

From concrete to community: How synthetic data can make urban digital twins more humane

Wei Zhai, Associate Professor of Public Affairs and Planning, University of Texas at Arlington

Published: November 25, 2025 8:26am EDT



How people behave is a critical element of how cities function.

Ahmed Deeb/picture alliance via Getty Images

When city leaders talk about making a town “smart,” they’re usually talking about urban digital twins. These are essentially high-tech, 3D computer models of cities. They are filled with data about buildings, roads and utilities. Built using precision tools like cameras and LiDAR – light detection and ranging – scanners, these twins are great at showing what a city looks like physically.

But in their rush to map the concrete, researchers, software developers and city planners have missed the most dynamic part of urban life: people. People move, live and interact inside those buildings and on those streets.

This omission creates a serious problem. While an urban digital twin may perfectly replicate the buildings and infrastructure, it often ignores how people use the parks, walk on the sidewalks, or find their way to the bus. This is an incomplete picture; it cannot truly help solve complex urban challenges or guide fair development.

To overcome this problem, digital twins will need to widen their focus beyond physical objects and incorporate realistic human behaviors. Though there is ample data about a city's inhabitants, using it poses a significant privacy risk. I'm a public affairs and planning scholar. My colleagues and I believe the solution to producing more complete urban digital twins is to use synthetic data that closely approximates real people's data.“

The privacy barrier

To build a humane, inclusive digital twin, it's critical to include detailed data on how people behave. And the model should represent the diversity of a city's population, including families with young children, disabled residents and retirees. Unfortunately, relying solely on real-world data is impractical and ethically challenging.

The primary obstacles are significant, starting with strict privacy laws. Rules such as the European Union's General Data Protection Regulation, or GDPR, often prevent researchers and others from widely sharing sensitive personal information. This wall of privacy stops researchers from easily comparing results and limits our ability to learn from past studies.

Furthermore, real-world data is often unfair. Data collection tends to be uneven, missing large groups of people. Training a computer model using data where low-income neighborhoods have sparse sensor coverage means the model will simply repeat and even magnify that original unfairness. To compensate for this, researchers can use the statistical technique of weighting the data in the models to make up for the underrepresentation.

Synthetic data offers a practical solution. It is artificial information generated by computers that mimics the statistical patterns of real-world data. This protects privacy while filling critical data gaps.

Synthetic data: Tool for fairer cities

Adding synthetic human dynamics fundamentally changes digital twins. It shifts them from static models of infrastructure to dynamic simulations that show how people live in the city. By generating synthetic patterns of walking, bus riding and public space use, planners can include a wider, more inclusive range of human actions in the models.

For example, Bogotá, Colombia, is using a digital twin to model its TransMilenio bus rapid transit system. Instead of relying only on limited or privacy-sensitive real-world sensor data, the city planners generated synthetic data to fill the digital twin. Such data artificially creates millions of simulated bus arrivals, vehicle speeds and queue lengths, all based on the statistical patterns – peak times, off-peak times – of actual TransMilenio operations.

This approach transforms urban planning in several crucial ways, making simulations more realistic and diverse. For example, planners can use synthetic pedestrian data to model how elderly and disabled residents would navigate a new urban design.

It also allows for risk-free testing of ideas. Planners can simulate diverse synthetic populations to see how a new flood evacuation plan would affect various groups, all without risking anyone's safety or privacy in the real world.

Making digital twins trustworthy

For all the promises of synthetic data, it can only be helpful if planners can trust it. Since they base major decisions on these virtual worlds, the synthetic data must be proved to be a reliable replacement for real-world data. Planners can test this by checking to see if the main policy decisions they reach using the synthetic data are the same ones they would have made using real-world data that puts people's privacy at risk. If the decisions match, the synthetic data is trustworthy enough to use for that planning task going forward.

Beyond technical checks, it's important to consider fairness. This means routinely auditing the synthetic models to check for any hidden biases or underrepresentation across different groups. For example, planners can make sure an emergency evacuation plan in the urban digital twin works for elderly residents with mobility issues.

Most importantly, I believe planners should include their communities. Establishing citizen advisory boards and designing the synthetic data and simulation scenarios directly with the people who live in the city helps ensure that their experiences are accurately reflected.

By moving beyond static infrastructure to dynamic environments that include people's behavior, synthetic data is set to play a critical role in urban planning. It will shape the resilient, inclusive and human-centered urban digital twins of the future.

Wei Zhai receives funding from National Science Foundation.

This article is republished from The Conversation under a Creative Commons license.