The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Volume L-4/W3-2026
10th International Conference on Smart Data and Smart Cities (SDSC), 29 September–1 October 2026, Sofia, Bulgaria

The Computational Evolution of Bikeability Assessment: from Static Gis Indices to GeoAI and Urban Digital Twins. A Scoping Review

Domenico D'Uva [1], Marco Seccaroni [1,*]

[1] Dept. of Architecture, Built Environment and Construction Engineering, Politecnico di Milano, Milan, Italy – marco.seccaroni@polimi.it

* Corresponding author

Keywords: Bikeability, Scoping Review, GeoAI, Street View Imagery, Digital Twin, Level of Traffic Stress.

Abstract

Urban bikeability assessment has undergone a profound computational transformation over the past two decades: from manual audits to GIS-derived indices, from GPS-validated demand models to computer vision applied to street-level imagery and its integration with geospatial analysis (GeoAI), and toward three-dimensional urban digital twins. This paper reports a scoping review conducted and reported according to the PRISMA extension for Scoping Reviews (PRISMA-ScR). Three open scholarly sources (Semantic Scholar, Crossref, Consensus) were searched using nine concept-cluster query strings; after deduplication and two-stage screening, 425 studies (2003–2025) were included and assigned to three computational generations, understood as methodological strata rather than exclusive publication periods: Generation I denotes static GIS-based indices, exemplified by the Level of Traffic Stress (LTS) framework; Generation II denotes behaviourally validated, demand-model approaches based on GPS telemetry and crowdsourced data; Generation III denotes GeoAI approaches applying computer vision to Street View Imagery (SVI), including automated LTS classification, GAN-based perspective correction, and biosensor fusion. Four persistent gaps emerge across all generations: a standardisation deficit, dynamic-variable blindness, a scalability–equity trade-off, and an unresolved validation gap between index computation and modal-shift outcomes. The evidence suggests a convergence toward hybrid approaches combining GIS network analysis, behavioural demand data, and SVI-derived perceptual indicators, with the urban digital twin as a plausible — but still under-validated — integration platform.

1. INTRODUCTION

The promotion of urban cycling as a sustainable, healthy, and low-emission transport mode (Pucher and Buehler, 2008; Pucher et al., 2010) has generated a substantial scientific literature on the measurement and evaluation of cycling-supportive environments. The concept of bikeability — broadly defined as the degree to which an urban environment supports cycling as a mode of transportation — has emerged as the central analytical construct for translating urban form into cycling propensity (Ewing and Cervero, 2010; Lowry et al., 2012). Its operationalisation, however, has proven elusive: despite decades of methodological development, no consensus index exists, and the literature spans at least three fundamentally different computational paradigms that are rarely treated in relation to one another.

Because the organising vocabulary of this review has been used inconsistently in the literature, three terms are defined at the outset. Static GIS denotes methods that compute bikeability exclusively from time-invariant vector representations of the road network and its attributes, without observational data on cycling behaviour. GeoAI is used here in its strict sense: the integration of artificial-intelligence outputs — typically computer-vision analysis of imagery — with geospatial data and spatial analysis. The application of AI to street imagery alone constitutes computer vision or photogrammetry; it becomes GeoAI only when its outputs are georeferenced, aggregated to spatial units, and analysed within a geospatial framework (Section 3.5). Urban Digital Twin (UDT) denotes a semantically enriched, three-dimensional digital representation of the city, synchronised with observational data flows and capable of simulating intervention scenarios (Section 3.7). These definitions position the paper within spatial information science rather than within computer vision per se.

The first paradigm, rooted in Geographic Information Systems (GIS), dominated the field from the early 2000s until approximately 2016. Its roots lie in the earlier bicycle level-of-service tradition (Lowry et al., 2012); its most influential contribution, the Level of Traffic Stress (LTS) framework of Mekuria et al. (2012), algorithmically classified road segments by perceived psychological stress, providing a universally applicable yet empirically unvalidated typology. Subsequent GIS-based indices proliferated across European and North American cities, combining infrastructure density, slope, connectivity, and destination accessibility into composite spatial indicators (Winters et al., 2013; Grigore et al., 2019; Hardinghaus et al., 2021).

A second paradigm emerged with the mass availability of GPS telemetry and crowdsourced data (Strava Metro, OpenStreetMap). Studies by Fosgerau et al. (2023) and Łukawska et al. (2023) demonstrated that infrastructure-induced cycling demand could be quantified through route choice models calibrated on real trajectories — overcoming the behavioural abstraction of static GIS. Concurrently, Vierø et al. (2025) exposed a critical data quality problem: neither OSM nor official cadastral datasets alone are sufficient for reliable network-based bikeability analysis, constraining scalability at the national level.

A third paradigm, now rapidly ascending, applies computer vision to Street View Imagery (SVI) within geospatial workflows. Ito and Biljecki (2021) demonstrated that a 34-indicator bikeability index can be derived primarily from SVI, establishing CV-SVI pipelines as scalable substitutes for manual audits. Bianconi et al. (2023) devise an index based on the segmentation of images taken outside urban centres. Lin et al. (2024) automated LTS classification through contrastive deep learning on 39,153 road segments; Ito et al. (2024) introduced GAN-based perspective correction to mitigate the vehicle-centric bias of standard Google Street View capture; and Ye et al. (2024) combined SVI with crash records and SHAP explainability to model perceived cycling safety across London's entire road network.

Existing reviews have catalogued bikeability instruments (Kellstedt et al., 2021; Ahmed et al., 2024), infrastructure design principles, the effects of bikeway networks on cycling (Buehler and Dill, 2016), network-connectivity measures (Schön et al., 2024), and SVI applications (Biljecki and Ito, 2021). However, they have not explicitly synthesised bikeability assessment as an evolution across GIS-based indices, behavioural-data validation, and GeoAI/SVI methods — leaving the intersecting fields of GIS, transportation science, and GeoAI without a unifying analytical framework.

This paper addresses this gap through a scoping review reported according to PRISMA-ScR. Our objectives are: (1) to map the computational evolution of bikeability assessment into a generational model; (2) to characterise each generation's methodological architecture, geographic coverage, and declared limitations; and (3) to identify structural gaps that define the research frontier and inform the design of future hybrid methods. Three review questions structure the synthesis. RQ1: Which computational methods have been used to operationalise urban bikeability? RQ2: What data sources, units of analysis, validation strategies, and geographic contexts characterise each method family? RQ3: Which methodological gaps limit comparability, equity, and validation against observed cycling uptake? Figure 1 anticipates, in conceptual form, the integrated methodological framework toward which the reviewed literature converges and which structures the discussion in Section 5.

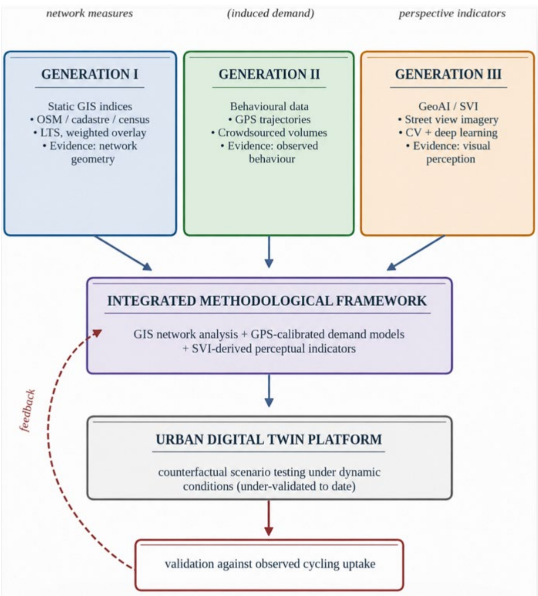

Figure 1. Conceptual diagram of the integrated methodological framework: the three computational generations contribute complementary evidence toward a hybrid, digital-twin-enabled assessment pipeline validated against observed cycling uptake.

2. METHODS: PRISMA-SCR PROTOCOL

2.1 Protocol Design

The review was designed following the Preferred Reporting Items for Systematic Reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR; Tricco et al., 2018). The research question was framed using the PCC (Population–Concept–Context) framework: Population = urban road networks and cycling infrastructure; Concept = computational methods for bikeability assessment; Context = urban environments across geographic scales. The review protocol — including the full query strings, eligibility rules, and exclusion codes — was documented prior to screening and data extraction and is available, together with the machine-readable screening decision log, from the corresponding author on request.

2.2 Search Strategy

Searches were executed in July 2026 against three open scholarly sources — the Semantic Scholar Academic Graph (bulk search API), Crossref (title-level relevance queries), and the Consensus search engine — using nine concept-cluster query strings designed a priori from the PCC blocks: the bikeability construct; bicycle level-of-service and compatibility instruments; the Level of Traffic Stress; street view imagery and computer vision; GPS, crowdsourced, and route-choice data; network connectivity and design; digital-twin simulation; suitability and friendliness indices; and AI-based assessment of cycling environments. Searches were restricted to English-language, peer-reviewed sources (2003–2025), with 2003 as the starting point following Kellstedt et al.'s (2021) identification of the first formal bikeability instrument. No reference-list hand-searching or citation tracking was performed; the implications are discussed in Section 5.

The complete source-specific query strings, filters, search dates, and per-source retrieval counts are documented in the supplementary search table. Because open scholarly indexes differ from subscription databases in coverage and matching behaviour, two declared mitigations were applied: the Semantic Scholar engine applies aggressive stemming (its single-token query for bikeability matches 32,484 records, including generic bike items), so a literal post-filter was applied to that query, and Crossref was queried at title level with local phrase filtering; Consensus contributed one complementary AI-based query. Publication type and peer-review status were assessed from record metadata and at title-level screening, and non-peer-reviewed or non-English items were excluded (Figure 2).

2.3 Screening and Inclusion Criteria

Records were screened in two stages against predefined eligibility criteria: a liberal title screen, followed by a stricter eligibility screen at title–abstract level; full texts were not systematically retrieved. Inclusion criteria required that studies: (i) propose, operationalise, or empirically apply a computational method for assessing cycling conditions; (ii) have a primary focus on urban or peri-urban environments; (iii) be peer-reviewed and published in English. Exclusion criteria: qualitative-only studies, health outcome studies without spatial methods, policy/advocacy documents without computational components, and walking-only active mobility studies. Criterion (i) was operationalised with an explicit boundary: studies had to produce an evaluative output about the cycling environment itself — an index, classification, stress or service level, suitability map, or validated assessment; purely behavioural estimation without such an output, routing algorithms, data-collection tools, and bike-sharing operations research were excluded under this rule. One non-peer-reviewed report (Mekuria et al., 2012) was retained as a documented exception to criterion (iii), given its foundational role in the LTS literature. Screening was performed by a single reviewer with the support of automated screening tools; every exclusion was assigned a reason code in the screening decision log, and the absence of double screening is declared as a limitation in Section 5.

2.4 Data Extraction

Data were extracted using a standardised matrix covering: authors and year, study type, geographic scope, unit of analysis, data sources, indicator categories, computational approach, weighting method, validation approach, and declared limitations. Structured metadata (year, venue, method family) were extracted for all included studies, and the full qualitative extraction was performed for a representative subset of 40 studies. Studies were subsequently assigned to one of three computational generations based on their primary methodological contribution. Transition studies bridging two generations (e.g., Cabral and Kim, 2022; Dai et al., 2023) are reported with both primary and secondary generation labels in the extraction matrix (I/II, II/III); the higher generation is used only for aggregate tabulation, and the transition is treated explicitly in the synthesis.

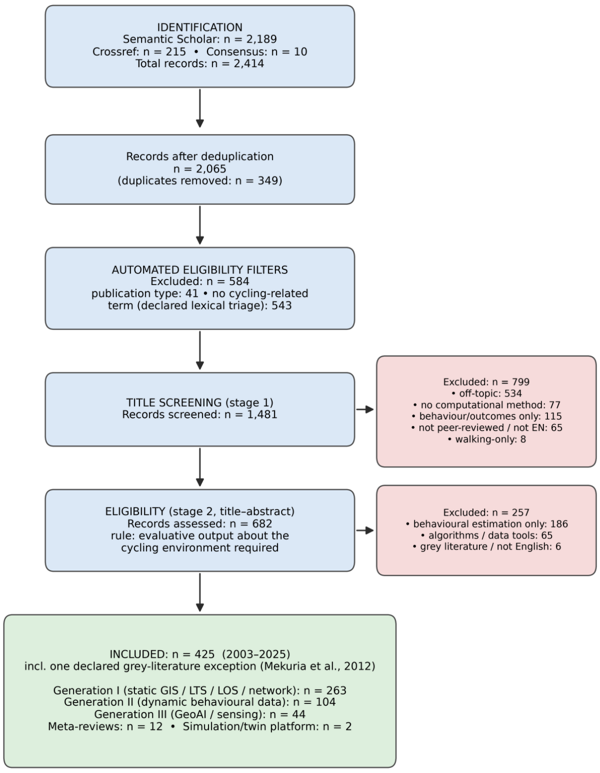

Figure 2. PRISMA-ScR flow diagram of the identification, screening, eligibility and inclusion stages.

2.5 Generational Classification Framework

The three-generation framework was developed inductively during the extraction phase. The generations are methodological strata — families of methods defined by their data epistemology, i.e., the type of evidence each treats as authoritative about cycling conditions — and not mutually exclusive publication periods: Generation I comprises static GIS approaches, which emerged mainly in 2003–2016 but remain in use; Generation II comprises behaviourally validated and demand-model approaches, which emerged mainly in 2017–2022; Generation III comprises SVI/GeoAI and sensor-fusion approaches, emerging from 2021 onward. This is why recent publications may be classified in Generation I. Generation I treats static geometric properties of the road network as the primary evidence; Generation II treats observed behaviour (GPS, crowdsourced activity) as authoritative validation; Generation III treats the cyclist's visual perception of the environment (SVI) as the primary evidence domain. This epistemological distinction structures the comparative analysis in Section 3.

3. RESULTS

3.1 PRISMA-ScR Flow: Search and Selection Process

The search and selection process is summarised in the PRISMA-ScR flow diagram (Figure 2). Source searches returned 2,414 records (Semantic Scholar 2,189; Crossref 215; Consensus 10). After deduplication (n = 2,065), automated eligibility filters removed 584 records (41 by publication type; 543 by a declared lexical triage requiring at least one cycling-related term). Title screening excluded 799 records under five mutually exclusive reason codes; 682 records were assessed at the stricter eligibility stage, of which 257 were excluded, yielding 425 included studies published between 2003 and 2025, including one declared grey-literature exception (Mekuria et al., 2012). Per-source identification counts, the full query strings, and the complete screening decision log are available from the corresponding author on request.

3.2 Data Extraction Matrix

The machine-readable data extraction matrix covering all 425 included studies is available from the corresponding author on request.

3.3 Generation I: Static GIS Indices (Emerged 2003–2016)

Generation I encompasses 263 studies — spanning the bikeability-index, traffic-stress, level-of-service, and network-connectivity families — whose core shares a common computational architecture: (a) a GIS layer representing road network attributes; (b) a set of indicators derived from static geometric and infrastructural properties; (c) a composite weighting procedure to aggregate indicators into a scalar index. The LTS framework (Mekuria et al., 2012; Furth et al., 2016) provides the operational backbone for connectivity assessment, while Winters et al. (2013, 2016) establish the canonical multi-factor structure replicated across subsequent European indices.

Three methodological sub-patterns emerge within this generation. The first is infrastructure-centric models, prioritising the presence and quality of dedicated cycling facilities, speed limits, and traffic volumes (Grigore et al., 2019; Schmid-Querg et al., 2021; Hardinghaus et al., 2021). The second is destination-accessibility models, incorporating POI density, public transport connectivity, and land-use mix as cycling attractors (Winters et al., 2013; Wysling and Purves, 2022). The third is user-stratified models, which begin to differentiate index weights by cyclist type, purpose, or sociodemographic group (Arellana et al., 2020; Porter et al., 2020; Cabral and Kim, 2022). A parallel applied strand extends this stratification to children: school bikeability has recently been formalised as a distinct construct, with dedicated indicator sets for assessing safe cycling to school (Paulusová and Sharmeen, 2025). The systematic review by Ahmed et al. (2024) confirms that only 3 of 15 surveyed Generation I indices incorporate all five CROW design principles (safety, comfort, attractiveness, directness, coherence), with coherence the most consistently neglected.

The generation's structural ceiling is defined by its data sources: with the exception of transition studies that bridge toward behavioural validation (e.g., Cabral and Kim, 2022, classified I/II), Generation I studies rely on static vector road network data

https://doi.org/10.5194/isprs-archives-L-4-W3-2026-49-2026

(typically OSM or official cadastres), supplemented by census and land-use layers, and none incorporate real-time or behavioural data. Validation is predominantly cross-sectional and correlational, without causal identification of infrastructure effects on cycling demand. The weighting problem — how to aggregate heterogeneous indicators without imposing arbitrary cardinal assumptions — remains unresolved: within the extraction matrix alone, equal weighting (Lowry et al., 2012; Wysling and Purves, 2022), expert-based weighting (Grigore et al., 2019; Schmid-Querg et al., 2021; Hardinghaus et al., 2021), survey-based AHP (Arellana et al., 2020; Castañon et al., 2024) and regression-derived weights (Winters et al., 2016; Codina et al., 2022) coexist without a dominant standard, and sensitivity analysis of weighting choices is rarely reported.

3.4 Generation II: Dynamic Data and Demand Modelling (Emerged 2017–2022)

Generation II is distinguished epistemologically by treating observed cycling behaviour as the primary validation evidence. The 104 included studies leverage GPS trajectories, Strava Metro ridership data, or full national network datasets to move beyond theoretical network quality toward empirical demand modelling.

The pivotal contribution of Fosgerau et al. (2023) demonstrates through Copenhagen GPS data that cycling infrastructure induces demand: the existing bicycle lane network increases cycling trips by 60% and cycling kilometres by 90% compared to a counterfactual without infrastructure. This estimate derives from a discrete route-choice model calibrated on GPS trajectories, in which the counterfactual is constructed by removing dedicated infrastructure attributes from the network and re-simulating route and mode choices; its external validity is bounded by Copenhagen's mature cycling culture, and the magnitudes should not be generalised to other contexts without qualification. Reggiani et al. (2022) extend this paradigm by introducing bikeability curves and Pareto-optimal routes, formalising user heterogeneity without relying on subjective typologies.

Crowdsourced data — primarily Strava Metro — enables spatial monitoring of ridership change at scale (Boss et al., 2018; Fischer et al., 2022; Imrit et al., 2024), but introduces a systematic representativeness problem: Strava users are disproportionately male, higher-income, and fitness-oriented, with documented spatial biases toward established cycling corridors. Nelson et al. (2021) and Lee and Sener (2021) characterise this as an irreducible limitation of all fitness-app GPS sources, requiring cross-validation with official counts or travel surveys. Among the included studies, the responses differ markedly: Boss et al. (2018) mitigate the bias by cross-validating Strava volumes against automated bicycle counts ($r = 0.76$–0.96), whereas Fischer et al. (2022) and Imrit et al. (2024) acknowledge the bias without demographic correction — a distinction between studies that mitigate and studies that merely acknowledge the problem.

The most consequential finding of Generation II for future method development is Vierø et al.'s (2025) national-scale data quality assessment of Danish bicycle infrastructure data: neither OSM nor the official GeoDanmark cadastre are individually sufficient, requiring conflation and topology repair for reliable network analysis. This finding is particularly significant given that the entire Generation I and much of Generation II literature assumes OSM completeness. Vierø and Szell (2025) subsequently confirm that even in Denmark — a global best-case scenario for cycling data — bikeability is spatially concentrated in urban areas, with rural infrastructure fragmentation unresolved.

3.5 Generation III: GeoAI and Street-Level Intelligence (Emerged 2021–Present)

Generation III is epistemologically anchored in the visual experience of cycling conditions: Street View Imagery (SVI) as the primary data source for environmental assessment. The 44 included studies apply a range of computer-vision and sensor-based methods — semantic segmentation, object detection, contrastive deep learning, GAN-based image translation, Random Forest with EDA physiological data — whose outputs are georeferenced and aggregated to network units, deriving bikeability-relevant indicators at fine spatial granularity without site visits. It is this geospatial integration, rather than the image analysis alone, that qualifies these pipelines as GeoAI (Section 1).

Ito and Biljecki (2021) establish the foundational result: a 34-indicator bikeability index derived primarily from SVI and computer vision, applied across Singapore and Tokyo, achieves finer spatial granularity and broader coverage than conventional non-SVI indices at a fraction of the cost of manual audits; agreement with non-SVI baselines is assessed through cross-geography comparison rather than behavioural ground truth. Lin et al. (2024) automate the LTS classification — previously a manual, data-intensive task — through contrastive deep learning on 39,153 Toronto road segments, demonstrating feasibility of scalable Generation I → III bridging. Critically, Ito et al. (2024) identify and quantify the vehicle-centric perspective bias inherent in standard Google Street View capture: green view index and semantic ratios differ substantially between road-centre and road-shoulder perspectives ($r = 0.35$–0.55), a structural artefact corrected through GAN-based image translation achieving $r = 0.81$–0.83.

The explainability frontier is represented by Ye et al. (2024), who combine SVI environmental features with crash records and SHAP values to produce a tri-tiered cycling safety prediction across London's full road network — quantifying the relative contribution of road geometry, greenery, speed infrastructure, and intersection design to perceived risk. Moser et al. (2025) extend this further into embodied sensing: 89 participants in Osnabrück provided 1,780 geo-referenced cycling trips with simultaneous Electrodermal Activity (EDA) measurements, enabling a Random Forest model to map physiological stress onto spatial infrastructure features — the first direct physiological grounding of bikeability assessment. Guo et al. (2025) apply YOLOv5 at 98.9% mAP@0.5 to detect slow-mobility bottlenecks across Wuhan's Third Ring Road, introducing a dynamic optimisation mechanism with transfer learning for temporal updates.

The generation's unresolved challenges are: (1) GSV temporal and geographic coverage — standard GSV is unavailable in many Global South cities, and temporal lag introduces anachronism in rapidly changing urban environments; (2) perspective and occlusion artefacts (vehicles, vegetation, weather) that degrade CV accuracy; (3) the gap between environmental perception and actual cycling behaviour — high SVI bikeability scores do not guarantee cycling uptake without integrating demand-side modelling from Generation II. Future SVI/GeoAI indices should therefore be externally validated against independent counts, GPS trajectories, route-choice models, and longitudinal before–after infrastructure interventions rather than only image-label accuracy or crash association.

https://doi.org/10.5194/isprs-archives-L-4-W3-2026-49-2026

3.6 Emerging Frontiers: Open-Vocabulary Semantic Understanding and Foundation Models

The Generation III studies reviewed above rely predominantly on task-specific, fully supervised computer vision: semantic segmentation and object detection models trained on fixed ontologies of street-scene classes (e.g., PSPNet, YOLOv5) or on manually labelled LTS data (Lin et al., 2024). Recent advances in computer vision and photogrammetry are, however, reshaping what semantic understanding of the cycling environment is feasible. Vision–language foundation models such as CLIP (Radford et al., 2021) enable zero-shot recognition of concepts described in natural language; promptable segmentation models such as the Segment Anything Model (Kirillov et al., 2023) delineate arbitrary objects without task-specific retraining; and open-vocabulary detectors such as Grounding DINO (Liu et al., 2024) localise categories unseen at training time.

For bikeability, these capabilities are consequential in three respects: (i) they can detect long-tail, cycling-specific categories absent from standard street-scene ontologies — kerb ramps, cycle-lane surface defects, temporary obstructions, bicycle parking — without the costly re-labelling that constrains AutoLTS-style pipelines; (ii) they reduce the city-specific training burden that currently limits transferability; and (iii) combined with photogrammetric advances in image-based 3D reconstruction, they can populate the geometric–semantic layers of urban digital twins directly from imagery. Their limitations are equally relevant: domain gaps between web-scale training data and street-level cycling contexts, spurious or hallucinated detections that are difficult to audit, substantial computational cost, and — crucially for this review's terminology — outputs that remain in image space until georeferenced and aggregated to network units, i.e., before any of this constitutes GeoAI. No included study yet employs these models, marking a clearly identifiable research frontier.

3.7 Urban Digital Twins as an Integration Platform

Within the reviewed corpus, digital twins appear as an adjacent research direction rather than as an included empirical strand: the intersection between digital-twin research and cycling terminology is occupied almost entirely by homonyms and product-level twins (battery cycling, bicycle product models, athlete physiology), and of 2,065 deduplicated records only two pertinent contributions remain — enabling work on dynamic 4D web-based visualisation of streetspace activities from semantic 3D city models (Beil et al., 2022) and a simulation-oriented bicycle infrastructure modelling contribution — alongside adjacent network-design and routing-optimisation work (Steinacker et al., 2022; Lonardi et al., 2025), rather than operational bikeability twins. To clarify what such a platform entails — and to balance the prominence of the term in the title — Figure 3 schematises how an urban digital twin (UDT) for bikeability is constructed. The base layers are geometric and semantic: a three-dimensional city model and a routable cycling network graph, whose reliability depends on the conflation of open and official datasets highlighted by Vierø et al. (2025). Generation III methods populate the semantic-enrichment layer, georeferencing CV-derived indicators to network segments; Generation II demand models provide the behavioural engine for agent-based simulation; and continuous synchronisation with sensor and count data distinguishes a twin from a static 3D model. The application layer supports the counterfactual scenario testing that Gap 4 (Section 4) identifies as the field's principal validation need. In this architecture the UDT is not a fourth generation but an integration platform in which the three data epistemologies are combined and validated — plausible but still unrealised in the bikeability literature.

3.8 Comparative Synthesis: Three Generations

Table 1 synthesises the three generations across data sources, computational approaches, validation methods and structural limitations, distinguishing each generation's period of emergence from the publication years actually observed in the extraction matrix — a distinction that prevents the methodological strata from being read as mutually exclusive historical periods.

4. IDENTIFIED GAPS AND RESEARCH DIRECTIONS

4.1 Gap 1: Standardisation Deficit

Across all three generations, no consensus operational definition of bikeability exists. Indicator selection, weighting procedures, and spatial units of analysis vary substantially, preventing cross-city or cross-study comparability: the extraction matrix shows, within Generation I alone, spatial units ranging from road segments and OD pairs to grid cells and neighbourhoods, and weighting procedures ranging from equal weights to expert surveys, AHP and regression-derived schemes. Ahmed et al. (2024) document this for Generation I indices; the Schön et al. (2024) review of network connectivity confirms that even the most fundamental measure — connectivity — has no standardised operationalisation. Generation III methods compound this problem by introducing domain-specific indicator sets (SVI-derived) incommensurable with GIS-based predecessors.

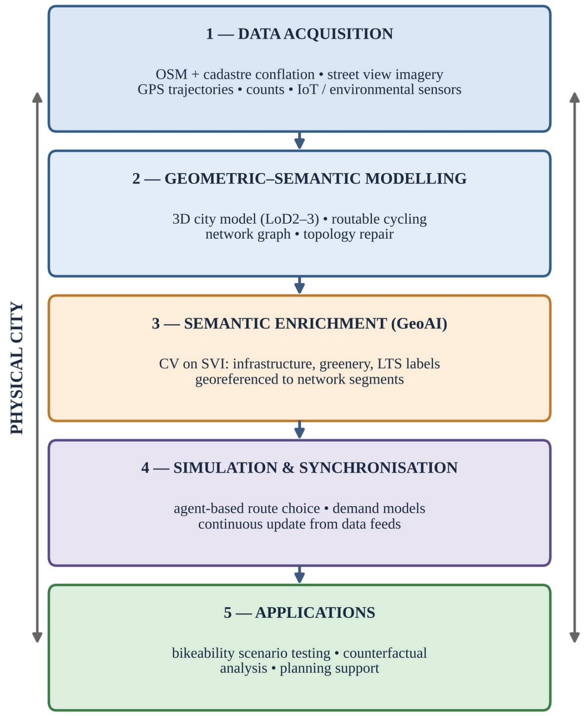

Figure 3. Construction of an urban digital twin for bikeability assessment: layered architecture from data acquisition to scenario-testing applications, with bidirectional synchronisation between the physical city and its digital counterpart.

Generation	n	Period of emergence	Years in matrix	Primary data sources	Computational approach	Validation method and strength	Structural limitation
Generation I (static GIS)	263	2003–2016	2003–2025	GIS, OSM, cadastre, census	Weighted overlay, AHP, LTS classification	Cross-sectional correlation with cycling rates; low	Static snapshot; no behavioural validation; standardisation deficit
Generation II (dynamic data)	104	2017–2022	2003–2025	GPS, Strava, crowdsourced OSM	Route choice models, Pareto optimisation, spatial autocorrelation	GPS trajectory hold-out, volume correlation, count validation; moderate–high (behavioural ground truth)	GPS user-selection bias; OSM data quality; rural under-coverage
Generation III (GeoAI)	44	2021–present	2021–2025	GSV, Mapillary, wearable sensors, crash data	Semantic segmentation, object detection, contrastive DL, GAN, RF+EDA	Cross-geography comparison, crash association, physiological sensing; moderate (behavioural validation absent)	GSV geographic/temporal gaps; Global South coverage; perception ≠ behaviour gap
Meta-reviews	12	—	2012–2025	Multi-database bibliographic	PRISMA / scoping review / narrative	n/a	Pre-2024 reviews miss Gen. III; no generational framework previously published

Table 1. Comparative synthesis of the three computational generations. n values refer to the full corpus of 425 included studies (two further simulation-platform contributions are discussed in Section 3.7); publication-year ranges refer to the full corpus.

4.2 Gap 2: Dynamic Variable Blindness

Current indices rarely integrate three categories of emerging urban cycling dynamics: (a) air quality and pollution exposure, documented as a primary cycling deterrent yet rarely integrated into the reviewed indices (Castañon and Ribeiro, 2021); (b) electric micromobility (e-bikes), which alters speed, route choice, and slope tolerance significantly; (c) bike-sharing system integration, which transforms the functional cycling network but is treated as external to bikeability measurement. Notably, these variables are not absent from adjacent literatures — cycling-exposure, health, and route-choice research measures them routinely — but they remain unintegrated into bikeability scoring. This omission is not methodologically incidental: it reflects the static data epistemology of Generation I and the fitness-user focus of Generation II GPS sources.

4.3 Gap 3: Scalability–Equity Trade-Off

Generation III methods offer superior spatial granularity but depend on GSV coverage and OSM completeness, both of which degrade in the Global South, rural areas, and lower-income urban neighbourhoods. Vierø and Szell (2025) confirm that even in Denmark, high bikeability is spatially concentrated in urban areas, effectively replicating rather than correcting infrastructure inequities. Gritton et al. (2024) document analogous bikeability disparities by race and socioeconomic status in Orange County, California. Generation III's analytical precision amplifies measurement of well-studied, well-covered areas — a scale-equity paradox with direct policy implications. Mitigation strategies are identifiable but rarely implemented: stratified SVI sampling, supplementation with crowdsourced imagery (Mapillary, KartaView) where GSV is unavailable, targeted field audits in low-coverage neighbourhoods, explicit uncertainty flags for missing infrastructure data, and equity-stratified validation against counts or surveys.

4.4 Gap 4: Validation Gap — Index to Modal Shift

The chain of evidence connecting bikeability measurement to actual modal shift remains fragile. Fosgerau et al. (2023) provide the strongest causal identification to date, but their Copenhagen results may not transfer to contexts without an established cycling culture. No reviewed Generation III study establishes a causal link between SVI-derived bikeability scores and cycling uptake. Closing this gap requires explicitly quasi-experimental designs: before–after studies, difference-in-differences analyses around infrastructure openings, panel travel surveys, and digital-twin scenario testing calibrated with observed counts. The integration of digital twin simulation — capable of testing infrastructure scenarios against demand models under dynamic conditions — remains nascent (Steinacker et al., 2022; Lonardi et al., 2025) and has not yet been applied to the full bikeability assessment pipeline (Section 3.7).

5. DISCUSSION

The three-generation model proposed in this review complements and extends existing systematic reviews. Kellstedt et al. (2021), Ahmed et al. (2024), and Castañon and Ribeiro (2021) offer important instrument catalogues and domain-specific gap analyses, but none characterise the computational evolution as a coherent epistemological progression. Biljecki and Ito (2021), conversely, provide a comprehensive cross-domain review of SVI technology across all urban analytics — of which cycling assessment is a peripheral application. The present work inverts this perspective: rather than mapping a technology across domains, it offers a structured synthesis of the major methodological families addressing a single analytical problem — to our knowledge, the first review organised around the specific GIS → behavioural-data → GeoAI computational transition.

The epistemological reading of the generational transitions clarifies why the gaps identified in Section 4 are persistent rather than incidental. The standardisation deficit (Gap 1) is an artefact of Generation I's paradigm: a field organised around bespoke local index construction will not converge on standard metrics without a shared coordination mechanism — which Generation III's automated scalability could theoretically provide if anchored in a shared indicator ontology. The dynamic variable blindness (Gap 2) is a data availability problem that Generation II's GPS approach partially addresses (e-bike trajectories and bike-share trip data are available but rarely integrated). The scalability–equity trade-off (Gap 3) reflects a genuine tension between Generation III's analytical power and its data infrastructure dependency — solvable in principle through crowdsourced SVI (Mapillary, KartaView) but requiring active equity-oriented sampling strategies. The validation gap (Gap 4) is the most methodologically consequential challenge: it requires not just methodological innovation but institutional access to mobility panel data and longitudinal before–after study designs.

A promising convergence path is visible in the literature: hybrid methods that combine GIS network analysis (Generation I's established validity) with GPS-calibrated demand models (Generation II's behavioural grounding) and SVI-based perceptual assessment (Generation III's scalability and cyclist-perspective orientation). Dai et al. (2023) approximate this in Xiamen, integrating OSM, deep neural network indicators, and GPS trajectories into a four-sub-index bikeability framework. The next step — represented in the reviewed literature only by adjacent contributions on network design and routing optimisation (Steinacker et al., 2022; Lonardi et al., 2025) rather than by a validated, full-pipeline implementation — is the integration of all three data epistemologies within a Digital Twin platform capable of counterfactual scenario testing (Figure 3). This convergence agenda sits squarely within the remit of ISPRS Technical Commission IV (Spatial Information Science), where GeoAI, 3D city modelling and urban digital twins are active research priorities; the semantic-understanding frontier discussed in Section 3.6 connects it equally to the photogrammetric computer-vision community.

This review has limitations. First, source coverage relied on open scholarly indexes (Semantic Scholar, Crossref, Consensus) rather than subscription databases; abstract coverage in these indexes is incomplete, no citation tracking was performed, and non-Anglophone and Global South contributions are likely under-represented — the restriction to English-language, peer-reviewed sources compounds this risk. Second, screening and extraction were performed by a single reviewer supported by automated screening tools, without double screening or inter-rater statistics; every decision carries a recorded reason code in the open decision log, but confirmation bias cannot be excluded. Third, the generational classification involves subjective judgements for transition-period studies, and the family assignment of the included corpus is lexicon-assisted with manual adjudication. Fourth, the PRISMA-ScR protocol, while appropriate for mapping an evolving field, does not support formal quality assessment of the included studies in the manner of a systematic review with risk-of-bias tools; claims about causal validity, comparative performance, and transferability should therefore be read as evidence mapping rather than evidence grading. Finally, the breadth of two decades of literature exceeds what a conference-format synthesis can report in full: the complete extraction matrix, query strings, and screening decision log are therefore available from the corresponding author on request.

6. CONCLUSIONS

To our knowledge, this scoping review is the first to map bikeability assessment as a three-part computational transition — from static GIS indices through behaviourally validated demand models to SVI/GeoAI methods — while identifying urban digital twins as a plausible but still under-validated integration platform. The 425 included studies, spanning 2003–2025, reveal a field in rapid methodological transition — with Generation III methods now capable of automating at scale what Generation I could only approximate manually — while four persistent gaps remain unresolved: standardisation, dynamic variable inclusion, scalability equity, and validation depth.

The generational model proposed constitutes an analytical vocabulary for researchers positioning new bikeability work and for practitioners evaluating methodological fit for specific urban contexts. Future research should prioritise: (1) cross-generational hybrid approaches combining GIS network analysis, GPS-calibrated demand models, and SVI-based perceptual assessment; (2) equity-oriented SVI sampling to extend Generation III coverage beyond well-documented cities; (3) longitudinal studies closing the bikeability-to-modal-shift validation loop through digital-twin counterfactual simulation. The immediate methodological priority is not another index alone, but a transparent validation pipeline linking infrastructure attributes, perceptual exposure, route choice, and observed cycling uptake across inequitable data-coverage contexts.

REFERENCES

Ahmed, T., Pirdavani, A., Wets, G., Janssens, D., 2024. Bicycle infrastructure design principles in urban bikeability indices: a systematic review. *Sustainability*, 16(6), 2545. https://doi.org/10.3390/su16062545

Ahmed, T., Pirdavani, A., Sitohang, I.F., Wets, G., Janssens, D., 2025. Assessment methods for bicycle environment safety and comfort: a scoping review. *Sustainable Futures*. 9, 100457, doi.org/10.1016/j.sftr.2025.100457.

Arellana, J., Saltarín, M., Larrañaga, A.M., González, V.I., Henao, C.A., 2020. Developing an urban bikeability index for different types of cyclists as a tool to prioritise bicycle infrastructure investments. *Transportation Research Part A*, 139, 310–334. https://doi.org/10.1016/j.tra.2020.07.010

Beil, C., Kendir, M., Ruhdorfer, R., Kolbe, T.H., 2022. Dynamic and web-based 4D visualization of streetspace activities derived from traffic simulations and semantic 3D city models. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, X-4/W2-2022, 29–36. doi.org/10.5194/isprs-annals-X-4-W2-2022-29-2022.

Bianconi, F., Filippucci, M., Seccaroni, M., Rolando, A., D'Uva, D., 2023. Machine learning and landscape quality. Representing visual information using deep learning-based image segmentation from street view photos. *SCIRES-IT - SCIentific RESearch and Information Technology* 13 (1), 117–134. doi.org/10.2423/i22394303v13n1p117

Biljecki, F., Ito, K., 2021. Street view imagery in urban analytics and GIS: a review. *Landscape and Urban Planning*, 215, 104217. https://doi.org/10.1016/j.landurbplan.2021.104217

Boss, D., Nelson, T., Winters, M., Ferster, C.J., 2018. Using crowdsourced data to monitor change in spatial patterns of bicycle ridership. Journal of Transport & Health 9, 226–233. https://doi.org/10.1016/j.jth.2018.02.008

Buehler, R., Dill, J., 2016. Bikeway networks: a review of effects on cycling. *Transport Reviews*, 36(1), 9–27. doi.org/10.1080/01441647.2015.1069908.

Cabral, L., Kim, A.M., Shirgaokar, M., 2019. Low-stress bicycling connectivity: assessment of the network build-out in Edmonton, Canada. Case Studies on Transport Policy 7 (2), 230–238. https://doi.org/10.1016/j.cstp.2019.04.002

Cabral, L., Kim, A.M., 2022. An empirical reappraisal of the level of traffic stress framework for segments. Travel Behaviour and Society 26, 143–158. https://doi.org/10.1016/j.tbs.2021.09.007.

Castañon, U.N., Ribeiro, P.J.G., 2021. Bikeability and emerging phenomena in cycling: Exploratory Analysis and Review, *Sustainability*, 13(4), 2394. https://doi.org/10.3390/su13042394

Castañon, U.N., Ribeiro, P.J.G., Mendes, J.F.G., 2024. Evaluating urban bikeability: a comprehensive assessment of

Póvoa de Varzim's network. *Sustainability*, 16(21), 9472. doi.org/10.3390/su16219472.

Codina, O., Maciejewska, M., Nadal, J., Marquet, O., 2022. Built environment bikeability as a predictor of cycling frequency: lessons from Barcelona. Transportation Research Interdisciplinary Perspectives 16, 100725. https://doi.org/10.1016/j.trip.2022.100725

Dai, S., Zhao, W., Wang, Y., Huang, X., Chen, Z., Lei, J., Stein, A., Jia, P., 2023. Assessing spatiotemporal bikeability using multi-source geospatial big data: a case study of Xiamen, China. *International Journal of Applied Earth Observation and Geoinformation* 125, 103539. doi.org/10.1016/j.jag.2023.103539.

Ewing, R., Cervero, R., 2010. Travel and the built environment: a meta-analysis. *Journal of the American Planning Association*, 76(3), 265–294. doi.org/10.1080/01944361003766766.

Fischer, J., Nelson, T., Winters, M., 2022. Riding through the pandemic: using Strava data to monitor the impacts of COVID-19 on spatial patterns of bicycling. Transportation Research Interdisciplinary Perspectives 15, 100667. doi.org/10.1016/j.trip.2022.100667

Fosgerau, M., Łukawska, M., Paulsen, M., Rasmussen, T.K., 2023. Bikeability and the induced demand for cycling. Proceedings of the National Academy of Sciences 120 (16), e2220515120. doi.org/10.1073/pnas.2220515120

Furth, P.G., Mekuria, M.C., Nixon, H., 2016. Network connectivity for low-stress bicycling. *Transportation Research Record*, 2587(1), 41–49. doi.org/10.3141/2587-06.

Gao, M., Fang, C., 2025. Pedaling through the cityscape: unveiling the association of urban environment and cycling volume through street view imagery analysis. Cities 156, 105573. doi.org/10.1016/j.cities.2024.105573.

Grigore, E., Garrick, N., Fuhrer, R., Axhausen, K.W., 2019. Bikeability in Basel. Transportation Research Record 2673 (6), 607–617. doi.org/10.1177/0361198119839982.

Gritton, J., Martinez, M.C., Bostean, G., Strong, M.T., 2024. Bikeability disparities in Orange County, California: intersection of place and demographics. SAGE Open 14 (2), 21582440241252477. doi.org/10.1177/21582440241252477.

Guo, Z., Xu, H., Lin, Q., 2025. Automated identification and spatial pattern analysis of urban slow-moving traffic bottlenecks using street view imagery and deep learning. ISPRS International Journal of Geo-Information 14 (9), 351. doi.org/10.3390/ijgi14090351.

Hardinghaus, M., Nieland, S., Lehne, M., Weschke, J.V., 2021. More than bike lanes—a multifactorial index of urban bikeability. Sustainability 13 (21), 11584. doi.org/10.3390/su132111584.

Imrit, A.A., Fischer, J., Chan, T.C.Y., Saxe, S., Bonsma-Fisher, M., 2024. A street-specific analysis of level of traffic stress trends in Strava bicycle ridership and its implications for low-stress bicycling routes in Toronto. Findings, 92109. doi.org/10.32866/001c.92109.

Ito, K., Biljecki, F., 2021. Assessing bikeability with street view imagery and computer vision. Transportation Research Part C: Emerging Technologies 132, 103371. doi.org/10.1016/j.trc.2021.103371.

Ito, K., Quintana, M., Han, X., Zimmermann, R., Biljecki, F., 2024. Translating street view imagery to correct perspectives to enhance bikeability and walkability studies. International Journal of Geographical Information Science 38 (12), 2514–2544. doi.org/10.1080/13658816.2024.2391969.

Kellstedt, D.K., Spengler, J.O., Foster, M., Lee, C., Maddock, J.E., 2021. A scoping review of bikeability assessment methods. Journal of Community Health 46 (1), 211–224. doi.org/10.1007/s10900-020-00846-4.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.-Y., Dollár, P., Girshick, R., 2023. Segment anything. In: Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV), pp. 4015–4026.

Lee, K., Sener, I.N., 2021. Strava Metro data for bicycle monitoring: a literature review. Transport Reviews 41 (1), 27–47. doi.org/10.1080/01441647.2020.1798558.

Lin, B., Saxe, S., Chan, T.C.Y., 2024. AutoLTS: automating cycling stress assessment via contrastive learning and spatial post-processing. Proceedings of the AAAI Conference on Artificial Intelligence 38 (20), 22222–22230. doi.org/10.1609/AAAI.V38I20.30227.

Liu, S., Zeng, Z., Ren, T., Li, F., Zhang, H., Yang, J., Jiang, Q., Li, C., Yang, J., Su, H., Zhu, J., Zhang, L., 2024. Grounding DINO: marrying DINO with grounded pre-training for open-set object detection. In: European Conference on Computer Vision (ECCV), Lecture Notes in Computer Science 15105. Springer, Cham, pp. 38–55. doi.org/10.1007/978-3-031-72970-6_3.

Lonardi, A., Szell, M., De Bacco, C., 2025. Cohesive urban bicycle infrastructure design through optimal transport routing in multilayer networks. Journal of the Royal Society Interface 22 (223), 20240532. doi.org/10.1098/rsif.2024.0532.

Lowry, M.B., Callister, D., Gresham, M., Moore, B., 2012. Assessment of communitywide bikeability with bicycle level of service. Transportation Research Record 2314 (1), 41–48. doi.org/10.3141/2314-06.

Łukawska, M., Paulsen, M., Rasmussen, T.K., Jensen, A.F., Nielsen, O.A., 2023. A joint bicycle route choice model for various cycling frequencies and trip distances based on a large crowdsourced GPS dataset. Transportation Research Part A: Policy and Practice 176, 103834. doi.org/10.1016/j.tra.2023.103834.

Mekuria, M.C., Furth, P.G., Nixon, H., 2012. Low-stress bicycling and network connectivity. Mineta Transportation Institute Report 11-19, San Jose, CA.

Moser, M.K., Schmidt, S., Graf, D.R.M., Keskin, M., Gandhi, S., Yap, W., Zeile, P., Heinke, M., Resch, B., 2025. Understanding the influence of urban characteristics on cyclists' stress measured through wearable sensors: a quantitative open data approach. Environment and Planning B: Urban Analytics and City Science 53 (2), 336–362. https://doi.org/10.1177/23998083251394426.

Nelson, T., Ferster, C., Laberee, K., Fuller, D., Winters, M., 2021. Crowdsourced data for bicycling research and practice. Transport Reviews 41 (1), 97–114. doi.org/10.1080/01441647.2020.1806943.

Paulusová, I., Sharmeen, F., 2025. School bikeability – what is it, and why is it important? An overview of key indicators and measurement. *Transport Reviews*, 45(1), 49–76. doi.org/10.1080/01441647.2024.2405218.

Porter, A.K., Kohl, H.W., Pérez, A., Reininger, B., Pettee Gabriel, K., Salvo, D., 2020. Bikeability: assessing the objectively measured environment in relation to recreation and transportation bicycling. *Environment and Behavior*, 52(8), 861–894. https://doi.org/10.1177/0013916518825289

Pucher, J., Buehler, R., 2008. Making cycling irresistible: lessons from The Netherlands, Denmark and Germany. *Transport Reviews*, 28(4), 495–528. doi.org/10.1080/01441640701806612.

Pucher, J., Dill, J., Handy, S., 2010. Infrastructure, programs, and policies to increase bicycling: an international review. *Preventive Medicine*, 50(S1), S106–S125. doi.org/10.1016/j.ypmed.2009.07.028.

Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., Sutskever, I., 2021. Learning transferable visual models from natural language supervision. In: Proceedings of the 38th International Conference on Machine Learning (ICML), Proceedings of Machine Learning Research 139, pp. 8748–8763.

Reggiani, G., van Oijen, T., Hamedmoghadam, H., Daamen, W., Vu, H.L., Hoogendoorn, S., 2022. Understanding bikeability: a methodology to assess urban networks. Transportation 49, 897–925. doi.org/10.1007/s11116-021-10198-0.

Rita, L., Nathvani, R., Peliteiro, M., Bostan, T.-C., Muller, E., Suel, E., Metzler, A.B., Tamagusko, T., Ferreira, A., 2023. Using deep learning and Google Street View imagery to assess and improve cyclist safety in London. Sustainability 15 (13), 10270. doi.org/10.3390/su151310270.

Schmid-Querg, J., Keler, A., Grigoropoulos, G., 2021. The Munich bikeability index: a practical approach for measuring urban bikeability. Sustainability 13 (1), 428. doi.org/10.3390/su13010428.

Schön, P., Heinen, E., Manum, B., 2024. A scoping review on cycling network connectivity and its effects on cycling. Transport Reviews 44 (4), 912–936. doi.org/10.1080/01441647.2024.2337880.

Semler, C., Sanders, M., Buck, D., Graham, J., Pochowski, A., Dock, S., 2017. Low-stress bicycle network mapping: the District of Columbia's innovative approach to applying level of traffic stress. Transportation Research Record 2662 (1), 31–40. doi.org/10.3141/2662-04.

Steinacker, C., Storch, D.-M., Timme, M., Schröder, M., 2022. Demand-driven design of bicycle infrastructure networks for improved urban bikeability. Nature Computational Science 2, 655–664. doi.org/10.1038/s43588-022-00318-w.

Tricco, A.C., Lillie, E., Zarin, W., O'Brien, K.K., Colquhoun, H., Levac, D., Moher, D., Peters, M.D.J., Horsley, T., Weeks, L., Hempel, S., Akl, E.A., Chang, C., McGowan, J., Stewart, L., Hartling, L., Aldcroft, A., Wilson, M.G., Garritty, C., Lewin, S., Godfrey, C.M., Macdonald, M.T., Langlois, E.V., Soares-Weiser, K., Moriarty, J., Clifford, T., Tunçalp, Ö., Straus, S.E., 2018. PRISMA extension for scoping reviews (PRISMA-ScR): checklist and explanation. Annals of Internal Medicine 169 (7), 467–473. doi.org/10.7326/M18-0850.

Vierø, A.R., Vybornova, A., Szell, M., 2025. How good is open bicycle network data? A countrywide case study of Denmark. Geographical Analysis 57 (1), 52–87. doi.org/10.1111/gean.12400.

Vierø, A.R., Szell, M., 2025. Network analysis of the Danish bicycle infrastructure: bikeability across urban-rural divides. Geographical Analysis 57 (4), 616–640. doi.org/10.1111/gean.70012.

Wang, K., Akar, G., Lee, K., Sanders, M., 2020. Commuting patterns and bicycle level of traffic stress (LTS): insights from spatially aggregated data in Franklin County, Ohio. Journal of Transport Geography 86, 102751. doi.org/10.1016/j.jtrangeo.2020.102751.

Winters, M., Brauer, M., Setton, E.M., Teschke, K., 2013. Mapping bikeability: a spatial tool to support sustainable travel. Environment and Planning B: Planning and Design 40 (5), 865–883. doi.org/10.1068/b38185.

Winters, M., Teschke, K., Brauer, M., Fuller, D., 2016. Bike Score®: associations between urban bikeability and cycling behavior in 24 cities. International Journal of Behavioral Nutrition and Physical Activity 13, 18. doi.org/10.1186/s12966-016-0339-0.

Wysling, L., Purves, R.S., 2022. Where to improve cycling infrastructure? Assessing bicycle suitability and bikeability with open data in the city of Paris. Transportation Research Interdisciplinary Perspectives 15, 100648. doi.org/10.1016/j.trip.2022.100648.

Ye, Y., Zhong, C., Suel, E., 2024. Unpacking the perceived cycling safety of road environment using street view imagery and cycle accident data. Accident Analysis & Prevention 205, 107677. doi.org/10.1016/j.aap.2024.107677.