Urban Planning
2025 • Volume 10 • Article 8613
https://doi.org/10.17645/up.8613

ARTICLE

Open Access Journal

The Potentials and Limitations of Agent-Based Models for Urban Digital Twins: Insights From a Surveillance and Behavioral Nudging Simulation

Sarah Shtaierman, Catarina Fontes, and Christoph Lütge

Institute for Ethics in Artificial Intelligence, Technical University of Munich, Germany

Correspondence: Catarina Fontes (catarina.fontes@tum.de)

Submitted: 3 May 2024 **Accepted:** 21 August 2024 **Published:** in press

Issue: This article is part of the issue "AI for and in Urban Planning" edited by Tong Wang (TU Delft) and Neil Yorke-Smith (TU Delft), fully open access at https://doi.org/10.17645/up.i388

Abstract

Although urban digital twins are still at an embryonic stage of development, their use cases are multiple, ranging from big data aggregation to simulations. Additionally, predictions can be rendered and quickly implemented using actuators to transform physical environments and influence urban life. In this article, we investigate the potential of an agent-based model in a smart city setting to predict emergent behavior in relation to the suppression of civil violence by implementing crowd management practices. To this end, we designed a simulation environment that includes cameras in public spaces and wearable sensors, and considers nudging and self-nudging processes supported by a surveillance apparatus. Building on Epstein's threshold-based model of civil violence, the proposed simulation is informed by surveillance theories and contemplates methods for crowd monitoring and social control. The experiments' results provide insights into how specific measures and combined actions may influence the suppression of civil violence in public spaces and can be useful to inform crowd management activities and policymaking. Moreover, we use the simulation to reflect upon the potentials and limitations of integrating agent-based models into urban digital twins and emphasize the imminent risks for individuals and democratic societies of employing a ubiquitous surveillance apparatus endowed with the autonomy to trigger actuators.

Keywords

agent-based model; crowd modeling; smart city; surveillance systems; urban digital twin; urban planning

1. Introduction

The rights to freedom of peaceful assembly and association are essential components of democracy, allowing citizens to hold meetings, sit-ins, strikes, rallies, events, or protests both offline and online. They guarantee freedom of expression, letting people take part in public affairs by, for instance, interacting and organizing to collectively express, promote, pursue, and defend common interests. These liberties are deemed vital enough to be enshrined in the Universal Declaration of Human Rights (UN General Assembly, 1948). Peaceful assembly and association are acquired rights and a manifestation of democratic values as they enact social change through the power of crowds. However, given that such movements mobilize crowds, they are often monitored by public authorities to prevent civil violence.

In urban planning, the idea of controlling crowds through surveillance has been addressed throughout centuries, informing many theories and the design of urban landscapes. For instance, the panopticon is proposed as a prison, in which each cell is observable by a tower placed at the center giving the impression that the cells are continuously watched and leading inmates to regulate their behavior in expectation of a potential gaze, orchestrating a ubiquitous form of surveillance that could be used to discipline them anytime (Bentham, 1791). Another example is the original plan of Brasília. Streets as a space of convergence between pedestrians and cars were to be abolished, while the low building density was intended to dissipate crowds, thus affecting immanent forms of socialization and assembly in public spaces (Holston, 1989).

As the smart city emerges, social control and the exercise of power increasingly rely on digital tools and technology. The smart city has become an umbrella concept that branches off in multiple approaches to the digitization of services, using, for instance, embedded sensors in public spaces or wearables carried by citizens. Urban-applied projects related to different spheres of planning and management are driven by technology-based solutions, often intertwined across many scales and connected through multiple networks (Batty, 2018; Batty et al., 2012). All these sensors feed a data collection apparatus that urban digital twins (UDTs) are intended to govern as a holistic digital replica, reproducing the real urban setting and activities within. UDTs are ultimately supposed to digitally represent and simulate scenarios based on a continuous feed of real-time sensor data, allowing distributed actuators within the city to automatically act on their predictions. Endeavors like the Cybersyn project (Espejo, 2014; Medina, 2011) and the City Brain (Liu et al., 2022; Zhang et al., 2019) are aligned with holistic visions for UDTs that entail leveraging ubiquitous sensing and relying on autonomous machines to infer and perpetrate urban management activities.

When proposed to capture urban complexity, crowd monitoring and social control are among the many use cases of UDTs (Caldarelli et al., 2023). Similarly, agent-based models (ABMs) are widely used in a variety of disciplines aiming to simulate complex systems and study emergent behavior, namely in relation to civil violence (Epstein, 2002; Fonoberova et al., 2012; Lemos et al., 2013). Nonetheless, the integration of ABMs into UDTs and their potential to predict and suppress civil violence leveraging surveillance systems and autonomous actuators inspired by a smart city scenario has not been fully explored. Addressing this gap in research, we propose a simulation investigating the agents' emergent behavior when encountering measures employed to suppress the escalation of civil violence based on a surveillance-driven approach, leveraging state-of-the-art technologies and nudging practices.

In the next section, we present the theoretical framework behind the design of an ABM to explore surveillance and behavioral nudging practices to suppress civil violence. In Section 3, we explain the model's design (further detailed in the Supplementary Material), the experiments' setup, and results. In Section 4, we discuss the potential of ABMs in relation to UDTs and the challenges both technologies and the underlying idea of surveillance represent to society. We finalize by highlighting how ABMs associated with UDTs can transform society and introduce a paradigm shift.

2. Urban Surveillance and Behavioral Nudging

2.1. Surveillance Theories

Surveillance in relation to urban spaces and urban life has led to several theories, some of which became particularly relevant to question the role of surveillance in social control and due to the novel frameworks introduced to address social change in relation to the ways spaces are scrutinized and designed in order to promote or dissuade specific activities. Below we present some of these theories, which inspired the conceptualization of the ABM proposed.

Foucault's (1977) theory addresses society through power structures and their relations to discipline and has its roots in the panopticon (Bentham, 1791). The panopticon is illustrated as a prison, designed to automize and deindividualize power, thereby rendering it insignificant whether someone is actually monitoring the prisoners. The same applies to the watchmen, who in turn fear being watched by their superiors at any given moment. According to Foucault, postindustrial societies have established institutions that function identically. Schools, universities, hospitals, or factories subject individuals to work in isolation and train them to internalize discipline and anticipate punishment with the goal of achieving social conformity and productivity (Foucault, 1977). Building on Foucault's theory, it has been proposed that society is trapped in a sort of prison created around a fixation on self-optimization and productivity (Han, 2015) or that synoptic surveillance allows mutual monitoring, including public and media scrutiny of politicians which may deter deviant behavior (Mathiesen, 2017).

Deleuze (2017) introduces the concept of "society of control," defined by the dissolution of spatial and temporal boundaries. In a society of control, individuals are no longer confined to a certain location where a task is performed for a certain amount of time. The theory rejects a holistic conception of individuals, which in turn are opposed to the masses, in favor of multiple, simultaneous dividuals that are transient and fluid. Fractured into multiple demographics, these abstract descriptors provide a rough facsimile of an individual when put together but fail to characterize an actual human being (Deleuze, 2017).

Whereas the subject of a disciplinary society lives in constant fear of punishment, the dividual of a society of control may even fail to realize their own subjugation, rendering organized resistance nigh impossible. Another difference is that a society of control moves away from centralized institutions visibly exerting their power on individuals to propose a dispersed and transient power structure, aligned with the ethos of smart city-enabled mass surveillance.

Haggerty and Ericson (2017) propose the surveillant assemblage, where surveillance manifests as the convergence of heterogeneous, unrelated systems that form an all-encompassing web of control. This assemblage consists of a unity of items that form a functional entity in the broadest sense and is inherently unstable and fluid, thereby impossible to institutionalize or codify. It is also rhizomatic in nature, thus it does not target or exempt certain parts of society but subjects everyone to varying kinds of comprehensive surveillance. The world and the humans populating it are described as a system of flows that must be broken up, redirected, and analyzed for the purpose of management, profit, and entertainment (Haggerty & Ericson, 2017). The reference to doubles, which roughly correspond to the concept of dividuals in the society of control, further illustrates how information and data flows are used to manage society as a system. A major purpose of the surveillant assemblage and surveillance in general is the classification and monitoring of people to manage crowd behavior through social control. The analysis, visualization, and bundling of disparate data streams is linked to urban management and governance practices, which, in turn, seeks to streamline the former. We consider that UDTs as a tool primarily devised for city management may become the culmination of surveillant assemblages.

Jacobs' conceptualization of surveillance emerges from the discussion on safety of urban spaces, underscoring the potential of crowds, i.e., continuous and large affluence of people and casual surveillance undertaken by both pedestrians and residents, to ensure urban safety (Jacobs, 1961). Casual surveillance refers to the friendly gaze of a curious onlooker, and citizens who develop a sense of belonging in relation to their neighborhood taking on an active role in monitoring their surroundings, promoting urban safety. Mutual surveillance becomes a natural activity between strangers who share the same space and the author suggests that such community-grounded mechanisms are more effective than police surveillance in fostering urban safety. The idea of citizens monitoring their surroundings transported to a smart city setting resembles activities related to crowdsourcing data and participatory practices leveraging digital platforms.

2.2. Surveillance and Automation in the Smart City

The smart city is driven by the digitalization and automation of services and infrastructures (Kitchin, 2017). Sensors and actuators are becoming dominant elements of urban landscapes and can work complementarily to enable the automation of components of the urban infrastructure. Sensors are deployed to collect data, i.e., to take measurements and convert them into an electrical current, while actuators work through electrical impulses converted into visible phenomena, thereby transforming urban environments (Arshi & Mondal, 2023). In extension, UDTs introduce a holistic vision for urban twinning processes leveraging ubiquitous computing, 3D reconstruction, big data, and artificial intelligence (AI) and promise numerous paradigm shifts by enabling the integration of multiple sources of heterogeneous data and moving away from control to prevention and emergence (Sadowski & Pasquale, 2015). UDTs are also considered to plan and manage cities as complex systems (Khajavi et al., 2019), harnessing citizens as both sensors and actuators (Wang et al., 2012). In the aftermath, citizens can be nudged towards decisions at the individual and societal level due to an apparatus of intelligent mass surveillance, managed with little to no human interference.

2.2.1. Sensors

CCTV systems including cameras, recorders, and monitors that capture, store, and view video footage were first introduced in public spaces in the 1980s and have been increasingly adopted to monitor urban spaces (Hempel & Töpfer, 2009). With AI they can be endowed with facial recognition features, enabling remote and real-time identity checks and tracking (Fontes & Perrone, 2021; Kumari et al., 2023; Norris & Armstrong, 2017).

There are multiple types of cameras that may be adapted to urban surveillance systems. Video cameras can be further categorized into fixed and pan-tilt-zoom cameras. The latter can have a flexible view range when controlled by a remote operator. Infrared and thermal cameras, particularly suited for poor lighting conditions, detect near-infrared and short-wavelength infrared radiation, while thermal cameras capture long-wave and far-infrared emissions. Radar (radio detection and ranging) and LiDAR (light detection and ranging) make use of radio waves or utilize lasers to detect their surroundings, delivering high accuracy and robustness under different lighting conditions and offering a large range of vision. Nonetheless, these promising technologies are still considered ill-suited for detecting humans or recognizing objects (Ibrahim, 2016; Kumari et al., 2023).

In public spaces, multiple cameras of the same (or different) type are strategically deployed to supplement each other's data. Auditory, ultrasonic, passive infrared, and pressure sensors, among many others, can be used complementarily (Elharrouss et al., 2021). Nonetheless, the integration of multi-sensor systems poses countless technical difficulties and challenges stemming from data fusion (Sreenu & Durai, 2019). Surveillance can also be undertaken through wearable sensors and smartphones. On-body sensing using smartphones and wearables has become a powerful source of both personal and environmental data, making use of, for instance, built-in accelerometers and humidity sensors. Smartphones are used to sense crowd densities and model their behavior through GPS, cellphone signal, and WiFi triangulation (Borean et al., 2015). In extension, the metadata produced by mobile telephony (e.g., the identification number of SIM cards, call duration, or whether a phone was switched on or off) can be utilized by government agencies to monitor suspicious behavior and track individuals under suspicion (Leistert, 2012).

A variant of on-body sensing is crowd-powered human sensing, i.e., crowdsourced data that is voluntarily provided through digital apps for diverse purposes including, for instance, noise mapping and air quality control (Rana et al., 2010; Zhuang et al., 2015). However, there are challenges in assuring data quality (Wang et al., 2012), which relate to existing infrastructure (Franke et al., 2015; Tewksbury, 2012) and lack of participants, although incentive mechanisms such as virtual credit are also used (Lee & Hoh, 2010).

2.2.2. Actuators

Actuators are instruments or processes that convert control or input signals into actual motion, force, or other desired actions (Arshi & Mondal, 2023). Under the smart city paradigm, actuators are meant to enact physical changes based on processed data collected by distributed sensors. This means, for instance, self-adapting roads in response to traffic jams or automatic water supply management (Pompigna & Mauro, 2022). Despite the existence of a variety of actuator types (for example hydraulic, pneumatic, and thermal), we focused on actuators related to crowd management, crowd control, or crowd steering. In Table 1, we summarize three main approaches to crowd management according to the literature.

2.3. Defining Centralized Surveillance, Coveillance, and Selfveillance

We propose a classification of surveillance types inspired by a combination of social and technology-based approaches to urban surveillance. We first divide surveillance types into centralized surveillance and decentralized surveillance. Centralized is enforced through an apparatus managed and controlled through identifiable power structures reinforcing governmental bodies, drawing on Foucault's theory. On the other hand, decentralized surveillance is inspired by Deleuze's society of control theory and Haggerty and Ericson's surveillance assemblage theory, relying on decentralized and more subtle forms of surveillance, leading to self-regulation, i.e., individuals watching their own behavior and self-nudging practices. The coveillance approach is inspired by casual surveillance drawing on Jacobs' theory and entails relying on communities to self-regulate themselves by having individuals watching others (see Section 2.1 and Table 2).

Table 1. Three approaches to crowd management.

Crowd management	Examples of actuators operating	References
Dynamic spatial changes	Manipulating entrances, exits, and barriers, the spatial configuration may either encourage or deter crowds from gathering. Manipulating train schedules or outright skipping stations to prevent people from assembling. Radio-frequency identification (RFID), GSM-based (i.e., phone-based) identification, and biometric door locks can prevent specific people from entering or leaving certain areas.	Franke et al. (2015); Sadowski and Pasquale (2015); Shetty et al. (2020)
Crowd communication	Wearable devices or smartphones to monitor locations and movements, as well as nudge their users to certain actions, i.e., expose them to conditions that subtly encourage a specific change in behavior through visual, tactile, and auditory cues. Text messages and warnings widely sent to citizens' private smartphones by the public authorities. Customized social and monetary incentives using smartphone applications. Phone jamming or tracking used to prevent communication among the crowd.	Benartzi et al. (2017); Sadowski and Pasquale (2015); Singla et al. (2015); Tewksbury (2012)
Physical dissuasion	Military technology such as drones or sublethal weapons like long-range acoustic devices (LRADs) can be used for crowd control.	Sadowski and Pasquale (2015)

Table 2. Classification of surveillance informing the design of an ABM.

Surveillance types		Examples of actuators operating	Representation in the ABM*
Centralized surveillance		Systems based on observation and data collection by, e.g., CCTV cameras deployed in public spaces.	Visible and hidden surveillance cameras. Phone tracking, phone jamming, and mass messaging.
Decentralized surveillance	Coveillance	Describes mutual surveillance between citizens, albeit nudged by technology intentionally or unintentionally promoting urban safety.	People inflow. Monetary reward to monitor peers.
	Selfveillance	Citizens monitor their own behavior despite a seeming absence of external surveillance, promoted under the guise of self-improvement and as a measure of self-protection.	Regular informational campaigns. Encouraging self-tracking apps.

Note: * See Table 1 in the Supplementary Material.

The ABM was designed to simulate how different combinations of the defined surveillance types affect the population agents, using pattern-oriented modeling (see Table 2 and Table 3). The approach is based on layering multiple observable phenomena to promote structural realism and aims to address a still unresolved question in the literature—which surveillance practices are most effective at preventing and suppressing emergent violent behavior in crowds.

In terms of how the surveillance types informed the ABM's design, the first decision was to translate the approaches into techniques supported by sensors and actuators mimicking a smart city setting. In centralized surveillance systems, visible or signaled cameras are more recognizable and therefore avoidable by who passes by, while hidden cameras represent both their namesake and miscellaneous sensors that a citizen cannot recognize, thus they cannot be avoided. In the ABM, visible cameras can be destroyed by violent agents under certain conditions whereas hidden cameras cannot be destroyed, unless located within the neighborhood of a destroyed visible camera. This models collateral damage and discoveries by violent agents sweeping the vicinity in search of other surveillance devices.

Phone tracking describes targeted phone espionage, while phone surveillance entails general mining of data generated by smartphones. Phone tracking covers the process of monitoring the location and activities of specific individuals. Additionally, phone jamming describes the interruption of phone communication between population agents, while mass messaging is a form of crowd communication by sending the citizens messages on the phone.

In decentralized surveillance systems, we consider people inflow a form of casual surveillance. In the ABM, the number of new agents entering or old agents exiting the grid can be artificially promoted and controlled through actuators. A monetary coveillance reward is introduced to simulate incentivized crowd-powered actuators, testing whether citizens receiving a reward for using apps that foster coveillance could be an effective measure. Selfveillance measures the efficacy of adopting an educational approach to foster self-nudging, which includes providing information material on public safety, encouraging self-tracking apps and promoting proximity between government and citizens.

Selfveillance was inspired by measures adopted during the Covid-19 pandemic, when contact tracing apps became widely used public surveillance systems, voluntarily accepted by citizens around the world (Fontes et al., 2022). Similarly, many other tools for crowd control and the monitoring of individuals' mandatory isolation were proposed, although ethical and legal concerns were raised (Fontes et al., 2023).

2.4. Civil Violence and Surveillance

In this article, we use the definition of civil violence as a spontaneous violent outburst against a central authority and aim to test measures employed that relate to authorities' ability to suppress it. In this sense, we underscore the strong relation between civil violence and surveillance. The latter is used by the state to ensure governability (Giddens, 1986), inextricably tied to policing activities, enforcing a social contract between citizens and authorities. Certain forms of surveillance are therefore established as an extension of the state's authority and in some cases refer to the exercise of power and a manifestation of authoritarianism. Thus, the emerging possibilities of surveillance linked to even early stages of UDTs might trigger resistance movements and community backlash, as seen in the case of face recognition technologies (Fontes & Perrone, 2021). Such resistance may involve the avoidance and destruction of surveillance devices (Ullrich & Knopp, 2018), which we explore on the proposed ABM.

3. Designing an ABM for Predicting and Preventing Civil Violence

3.1. ABMs of Civil Violence

Agent-based modeling is a way of investigating complex systems, initially proposed in the 1980s, based on individual actors (agents) that act according to pre-programmed rules within a certain environment. The goal is to observe and explore significant emergent behavior while varying environmental or starting conditions in each experiment (Railsback, 2019). Regarding ABMs of civil violence, such as protests, riots, and uprisings, there are two distinct approaches. The first is the so-called rational behavior model, which iteratively optimizes a utility function. The second option yields rule-based models, in which agents change their behavior when certain thresholds are crossed. This approach is widely used when studying collective behavior and contagion effects (Lemos et al., 2013). The most influential rule-based model of civil violence was developed by Epstein (2002), modeling a central authority attempting to resist a decentralized rebellion. The results proved to capture many real phenomena qualitatively, such as deceptive behavior of population agents and sudden outbursts of violence after crossing certain tipping points (Epstein, 2002).

We build on Epstein's model and consider relevant criticism that followed the presentation of this ABM. Indeed, significant shortcomings were identified in follow-up research, such as unrealistic agent movement, simplistic cop modeling, and the agents possessing no memory and not having the capacity to learn from past events (Lemos et al., 2013). Consequently, improvements were proposed to investigate how many policemen are required to maintain low crime rates in urban settings (Fonoberova et al., 2012) or by adding a personality vector and a third class of media agents to simulate street protests (Lemos et al., 2013). Other approaches looked at modeling more realistic social contagion processes of civil disorder during public demonstrations (Kurland & Chen, 2016) or replacing the arrest probability with an Iterated Prisoner Dilemma game and introducing more granular parameters for population agents (Goh et al., 2006). Overall, a major concern in designing an ABM refers to the selection of components and rules to capture reality, as overly complex models are hardly interpretable (Railsback, 2019).

3.2. Entities and Environment

The proposed ABM simulation takes place within a two-dimensional 40x40 patches large grid, modeling an open public space such as a town square. Each discrete time step (tick) corresponds to 10 seconds. All experiments are set to run for four hours each (i.e., 1,440 ticks). Population agents refer to citizens who move across the grid. They can be created and removed from the simulation during runtime, i.e., enter or leave the public space. Although place-making factors are not explicitly included in the design of the grid, controlling the influx of new population agents as well as the speed at which they move across the board mimics the dynamics of public spaces and their importance to a local community. Furthermore, population agents can switch between four different states (quiet, agitated, violent, and jailed) according to certain rules, further determining their role and behavior in the simulation (Table 3). For more details see the Supplementary Material provided.

The objective of the simulation is to predict and prevent civil violence. This translates into maximizing the number of quiet agents, i.e., reducing the amount of agitated and violent agents. Quiet, agitated, and violent agents may switch back and forth between these states. However, only agitated and violent agents may be

Table 3. Types of agents and their characteristics.

Type of agents	Characteristics
Quiet agent	Neither agitated nor violent, they do not interact meaningfully with their environment beyond moving across the lattice.
Agitated agent	Do not behave differently from quiet agents, but are at higher risk of turning violent.
Violent agent	Avoid visible surveillance cameras and cops. Might destroy the former and incapacitate the latter when certain conditions are met.
Jailed agent	Remains idle for a specified number of ticks before becoming active again. By adjusting the jail term we can determine whether to model full-fledged arrests or only brief encounters with cops.
Policeman	Randomly move across the grid or stay on their patch to arrest violent agents, depending on the user settings. They do not meaningfully interact with surveillance devices.

jailed. This has no immediate influence on which state an agent will adopt once active again. As we are modeling a smart city scenario, two types of static surveillance cameras (visible and not visible) are randomly placed across the grid for each simulation. Every 60 ticks the camera identifies whether violent agents are within its range of vision. If that is the case, those violent agents are permanently removed from the simulation. Additionally, all centralized and decentralized forms of surveillance were included in the simulation (see Table 2 and description of the experiments below).

3.3. Experiments and Results

3.3.1. Effect of Surveillance Cameras

We wanted to investigate the effects of both visible and hidden cameras on the sum of agitated and violent agents divided by the total number of agents, which for this research we interpret as the civil violence rate, over the course of four hours (1,440 ticks). For this purpose, we run several experiments (Figures 1 and 2), modeled according to the following rules:

- Varying the number of mixed visible and hidden cameras (both within a range of [10,40]), thereby investigating combinations of visible and hidden cameras at different ratios;
- Varying the number of visible cameras within a range of [10,50] in steps of 10, without hidden cameras;
- Varying degrees of aggressive police interference in a range of [0.2, 0.8] in steps of 0.2 for a fixed number of visible and hidden cameras (both either 20 or 40);
- Varying radii of camera vision within a range of [1,4] in steps of 1 with a threshold of –0.25, and 10 visible and hidden cameras respectively;
- In all the cases, coveillance is set to 0.51, while selfveillance is set to 1.02, i.e., a low level for both. All cameras remove violent agents by default.

The effect of hidden cameras on civil violence results in a step function that reduces the civil violence rate linearly over time, which sometimes approaches zero by the end of the simulation. This can be interpreted through violent agents' behavior, who would avoid visible cameras but have no means of escaping hidden cameras. Therefore, they are removed from the simulation more frequently, leading to a sudden dip in civil violence, before it starts rising again. A higher number of cameras does not mean that civil violence can be suppressed more easily, especially in the very early stages. Instead, they seem to cause stronger fluctuations

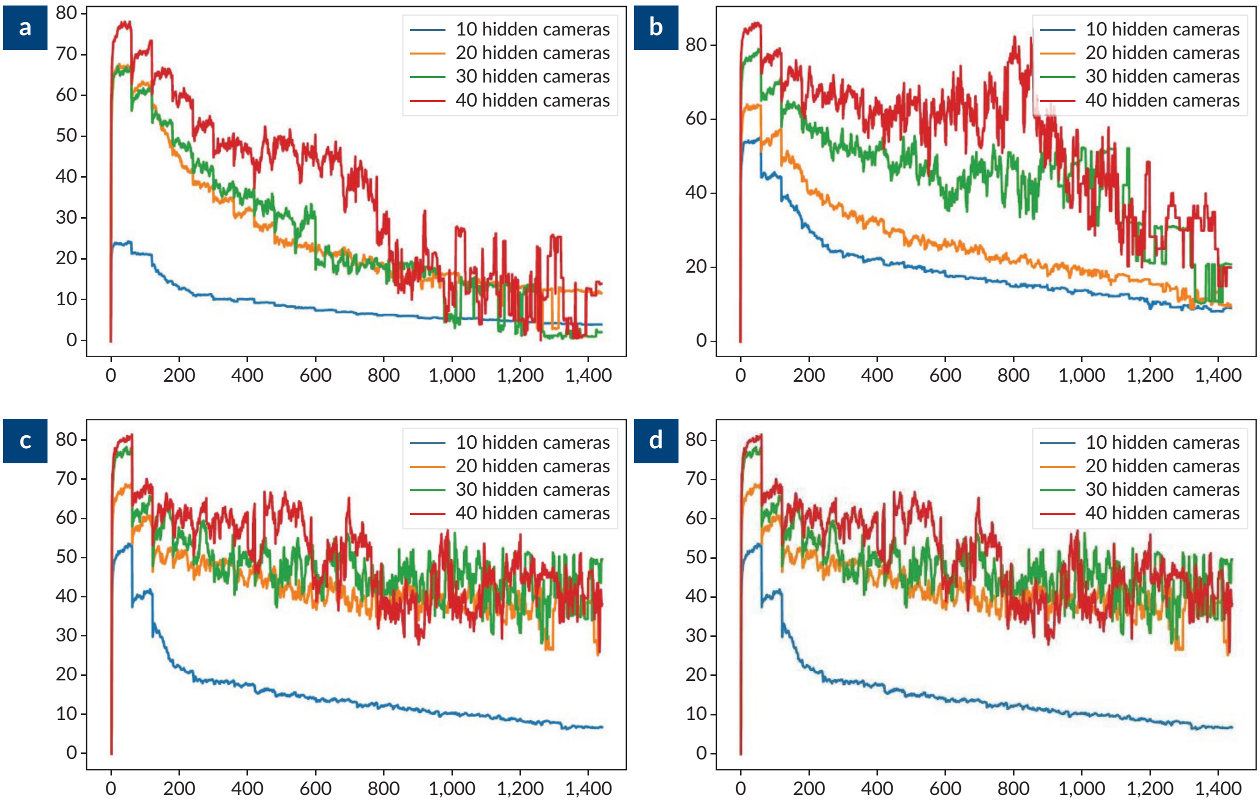

Figure 1. Effect of mixed camera types on the sum of agitated and violent agents divided by the total number of agents over the course of four hours with (a) 10 visible cameras, (b) 20 visible cameras, (c) 30 visible cameras, and (d) 40 visible cameras.

in the civil violence rate. As more agents turn violent due to the high surveillance level, more of them will be removed due to the cameras detecting them. In conclusion, it appears that a low number of hidden cameras are more effective at suppressing civil violence than visible cameras, though they also produce more general unrest, as the number of violent agents is in constant flux. Combining hidden and visible cameras demonstrates that, depending on the ratio, the characteristics of each type may dominate the cumulative effect. For instance, we observed that the higher the number of visible cameras, the lower the civil violence rate. At 40 visible cameras (Figure 1d), this rate seems to grow constant as the simulation pursues. Meanwhile, at 10 visible cameras (Figure 1a), the civil violence rate sinks monotonously as hidden cameras mostly dominate. According to the simulation, the most effective method appears to be a low number of hidden and visible cameras.

In terms of using exclusively visible cameras, the results show that the drop in civil violence rate is roughly 20–30% and the higher the number of visible cameras, the less effective they are (Figure 2a). (Smart) CCTV systems are widely studied in the literature and many authors advocate the little influence they have on public disorder offenses, becoming less effective over time and introducing significant challenges for the use of public spaces and individual rights (Fontes & Lütge, 2021; Fontes & Perrone, 2021; Kostka et al., 2021).

The involvement of police introduces a nigh oscillating behavior. If more policemen are involved, the civil violence rate decreases (Figures 2b and 2c). Indeed, a very large number of police officers, coupled with heavy surveillance, are necessary to keep the civil violence rate at a relatively low level (Figure 2c). When compared to the use of hidden cameras alone, the simulation shows that police involvement would render less effective.

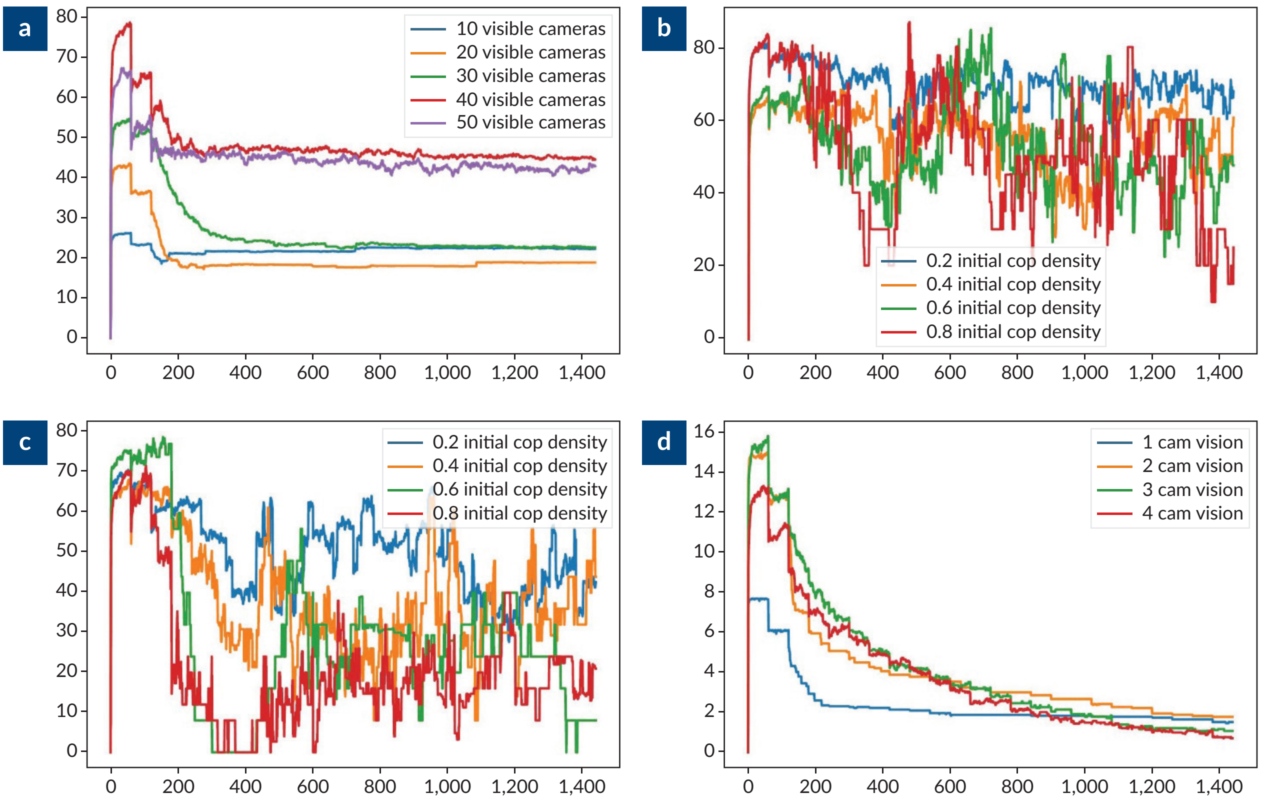

Figure 2. Effects of varying the conditions of the experiment: (a) effect of visible cameras on the sum of agitated and violent agents divided by the total number of agents over the course of four hours; (b) effect of police officers on the sum of agitated and violent agents divided by the total number of agents over the course of four hours for a fixed number of mixed type cameras with 20 visible and 20 hidden cameras; (c) effect of police officers on the sum of agitated and violent agents divided by the total number of agents over the course of four hours for a fixed number of mixed type cameras with 40 visible and 40 hidden cameras; and (d) effect of the cameras' range of vision on the sum of agitated and violent agents divided by the total number of agents over the course of four hours for 10 visible and 10 hidden cameras.

In terms of the range of vision, the more effective solution seems to be a smaller range of vision for a combination of hidden and visible cameras, although the results show that it is only significant for cam-vision = 1 (Figure 2d), implying that deterrence has little to do with the actual technical capabilities of the cameras. Indeed, CCTV systems do not even have to be operational in order to achieve a deterrence effect.

3.3.2. Effect of Coveillance

Additionally, we investigated the effect of coveillance (as defined in Table 2) on suppressing civil violence. For this purpose, we run several experiments (Figure 3), modeled according to the following rules:

- Varying the strength of coveillance within a range of [0,3] in steps of 1, to investigate the influence of coveillance on the number of violent agents;
- Varying the level of people inflow within a range of [1,4] in steps of 1 to observe the civil violence rate over time;
- Varying the agent vision radius within a range of [1,3] in steps of 1 to observe the civil violence rate over time;

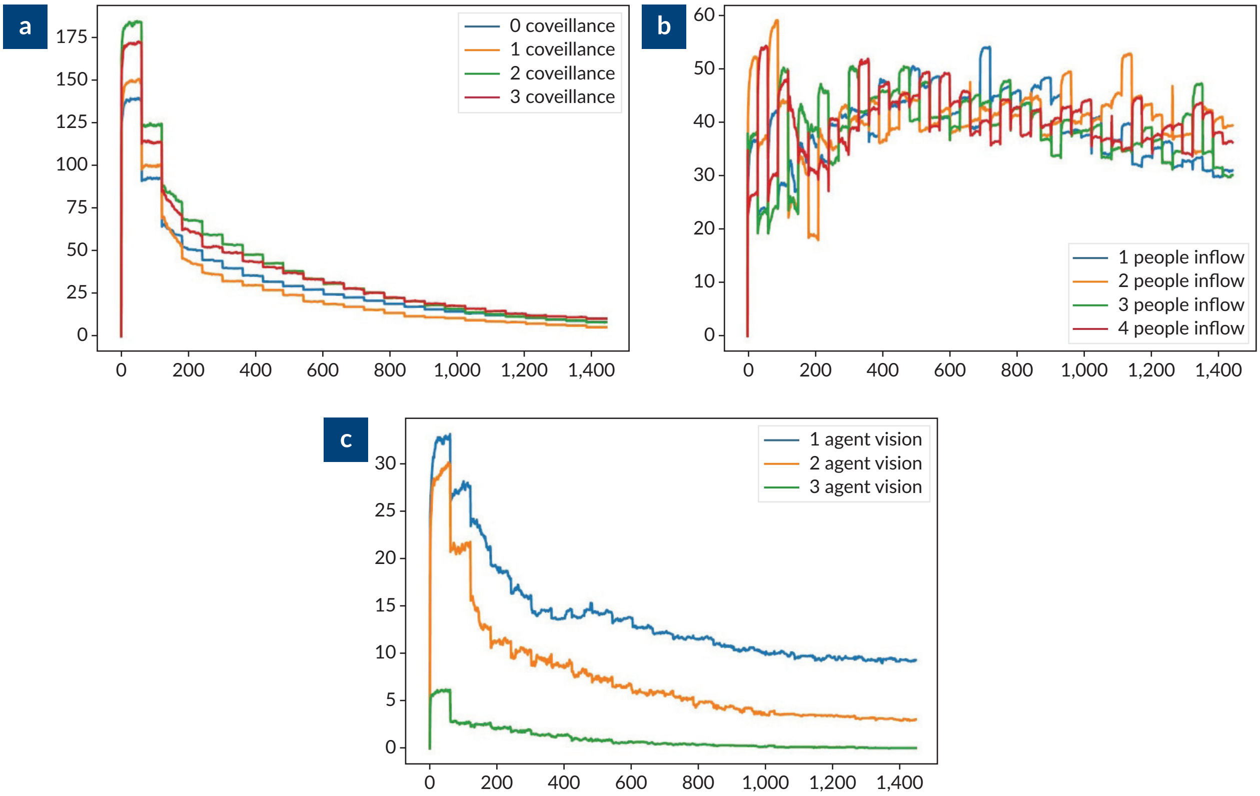

Figure 3. Effects of varying the conditions of the experiment: (a) effect of coveillance on the number of violent agents over time; (b) effect of people inflow on the civil violence rate over time; and (c) effect of the agents' range of vision on the civil violence rate over time.

- In all cases, we added 10 visible and 10 hidden cameras, building on the results of the previous experiments.

The conceptualization of coveillance is based on people inflow, i.e., how many new agents are entering the grid, and agent vision, i.e., an agent's range of vision, mimicking for the experiment the dynamics of public space and casual surveillance (see Table 2). The results show that coveillance can reduce the number of violent agents exponentially, although the employed strength makes little difference (Figure 3a). Figure 3b shows how increasing population inflow does not have an effect on the civil violence rate. However, a larger range of vision for agents seems effective at reducing it (Figure 3c). In the simulation, the agent's range of vision determines its range of movement within a single time step, which corresponds to the speed at which an agent moves across the grid. Therefore, the simulation leads us to conclude that quickly moving about and having a large range of vision is more effective than having a high number of agents. The results also refer to the importance of who is watching, as the range of vision was deemed insignificant for surveillance cameras but highly important for population agents.

3.3.3. Effect of Selfveillance Combined With Smartphone Interference

Next, we investigated the effect of selfveillance (as defined in Table 2) combined with smartphone interference. For this purpose, we run several experiments (Figure 4), modeled according to the following rules:

- Varying the strength of selfveillance within a range of [0,3] in steps of 1 to explore its effect on the number of violent agents;
- Varying the strength of selfveillance within a range of [0,3] in steps of 1, while setting phone jamming to values in a range of [0.1,0.3] in steps of 0.1. Since phone jamming tends to yield high, constant civil violence rates, we tested whether selfveillance is effective in the face of a particularly strong civil violence outbreak.

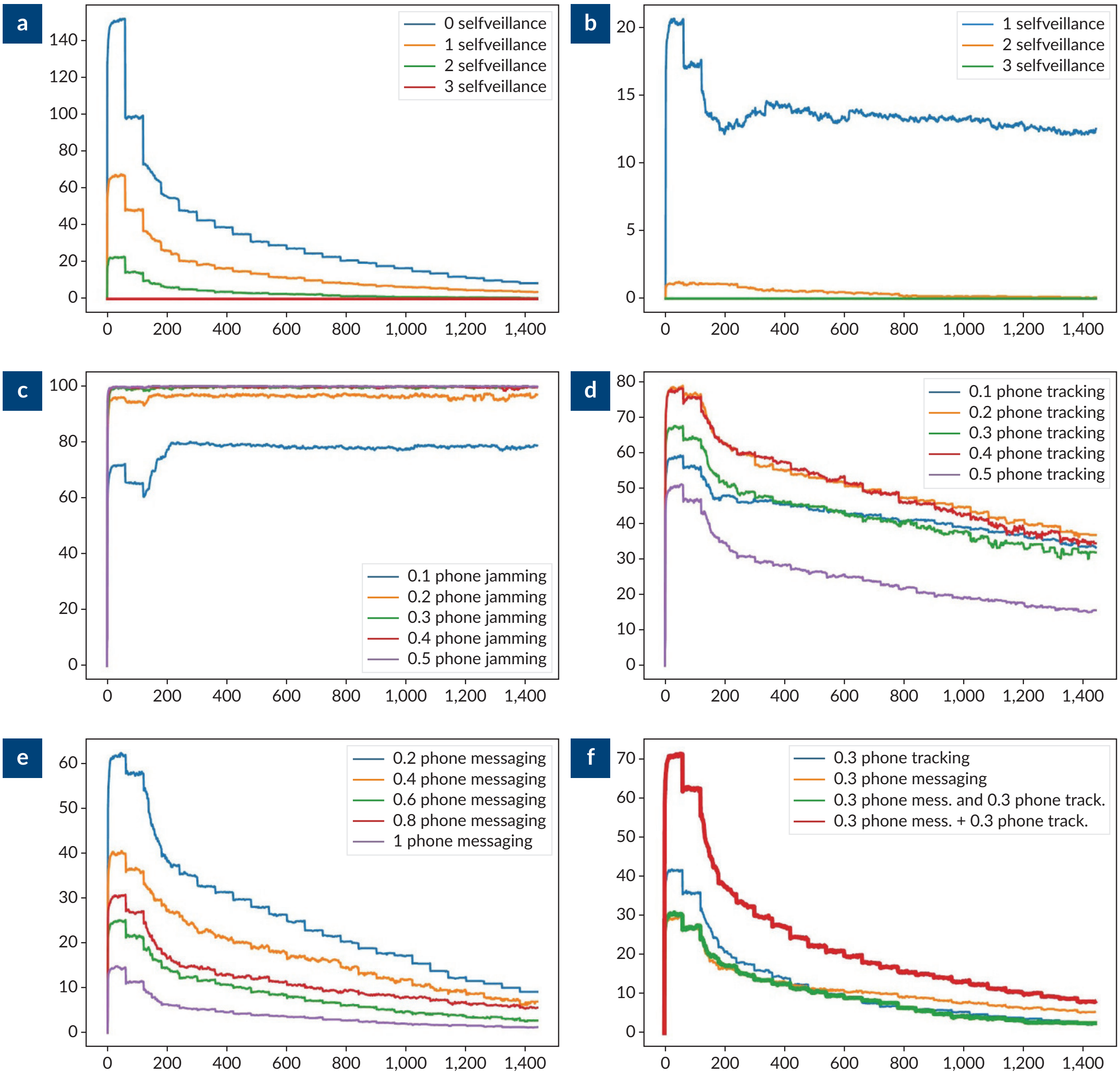

Figure 4. Effects of varying the conditions of the experiment: (a) effect of selfveillance on the number of violent agents over time; (b) effect of selfveillance in conjunction with 0.2 phone jamming on the civil violence rate over time; (c) effect of phone jamming on the civil violence rate over time; (d) effect of phone tracking on the civil violence rate over time; (e) effect of phone messaging on the civil violence rate over time; and (f) effect of phone tracking combined with phone messaging on the civil violence rate over time.

We also looked into the effects of smartphone interference alone by:

- Varying the level of phone surveillance within a range of [0,3] in steps of 1;
- Varying the level of phone tracking within a range of [0.1,0.5] in steps of 0.1;
- Varying the strength of phone jamming within a range of [0.1,0.5] in steps of 0.1;
- Varying the level of phone messaging within a range of [0.2,1.0] in steps of 0.2;
- Observing whether any synergetic effects appear when employing multiple phone-based actuators (i.e., phone messaging and phone tracking, both at level 0.3) at once;
- In all cases, we added 10 visible and 10 hidden cameras.

As the level of selfveillance increases, the results show an exponential decay of violent agents (Figure 4a), rendering selfveillance highly effective, similar to what we have seen with coveillance. The simulation also showed that 0.2 phone jamming and low selfveillance can significantly reduce the civil violence rate (Figure 4b). However, only phone jamming at 0.2 yields a 90% rate of civil violence (Figure 4c). If we increase the level of selfveillance, the rate of civil violence tends to zero (Figures 4a and 4b). Although unlikely that civil violence can ever be fully eradicated, the results indicate that selfveillance has great potential and its value might have been underestimated, at least until the recent event of the Covid-19 pandemic, as seen in the case of contact tracing apps (Fontes et al., 2022). Moreover, according to the results of the experiments, in comparison to other methods such as phone jamming, phone tracking, and phone messaging (as forms of centralized surveillance and nudging citizens), selfveillance can be more effective in suppressing civil violence (Figures 4c–4f).

We also explored whether a combination of phone-based actuation techniques could work in synergy, but results showed that they do not compound additively. For instance, employing both phone tracking and phone messaging does not yield better results than using the latter alone (Figure 4f).

4. Potentials and Limitations: A Socio-Technical Approach to ABMs for UDTs and Urban Surveillance

4.1. Key Takeaways of an ABM and Potential Integration Into a UDT

The results described in the previous section seem promising and encourage further research and testing in closer to real environments. Indeed, UDTs can provide the context data for the set-up environment of an ABM and together they can be employed to explore and experiment in a simulation environment before implementing policies and measures in practice.

However, while ABMs are useful for exploring emergent behavior and mainly valuable for understanding trends and relationships, they cannot be trusted for quantitative measurements. Additionally, complex ABMs may yield difficult and less reliable interpretations of cause–effect relationships. ABMs' potential to assist decision-makers lies in the enabled simulations and predictions, which can be complemented and tested using other methods and approaches (namely empirical and leveraging urban big data). As we have demonstrated, the use cases and experiments enabled by these models are numerous. Another advantage lies in the fact that simulation environments are considered safe spaces to conduct experiments before exposing communities to greater risks and potential harms.

According to this simulation's results, we can present the following takeaways, which could guide future research and inform policymaking.

- A low number of visible and hidden surveillance cameras are most effective at suppressing civil violence.
- Surveillance cameras do not require a high range of vision to be effective. Conversely, a heightened range of vision renders citizens more effective at reducing civil violence through coveillance.
- Selfveillance can be highly effective at suppressing civil violence.
- The simultaneous use of multiple phone-based actuators does not yield better results than the individual measures.

We emphasize that despite the promising results, embarking on intrusive surveillance and ubiquitous data collection in public spaces raises significant ethical concerns that should be addressed and assessed to ensure that such policies work in the best interest of impacted communities. The use of AI for law enforcement, as well as other critical activities that overlap with urban planning, and considering nudging as a means to govern societies, are not without a toll on individuals and communities. Below, we explore some of the implications.

4.2. Ethical Implications for Individuals and Society

Surveillance and tracking systems have been moving away from humans-in-the-loop towards fully automated intelligent systems associated with many socio-technical challenges that currently prevent a full realization of their potentials. For instance, robust computer vision cannot be taken for granted, while misclassifications of abnormal crowd behavior can have severe consequences (Wang et al., 2012). Additionally, a demand for real-time processing and communication coupled with a large amount of data yields high computational costs, requiring intensive resources and underscoring the importance of creating a general framework for distributed surveillance systems (Kumari et al., 2023; Valera & Velastin, 2005). Such technical limitations render highly autonomous surveillance systems infeasible for now. Nonetheless, researchers are identifying ways to navigate and overcome such problems and autonomous surveillance systems are quickly becoming more robust, reliable, and ubiquitous. In terms of actuation, it is worth mentioning that the smart city paradigm often considers humans as an integral part of the system (see for instance Fontes et al., 2024). When it comes to complex tasks such as surveillance and crowd control, the value of human input is not diminished by the automation of surveillance systems. As the simulation confirmed, nudging and self-nudging can be an integral part of social control assisted by technology-based systems. Additionally, there are social and cultural aspects in terms of how power structures interact with surveillance practices, which cannot be underestimated lest that efficiency is undermined by lack of trust and acceptance or lead to the erosion of democratic values (Fontes et al., 2022; Wood, 2009). Indeed, even in the scenario of intelligent autonomous surveillance systems, humans will somehow remain in the loop; the question is whether they are in control or controlled by technology.

In parallel, UDTs are evolving and integrating data from several sources. Indeed, ubiquitous surveillance is not only focused on physical spaces but also on digital social spaces. However, despite the advancements in crowd characterization using social media data (Duan et al., 2020; Gong et al., 2020), current UDTs are still mainly used to model physical space and a few specific activities taking place there (often related to mobility). The convergence between physical and cyberspace further complicates complex systems (Fontes & Dubey, in press), deeming the full realization of UDTs difficult to accomplish due to data gaps or the underestimation of the impact of social media on crowd behavior.

We also underscore that the social implications of autonomous surveillance systems for individuals and society are beyond technical feasibility and the overall performance limitations of a system. Surveillance is a means of bestowing power on the watcher over the watched. When conducted as mass surveillance, it can deepen existing or create new power imbalances, as exposed by the surveillance capitalism theory (Zuboff, 2023). Moreover, it allows profiling and identification of individuals and can lead to exploiting certain vulnerabilities while employing behavioral nudging techniques, sometimes hardly distinguishable from manipulative practices (Fontes et al., 2022).

Public authorities may have plausible arguments to justify mass surveillance, namely to suppress civil violence. According to some studies, people may actually accept being exposed to intrusive surveillance technologies in some cases (Kostka et al., 2021). Nonetheless, acceptance and trust in public authorities do not efface risks related to fairness, transparency, and proportionality. Individual consent might not prevent abusive use of personal data, if prerequisites such as individual autonomy and transparency are overlooked in the first place (Fontes et al., 2022). In extension, some of these systems may be beyond the rule of law (Ogasawara, 2022; Van Brakel, 2021), pose a threat to fundamental and human rights (Pauwels, 2020), or lead to the undermining of democratic values due to a potential chilling effect (Barkane, 2022; Selinger & Hartzog, 2020) and self-censorship (Büchi et al., 2022). They can also impact accessibility and the expectation of anonymity in public spaces, thereby promoting the exclusion of disproportionately affected vulnerable groups (Fontes & Lütge, 2021; Hirose, 2016). Mass and automated surveillance of individuals and modeling crowd behavior in a realistic and holistic manner may radically change the way society functions by trivializing the use of personal data, leading to a paradigm shift in the perception of privacy in relation to personal data (Fontes et al., 2024).

On the other hand, the empowerment of public authorities over citizens may result in coercing populations into accepting a restriction of rights, which resonates with how authoritarian regimes and autocratic forms of exercising power. Thus, mainstreaming surveillance governed by autonomous systems may widen the rift between citizens and public authorities. Raising awareness among public authorities and populations about the impacts of autonomous systems should be a priority, namely when discussing holistic forms of twinning reality through UDTs and human digital twins, which might as well become agents of ABMs not only inspired on real cities but run on their digital replicas (Fontes et al., 2024).

5. Conclusion

While the contextual data UDTs aggregate could be leveraged to inform more realistic ABMs for simulations on urban surveillance and security, consequently yielding more sophisticated interpretations to support urban planning and management, the implications of mass surveillance are serious for individuals and society. Therefore, instead of being lured into creating a holistic model of the real world with the purpose of rendering all events predictable and undesirable outcomes preventable, we should question how technology is replacing humans when it comes to interpreting the world. We should investigate how such models can incept values in communities that may be in tension with existing values and norms. Moreover, surveillance invokes a paternalistic conception of caretaking and being taken care of. If UDTs will have the ability of transforming physical space to manage crowds and urban life, we need to ask on behalf of whom. Democratic systems entail changes in the structures of power and social mobility.

As demonstrated, simulations can hint at ways of leveraging surveillance in many forms to nudge behavior and control crowds. However, the assumption that autonomous surveillance systems can be run through simulations represents a turning point for society. There is the need for more research and impact assessments in order to find a balance at the intersection of simulation-based environments and holistic data models in relation to the automation of decisions affecting real cities and citizens.

Acknowledgments

This research results from a collaboration between the first author and the Institute for Ethics in Artificial Intelligence (Technical University of Munich) and builds on the first author's master thesis.

Funding

Contributions by the second author were provided in relation to the project Ethics for the Smart City, funded by the Institute for Ethics in Artificial Intelligence (Technical University of Munich).

Conflict of Interests

The authors declare no conflict of interests.

Data Availability

Data can be made available upon request.

Supplementary Material

Supplementary material for this article is available online in the format provided by the authors (unedited).

References

Arshi, O., & Mondal, S. (2023). Advancements in sensors and actuators technologies for smart cities: A comprehensive review. *Smart Construction and Sustainable Cities*, *1*(1), Article 18.

Barkane, I. (2022). Questioning the EU proposal for an Artificial Intelligence Act: The need for prohibitions and a stricter approach to biometric surveillance. *Information Polity*, *27*(2), 147–162.

Batty, M. (2018). Artificial intelligence and smart cities. *Environment and Planning B: Urban Analytics and City Science*, *45*(1), 3–6.

Batty, M., Axhausen, K. W., Giannotti, F., Pozdnoukhov, A., Bazzani, A., Wachowicz, M., Ouzounis, G., & Portugali, Y. (2012). Smart cities of the future. *The European Physical Journal Special Topics*, *214*, 481–518.

Benartzi, S., Beshears, J., Milkman, K. L., Sunstein, C. R., Thaler, R. H., Shankar, M., Tucker-Ray, W., Congdon, W. J., & Galing, S. (2017). Should governments invest more in nudging? *Psychological Science*, *28*(8), 1041–1055.

Bentham, J. (1791). *Panopticon: Or, the inspection-house. Containing the idea of a new principle of construction applicable to any sort of establishment, in which persons of any description are to be kept under inspection.* Thomas Byrne.

Borean, C., Giannantonio, R., Mamei, M., Mana, D., Sassi, A., & Zambonelli, F. (2015). Urban crowd steering: An overview. In G. Di Fatta, G. Fortino, W. Li, M. Pathan, F. Stahl, & A. Guerrieri (Eds.), *Internet and Distributed Computing Systems: 8th International Conference, IDCS 2015, Windsor, UK, September 2–4, 2015. Proceedings* (pp. 143–154). Springer.

Büchi, M., Festic, N., & Latzer, M. (2022). The chilling effects of digital dataveillance: A theoretical model and an empirical research agenda. *Big Data & Society*, *9*(1). https://doi.org/10.1177/20539517211065368

Caldarelli, G., Arcaute, E., Barthelemy, M., Batty, M., Gershenson, C., Helbing, D., Mancuso, S., Moreno, Y., Ramasco, J. J., Rozenblat, C., Sánchez, A., & Fernández-Villacañas, J. L. (2023). The role of complexity for digital twins of cities. *Nature Computational Science*, *3*(5), 374–381.

Deleuze, G. (2017). Postscript on the societies of control. In D. Wilson & C. Norris (Eds.), *Surveillance, crime and social control* (pp. 35–39). Routledge.

Duan, J., Zhai, W., & Cheng, C. (2020). Crowd detection in mass gatherings based on social media data: A case study of the 2014 Shanghai New Year's Eve stampede. *International Journal of Environmental Research and Public Health*, *17*(22), Article 8640.

Elharrouss, O., Almaadeed, N., & Al-Maadeed, S. (2021). A review of video surveillance systems. *Journal of Visual Communication and Image Representation*, *77*, Article 103116.

Epstein, J. M. (2002). Modeling civil violence: An agent-based computational approach. *Proceedings of the National Academy of Sciences*, *99*(Suppl. 3), 7243–7250.

Espejo, R. (2014). Cybernetics of governance: The Cybersyn project 1971–1973. In G. S. Metcalf (Eds.), *Social systems and design* (pp. 71–90). Springer.

Fonoberova, M., Fonoberov, V. A., Mezic, I., Mezic, J., & Brantingham, P. J. (2012). Nonlinear dynamics of crime and violence in urban settings. *Journal of Artificial Societies and Social Simulation*, *15*(1), Article 2.

Fontes, C., Carpentras, D., & Mahajan, S. (2024). Human digital twins unlocking Society 5.0? Approaches, emerging risks and disruptions. *Ethics and Information Technology*, *26*, Article 54.

Fontes, C., Corrigan, C., & Lütge, C. (2023). Governing AI during a pandemic crisis: Initiatives at the EU level. *Technology in Society*, *72*, Article 102204.

Fontes, C., & Dubey, R. K. (in press). *Urban futures: Possibilities and challenges for ethical virtual cities* (chapter of an handbook on digital ethics). Edward Elgar.

Fontes, C., Hohma, E., Corrigan, C. C., & Lütge, C. (2022). AI-powered public surveillance systems: Why we (might) need them and how we want them. *Technology in Society*, *71*, Article 102137.

Fontes, C., & Lütge, C. (2021). Surveillance and power relations. The use of facial recognition technologies and remote biometric identification in public spaces and impacts on public life. *Revista de Direito Público*, *18*(100), 91–116.

Fontes, C., & Perrone, C. (2021). *Ethics of surveillance: harnessing the use of live facial recognition technologies in public spaces for law enforcement*. Technical University of Munich.

Foucault, M. (1977). *Discipline and punish: The birth of the prison*. Allen Lane.

Franke, T., Lukowicz, P., & Blanke, U. (2015). Smart crowds in smart cities: Real life, city scale deployments of a smartphone based participatory crowd management platform. *Journal of Internet Services and Applications*, *6*, Article 27.

Giddens, A. (1986). The nation-state and violence. *Capital & Class*, *10*(2), 216–220.

Goh, C. K., Quek, H., Tan, K. C., & Abbass, H. A. (2006). Modeling civil violence: An evolutionary multi-agent, game theoretic approach. In *2006 IEEE International Conference on Evolutionary Computation* (pp. 1624–1631). IEEE.

Gong, V. X., Daamen, W., Bozzon, A., & Hoogendoorn, S. P. (2020). Crowd characterization for crowd management using social media data in city events. *Travel Behaviour and Society*, *20*, 192–212.

Haggerty, K. D., & Ericson, R. V. (2017). The surveillant assemblage. In D. Wilson & C. Norris (Eds.), *Surveillance, crime and social control* (pp. 61–78). Routledge.

Han, B. C. (2015). *The burnout society*. Stanford University Press.

Hempel, L., & Töpfer, E. (2009). The surveillance consensus: Reviewing the politics of CCTV in three European countries. *European Journal of Criminology*, *6*(2), 157–177.

Hirose, M. (2016). Privacy in public spaces: The reasonable expectation of privacy against the dragnet use of facial recognition technology. *Connecticut Law Review*, *49*(5), 1591–1620.

Holston, J. (1989). *The modernist city: An anthropological critique of Brasília*. University of Chicago Press.

Ibrahim, S. W. (2016). A comprehensive review on intelligent surveillance systems. *Communications in Science and Technology*, *1*(1), 7–14.

Jacobs, J. (1961). *The death and life of great American cities*. Vintage.

Khajavi, S. H., Motlagh, N. H., Jaribion, A., Werner, L. C., & Holmström, J. (2019). Digital twin: Vision, benefits, boundaries, and creation for buildings. *IEEE Access*, *7*, 147406–147419.

Kitchin, R. (2017). Data-driven urbanism. In R. Kitchin, T. P. Lauriault, & G. McArdle (Eds.), *Data and the city* (pp. 44–56). Routledge.

Kostka, G., Steinacker, L., & Meckel, M. (2021). Between security and convenience: Facial recognition technology in the eyes of citizens in China, Germany, the United Kingdom, and the United States. *Public Understanding of Science*, *30*(6), 671–690.

Kumari, B., Jaiswal, R., Kumar, K., Mohit, R., & Singh, K. (2023). Automated CCTV surveillance system. *Journal of Optoelectronics and Communication*, *5*(1), 36–41.

Kurland, J., & Chen, P. (2016). Simulating civil disorder: An agent-based modeling approach. *Oriental Journal of Computer Science and Technology*, *9*(3), 153–164.

Lee, J. S., & Hoh, B. (2010). Dynamic pricing incentive for participatory sensing. *Pervasive and Mobile Computing*, *6*(6), 693–708.

Leistert, O. (2012). Resistance against cyber-surveillance within social movements and how surveillance adapts. *Surveillance & Society*, *9*(4), 441–456.

Lemos, C., Coelho, H., & Lopes, R. J. (2013). Agent-based modeling of social conflict, civil violence and revolution: State-of-the-art-review and further prospects. In *Proceedings of the 11th European Workshop on Multi-Agent Systems (EUMAS 2013)*. CEUR-WS.

Liu, W., Mei, Y., Ma, Y., Wang, W., Hu, F., & Xu, D. (2022). City Brain: A new model of urban governance. In M. Li, G. Bohács, A. Huang, D. Chang, & X. Shang (Eds.), *IEIS 2021: Proceedings of 8th International Conference on Industrial Economics System and Industrial Security Engineering* (pp. 107–115). Springer Nature.

Mathiesen, T. (2017). The viewer society: Michel Foucault's 'panopticon'revisited. In D. Wilson & C. Norris (Eds.), *Surveillance, crime and social control* (pp. 41–60). Routledge.

Medina, E. (2011). *Cybernetic revolutionaries: Technology and politics in Allende's Chile*. MIT Press.

Norris, C., & Armstrong, G. (2017). CCTV and the social structuring of surveillance. In D. Wilson & C. Norris (Eds.), *Surveillance, crime and social control* (pp. 81–102). Routledge.

Ogasawara, M. (2022). Legalizing illegal mass surveillance: A transnational perspective on Canada's legislative response to the expansion of security intelligence. *Canadian Journal of Law and Society/La Revue Canadienne Droit et Société*, *37*(2), 317–338.

Pauwels, E. (2020). *Artificial Intelligence and data capture technologies in violence and conflict prevention*. Global Centre on Cooperative Security.

Pompigna, A., & Mauro, R. (2022). Smart roads: A state of the art of highways innovations in the Smart Age. *Engineering Science and Technology, an International Journal*, *25*, Article 100986.

Railsback, S. F. (2019). *Agent-based and individual-based modeling: A practical introduction* (2nd ed.). Princeton University Press.

Rana, R. K., Chou, C. T., Kanhere, S. S., Bulusu, N., & Hu, W. (2010). Ear-phone: An end-to-end participatory urban noise mapping system. In T. Abdelzaher, T. Voigt, & A. Wolisz (Eds.), *IPSN '10: Proceedings of the 9th ACM/IEEE International Conference on Information Processing in Sensor Networks* (pp. 105–116). Association for Computing Machinery.

Sadowski, J., & Pasquale, F. A. (2015). *The spectrum of control: A social theory of the smart city* (SSRN Scholarly Paper No. ID 2653860). https://papers.ssrn.com/abstract.2653860

Selinger, E., & Hartzog, W. (2020). The inconsentability of facial surveillance. *Loyola Law Review*, *66*, 33–54.

Shetty, S., Shetty, S., Vishwakarma, V., & Patil, S. (2020). Review paper on door lock security systems. In *2020 International Conference on Convergence to Digital World – Quo Vadis (ICCDW)*. IEEE. https://ieeexplore.ieee.org/document/9318636

Singla, A., Santoni, M., Bartók, G., Mukerji, P., Meenen, M., & Krause, A. (2015). Incentivizing users for balancing bike sharing systems. In *Proceedings of the AAAI Conference on Artificial Intelligence* (Vol. 29, No. 1, pp. 723–729). AAAI.

Sreenu, G. S. D. M. A., & Durai, S. (2019). Intelligent video surveillance: A review through deep learning techniques for crowd analysis. *Journal of Big Data*, *6*(1), Article 48.

Tewksbury, D. (2012). Crowdsourcing homeland security: The Texas virtual borderwatch and participatory citizenship. *Surveillance & Society*, *10*(3/4), 249–262.

Ullrich, P., & Knopp, P. (2018). Protesters' reactions to video surveillance of demonstrations: Counter-moves, security cultures, and the spiral of surveillance and counter-surveillance. *Surveillance & Society*, *16*(2), 183–202.

UN General Assembly. (1948). *Universal declaration of human rights*.

Valera, M., & Velastin, S. A. (2005). Intelligent distributed surveillance systems: A review. *IEE Proceedings – Vision, Image and Signal Processing*, *152*(2), 192–204.

Van Brakel, R. (2021). How to watch the watchers? Democratic oversight of algorithmic police surveillance in Belgium. *Surveillance & Society*, *19*(2), 228–240.

Wang, B., Ye, M., Li, X., Zhao, F., & Ding, J. (2012). Abnormal crowd behavior detection using high-frequency and spatio-temporal features. *Machine Vision and Applications*, *23*, 501–511.

Wood, D. M. (2009). The surveillance society' questions of history, place and culture. *European Journal of Criminology*, *6*(2), 179–194.

Zhang, J., Hua, X. S., Huang, J., Shen, X., Chen, J., Zhou, Q., Fu, Z., & Zhao, Y. (2019). City brain: Practice of large-scale artificial intelligence in the real world. *IET Smart Cities*, *1*(1), 28–37.

Zhuang, Y., Lin, F., Yoo, E. H., & Xu, W. (2015). Airsense: A portable context-sensing device for personal air quality monitoring. In E. Baccelli, H. Ghasemzadeh, G. Marfia, & K. Venkatasubramanian (Eds.), *MobileHealth '15: Proceedings of the 2015 Workshop on Pervasive Wireless Healthcare* (pp. 17–22). Association for Computing Machinery.

Zuboff, S. (2023). The age of surveillance capitalism. In W. Longhofer & D. Winchester (Eds.), *Social theory re-wired* (pp. 203–213). Routledge.

About the Authors

Sarah Shtaierman studied electrical engineering and information technology at the Technical University of Munich and the University of Waterloo (BSc and MSc), graduating in 2024. She is currently pursuing further studies at the Munich School of Philosophy. Her research interests include the impact of socio-technical systems on pre-existing inequalities and power dynamics.

Catarina Fontes is a postdoctoral researcher at the Technical University of Munich – Institute for Ethics in Artificial Intelligence. Her current research project Ethics for the Smart City looks at risks, benefits, and governance aspects to inform value-aligned socio-technical frameworks aimed to assist cities in the ethical implementation of AI-enabled systems with a focus on Smart City agendas. Catarina Fontes holds a PhD in urban studies from the Nova University of Lisbon and the University Institute of Lisbon.

Christoph Lütge is full professor of business ethics at Technical University of Munich (TUM) and the director of the TUM Institute for Ethics in Artificial Intelligence. Christoph Lütge is distinguished visiting professor at Tokyo University and has held further visiting positions at Harvard (Berkman Klein Center), Taipei, and Kyoto, and has been a member of the European AI ethics initiative AI4People and of the German Ethics Commission on Automated and Connected Driving (2016–2017). Since 2020, he is consortium leader of the Global AI Ethics Consortium.