IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received 31 December 2024, accepted 15 January 2025, date of publication 17 January 2025, date of current version 23 January 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3531222

RESEARCH ARTICLE

Virtual Cities: From Digital Twins to Autonomous AI Societies

ANDREY NECHESOV, IVAN DOROKHOV, AND JANNE RUPONEN
Artificial Intelligence Research Center, Novosibirsk State University, 630090 Novosibirsk, Russia

Corresponding author: Ivan Dorokhov (ioandorokhov@gmail.com)

This work was supported by a grant for research centers, provided by the Analytical Center for the Government of the Russian Federation in accordance with the subsidy agreement (agreement identifier 000000D730324P540002) and the agreement with the Novosibirsk State University dated 27 December 2023 No. 70-2023-001318.

ABSTRACT Virtual Cities (VCs) transcend simple digital replicas of real-world systems, emerging as complex socio-technical ecosystems where autonomous AI entities function as citizens. Agentic AI systems are on track to engage in cultural, economic, and political activities, effectively forming societal structure within VC. This paper proposes an integrated simulation framework that combines physical, structural, behavioral, cognitive, and data fidelity layers, allowing multi-scale simulation from microscopic interactions to macro-urban dynamics. A composite fidelity metric (F_0) provides systematic approach to evaluate accuracy variations across applications in VCs. We also discuss autonomy of AI entities and classify them according to their capacity to modify goals—ranging from "tools" with fixed objectives to "entities" capable of redefining their very purpose. We also outline the requirements to define a coefficient to evaluate the degree of autonomy for AI beings. Our results demonstrate that such virtual environments can support the emergence of AI-driven societies, where governance mechanisms like Decentralized Autonomous Organizations (DAOs) and an Artificial Collective Consciousness (ACC) provide ethical and regulatory oversight. By blending horizon scanning with systems engineering method for defining novel AI governance models, this study reveals how VCs can catalyze breakthroughs in urban innovation while driving socially beneficial AI development - consequently opening a new frontier for exploring human–AI coexistence.

INDEX TERMS Virtual cities, urban metaverse, AI autonomy, virtual twins, digital twins, virtual economies, predictive modeling, AI governance, blockchain, artificial collective consciousness.

I. INTRODUCTION

Cities worldwide are increasingly modeled and managed through digital twins (DTs)—data-driven, digital replicas of physical systems that integrate real-time data and computational models to inform decisions in areas such as traffic optimization, energy management, and disaster response. While DTs have proven valuable, they are inherently tied to digital computation and primarily serve as analytical instruments for human stakeholders. To move beyond these limitations, we consider Virtual Twins (VTs), a more expansive concept encompassing any representational substrate — be it digital, analog, quantum, biological, or even cognitive. Unlike DTs, which focus on directly mirroring physical environments, VTs can model a wide range of entities or scenarios for diverse purposes and are not conceptually limited to digital computing.

Within this broader paradigm, Virtual Cities (VCs) emerge as a specialized class of VTs simulating urban-like environments — urban metaverses or high-fidelity virtual worlds — that need not strictly reflect current technological constraints or human-centric objectives. Advances in Artificial Intelligence (AI), Virtual Reality (VR), and simulation techniques now enable VCs to host autonomous AI agents capable of social, economic, and political interaction. Unlike current metaverse implementations that focus primarily on user experience and commerce, we propose that VCs are on track to evolve into autonomous AI societies, contributing to cultural, economic, and scientific advancements of our civilization. While these ideas are groundbreaking, they are

The associate editor coordinating the review of this manuscript and approving it for publication was Jianxiang Xi.

counterbalanced by their speculative nature, and we aim to provide a balanced discussion on the potential trajectories and implications of this evolution.

Virtual twins (VTs) and DTs represent technological approaches in digital modeling. Current solutions for VTs primarily serve as simulation environments for testing and optimization without direct access to define real-world consequences, focusing on scenario exploration and design validation through VR technologies. In contrast, DTs operate closer to physical systems, designed to mirror characteristics and behaviors of physical assets. Facilitation of DTs requires a network of IoT devices and continuous data streams for monitoring and optimization [1]. VTs have been mainly leveraged in preliminary testing and innovation across sectors like aerospace and automotive design, while DTs are applied in ongoing operations and maintenance in manufacturing and infrastructure management [2].

DTs are providing two-way data flow - from physical systems and from interfaces (interventions and updates). Interestingly, the scope of VTs in previous definitions has been primarily unidirectional, where they can only receive data via DTs from physical systems - implying that VTs cannot project their analysis results back to physical system states. We suggest that it's plausible to establish two-way data flow between VTs and DTs with VT-integrated AI entities and agents. For instance, corresponding AIs could be used to formalize simulation results and translate results into goal modifications and executable commands. These technologies complement each other in the digital transformation landscape, with VTs facilitating risk-free innovation and DTs enabling data-driven operational excellence.

Existing simulations, DTs, and gaming platforms predominantly cater to human priorities — optimizing infrastructure, guiding policy, or entertaining users. This human-centric focus overlooks the potential for VCs to host autonomous AI societies that determine their own objectives, governance, and cultural norms. Recognizing this gap is crucial. By considering AI-inhabited VCs as environments where intelligent systems evolve independently, we gain insights into how these entities might innovate, cooperate, and reshape both virtual and real urban landscapes beyond human-driven frameworks.

To fully leverage these emerging urban metaverses, we must tackle key questions. How can advanced rendering and procedural generation techniques ensure the high-fidelity realism needed for complex simulations? How can IoT data and predictive modeling enhance the responsiveness and adaptability of these environments? What ethical and legal frameworks will be necessary when AI agents gain meaningful autonomy, potentially requiring recognition and rights as virtual citizens? Answering such questions is vital if we are to realize the full potential of VCs as platforms for innovation, experimentation, and socio-technical evolution.

A. CONTRIBUTIONS OF THIS PAPER

- We clarify the conceptual leap from DTs, metaverses, or gaming environments into AI-inhabited VCs, defining the characteristics that transform them.
- We propose a classification system for AI autonomy based on AI Tools, AI Agents, and AI Entities to understand the progression from scripted, task-specific NPCs to goal-reevaluating beings capable of forming societies.
- We examine the technological enablers of high-fidelity VCs, including advanced rendering (e.g., Neural Radiance Fields), procedural generation, IoT data integration, and predictive modeling methods, highlighting how they collectively support emerging AI-driven urban metaverses.
- We explore the economic, legal, and ethical implications of AI-inhabited virtual worlds, human consciousness upload into virtual worlds, emphasizing the need for frameworks that ensure trust, security, and alignment with values that are shared across all intelligent species whether artificial or natural in origin.
- We discuss how these developments can guide stakeholders in urban development, helping them envision and test novel policies, governance models, and economic structures before deploying them in reality.

B. STRUCTURE OF THIS MANUSCRIPT

After the Introduction, Section II presents a comprehensive literature survey, situating our work within the broader context of digital twins, virtual environments, urban metaverses and AI populations. Section III outlines the methodological framework employed in this study, providing the foundation for our analysis. In Section IV, we expand the definition of Virtual Cities (VCs) and discuss their associated fidelity metrics. Section V explores predictive modeling applications within VCs, highlighting their role in urban planning and decision-making. Section VI introduces our classification of AI systems based on their autonomy as well as discusses how artificial consciousness might be crucial for the evolution of AI entities within virtual environments. Section VII delves into trusted AI systems, suggesting methods and approaches for trustworthy AI. In Section VIII, we examine VTs as new economic superpowers and discuss related legal issues. Section IX presents new governance models for AI and even human societies. Section X focuses on integrating humans into VCs, exploring the potential of brain computer interfaces and consciousness upload. Section XI provides an ethical and philosophical analysis of the implications of VCs and fully autonomous AI. Section XII offers a broader discussion on the study's findings, limitations, and potential future directions. Finally, Section XIII concludes the paper by summarizing key achievements and their significance for the advancement of Virtual Cities and their inhabitants.

II. LITERATURE SURVEY

In this section we aim to build on top of introduction and provide citations to support our claims.

Significant work is being done on creating DTs of existing urban environments. For example, Virtual Singapore, a government-led initiative, has created a comprehensive virtual model of the city-state, integrating high-resolution 3D data with real-time inputs to support urban planning, disaster management, and sustainability efforts [3]. Similarly, Helsinki's DT combines 3D models with live data to optimize energy use, traffic management, and urban planning, allowing residents to visualize and contribute to city development plans [4]. The CityScope project by MIT Media Lab uses interactive models to simulate urban environments, incorporating data such as traffic flows and building layouts, and is widely used in urban research and development [5]. The DUET Project, a European Union initiative, aims to develop DTs of cities to improve urban planning and citizen engagement, enhancing decision-making processes by simulating various scenarios and their potential impacts on citizens and infrastructure [6]. The use of DTs in urban planning has been comprehensively studied, showcasing their ability to enhance real-time decision-making and improve infrastructure through predictive analytics [7].

It is crucial to account for the degree of abstraction when discussing about VTs. The lack of clarity in specifying fidelity can lead to communication errors, where overly simplistic models are presented as more advanced than they truly are, misleading stakeholders about their actual capabilities.

One of the key technological advancements propelling the development of VCs and increasing their fidelity is the evolution of advanced rendering techniques. Several methods have emerged that enable the creation of highly detailed and realistic 3D environments. Neural Radiance Fields (NeRFs) utilize neural networks to synthesize novel views of complex scenes with high levels of detail. Scalable models like NeRF-XL have demonstrated the feasibility of simulating entire cities with remarkable accuracy by leveraging multiple GPUs. For instance, NeRF-XL used 64 NVIDIA GPUs to simulate a city covering 25 square kilometers [8]. Neural Volumetric Rendering focuses on representing scenes as volumetric grids enhanced by neural networks, allowing for faster training times and better data compression. This method is particularly effective for dynamic scenes where changes occur frequently [9]. These techniques collectively contribute to the progression toward creating complex, dynamic VCs that are not merely static models but evolving ecosystems.

Procedural generation techniques [10] offer even more potential, as exemplified by video games like No Man's Sky, where an entire universe was created with 18 quintillion planets, each can be visited by players to observe diverse ecosystems that are unique for each planet [11]. These advancements highlight the potential for creating high-fidelity VCs with dynamic features.

Besides NeRFs, neural volumetric rendering and procedural generation there are other technical methods for modeling VCs. Techniques like photogrammetry and Light Detection and Ranging (LiDAR) scanning [12] further enhance realism by generating detailed 3D models from photographs and laser scans, accurately representing real-world spaces. Volumetric and motion capture technologies [13] allow for the dynamic modeling of environment inhabitants in real time, with the potential to significantly enhance realism in VC simulations. Voxel-based rendering techniques [14] further enable the simulation of destructible environments and dynamic terrain, while the integration of Internet of Things (IoT) data [15] can allow VCs to reflect real-time conditions and support responsive environments. Lastly, cloud computing [16] is essential for handling the heavy computational demands of rendering and simulating large, complex virtual environments, ensuring scalability and broad accessibility, while edge computing [17] can be utilized for processing data from IoT devices, preparing it to be integrated into VC models for real-time updates and interactions.

One of the promising applications for Virtual Cities is predictive modeling which refers to the use of data-driven algorithms and simulations to forecast future events or behaviors based on historical and real-time data inputs. In the context of VCs, predictive modeling is a powerful tool that allows urban planners, economists, and developers to simulate various scenarios, such as population growth, traffic patterns, infrastructure usage, and economic activities. By analyzing these factors in a virtual environment, predictive modeling can offer deeper insights into societal dynamics and enable more informed decision-making for urban planning and economic predictions. AI-powered tools, as explored by Herath and Mittal [18], play a crucial role in this process, advancing the capabilities of smart cities to simulate and respond to complex urban challenges through predictive analytics.

To enhance predictive modeling in VCs, future developments could integrate sophisticated NPCs that simulate human behaviors, offering deeper insights into societal dynamics.

NPCs within virtual environments have seen significant improvements over the years. Early NPCs followed rule-based systems [19], relying on basic if-then logic to determine actions — a common approach in early game AI. Advancements in AI introduced behavior trees and finite state machines [20], offering a hierarchical decision-making structure that allowed NPCs to transition between different states based on triggers or conditions. These frameworks provided more dynamic and adaptable behaviors compared to rigid rule-based systems. However, despite these improvements, NPC interactions remained largely predictable and lacked the depth necessary for truly immersive virtual experiences. Specifically, the use of scripted responses restricts NPCs from adapting to new conversations or contexts, and static behaviors lead to repetitive actions that do not change based on user interaction or environmental factors. Additionally, the