Ph.D. degrees (Hons.) from the Technical University of Berlin, Germany. He is currently a Full Professor with INRS, Montréal, Canada. He was a co-recipient of the 2009 IEEE Communications Society Best Tutorial Paper Award. Furthermore, he was a Marie Curie IIF Fellow of the European Commission, from March 2014 to February 2015. In March 2017, he received the Friedrich Wilhelm Bessel Research Award from the Alexander von Humboldt Foundation in recognition of his accomplishments in research on FiWi enhanced networks. In May 2017, he was named one of the three most promising scientists in the category "Contribution to a Better Society" of the Marie Skłodowska-Curie Actions (MSCA) 2017 Prize Award of the European Commission. In Winter 2019 and Winter 2020, he held a UC3M-Banco de Santander Excellence Chair of the Universidad Carlos III de Madrid (UC3M). He is the coauthor of the book *Toward 6G: A New Era of Convergence* (Wiley-IEEE Press, January 2021).