Street Art & Urban Creativity | Vol. 11, No. 7, 2025 | ISSN 2183-9956
Street Art & Urban Creativity. Scientific Journal
https://doi.org/10.62161/sauc.v11.5978

BARRIERS TO THE EFFECTIVE IMPLEMENTATION OF AI IN PREDICTING PUBLIC WORKS

ELENA USOBIAGA[1], PATRICIA MOLINA-COSTA[1], BEGOÑA ISPIZUA [1], JOSE LUIS IZKARA [2]

[1] TECNALIA, BASQUE RESEARCH AND TECHNOLOGY ALLIANCE (BRTA), ESPAÑA
[2] UNIVERSIDAD DE DEUSTO, SPAIN

KEYWORDS
Artificial Intelligence (AI)
Urban Planning
Public Works
Prediction
Public Data Quality
Data Sharing
Smart Cities

ABSTRACT
Public data quality issues often hinder AI applications in urban planning, affecting model applicability, effectiveness, and results. A case study on predicting public works impacted by network infrastructure and its city-wide impact is presented in this paper. This case has served to identify barriers and limitations to AI adoption in this subdomain, allowing to inform a set of recommendations to improve public data production and sharing, paving the way for future AI modelling in public works prediction. By addressing these challenges, cities can unlock the full potential of AI-driven urban planning and decision-making.

RECEIVED: 06 / 08 / 2025
ACCEPTED: 31 / 10 / 2025

1. Introduction

Artificial intelligence (AI) is a key technology that is becoming increasingly relevant in urban environments. The rise of local digital twins (LDT) together with AI opens up new opportunities for experimentation and prediction in the field of urban planning (Villanueva-Merino et al., 2024). Indeed, the integration of AI into urban planning has led to the emergence of "smart cities" —urban environments that leverage AI's capacity to analyse vast amounts of data and support informed decision-making. Technological advancements have fueled the evolution of intelligent urban systems, and a shift is now underway toward self-learning, self-improving cities (Wu, 2025).

Cities face significant challenges in implementing AI due to the lack of integration between departments, resulting in scattered or incomplete information that is often siloed within different departments, hindering efficient data sharing and coordination (Urrutia-Azcona et al., 2021). Moreover, in the field of urban planning, accessing high-quality, disaggregated data with relevant metadata remains a substantial challenge (Urra-Uriarte et al., 2024). To overcome these hurdles, it is essential to adopt a centralised data-sharing vision that prioritises accurate and actionable data, aligned with user needs and requirements (Villanueva-Merino et al., 2024).

This article specifically focuses on the possibilities and limitations of applying AI to predict works in public spaces by combining public data with private data on water, gas, electricity, telecommunications, and municipal network infrastructure, intending to minimize potential incidents or disruptions caused by public works (such as, Prolonged outages in the supply of electricity, water, or fiber optic networks, as well as disruptions in gas distribution systems, pose significant safety risks and operational challenges).

In particular, this research uses private data from Inkolan, an organisation that provides digital information from urban networks and service providers. This organisation operates in Spain, so the scope of application is framed within this country. After analysing the availability of public data in some municipalities, the case study focuses on the city of Madrid, working with open data on completed public works. The case study allows us to address the difficulties encountered in applying AI models in the field of public works prediction in public spaces, with the goal of extracting conclusions and recommendations for improvement that facilitate the production of high-quality data to improve the applicability of future AI models in this subdomain of urban planning. The structure of the article includes a bibliographic review of the state of the art, a description of the case study and the methodology employed, an analysis of the results obtained, and concludes with a discussion section that addresses some recommendations for improvement, as well as general conclusions.

2. State of the art

2.1. Data requirements for AI application

The exponential growth in the capabilities and applications of AI in recent years can be attributed to the convergence of three key factors. First, the rapid increase in processing power,partly driven by the principles outlined in Moore's Law (Garg, 2021; Moore, 2006). Second, the proliferation of digital data from multiple sources (Gao et al., 2024; Liu et al., 2018). Finally, the development of advanced machine learning algorithms, particularly deep learning models based on artificial neural networks with billions of interconnected parameters (Damioli et al., 2025; Lu et al., 2018). Together, these factors have transformed AI into a general-purpose technology with broad applications.

Data is the essential element for the functioning of artificial intelligence systems. Machine learning models are trained, validated, and optimised through data. The quality, quantity, and diversity of that data directly influence the performance, accuracy, and generalizability of AI applications. Numerous studies have proposed different data quality dimensions depending on their evaluation objectives and the stages of the dataset lifecycle (Gong et al., 2023). A practical and widely used model considers five key dimensions: accuracy, relevance, completeness, timeliness, and consistency (Fernández Álvarez & Garaizar Sagarminaga, 2024). Accuracy implies that data must faithfully represent the phenomena they describe; relevance requires data to be pertinent to the problem under study; completeness ensures that critical records are not missing; timeliness refers to the availability of up-to-date data; and consistency guarantees coherence across sources and over time. Deficiencies in any of these dimensions can lead to less accurate, less robust, and less generalizable models, whereas systematic improvements in data quality enhance the reliability and value of AI solutions (Mohammed et al., 2025). For this reason, active data quality management is considered a foundational pillar for the development of trustworthy and effective artificial intelligence.

These principles become especially critical in complex domains such as urban planning, which is inherently multidimensional, long-term and involves different dimensions and multiple stakeholders, including public administrations, citizens, and private agents. The urban planning process is dynamic and context-dependent, requiring heterogeneous datasets such as geospatial and cadastral information, socio-economic and environmental indicators, sensor data from IoT infrastructures, satellite imagery, and inputs from citizen participation (e.g., surveys or social media). The integration and quality assurance of such varied sources are critical for the successful deployment of AI-based urban solutions.

2.2. AI-driven approaches for urban planning

AI is increasingly transforming urban planning, particularly in the design and management of public spaces. AI tools and algorithms are being used to analyse large datasets, predict urban trends, optimise land use, and improve the quality and accessibility of public spaces. These technologies help planners make more informed, efficient, and creative decisions, but also raise new challenges related to ethics, data governance, and community engagement.

In this field, tools such as machine learning and deep learning are widely applied for pattern recognition, classification, and prediction of urban growth, land use, and environmental impacts. Algorithms such as random forests, convolutional neural networks (CNNs), and generative adversarial networks (GANs) are commonly used to analyse earth observation data and simulate urban patterns (Anwar & Sakti, 2024; Chaturvedi & de Vries, 2021; He & Chen, 2024; Park et al., 2023).Additionally, AI-integrated Geographic Information Systems (GIS) allow for enhanced spatial analysis, supporting the optimisation of public infrastructure, mobility, and microclimatic conditions in urban spaces (Anwar & Sakti, 2024; Korobeinikova et al., 2024).

AI has also been used to address age-related issues, such as creating age-friendly urban environments. Data-driven models, including deep convolutional neural networks, have been used to evaluate deficiencies in urban design and propose improvements (Delgado-Enales et al., 2022). AI-based urban planning tools, such as the Age-Friendly Index Simulator (AFIS) integrates image recognition, machine learning and the hybridization of metaheuristics and simulation-based optimisation to identify accessibility issues and propose effective solutions (Delgado-Enales et al., 2022; Villanueva-Merino et al., 2024). Other planning and decision support systems combine objective data with stakeholder participation, enabling policymakers to evaluate interventions with greater transparency and legitimacy towards sustainable urban planning (Delgado Enales, 2024; Park et al., 2023; Shulajkovska et al., 2024). Furthermore, AI-driven design tools, such as neural networks and clustering algorithms, can generate personalised design schemes, such as colour palettes for public spaces, tailored to local culture, climate, and user preferences (Xiao Ting Cheng et al., 2025).

However, the adoption of AI in urban planning also faces significant challenges. Technical limitations include scarce data infrastructure, complex models' requirements, and data quality. Organizational barriers such as resistance to change, lack of interdisciplinary collaboration, and insufficient expertise and training among planners (Bibri et al., 2024; Koutra & Ioakimidis, 2023; Rjab et al., 2023). Critical literature emphasizes that Big Data and IA algorithms often function as "weapons of math destruction" (O'Neil, 2016) as their opacity and the structural biases inherent in their datasets (such as race, gender, or age) amplify social inequalities (D'Ignazio & Klein, 2020). This issue also applies to human-in-the-loop systems, where evidence shows that exposure to biased algorithmic decisions can modify and consolidate human judgment, converting supervision into a channel that perpetuates and transfers the AI's own biases instead of correcting them (Agudo et al., 2024; Vicente & Matute, 2023). To mitigate these risks, explainability, governance and transparency are critical. Explainable AI (XAI) methods allow humans to interpret the functioning and impact of models, providing insight into their logic, limitations, and potential biases. Given the public relevance of urban planning, ensuring the interpretability and accountability of AI systems is essential for citizen trust (Molina-Costa, 2024).

In conclusion, AI offers opportunities to improve efficiency, personalisation, and community engagement in urban planning. AI can automate routine tasks, improve forecasting, and optimise spatial planning, leading to more sustainable and resilient cities (He & Chen, 2024; Mashhood et al., 2023; Zheng et al., 2023). However, it is essential to address the challenges and limitations associated with its adoption. Investing in data infrastructure, fostering interdisciplinary collaboration, and establishing ethical and transparent governance frameworks are key to maximising the opportunities offered by AI in urban planning.

A specific application where AI can be particularly impactful is in the coordination of public works. Public works have an impact on the economic activities in the affected area, as access to businesses becomes difficult. This also causes inconvenience to citizens, who experience reduced accessibility, less available public space, and disturbances in air and noise quality (Correia & Roseland, 2022). In the specific case of public works, the prediction of this type of activity can help improve their coordination (Hamann et al., 2023; Pericault et al., 2023). This coordination is key to increasing efficiency, aligning the various stakeholders, and reducing resource usage by avoiding repeated openings of the same street at different times.

3. Materials and methods

3.1. Use case description

The main objective of this case study was to explore the possibilities of AI to offer a new approach to predicting and impacting public space works based on the combination of public data from municipalities and private data offered by Inkolan. Inkolan is an organisation made up of most of Spain's major public service operators, which facilitates the supply of digital information on water, gas, electricity, telecommunications and municipal network infrastructure.

This information is key to carrying out works in public spaces to avoid breaks and incidents in service networks. As part of its activity, Inkolan has over 20 years of record of the data downloads made and their metadata (who makes the download, when, with what purpose, for what location, etc.). This information is considered prone to applying AI at the urban level. This research focuses on two objectives:

- As a main objective, to develop an AI algorithm that predicts the works carried out in public space based on Inkolan's download information.
- As a secondary objective, to develop an AI algorithm that can predict Inkolan downloads based on historical data.

In light of the application domain, it is crucial to contextualise the data using urban variables that help enrich the analysis and better understand the spatial impacts of each public work.

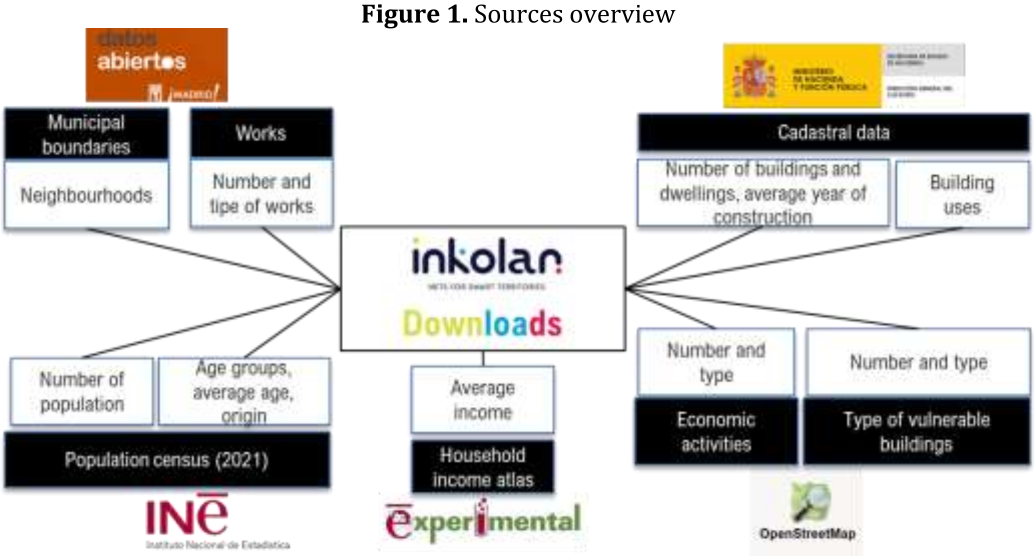
3.2. Data sources

This study combines georeferenced public and private data sources to explore the feasibility of applying AI models to predict public space works and analyse their urban impact. The geolocation or the possibility of georreferencing them is crucial to support the spatial comparisons required for the project.

The main data source for the analysis is Inkolan's internal dataset, It consist on metadata associated with downloads of underground infrastructure information. Each record includes various attributes such as the profile of the requesting entity, the purpose of the request, whether it is linked to a public work, the expected execution date, and a geospatial boundary (referred to as a "fence") that defines the geographic area of interest. In addition, the study incorporates multiple public data sources, grouped into two categories:

- Public works data: These datasets describe the execution of public space interventions and include licenses and records of completed works. When properly georeferenced (e.g., with coordinates or addresses), they can be linked to Inkolan's records to validate and train predictive models.
- Urban contextual data: These datasets provide relevant information about the physical, social, and economic environment in which public works occur. They help enrich the interpretation of predictions and analyse potential impacts. Sources include administrative boundaries, census data, cadastral information, household income indicators, and open geospatial data.

Figure 1. Sources overview



Source: Own elaboration, 2025.

All this sources were interconnected with Inkolan download records as represented in Figure 1. Table 1 summarises the different data sources used in this study, along with their main characteristics and geolocation capabilities.

Table 1. Overview of data sources used in the study

Source	Category	Type of Data	Key Attributes	Geo-referenced
Inkolan download records	Private data	Metadata about digital infrastructure requests	User identity, type of user, timestamp, purpose, work-related flag, expected work date, company profile, geofence	Yes
City council / Opendata	Finished works on public space data	Works permits and records	Work type, execution date, location	No
National Institute of Statistics	Urban contextual data	Population census (2021)	Demographics, population density, and age distribution by census track	Yes

Cadastral data	Urban contextual data	Building information	Building use, surface area, year of construction, number of dwellings	Yes
Household Income Atlas	Urban contextual data	Socioeconomic indicators	Household income by census track	Yes
OpenStreetMap (OSM)	Urban contextual data	Open geospatial data of urban infrastructure and amenities	Economic activities, type of buildings	Yes
Municipal boundaries	Urban contextual data	Administrative boundaries	Names, codes, boundaries of districts and neighborhoods	Yes

Source: Own elaboration, 2025.

3.3. Methodology

The methodology adopted in this study is framed within a mixed-methods approach for research in Urban Data Science, as proposed by Batty (2013) in The New Science of Cities. This approach recognizes the need to combine the collection and analysis of large volumes of urban data with qualitative contextualization and the participation of different urban stakeholders. In this way, it facilitates a deeper and more realistic understanding of urban phenomena, incorporating both technical and social dimensions. The integration of georeferenced quantitative data with knowledge contributed by stakeholders allows the design of predictive models and artificial intelligence solutions better suited to the complexity and heterogeneity of the urban environment.

The methodology followed consists of several steps as presented in Figure 2 and explained below.

Figure 2. Methodology

Source: Own elaboration, 2025.

Initially, a data collection exercise is conducted. Inkolan is contacted to provide data on download records and associated metadata. In parallel, collaborations are established with the municipalities of Zaragoza and Ermua to obtain data on works in public spaces. However, due to data quality and completeness issues, the project ultimately relies on open data from the city of Madrid. Additional publicly available datasets from the National Institute of Statistics and the National Cadastre are used to contextualise download records and data on works, highlighting the importance of open data sources to overcome limitations and ensure data availability.

Secondly, all the collected information is preprocessed by georeferencing the data lacking precise coordinates through the Google Maps API, and interrelating the data to provide context through geo-algorithms developed through FME Form (a geospatial ETL tool for data processing and integration tasks). For instance, Inkolan downloads are linked to their corresponding neighborhood and are enriched with contextual information such as building data and population. The public works are also localised by neighborhood and spatially related to the Inkolan downloads. This careful preprocessing is crucial to address data inconsistencies and ensure the high-quality, interrelated, and geolocated dataset necessary for the effective performance of AI models, especially given the common challenges of data heterogeneity and incompleteness in real-world applications.

Thirdly, a comprehensive exploratory data analysis is undertaken to identify patterns and trends within the Inkolan dataset. Additionally, analyses are conducted to examine the relationships between the Inkolan download data and the public works data, including spatial and temporal correlations, to identify potential associations and dependencies for predictive models.

Fourthly, based on the previous analysis, an ARIMA predictive model was developed to forecast Inkolan downloads using historical patterns. This algorithm uses recent download frequencies to estimate the number of future downloads expected to occur. The model parameters (p, q, d) were estimated through an optimization process that automatically selects best values based in criteria such as the Arkaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC), ensuring balance between accuracy and generalizability. Additionally, a comprehensive dashboard is developed in Power BI to facilitate the visualisation of patterns and trends in the relationship between downloads, public works, and their associated contextual factors.

Fifthly, a synthesis of outcomes is performed , offering a comprehensive overview of the study's results, the encountered limitations, and the potential for future development and application.

As a final step and with the aim of cross-checking the conclusions obtained, a series of interviews were conducted with relevant stakeholders to better understand the relationship between the Inkolan downloads, the request for building licences, and the subsequent execution of the work. These interviews served to validate the findings and assess the feasibility of continuing the research on predicting public space interventions.

4. Results

4.1. Data collection: gaps and lacks

As mentioned in the methodology, the availability of data related to public works licenses or construction records was contrasted with the municipalities of Zaragoza and Ermua. Some of the difficulties encountered in this process include: the intricacies of identifying suitable organizational interlocutors, the diversity of data formats (in some cases, despite being digitized, the relevant information was scattered across different PDF documents attached to a construction file) and the most relevant and common to both, the insufficient volume of records that hinders the effective application of AI algorithms (between 100 and 500 total records depending on the consulted record).

Given this fact, the available data were analyzed in conjunction with public datasets that could provide a sufficient volume to support the implementation of AI models, identifying a promising opportunity in the city of Madrid. A dataset of completed construction permits carried out in the public domain—covering sewage works, emergency repairs, canalization, and scheduled trenching between 2017 and 2021—was selected. This dataset contained more than 55,000 records, including the month and year of completion, the type of work, and a reference postal address for each project. However, it lacked additional relevant information such as the start date, duration, scope of the work, or the identity of the designer or developer involved.

While the volume of data was adequate for exploratory analysis and initial model development, this highlights the need to balance data quantity with quality, as well as the importance of integrating multiple data sources to enrich the contextual understanding and improve the accuracy of predictive models. By combining datasets from different origins—such as open data from Madrid, municipal records, and third-party sources like Inkolan—this study aimed to overcome the limitations of individual datasets and enhance the robustness of the overall analysis.

4.2. Data preprocessing

The preprocessing of public construction data involved adapting and unifying files from different years into a single file, geocoding records based on the postal address indicated for each work, and subsequently cleaning the data to discard records without assigned georeferencing or poorly geolocated (those that had been geolocated outside the municipal boundary). Furthermore, the database of works was spatially related to the Inkolan downloads database to identify the works contained within the perimeter of one or several downloads.

Moreover, the Inkolan downloads, in addition to being related to the projects, were also related to contextual information to identify the neighbourhoods where they occur, as well as some contextual indicators related to the population potentially affected in the event of a construction project within the download perimeter, the sensitive equipment affected, information on the income level of the residents, or the economic activities that may be affected. The objective of including these data was to enrich subsequent analyses and provide context for the visualisation of results.

4.3. Data analysis main conclusions

After preprocessing the data, two comprehensive datasets were obtained: one containing public works and the other comprising Inkolan downloads, both enriched with spatial and temporal attributes.

However, as previously mentioned, no common field exists between the two databases that could establish a one-to-one relationship—that is, one that links a given work to a specific download. Therefore, the first step involved defining criteria to approximate this one-to-one work-download association.

Initially, each download was linked to all works whose geolocations fell within the polygon specified in the corresponding download record, as any other association would have lacked logical consistency. Subsequently, the Euclidean distance between the polygon's centroid and the locations of the works was calculated, generating for each download a ranked list of potentially related works based on spatial proximity.

In addition, the downloads dataset included the download date, along with user-provided estimates for the start and end dates of the works. It is important to emphasise that only the download date was fully reliable. Conversely, the public works dataset provided verified information on project completion dates. For this reason, a second criterion—based on temporal proximity—was also introduced to refine the association process. This proximity was calculated by subtracting the download date from the work's completion date.

Nevertheless, a one-to-one relationship—linking a single work to a single download—had not yet been established. To address this, some hypotheses were formulated to guide the creation of one-to-one relationships. This involved defining thresholds for the maximum permissible time difference between a work's completion date and the corresponding download date, as well as for the maximum allowable Euclidean distance between their centroids.

A one-to-one relationship occurs when a work is located within a single download polygon, and that polygon contains no other works. However, this is uncommon given the size of the datasets: approximately 50,000 works over 4 years versus 40,000 downloads over a decade. Therefore, three scenarios were analyzed to define these criteria:

1. Downloads with a single work inside the polygon: A sample of 4,154 records showed that such downloads were located at most 75 meters from the associated work and typically occurred up to 1,500 days before the work's completion.
2. Works within a single polygon: From 1,827 works analyzed, no additional filtering criteria emerged, but the distance and time thresholds from case 1 were generally satisfied.
3. Strict one-to-one matches: Only 45 cases met the strict condition of one work within one polygon, with no overlaps elsewhere. Distances rarely exceed 100m (mostly below 75m), while time differences vary widely between −2,000 and 3,000 days. Due to the small sample, this pattern was not considered generalizable. However, combining this with case 1, the time threshold was adjusted to [−800, 1500] days to account for negative date differences observed in these matches.

As a conclusion after this analysis, it was determined that a work would be associated with a discharge if it was located within 75 meters of the centroid of the area and had a date difference between -800 and 1,500 days.

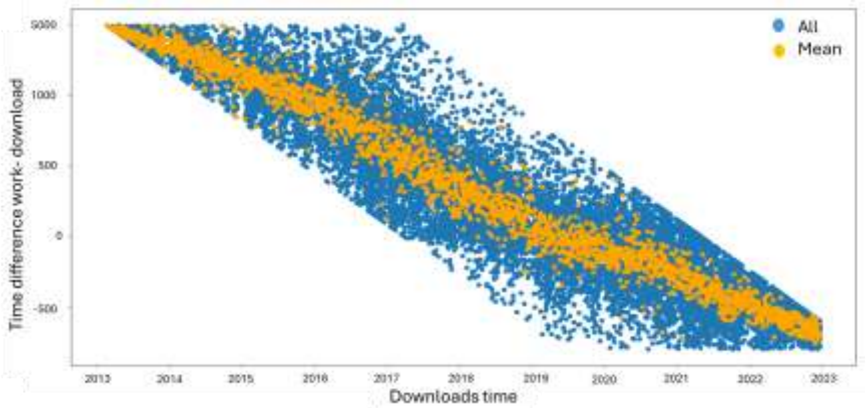
4.4. Modelling and visualisation results

Once the data were preprocessed and relationships between them were defined, two approaches were pursued: identifying works associated with each download and predicting future downloads. The following subsections explain these models in detail.

4.4.1. Work–Downloads relationship patterns

From the previous analysis, it was determined that, in general, a work may be completed between 800 days before and 1500 days after a download is made. Using this insight, the download–work matching process was refined, and the temporal evolution of downloads versus the time difference to associated works was examined. Figure 3 shows, for each download date, the difference in days between that download date and the completion date of the temporally and spatially associated works.

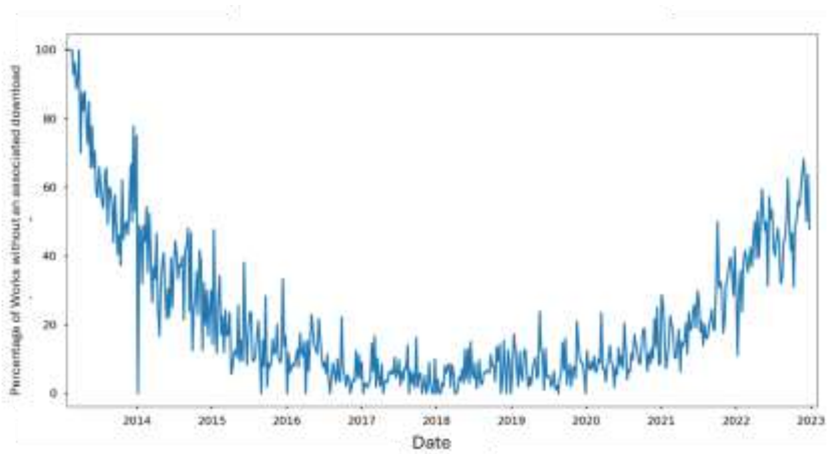
Figure 3. Download Dates Compared to Time Remaining Until Work Completion



Source: Own elaboration, 2025.

The analysis revealed a linear trend: over time, the gap between a download date and the corresponding project completion date decreases. This is due to a data bias—as previously mentioned, downloads span from 2010 to 2022, while project completions are only recorded between 2017 and 2021—resulting in more frequent associations of recent downloads with past projects. Consequently, a direct relationship cannot yet be confirmed. Nevertheless, when applying spatial and temporal filters, some downloads had no matching works, allowing for the calculation of the percentage of unmatched downloads over time, as illustrated in Figure 4.

Figure 4. Percentage of downloads without an associated work

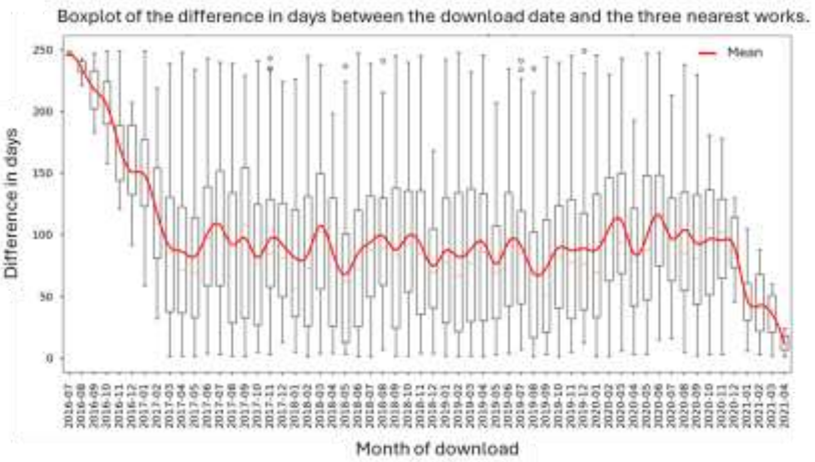


Source: Own elaboration, 2025.

To continue refining the work-download relationship, the three closest works were selected for each download, and the average time until the work completion dates was calculated. This provided a “days to completion” value for each download. The results still showed a downward trend due to the differing date ranges between the datasets, which led to the application of stricter filtering. However, since download type affects timing, the data were split into work-related and project-related downloads, which are characteristics specified within the dataset itself. Frequency analysis indicated:

- Work downloads typically occur 0–250 days before the project starts, with centroids within 75m. As shown in Figure 5, the series remained stationary between 2017–2021, with an interquartile error margin of approximately 100 days (~3 months), suggesting reasonably reliable one-to-one matches.

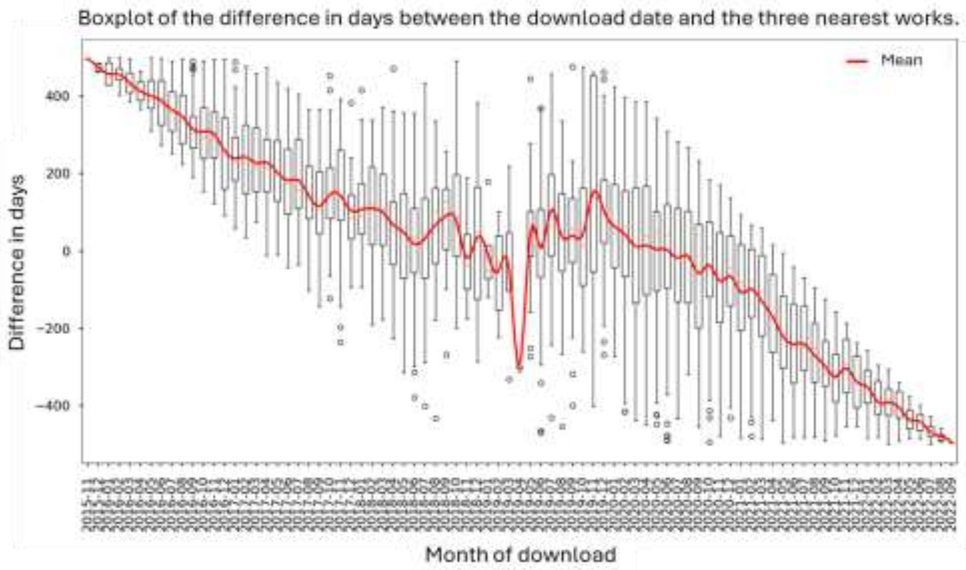
Figure 5. Boxplot of the difference in days between the dates of work-type downloads and the dates of the three nearest works.



Source: own elaboration.

- Project downloads: typically occur 500 days before to 500 days after project start, also within 75m. However, no stable temporal pattern emerged (see Figure 6), making one-to-one matching unreliable. With more project completion data and narrower interquartile ranges, predictive or autoregressive models could be developed in the future.

Figure 6. Boxplot of the difference in days between the dates of project-type downloads and the dates of the three nearest works.



Source: own elaboration.

Based on the above, it is demonstrated that, first and foremost, due to the difference in date ranges, it is challenging to obtain a valid ground truth for training a model capable of predicting the relationship between a work and a download. Nevertheless, using the defined temporal and spatial criteria, a meaningful measure of the temporal evolution of downloads without an associated work has been established. Additionally, these criteria have enabled the association of a reduced set of works with downloads, creating a certain systematic work-download relationship. From this, it can be inferred that, in the particular case of work-type downloads, there is some stationarity, confirming that approximately 100 days typically elapse between a download and the completion of the work.

4.4.2. Download predictive model

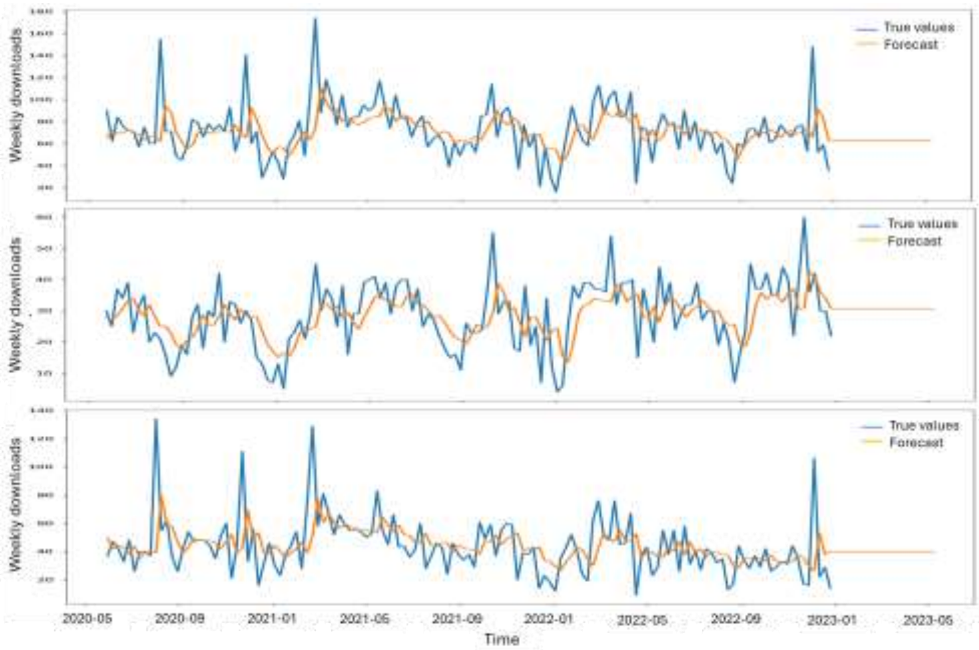
The objective of this second modelling approach was to forecast the future number of downloads from Inkolan using historical data. To achieve this, an ARIMA (Autoregressive Integrated Moving Average) model was developed (Shumway & Stoffer, 2005). This time series model captured the temporal dynamics of download activity and demonstrated the ability to generate forecasts of future values based on information from previous time points.

Three experiments were conducted to evaluate the predictive performance of the model. First, the temporal trends of the entire download dataset were analyzed and forecasted. Second, due to significant differences in download frequencies between project- and work-type downloads, the dataset was stratified into these two categories, and separate predictive models were evaluated for each. In all experiments, the dataset was partitioned into training and test subsets, with 80% of the data used for model training and 20% reserved for out-of-sample evaluation.

The results (see Figure 7) indicated strong predictive accuracy for intermediate download values. However, due to the moving-average nature of the ARIMA model, it was unable to accurately forecast extreme peaks—instances where download counts deviated significantly from the mean and standard deviation—with an RMSE of 23,95. Moreover, predictive accuracy improved when modelling more specific subsets of data, with RMSE values of 9,21 for works and 20,30 for projects.

Finally, the works dataset was further disaggregated by type, a feature also presents in the data. As anticipated, increased specificity improved forecast precision. Nonetheless, working with smaller subsets resulted in reduced sample sizes for training, potentially limiting the model's ability to generalise to unseen data.

Figure 7. Prediction of Future Downloads Using the ARIMA Model for the Complete Dataset, and Separately for Work-Type and Project-Type Downloads

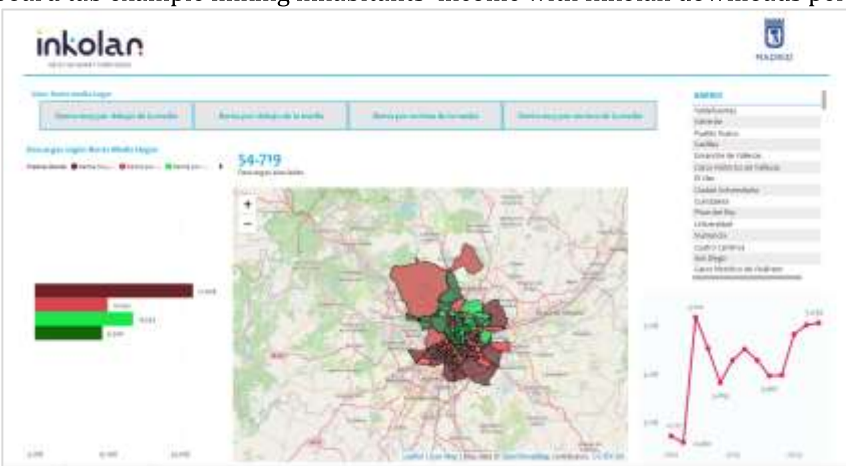


Source: Own elaboration, 2025.

4.4.3. Visualization dashboard

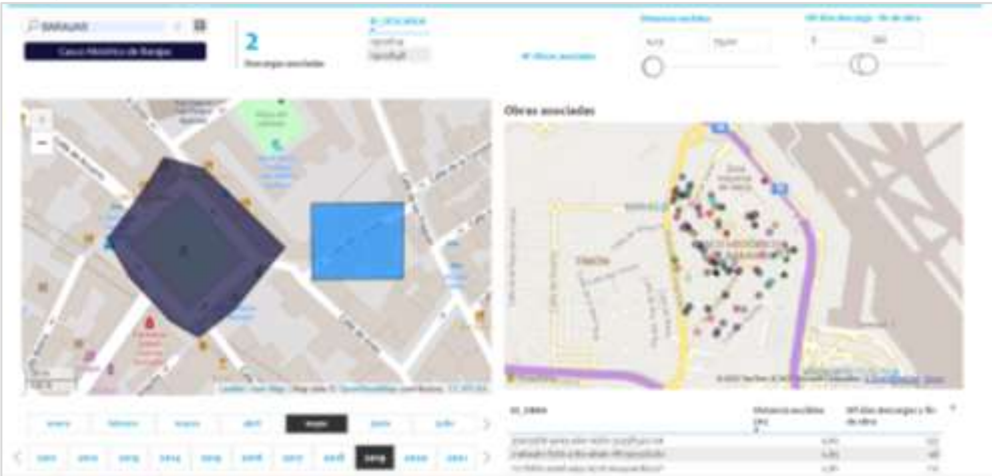
For the visualisation of results, both construction data and downloads data are integrated with associated contextual information and conclusions obtained about the relationship patterns between works and downloads into a dashboard developed with PowerBI. The dashboard consists of several tabs through which information on Inkolan downloads and open data construction can be analyzed separately, incorporating different possibilities for relating Inkolan downloads information to its context (to analyze the influence of inhabitant income or the construction period of city neighborhoods on downloads, see Figure 8) and a specific tab that allows relating Inkolan downloads to their potential works based on the parameters defined from the exploratory analysis (see Figure 9).

Figure 8. Dashboard tab example linking inhabitants' income with Inkolan downloads per neighbourhood



Source: Own elaboration, 2025.

Figure 9. Dashboard tab example linking Inkolan downloads with open data works



Source: Own elaboration, 2025.

4.5. Synthesis of Outcomes

Nevertheless, a predictive model of public works based on Inkolan's downloads could not be consolidated. Therefore, progress was made in relating each download to a work based on date and position criteria. In the event of a direct relationship (one-to-one) between downloads and works, it would be possible to determine the expected temporal difference for future downloads. In this sense, due to the lack of this direct relationship, the following hypotheses were formulated:

1. A work is associated with a download if it is located within the polygon, with a distance from the centroid of less than 75m and with a temporal difference from the download between -800 and 1,500 days. This suggests that the work is completed between 2 years before the download and 4 years after.
2. For downloads of type "work", it was observed that downloads occur within a range of 0 to 250 days before the start date of the work. Additionally, it was found that the centroids must be within 75m of each other.
3. In contrast, for downloads of type "project", downloads occur within a range of 500 days before to 500 days after the project start date. Similarly, it was found that the centroids must be within 75m of each other.

After these pairing approximations, the difference in days between the download and the possible associated works was analysed. In summary, the following results were obtained:

- With a common field between the download and work databases, a one-to-one relationship can be defined, which would allow for a prediction of when the work related to the download will be completed.
- The proportion of downloads that will not lead to work is known, based on the defined maximum dates and distances.
- A time interval is provided for when the work is expected to be completed with a download date.

These results suggest that, although a predictive model of public works based on Inkolan's downloads could not be consolidated, significant progress was made in relating downloads to works. These findings can be useful for improving the planning and management of public works.

4.6. Contrast with stakeholders

Given the results obtained from trying to relate the Inkolan download file to the works file, it was considered necessary to conduct a series of interviews with key agents to better understand the relationship between download and work, the process from the license request to the execution of the work, the agents involved, and the possibilities that this may open up for the continuation of research in this field.

Contrast interviews were conducted with departments of license management for works from two city councils (Bilbao and Madrid) and with responsible parties for the execution of works from Inkolan's partners in different territories of Spain. Additionally, the norms for managing works from 10 Spanish municipalities related to the areas where the interviewed agents operate are reviewed.

The results of the analysis provided some relevant conclusions. Firstly, it should be mentioned that there is no single procedure for granting a work license in Spain, as they differ from one municipality to another. Secondly, the time period between the license request and the granting date depends on many variables, including the municipality's agility and, depending on the effects of the work, may require the granting of other licenses/authorisations linked to the project (Coastal, Roads, Provincial, Industry, etc.). A City Council, in general, waits for the applicant to have all the necessary permits before granting the license. There are no integrated authorisations or one-stop shops. Permits must be requested from each organisation, which can take up to one year or more (depending on the organisation, municipality, etc.). Thirdly, there is currently no common identifier that allows associating an Inkolan download with a municipal work registry. Fourthly, in the download-work process, several agents intervene, and sometimes the study or documentation preparation for the work license application is outsourced to a company (which usually involves downloading Inkolan to perform the impact studies), which means that the entity requesting the license may not be the same as the one that performed the download. Sometimes, there may be three entities with different roles (designer, developer, and contractor) linked to the download and work process, and it is not always the designer who performs the download. Situations arise where the download is performed later for checks or even at different moments in the process, and/or by different agents.

5. Discussion

The study highlights critical challenges in applying AI models to predict public works based on the available data, emphasising data completeness as the most significant limitation. The absence of common identifiers linking Inkolan downloads with municipal records made it difficult to establish one-to-one relationships, which required the use of spatial and temporal criteria to associate the data. Although the other quality dimensions (accuracy, relevance, timeliness, and consistency) were generally better addressed, they also presented limitations such as geolocation errors, temporal mismatches between datasets, and a lack of standardised structures. These deficiencies affect the effectiveness of predictive modelling and analysis. The results underline that having large volumes of data is not enough; it is important to ensure that the available data is of high quality. However, considering that transparency and equity are key elements in AI-driven urban decisions, even accurate and consistent datasets may perpetuate systemic inequities if they reflect historical biases. The stakeholder consultations and regulatory analysis confirm the difficulty of obtaining quality data in the context of urban planning. This work offers an initial approach to addressing these challenges and guides future improvements in data integration and quality.

As highlighted in the literature review, a key limitation in urban planning is the reliance on existing datasets that were not originally designed for analytical or predictive purposes. Consequently, these datasets often fail to meet the requirements of contemporary research and technological applications. This study underscores the need to generate data specifically tailored to the research objectives, which necessitates collaboration among multiple stakeholders — a principle already emphasised as essential in urban governance and planning.

Regarding the challenges identified by public administrations in data production, this paper proposes a set of recommendations to facilitate the integration of AI algorithms in future urban planning initiatives. First and foremost, it is essential to digitize data from the outset, ensuring that they are structured with future analytical use in mind. In particular, information on construction permits and their follow-up should be collected in digital formats that allow for seamless export to databases. Digitization efforts must also prioritize transparency, such as documenting the sources, limitations, and potential biases of datasets. This aligns with the principles of Explainable AI (XAI), which emphasize making algorithmic decision-making processes accessible to non-technical stakeholders.

Given the complexity of the permit and construction process, it is crucial to systematically interrelate the various stages — including procedures, dates, incidents, delays (e.g., requests for modifications or additional information), outcomes (e.g., rejections, approvals, or unexecuted works), the scale of the work (budget, estimated duration), and the geolocation of the site (with coordinates and, preferably, a defined perimeter). Additionally, it is important to record the roles of the involved stakeholders at each stage. This structured approach can help identify patterns of inequity, such as disproportionate delays or rejections in marginalized neighborhoods, which may indicate algorithmic or procedural biases. Furthermore, documenting stakeholder roles enhances accountability by clarifying who is responsible for data quality and decision-making at each step.

The proper structuring and formatting of such data would enable the development of a robust database on public works, facilitating the identification of key milestones in the process from permit application to execution. Furthermore, since the submission of service network impact plans is often mandatory, it would be beneficial to systematically record whether such plans are included in the permit application and, if so, to note their reference codes. This would support the integration of public records with the data downloaded from Inkolan. However, this integration must be accompanied by safeguards to prevent the reinforcement of biases. For example, if AI models prioritize permits in areas with higher property values due to data availability, planners must actively audit these outcomes to ensure equitable resource allocation.

A fundamental requirement for the development of a predictive algorithm is the availability of a dataset that clearly links works with their associated downloads, serving as a ground truth to refine the relationship patterns and train the algorithm. This could enable the prediction of future permit applications based on download data, the estimation of work duration, and the analysis of whether a given download will result in an actual work, depending on its characteristics. Yet, the ethical implications of such predictions are significant.

However, due to the complexity of the process and the variability in timeframes influenced by factors such as municipal processing speed, the scale and type of impact of the work, the algorithm’s validity would be context-specific. Its accuracy may vary depending on the nature of the work (e.g., its scale or impact), and its application to other contexts would require adaptation based on locally relevant data. This context-specificity highlights the importance of proactively involving local stakeholders in the validation of results to mitigate potential algorithmic biases.

6. Limitations of the Study and Future Research Directions

This study has several limitations that warrant consideration. First, the scarcity of publicly available data on public works licenses and construction records significantly constrained the scope of analysis. Second, the absence of standardized procedures for granting work licenses across Spanish municipalities introduced variability in data collection and interpretation. This lack of harmonization complicates cross-jurisdictional comparisons and limits the potential for scalable solutions. Finally, the current fragmented nature of permit processes, requiring separate authorizations from multiple agencies, introduces complexity in tracking and analyzing workflows, as there is no centralized or integrated system to monitor progress or outcomes.

To address these limitations, future research may benefit from focusing on the following areas:

1. Standardization of Data Collection: Establish unified frameworks for collecting and sharing public works data across municipalities, including common identifiers to link disparate systems (e.g., municipal registries and external databases).
2. Leveraging Indirect Data Sources: Explore alternative data streams (e.g., public tenders, procurement records) as proxies for construction Works licenses.
3. Enhanced Data Infrastructure: Investigate centralized digital platforms or "one-stop shops" to streamline permit applications and reduce administrative fragmentation. Pilot

projects could test integrated workflows and assess their impact on processing times and transparency.

4. AI and Predictive Analytics: Explore AI models to predict delays, identify procedural bottlenecks, or detecting patterns of inequity (e.g., disproportionate rejections in marginalized neighborhoods). Expanding datasets through municipal collaboration and synthetic data generation, while aligning with Explainable AI (XAI) principles, would enhance transparency and trust.
5. Policy and Governance Analysis: Address socio-political barriers to harmonization, such as stakeholder resistance, legal discrepancies, and resource disparities. Policy reforms could incentivize inter-agency collaboration and establish accountability mechanisms for data quality and decision-making.

By addressing these limitations and expanding the scope of inquiry, future research can contribute to more equitable, efficient, and data-driven public works management systems.

7. Conclusions

This study has revealed the complexity of the public works permit process in Spain, as well as the absence of a unified and standardised procedure across jurisdictions. Moreover, it has underscored the critical importance of inter-agency collaboration among all stakeholders involved in the process. Additionally, the necessity of generating purpose-specific data has been emphasised, with particular emphasis on the need to establish a systematic and reliable linkage between Inkolan download records and municipal public works registries.

The development and application of artificial intelligence (AI) in the context of public works and urban planning have shown great potential. However, the success of these applications is highly dependent on the quality and availability of data. From the analysis conducted in this study, it is evident that the data required for AI implementation must meet specific quality dimensions such as accuracy, relevance, completeness, timeliness, and consistency. These dimensions are essential to ensure that the data used for training AI models is reliable and representative of the real-world scenarios they aim to replicate or predict.

One of the key findings of this study is the need for structured and digitised data from the outset. Information related to construction permits and their follow-up must be collected in digital formats that allow for seamless integration into databases. This structured approach facilitates the identification of key milestones in the process, from permit application to execution, and supports the development of robust databases on public works.

Additionally, the study highlights the importance of systematically recording relevant information such as service network impact plans, stakeholder roles, and geolocation data. These details are crucial for building a comprehensive dataset that can serve as a ground truth for training predictive algorithms. The integration of public records with data from platforms like Inkolan further enhances the utility and scope of the data.

Despite the potential benefits, the study also identified several challenges, including the insufficient volume of records available in the municipalities of Zaragoza and Ermua. The limited data hindered the effective application of AI algorithms, emphasising the need for more extensive and standardised data collection practices.

In conclusion, while AI has the potential to revolutionise urban planning and public works management, its successful implementation requires a concerted effort to improve data quality, availability, and standardisation. Addressing these challenges will be essential for leveraging AI to its full potential and achieving more efficient and effective urban development.

8. Acknowledgements

The authors wish to thank INKOLAN and their partners, as well as the municipalities and stakeholders who have participated in this process.

References

Agudo, U., Liberal, K. G., Arrese, M., & Matute, H. (2024). The impact of AI errors in a human-in-the-loop process. *Cognitive Research: Principles and Implications, 9*(1). https://doi.org/10.1186/s41235-023-00529-3

Anwar, M. R., & Sakti, L. D. (2024). Integrating Artificial Intelligence and Environmental Science for Sustainable Urban Planning. *IAIC Transactions on Sustainable Digital Innovation (ITSDI)*, *5*(2), 179-191. https://doi.org/10.34306/itsdi.v5i2.666

Batty, M. (2013). *The New Science of Cities*. The MIT Press. https://www.jstor.org/stable/j.ctt9qf7m6

Bibri, S. E., Huang, J., Jagatheesaperumal, S. K., & Krogstie, J. (2024). The synergistic interplay of artificial intelligence and digital twin in environmentally planning sustainable smart cities: A comprehensive systematic review. *Environmental Science and Ecotechnology*, *20*, 100433. https://doi.org/10.1016/j.ese.2024.100433

Chaturvedi, V., & de Vries, W. T. (2021). Machine Learning Algorithms for Urban Land Use Planning: A Review. *Urban Science, 5*(3), 68. https://doi.org/10.3390/urbansci5030068

Correia, C. R., & Roseland, M. (2022). Addressing Negative Externalities of Urban Development: Toward a More Sustainable Approach. *Urban Science, 6*(2), 38. https://doi.org/10.3390/urbansci6020038

Damioli, G., Van Roy, V., Vertesy, D., & Vivarelli, M. (2025). Is artificial intelligence leading to a new technological paradigm? *Structural Change and Economic Dynamics, 72*, 347-359. https://doi.org/10.1016/j.strueco.2024.12.006

Delgado Enales, I. (2024). *Advanced optimization and data modeling techniques to improve accesibility and thermal comfort in urban plannning*. [Tesis Doctoral, Universidad del País Vasco] http://hdl.handle.net/10810/73521

Delgado-Enales, I., Molina-Costa, P., Osaba, E., Urra-Uriarte, S., & Del Ser, J. (2022). Improving the Urban Accessibility of Older Pedestrians using Multi-objective Optimization. *2022 IEEE Congress on Evolutionary Computation (CEC)*, 1-8. https://doi.org/10.1109/CEC55065.2022.9870432

D'Ignazio, C., & Klein, L. F. (2020). *Data Feminism*. The MIT Press. https://doi.org/10.7551/mitpress/11805.001.0001

Fernández Álvarez, L., & Garaizar Sagarminaga, P. (2024). *La inteligencia artificial explicada a todos los públicos (castellano)*. A Fin de Cuentos.

Gao, R. X., Krüger, J., Merklein, M., Möhring, H.-C., & Váncza, J. (2024). Artificial Intelligence in manufacturing: State of the art, perspectives, and future directions. *CIRP Annals, 73*(2), 723-749. https://doi.org/10.1016/j.cirp.2024.04.101

Garg, P. K. (2021). Overview of Artificial Intelligence. In L. Sharma y p. Kumar Garg (Eds.) *Artificial Intelligence*. Chapman and Hall/CRC.

Gong, Y., Liu, G., Xue, Y., Li, R., & Meng, L. (2023). A survey on dataset quality in machine learning. *Information and Software Technology, 162*, 107268. https://doi.org/10.1016/j.infsof.2023.107268

Hamann, K. R. S., Bertel, M. P., Ryszawska, B., Lurger, B., Szymański, P., Rozwadowska, M., Goedkoop, F., Jans, L., Perlaviciute, G., Masson, T., Fritsche, I., Favaro, T., Hofer, A., Eisenberger, I., Gutschi, C., Grosche, C., Held, J., Athenstaedt, U., & Corcoran, K. (2023). An interdisciplinary understanding of energy citizenship: Integrating psychological, legal, and economic perspectives on a citizen-centred sustainable energy transition. *Energy Research & Social Science, 97*, 102959. https://doi.org/10.1016/j.erss.2023.102959

He, W., & Chen, M. (2024). Advancing Urban Life: A Systematic Review of Emerging Technologies and Artificial Intelligence in Urban Design and Planning. *Buildings, 14*(3), 835. https://doi.org/10.3390/buildings14030835

Korobeinikova, A., Danilina, N., & Teplova, I. (2024). Planning Public Space Climate Comfortability: A GIS-Based Algorithm for the Compact Cities of the Far North. *Land, 13*(11), 1763. https://doi.org/10.3390/land13111763

Koutra, S., & Ioakimidis, C. S. (2023). Unveiling the Potential of Machine Learning Applications in Urban Planning Challenges. *Land, 12*(1), 83. https://doi.org/10.3390/land12010083

Liu, J., Kong, X., Xia, F., Bai, X., Wang, L., Qing, Q., & Lee, I. (2018). Artificial Intelligence in the 21st Century. *IEEE Access, 6*, 34403-34421. https://doi.org/10.1109/ACCESS.2018.2819688

Lu, H., Li, Y., Chen, M., Kim, H., & Serikawa, S. (2018). Brain Intelligence: Go beyond Artificial Intelligence. *Mobile Networks and Applications, 23*(2), 368-375. https://doi.org/10.1007/s11036-017-0932-8

Mashhood, M., Salman, H., Amjad, R., & Nisar, H. (2023). The Advantages of Using Artificial Intelligence in Urban Planning – A Review of Literature. *JOURNAL OF STATISTICS, COMPUTING AND INTERDISCIPLINARY RESEARCH, 5*(2), 1-12. https://doi.org/10.52700/scir.v5i2.125

Mohammed, S., Budach, L., Feuerpfeil, M., Ihde, N., Nathansen, A., Noack, N., Patzlaff, H., Naumann, F., & Harmouch, H. (2025). The effects of data quality on machine learning performance on tabular data. *Information Systems, 132*, 102549. https://doi.org/10.1016/j.is.2025.102549

Molina-Costa, P. (2024). *La planificación urbana en la era de la digitalización. In Ciudad digital Conectividad y (des)aceleración.* CATARATA. https://www.catarata.org/libro/ciudad-digital_150546/

Moore, G. E. (2006). Cramming more components onto integrated circuits, Reprinted from Electronics, volume 38, number 8, April 19, 1965, pp.114 ff. *IEEE Solid-State Circuits Society Newsletter, 11*(3), 33-35. https://doi.org/10.1109/N-SSC.2006.4785860

O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy.* Crown Publishing Group.

Park, C., No, W., Choi, J., & Kim, Y. (2023). Development of an AI advisor for conceptual land use planning. *Cities, 138*, 104371. https://doi.org/10.1016/j.cities.2023.104371

Pericault, Y., Viklander, M., & Hedström, A. (2023). Modelling the long-term sustainability impacts of coordination policies for urban infrastructure rehabilitation. *Water Research, 236*, 119912. https://doi.org/10.1016/j.watres.2023.119912

Rjab, A. B., Mellouli, S., & Corbett, J. (2023). Barriers to artificial intelligence adoption in smart cities: A systematic literature review and research agenda. *Government Information Quarterly, 40*(3), 101814. https://doi.org/10.1016/j.giq.2023.101814

Sanchez, T. W., Fu, X., Yigitcanlar, T., & Ye, X. (2024). The Research Landscape of AI in Urban Planning: A Topic Analysis of the Literature with ChatGPT. *Urban Science, 8*(4), 197. https://doi.org/10.3390/urbansci8040197

Shulajkovska, M., Smerkol, M., Noveski, G., & Gams, M. (2024). Enhancing Urban Sustainability: Developing an Open-Source AI Framework for Smart Cities. *Smart Cities, 7*(5), 2670-2701. https://doi.org/10.3390/smartcities7050104

Shumway, R. H., & Stoffer, D. S. (2005). *Time Series Analysis and Its Applications*. Springer.

Urra-Uriarte, S., Herranz-Pascual, K., Molina-Costa, P., Martin Roncero, U., & Glodeanu, A. (2024). An Age-Friendly Neighbourhood Index as a Long-Term Urban Planning Decision-Making Tool. *Urban Science, 8*(3), 127. https://doi.org/10.3390/urbansci8030127

Urrutia-Azcona, K., Usobiaga-Ferrer, E., De Agustín-Camacho, P., Molina-Costa, P., Benedito-Bordonau, M., & Flores-Abascal, I. (2021). ENER-BI: Integrating Energy and Spatial Data for Cities' Decarbonisation Planning. *Sustainability, 13*(1), 383. https://doi.org/10.3390/su13010383

Vicente, L., & Matute, H. (2023). Humans inherit artificial intelligence biases. *Scientific Reports, 13*(1), 15737. https://doi.org/10.1038/s41598-023-42384-8

Villanueva-Merino, A., Urra-Uriarte, S., Izkara, J. L., Campos-Cordobes, S., Aranguren, A., & Molina-Costa, P. (2024). Leveraging Local Digital Twins for planning age-friendly urban environments. *Cities, 155*, 105458. https://doi.org/10.1016/j.cities.2024.105458

Wu, S. Z. (2025). *The AI City*. Springer Nature. https://doi.org/10.1007/978-981-96-2560-4

Xiao Ting Cheng, Ming Wang, & XingQi Fan. (2025). Generation of Personalized Urban Public Space Color Design Scheme Assisted by Artificial Intelligence. *International Journal of High Speed Electronics and Systems*, *7*(2). https://doi.org/10.1142/S0129156425401615

Zheng, Y., Lin, Y., Zhao, L., Wu, T., Jin, D., & Li, Y. (2023). Spatial planning of urban communities via deep reinforcement learning. *Nature Computational Science, 3*(9), 748-762. https://doi.org/10.1038/s43588-023-00503-5