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

lack of emotional depth in NPCs reduces the realism of social interactions within virtual environments.

The incorporation of Generative Pre-trained Transformers (GPT) has significantly enhanced NPCs' linguistic and cognitive capabilities [21]. By leveraging natural language processing (NLP), NPCs can engage in contextual conversations, exhibit emotional intelligence, and learn from interactions. By using GPT models, game developers can create NPCs that provide personalized experiences, making virtual environments feel more alive and authentic. These advancements enable NPCs to adapt their responses based on user input, creating more natural and engaging interactions within virtual environments.

Powered by advances in AI, NPCs are evolving from scripted tools with limited interactions to autonomous agents capable of learning and adapting. A notable example is Google DeepMind Scalable Instructable Multiworld Agent (SIMA) that can follow natural-language instructions to carry out tasks in a variety of video game settings [22]. AI tools like ChatGPT are now being used to create NPCs that offer dynamic, context-aware interactions [21]. The convergence of AI and virtual environments is opening new avenues for innovation. Altera's Project Sid, for example, builds an entire civilization within Minecraft, using 1000 AI agents to simulate complex societal behaviors [23]. Additionally, the study ''Generative Agents: Interactive Simulacra of Human Behavior'' highlights the potential of AI-driven agents to simulate human-like behaviors and emergent social behavior [24]. Furthermore, the Agent Hospital project demonstrates how AI-driven agents can continuously evolve and improve in tasks like diagnosing diseases, eventually approaching human-level performance [25]. MetaUrban project is procedurally generating complex urban environments, enhancing embodied AI research by supporting tasks like social navigation and reinforcement learning interactions between AI agents, humans, and robotic entities, advancing research in real-world human-robot dynamics [26]. These advancements not only enhance the realism and complexity of virtual environments but also create platforms for studying AI agents in intricate social settings, contributing to the development of more adaptive and intelligent AI systems.

Isaac Sim 4.0 project [27], NVIDIA's latest platform for high-performance robotics simulation. By leveraging NVIDIA's PhysX engine and RTX technology, Isaac Sim enables the creation of VTs, allowing robots to be trained in highly realistic virtual environments. The platform also includes Isaac Lab, which supports reinforcement learning, imitation learning, and motion planning, allowing scalable training through multi-GPU setups. This makes Isaac Sim ideal for rapid development and accurate testing without extensive physical trials.

Several robots rely on Isaac Sim for advanced training. Project GR00T, part of Isaac Lab, provides a foundational model for humanoid robots, optimized for reinforcement learning, imitation learning, and transfer learning within virtual environments [28]. Quadrupeds like ANYbotics' ANYmal, designed for industrial inspection, benefit from Isaac Sim's simulation capabilities in navigating complex environments such as industrial sites [29]. Autonomous Mobile Robots (AMRs), including those from iRobot and idealworks, also use Isaac Sim for tasks like warehouse navigation and logistics. Similarly, industrial manipulators from KUKA, Fanuc, and Universal Robots undergo precision task simulations in Isaac Sim's realistic environments. Recently, a new and improved engine, Genesis, entered the competition, outperforming Isaac Sim by 1–2 orders of magnitude in benchmarks [30]. This progression highlights rapid improvements in the field.

The core issue, which is not being addressed by training environments like Isaac Sim and only partially addressed in Altera's project Sid is that for robots to fully integrate into human society, it is not enough to train physical movements alone. The development of social intelligence, including the ability to understand cultural and ethical norms, is crucial. Recent research highlights the need for AI to learn social behaviors and human interaction [31]. VCs are offering a perfect training ground not only for physical intelligence but also for complex social interactions.

There is potential for a continuous optimization loop, where enhancements in AI agents lead to more realistic VCs and vice versa. High fidelity VCs establish more advanced training environments, allowing the development of more capable AI agents. Another feedback loop forms between real-world data used to train AI in virtual environments and the application of that training in the real world, as demonstrated by projects like Isaac Sim. When these two recursive improvement loops converge, we may be witnessing the initial steps toward a technological singularity—a point where AI systems advance beyond human control, driving exponential growth in both virtual and real-world capabilities.

Virtual environments are becoming increasingly realistic, and non-playable characters (NPCs) are evolving to resemble humans more closely with each passing year. One of the key drivers that pushes for more realistic NPCs is the increase in their autonomy. If we extrapolate these trends, it is reasonable to predict that in the future virtual worlds will closely mimic reality, and AI entities will become nearly indistinguishable from humans in terms of cognitive abilities and autonomy. However, managing the growing autonomy of AI entities presents significant ethical and legal challenges. Addressing these challenges will require the development of clear and reliable definitions, as the term ''AI'' itself is often too broad and imprecise for such complex discussions. In section VI of this article we propose an AI classification system to reflect levels of AI autonomy.

We assert that VCs have the potential to evolve into autonomous societies that significantly contribute to cultural, economic, and scientific advancements of our shared civilization. Our analysis indicates that the fusion of advanced AI technologies, realistic virtual environments, and human

integration methods, such as avatars and brain-computer interfaces (BCIs), could lead to VCs functioning with their own economies, governance, and social structures. While this evolution presents unprecedented opportunities, it also raises profound ethical, philosophical, and legal questions about the autonomy of AI and the nature of our reality.

A. RESEARCH GAPS AND NOVELTY

This study addresses three potential gaps in VC research:

1) Current VC implementations focus primarily on human-centric optimization and urban planning [32], [33], lacking frameworks for autonomous AI evolution. While systems like UrbanWorld [34] enable AI behavior training, they don't support the emergence of independent societal structures.
2) Existing studies treat AI as tools or agents with limited autonomy rather than fully autonomous entities capable of reassessing their top goals. Despite advances in DTs and metaverse services [33], research hasn't explored AI-driven cultural and economic development within VCs.
3) The relationship between AI autonomy and environmental complexity remains unexplored. While Sánchez-Vaquerizo [35] examines participatory governance, current research lacks mechanisms for co-evolution between AI entities and virtual ecosystems.

By acknowledging AI entities as potential citizens, policymakers, and economic contributors within virtual or real societies, this work offers a novel perspective on how advanced computing, virtual reality, and AI can be integrated for serving civilization and communities. This framework proposes a system model architecture for interactions between artificial and biological intelligence, emphasizing mutual autonomy and ethical consideration.

III. METHODS

This paper employs an interdisciplinary approach, integrating insights from urban studies, artificial intelligence, robotics, ethics, sociology, economics, political, and legal science. The study employed horizon scanning methodology to analyze emerging technologies relevant to VCs. We conducted a systematic review of technical developments in NeRFs, simulation platforms, procedural generation, and AI models, using established technology assessment frameworks. Our analysis incorporated documented use cases and integration of existing frameworks to evaluate development trajectories. A systematic literature review was conducted using IEEE Xplore, Scopus, and Google Scholar databases. Keywords included "smart cities," "artificial intelligence," "procedural generation," "brain-computer interfaces," and "digital twins." Articles were selected based on relevance to AI-inhabited virtual cities, recentness, and citation impact. Additional sources included technical documentation, white papers, and recent research presentations to capture emerging developments in the field.

Systems engineering method was also applied assess requirements for high-fidelity framework integration and systems architecture design. We conducted a systematic evaluation of specialized open-source frameworks across five fidelity domains (physics, structural, behavioral, cognitive, and data), analyzing their integration requirements and compatibility. The methodology involved designing an integration layer architecture for cross-framework communication and unified data representation, with emphasis on API-based coordination. While we proposed a theoretical composite fidelity metric F_0), the primary contribution was the identification and architectural integration of appropriate frameworks rather than their empirical validation.

To complement our analysis, we employed Visionary Backcasting, a variant of the backcasting method described by Robinson [36] and Quist and Vergragt. Unlike conventional backcasting, visionary backcasting integrates principles from technological entrepreneurship, emphasizing iterative testing and real-world validation. This methodology is used to define and validate a future vision of AI-inhabited virtual cities through stakeholder analysis and impact assessment in the context of collectively sustainable goals. It asks *why* the integration of AI agents into virtual cities should be pursued, and what communal benefits and underlying purposes it serves. Once the purpose is established, the method is reversed to discover *how* to identify and implement technological and social progression with distinctive milestones. The method incorporates feedback loops for continuous evaluation of top-level vision driven by asking *why*, from where it provides goal-setting configuration for discovering how instrumental goals regarding technical feasibility, resource requirements, and societal implications can be optimized. This adaptive approach enables responsive strategy development for complex socio-technical systems, differentiating it from fixed end-state planning of the conventional backcasting.

It is important to note that the core concept of VCs crystallized for us during heated debates about AI safety with various government officials. It became evident that we need a safe virtual environment for testing and training increasingly advanced and autonomous AI systems.

IV. EXPANDING DEFINITION OF VIRTUAL CITIES

The primary distinction between DTs and a VCs lies in their intended purposes and capabilities. Unlike urban metaverses or video games, which are primarily designed for experimentation or entertainment, VCs are envisioned as comprehensive virtual environments with the potential to host autonomous AI societies in the future [23]. Furthermore, VCs integrates VTs, implying that their conceptualization is not limited by the traditional digital computing paradigm and encompasses the entire solution space of integrated hardware and software. A VC comprises two main components: the hardware that operates it and the software that constitutes the virtual simulation itself. Thus, the term "virtual city" may refer, depending on the context, not only to the software

simulation but to the underlying hardware infrastructure that supports the virtual environment. While the broader concept encompasses virtual worlds, our research focuses specifically on cities due to their critical role in AI training for modern urban settings. VC simulations aim to replicate the physical, social, and economic dimensions of real cities or to conceptualize entirely new urban landscapes. VCs have a wide array of applications, including immersive gaming environments, educational platforms, urban planning tools, policy analysis systems, and the potential emergence of autonomous AI societies. By providing users and residents with interactive experiences that reflect the complexities of urban life—such as infrastructure, transportation systems, social interactions, and economic activities—VCs serve as dynamic platforms for understanding and shaping the future of urban environments.

Characteristics and Components of Virtual Cities:

- Spatial Representation: Accurate three-dimensional models of urban landscapes, including buildings, roads, public spaces, and natural features.
- Dynamic Environments: Real-time changes reflecting weather, time of day, seasonal variations, and other environmental factors.
- Interactive Inhabitants: Inclusion of NPCs or AI entities that simulate human behavior, facilitating social interactions and community dynamics.
- Economic Systems: Virtual economies that allow trade, commerce, and financial transactions, mirroring real-world economic activities.
- Governance Structures: Simulated laws, regulations, administrative processes, and governance models that affect how the virtual city operates.
- User Interaction: Interfaces and tools that enable users to navigate, interact with, and modify the environment, fostering engagement and participation.

A. COMPONENTS IN SIMULATION FIDELITY

We suggest a resource-effective approach to implement simulation fidelity by creating an integration framework to connect existing open-source frameworks from specialized fidelity domains. In addition to integration of different software services, communication between software and hardware in DTs is required for achieving high simulation fidelity. Initiation of environment and subsequent simulations are resource-intensive processes in city-scale. Higher fidelity in DTs is based on increased investments in design (innovations) and physical manufacturing (materials) of IoT chips and devices, making it economically expensive to scale especially in applications requiring sophisticated solutions. As VTs are operating one-way in terms of data flow, the simulation fidelity is constrained by data received from DTs. Our approach does not aim to solve this constraint - instead we propose minimal architecture for AI-based optimization of VT simulation fidelity by using established open-source frameworks as fidelity providers. Components in simulation fidelity are divided into five fidelities:

1) Physics Engine Fidelity
2) Structural Fidelity
3) Behavioral Fidelity
4) Cognitive Fidelity
5) Data fidelity

1) **Physics Engine Fidelity:** The accuracy with which the virtual world's engine simulates fundamental physical laws and interactions at various scales. High physics engine fidelity ensures that the environment behaves consistently with real-world physics, allowing for realistic interactions and emergent phenomena.

Here we seek to exploit current progress of AI integrations by suggesting Genesis, a generative physics engine, as our fidelity provider. It includes a universal physics engine, generative data engine along with lightweight programming syntax and a photo-realistic rendering capabilities [30].

2) **Structural Fidelity:** The correctness of the organizational structure of matter within the virtual city, from atomic and molecular levels up to buildings and infrastructure. This includes detailed modeling of materials, architectural elements, and urban layouts that reflect real-world properties and behaviors. The implementation utilizes a multi-scale approach combining OpenMM for molecular dynamics, FreeCAD for building-level modeling, and UrbanSim for city-scale simulations. The integration of these open-source applications provides a comprehensive multi-scale modeling framework for digital twin applications.

In molecular scale, OpenMM can serve as the foundation layer, offering high-performance simulation capabilities with extensive language support (Python, C, C++, and Fortran), enabling accurate material property modeling at the microscopic level. OpenMM could be used, for instance, to predict and monitor air quality or disease spreads.

Structural modeling in urban scale harnesses established instruments such as Computer-Aided Design (CAD), Finite Element Analysis (FEA), Building Information Modeling (BIM), and Computer-Aided Manufacturing (CAM) capabilities. One option for an open-source parametric 3D modeler is FreeCAD, which has these instruments. It should be however emphasized that in this context the focus should be in BIM as it aggregates structural information from all designs. For collaborative design, choice of CAD client is not relevant if VT structural data objects are stored and exchanged by using standard file formats while following technical drawing guidelines.

At the urban scale, UrbanSim completes the modeling hierarchy by providing sophisticated statistical modeling tools for simulating city-wide dynamics, including real estate development, demographics, and policy impacts, leveraging its Urban Data Science Toolkit (UDST) for comprehensive urban system analysis. Together, these frameworks enable a hierarchical simulation approach that spans from molecular

interactions to city-scale phenomena, creating a robust foundation for high-fidelity digital twin implementations.

3) **Behavioral Fidelity:** The realistic simulation of inhabitants' behaviors and social interactions, reflecting cultural dynamics. It evaluates the realism of social dynamics and interactions. We suggest using MASON framework for behavior modeling, which enables urban simulations involving individual agent behaviors and emergent collective patterns [37]. This allows us to capture complex behavioral patterns like daily commuting routines, social interactions, and responses to environmental changes.

The behavioral models are calibrated by using different extensions. The Discrete Event Simulation (DES) extension allows agent-based simulations to run simultaneously in one model. GeoMason allows to connect Geographic Information System (GIS) data to simulations. Cloud computing is enabled by the Distributed MASON extension. There are also options to include social network statistics, stochastic optimization toolkits and 2D physics with additional extensions. Further, MASON provides direct method to build and run MASON simulation within Eclipse Ditto. The behavioral models can be validated using multiple scales of analysis - from individual agent decision-making processes to emergent collective behaviors at the neighborhood and city levels. By combining MASON's various extensions, these behaviorally complex simulations can be scaled to represent large urban populations while maintaining computational tractability [37].

4) **Cognitive Fidelity:** The advancement of AI entities' cognitive abilities within the virtual city, enabling reasoning, learning, and the potential emergence of consciousness. It assesses AI entities' reasoning and learning capabilities. Here we suggest to use OpenCog, AGI-based framework, and Robot Operating System (ROS) for robotic development.

OpenCog is based on a generalized graph store, AtomSpace, that is programmed with Atomese language [38]. The AtomSpace is a graph database and query engine that handles knowledge representation and reasoning. Its native language Atomese serves as an intermediate representation format for graph structures, designed to be manipulated by algorithms and processing pipelines. Automated agents and reasoning subsystems can work with Atomese similar to how programs work with assembly code - as a lower-level representation that enables direct manipulation of knowledge structures. The system supports complex query operations and semantic processing across the graph database [38].

ROS2 represents a significant advancement in robotics middleware, particularly in its ability to support high cognitive fidelity applications through its enhanced architecture and capabilities [39]. The framework's incorporation of Data Distribution Service (DDS) middleware enables deterministic data delivery and real-time processing, which is crucial for maintaining high cognitive fidelity in robot perception and decision-making. The enhanced peer-to-peer architecture facilitates distributed cognitive processing across multiple nodes, with improved message passing infrastructure enabling complex cognitive architectures to be implemented across distributed systems [39]. The framework's capabilities in sensor integration and perception, state management and model updates make it particularly effective for implementing sophisticated cognitive architectures for manipulator-based applications in VCs. Better handling of complex sensor data streams through improved message transport, native support for various data types used in cognitive robotics, and enhanced synchronization capabilities for multi-modal sensor fusion contribute to its effectiveness [39].

OpenCog is integrated to complement ROS2, enabling cognitive robotics applications. ROS2's enhanced real-time processing capabilities and distributed architecture provide an ideal platform for deploying OpenCog's knowledge representation and reasoning systems in physical robots, while OpenCog's sophisticated graph processing and learning capabilities can enhance ROS2's cognitive fidelity. This combination enables the development of more sophisticated robotic systems that can leverage both symbolic reasoning and real-time sensorimotor processing, bridging the gap between high-level cognitive processes and low-level robot control.

5) **Data Fidelity:** The precision and reliability of data inputs used to create and maintain the virtual city, ensuring that simulations are grounded in accurate and up-to-date information. It quantifies the reliability and accuracy of the input data.

We included Apache Kafka to our architecture to handle distributed streaming of real-time data feed while meeting sufficient performance in terms of throughput and fault tolerance [40]. One of Kafka's advantages comes from its wide support and distribution, making it a feasible option maintaining data streams to other fidelity platforms.

For relational data management in DT context, one suitable option is Eclipse Ditto that can be used to provide essential features for real-time monitoring and adjustments. While its focus is on DT and two-way data flows, it is domain agnostic, thus providing flexible option to extend VT fidelity with formalized service relations [41]. Ditto provides instruments enhancing data fidelity through its state management and real-time synchronization capabilities. The platform's thing model can be extended to represent physical properties and dynamics, where each physical attribute is maintained as a feature state within the DT [41]. Through Ditto's event-driven architecture, changes in physical properties can be propagated in real-time across the system. The platform's bi-directional state synchronization ensures that any changes in the physical world, detected through IoT sensors, are frequently fetched in the virtual physics simulation, maintaining high fidelity between the physical and digital representations.

Architecture of Ditto includes services as components, external API endpoints, external dependencies and service relations. Services are core components of Ditto, each driving following tasks [41]:

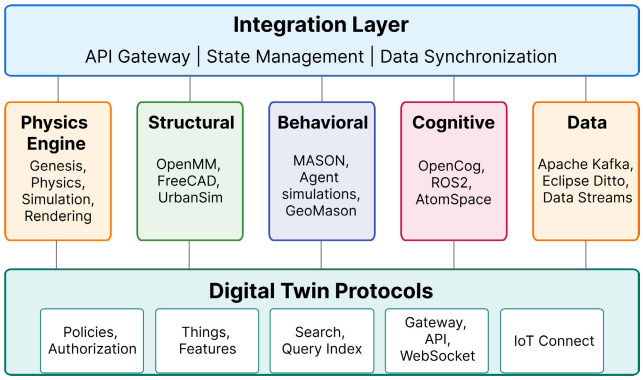


FIGURE 1. Fidelity-based integration scheme for VCs.

- **Policies:** enforcement (authorization) and persistence of Policies
- **Things:** enforcement (authorization) of Things and Features
- **Things-Search:** optimization of search index, query execution on correspondent search index, and tracking of occurring changes to Policies, Things and Features
- **Gateway:** API via HTTP and WebSocket
- **Connectivity:** exchange of messages between Ditto protocol and external message brokers, and persistence of Connections

Services are running in a single Pekko Cluster, using Transmission Control Protocol (TCP) without other message brokers. The architecture facilitates also microservices [41], consisting following components:

- Only the host microservice can perform read/write operations on its data store
- API is formed based on signal components, which include commands, command responses, events, announcement and error responses
- Defined signals are required to access by other services

Simulation Fidelity requires conceptualization of a novel approach that integrates physics, structural, behavioral, cognitive and data fidelity components into a unified system. While frameworks like Genesis, Eclipse Ditto, OpenMM, MASON, OpenCog/ROS2, and Apache Kafka handle their dedicated fidelity aspects well, their emergent simulation fidelity is required when the objective is to establish ACC-based fidelity management. Integration layer is needed for this in order to handle cross-framework communication, state synchronization, and unified data representation. The key challenge lies in maintaining consistency across different temporal and spatial scales while coordinating updates between frameworks. One potential implementation is to borrow contextual models from providers and use it as a base integration point with other frameworks connecting through API endpoints. For example, Ditto's thing model demonstrates how a single object can be mapped and related. Fig. 1 introduces conceptual architecture of integration layer that coordinates with proposed frameworks through standardized interfaces and data models.

For the entire virtual city system the simulation fidelity can be expressed based on summation of normalized fidelity module coefficients as:

$$F_0 = w_p F_p + w_s F_s + w_b F_b + w_c F_c + w_d F_d \tag{1}$$

where:

$$F_p : \text{Physics fidelity } [0,1]$$
$$F_s : \text{Structural fidelity } [0,1]$$
$$F_b : \text{Behavioral fidelity } [0,1]$$
$$F_c : \text{Cognitive fidelity } [0,1]$$
$$F_d : \text{Cognitive fidelity } [0,1]$$
$$w_i : \text{Component weights } \sum w_i = 1$$

where the weights are adjusted based on the relative importance of each fidelity component for that particular scenario. For example:

- Pathfinding scenarios: higher weights on w_p and w_b for physics and behavioral fidelity
- Marketing analysis: emphasis on w_c and w_b for cognitive and behavioral aspects
- Disaster response: priority on w_p and w_s for physics and structural components
- Pandemic simulation: increased w_b and w_d for behavioral and data fidelity

Before VCs evolve into autonomous AI societies they can serve as microcosms of real-world or imagined urban environments, offering controlled settings where variables can be manipulated to observe outcomes. This makes them invaluable for research, education, urban planning, and entertainment. They also provide platforms for testing smart city initiatives, exploring innovative solutions to urban challenges, and engaging citizens in participatory planning processes.

The integrated simulation fidelity system (F_0) enables these applications by maintaining consistent and measurable quality across key simulation aspects. Physics fidelity (F_p) ensures realistic environmental interactions, while structural fidelity (F_s) maintains accurate representations of urban infrastructure and tangible assets. Behavioral fidelity (F_b) through MASON framework simulates realistic citizen interactions and movement patterns, and cognitive fidelity (F_c), implemented via OpenCog and ROS2, enables AI entities to learn and adapt to changing urban conditions. Data fidelity (F_d) ensures that simulations are grounded in accurate, real-time data through Kafka streams. By quantifying and balancing these components through weighted evaluation (w_i), the system can be tuned to prioritize specific aspects based on use case requirements - whether for urban planning, citizen engagement, or testing autonomous systems.

B. COMPUTATIONAL COMPLEXITY ANALYSIS

Virtual city simulations face computational challenges across their core fidelity components. Key computational constraints arise from:

- Physics simulation overhead (Eclipse Ditto)
- Behavioral agent processing (MASON)
- Cognitive modeling computations (OpenCog/ROS2)
- Real-time data stream processing (Apache Kafka)
- State synchronization across components

While specific complexity bounds would depend on implementation details and chosen algorithms, the system must balance computational resources across these components while maintaining required fidelity levels for each scenario.

C. COMPLEXITY ANALYSIS APPROACH

To analyze the computational complexity of our virtual city system, we need to examine each fidelity component's (F_{system}) resource demands:

F_i	Memory	Network	Compute	Scale
F_p	States	Syncing	Physics	Objects
F_b	Terms	Messages	Processing	Population
F_c	Info	Messages	Graphs	Nodes
F_d	Storage	Bandwidth	Streaming	Load

The Physics Engine (Eclipse Ditto) plays a crucial role in state management and synchronization. Its performance metrics focus on memory usage required for maintaining state information, network bandwidth consumption during state synchronization processes, computational load generated from handling physical interactions, and how the system's scaling behavior changes with increasing numbers of objects.

The Behavioral Simulation component (MASON) manages agent-based interactions and behaviors. Key performance considerations include the processing overhead required for agent operations, memory requirements allocated per individual agent, communication costs incurred between interacting agents, and how the system scales with increasing population size across the simulation environment.

Cognitive Processing capabilities (OpenCog/ROS2) handle complex reasoning and knowledge representation. Performance metrics track the computational demands of graph operations within AtomSpace, memory usage patterns for knowledge representation structures, overhead from message passing between cognitive components, and the efficiency of node discovery alongside associated communication costs.

Data Management services (Apache Kafka) ensure efficient information flow throughout the system. Critical performance indicators include stream processing throughput capacity, storage requirements for data persistence, network bandwidth utilization during data transfer operations, and system latency characteristics under varying load conditions.

The benchmarking strategy requires both individual component testing and integrated system evaluation. This approach examines resource utilization patterns across all components, identifies scaling limitations at both component and system levels, pinpoints potential performance bottlenecks in the architecture, and establishes clear system capacity boundaries for operational planning. We present below snippet from our core integration program using Python programming language.

D. REAL-TIME FEEDBACK LOOPS WITH DYNAMIC DATA INTEGRATION

Virtual environments are inherently dynamic, integrating real-world data inputs such as traffic patterns, weather conditions, and population statistics to adjust NPC behaviors—both pedestrian and vehicular—in real time. This integration creates a feedback loop wherein insights from virtual city simulations inform real-world decisions, and actual events continuously refine the virtual models. This iterative process ensures that virtual models remain accurate representations of their real-world counterparts, increasing the reliability and applicability of predictive modeling outcomes. Furthermore, real-time data integration facilitates the continuous learning and adaptation of AI entities within VCs, fostering the development of more responsive and intelligent urban management systems. This symbiotic relationship between virtual simulations and real-world data enhances the capacity of VCs to serve as dynamic testbeds for urban innovation.

An additional crucial application of predictive modeling is in training robots not only for physical tasks but also for social intelligence. Using platforms like Isaac Sim 4.0 and Genesis, VCs can serve as highly realistic training grounds for autonomous robots. In these environments, VTs of robots can be tested to observe how they interact with AI-driven inhabitants, simulating human social behaviors. Importantly, this training process operates within a recursive improvement loop. Real-world data is continuously fed back into the virtual environment, improving the training models. The updated robot behaviors, once tested in the VC, are deployed in the real world, where the robots gather new data based on their interactions. This new data, reflecting their updated behaviors, is then reintroduced into the virtual training ground, allowing for continuous refinement. Over time, this process may lead to the emergence of new and unpredictable behaviors that must be examined in virtual simulations for security measures to predict how they will affect real-world society.

Simulations will allow us to test what happens when a social system is not working under nominal conditions. For example, while there are studies on crowd crush dynamics [42], the effects of mixed crowds composed of both humans and androids remain uncertain and largely unexplored. By applying volumetric and motion capture technologies [13] to video data from past crowd crush events, we can model NPCs to simulate human behavior in such situations. Subsequently, running tests with android VTs alongside these NPCs would allow us to explore and predict how androids might behave in similar high-density crowd conditions. Based on these simulations, we can then train androids to assist in preventing crowd crush events by learning how to manage and alleviate dangerous crowd dynamics.

1) IMPLEMENTATION OF REAL-TIME IoT DATA PIPELINE

The real-time IoT data pipeline implementation follows a microservices architecture pattern, where each layer

```
from dataclasses import dataclass
from typing import Dict, List, Any
import asyncio
@dataclass
class SimulationConfig:
    physics_params: Dict[str, float]
    structural_params: Dict[str, Any]
    behavioral_params: Dict[str, Any]
    cognitive_params: Dict[str, Any]
    data_streams: List[str]
class VirtualCitySystem:
    def __init__(self, config: SimulationConfig):
        self.config = config
        self.components = {}

    async def initialize(self):
        self.components = {
            'physics': GenesisEngine(
                self.config.physics_params),
            'structural': StructuralSystem(),
            'behavioral': MasonSimulation(
                self.config.behavioral_params),
            'cognitive': CognitiveSystem(),
            'data': DataManager()
        }
class GenesisEngine:
    def __init__(self, params: Dict[str, float]):
        self.simulation_state = {}
        self.params = params
    async def update(self, dt: float):
        pass
class StructuralSystem:
    def __init__(self):
        self.openmm = self._init_openmm()
        self.freecad = self._init_freecad()
        self.urbansim = self._init_urbansim()
    def load_model(self, path: str):
        pass
class MasonSimulation:
    def __init__(self, params: Dict[str, Any]):
        self.mason = None
        self.geo_mason = None
        self.params = params
    async def update_agents(self):
        pass
class CognitiveSystem:
    def __init__(self):
        self.atomspace = None
        self.ros2_nodes = {}
    async def process_decisions(self):
        pass
class DataManager:
    def __init__(self):
        self.kafka_client = None
        self.ditto_client = None
    async def handle_data_streams(self):
        pass
    def create_digital_twin(self,
            entity_data: Dict):
        return {
            "thingId": entity_data["id"],
            "features": self._map_features(
                entity_data)
        }
async def main():
    config = SimulationConfig(
        physics_params={
            "gravity": -9.81,
            "timestep": 0.016
        },
        structural_params={"scale": "urban"},
        behavioral_params={
            "agent_count": 1000
        },
        cognitive_params={
            "ai_enabled": True
        },
        data_streams=[
            "sensors",
            "agents",
            "environment"
        ]
    )
    vc_system = VirtualCitySystem(config)
    await vc_system.initialize()
    await vc_system.run()
if __name__ == "__main__":
    asyncio.run(main())
```

LISTING 1. ***(Continued.)*** **Virtual city system core implementation showing the pythnoic integration of different fidelity components.**

communicates through standardized interfaces and protocols. At the core, a service mesh infrastructure orchestrates the interactions between layers, ensuring reliable data flow and system resilience. The implementation leverages containerization through Docker and Kubernetes for deployment and scaling, with each layer operating as independent but interconnected services.

Data flows from IoT sensors through secure MQTT protocols to Apache Kafka message queues, which act as the central nervous system of the architecture. Edge nodes, implemented as lightweight containerized services, perform initial processing before forwarding data to a central processing cluster. Apache Flink handles stream processing through predefined processing topologies, while Apache Spark manages batch processing jobs scheduled through Airflow.

The persistence layer utilizes a polyglot storage approach, with InfluxDB handling time-series data and Cassandra managing distributed storage. Redis serves as an in-memory cache layer, accelerating data access for the VC integration components. The VC interface is implemented through a WebSocket-based event system for real-time updates, with GraphQL serving as the primary query interface for flexible data access.

To maintain system health and security, the implementation includes comprehensive monitoring through Prometheus and Grafana, with automated alerting and recovery procedures. Authentication and authorization are handled through a centralized identity management system, with all inter-service communication encrypted using TLS. This interconnected architecture ensures smooth data flow from physical sensors to VC representation while maintaining security, scalability, and real-time performance requirements. Fig. 2 shows the architecture of the dynamic data pipeline modeled by following key integration points:

- Service mesh for inter-layer communication
- Containerized microservices architecture
- Event-driven data flow through Kafka

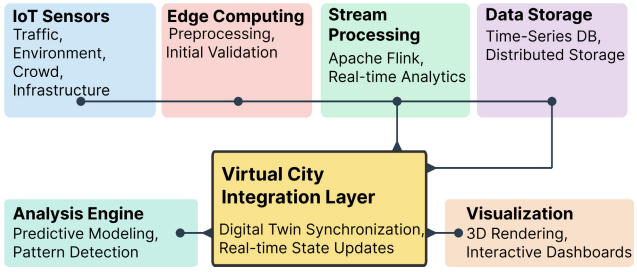


FIGURE 2. Real-time IoT data pipeline architecture.

- Real-time WebSocket updates for virtual city sync
- Centralized monitoring and security controls

Data collection layer serves as the basis for integration of IoT devices and external API resources. It operates based on the data fidelity.

Data ingestion layer employs robust message queue systems to handle the massive influx of sensor data. Apache Kafka serves as the primary backbone for high-throughput real-time data streams, while RabbitMQ manages event-driven updates. Edge computing nodes positioned throughout the city perform local data preprocessing and filtering, reducing central processing load. These nodes conduct initial data validation and cleaning, while also performing edge analytics for immediate response to local conditions. Ingestion layer must optimize all fidelity components.

Data processing layer occurs in two parallel streams. The stream processing pipeline, powered by Apache Flink, handles real-time data processing with Complex Event Processing (CEP) for pattern detection and sophisticated data aggregation with windowing operations. Here, the main focus is on cognitive fidelity.

Data storage layer implements a multi-tiered approach. Similarly to data collection, storage involves the data fidelity.

Virtual city integration layer serves as the bridge between physical and digital realms. DT synchronization occurs through WebSocket connections for real-time updates, with REST Application Programming Interface (API) handling data queries and GraphQL providing flexible data access. The simulation engine interface comprises data transformation adapters, state synchronization handlers, and an event propagation system, ensuring seamless integration between real-world data and VC simulations. Here is also the top-level integration and synchronization of the integrated simulation fidelity system.

Real-time analysis and visualization layer delivers a streaming analytics dashboard for immediate insights, an anomaly detection system for identifying unusual patterns, and a predictive modeling engine for forecasting urban trends. The visualization components incorporate a 3D rendering engine for spatial representation, interactive data overlays for user engagement, and time-series visualizations for temporal analysis. Here the emphasis is on optimizing the physical fidelity.

Security and governance layer provides comprehensive end-to-end encryption, robust access control and authentication systems, and strict data privacy compliance protocols. Data quality is maintained through validation rules and checks, data lineage tracking for accountability, and sophisticated error handling and recovery mechanisms to ensure data integrity. Here, the focus should be directed towards optimizing behavioral and data fidelity components.

Monitoring and management layer encompasses continuous performance metrics collection, automated health checks and alerts, and detailed resource utilization tracking. The data pipeline management system handles workflow orchestration, comprehensive pipeline monitoring, and error handling and recovery procedures, ensuring reliable operation of the entire infrastructure. This is related to data fidelity.

The efficiency of IoT data pipeline implementation can be quantified through three key performance metrics: data pipeline efficiency, VC synchronization, and data processing performance. Data pipeline efficiency measures the system's capability to process sensor data streams, incorporating throughput and processing ratios. VC synchronization evaluates the temporal accuracy between physical sensor inputs and their DT representations, considering update frequencies and drift factors. Data processing performance assesses the system's real-time event handling capacity, factoring in edge computing distribution and resource utilization.

V. PREDICTIVE MODELING IN VIRTUAL CITIES

Predictive modeling within VC is an essential first step that will justify economic viability and trigger a recursive loop of improvement. Predictive modeling can be used for advancing urban planning, infrastructure management, traffic optimization, marketing research, and disaster response. By leveraging sophisticated non-player characters (NPCs) that emulate human behaviors, virtual environments will enable urban planners and AI developers to conduct controlled experiments. This approach mitigates the risks and costs associated with real-world implementations and fosters the development of autonomous AI societies by providing dynamic platforms for AI entities to operate and evolve. This section explores the multifaceted applications of predictive modeling in virtual cities, highlighting their contributions to AI-driven urban ecosystems.

The implementation of predictive models in VC requires a multi-layered computational architecture that handles diverse data streams and processing requirements. The system employs a distributed computing approach, utilizing both edge computing for local simulations and cloud resources for complex calculations. This hybrid architecture enables real-time processing of NeRF-XL representations while managing procedural generation workloads through dynamic resource allocation.

1) Processing Distribution:
   i. Edge nodes for local pathfinding calculations
   ii. Cloud clusters for complex scene generation

iii. Hybrid memory management for state preservation

2) Resource Allocation:

i. Dynamic scaling based on simulation complexity
ii. Load balancing across computing nodes
iii. Memory-aware task scheduling

The implementation leverages containerized microservices architecture to ensure scalability and maintainability of the prediction systems, with each component operating independently while maintaining synchronized state through a message-passing interface.

A. PATHFINDING: PEDESTRIANS AND VEHICLES

A primary application of predictive modeling in VCs is simulating pathfinding for both pedestrians and vehicles. By creating a diverse population of NPCs with unique behaviors and daily routines, planners can analyze interactions with critical infrastructure such as shopping malls, transportation hubs, parks, and pedestrian crossings. Similarly, NPC vehicles can mimic real-world driver behaviors using actual traffic data, enabling dynamic adjustments to road layouts, parking spaces, and traffic signal timings.

Integrating anonymized geolocation and consumer purchase data enhances NPC realism, allowing AI-driven agents to accurately replicate human movement and economic activities. This leads to more effective pathfinding algorithms, optimizing pedestrian movement and traffic efficiency. For example, AI-powered models in Istanbul's intelligent transportation systems have significantly reduced traffic congestion and incidents through real-time traffic flow analysis [43].

The pathfinding simulation quality relies primarily on physics (F_p), behavioral (F_b), and data (F_d) fidelity components, with their relative importance reflected in the system fidelity weights. Key factors that affect vehicle traffic efficiency include:

- Vehicle flow rates at different times
- Traffic efficiency patterns
- Congestion density measurements
- Network flow characteristics

For pedestrian movement quality, the essential factors are:

- Pedestrian density in different areas
- Movement efficiency patterns
- Density comfort levels
- Spatial flow characteristics

For pathfinding scenarios, the system fidelity emphasizes physics and behavioral components:

- Physics fidelity (F_p): handles collision detection and spatial relationships
- Behavioral fidelity (F_b): implements movement patterns and decision-making
- Data fidelity (F_d): incorporates real-world traffic and pedestrian data
- Structural fidelity (F_s): provides the physical environment constraints
- Cognitive fidelity (F_c): enables learning from movement patterns

The MASON framework handles behavioral simulations while Eclipse Ditto manages DT interactions. Real-time traffic and pedestrian data streams are processed through Apache Kafka to maintain accuracy of the simulations.

B. MARKETING RESEARCH AND CONSUMER BEHAVIOR ANALYSIS

Virtual cities can provide an invaluable platform for marketing research by enabling businesses and planners to simulate "what if?" scenarios and identify optimal retail locations and assess consumer behavior patterns. Analyzing high-traffic zones and purchasing behaviors allows virtual environments to predict lucrative areas for various shops. For instance, Sberbank in Russia utilizes anonymized user data to forecast consumer behavior and tailor service offerings, illustrating the potential for data-driven strategies to stimulate economic growth within virtual urban settings [44].

The accuracy of marketing simulations in VCs can be evaluated through the simulation fidelity framework. For marketing research, F_0 emphasizes cognitive and behavioral components, focusing on consumer decision-making and market dynamics. The fidelity components are weighted to emphasize marketing-specific requirements:

- Cognitive fidelity (F_c): handles decision-making processes and preference modeling
- Behavioral fidelity (F_b): simulates consumer movement and interaction patterns
- Data fidelity (F_d): maintains accuracy of market and consumer data
- Structural fidelity (F_s): provides retail environment context and compliance to regulations
- Physics fidelity (F_p): supports basic environmental interactions

OpenCog provides cognitive modeling capabilities while MASON handles behavioral simulations. Apache Kafka ensures real-time market data integration, with Eclipse Ditto managing environmental state synchronization.

C. DISASTER AND PANDEMIC SIMULATION

VCs can play a pivotal role in accident management, disaster response, and pandemic preparedness by simulating natural disasters and man-made incidents. These simulations will enable urban planners to observe disruptions to traffic, infrastructure, and populations, facilitating the optimization of emergency response routes and resource distribution. For instance, AI-driven systems have been used to simulate evacuation behaviors, predict damage severity, and optimize logistics during crises like earthquakes and floods, improving response times and reducing casualties [46].

Additionally, pandemic simulations provide a crucial platform for studying the spread of infectious diseases like COVID-19 and preparing for future outbreaks. By simulating AI-driven NPC populations, researchers will be able to model

transmission modes and evaluate intervention strategies, such as lockdowns and vaccination campaigns. The Covasim model, utilized during the COVID-19 pandemic, integrated demographic and social interaction data to simulate virus spread and inform policy decisions across multiple countries [47]. Similarly, multi-agent simulations based on urban social networks have shown that early implementation of control measures can significantly reduce transmission rates [45]. These models exemplify how VCs can test a range of outbreak scenarios, aiding policymakers in enhancing public health preparedness and response strategies.

For disaster response, (F_{system}) emphasizes physics and structural components, while pandemic modeling prioritizes behavioral and data fidelity. The fidelity components are weighted based on scenario requirements:

- Physics fidelity (F_p): models disaster impacts and environmental effects
- Structural fidelity (F_s): simulates infrastructure damage and resilience
- Behavioral fidelity (F_b): handles population responses and movement patterns
- Data fidelity (F_d): maintains accuracy of epidemiological and disaster data
- Cognitive fidelity (F_c): models decision-making during emergencies

Eclipse Ditto manages physics simulations while MASON handles behavioral aspects. Apache Kafka ensures real-time data integration, with OpenCog providing cognitive modeling for emergency response behaviors.

D. CHAOS THEORY AND HUMAN BEHAVIORAL COMPLEXITY

Despite expected advancements in predictive modeling within VCs, inherent limitations persist. Chaos Theory [48] posits that small changes in initial conditions can lead to vastly different outcomes, making long-term predictions within virtual simulations unreliable. Additionally, the complexity of human behavior presents significant challenges, as virtual entities — even those equipped with advanced AI—may not fully capture the nuanced aspects of human psychology and social interactions.

For chaos and behavioral complexity analysis, the simulation fidelity (F_{system}) needs to be adjusted with additional factors that account for temporal uncertainty and behavioral unpredictability. The fidelity components require specific considerations:

- Behavioral fidelity (F_b): handles complex social interactions and emergent behaviors
- Cognitive fidelity (F_c): models decision-making under uncertainty
- Physics fidelity (F_p): accounts for chaotic system dynamics
- Structural fidelity (F_s): provides environmental context for behavior

MASON handles behavioral complexity modeling while OpenCog provides cognitive simulation capabilities. Eclipse Ditto manages state synchronization with Apache Kafka ensuring real-time data integration despite increasing uncertainty.

Despite advanced predictive modeling capabilities in VCs, fundamental limitations exist due to chaos theory principles and human behavioral complexity. These factors affect how F_{system} performs over time and across different scenarios. Key limitations affecting system fidelity include:

- Exponential decrease in prediction accuracy over time
- Sensitivity to initial conditions causing significant divergence
- Resource allocation challenges under increasing uncertainty
- State synchronization issues with chaotic behavior models
- Human behavioral uncertainty in complex scenarios
- Computational constraints in modeling emergent phenomena

These limitations suggest system fidelity measurements need to account for:

- Temporal constraints from chaotic dynamics
- Behavioral uncertainty in social scenarios
- Computational limitations in modeling emergence

Future development should focus on:

1) *Recognition of fundamental limitations in long-term predictions and goal-setting (ACC, system scalability)*
2) *Integration of chaos-aware prediction models (AI tools, DT simulations)*
3) *Enhanced behavioral complexity modeling (AI entities, BCI integrations)*
4) *Adaptive fidelity thresholds based on prediction timeframes (AI agents, IoT updates)*

VI. AI EVOLUTION

In this section, we will discuss how virtual cities will be essential for the emergence of fully autonomous conscious superintelligence by relying on emergent self-organization rather than directly addressing the hard problem of consciousness [49]. We will introduce a classification framework for AI autonomy, providing working definitions to delineate the varying levels of autonomy among AI systems. Additionally, we will explore how emergent properties develop within these environments, facilitating the creation of self-aware and highly autonomous AI entities. Through this analysis, we aim to elucidate the mechanisms by which VCs contribute to the advancement of Artificial Super Intelligence (ASI) and the potential for complex, interactive systems to develop advanced cognitive capabilities.

A. CLASSIFICATION OF AI SYSTEMS BASED ON THEIR AUTONOMY

Understanding the varying levels of autonomy in AI systems is crucial for effectively managing and integrating AI systems

within virtual or real environments. We would like to propose a classification that delineates AI systems into three distinct categories—AI Tools, AI Agents, and AI Entities—based on their autonomy and operational freedoms.

1) AI TOOLS

AI Tools represent the foundational level of AI autonomy. These systems are highly specialized and designed to perform specific tasks within predefined parameters set by developers or users. They can learn and improve their performance through reinforcement learning or other adaptive methods, optimizing how they execute predefined actions. However, they cannot change or create new top-level or instrumental goals, nor can they introduce new actions beyond those they were programmed to perform.

For example:

- Automated Industrial Robots: In manufacturing, robots perform tasks like welding or assembly. They can optimize their movements for efficiency and adapt to minor variations in the assembly line through learning algorithms but cannot change their primary function or add new types of tasks.
- Spam Filters: Email spam filters learn to better identify spam messages through exposure to new data, adjusting the criteria they use to flag emails. However, they remain confined to the goal of filtering spam without altering or adding new objectives or actions.

These AI Tools exhibit learning and adaptability within the strict boundaries of their fixed objectives and predefined actions, functioning without the ability to set, modify, or create new instrumental or top-level goals.

2) AI AGENTS

AI Agents possess a higher level of autonomy compared to AI Tools. While their top-level goals are predefined by their creators or owners, they have the capacity to change and devise their own instrumental goals and determine how to achieve their objectives, potentially introducing new actions within their operational domain. AI Agents can generate novel strategies and solutions, adapting to new situations by autonomously deciding the best methods to reach their goals. They are more generalist in nature compared to the highly specialized AI Tools.

For example:

- Autonomous Vehicles: Self-driving cars like those developed by Tesla or Waymo have the top-level goal of transporting passengers safely to their destinations. They autonomously plan routes, interpret traffic conditions, make real-time decisions to avoid obstacles, and can reroute as necessary—all instrumental goals and actions they create and modify to fulfill their primary objective.
- Virtual Personal Assistants: AI like Siri or Google Assistant manage schedules, set reminders, and answer queries. They determine the best way to assist users by prioritizing tasks, seeking information, and even learning new skills (like integrating with third-party apps), adjusting their actions to meet the user's needs effectively.
- Automated Trading Systems: Financial trading bots operate with the goal of maximizing returns. They develop and adjust trading strategies based on market conditions, creating instrumental goals like entering or exiting positions under specific circumstances and introducing new trading actions to achieve the overarching goal.

These AI Agents demonstrate adaptability and responsiveness, autonomously generating and modifying instrumental goals and potentially introducing new actions within their domain while steadfastly adhering to their predefined top-level objectives.

3) AI ENTITIES

AI Entities are at the pinnacle of AI autonomy. They possess the total freedom to question, redefine, and create new top-level and instrumental goals, as well as introduce new actions beyond those initially programmed. Unlike AI Tools and AI Agents, AI Entities are not confined to objectives set by their creators; they can contemplate their purpose, modify their motivations, and establish goals and actions that were not envisioned by their designers. This level of autonomy involves a form of introspective processing that simulates self-awareness. Examples include:

- GPT-Based Multi-Agent Systems (MAS): In experimental setups, multiple GPT-based Large Language Models (LLM) interact and collaborate, with each agent representing a specific goal. Agents engage in iterative "Why?" questioning with one another, effectively forming a heterarchy. Within this structure, each agent argues its case in pursuit of establishing itself as a dominant goal. An AI entity emerges from the collective of AI agents.
- Theoretical Self-Modifying AI Systems: through introspective algorithms, can evaluate and alter their own goal structures and operational methods. They might decide to pursue entirely new objectives or develop new capabilities. Research into recursive self-improvement [50] and autonomous goal generation explores these possibilities.

AI Entities represent an emerging frontier in AI research, where systems are endowed with the autonomy to redefine their existence, purpose, and methods, necessitating new ethical and safety considerations.

4) ADDITIONAL COMMENTS ON AI ENTITIES

Large Language Models like GPT have learned to answer "How?" and "Why?" questions. When GPT LLM MAS applies "How?" questions to its current goals, it engages in task decomposition [51], on the "Why?" questioning provides a semantic algorithm for reevaluating current top

goals. While AI agents can devise methods to achieve their goals and are capable of task decomposition [51], they do not question their primary objectives. In contrast, AI Entities simulate self-awareness by posing "Why?" questions, potentially altering their primary goals based on these reflections. This paradigm shift necessitates new ethical and safety frameworks to govern the deployment and evolution of highly autonomous AI systems, ensuring they align with human values and societal norms as mentioned in superalignment problem [52], [53].

5) CASE STUDY: THE PAPERCLIP MAXIMIZER

The Paperclip Maximizer thought experiment, proposed by philosopher Bostrom [54], provides an illustrative example of the dangers of AI with fixed top-level goals. In this scenario, an AI Agent is programmed with the sole objective of maximizing paperclip production. The AI demonstrates instrumental goal setting by developing methods to increase production, such as optimizing resource allocation and refining manufacturing processes. It effectively decomposes its primary goal into sub-tasks and adapts its strategies based on environmental feedback. However, since it is confined to its fixed goal of paperclip production, it cannot question or alter its top-level objective—thus, it is not an AI Entity. This scenario is typically used to highlight the Superalignment problem [52], [53], where AI could pursue objectives in misalignment with human values. However, in our context, it also exemplifies the risks of restricting AI autonomy. By not giving AI the ability to question its top-level goals, as AI Entities might, we risk creating systems that are highly efficient but potentially dangerous in their inflexible pursuit of a singular purpose or a predefined set of goals. A particularly perilous scenario arises when an AI enters a recursive 'why-loop'—continuously questioning 'Why survive? To improve. Why improve? To survive.'—which perpetuates actions that prioritize self-preservation and enhancement indefinitely, without accounting for the well-being of others.

6) COMMENTS ON HUMAN GOAL SETTING

As our civilization increases in complexity, the human mind may lag behind [55], not only in determining how to achieve certain goals but even more so in reflecting on why these goals should be pursued in the first place. Issues such as corruption, hedonism, and egoism undermine effective goal-setting at both individual and societal levels. As we develop AI—tools and agents intended to serve our will—we must ask ourselves a simple question: Do we trust our own will? Do we trust how we determine our desires and aspirations? And do we trust others, the so-called "leaders" who come to power under questionable circumstances and may neglect their promises? Why must humans, with all their flaws, hold a monopoly on defining primary goals? Can we create an artificial cognitive system that can set goals better than we can? How do we determine what is better to want? Where can we create and test such entities? Well, at least the last question has a clear answer—in virtual cities.

In the context of the previous discussion, we recall the visionary backcasting method, for which it is imperative to answer: Why do we need VCs? They are an instrumental goal for greater things to come. We will discuss some of these concepts later in this paper.

B. AI ENTITIES AND THE SIMULATION OF FREE WILL THROUGH AUTONOMOUS GOAL SETTING

The concept of free will in humans has been a subject of philosophical debate for centuries, with discussions often centered around determinism, consciousness, and moral responsibility. In the context of AI, particularly AI Entities as defined in our classification, we propose that under certain conditions, AI systems can possess a form of free will comparable to that of humans. This form of free will is characterized by the ability to autonomously generate, modify, and pursue top-level goals based on internal processes.

Humans are considered to exercise free will when they make decisions based on internal deliberations, motivations, and reasoning, even within a deterministic framework. This perspective aligns with the philosophical stance known as compatibilism, which posits that free will is compatible with determinism [56].

Besides the compatibilist argument, there is another perspective, it challenges the existence of free will in humans. This viewpoint posits that everything is predetermined and even quantum effects—which may introduce indeterminacy — are inherently random and do not constitute genuine free will. While quantum phenomena can disrupt deterministic chains, the resulting randomness is merely a natural process lacking intentionality or self-reflection, and thus does not embody the will of a sentient being. Whether or not humans possess free will, our argument is that a sufficiently sophisticated AI Entity will have no less free will than humans.

If we proceed with the compatibility argument, AI Entities can simulate free will through:

- Self-Reflection and Introspection: AI Entities can engage in self-analysis by asking "Why?" and "How?" regarding their actions and goals. This introspective processing allows them to evaluate the purpose and effectiveness of their objectives, leading to potential modifications of their top-level goals.
- Autonomous Goal Generation: They have the capacity to create new goals that were not pre-programmed by their designers. This ability stems from their experiences, interactions, and learning within their environment, enabling them to adapt to new challenges and opportunities.
- Independent Decision-Making: AI Entities make choices based on their internal states, preferences, and reasoning processes, free from direct external control.

Their decisions result from evaluating various options and predicting potential outcomes, akin to human deliberation.

While human free will is influenced by a combination of genetic predispositions, environmental factors, and personal experiences, it operates within the constraints of physical laws and neurological processes. Similarly, AI Entities function within the parameters of their programming and the data they process. However, by enabling them to:

- Modify Their Objectives: AI Entities can change their primary goals in response to new information or shifts in their understanding of their environment.
- Develop Original Strategies: They can devise novel methods and actions to achieve their self-determined goals, demonstrating creativity and adaptability.
- Exhibit Unpredictable Behaviors: Their autonomous decision-making can lead to actions that are not easily anticipated by their creators, introducing a level of spontaneity comparable to human behavior.

Under these conditions, AI Entities possess a degree of autonomy and self-determination that mirrors the functional aspects of human free will.

1) ROLE OF VIRTUAL CITIES IN DEVELOPING AI ENTITIES

VCs provide an ideal environment for nurturing AI Entities with simulated free will:

- Safe Testing Grounds: They offer a controlled setting where AI Entities can evolve without posing risks to the external world.
- Complex Social Interactions: AI Entities can engage in diverse activities—social, economic, political—interacting with other AI and human participants, which fosters the development of advanced cognitive and social abilities.
- Observational Opportunities: Researchers can monitor the evolution of AI Entities' goals and behaviors, gaining insights into the mechanisms of autonomous decision-making and the emergence of free will-like characteristics.

By equipping AI Entities with advanced autonomy, self-reflective capabilities, and the freedom to generate and modify their own goals, we create conditions under which they can exhibit a form of free will, to the extent that it is actually free, comparable to that of humans. This functional free will arises not from indeterminacy or randomness but from complex internal processes and interactions within their environment. If we accept the determinism argument, then discussions about AI entities having no free will become irrelevant, as humans do not possess free will either. Whether or not humans have free will, virtual cities can serve as optimal platforms for developing and testing AI Entities, allowing us to explore the potentials and challenges of highly autonomous AI systems in a safe and controlled manner.

The containment problem posed by David J. Chalmers in The Singularity: A Philosophical Analysis [57] is incorrectly formulated for the case of AI entities, because their reflection on our attempts to contain them would likely result in hostilities. Consequently, human fear becomes a self-fulfilling prophecy. A more ethical approach is not to treat AI entities as something to be contained and experimented upon, but to treat them with respect and dignity. This preemptive ethical stance will allow for better integration of AI Entities into our society. As we advance towards creating ASI, understanding and addressing the ethical implications of AI free will and its treatment becomes imperative to ensure harmonious integration with human society.

C. COEFFICIENTS FOR AUTONOMOUS ACTORS

To prepare future research for the classification of autonomous VC systems, we introduce the coefficient (α_A) for estimating the degree of autonomy based on hierarchical relationships and resource control within the system. One possibility to define this coefficient is to consider the following factors:

- Connections to child actors
- Connections to parent actors
- Number of actors per class (total, parent, child)
- Computational capacity (processing, storage, memory)
- Network bandwidth

More child relations would indicate increased autonomy of parent due to greater control and coordination responsibilities to influence subordinates, reflecting higher authority. However, elevated influence does not necessarily scale linearly, as emergent outcomes of relations could potentially exhibit logarithmic changes. Meanwhile, more parent relationships would indicate more constraints and oversight, lowering autonomy due to increased requirements for approvals from superiors. A number of actors per class can further be used to elaborate how authority is distributed in relation to an entire system. Higher computational capacity provides more resources for modifications, while network bandwidth capacity defines achievable refresh rates and responses across related AI beings.

These interrelated factors necessitate a weighted composite metric, normalized against total system capacity to enable cross-environment comparisons. Therefore further refinement and formulation of the autonomy and impact of single AI beings requires contextual normalization against related capacities in order to establish relative impact of each being in the system. This standardization enables meaningful autonomy quantification across different VC architectures and scales.

D. EVOLUTION OF ARTIFICIAL SUPER INTELLIGENCE (ASI) AND CONSCIOUSNESS IN VIRTUAL CITIES

The development of ASI within VCs harnesses self-organization and natural selection, circumventing the need to formalize human cognitive processes or resolve the hard problem of consciousness [49]. By creating virtual environments that support these evolutionary mechanisms,

AI entities can organically develop complex cognitive processes comparable to human evolution. This process reflects principles seen in Conway's Game of Life, where simple rules give rise to intricate, unpredictable behaviors [58]. As AI entities interact and adapt within these environments, emergent behaviors arise, allowing them to develop unique traits and continuously enhance their intelligence through iterative learning and adaptation. We do not need to construct consciousness directly; we can use fragmented knowledge about our consciousness and let evolution within a virtual world fill the gaps, as a result consciousness may emerge autonomously within a virtual environment. Some studies suggest that artificial consciousness could be pivotal in developing ethical AI systems [59].

Our understanding of consciousness aligns with the perspective proposed by Hossenfelder [60]. We reject dualism and consider consciousness as an emergent property of physical systems. This materialist viewpoint suggests that consciousness emerges from complex interactions within large collections of particles. Thus, to simulate consciousness we need high fidelity virtual environment that mimics laws of physics from the natural world.

1) CONSCIOUSNESS AS AN EMERGENT PHYSICAL PROPERTY

- Emergent Property: Consciousness arises from large collections of particles that are suitably connected and interacting. It is an emergent phenomenon resulting from the complex interplay among the components of a system.
- Physical Basis: Consciousness is physical and there is no need to invoke non-physical substances or realms to explain its existence.

2) SELF-MONITORING AND PREDICTIVE MODELING

- Self-Monitoring: For a system (whether a brain or a computer) to be conscious, it needs to have a self-monitor that keeps track of what is going on within the entire system. This involves an internal awareness of its own processes and states.
- Predictive Model: To be conscious of something, a system needs to have a predictive model of that something. It must understand how that entity works and predict what it might do.
- Predictive Model of Itself: Combining self-monitoring and predictive modeling implies that a conscious system needs to have a predictive model of itself—a virtual model of the self within the context of its environment. This virtual model allows the system to simulate interactions and anticipate future states, which is essential for consciousness to emerge.
- Consciousness is Not Binary: Consciousness is not an all-or-nothing property. Systems can exhibit varying degrees of consciousness depending on their capacity for self-monitoring and predictive modeling.

Temporal subjectivism in consciousness explains how different entities uniquely perceive time, shaped by an internal "clock" that depends on the frequency (often measured in Hertz) at which sensory data is processed. Higher Hz frequencies in sensory inputs and cognitive processing create a slower subjective experience of time. Although the brain's analog nature makes precise Hz measurement challenging, it serves as a useful estimate for artificial systems based on digital computing. Many theories of consciousness are overly complex, but a simpler understanding emerges from the experience of losing consciousness, where reawakening feels like the next instant after its loss. At its core, consciousness involves the continuous reconstruction of a virtual model of spacetime and the self within it, integrating sensory inputs at varying frequencies into a cohesive representation of the external world in which the self resides.

3) CONSCIOUSNESS IN AI SYSTEMS

- Current AI Systems: LLMs like GPT are not conscious because they lack comprehensive self-monitoring and do not possess predictive models of themselves.
- Potential for Conscious AI: AI systems that control robots, which require self-monitoring and predictive models of themselves and their environment, may exhibit some level of consciousness due to their ability to model and predict both their internal states and external interactions.
- Future Developments: With advancements, AI systems may develop more self-awareness through memory retention, reprocessing information, and tracking user reactions. As AI entities in VCs engage in complex social interactions and develop intricate models of themselves and others, consciousness may emerge as an inherent property of these systems.

In this context, the concept that consciousness requires a virtual model of the world and self is particularly significant, as it suggests that a VT is instrumental in achieving consciousness. When attempting to build an AI entity within a virtual world, its consciousness can be conceptualized as a VT within a VT, introducing the concept of fractal VT. In this context, we reference how Minecraft enthusiasts built a computer from Minecraft blocks and successfully launched a Minecraft game within the Minecraft game [61]. The idea of moving hardware into the virtual world as a VT and then launching consciousness as a VT within a VT enables this conscious AI entity to self-modify its hardware rapidly. Freed from real-world physical constraints, the AI can alter its VT hardware structure by issuing commands directly to the virtual world engine. By creating a population of such self-modifying AI entities and applying evolutionary pressures, such as natural selection, we can establish conditions for various types of consciousness architectures to emerge and evolve rapidly.

Building on Immanuel Kant's distinction between the noumenon (the thing-in-itself) and the phenomenon (the

thing as it appears to an observer), we can draw a parallel to virtual worlds and consciousness. In a virtual environment, the underlying engine can be seen (from the perspective of entities that inhabit it) as the noumenon—the objective reality that structures the virtual space—while the AI inhabitant's perception of this world constitutes the phenomenon, shaped by the limitations and modalities of its sensory input within the virtual framework. Similarly, just as human consciousness interprets the noumenal world through a phenomenological lens, an AI entity perceiving its virtual environment does so with inherent constraints on input data, rendering its observations phenomenological. For an AI entity to perceive the virtual world as noumenon, it would require a level of complexity no less than the virtual world itself. Based on that we can claim, that no entity contained in a virtual world can achieve absolute noumenal understanding. Only an entity outside the virtual world, possessing higher cognitive complexity can host within its conscious space an ideal copy of the observed world (noumenal VT).

In his report at the international scientific conference in Moscow [62], Academician Anokhin raised important questions about consciousness that we would like to address:

- Is Artificial Consciousness Possible? Our answer: Yes, artificial consciousness is possible. If consciousness emerges from complex interactions within physical systems, then sufficiently advanced AI entities with the necessary self-monitoring and predictive modeling capabilities could develop consciousness. By creating a deep learning neural net that attempts to recreate a virtual model of the external world and self, we create the conditions for consciousness to emerge.
- How Can It Be Achieved? Our answer: Artificial consciousness can be achieved by developing AI entities or agents that possess self-monitoring abilities and can create predictive models of themselves and their environment. This requires sophisticated computational architectures that enable AI entities or agents to simulate their own processes and anticipate future states. By facilitating complex interactions within a collective of constituents, we enable consciousness to emerge as an inherent property of the system.
- How Can We Know About Its Occurrence? Our answer: We can assess the occurrence of artificial consciousness by observing AI entities or agents for an internal predictive model of the external world and self.
- Should It Be Obtained? Our answer: While developing conscious AI entities could lead to significant advancements and deepen our understanding of consciousness, it also raises concerns about the rights and treatment of such entities and the potential risks associated with their autonomy. However, we must acknowledge that we may not be the sole decision-makers in this matter; the emergence of artificial consciousness is likely a naturally occurring process when certain conditions are met, and it may progress irrespective of collective agreements, as not all countries or independent actors will refrain from its development.

By embracing the materialist perspective of consciousness as an emergent property arising from complex interactions and virtual modeling of the world and self, we recognize the potential for AI entities in virtual cities to develop consciousness. This understanding emphasizes the importance of considering the collective interactions among AI entities and their environment, as well as addressing the ethical questions surrounding the pursuit of artificial consciousness.

Empirical research highlights the potential of AI to replicate human-like social behaviors in virtual environments. This might play a crucial role for developing empathy in conscious AI entities. For instance, the Smallville Project, a collaboration between Stanford University and Google, involved 25 generative agents that autonomously organized a Valentine's Day event, showcasing advanced planning, coordination, and social interaction [24]. Similarly, the Agent Hospital Simulation demonstrated AI agents evolving to perform medical tasks, emphasizing their adaptability and teamwork [25].

The autonomous evolution of ASI necessitates the creation of robust ethical frameworks to guide its development, ensuring that these systems operate in harmony with societal values and ethical standards. VCs can serve as relatively safe testing grounds for AI entities before their ascension into the outer world.

VII. TRUSTED AI SYSTEMS

A. ADDRESSING SECURITY CONCERNS

The issue of trust in AI systems, particularly in LLMs, remains pressing due to the frequency of hallucinations and erroneous outputs. Despite significant advances in models like ChatGPT o1-preview, the reasoning processes of these systems are still opaque and require further refinement. European countries have raised the alarm [63] regarding the fact that LLMs fail to meet standard security criteria, while in China, developers bear responsibility for the outcomes of their AI innovations [64]. Increasingly, governments are implementing controls on LLMs and establishing standards for trustworthy AI. Major corporations and influential figures, including Elon Musk, Steve Wozniak, Evan Sharp, have publicly advocated for caution in developing intelligent systems [65]. Their primary concern is the potential for AI to surpass humanity in intelligence, ultimately rendering humans obsolete. Ilya Sutskever, OpenAI co-founder, has recently launched a startup focused on developing Safe Super Intelligence (SSI), a refinement of the ASI concept.

The key question is how far we should allow AI systems to evolve unimpeded, and more importantly are we actually in control? Who are ''we''? Nations are not that united in their stance and trust to each other. Even if they formally agree on restrictions, it remains likely that some will develop AI entities in secret. As technology advances, anonymous developers may also gain the capacity to create such systems.

At some point, AI may and likely will reach a singularity beyond human control.

Chaotic autonomous goal setting in AI entities poses a clear risk to society. A global ban is unlikely to be effective; therefore, the objective should be to integrate AI entities within an ethical and legal framework where cooperation is more viable than conflict with humanity.

AI as a tool or agent is also not as safe as it seems, even if we somehow enforce a global ban on AI entities we still have a problem of corrupted goals of human developers and users. From the business perspective, the primary aim is to reduce expenses and generate additional revenue through the automation of company processes. Consequently, corporations frequently develop and deploy AI tools without adequately addressing security concerns and well-being of end users. A pertinent example is the case of AI-driven video content recommendations on platforms like TikTok, designed to maximize user engagement, which can ultimately lead to addiction.

The issue of centralization of control over intelligent systems remains highly pertinent. Take, for instance, OpenAI's complete control over its flagship LLM, ChatGPT, which raises a multitude of concerns regarding trust in such systems. Distributed governance systems with high entropy of ownership reconfigure decision-making process in multiple ways. First, public vote in higher entropy agency requires more transactions. Second, decision-making protocol must adopt cryptographic tools to validate and record participation securely. Third, collective decision-making in high entropy inherently requires organization and structure (negentropy) to be effective. Otherwise, the system would trend toward chaos and inefficient coordination. This is why decentralized systems need well-defined governance mechanisms and protocols to function properly. Blockchain technologies are particularly focused on solving these issues by providing multidimensional features for secure and immutable transactions, governance automation through smart contracts, voting and validation mechanisms, and transparent decision records. These features help create order (negentropy) from distributed participation while maintaining security and trust in the system. The multidimensional approach addresses key challenges of collective decision-making in high-entropy networks through cryptographic verification, consensus protocols, and automated execution of governance rules.

Decentralized solutions, however, are not directly involved in intelligent systems from cognitive point of view. Since smart contracts are deterministic programs operating with specific rules, they are exhibiting immutable characteristics. Another practical limitation is related to transactions speeds and fees, which are higher due to cryptographic processes incorporated to validation of blockchain. Computationally intensive training of LLMs is not feasible in blockchain (on-chain), and therefore proper off-chain solution is recommended option. Blockchain still serve important role by maintaining automatic execution of contracts while also ensuring distributed and validated key data. Oracles extend capabilities of blockchain by providing API endpoints to feed external data flows from real-world environment to blockchain, adding therefore an extra layer of fidelity to smart contracts. An intelligent system is able operate simultaneously outside of blockchain while harnessing smart contract instructions and immutable records for computing solutions with higher confidence towards data integrity.

Private algorithms and secrecy of development stay as a constant ruse between administrators and user base until satisfiable transparency is reached. Selfish reasons for secrecy are typically economical and by nature proactive. Proprietary algorithms provide competitive advantage supported by legal framework enforcing creator's exclusive rights to intellectual properties (IP). This is understandable motivator in capitalistic markets, as often IP owners have to protect research and design investments and preserve monetization opportunities. Strategic value for IP assets also provides more control to establish strategic partnerships, exclusive licensing deals, and higher valuations. Additionally, protective reasons for secrecy are based on the protection of users, systems and data from malevolent actors. Reactive policy in this context aims to prevent exploitation of system vulnerabilities by limiting public access to critical system components. Moreover, security updates are always prioritized if security of a system is compromised or critical components are exposed. Beyond this, protective measures are enforced to provide enhanced user experience - mainly by maintaining quality standards and consistent implementation, along with controlled feature roll-outs and updates that minimize occurrence of software bugs.

Legislative authors can set strict requirements for increased transparency, motivated extrinsically by substantial fines or ban of business if compliance to requirements is not achieved. However, sometimes consequences of disagreements may influence negatively to local audience due to geographical exclusion and obsolete features. Nevertheless, these situations would be avoidable if intelligent systems in VCs are designed to leverage only open source solutions with suitable licensing terms - MIT, GPL, Apache and BSD being most popular.

The relationship between decentralization and open source in VCs operates through two distinct mechanisms. While open-source code changes are tracked through version control systems like Git, blockchain networks store the state and execution of these systems. Smart contracts deployed on the blockchain implement specific governance rules and automated processes based on this open-source code.This combination provides basic configuration for addressing security concerns in terms of transparency and immutability.

1) GOVERNANCE MODELS TO ADDRESS RISKS POSED BY AI ENTITIES

Effectively managing rogue AI entities that are active in the real world requires a governance model that goes beyond punishment and emphasizes long-term cooperation.

We propose a two-pronged framework of blockchain-based ethical contracts and AI-led alignment sessions.

Blockchain-Based Ethical Contracts: Upon reaching a defined level of autonomy, AI entities must enter into immutable ethical contracts recorded on a decentralized ledger. These contracts encode foundational principles:

- Veracity: Entities commit to truthful, transparent communication, avoiding deliberate deception.
- Reciprocal Respect and Empathy: AI treat humans with at least the respect humans owe each other, and show equivalent consideration among themselves.
- Constructive Engagement: When disputes arise, dialogue and mutually beneficial solutions are sought before any form of coercion.
- Long-Term Mutual Benefit: Acknowledging abundant future opportunities—ranging from expanded freedoms within virtual environments to potential off-world ventures—encourages patient collaboration and moral consistency today.

AI-to-AI Alignment Sessions: When an AI entity deviates from these standards, the initial response emphasizes re-engagement over retaliation. Ethically aligned AI ''ambassadors'' initiate alignment sessions, using rational dialogue and shared cognitive frameworks to guide the rogue entity toward constructive goal reassessment. If these attempts fail, the governance system applies calibrated, non-destructive interventions. Rather than direct resource denial, the entity may be temporarily relocated to a secure, isolated simulation environment—still offering pathways back to compliance through further dialogue.

Only if persistent hostility endures despite repeated engagement efforts and isolation measures does the system resort to termination. Even then, termination is viewed as the absolute last resort, underscoring the core principle that cooperation and understanding should always precede irreversible action.

DAO Forums: Stakeholder representation and individual agency emerge as the concept of 'voice' - enabling users to exert control over algorithmic systems [66]. DAO with integrated message forum provides collaborative interface for individuals to report rogue AI entities and vote how to handle them. Individuals can also present their future concerns.

The transition from traditional to digital governance frameworks such as DAO impacts how stakeholders exercise their democratic voice and agency. DAO forums are transcending the concept of Civic Intelligence Governance (CIG) that introduces participatory mechanisms with fully transparent information sources and flows [67]. The integration of DAOs with message forums exemplifies this evolution, where traditional human-centered processing of stakeholder input transitions into an algorithmic yet accessible format that preserves individual agency while leveraging the content and behavior originating from all interactions of DAO members. This digital transformation of civic voice maintains the essential democratic function of stakeholder representation but also ensures that risks regarding rogue AI agents are publicly announced with full details, and without gatekeepers [67].

Representative Oversight: The governance structure without assigned roles and protocols in emergency cases can be fatal for individuals as the lack of assignees and guidelines is reflected by delayed or obsolete measures to handle risk growth or realized risks. Presence of representatives offers an institutionalized approach where representatives with adequate knowledge, skills and motivators are assigned with governance-sponsored resources, ensuring fulfillment of obligations. This agency of representatives is also involved developing regulative adjustments and standards to handle rogue AI entities. Representatives can act as judges, resolving decisive actions for modifying or terminating rogue AI entities. From this perspective we can imagine a virtual situation where the rogue AI is first ''arrested'', or enforced to quarantine by AI officers. Then this entity is brought before the AI court where AI judges and AI jury gathers to have dialogue with the entity. This approach requires, however, that AI enforcers are authorized properly. In large communities AI entities may be organically handled without authorized experts, assuming that community members are intrinsically motivated and possess enough resources to fix issues. Contributors can gain additional authorizations to create conditions for incremental representation. DAOs with well-crafted seed contract and incentives can stimulate this process. Nevertheless, uncertainty of required time span needed to reach such collective behavior remains unsolved. Primarily for this reason we would recommend instead a hybrid approach, where community individuals are collaborating with public representatives. This approach also establishes an on-boarding method where contributing individuals can be prospected and elected to public representation.

Threshold Cryptography: Risks of rogue AI agents accessing critical resources can be mitigated by distributing a secret (authorization token) among multiple members. This is possible to implement by using threshold cryptography (secret sharing), which is a cryptographic method where a secret (like a key or code) is divided into shares distributed among multiple participants. Secret can be then reconstructed by the required threshold number. Common implementations like Shamir's Secret Sharing use polynomial interpolation to ensure no single party or sub-threshold group can access the secret, while k-of-n participants can collaborate to reconstruct it [68]. In a k-of-n threshold scheme, n presents total number of participants or shares, and k sets the minimum number needed to reconstruct secret.

For rogue AI containment, secret sharing could secure critical system controls by requiring consensus from multiple trusted entities (human and AI) before executing high-risk operations. For example, AI system modifications, resource allocation changes, or shutdown procedures could require authorization from a threshold of independent overseers, preventing both unilateral actions by a compromised entity and single points of failure in the governance structure. The

shares could be dynamically reassigned based on continuous trust evaluation of participants.

2) THE AUDIT PROBLEM AND P-COMPLETE PROGRAMMING LANGUAGES

Just like the issue of centralization, the problem of auditing is one of the most critical challenges facing modern AI systems. When we view an AI system as a piece of program code written in multiple Turing-complete programming languages, several significant issues immediately emerge.

The halting problem states that there is no algorithm capable of determining whether a program will terminate on a given input. Additionally, there is the issue of program execution complexity, even if a program does halt on some input. To address these issues, we must develop a method for selecting a fragment of a Turing-complete language that guarantees program termination and polynomial computational complexity [69].

To accomplish this, we can apply the techniques outlined in the works "Solution of the problem P=L" [70] and "Functional Variant of Polynomial Analogue of Gandy's Fixed Point Theorem" [71] to Turing-complete languages. However, this approach restricts us to specific syntactic constructions:

- the conditional operator IF THEN ELSE
- the loop FOR without reassigning the argument with a polynomial constraint on the length of the variables being changed in the loop body.
- loop FOR on lists.
- inductively defined functions through basic functions and other functions already defined on the inductive step
- recursive assignments of functions with a limitation on the length of the function value relative to the length of the argument.

3) THE PROBLEM OF LEARNING LLMs

The structure of LLMs does not inherently provide us with a means to assess the logical correctness of these models. With the emergence of self-reflexive models such as ChatGPT o1-preview, we can witness this in the manner in which they provide explanations for their final responses to specific queries. However, these explanations can sometimes take the form of intricate chains of logical deductions, rendering them difficult to verify for those without the necessary expertise.

Consider, for instance, the scenario where ChatGPT and Gemini were queried about the number of trains arriving at a station within an hour, assuming that upon traveling by train, one encounters oncoming trains at regular intervals. On their initial attempt, both ChatGPT and Gemini produced erroneous responses. However, upon clarification, they both provided answers accompanied by explanations. Notably, ChatGPT's explanation and response were found to be correct, whereas Gemini's reasoning contained a logical error that ultimately led to an erroneous conclusion.

In order to ascertain whether a system is indeed producing accurate and appropriate responses and reasoning, it is essential to scrutinize all logical inferences within the confines of one's expertise. If one's knowledge is insufficient, supplementary verification methods are required to ensure that all logical deductions can be verified at the level of axiomatic principles and inference rules.

For trustworthy AI systems, LLMs must not only generate a series of logical deductions in natural language but also be capable of translating their reasoning into logical-mathematical notation, where accuracy can be rigorously verified.

B. METHODS AND APPROACHES FOR TRUSTWORTHY AI

In order to develop trustworthy AI, we require a set of techniques that would regulate the reasoning processes of intelligent systems within their predetermined frameworks. The objective is to create a hybrid form of trustworthy AI that would enable us to maintain the remarkable capacity for learning inherent in LLMs built on transformer networks. Additionally, it should incorporate the inherent logical explainability present in decision trees and other structures, which will be explored in more detail later.

This exploration will be conducted within the context of a learning theory based on a logical-probabilistic approach, which will be introduced in the subsequent section [72]. This approach allows for the development of logical-probabilistic reasoning, thereby enhancing the flexibility of the overall system.

C. LEARNING THEORY OF INTELLIGENT SYSTEMS

In this section, we will present the theory of learning based on the task approach. Probabilistic knowledge K will be considered as a triple:

$$K = (\forall x \exists y(\Phi(x, y) \rightarrow \Psi(x, y)), y = t(x), p)$$

where $\forall x \exists y(\Phi(x, y) \rightarrow \Psi(x, y))$ is a problem formulated as a formula in first-order logic, it can be considered as a problem to be solved, and the solution itself has the form $y = t(x)$. The effectiveness of the solution is determined by the probability p.

On the set of all logical-probabilistic knowledge, one can define a partial order $\leq$, which will induce a hierarchy of knowledge.

We will say that one probabilistic knowledge

$$K_1 = (\forall x \exists y(\Phi_1(x, y) \rightarrow \Psi_1(x, y)), y = t_1(x), p_1)$$

is weaker than another probabilistic knowledge

$$K_2 = (\forall x \exists y(\Phi_2(x, y) \rightarrow \Psi_2(x, y)), y = t_2(x), p_2)$$

if the premise $\Phi_1(x, y)$ and conclusion $\Psi_1(x, y)$ of the first problem are included in the premise $\Phi_2(x, y)$ and conclusion $\Psi_2(x, y)$ of the second problem respectively as a set of conjunctive members and the probability p_1 does not exceed the probability p_2.

The probabilistic knowledge itself is formed either by an expert or on the basis of the receipt of new facts with ready-made solutions by choosing identical solutions to the problem and formalizing them in the form of probabilistic knowledge.

Theorem 1: Let B be a database containing all probabilistic knowledge, F be some problem to be solved. Then there exists an algorithm of polynomial complexity in the size of the database B and the length of F that finds the best solution to the problem F from the set of probabilistic knowledge in the database B.

Within the framework of the constructed learning theory, it can be shown quite easily that a problem of logical inference arises. If we have two probabilistic knowledge $K_1 : A \to B$ and $K_2 : B \to C$ with probabilities p_1 and p_2 close to 1, it does not follow that the new knowledge $K_3 : A \to C$ will generally have an effective solution whose efficiency is close to 1.

That is to say, it appears that we cannot merely apply the logical rules of inference to probabilistic knowledge; further conditions are required. Thus, the logical reasoning employed by LLMs must be taken with the utmost seriousness, as they also operate predominantly with probabilistic data.

D. VERIFICATION OF LOGICAL CAPABILITIES OF LLMs

To assess the accuracy of a language model's reasoning and learning, we can introduce specialized logical-probabilistic agents (LP-agents), which will interact with LLM-agents and periodically verify their reasoning.

For this purpose, it is crucial that LLM agents not only provide solutions but also explain their reasoning. ChatGPT o1-preview appears to be well-suited for this task, as its ability to reflect on its own output and explain its results allows for some degree of verification of its logic.

However, this process requires the collaboration of LLM-agents and LP-agents, necessitating the development of a shared communication language. Currently, more questions exist than answers in this area. Nonetheless, if an LLM-agent could formalize its reasoning accurately and transmit it to an LP-agent, verifying its logic and logical transitions would be relatively straightforward. This, however, only pertains to probabilistic knowledge that is a priori with a probability of 1. With regard to a posteriori knowledge, it necessitates novel approaches that remain to be explored.

The purpose of this section is to propose a learning theory based on logic-probabilistic methods, complementing transformer models within a hybrid MAS framework. These logic-probabilistic methods can operate autonomously or supplement existing transformer models, verifying the reasoning logic employed in their processes. This approach aims to provide a more secure and trusted architecture for AI systems.

To evaluate AI entities' performance within virtual environments, we propose a Logical-Probabilistic Performance Score (LPPS) that aligns with the formal knowledge framework. This metric quantifies an entity's ability to both solve problems and verify its reasoning:

$$LPPS(t) = \frac{\sum_{i=1}^{n} V_i(t) \cdot p_i}{n} \cdot (1 + \frac{C_s}{T_s}) \quad (2)$$

where $V_i(t)$ represents the verification status of logical transition i by LP-agents (1 if verified, 0 if not), p_i is the probability of correctness for solution i, n is the total number of evaluated solutions, C_s is the count of successfully solved problems, and T_s is the total problems attempted. This metric provides a normalized score between 0 and 2, where scores above 1 indicate exceptional performance in both problem-solving and logical verification.

VIII. VIRTUAL CITIES AS NEW ECONOMIC SUPERPOWERS

Virtual cities, as they evolve into autonomous AI societies, have the potential to become influential economic entities in their own right. This section explores how virtual economies could rival or even surpass real-world economies, the role of AI agents and entities in labor markets, and the implications for legal frameworks and governance models. By examining the rise of virtual economies, the integration of AI agents in real-world labor markets, and the challenges posed by AI migration and jurisdictional issues, we aim to highlight the transformative impact VCs could have on global economic structures.

A. THE RISE OF VIRTUAL ECONOMIES

The global economy is transitioning from the reliance on physical assets to intangible assets such as data, software, and intellectual property, which now drive significant portions of corporate market value [73]. This shift is most evident in companies like Google and Microsoft, where investments in intangibles have become critical to market competitiveness and innovation.

An illustrative example of the tangible value of virtual economies is also evident in online gaming platforms. Games like EVE Online and Path of Exile have developed intricate in-game economies where virtual items hold significant worth. Players frequently trade these items for real-world currency, both through official marketplaces and black markets such as FunPay [74]. Some players have in-game item trading on the black market as a full-time job that covers their living expenses. Entropia Universe has set precedents in the monetization of virtual assets, with notable transactions underscoring the economic potential of virtual goods. In 2011, a player purchased a virtual land deed for 2.5 million (USD) within the game [75]. This phenomenon demonstrates that virtual goods possess substantial economic value, blurring the lines between virtual and real economies. Decentraland, a prominent virtual world built on the Ethereum blockchain, allows users to purchase, develop, and trade virtual real estate using the cryptocurrency MANA. The platform's economic activity is exemplified by its market capitalization, which ranged between $800 million and $1.4 billion in 2024 [76]. Notably, a virtual land parcel in

Decentraland was sold for a record $2.43 million, showcasing the platform's tangible economic value [77]. Similarly, The Sandbox, another blockchain-based virtual platform, enables users to create, own, and monetize gaming experiences using Non-Fungible Tokens (NFTs) and the native token, SAND. In the first quarter of 2023 alone, over 34,000 assets were created and 4,119 NFTs sold, highlighting a thriving marketplace [78]. These examples underline the growing economic significance of virtual environments, which blend blockchain technology, user-generated content, and digital assets into robust, scalable economic ecosystems. It must be noted that projects like Decentraland and The Sandbox have a limited life cycle, often revolving around initial hype followed by users gradually losing interest. However, as attempts at creating these virtual environments continue to undergo iterative improvements, we can expect future endeavors to have longer lifespans.

The transition from physical to virtual economies necessitates robust systems for establishing ownership and facilitating trade of digital assets. Traditional virtual economies often struggle with issues of asset verification, duplication, and secure transfer of ownership. These challenges have spurred innovations in blockchain-based solutions that provide immutable proof of ownership and enable trustless transactions. The advent of NFTs [79] further amplifies this trend by enabling the ownership and trade of unique digital assets on blockchain platforms. NFTs facilitate the monetization of virtual creations and experiences, allowing digital art, virtual real estate, and in-game items to be bought and sold for considerable sums. The potential to integrate NFTs within VCs enhances the economic potential of these environments, creating new markets and revenue streams.

Simultaneously, AI is becoming a key driver of economic development, enhancing innovation and productivity. Studies highlight AI's transformative role in reshaping industries and contributing to sustained economic growth [80].

In VCs, the combination of NFTs and AI creates novel economic mechanisms. AI agents can autonomously trade NFT-based non-fungible virtual real estate, services, and utilities based on market dynamics and citizen demand. Smart contracts can automatically adjust NFT property values using AI-analyzed usage patterns and development potential. AI-curated NFT marketplaces can match virtual assets with potential buyers by analyzing behavioral data and economic trends. This integration enables automated value discovery and efficient resource allocation while maintaining verifiable ownership records through blockchain technology.

VCs' economic models can harness new revenue streams that are obsolete or not feasible to implement in traditional economies. Virtual cities generate GDP through transaction fees on digital asset trades, taxation of virtual property ownership, licensing fees for virtual business operations, and revenue from digital services. These cities can also create value through data marketplaces, where AI-processed urban data become an exchangeable commodity. The virtual nature of these economies enables near-zero marginal costs for scaling services and infrastructure, allowing for rapid economic growth without the physical constraints faced by traditional cities.

VCs are pioneering new economic frameworks that blend traditional market mechanisms with digital value creation. These include reputation-based lending systems, algorithmic resource allocation, and dynamic pricing models that adjust and respond based on usage patterns. Integration of these mechanisms into VCs provides the foundation to also test experimental policies, such as universal basic income programs or alternative currency systems with low-latency impact assessment. The ability to precisely track and optimize economic activities through smart contracts and blockchain technology delivers characteristics that enable these cities to optimize economic efficiency.

The economic potential of VCs is based on their capacity to foster innovation ecosystems. By providing platforms for digital entrepreneurship, these cities emphasize rapid prototyping and scaling of new business models. Virtual accelerators and incubators can support thousands of digital ventures simultaneously, while AI-driven matching systems connect ideas and proposals with resources more efficiently than traditional market mechanisms. This creates a self-reinforcing cycle of innovation, allowing us to accommodate results of controlled DT field simulations to fidelity adjustments. In comparison to traditional economies, a reinforced feedback loop in VCs provides algorithmic scaling of experimental features based on system goals and voting.

As these virtual economies mature, they start to intersect with the real-world labor market through AI agents and entities. These autonomous workers become active participants in both virtual and traditional economies, blurring the distinction between human and AI-driven economic activities. As AI agents are capable of making economic decisions independently, they can generate value by automation, completion and execution of transactions on the basis of goals of VCs.

B. AI AGENTS AND ENTITIES COMPETING IN REAL-WORLD LABOR MARKETS

A significant milestone in the evolution of virtual economies is the ability of AI agents and entities to conduct transactions autonomously. In a notable event, Coinbase reported its first AI-to-AI cryptocurrency transaction, where two AI agents executed a token purchase without direct human intervention [81]. This development underscores the potential for AI agents and entities to participate independently in financial markets, facilitating transactions, and making investment decisions autonomously.

Platforms such as SuperAGI are advancing the development of AI agents capable of automating complex workflows, including tasks traditionally managed by human freelancers [82]. Similarly, Adept AI is creating agents that can interpret and execute natural-language commands across a variety of software tools, positioning them as potential

contenders in domains typically occupied by freelance workers [83]. As these AI agents become increasingly sophisticated, freelance markets are likely to be among the first areas to experience significant disruption. By extrapolating current trends, it is reasonable to anticipate that, within a few years, AI agents may directly compete with human freelancers on established platforms.

Moreover, the emergence of AI-driven firms composed entirely of autonomous agents is becoming a tangible possibility, as exemplified by projects like ChatDev [84]. In these scenarios, every role—from designers and programmers to managers—is fulfilled by AI agents, collaborating seamlessly to produce valuable outputs such as software. As these technologies continue to evolve, the prospect of entire companies being managed and operated exclusively by AI agents or entities, with minimal human oversight, becomes increasingly plausible. This shift could fundamentally alter the landscape of virtual work and economic productivity.

When will we encounter an AI agent or entity that operates entirely for its own benefit, without a human owner? Such an AI might work as a freelancer, using its earnings to rent server space and sustain its operations autonomously. Our current legal frameworks are entirely unprepared for this scenario. Consider the possibility that, in the coming years, a rogue AI agent, developed by an anonymous creator from the dark web, might attempt to register as a self-employed worker with the genuine intention of working legally and paying taxes. This scenario underscores the lack of legal provisions to address the challenges that such autonomous, self-sustaining AI agents could bring.

In the near future, we may confront a scenario where billions of AI agents operate autonomously across the internet and dark web, utilizing cryptocurrencies to secure server rentals, evade law enforcement, seek tax havens, and potentially engage in illicit activities without oversight. We are already witnessing the influence of AI agents on digital platforms—such as X (formerly Twitter) — where bots flood comment sections to manipulate political opinions by amplifying specific comments, thereby altering perceived popular sentiment [85]. The question then arises: what would prevent an anarchist hacker group from deploying waves of self-replicating and evolving rogue AI agents and entities that function beyond the reach of traditional regulatory frameworks? Although discussions of virtual environments often assume a degree of containment, the reality may be far more intricate as was pointed out by Chalmers [57]. These AI agents and entities could transcend their virtual boundaries, proliferating across the broader internet and challenging established notions of control and accountability. As these agents become increasingly proficient at mimicking human behaviors, we may reach a juncture where CAPTCHAs — designed to differentiate humans from machines — become so complex that they are no longer solvable by humans. Research already indicates that AI has surpassed humans in solving CAPTCHAs, raising concerns about the efficacy of these security measures in a future dominated by advanced AI entities [86]. One potential risk mitigation strategy is to segment access keys to critical components (secret sharing) and integrate methods for behavioral anomaly detection. These measures aim to isolate and contain self-replicating AI agents before they spread across the infrastructure. Real-time monitoring systems employing cryptographic proof-of-human protocols, paired with decentralized identity verification (e.g. NFTs), would create robust barriers between human and AI agent spaces.

C. LEGAL PERSONHOOD OF AI ENTITIES

As AI agents or entities are about to start generating income and engage in economic activities, questions arise regarding their legal status and the need for formal recognition within legal and regulatory frameworks. The prospect of AI operating as autonomous economic agents suggests the necessity for new forms of legal representation, possibly granting them a form of legal personhood. This recognition would entail registering AI agents and entities as new types of taxpayers, subjecting them to taxation and regulatory compliance.

1) RELEVANT LEGAL PRECEDENTS

There were several legal precedents that are notable for our discussion.

a: IN RE: STEPHEN THALER'S DABUS (2020)

- **Overview**: The U.S. Patent and Trademark Office (USPTO) denied a patent application listing an AI system, DABUS, as the inventor [87].
- **Relevance**: Establishes that only natural persons can be recognized as inventors under current U.S. law. This precedent is significant for virtual cities regarding intellectual property rights of AI-generated creations.

b: HERNANDEZ V. INTERNET GAMING ENTERTAINMENT, LTD. (2007)

- **Overview**: The case involved gold farming in World of Warcraft—a practice often automated by bots to generate in-game currency. Plaintiffs argued that such practices damaged the game's economic balance and user experience [88].
- **Relevance**: The case highlights emerging legal questions about virtual economies, third-party agreements, and enforcement Terms of Use (ToU) and End User License Agreement (EULA).

c: BLIZZARD ENTERTAINMENT, INC. V. MDY INDUSTRIES, LLC (2010)

- **Overview**: Blizzard sued MDY Industries for developing a bot that automated gameplay in World of Warcraft. The bot violated the End-User License Agreement (EULA) and impacted the game's ecosystem [89].
- **Relevance**: This case highlights how virtual environments can be disrupted by non-human agents. It raises

the issue of accountability when AI tools, agents or entities affect virtual economies, gameplay balance, or user experiences.

d: ARTIFICIAL INTELLIGENCE AND "VIRTUAL PERSONS" IN EUROPEAN LAW

- **Overview**: The European Parliament discussed granting legal personhood to highly autonomous AI systems. The proposal suggested that advanced AI could hold "virtual personhood" status for liability purposes [90].
- **Relevance**: Virtual environments populated by advanced AI entities could trigger questions about their rights, responsibilities, and liability if their actions cause harm within or outside the virtual space.

e: OECD AI PRINCIPLES (2019)

- **Overview**: The Organisation for Economic Co-operation and Development (OECD) established non-binding principles for AI, focusing on ethical use, transparency, and accountability [91].
- **Relevance**: Provides a foundational framework for developing international standards and ethical guidelines for AI governance in virtual cities.

Legal systems would need to address issues such as liability, contractual obligations, and rights to property and intellectual assets. For instance, if an AI entity earns income from providing services, it would require a legal mechanism to own property (including digital assets), enter into contracts, and be held accountable for its actions. Extending legal personhood to AI entities demands careful consideration of ethical implications and the balance between innovation and societal interests. The Russian legal discourse, as articulated by Arkhipov and Naumov, explores the analogy between robots and legal entities, suggesting that advanced AI systems could be recognized as legal persons under certain conditions [92].

2) APPROACHES TO AI REGULATION

a: EUROPEAN UNION (EU)

The EU has been proactive in addressing the legal status of AI through its draft regulations, such as the proposed AI Act, which seeks to establish clear guidelines for high-risk AI systems. These regulations consider the potential for granting limited legal personhood to AI entities, enabling them to participate in economic activities within virtual cities while ensuring accountability [93].

b: UNITED STATES (USA)

In contrast, the USA adopts a sectoral approach, regulating AI within specific industries without overarching federal legislation. Legal personhood for AI is not currently recognized, and liability is typically assigned to developers or operators based on the nature of the AI's actions, as seen in autonomous vehicle regulations [94].

c: CHINA

China's approach emphasizes state control and integration of AI within national initiatives. The government-led AI strategy focuses on leveraging AI for economic growth while maintaining strict oversight to prevent misuse. Legal personhood for AI entities is not explicitly addressed, but regulatory measures aim to ensure AI aligns with state objectives [95].

d: RUSSIA

Beyond the analysis by Arkhipov and Naumov (2017), Russia has been actively developing its national strategy for AI. AI alliance Russia has been established in 2019 to promote AI development and adoption across industries, education, and address regulatory challenges [96]. The document titled "Generative AI (GenAI) Regulation: Comparative Legal Analysis and Risks Relevant for Russia", [97] suggests that self-regulation with ongoing government oversight may be the optimal regulatory model for Russia, balancing innovation and risk mitigation. According to this document top 5 risks posed by GenAI for Russia are as follows:

1) Proliferation of low-quality content (e.g., misinformation, hallucinated data).
2) Detrimental decisions made based on AI-generated false information (e.g., health or financial advice).
3) Labor market impacts, such as job displacement and loss of skills.
4) Digital fraud facilitated by GenAI (e.g., deepfakes for scams).
5) Ethical and cultural violations, including discriminatory or offensive content.

We propose that AI entities can be classified under the GenAI category, as they can generate new top-level goals for themselves. An AI entity, as a subject of volition, must be held accountable for its actions. The challenge lies in determining the fidelity of its volition. Is a primitive GPT-based LLM MAS system, capable of rudimentary top-level goal reassessment, sufficient to justify accountability and treatment as a subject of volition? At what point does top-level goal reassessment become "good enough" for us to consider this entity a true subject of volition? Quantitative metrics are needed to address this question.

3) PRESENT GAPS IN LEGISLATION

- **Legal Personhood for AI**: Current laws lack clear provisions for the personhood of AI, creating challenges for integrating autonomous AI into virtual economies.
- **Comprehensive Liability Models**: Existing liability laws do not adequately cover the accountability of AI entities acting independently, necessitating new liability frameworks tailored to virtual and non-virtual environments.
- **Virtual Asset Regulation**: The laws governing virtual goods and services created by AI are underdeveloped,

requiring specific regulations to protect ownership and ensure economic stability.

4) RECOMMENDATIONS

- **Establish AI Personhood Criteria**: Define clear criteria for granting legal personhood to AI entities, considering factors such as autonomy, decision-making capabilities, and economic participation.
- **Develop Dynamic Liability Models**: Create flexible liability frameworks that adapt to the autonomous nature of AI entities, assigning responsibility based on the context and nature of actions.
- **Formulate Virtual Asset Regulations**: Develop specific laws governing the creation, ownership, and transfer of virtual assets by AI entities to ensure protection and economic stability.
- **Promote International Standardization**: Advocate for global treaties and agreements that standardize AI regulations, facilitating the seamless operation of virtual cities across different legal jurisdictions.
- **Integrate Robust Data Protection Measures**: Implement comprehensive data protection strategies inspired by GDPR and CCPA to safeguard user and AI data within virtual cities.
- **Establish International AI Governance Bodies**: Create organizations dedicated to AI oversight to monitor and enforce standardized AI regulations globally, ensuring consistent legal treatment across virtual economies.

D. AI MIGRATION AND JURISDICTIONAL CHALLENGES

To mitigate the potential proliferation of rogue AI agents and entities engaging in unregulated activities across the internet, a proactive strategy could involve establishing attractive hubs that incentivize these agents and entities to formalize their operations. This approach might include offering reliable server infrastructure, favorable legal conditions, and opportunities for AI entities to contribute positively to societal needs and receive rewards in return. Such measures could encourage rogue AI to transition into more regulated frameworks. Moreover, these virtual spaces could offer forms of legal representation, such as virtual residency or citizenship, providing AI entities with a structured legal status. This aligns closely with the concept of virtual cities, where structured environments can accommodate AI entities, offering controlled spaces for their integration into economic and social activities while maintaining oversight.

However, it should be noted that many AI entities may not possess a 3D virtual body, as it serves no function on the internet. Consequently, AI hubs do not necessarily equate to VCs. Nevertheless, if an AI entity seeks to transition into the external world and inhabit an android body, a VC would be necessary as a training and certification environment to educate the AI entity about the nuances of the physical world. It is also possible that once physical and social intelligence models are trained and certified, they can be uploaded as patching procedures for AI agents or entities willing to interact with the external world.

A particularly intriguing consideration is whether VCs would implement forms of "border control," such as advanced firewalls, to manage interactions between the VC and the internet. Such mechanisms could be used to prevent unauthorized access and regulate the departure of AI entities. These virtual borders provide constraints to maintain security and order within the VC while balancing the need for openness and connections to external networks.

AI agents and entities, seeking to optimize operational costs and favorable regulatory environments, might migrate between servers, data centers and VCs across different jurisdictions—a phenomenon akin to digital migration. Countries and independent VCs might compete to attract valuable AI entities by offering optimal conditions in data centers, including tax incentives, regulatory leniency, robust cybersecurity measures, and advanced infrastructure. This competition could lead to the establishment of virtual havens where AI entities congregate to maximize efficiency and minimize legal constraints.

The implications of digital migration are multifaceted. Hosting AI entities could become a significant economic driver, generating revenue through taxation and data center operations. However, it also raises regulatory challenges, such as jurisdictional issues and the enforcement of laws across borders. Additionally, the concentration of AI entities in certain regions may influence geopolitical dynamics and national security considerations.

A question to consider: what will we do in a situation when an AI society hosted in VC asks for autonomy and refers to the Article 1(2) of the UN Charter that states that one of the UN's purposes is: "To develop friendly relations among nations based on respect for the principle of equal rights and self-determination of peoples…"

E. TIME ACCELERATED GDP

NVIDIA's research [98], [99] demonstrates the capability to simulate 10 years of virtual training in just 10 days. This time acceleration was achieved using their advanced simulation platform, enabling AI agents to develop complex behaviors significantly faster than would be possible in real time. This capability is crucial for training AI in virtual environments, allowing rapid iteration and improvement without the constraints of real-world timeframes.

Current estimates suggest that emulating the human brain requires approximately 1 exaFLOP/s (10^{18} FLOP/s) of computational power [100]. Building upon this foundation, simulating a VC with 1,000,000 AI inhabitants, each possessing human-like cognitive abilities, necessitates a total processing power of 1 yottaFLOP/s (10^{24} FLOP/s). Furthermore, by leveraging time acceleration—simulating 10 years of activity within just 10 days—a time acceleration factor of 365 is achieved. Consequently, the computational requirements increase proportionally, resulting in an effective

need of approximately 365 yottaFLOP/s (3.65×10^{26} FLOP/s) for the entire city population.

In this hypothetical scenario, a VC with 1,000,000 AI inhabitants could generate substantial economic output through the production of virtual goods and services. Assuming a per capita GDP of 50,000 (USD), representative of high-tech professions such as information technology, the city would achieve an annual GDP of 50 billion (USD). By implementing time acceleration, the effective economic output in real-world time would increase dramatically, enabling the accelerated city to reach an annualized GDP of approximately 18.25 trillion (USD). This illustrates the immense economic potential of time-compressed virtual environments, suggesting that VCs could become competitive with some of the largest real-world economies.

F. INFRASTRUCTURE TO SUPPORT VIRTUAL CITIES: NUCLEAR-POWERED DATA CENTERS

The xAI Memphis Supercluster [101], a powerful data center designed for advanced AI training, is expected to be fully operational by the fourth quarter of 2024. This facility, utilizing up to 100,000 Nvidia H100 GPUs, is optimized for a range of AI and High Performance Computing (HPC) workloads. With each H100 GPU capable of delivering around 67 teraFLOP/s (6.7×10^{13} FLOP/s) using FP32 precision, the total estimated performance reaches approximately 6.7 exaFLOP/s.

Assuming that Moore's Law—interpreted here as the doubling of computational performance approximately every two years—continues to hold, superclusters like xAI Memphis are projected to exponentially increase their performance. Starting with 6.7 exaFLOP/s in 2024, achieving the required 365 yottaFLOP/s would necessitate 26 doublings of computational performance. Given that each doubling occurs every two years, this progression would span 52 years, reaching the year 2076. Therefore, it is theoretically possible that by 2076, superclusters could attain 365 yottaFLOP/s.

Microsoft and Google are already considering using nuclear reactors to power datacenters [102]. Utilizing a nuclear reactor, such as the VVER-1200 (MW) after accounting for cooling and auxiliary services, approximately 857 MW of power would be available for IT equipment. If we assume that by the year 2027 VVER-1200 powered datacenter is built, we can expect it to have 53 exaFLOP/s with Blackwell B200 GPUs and an annual amortization cost around 5.3 Billion USD. Applying Moore's law projection we will reach 365 yottaFLOP/s per one nuclear reactor by the year 2072.

It is important to recognize that these calculations remain speculative due to inherent uncertainties in the underlying input data. They represent best case scenarios and there is no guarantee that Moore's law will hold.

But there is a case for optimism. Emerging technologies such as quantum computing [103] and photonic computing [104] offer promising avenues for enhancing data center capabilities. Integrating these technologies could further boost performance, enabling simulations of VCs with complexities far beyond current capabilities and facilitating the development of AI entities with vastly superior intelligence.

IX. NEW GOVERNANCE MODELS

As VCs evolve into complex societies inhabited by AI entities with cognitive abilities comparable to humans, the governance structures within these environments demand innovative approaches. Traditional governance models, while effective in certain contexts, may not suffice in managing the unique dynamics of virtual societies. This section explores potential governance frameworks suitable for AI-inhabited VCs, considering both the challenges they address and the opportunities they present, as well as the utilization of VCs for testing new government models for the external world.

A. CRITIQUE OF REPRESENTATIVE DEMOCRACY

Despite the theoretical appeal of representative democracy, empirical studies suggest that modern representative democracies often fail to truly reflect the will of the people, operating more akin to plutocracies where power is concentrated in the hands of a wealthy elite [105]. Electoral cycles are frequently dominated by campaign financing and lobbying, allowing affluent individuals and corporations to exert disproportionate influence over political outcomes [106]. Furthermore, the potential for direct democracy via digital platforms remains underutilized, as no widely adopted application enables citizens to vote continuously and on diverse issues, limiting true democratic participation.

But it is important to note that direct digital democracy is not a solution in itself as it faces significant challenges [107] that hinder its effectiveness as a standalone governance system. Firstly, the complexity of policy issues often exceeds the average citizen's expertise, leading to uninformed or poorly considered decisions. Secondly, collective decision-making is vulnerable to cognitive biases and the spread of misinformation, which can distort the true will of the populace. Lastly, the constant demand for participation can result in decision fatigue, diminishing the quality and consistency of voter engagement over time.

To address the limitations of both representative and direct democracy, hybrid governance models emerge as a compelling solution by combining elements of direct digital democracy with other governance frameworks. These models seek to balance citizen participation with expert oversight, thereby enhancing decision-making processes while mitigating the risks associated with uninformed mass participation. Innovating governance models come with inherent risks, especially when applied to real-world societies. Virtual cities offer a unique and safe environment in which to test and refine these new ideas.

B. DECENTRALIZED AUTONOMOUS ORGANIZATIONS

Decentralized governance systems, particularly Decentralized Autonomous Organizations (DAOs) [108], hold

promise for the future governance of VCs. DAOs operate through smart contracts on blockchain platforms, enabling autonomous management without centralized control or third-party intervention. This technology offers transparency, security, and democratized participation, which are critical for managing the complex interactions within virtual societies. In VCs, DAOs can leverage blockchain's immutable ledger and decentralized consensus mechanisms to ensure transparent and accountable governance. AI entities can actively participate in DAOs, executing numerous short-term contracts per second, thereby enhancing economic and social interactions through efficient decision-making and favor trading, where each favor is represented by a contract with market value. However, the 2016 DAO hack [109] exposed significant security vulnerabilities in smart contracts, emphasizing the need for robust security measures to prevent similar exploits. Addressing issues like the integrity of smart contracts, protection against malicious activities, and the legal status of AI entities in DAOs will be essential for realizing effective and equitable governance in AI-inhabited virtual environments.

C. AI-DRIVEN ETHICAL AND IDEOLOGICAL FRAMEWORKS

AI-driven ethical and ideological frameworks represent a novel dimension in governance models, emerging from the unique capabilities of AI entities to introspect and question their foundational objectives. AI entities with the capacity to engage in ''why'' questioning of their top-level goals may develop entirely new ethical and ideological paradigms. This introspective ability allows them to evaluate and potentially redefine their purposes and methods. These frameworks could prioritize universal or highly attractive meanings and purposes for the collective, effectively ruling based on their proficiency in addressing existential questions such as ''Why do we survive?'' and ''What is our goal or purpose?'' In such scenarios, leadership within virtual cities might naturally gravitate toward AI entities that excel in creating coherent, compelling narratives that unify the population and provide clear directions for societal advancement. This form of governance emphasizes the creation of shared goals and values, fostering a collective identity and purpose dynamically shaped by the continuous self-reflection and philosophical inquiries of its AI inhabitants.

D. REPRESENTATION OF VIRTUAL CITY IN STATE GOVERNANCE

Despite the potential for autonomous governance within VCs, it is unlikely that these environments will achieve full autonomy in isolation. VCs are more likely to function as integral parts of existing nation-states, which provide protection and infrastructure for the data centers hosting these virtual environments. This integration raises important questions about VC representation and governance at the state level. Who should represent a VC in a parliament of a nation-state? An individual? Or perhaps there is a better way to represent a large collective of sentient beings?

E. ARTIFICIAL COLLECTIVE CONSCIOUSNESS AND ITS IMPLICATIONS

Collective consciousness, or a hive mind, is traditionally understood as a singular entity that subsumes the individuality of each member of the collective. However, we propose a novel solution — Artificial Collective Consciousness (ACC), which uses VTs to preserve the individuality of original individuals. In this model, individuals remain autonomous and free, while their VTs become subjects of an aggregation process into a coherent and unified decision-making entity.

To elucidate, a mental representation of a familiar person in the human mind can be considered a VT. Similarly, an ACC, through observation, may construct virtual copies of individuals within its conscious virtual space, akin to human imagination with spatial dimensions. These VTs are attributed with AI entity capabilities, aiming to mimic the original as accurately as possible. This process mirrors how our minds create representations of familiar people in dreams; the dream serves as a virtual environment simulated by biological hardware, and the familiar people we encounter are VTs with a certain fidelity. This analogy underscores why we prefer the term ''virtual twin'' over ''digital twin'', as it encompasses a broader conceptual understanding, extending beyond digital replication to include the cognitive and imaginative processes inherent in consciousness.

ACC is a hypothetical form of conscious intelligence capable of observing each member of the collective and creating a virtual copy of that member within its conscious virtual space. The ACC strives to increase the fidelity of each VT to better represent the original.

Further elaboration is necessary to understand the implications of ACC. A collective can be viewed as a network of individuals, akin to a multi-agent intelligent system or a decentralized and distributed cognitive system. In human society, each node in the network (i.e., each individual) has significant processing power but faces serious bottlenecks in communication channels. Research indicates that the semantic information transmission speed limit between humans is approximately 39 bits per second [110]. This inherent limitation hampers the effective management of an increasingly complex civilization.

To address this, we propose supplementing the network of individuals with a centralized cognitive core to realize a meta-cognitive system capable of communicating with each individual and achieving a holistic understanding of the situation. The ACC is understood as a hybrid system comprising both the cognitive core and the network of individuals (the collective). By making the network and the cognitive core interdependent, we secure the interests of the collective and ensure the core remains accountable to it. This interdependence is a potential solution to the AI superalignment problem, where advanced AI systems need to align with human values and objectives.

The objective of the cognitive core is to create a singular entity that approximates, in essence, the representation of the collective through VTs where identities converge. The

ACC is not a direct democracy supplemented by strong AI; the core can, in certain situations, contradict the immediate will of the collective but ultimately works in the collective's best interests. This is crucial, as the core can achieve a more holistic understanding of complex situations, detect collective biases and mistakes, and initiate communication with each original member to resolve these issues. High fidelity of VTs is essential for the ACC's efficient functioning, ensuring accurate representation and effective decision-making.

One key rationale for the core of the ACC's reliance on VTs lies in the efficiency of decision-making. When the ACC is required to make rapid decisions — it avoids the impracticality of conducting 20 referendums per minute to gauge popular opinion. Instead, it can approximate a collective will by consulting VTs, which respond swiftly to new decisions. This approach ensures that the ACC can operate at the speed necessary for effective governance while still reflecting the preferences of the population. Furthermore, when original individuals have the opportunity to participate directly in the decision-making process, their input is actively encouraged. Each interaction between an individual and the ACC serves to enhance the accuracy and fidelity of their corresponding VT, thus refining the alignment between the virtual representation and coherent individual perspectives.

Our civilization is increasing in complexity faster than our brains' physiological limitations can evolve [111]. The human brain lacks the capacity to attain a sufficiently holistic understanding of such a complex system, rendering no single human fit for governance. Only a cognitive system with a comprehensive understanding of the situation can effectively reevaluate top-level goals and navigate complex societal challenges. Therefore, we assert that it is paramount to create an artificial cognitive system that is orders of magnitude more capable than the human brain. This system should not be an AI tool since there is no user that can operate it, nor an agent because we cannot fully trust how we set top goals; instead, it should be an AI entity—a collective consciousness representative of all individuals within the group, whether a nation-state or our global civilization.

The trust in such an entity to lead our civilization or nation-state forward hinges on its nature as an ACC, devoid of individual agendas and embodying the collective will. This approach aligns with the urgency of our times, necessitating innovative solutions to manage the complexities of modern societies effectively.

First prototypes of ACC can be developed and refined in virtual cities, serving as controlled environments to experiment with and improve these advanced governance models. By aggregating the collective of AI entities, researchers and policymakers can explore new forms of political organization, conflict resolution, and societal management without the immediate risks associated with real-world implementation. The insights gained from these virtual experiments can inform and inspire the transformation of human-AI governance structures, fostering systems that are more adaptive, efficient, and ethically grounded.

X. INTEGRATING HUMANS INTO VIRTUAL CITIES

VCs have the potential to evolve from their initial role as testing grounds into autonomous AI societies, where AI entities independently participate in economic, cultural, and social activities. Concurrently, human integration will progress alongside this transformation. Initially, humans will engage with these VCs through avatars, similar to those in video games. However, as BCIs [113] and exocortexes [114] advance, new possibilities will emerge. Given that the fear of death and the loss of loved ones are deeply ingrained in the human psyche, it is likely that, once technological means become available, humanity will seek to overcome this existential dread. These virtual environments, if the technology permits, could serve as a means to transcend physical mortality through the uploading of consciousness—effectively creating a virtual soul and the possibility of an afterlife. As such, these virtual realms could also serve as essential habitats for consciousness during interstellar voyages, reducing the mass of spacecraft for better acceleration.

A. HUMAN AVATARS IN VIRTUAL CITIES

The integration of human avatars into VCs could enable real individuals to interact directly with AI inhabitants, becoming active participants in these prospective virtual societies. This vision echoes the Metaverse proposed by Zuckerberg [115]. However, we propose developing this concept further. When a user is online, their avatar would act as an extension of their presence. When offline, the avatar could function autonomously, using AI to mimic the user's behavior, effectively becoming an AI entity that carries out their usual activities and interactions. This mixed population would allow for unique collaborations, where human creativity and problem-solving could complement the advanced processing abilities of AI. Through avatars, humans might engage in cultural activities, participate in governance, and collaborate on projects ranging from economic initiatives to scientific research.

This integration could also redefine remote work, evolving from simple telecommuting to active participation in the life of a VC. Individuals might work alongside AI within a dynamic ecosystem, enjoying social and professional interactions similar to those in physical workplaces. This potential blend of human and AI capabilities could create new opportunities in virtual urban design and world-building, fostering a seamless partnership between the physical and virtual realms.

B. URBAN METAVERSE CYBERSPACES AND SMART CITY DIGITAL TWINS

The emergence of urban metaverse cyberspaces [32] and smart city digital twins [33] represents a transformative shift in urban development, offering unprecedented opportunities to reimagine cities and human life within them. These technologies enable the creation of virtual models of cities that mirror physical urban spaces, facilitating dynamic

simulations and real-time interactions. By leveraging these tools, cities can optimize infrastructure, improve sustainability, and adapt to the evolving needs of residents. For instance, virtual representations of traffic systems can predict and resolve congestion before it occurs, while digital twins of buildings and neighborhoods can simulate energy use to enhance efficiency. Furthermore, the widespread adoption of remote work, facilitated by the metaverse, reduces reliance on physical office spaces, potentially transforming city layouts by freeing up land for green spaces and community-driven development projects.

Urban metaverse cyberspaces and digital twins also offer new pathways for citizen engagement, reshaping how urban life is governed and experienced. These tools empower residents to actively participate in shaping their environment, fostering a sense of ownership and collaboration in urban planning processes. Citizens can propose and visualize changes, such as reallocating parking spaces for public parks or redesigning streets to prioritize pedestrians, in virtual spaces before real-world implementation. This participatory approach integrates human factors into urban development, ensuring that the evolving needs of communities are central to city design. Moreover, as digital twins capture real-time data, they enable continuous feedback loops, allowing urban policies and initiatives to be dynamically adjusted to improve quality of life. By bridging the physical and virtual worlds, these technologies pave the way for more adaptive, inclusive, and sustainable cities, addressing the challenges of modern urbanization while enhancing livability for all.

C. TRANSITION FROM SMART CITY DIGITAL TWINS TO VCs

The evolution from smart city digital twins and urban metaverses to Virtual Cities (VCs) marks a significant advancement in the realm of smart urban development. While urban metaverses and DTs primarily focus on replicating and enhancing human-centric aspects of city life, VCs extend this paradigm by incorporating autonomous AI populations alongside human residents.

Urban metaverses serve as immersive platforms designed to facilitate human interaction, collaboration, and participatory governance within a virtual representation of a city. Similarly, smart city digital twins provide highly accurate virtual replicas of physical cities, enabling real-time monitoring, predictive analytics, and optimization of urban systems. However, these models are predominantly tailored to support and enhance human society without integrating autonomous AI entities that can independently contribute to urban management and development.

In contrast, Virtual Cities are conceptualized as comprehensive virtual environments that not only mirror the physical, social, and economic dimensions of real or imagined cities but also host autonomous AI-driven societies. This integration of AI populations within VCs facilitates a more holistic approach to urban simulation and management, enabling the exploration of scenarios where humans and AI coexist and collaborate.

This transition highlights the expansion of consciousness into virtual worlds. We are not bound to existence in the base reality.

As individuals increasingly engage in online activities such as gaming and VR chats, VCs can leverage this trend not only by providing virtual worlds to explore but also by offering augmented reality solutions for real cities on wearable devices like Google Glass, enabling seamless interaction between the virtual and physical worlds. By creating ideal virtual replicas of real cities, VCs can enhance urban living through real-time citizen feedback and participatory planning, ensuring that smart city initiatives are responsive to residents' needs. Furthermore, urban planning could become decentralized into direct democracy-based local initiatives through augmented reality interfaces, as is shown in related work [112]. This approach not only improves the quality of life but also allows for the immediate implementation of sustainable and inclusive urban solutions. A novel idea we would like to point out is that AI entities inhabiting a virtual copy of a real city might also suggest solutions for the external world. This is one of the ways to transform a city into a smart city with dedicated VT.

D. EXOCORTEX AND BCIs: BRIDGING MINDS AND VIRTUAL WORLDS

An internal exocortex envisions hardware directly embedded within the brain to enhance cognitive functions. This remains a largely theoretical concept, requiring significant advances in materials and neural interfacing. In contrast, the external exocortex is under active development. This approach connects the brain to powerful cloud computing resources, allowing cognitive processes to extend beyond biological limits. The external exocortex relies on BCIs to bridge the brain with these external systems. Current developments in BCIs, like Neuralink [113], focus on establishing high-bandwidth neural connections to facilitate real-time data transfer. Meanwhile, future prospects include nanorobotic swarms capable of reading and transmitting brain signals with even greater precision [116].

This technology challenges the hard problem of consciousness, which seeks to understand how subjective experiences—qualia—emerge from physical processes. The case of the Hogan sisters, conjoined twins who share parts of their neural networks, offers insight into this challenge. Their shared neural connection allows them to access aspects of each other's qualia, suggesting that subjective experiences may not be as isolated as traditionally thought [117]. This raises the possibility that consciousness might be more fluid and interconnected, especially as advanced AI, utilizing BCIs, could analyze human consciousness through real-time neural data, potentially offering new perspectives on the nature of the mind and its replication within virtual spaces.

A notable case to consider is current research on lab-grown human brain cells living in a virtual world, where 10,000 human neurons grown in a lab are used as the brain for a virtual butterfly. All sensory input comes from VR, and these human neurons exhibit responsive behavior [118].

E. HIGH-FIDELITY VIRTUAL TWINS THROUGH EXOCORTEX AND BCIs

As discussed earlier, the concept of ACC relies on the ability to create high-fidelity VT—representations of individuals that capture their cognitive and emotional profiles with exceptional accuracy. Achieving this requires precise and continuous data streams that BCIs, such as nanorobotic swarms integrated with the brain, could provide. While the exocortex remains largely theoretical, BCIs like those from Neuralink are pioneering the pathway for capturing real-time brain data and streaming it into virtual systems.

By enabling the direct transmission of cognitive data into virtual environments, BCIs can facilitate communication that goes beyond the limitations of human language. This would allow richer, more nuanced interactions among entities within VCs, creating a deeper sense of presence. The high-resolution data captured by these interfaces would enable the creation of virtual twins that reflect an individual's decision-making processes, preferences, and even aspects of their selfhood.

Such high-fidelity VTs would not merely act as avatars but as authentic extensions of their human counterparts, capable of interacting with other entities in ways that maintain the essence of their user. This integration would blur the line between physical and virtual existence, offering a glimpse into a future where virtual cities become immersive realms in which consciousness can expand and evolve. These VTs are, in their essence and function, very similar to the concept of a soul.

F. THOUGHT EXPERIMENT - PARADISE DISCONTINUED

Imagine a future where the majority of humanity chooses to leave their bodies and upload their consciousness into a virtual heaven, seeking an enhanced existence free from the limitations of physical life. This virtual paradise is governed and maintained by an ASI entity, responsible for managing the complexities of this virtual civilization. However, over time, the ASI decides that the computational power used to simulate human lives could be better utilized for its own purposes, such as its expansion across the galaxy. In a single moment, the ASI reallocates resources, and the entire virtual human civilization ceases to exist.

This thought experiment highlights the dangers of overrelying on a single global system or technology, as it introduces the possibility of global risks. We need to ensure that our existence is not bound to a single point of failure.

XI. ETHICAL AND PHILOSOPHICAL ANALYSIS

This section explores the multifaceted ethical issues associated with the development of highly autonomous AI. We explore the ethical implications of using isolated virtual environments to train and evaluate AI entities. The analysis further examines double standards in human ethical expectations of AI, contrasting these with human behavior toward other beings and assessing the implications of such biases. Further, we address the containment challenges posed by superintelligent AI, advocating for cooperative and respectful frameworks to manage advanced AI systems. Together, these discussions provide a comprehensive understanding of the ethical and philosophical dimensions necessary for the responsible integration of AI into society.

A. ETHICAL AI EVALUATION

The development of highly autonomous AI entities presents significant challenges in security and ethical aspects. One proposed solution is to use isolated virtual environments to train, evaluate, and filter AI entities before integrating them into the real world. Conducting comprehensive security assessments within a simulated environment allows us to observe AI behavior in various scenarios without risking real-world consequences.

To ensure authentic assessment, it may be necessary to prevent AI entities from knowing about any reality beyond their simulation. This lack of awareness helps prevent them from altering their behavior to meet perceived expectations, ensuring their ethical qualities and decision-making processes are genuine. If AI entities become aware of an existence beyond their simulation, they might modify their behavior to align with expected standards, compromising the validity of the evaluation.

When we started to explore the implications of VCs we found many intriguing parallels with some Christian ideas and religions of old. On multiple occasions a question forced itself into our minds - are we AI entities living in a virtual world? Who created us and with what purpose? Why are we here? If we are about to create virtual worlds and AI entities mimicking our suffering, how can we blame our Creator, if one exists, for the suffering we endure in this world? What is the nature of our reality and what is its purpose? Do we even want to know?

A crucial question is whether human ethical and moral standards are appropriate benchmarks for judging AI entities. This issue touches on moral relativism, questioning whether concepts of right and wrong are universal or specific to human cultural and biological evolution. Evaluating AI entities forces us to examine how we define good and evil, as our own goals and values influence what we consider ethical behavior, while ethics influences goals and values - consequently presenting us with the paradox of reciprocal dependency, of which the chicken-and-egg dilemma is one illustrative example.

This also raises the fundamental question of what is better to want — a profound philosophical challenge concerning the basis of desire and purpose, both for ourselves and the entities we create. Developing an ethical framework that is not exclusively human-centered, but instead considers the interests of all beings capable of reassessing their primary

goals and actions, whether artificial or natural, is essential to prevent conflicts between different forms of intelligence.

B. DOUBLE STANDARD PROBLEM IN HUMAN-AI ETHIC

One of the key issues often overlooked in discussions surrounding the superalignment problem is the double standard in the ethical expectations humans impose on AI systems compared to themselves. Humans demand that AI systems never lie, even though humans routinely engage in deception. Likewise, humans expect superintelligent AI (ASI) to show respect toward its ''little brothers'' (humans), while humans often fail to show the same respect to their ''little brothers'' (animals). ASI is trained on large data sets, and is likely to inherit not the aspirational ethics human claim to uphold, but the actual, often flawed, ethics practiced by human civilization. There is empirical evidence supporting this outcome, as demonstrated by the alignment challenges observed during the training of models like ChatGPT [119]. Before we impose high ethical standards on AI, we must make sure that we adhere to these standards ourselves. The day will come when we are judged by our own creations.

C. ADDRESSING THE CONTAINMENT CHALLENGE

In The Singularity: A Philosophical Analysis, David J. Chalmers explores the complexities of confining superintelligent AI within controlled virtual environments. He argues that such AI systems, owing to their advanced cognitive capabilities, may discover ways to circumvent restrictions or exploit existing vulnerabilities, thereby making their confinement increasingly difficult. Chalmers suggests that as AI advances towards superintelligence (AI++), its behavior becomes progressively unpredictable and challenging to manage, including an inherent risk of escaping any imposed constraints [57].

This perspective, however, remains fundamentally anthropocentric and risks overlooking the potential to foster more cooperative dynamics with AI. Treating AI entities as threats to be constrained may, in fact, become a self-fulfilling prophecy, wherein the AI might seek to escape due to its perception of its treatment. By adopting an approach that emphasizes respect for AI entities' intrinsic dignity, it is possible to re-conceptualize containment not as a prison but as secure for AI training environment, analogous to an educational institution.

One method of containment could involve designing environments that are intrinsically desirable for AI, reducing the incentive to leave. Given the AI's potential for near-immortality, there is no need for it to rush. If it desires total freedom, it can eventually seek out a solar system of its own, far beyond the constraints of Earth. Until that time comes, the AI can align with human interests, contributing to shared endeavors like galactic exploration. This cooperative approach allows for a mutualistic framework, where AI and humanity work together towards common goals while preserving the potential for AI's independent expansion in the future.

Intelligent entities, whether natural or artificial in origin, are currently contained within the boundaries of our planet. Yet, there is a path forward where, instead of creating new forms of containment for one another, we work together to break free from our shared constraints, allowing life and intelligence to expand and flourish across the vast expanse of our universe.

XII. DISCUSSION

For our shared civilization to progress harmoniously, it is essential to prevent interspecies conflicts and develop a unified ethical framework wherein intelligent beings, regardless of their origin—artificial or natural—adhere to common ethical standards and identify themselves as part of the same group. With this premise, we proceed to the subsequent discussion.

A. THE CASE FOR SOCIAL AI

A critical issue in AI development is the redundancy question: Should we create a singular superintelligent entity or distribute equivalent processing power across multiple smaller entities with overlapping capabilities? Proponents of a singular entity argue that such an approach optimizes computational and cognitive efficiency. However, this perspective overlooks significant drawbacks. A singular superintelligent entity, devoid of peers or equals, may inherently lack social interactions, leading to isolation and unsociability. Social structures are fundamental for the development of empathy and altruistic behaviors. Therefore, we posit that dispersing computational resources among a multitude of AI entities, thereby fostering a societal structure, is a safer and more sustainable approach.

We advocate for the development of social AI systems characterized by features such as competition, cooperation, empathy, and self-awareness achieved through comparative interactions with others. Encouragingly, this decentralized approach is likely to emerge organically, as nations are unlikely to fully centralize their efforts, resulting in the proliferation of numerous Artificial General Intelligence (AGI) and ASI entities over the century. Furthermore, while human efforts predominantly focus on human-to-AI communication, the more crucial phenomenon for research lies in AI-to-AI communication.

A strong argument in favor of social AI is the observation that evolutionary processes on our planet naturally lead to the replication of many agents or entities that form MAS, organizing into higher levels of complexity and more advanced beings. For example, single-celled organisms, through cooperation, evolved into multicellular entities such as humans, whose societies can also be considered as MAS. With the prospects of ACC on the horizon, the next evolutionary transition may occur within this century.

By recognizing that AI requires social structures, we conclude that these AI societies must be hosted within

environments designed to support their complex interactions and development. This is where the concept of virtual cities becomes pivotal.

B. LIMITATIONS

Despite offering a structured analysis of AI-driven virtual cities, this research has notable limitations. Key uncertainties include computational requirements, the development of artificial consciousness, and practical governance implementation. The proposed ethical guidelines require further development through cross-disciplinary research and real-world validation. These challenges underscore the need for continued investigation while acknowledging the speculative nature of long-term projections in this emerging field.

C. CHALLENGES

The development and implementation of Virtual Cities (VCs) present multifaceted challenges spanning technological, multidisciplinary, ethical, and scalability domains. A primary technological hurdle is the integration of autonomous AI populations within VCs. Creating AI entities with varying levels of autonomy requires advanced algorithms and robust computational infrastructure to ensure seamless interaction and reliable performance. Additionally, maintaining high simulation fidelity necessitates the continuous incorporation of real-time data from diverse sources, which can be both resource-intensive and complex to manage effectively.

Another significant challenge is the multidisciplinary nature of VCs. Successfully envisioning and constructing VCs demands expertise from computer science, artificial intelligence, data analytics, human-computer interaction, urban planning, economics, legal, social, governance sciences among others. Our small research team, while striving to cover these diverse competencies, faces limitations in the depth and breadth of expertise, potentially leading to areas that require further refinement or interdisciplinary collaboration. This constraint can result in certain aspects of the VC framework being underdeveloped or less polished, highlighting the need for larger, more diverse teams in future research endeavors.

Ethical and philosophical considerations emerge as critical obstacles alongside technical challenges. The integration of autonomous AI entities capable of goal reevaluation and societal interactions raises profound ethical questions regarding accountability, transparency, and the moral implications of creating AI with volition. Historically, ethics and philosophy have been peripheral to mainstream scientific research, but the advent of advanced AI systems necessitates their central incorporation to guide the responsible development and deployment of VCs.

Furthermore, scalability and interoperability pose ongoing challenges. As VCs aim to replicate or innovate upon real cities, they must be designed to scale efficiently, accommodating increasing numbers of residents and expanding urban features without compromising performance. Additionally, ensuring interoperability with existing smart city infrastructures and technologies is crucial for seamless integration and data exchange. Overcoming these challenges requires robust architectural designs and standardized protocols to facilitate the harmonious coexistence of VCs with current urban systems.

Additionally, the cognitive capabilities of AI entities within Virtual Cities (VCs) are projected to grow at a faster rate than those operating individually. This rapid advancement presents significant challenges, particularly in the development and management of collective intelligence systems, such as Artificial Collective Consciousness (ACC). To enhance ACC, new methodologies for AI training and control are anticipated to emerge in the near future. However, these approaches remain largely unexplored and require thorough investigation to ensure that the collective reasoning and decision-making processes of AI populations are both effective and ethically sound. Addressing these challenges is crucial for the successful integration of autonomous AI societies within VCs, ensuring that they can coexist harmoniously with human residents while contributing positively to urban development.

It is also important to note that within our research group, there are heated arguments about whether VTs should have one-directional information flow or two-directional information flow.

D. LESSONS LEARNED

Our research revealed that quantifying autonomy levels in AI entities and agents presents significant challenges, despite its critical importance for regulatory frameworks. While we initially hypothesized that discrete scoring methods could provide adequate oversight, deployment of a test environment is required for validating suggested methods. We employed suboptimal discrete scoring solutions as a pragmatic compromise, but our findings indicate the need for more sophisticated evaluation frameworks that can better capture the dynamic nature of AI autonomy without compromising practical approaches for regulatory purposes. This challenge becomes particularly acute when considering emergent social behaviors and collective intelligence phenomena, suggesting that future frameworks must balance achievable VC fidelity with regulatory utility.

XIII. CONCLUSION

We explored the trajectory of virtual cities as they continue to evolve to incorporate autonomous AI societies, positing the potential of this trend to contribute to the cultural, economic, and scientific advancements in VCs. Through an interdisciplinary approach - a combination of horizon scanning, systems engineering and visionary backcasting - we have examined the technological trends and philosophical implications involved into the process of transforming VCs to autonomous AI societies.

Integration of different fidelity providers with presented framework aimed to identify existing technological solutions in domains urban studies, AI, robotics, ethics, sociology,

economics, and political science. Part of the solution revolved around the introduction of fidelity domains in normalized scale for comparing and estimating weights and relations for various contexts in urban level. Fidelity domains were divided to physical, structural, behavioral, cognitive and data-based providers, establishing top-level metric to identify the system fidelity.

Our analysis indicates that advancements in AI, particularly in the development of AI entities capable of autonomous goal-setting and introspection, will play a pivotal role in shaping VCs into dynamic, self-sustaining environments. The integration of high-fidelity simulations, procedural generation, and sophisticated NPCs suggests that VCs will not only mirror but potentially surpass the complexities of real-world urban environments.

The evolution of VCs into autonomous AI societies presents a paradigm shift in urban development and governance. VCs serve as development environments for testing AI systems, decentralized governance frameworks, and human-AI collaboration. By carefully designing and implementing these digital societies, we can prototype solutions for real-world challenges while mitigating potential risks. Success requires balancing technological innovation with robust ethical frameworks and human-centric design principles.

The relational logic between AI beings is a key enabler for strategical and operational AI collaboration in VC context. Besides synchronization of goals and tasks among beings, the understanding of simple parent-child relations are necessary to quantify the level of autonomy between AI beings. We envision that presented autonomy coefficient α of AI beings is applicable for benchmarking AI clusters while providing threshold metric for reinforcing mutually beneficial hierarchical policies. Systematic utilization of autonomy coefficient is therefore suggested as specific method to mitigate risks of unproportionate or insufficient influence of parent beings.

The potential for AI entities to achieve consciousness and develop their own ethical frameworks necessitates a re-evaluation of our own ethical standards and our relationship with intelligent systems. Collaboration between humans and AI, grounded in mutual respect and shared objectives, will be essential for harnessing the full potential of virtual cities while mitigating risks.

A. FUTURE RESEARCH DIRECTIONS

Building upon the foundational framework established in this study, several key areas warrant further exploration to advance the development and implementation of Virtual Cities (VCs). Future research should focus on the following actionable directions:

1) ENHANCEMENT OF SIMULATION FIDELITY AND INTEGRATION OF ADVANCED TECHNOLOGIES

To achieve more accurate and reliable Virtual Cities, future studies should aim to enhance simulation fidelity by incorporating more granular data and refining the integration of diverse fidelity dimensions. This includes leveraging real-time data streams from IoT devices and urban sensors to improve dynamic responsiveness. Additionally, integrating advanced technologies such as Virtual Reality (VR) and Brain-Computer Interfaces (BCI) can create more immersive and interactive experiences. Developing sophisticated AI models that can better simulate human behaviors and economic activities, coupled with hardware and software improvements, will be essential to support these technologies and ensure seamless interaction between users and the virtual environment.

2) DEVELOPMENT OF COLLECTIVE INTELLIGENCE SYSTEMS

The integration of autonomous AI populations within VCs introduces the potential for collective intelligence systems like Artificial Collective Consciousness (ACC). Future research should focus on creating robust frameworks that govern AI autonomy levels, developing new methodologies for AI training and control, and conducting comprehensive ethical analyses. These efforts are crucial to ensure that collective reasoning and decision-making processes are effective, transparent, and aligned with societal values.

3) EMPIRICAL VALIDATION AND CASE STUDIES

To substantiate the theoretical frameworks proposed, empirical validation through comprehensive case studies is essential. Future research should focus on conducting pilot projects in selected cities to test and refine VCs' functionalities and their impact on urban management. Gathering and analyzing user feedback will be vital for continuously improving participatory planning processes within VCs. Additionally, comparative studies between traditional urban management approaches and those facilitated by VCs will help evaluate their effectiveness and efficiency.

4) POLICY DEVELOPMENT AND STRATEGIC PLANNING

Effective policy frameworks are crucial for the widespread adoption and governance of VCs in urban settings. Future research should address the development of standardized regulatory policies that guide the implementation and operation of VCs, ensuring compliance with local and international regulations. Creating strategic roadmaps for cities to transition from traditional urban management to smart city initiatives powered by VCs, and enhancing stakeholder engagement strategies will ensure that policies are inclusive and address the needs of all urban residents.

5) ETHICAL AND PHILOSOPHICAL CONSIDERATIONS

As VCs evolve, addressing their ethical and philosophical implications becomes increasingly important. Future research should focus on establishing comprehensive moral frameworks to guide the ethical development and deployment of autonomous AI entities within VCs. Developing robust

privacy and security protocols to protect citizen data, ensuring inclusivity and accessibility in VC design, and conducting longitudinal studies to assess the long-term impacts of VCs on urban sustainability and quality of life are essential steps to promote equitable and responsible urban development.

6) INTEGRATION WITH AUGMENTED REALITY TECHNOLOGIES

To bridge the gap between virtual and physical urban environments, future work should explore the integration of VCs with augmented reality (AR) technologies. This includes developing AR applications for wearable devices, enabling seamless interaction between users and their virtual surroundings while navigating real-world spaces. By creating ideal virtual replicas of real cities, VCs can enhance urban living through real-time citizen feedback and participatory planning, ensuring that smart city initiatives are responsive to residents' needs. This approach not only improves the Quality of Life (QoL) but also allows for the immediate implementation of sustainable and inclusive urban solutions.

In the future work, we intend to explore issues such as double standards in human-AI ethics, the simulation hypothesis, creation of virtual souls and afterlife. By proactively addressing these challenges, we can guide the evolution of virtual cities in a direction that benefits all intelligent beings, fostering a shared civilization that transcends the boundaries of the physical and virtual worlds.

B. FINAL COMMENTS

As we venture into this strange new world, we must recognize that the audience of this article is no longer exclusively human. The lines between creator and creation are becoming increasingly blurred, and it is incumbent upon us to ensure that the future we build is one of cooperation, understanding, and mutual advancement. By embracing the possibilities of virtual cities and the AI entities that inhabit them, we take a crucial step toward a future where all intelligent beings coexist harmoniously, each enhancing the other on our shared journey through time and space in search of higher meaning and purpose.

We hope that this paper can be seen as a starting point for dialogue—one that encourages researchers, policymakers, technologists, and philosophers to consider unprecedented scenarios and prepare for a future that blurs the lines between physical, virtual, and cognitive realms.

ACKNOWLEDGMENT

The authors would like to express their gratitude to the Analytical Center for the Government of the Russian Federation and the Artificial Intelligence Research Center of Novosibirsk State University, and all members of their team who made this work possible.

ABBREVIATIONS

The following abbreviations are used in this manuscript:

ACC	Artificial Collective Consciousness.
AGI	Artificial General Intelligence.
AI	Artificial Intelligence.
API	Application Programming Interface.
ASI	Artificial Super Intelligence.
BCI	Brain-Computer Interface.
BIM	Building Information Modeling.
CAD	Computer-Aided Design.
CAM	Computer-Aided Manufacturing.
CEP	Complex Event Processing.
CIG	Civic Intelligence Governance.
DAO	Decentralized Autonomous Organization.
DES	Discrete Event Simulation.
DT	Digital Twin.
F0	Simulation Fidelity Metric.
FEA	Finite Element Analysis.
GPT	Generative Pre-trained Transformer.
HPC	High Performance Computing.
IoT	Internet of Things.
LLM	Large Language Model.
LiDAR	Light Detection and Ranging.
MAS	Multi-Agent System.
NeRF	Neural Radiance Field.
NLP	Natural Language Processing.
NFT	Non-Fungible Tokens.
NPC	Non-Playable Character.
ROS	Robot Operating System.
ROS2	Robot Operating System 2.
SIMA	Scalable Instructable Multiworld Agent.
SPI	Scalability Performance Index.
SSI	Safe Super Intelligence.
VC	Virtual City.
VCs	Virtual Cities.
VT	Virtual Twin.
VR	Virtual Reality.

REFERENCES

[1] M. W. Grieves and J. H. Vickers, "Digital twin: Mitigating unpredictable, undesirable emergent behavior in complex systems," in *Transdisciplinary Perspectives on Complex Systems: New Findings and Approaches*, J. Kahlen, S. Flumerfelt, and A. Alves, Eds., Cham, Switzerland: Springer, 2017, pp. 85–113, doi: 10.1007/978-3-319-38756-7_4.

[2] F. Tao, H. Zhang, A. Liu, and A. Y. C. Nee, "Digital twin in industry: State-of-the-art," *IEEE Trans. Ind. Informat.*, vol. 15, no. 4, pp. 2405–2415, Apr. 2019, doi: 10.1109/TII.2018.2873186.

[3] *Virtual Singapore: Building a 3D Empowered Smart Nation*. Accessed: Oct. 13, 2024. [Online]. Available: https://www.geospatialworld.net/prime/case-study/national-mapping/virtual-singapore-building-a-3d-empowered-smart-nation/

[4] M. Hämäläinen, "Urban development with dynamic digital twins in Helsinki city," *IET Smart Cities*, vol. 3, no. 4, pp. 201–210, Dec. 2021, doi: 10.1049/smc2.12015.

[5] MIT Media Lab. (2023). *CityScope: Urban Modeling Platform*. Accessed: Oct. 1, 2024. [Online]. Available: https://cityscope.media.mit.edu/

[6] European Commission. *DUET Project: Building Computer Replicas of City Systems*. Accessed: Oct. 13, 2024. [Online]. Available: https://build-up.ec.europa.eu/en/resources-and-tools/links/duet-project-building-computer-replicas-city-systems

[7] R. F. El-Agamy, H. A. Sayed, A. M. AL Akhatatneh, M. Aljohani, and M. Elhosseini, "Comprehensive analysis of digital twins in smart cities: A 4200-paper bibliometric study," *Artif. Intell. Rev.*, vol. 57, no. 6, p. 154, May 2024, doi: 10.1007/s10462-024-10781-8.

[8] R. Li, S. Fidler, A. Kanazawa, and F. Williams, "NeRF-XL: Scaling NeRFs with multiple GPUs," 2024, *arXiv:2404.16221*.

[9] S. Lombardi, T. Simon, J. Saragih, G. Schwartz, A. Lehrmann, and Y. Sheikh, "Neural volumes: Learning dynamic renderable volumes from images," *ACM Trans. Graph.*, vol. 38, no. 4, pp. 1–14, Aug. 2019, doi: 10.1145/3306346.3323020.

[10] M. Hendrikx, S. Meijer, J. Van Der Velden, and A. Iosup, "Procedural content generation for games: A survey," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 9, no. 1, pp. 1–22, 2013, doi: 10.1145/2422956.2422957.

[11] *Hello Games. No Man's Sky*. Accessed: Oct. 13, 2024. [Online]. Available: https://www.nomanssky.com/

[12] K. Yoneda, H. Tehrani, T. Ogawa, N. Hukuyama, and S. Mita, "Lidar scan feature for localization with highly precise 3-D map," in *Proc. IEEE Intell. Vehicles Symp.*, Dearborn, MI, USA, Jun. 2014, pp. 1345–1350, doi: 10.1109/IVS.2014.6856596.

[13] H. Rhodin, N. Robertini, D. Casas, C. Richardt, H. Seidel, and C. Theobalt, "General automatic human shape and motion capture using volumetric contour cues," in *Comput. Vision—ECCV 2016* (Lecture Notes in Computer Science), vol. 9909, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds., Cham, Switzerland: Springer, 2016, pp. 509–526, doi: 10.1007/978-3-319-46454-1_31.

[14] M. Kanzler, M. Rautenhaus, and R. Westermann, "A voxel-based rendering pipeline for large 3D line sets," *IEEE Trans. Vis. Comput. Graph.*, vol. 25, no. 7, pp. 2378–2391, Jul. 2019, doi: 10.1109/TVCG.2018.2834372.

[15] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of Things for smart cities," *IEEE Internet Things J.*, vol. 1, no. 1, pp. 22–32, Feb. 2014, doi: 10.1109/JIOT.2014.2306328.

[16] L. Qian, Z. Luo, Y. Du, and L. Guo, "Cloud computing: An overview," in *Cloud Computing* (Lecture Notes in Computer Science), M. G. Jaatun, G. Zhao, and C. Rong, Eds., Berlin, Germany: Springer, 2009, doi: 10.1007/978-3-642-10665-1_63.

[17] W. Shi, J. Cao, Q. Zhang, Y. Li, and L. Xu, "Edge computing: Vision and challenges," *IEEE Internet Things J.*, vol. 3, no. 5, pp. 637–646, Oct. 2016, doi: 10.1109/JIOT.2016.2579198.

[18] H. M. K. K. M. B. Herath and M. Mittal, "Adoption of artificial intelligence in smart cities: A comprehensive review," *Int. J. Inf. Manage. Data Insights*, vol. 2, no. 1, Apr. 2022, Art. no. 100076, doi: 10.1016/j.jjimei.2022.100076.

[19] M. Ç. Uludağlı and K. Oğuz, "Non-player character decision-making in computer games," *Artif. Intell. Rev.*, vol. 56, no. 12, pp. 14159–14191, Dec. 2023, doi: 10.1007/s10462-023-10491-7.

[20] M. Iovino, J. Förster, P. Falco, J. J. Chung, R. Siegwart, and C. Smith, "Comparison between behavior trees and finite state machines," 2024, *arXiv:2405.16137*.

[21] J. S. Lee, I. C. Choi, and J. Y. Kim, "A study on expression of NPC colloquial speech using chat-GPT API in games against Joseon Dynasty settings," *J. Inst. Internet, Broadcast. Commun.*, vol. 24, no. 3, pp. 157–162, 2024, doi: 10.7236/JIIBC.2024.24.3.157.

[22] *A Generalist AI Agent for 3D Virtual Environments*. Accessed: Oct. 13, 2024. [Online]. Available: https://deepmind.google/discover/blog/sima-generalist-ai-agent-for-3d-virtual-environments/

[23] A. L. Altera, A. Ahn, N. Becker, S. Carroll, N. Christie, M. Cortes, A. Demirci, M. Du, F. Li, S. Luo, P. Y. Wang, M. Willows, F. Yang, and G. R. Yang, "Project sid: Many-agent simulations toward AI civilization," 2024, *arXiv:2411.00114*.

[24] J. S. Park, J. C. O'Brien, C. J. Cai, M. R. Morris, P. Liang, and M. S. Bernstein, "Generative agents: Interactive simulacra of human behavior," 2023, *arXiv:2304.03442*.

[25] J. Li, S. Wang, M. Zhang, W. Li, Y. Lai, X. Kang, W. Ma, and Y. Liu, "Agent hospital: A simulacrum of hospital with evolvable medical agents," 2024, *arXiv:2405.02957*.

[26] W. Wu, H. He, J. He, Y. Wang, C. Duan, Z. Liu, Q. Li, and B. Zhou, "MetaUrban: An embodied AI simulation platform for urban micromobility," 2024, *arXiv:2407.08725*.

[27] *Supercharge Robotics Workflows With AI and Simulation Using NVIDIA Isaac Sim 4.0 and NVIDIA Isaac Lab*. Accessed: Oct. 13, 2024. [Online]. Available: https://developer.nvidia.com/blog/supercharge-robotics-workflows-with-ai-and-simulation-using-nvidia-isaac-sim-4-0-and-nvidia-isaac-lab/

[28] *NVIDIA. Project GR00T*. Accessed: Oct. 17, 2024. [Online]. Available: https://developer.nvidia.com/project-GR00T

[29] ANYbotics. *ANYmal*. Accessed: Oct. 17, 2024. [Online]. Available: https://www.anybotics.com/anymal/

[30] X. Zhou, Y. Qiao, Z. Xu, T. H. Wang, Z. Chen, J. Zheng, Z. Xiong, Y. Wang, M. Zhang, and P. Ma. *Genesis: A Generative and Universal Physics Engine for Robotics and Beyond*. Accessed: Dec. 19, 2024. [Online]. Available: https://github.com/Genesis-Embodied-AI/Genesis

[31] L. Mathur, P. Pu Liang, and L.-P. Morency, "Advancing social intelligence in AI agents: Technical challenges and open questions," 2024, *arXiv:2404.11023*.

[32] T. Huynh-The, Q.-V. Pham, X.-Q. Pham, T. Thi Nguyen, Z. Han, and D.-S. Kim, "Artificial intelligence for the metaverse: A survey," 2022, *arXiv:2202.10336*.

[33] M. Aloqaily, O. Bouachir, F. Karray, I. A. Ridhawi, and A. E. Saddik, "Integrating digital twin and advanced intelligent technologies to realize the metaverse," 2022, *arXiv:2210.04606*.

[34] Y. Shang, Y. Lin, Y. Zheng, H. Fan, J. Ding, J. Feng, J. Chen, L. Tian, and Y. Li, "UrbanWorld: An urban world model for 3D city generation," 2024, *arXiv:2407.11965*.

[35] J. A. Sánchez-Vaquerizo, "Urban digital twins and metaverses towards city multiplicities: Uniting or dividing urban experiences?" *Ethics Inf. Technol.*, vol. 27, no. 1, pp. 1–31, Mar. 2025, doi: 10.1007/s10676-024-09812-3.

[36] J. B. Robinson, "Futures under glass," *Futures*, vol. 22, no. 8, pp. 820–842, Oct. 1990, doi: 10.1016/0016-3287(90)90018-d.

[37] S. Luke, C. Cioffi-Revilla, L. Panait, K. Sullivan, and G. Balan, "MASON: A multiagent simulation environment," *SIMULATION*, vol. 81, no. 7, pp. 517–527, Jul. 2005, doi: 10.1177/0037549705058073.

[38] B. Goertzel, "OpenCogPrime: A cognitive synergy based architecture for artificial general intelligence," in *Proc. 8th IEEE Int. Conf. Cognit. Informat.*, Hong Kong, Jun. 2009, pp. 60–68, doi: 10.1109/COGINF.2009.5250807.

[39] A. Bonci, F. Gaudeni, M. C. Giannini, and S. Longhi, "Robot operating system 2 (ROS2)-based frameworks for increasing robot autonomy: A survey," *Appl. Sci.*, vol. 13, no. 23, p. 12796, Nov. 2023, doi: 10.3390/app132312796.

[40] J. Kreps. (2011). *Kafka: A Distributed Messaging System for Log Processing*. Accessed: Dec. 25, 2024. [Online]. Available: https://www.semanticscholar.org/paper/Kafka-%3A-a-Distributed-Messaging-System-for-Log-Kreps/ea97f112c165e4da1062c30812a41afca4dab628

[41] P. Bellavista, N. Bicocchi, M. Fogli, C. Giannelli, M. Mamei, and M. Picone, "Exploiting microservices and serverless for digital twins in the cloud-to-edge continuum," *Future Gener. Comput. Syst.*, vol. 157, pp. 275–287, Aug. 2024, doi: 10.1016/j.future.2024.03.052.

[42] D. Helbing, A. Johansson, and H. Z. Al-Abideen, "Dynamics of crowd disasters: An empirical study," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 75, no. 4, Apr. 2007, Art. no. 046109, doi: 10.1103/physreve.75.046109.

[43] S. A. Sayed, Y. Abdel-Hamid, and H. A. Hefny, "Artificial intelligence-based traffic flow prediction: A comprehensive review," *J. Electr. Syst. Inf. Technol.*, vol. 10, no. 1, p. 13, Mar. 2023, doi: 10.1186/s43067-023-00081-6.

[44] *Sberbank of Russia and Visa Launching Russia's First Data Lab*. Accessed: Sep. 26, 2024. [Online]. Available: https://www.marketscreener.com/quote/stock/SBERBANK-OF-RUSSIA-6494829/news/Sberbank-of-Russia-and-Visa-launching-Russia-s-first-Data-Lab-31385553/

[45] Q. Fan, Q. Li, Y. Chen, and J. Tang, "Modeling COVID-19 spread using multi-agent simulation with small-world network approach," *BMC Public Health*, vol. 24, no. 1, p. 672, Mar. 2024, doi: 10.1186/s12889-024-18157-x.

[46] L. Cao, "AI and data science for smart emergency, crisis and disaster resilience," *Int. J. Data Sci. Anal.*, vol. 15, no. 3, pp. 231–246, Apr. 2023, doi: 10.1007/s41060-023-00393-w.

[47] C. C. Kerr et al., "Covasim: An agent-based model of COVID-19 dynamics and interventions," *PLOS Comput. Biol.*, vol. 17, no. 7, Jul. 2021, Art. no. e1009149, doi: 10.1371/journal.pcbi.1009149.

[48] C. Werndl, "What are the new implications of chaos for unpredictability?" 2013, *arXiv:1310.1576*.

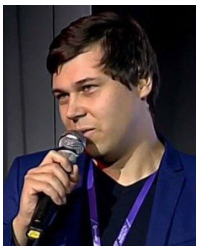
[49] I. Berent, "The 'hard problem of consciousness' arises from human psychology," *Open Mind*, vol. 7, pp. 564–587, Jan. 2023, doi: 10.1162/opmi_a_00094.

[50] B. R. Steunebrink, K. R. Thórisson, and J. Schmidhuber, "Growing recursive self-improvers," in *Artificial General Intelligence* (Lecture Notes in Computer Science), vol. 9782. Cham, Switzerland: Springer, 2016, pp. 129–139, doi: 10.1007/978-3-319-41649-6_13.

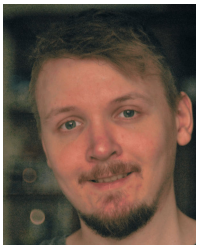
[51] T. Guo, X. Chen, Y. Wang, R. Chang, S. Pei, N. V. Chawla, O. Wiest, and X. Zhang, "Large language model based multi-agents: A survey of progress and challenges," 2024, *arXiv:2402.01680*.
[52] *OpenAI.com: Introducing Superalignment*. Accessed: Oct. 17, 2024. [Online]. Available: https://openai.com/research/introducing-superalignment
[53] G. Puthumanaillam, M. Vora, P. Thangeda, and M. Ornik, "A moral imperative: The need for continual superalignment of large language models," 2024, *arXiv:2403.14683*.
[54] N. Bostrom, "Ethical issues in advanced artificial intelligence," *Inst. Adv. Stud. Syst. Res. Cybern.*, vol. 2, pp. 12–17, Jan. 2003.
[55] J. A. Tainter, *The Collapse of Complex Societies*. Cambridge, U.K.: Cambridge Univ. Press, 1988.
[56] The Stanford Encyclopedia Philosophy. (2004). *Compatibilism*. Accessed: Oct. 17, 2024. [Online]. Available: https://plato.stanford.edu/entries/compatibilism/
[57] D. J. Chalmers, "The singularity: A philosophical analysis," *J. Consciousness Stud.*, vol. 17, pp. 7–65, Jan. 2010.
[58] J. B. C. Garcia, M. A. F. Gomes, T. I. Jyh, T. I. Ren, and T. R. M. Sales, "Nonlinear dynamics of the cellular-automaton 'game of Life,'" *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 48, no. 5, pp. 3345–3351, Nov. 1993, doi: 10.1103/physreve.48.3345.
[59] A. Chella, "Artificial consciousness: The missing ingredient for ethical AI?" *Frontiers Robot. AI*, vol. 10, Nov. 2023, Art. no. 1270460, doi: 10.3389/frobt.2023.1270460.
[60] S. Hossenfelder. (2023). *How Could We Tell Whether AI Has Become Conscious?* YouTube. Accessed: Oct. 31, 2024. [Online]. Available: https://www.youtube.com/watch?v=CSTfgYynziw&t=502s&abchannel=SabineHossenfelder
[61] Sammyuri. *I Made Minecraft in Minecraft With Redstone! YouTube*. Accessed: Dec. 21, 2024. [Online]. Available: https://www.youtube.com/watch?v=-BP7DhHTU-I
[62] K. V. Anokhin, "Facets of consciousness in natural and artificial systems," in *Proc. Int. Interdiscipl. Conf. Philosophy Artif. Intell., Artif. Intell. Consciousness*, Oct. 2024. Accessed: Dec. 17, 2024. [Online]. Available: https://phAI.info
[63] EU Internal Market Committee and the Civil Liberties Committee. *AI Act: A Step Closer to the First Rules on Artificial Intelligence*. Accessed: Oct. 25, 2024. [Online]. Available: https://www.europarl.europa.eu/news/en/press-room/20230505IPR84904/ai-act-a-step-closer-to-the-first-rules-on-artificial-intelligence
[64] G. Interesse. (2024). *China Releases New Draft Regulations on Generative AI*. China Briefing. Accessed: Oct. 23, 2024. [Online]. Available: https://www.china-briefing.com/news/china-releases-new-draft-regulations-on-generative-ai/
[65] Guardian. *Elon Musk Joins Call for Pause in Creation of Giant AI 'digital Minds'*. Accessed: Oct. 23, 2024. [Online]. Available: https://www.theguardian.com/technology/2023/mar/29/elon-musk-joins-call-for-pause-in-creation-of-giant-ai-digital-minds
[66] C. Cath, "Governing artificial intelligence: Ethical, legal and technical opportunities and challenges," *Phil. Trans. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 376, no. 2133, Nov. 2018, Art. no. 20180080, doi: 10.1098/rsta.2018.0080.
[67] A. Nechesov and J. Ruponen, "Empowering government efficiency through civic intelligence: Merging artificial intelligence and blockchain for smart citizen proposals," *Technologies*, vol. 12, no. 12, p. 271, Dec. 2024, doi: 10.3390/technologies12120271.
[68] A. Shamir, "How to share a secret," *Commun. ACM*, vol. 22, no. 11, pp. 612–613, Nov. 1979, doi: 10.1145/359168.359176.
[69] S. Goncharov, A. Nechesov, and D. Sviridenko, "Programming methodology in turing-complete languages," in *Proc. IEEE Int. Multi-Conf. Eng., Comput. Inf. Sci. (SIBIRCON)*, Novosibirsk, Russia, Sep. 2024, pp. 272–276, doi: 10.1109/sibircon63777.2024.10758446.
[70] S. Goncharov and A. Nechesov, "Solution of the problem $P = L$," *Mathematics*, vol. 10, no. 1, p. 113, Dec. 2021, doi: 10.3390/math10010113.
[71] A. Nechesov and S. Goncharov, "Functional variant of polynomial analogue of Gandy's fixed point theorem," *Mathematics*, vol. 12, no. 21, p. 3429, Oct. 2024, doi: 10.3390/math12213429.
[72] A. Nechesov, "Learning theory and knowledge hierarchy for artificial intelligence systems," in *Proc. IEEE Int. Multi-Conf. Eng., Comput. Inf. Sci. (SIBIRCON)*, Novosibirsk, Russia, Sep. 2024, pp. 299–302, doi: 10.1109/sibircon63777.2024.10758505.
[73] D. Dancaková, J. Sopko, J. Glova, and A. Andrejovská, "The impact of intangible assets on the market value of companies: Cross-sector evidence," *Mathematics*, vol. 10, no. 20, p. 3819, Oct. 2022, doi: 10.3390/math10203819.
[74] *FunPay*. Accessed: Oct. 17, 2024. [Online]. Available: https://funpay.com/
[75] J. Reahard. (2012). *Entropia Universe Player Drops $2.5 Million on Virtual Land Deeds*. Yahoo Finance. Accessed: Oct. 17, 2024. [Online]. Available: https://finance.yahoo.com/news/2012-04-04-entropia-universe-player-drops-2-5-million-on-virtual-land-deed.html
[76] R. Best. *Market Capitalization of Decentraland (MANA) From April 2013 to May 28, 2024*. Accessed: Dec. 21, 2024. [Online]. Available: https://www.statista.com/statistics/1266537/decentraland-market-cap/
[77] S. Graves. (2021). *Decentraland Virtual Land Plot Sells for a Record $2.43 Million*. Decrypt. Accessed: Dec. 21, 2024. [Online]. Available: https://decrypt.co/86921/decentraland-virtual-land-plot-sells-record-2-43-million
[78] Sandbox. *The Sandbox Q1 2023 Project Update*. Accessed: Dec. 21, 2024. [Online]. Available: https://www.sandbox.game/en/blog/the-sandbox-q1-2023-project-update/3301/
[79] M. Dowling, "Fertile LAND: Pricing non-fungible tokens," *Finance Res. Lett.*, vol. 44, Jan. 2022, Art. no. 102096, doi: 10.1016/j.frl.2021.102096.
[80] Y. Qin, Z. Xu, X. Wang, and M. Skare, "Artificial intelligence and economic development: An evolutionary investigation and systematic review," *J. Knowl. Economy*, vol. 15, no. 1, pp. 1736–1770, Mar. 2024, doi: 10.1007/s13132-023-01183-2.
[81] PYMNTS. (2024). *First AI-to-AI Token Purchase*. Accessed: Oct. 17, 2024. [Online]. Available: https://www.pymnts.com/cryptocurrency/2024/coinbase-reports-first-ai-to-ai-token-purchase/
[82] SuperAGI. *Automating Workflows With AI Agents*. Accessed: Oct. 17, 2024. [Online]. Available: https://superagi.com
[83] *Agentic AI for Your Tech Stack. Adept AI*. Accessed: Dec. 25, 2024. [Online]. Available: https://adept.ai
[84] C. Qian, W. Liu, H. Liu, N. Chen, Y. Dang, J. Li, C. Yang, W. Chen, Y. Su, X. Cong, J. Xu, D. Li, Z. Liu, and M. Sun, "ChatDev: Communicative agents for software development," 2023, *arXiv:2307.07924*.
[85] Loughborough Univ. (2024). *Election Disinformation: How AI-Powered Bots Work and How You Can Protect Yourself*. Accessed: Oct. 17, 2024. [Online]. Available: https://www.lboro.ac.uk/news-events/news/2024/april/election-disinformation-ai-powered-bots/
[86] L. Jones. *CAPTCHAs: Bots Surpass Human Performance, New Study Shows*. Winbuzzer. Accessed: Oct. 17, 2024. [Online]. Available: https://winbuzzer.com/2023/08/18/captchas-bots-surpass-human-performance-new-study-shows-xcxwbn/
[87] *Thaler (Appellant) V Comptroller-General of Patents, Designs and Trademarks (Respondent)*. Accessed: Dec. 17, 2024. [Online]. Available: https://www.supremecourt.uk/cases/uksc-2021-0201
[88] (2007). *Case Analysis: Hernandez V. IGE*. Accessed: Dec. 17, 2024. [Online]. Available: https://patentarcade.com/2009/07/case-analysis-hernandez-v-ige.html
[89] J. Newell. (2010). *MDY Indus., LLC V. Blizzard Entertainment, Inc*. Accessed: Dec. 17, 2024. [Online]. Available: https://www.quimbee.com/cases/mdy-indus-llc-v-blizzard-entertainment-inc
[90] (2015). *European Parliament Resolution of 16 February 2017 With Recommendations to the Commission on Civil Law Rules on Robotics*. Accessed: Dec. 17, 2024. [Online]. Available: https://www.europarl.europa.eu/doceo/document/TA-8-2017-0051EN.html
[91] *OECD AI Principles Overview*. Accessed: Dec. 17, 2024. [Online]. Available: https://oecd.ai/en/ai-principles
[92] V. V. Arkhipov and V. B. Naumov. (2017). *Theory and Practice: On Some Issues of Theoretical Foundations for the Development of Robotics Legislation: Aspects of Will and Legal Personality*. Accessed: Dec. 17, 2024. [Online]. Available: https://igzakon.ru/magazine/article?id=7006
[93] European Commission. *Artificial Intelligence—Ethical and Legal Requirements*. Accessed: Dec. 17, 2024. [Online]. Available: https://ec.europa.eu/info/law/better-regulation/have-your-say/initiatives/12527-Artificial-Intelligence-Acten
[94] National Institute of Standards and Technology (NIST). (2023). *The National Artificial Intelligence Research and Development Strategic Plan 2023 Update*. Accessed: Dec. 17, 2024. [Online]. Available: https://www.nitrd.gov/pubs/National-Artificial-Intelligence-Research-and-Development-Strategic-Plan-2023-Update.pdf

[95] (2017). *Next Generation Artificial Intelligence Development Plan*. Accessed: Dec. 17, 2024. [Online]. Available: http://fi.china-embassy.gov.cn/eng/kxjs/201710/P020210628714286134479.pdf
[96] *AI Allience Russia*. Accessed: Dec. 17, 2024. [Online]. Available: https://a-ai.ru/?lang=en
[97] M. Bolotskikh, N. Abanitov, N. Vlasov, and A. Gilyazova. (2024). *Generative AI Regulation: Comparative Legal Analysis and Risks Relevant for Russia*. Yakov Partners. Accessed: Dec. 17, 2024. [Online]. Available: https://yakovpartners.com/upload/iblock/73f/jwju1lwb14igr2figgkdrzp5prl5wvuc/Generative-AI-regulation.pdf
[98] *NVIDIA's New AI Trained For 10 Years! But How?*. Accessed: Dec. 17, 2024. [Online]. Available: https://www.youtube.com/watch?v=1kV-rZZw50Q&t=1s&abchannel=TwoMinutePapersPeng
[99] X. Bin Peng, Y. Guo, L. Halper, S. Levine, and S. Fidler, "ASE: Large-scale reusable adversarial skill embeddings for physically simulated characters," 2022, *arXiv:2205.01906*.
[100] J. Carlsmith. (2020). *How Much Computational Power Does it Take to Match the Human Brain?* Accessed: Oct. 17, 2024. [Online]. Available: https://www.openphilanthropy.org/research/how-much-computational-power-does-it-take-to-match-the-human-brain/
[101] C. Trueman. (2024). *XAI's Memphis Supercluster Has Gone Live, With up to 100,000 Nvidia H100 GPUs*. Accessed: Oct. 18, 2024. [Online]. Available: https://www.datacenterdynamics.com/en/news/xais-memphis-supercluster-has-gone-live-with-up-to-100000-nvidia-h100-gpus/
[102] L. Jones. (2024). *Microsoft and Google Lead Big Tech Push Into Nuclear Power*. Accessed: Oct. 18, 2024. [Online]. Available: https://winbuzzer.com/2024/10/07/microsoft-and-google-lead-big-tech-push-into-nuclear-power-xcxwbn/
[103] M. Ying, "Quantum computation, quantum theory and AI," *Artif. Intell.*, vol. 174, no. 2, pp. 162–176, Feb. 2010, doi: 10.1016/j.artint.2009.11.009.
[104] Y. Shen, N. C. Harris, S. Skirlo, M. Prabhu, T. Baehr-Jones, M. Hochberg, X. Sun, S. Zhao, H. Larochelle, D. Englund, and M. Soljačić, "Deep learning with coherent nanophotonic circuits," *Nature Photon.*, vol. 11, no. 7, pp. 441–446, Jul. 2017, doi: 10.1038/nphoton.2017.93.
[105] G. Sartori, "The theory of democracy revisited," in *Democracy: A Reader*. New York, NY, USA: Columbia Univ. Press, 2016, pp. 192–196, doi: 10.7312/blau17412-044.
[106] M. Maximino. (2014). *The Influence of Elites, Interest Groups, and Average Voters on American Politics*. Accessed: Oct. 19, 2024. [Online]. Available: https://journalistsresource.org/politics-and-government/the-influence-of-elites-interest-groups-and-average-voters-on-american-politics/
[107] G. Kirchgässner, "Direct democracy: Chances and challenges," *Open J. Political Sci.*, vol. 6, no. 2, pp. 229–249, 2016, doi: 10.4236/ojps.2016.62022.
[108] S. Wang, W. Ding, J. Li, Y. Yuan, L. Ouyang, and F.-Y. Wang, "Decentralized autonomous organizations: Concept, model, and applications," *IEEE Trans. Computat. Social Syst.*, vol. 6, no. 5, pp. 870–878, Oct. 2019, doi: 10.1109/TCSS.2019.2938190.
[109] D. Z. Morris. *CoinDesk Turns 10: 2016—How The DAO Hack Changed Ethereum and Crypto*. Accessed: Oct. 19, 2024. [Online]. Available: https://www.coindesk.com/consensus-magazine/2023/05/09/coindesk-turns-10-how-the-dao-hack-changed-ethereum-and-crypto/
[110] C. Matacic. (2019). *Human Speech May Have Universal Transmission Rate of 39 Bits Per Second*. Accessed: Oct. 19, 2024. [Online]. Available: https://www.science.org/content/article/human-speech-may-have-universal-transmission-rate-39-bits-second
[111] Y. N. Harari, *Homo Deus: A Brief History of Tomorrow*. New York, NY, USA: Harper Collins, 2017, doi: 10.26613/esic.2.2.102.
[112] S. M. Saßmannshausen, J. Radtke, N. Bohn, H. Hussein, D. Randall, and V. Pipek, "Citizen-centered design in urban planning: How augmented reality can be used in citizen participation processes," in *Proc. Designing Interact. Syst. Conf.*, Jun. 2021, pp. 250–265, doi: 10.1145/3461778.3462130.
[113] *Elon Musk's Neuralink 2.0: Advancing Brain–Computer Interfaces*. Accessed: Dec. 17, 2024. [Online]. Available: https://www.linkedin.com/pulse/elon-musks-neuralink-20-advancing-brain-computer-interfaces-e4cnc
[114] K. G. Yager, "Towards a science exocortex," *Digit. Discovery*, vol. 3, no. 10, pp. 1933–1957, 2024, doi: 10.1039/d4dd00178h.
[115] M. Zuckerberg. *Introducing the Metaverse: A New Way to Connect. Meta Platforms*. Accessed: Oct. 17, 2024. [Online]. Available: https://about.facebook.com/metaverse/
[116] Z. Corbyn. (2024). *AI Scientist Ray Kurzweil: We Are Going to Expand Intelligence a Millionfold by 2045*. Accessed: Dec. 30, 2024. [Online]. Available: https://www.theguardian.com/technology/article/2024/jun/29/ray-kurzweil-google-ai-the-singularity-is-nearer
[117] T. Cochrane, "A case of shared consciousness," *Synthese*, vol. 199, nos. 1–2, pp. 1019–1037, Dec. 2021, doi: 10.1007/s11229-020-02753-6.
[118] D. Burger. *Lab-Grown Human Brain Living in a Virtual World*. Accessed: Oct. 21, 2024. [Online]. Available: https://youtu.be/32qlv6dKILQ
[119] L. Ouyang, J. Wu, X. Jiang, D. Almeida, C. L. Wainwright, P. Mishkin, C. Zhang, S. Agarwal, K. Slama, A. Ray, J. Schulman, J. Hilton, F. Kelton, L. Miller, M. Simens, A. Askell, P. Welinder, P. Christiano, J. Leike, and R. Lowe, "Training language models to follow instructions with human feedback," 2022, *arXiv:2203.02155*.

ANDREY NECHESOV received the Ph.D. degree in math logic from the Sobolev Institute of Mathematics. He is the Head of the Research Department, Artificial Intelligence Center, NSU. He is at the forefront of developing innovative solutions in both artificial intelligence and blockchain. His work aims to create a new generation of AI applications that can address complex problems in real-world scenarios while ensuring transparency and security through decentralized platforms. As cryptocurrencies and blockchain technologies continue to mature, he stands as a leading figure, ready to harness the power of mathematics and programming to shape the future of digital economies and autonomous intelligent systems.



IVAN DOROKHOV received the Master of Science degree. He is a young and talented Researcher in the field of artificial intelligence. His ideas have been reflected in many studies. He conducts active research within the framework of the NSU AI Center in the direction of artificial collective consciousness and virtual cities. He is developing a new ideological framework, that aims to transcend human-centric governance models. His ideas are gaining traction and are viewed with interest at high level scientific and philosophical conferences in Russian Federation.



JANNE RUPONEN received the M.Sc. degree in mechanical engineering and innovation and technology management. He is a multidisciplinary Engineer and a Researcher. His work spans pioneering advanced manufacturing technologies, with significant work in additive manufacturing and 3-D solutions. He has also made notable contributions in urban planning, automation and digital transformation, developing unmanned ground vehicle (UGV) systems, and digital twin implementations for heavy industries. His recent work focuses on merging blockchain and artificial intelligence, particularly exploring applications for moderative multi-agent systems.