Supplementary Materials

This appendix contains the supporting documentation for the systematic review reported in the main text: the complete search strategy (Table S1), the data-extraction and coding dictionary (Table S2), the characteristics and individual results of the included studies (Table S3a), the study-level critical appraisal (Table S3b), the retrospective independent-verification agreement summary (Table S3c), the completed PRISMA 2020 checklist (Table S4), and the full-text exclusion summary (Table S5). Section numbers refer to the main text.

Table S1. Information sources and complete database search strategy.

Identical Boolean expression entered in each of the five databases:

("augmented reality" OR "virtual reality" OR "mixed reality" OR "extended reality" OR XR OR WebXR OR "digital twin" OR immersive OR "gaussian splatting" OR "neural rendering" OR "procedural model*") AND ("archaeological park" OR "archaeological site" OR "ancient city" OR "ancient cities" OR "historic district" OR "urban heritage" OR "historic city" OR insula* OR "city block*" OR "street network" OR cityscape OR "urban fabric" OR "urban planning") AND (reconstruct* OR heritage OR archaeolog* OR "cultural" OR planning OR navigat*)

Database	Platform/interface	Fields	Coverage	Interface filters/limits	Screening language	Last search	Records
Scopus	Elsevier Scopus web interface	Title, abstract, keywords	2010-2026	No language or report-type interface filter	English full reports required at screening	15 July 2026	2,660
Web of Science Core Collection	Clarivate Web of Science web interface	Title, abstract, keywords	2010-2026	No language or report-type interface filter	English full reports required at screening	15 July 2026	1,122
IEEE Xplore	IEEE Xplore web interface	Title, abstract, keywords	2010-2026	No language or report-type interface filter	English full reports required at screening	15 July 2026	771
ACM Digital Library	ACM Digital Library web interface	Title, abstract, keywords	2010-2026	No language or report-type interface filter	English full reports required at screening	15 July 2026	2,622
SpringerLink	Springer Nature SpringerLink web interface	Title, abstract, keywords	2010-2026	No language or report-type interface filter	English full reports required at screening	15 July 2026	3,615
Total	Five database interfaces	Same Boolean expression and field scope	2010-2026	No interface language/report-type limit	English full reports required at screening	15 July 2026	10,790

No supplementary hand-searching of journals or conference proceedings and no backward or forward citation searching were

conducted. Google Scholar was not used for study identification. Database exports were combined before deduplication and screening.

Table S2. Thirty-five-field data-extraction and coding dictionary

This dictionary operationalizes the five blocks described in Section 2.5. NR = not reported; N/A = not applicable. Missing or unclear information is not imputed.

No.	Field	Definition	Allowed values/coding rule
1	Study ID	Unique corpus identifier linked to the manuscript reference number.	Reference number / internal ID
2	Full citation	Authors and bibliographic citation.	Text
3	Publication year	Year of first peer-reviewed publication, including online-first publication.	2010–2026
4	Venue and report type	Journal/conference and eligible report format.	Journal article / full conference paper
5	DOI/URL and source	Persistent identifier and database in which the record was retrieved.	DOI/URL; database
6	Application domain	Primary domain addressed by the report.	Archaeological site/park / historic district / urban planning / supporting
7	Country and case study	Geographic location and named site/city/project.	Text / NR
8	Context typology	Type of place or planning context.	Ancient city / park / building / district / contemporary city / landscape
9	Spatial scale	Highest level at which meaningful interaction or analysis occurs.	S1 object/room; S2 building; S3 block; S4 itinerary; S5 district; S6 city/territory
10	Multi-scale capability	Whether users or analyses move between more than one scale.	Yes / No / NR
11	Delivery setting	Relationship between system use and the physical place.	On-site / off-site / hybrid or remote-collaborative / not applicable
12	Purpose/RQ contribution	System purpose and review question(s) informed.	Interpretation / education / planning / management / navigation / enabling
13	XR modality	Immersive modality used or enabled.	AR / VR / MR / multimodal XR / immersive digital twin / GIS-only enabling
14	Acquisition/data source	Source of geometry, imagery, documents, or urban data.	Survey / photogrammetry / laser scan / archive / GIS / BIM / mixed / NR
15	Modeling/reconstruction	Method used to create extant, reconstructed, or proposed content.	Manual / parametric / procedural / photogrammetric / hybrid / NR
16	Integration layer	Connection to structured information or simulation.	GIS / BIM/HBIM / digital twin / semantic layer / none / NR
17	Software/engine	Principal authoring, analysis, or rendering environment.	Named software/version / NR
18	Delivery device/platform	User-facing hardware or delivery channel.	Mobile / HMD / CAVE / desktop / web / UAV / hybrid / not applicable

No.	Field	Definition	Allowed values/coding rule
19	Tracking/localization	Method used to register or position content.	Marker/image / markerless / GNSS-inertial / depth / none / NR
20	Interaction/navigation	How users move, select, manipulate, or collaborate.	Text
21	Content/narrative/gamification	Interpretive content and experience-design paradigm.	POI / narrative / game / simulation / co-design / none / NR
22	Target users/stakeholders	Intended users or participating stakeholder group.	Visitors / students / experts / planners / citizens / mixed / NR
23	Evaluation design	Human or technical evaluation design.	Randomized / comparative / within-subject / questionnaire / qualitative / benchmark / descriptive
24	Sample size	Number of participants or technical test observations.	Integer / technical dataset size / NR / not applicable
25	Sample composition/recruitment	Participant characteristics, recruitment, and intended-population fit.	Text / NR / not applicable
26	Outcome domains/instruments	Human outcomes, technical metrics, and named instruments.	Text / NR
27	Follow-up/time points	Timing of outcome measurement or repeat exposure.	Immediate / specified interval / none / NR
28	Quantitative results	Numerical results retained with group, time point, and statistic.	As reported / NR
29	Qualitative results	Themes, observations, or author-reported qualitative findings.	As reported / NR
30	Technical performance	Frame rate, tracking error, latency, accuracy, complexity, or related metrics.	As reported / NR / not applicable
31	Reported benefits	Benefits supported or claimed by the study.	Text
32	Limitations/critical issues	Study limitations and implementation constraints.	Text / NR
33	Author-proposed future research	Future work stated by the study authors.	Text / NR
34	Evidence class	Role of the report in the review.	A direct application / B enabling method / C foundational survey
35	Evidence strength and rationale	Study-level judgment used to weight narrative claims.	High / medium / low / indirect + reason

Table S3a. Characteristics and principal individual results of the 21 included studies

Reference numbers correspond to the main manuscript. Class A = direct application; Class B = enabling method; Class C = foundational survey. S1 = object/room; S2 = building; S3 = block/insula; S4 = itinerary; S5 = district; S6 = city/territory. NR = not reported; N/A = not applicable.

Study	Year/class	Domain/case	Technology	Scale/mode	Evaluation/sample	Principal result and contribution to synthesis
[14] Bekele et al.	2018 / C	Cultural heritage; multiple settings	Survey of AR, VR, and MR	Various / N/A	Foundational survey; no participant sample applicable	Taxonomy of technologies, immersion levels, and application purposes; used for background and definition alignment.
[26] Yağcı and Bölükbaşı Ertürk	2025 / A	Konya People's House, Türkiye; urban heritage	Mobile AR; archive-to-3D workflow	S2 / on-site	Comparative technical assessment of three AR platforms; no participant sample	Documents and drawings were transformed into a 3D reconstruction and AR interface; Unity-Vuforia was judged the most comprehensive option, but testing was limited in devices/profiles and relied on expert observation/ordinal scores.
[27] Pagano et al.	2020 / A	Paestum, Italy; archaeological park	HMD-VR narrative experience	S2 / on-site installation	Two observational /questionnaire evaluation sessions; n = 145 in total	Participants were generally positive; reported highlights included natural interaction, strong immersion, limited discomfort, and positive memorability feedback, without a controlled comparator.
[28] Natephra et al.	2017 / A	Building and urban-lighting design	BIM-to-VR lighting workflow	S2 / off-site	Technical validation at six points under three lighting conditions	The BIM/VR lighting workflow was quantitatively compared with physical luxmeter measurements, Radiance, and 3ds Max; errors were reported for the technical claim.
[29] Campanaro and Landeschi	2022 / A	Pompeii, Italy; House of the Greek Epigrams	VR eye tracking and 3D GIS	S3 / off-site laboratory	Exploratory within-participant study; n = 5	Gaze and fixation patterns varied across reconstructed domestic spaces and lighting conditions; the sample was exploratory.
[30] Imottesjo and Kain	2022 / A	Urban CoCreation Lab; urban planning	Web desktop, HMD-VR, and mobile AR	S5 / hybrid and remote-collaborative; multiscale	Iterative stakeholder workshops: phase 1 n = 8; phase 2 n = 2	Stakeholder trials informed minimum viable/marketable product specifications; synchronized volumetric co-design changes propagated in under approximately five seconds.
[31] Melelli et al.	2023 / A	Perugia, Italy; urban scientific heritage	Personalized mobile AR trails	S4 / on-site	Descriptive application;	Profile-based routes and points of interest connect geological and

Study	Year/class	Domain/case	Technology	Scale/mode	Evaluation/sample	Principal result and contribution to synthesis
					sample size NR	naturalistic heritage across the city.
[32] Panou et al.	2018 / A	Chania old town, Greece; urban heritage	Mobile outdoor AR	S4 / on-site	Technical tests at three monuments; no human sample	Tracking and registration were tested outdoors in Chania, but no user sample or robust quantitative error, latency, or stability measures were reported.
[33] Zhao et al.	2026 / A	Hubu Mountain, Xuzhou, China; historic district	Mobile narrative AR game	S5 / on-site	Randomized controlled experiment; n = 80	AR improved cultural knowledge and dissemination intention; the reported difference in immersion was not statistically significant.
[34] Kyrlitsias et al.	2020 / A	Choirokoitia, Cyprus; archaeological settlement	HMD-VR versus desktop virtual tour	S5 / off-site	Controlled between-group comparison; n = 40	HMD-VR increased presence and user experience, while the desktop condition produced the larger reported knowledge gain (d approximately 0.71).
[35] J. Chen et al.	2025 / B	Historic streetscape, Suzhou, China	Big/small-data decision support for AR	S5 / N/A	Mixed big/small-data observational study; visitor/expert questionnaire n not verified	Social-media, physical-spatial, questionnaire, GIS, and OLS data identify demand and candidate locations for future AR interventions; this does not test an implemented AR intervention.
[36] Rzeszewski and Orylski	2021 / A	Poznań, Poland; urban planning	WebXR planning visualization	S5 / off-site web	Exploratory usability study; purposive sample n = 15	Six WebXR environments (four VR and two AR) were tested. SUS scores ranged from 65 to 90; three participants scored below 68, and device/browser limitations involved anchoring, interface intrusiveness, and legibility.
[37] Yang et al.	2026 / A	Urban areas from an aerial perspective; planning	Depth-aware UAV augmented reality	S5 / aerial/on-site	Technical benchmark; 9,476 RGB-depth training pairs; no human participants	The manuscript reports 27 fps and occlusion IoU values of 0.973 offline and 0.892 in flight.
[38] Sánchez Berriel et al.	2023 / A	San Cristóbal de La Laguna,	City-scale mobile AR	S6 / on-site	Technical and heuristic evaluation	The application sustained more than 60 fps; the square-circuit

Study	Year/class	Domain/case	Technology	Scale/mode	Evaluation/sample	Principal result and contribution to synthesis
		Spain; historic city			plus qualitative test; n = 5 users	positional error was 3.63 +/- 0.28 m, with GPS and compass behavior the main user-reported problems.
[39] Dylla et al.	2010 / A	Rome Reborn; ancient city	Procedural virtual-city reconstruction	S6 / off-site	Technical case study	The model contains more than 7,000 buildings: approximately 250 documented Class I buildings and approximately 6,750 procedurally generated Class II buildings.
[40] Pietroni et al.	2023 / A	Eight Italian archaeological parks	Web/VR platform, ATON, digital models, and 360-degree panoramas	S6 network / hybrid	National-program case study; no ex-post user evaluation	Open infrastructure, FAIR metadata, and reliability coding support access and maintainability across eight parks; the authors state that ex-post UX evaluation remains to be conducted.
[41] S. Chen et al.	2022 / A	New York, Tokyo, and London; urban navigation	UrbanRama HMD-VR	S6 / off-site	Expert study n = 11 and experiment n = 30	Focus-plus-context navigation increased street-level time and reduced perspective switching while preserving spatial judgments.
[42] Dembski et al.	2020 / A	Herrenberg, Germany; urban digital twin	CAVE/HMD digital twin	S6 / hybrid workshops	Public workshops and non-representative questionnaire; n = 39	The study reports that 95% perceived a positive effect on local planning and 97% on public participation; no implemented decision outcome was measured.
[43] Moreira et al.	2026 / A	Pelourinho, Salvador, Brazil; historic district	Survey-to-model mobile AR	S5 / on-site	Design Science Research implementation case; no end-user testing	Presents an end-to-end digitization and AR workflow; comprehensive user testing is described as future work.
[44] Limongiello et al.	2025 / B	Ancient Abellinum archaeological area, Italy	HGIS/GIS, UAV, TLS, and HBIM	S6 / N/A	Integrated-survey method	A multiscale documentation and management system integrates approximately 25 ha of survey data, UAV photogrammetry, TLS, GIS, and HBIM.
[45] Peng et al.	2025 / B	West Lake, China; urban-	GIS-based visual analysis	S6 / N/A;	Spatial-analysis case study	The study identifies 37 landscape types and reports that only 2 of

Study	Year/class	Domain/case	Technology	Scale/mode	Evaluation/sample	Principal result and contribution to synthesis
		heritage landscape		multiscale		10 potential city-lake visual corridors were unobstructed; used as enabling evidence.

Table S3b. Study-level critical appraisal of the 21 included studies

D1 = appropriateness of design for the claim; D2 = adequacy of participant sample or technical test data; D3 = validity/completeness of outcome measures or performance metrics; D4 = comparator, control, or technical baseline; D5 = transparency and reproducibility. Ratings: Yes, Partly, No, N/A, or NR. High = at least four applicable Yes ratings with no critical No; Medium = at least three Yes ratings with one important limitation; Low = descriptive, exploratory, or uncontrolled evidence; Indirect = Class B/C evidence not used for direct effectiveness claims. The original appraisal was carried out by a single reviewer. Following peer review, a second reviewer independently applied the D1–D5 rubric and the overall evidence-strength classification to the included reports. All disagreements were subsequently resolved by discussion. The ratings reported in this table represent the reconciled consensus judgments, while pre-reconciliation agreement statistics are reported in Table S3c. Applying the rubric to the 21 included studies yields five studies rated high, two medium, ten low, and four indirect. These are study-level judgments of design adequacy and reporting completeness, not body-of-evidence certainty ratings.

Study	D1	D2	D3	D4	D5	Overall	Reason
[14]	Yes	Partly	N/A	N/A	Yes	Indirect	Foundational survey used for taxonomy, not for direct effectiveness claims.
[26]	Partly	No	Partly	Yes	Partly	Low	Three AR platforms were compared using expert observation and ordinal judgments, but devices/profiles were limited and quantitative validation was absent.
[27]	Partly	Partly	Partly	No	Partly	Low	Two uncontrolled event-based observational/questionnaire sessions (approximately 70 and 75 participants) provide useful feedback but no causal comparator.
[28]	Yes	Yes	Yes	Yes	Yes	High	Technical validation used six points, three lighting conditions, physical measurements, and two established simulation baselines; High applies to its technical claim, not user effectiveness.
[29]	Partly	No	Partly	Partly	Partly	Low	Pilot integration of objective eye tracking and 3D GIS with n = 5; results and formal comparisons were limited.
[30]	Yes	Partly	Partly	N/A	Yes	Low	The formative co-design method fits specification development, but

Study	D1	D2	D3	D4	D5	Overall	Reason
							the two workshop phases used small samples (n = 8 and n = 2).
[31]	Partly	NR	NR	N/A	Partly	Low	Project description without a formal user study or outcome/performance metrics.
[32]	Partly	Partly	Partly	Partly	Partly	Low	Outdoor technical tests at three monuments lacked a user sample and robust quantitative performance measures.
[33]	Yes	Yes	Partly	Yes	Yes	High	Random allocation, n = 80, comparator, explicit outcomes, and statistical analysis.
[34]	Yes	Partly	Yes	Yes	Yes	High	Randomized n = 40 comparison with presence, experience, and knowledge outcomes.
[35]	Yes	Partly	Partly	N/A	Partly	Indirect	Mixed-data demand analysis with visitor/expert questionnaires and OLS; no implemented AR intervention was evaluated.
[36]	Yes	Partly	Yes	Partly	Yes	Medium	Named usability instrument and informative device observations; sampling and comparator limitations constrain inference.
[37]	Yes	Yes	Yes	Yes	Partly	High	Large technical dataset, defined baseline, real-time and flight tests, and explicit performance metrics; some implementation detail remains limited.
[38]	Yes	Partly	Yes	Partly	Yes	Medium	Defined technical metrics and field constraints; human evaluation remains preliminary.
[39]	Yes	Partly	Partly	N/A	Partly	Low	Technical city-reconstruction case without a user-effect evaluation; reporting is sufficient for the main procedural-model claim but incomplete for reproducibility.
[40]	Partly	NR	NR	N/A	Yes	Low	Detailed multi-site program report; the authors state that ex-post UX evaluation remains to be performed.
[41]	Yes	Partly	Yes	Yes	Yes	High	Qualitative n = 11 and randomized quantitative n = 30 studies with comparative spatial-behavior outcomes; the written threshold yields High.
[42]	Partly	Partly	Partly	No	Yes	Low	Non-representative questionnaire n = 39 attached to workshops; no comparator or implemented decision outcome.
[43]	Yes	Partly	Partly	N/A	Yes	Low	Transparent Design Science Research workflow; end-user evaluation is future work.
[44]	Yes	Yes	Yes	Partly	Yes	Indirect	Strong enabling GIS/HBIM survey method, but not direct XR effectiveness evidence.

Study	D1	D2	D3	D4	D5	Overall	Reason
[45]	Yes	Yes	Yes	Partly	Partly	Indirect	Enabling spatial-analysis evidence; transfer to XR remains indirect.

Table S3c. Pre-reconciliation inter-reviewer agreement in the retrospective independent verification of the included reports.

Observed agreement and Cohen's κ are based on the two independent codings before consensus discussion. Eligibility κ is not estimable because both reviewers classified every included report as eligible. Spatial scale is ordinal; therefore, weighted κ values are additionally reported. Overall evidence strength is based on 20 valid paired ratings because one non-valid second-reviewer entry was treated as missing for the agreement calculation. All disagreements were subsequently discussed and resolved by consensus, and the reconciled classifications are reported in Tables S3a and S3b.

Variable	Valid paired ratings, n	Observed agreement (%)	Cohen's κ	Linear weighted κ	Quadratic weighted κ
Eligibility	21	100.0	N/E	—	—
Evidence class	21	100.0	1.000	—	—
Spatial scale	21	61.9	0.507	0.712	0.871
Delivery mode	21	81.0	0.745	—	—
D1–D5 appraisal domains (pooled)	105	66.7	0.478	—	—
Overall evidence strength	20	70.0	0.600	—	—

Note: κ = Cohen's kappa coefficient; N/E = not estimable because both reviewers assigned all reports to a single category. Weighted κ values are reported only for spatial scale because the S1–S6 classification is ordinal. The D1–D5 row pools the five appraisal-domain judgments across the 21 included reports (5 × 21 = 105 paired ratings). Overall evidence strength is based on 20 valid paired ratings because one non-valid second-reviewer entry was excluded from the agreement calculation. All statistics refer to the independent ratings obtained before reconciliation.

Table S4. Completed PRISMA 2020 checklist

The checklist below refers to the revised manuscript. Section numbers refer to the revised main text; page numbers should be confirmed against the final typeset proof. Entries recorded as "Not applicable" identify PRISMA items that do not apply because no statistical synthesis, certainty rating, or review protocol was undertaken; the single item recorded as "Not met" is reported together with its reason.

Item	PRISMA 2020 requirement	Status	Location/comment
1	Identify the report as a systematic review in the title.	Complete	Title.
2	Report all PRISMA for Abstracts items.	Complete	Abstract. Reports the objectives, information sources and final search date, eligibility criteria,

Item	PRISMA 2020 requirement	Status	Location/comment
			synthesis method, number of included studies, principal results, limitations, funding, and registration status.
3	Describe the rationale in the context of existing knowledge.	Complete	Sections 1.1-1.3.
4	State explicit objectives or questions.	Complete	Section 1.4 (RQ1-RQ6).
5	Specify eligibility criteria and how studies were grouped for synthesis.	Complete	Sections 2.3 and 2.7. Section 2.3 states the inclusion and exclusion criteria and the analytical groups used for the synthesis (evidence class, application domain, XR modality, spatial scale, delivery mode); the eligible corpus comprises direct applications, enabling methods, and one foundational survey.
6	List all information sources, platforms, and the date each was last searched.	Complete	Section 2.2 and Table S1.
7	Present the full search strategy, limits, and filters for every source.	Complete	Section 2.2 and Table S1; the same expression and field scope are reported for all five databases.
8	Describe the selection process, reviewers, independence, and automation.	Complete	Section 2.4. The original title/abstract and full-text screening was performed by one reviewer. Following peer review, a second reviewer independently re-assessed the eligibility of all 21 included reports; agreement was 100%. The excluded-record pool was not independently re-screened. No screening-automation tool was used.
9	Describe the data-collection process, reviewers, clarification, and automation.	Complete	Section 2.5 and Table S3c. Original data extraction was performed by one reviewer. Following peer review, a second reviewer independently recoded the key synthesis variables for all 21 included reports. Disagreements were resolved by discussion. Study authors were not contacted and no automation or figure-extraction software was used.
10a	List and define all outcomes and rules for selecting results.	Complete	Section 2.5 and Table S2. The 35-field dictionary reports the name, definition, and admissible values or coding rule for every extracted field.
10b	List other variables and assumptions about missing or unclear information.	Complete	Section 2.5 and Table S2. Missing or unclear information was recorded as "not reported" and was not imputed; "not applicable" is defined in Table S2.
11	Describe study-level critical-appraisal methods and reviewer process.	Complete	Section 2.6 and Tables S3b–S3c. The D1–D5 rubric was initially applied by one reviewer and subsequently applied independently by a second reviewer to the included corpus. Pre-reconciliation agreement is reported in Table S3c; all disagreements were resolved by discussion and Table S3b reports the final consensus judgments.
12	Specify effect measures for each outcome.	Complete	Section 2.7 states that human and technical results were retained in their reported form and that no effects were calculated or converted by the reviewer.

Item	PRISMA 2020 requirement	Status	Location/comment
13a	Explain how studies were assigned to each synthesis.	Complete	Sections 2.3 and 2.7; Table S3a.
13b	Describe data preparation and handling of missing statistics.	Complete	Sections 2.5 and 2.7; NR, no imputation, and no conversion.
13c	Describe tabulation and visual-display methods.	Complete	Section 2.7; main Table 1, Figures 2-4, and Table S3a.
13d	Describe and justify synthesis methods and software.	Complete	Section 2.7. Structured narrative synthesis stratified by modality, domain, spatial scale, and delivery setting; the section states explicitly that no meta-analysis was performed because designs, participants, interventions, outcome constructs, and instruments were not sufficiently comparable.
13e	Describe methods used to explore causes of heterogeneity.	Complete	Narrative stratification by evidence class, modality, domain, scale, setting, pipeline, and evaluation family; no statistical heterogeneity analysis.
13f	Describe sensitivity analyses.	Not applicable	No statistical synthesis was performed.
14	Describe assessment of bias due to missing results.	Complete	Section 2.7 explains why no funnel plot or outcome-level statistical tool was used and discusses residual risk in Section 8.6.
15	Describe assessment of certainty in the body of evidence.	Complete	Section 2.7 states that no GRADE-equivalent body-of-evidence rating was performed and explains why.
16a	Report study-selection results, ideally with a flow diagram.	Complete	Section 2.4 and Figure 1. PRISMA 2020 flow diagram for a new review (databases and registers), reporting records identified, duplicates removed, records screened and excluded, reports sought and not retrieved, reports assessed and excluded with reasons, and studies included.
16b	Cite reports that appeared eligible but were excluded and explain why.	Partially met	Tables S5a-S5b and Section 8.6. The study-level exclusion log was not retained, so the aggregate counts in Table S5a cannot be resolved into individual records. Table S5b cites reports within the scope of the search that could reasonably appear eligible, with the primary reason each was not included; the list is illustrative rather than exhaustive.
17	Cite each included study and present its characteristics.	Complete	Section 3.6 and Table S3a. All 21 included studies are cited, with evidence class, application domain and case study, technology, spatial scale, delivery mode, evaluation design and sample where available, and principal result.
18	Present study-level critical-appraisal results.	Complete	Section 3.6 and Table S3b. Domain ratings, overall evidence strength, and reasons are reported for every included study; Section 3.6 summarizes five high, two medium, ten low, and four indirect.
19	Present individual-study results for all relevant outcomes.	Complete	Sections 4-7 and Table S3a. Results are reported in the units used by the original studies; "NR" marks details that the source reports did not provide.

Item	PRISMA 2020 requirement	Status	Location/comment
20a	Summarize characteristics and appraisal of studies contributing to each synthesis.	Complete	Sections 3-8 and Tables S3a-S3b, which identify the studies contributing to each synthesis and their appraisal.
20b	Present results of each statistical synthesis.	Not applicable	No statistical synthesis was performed.
20c	Present investigations of heterogeneity.	Complete	Narrative comparisons by domain, modality, scale, and setting in Sections 3-8; no statistical heterogeneity analysis.
20d	Present sensitivity analyses.	Not applicable	No statistical synthesis was performed.
21	Present assessments of bias due to missing results.	Complete	No formal assessment; rationale and residual publication/selective-reporting risk are reported in Sections 2.7 and 8.6.
22	Present certainty assessments.	Not applicable	No body-of-evidence certainty rating was performed; the rationale and limits to confidence are stated in Sections 2.7 and 8.6.
23a	Interpret results in the context of other evidence.	Complete	Sections 7-8.
23b	Discuss limitations of the included evidence.	Complete	Section 8.6.
23c	Discuss limitations of the review process and their likely impact.	Complete	Section 8.6, including single-reviewer screening/extraction, no supplementary searching, no registration/protocol, and lack of exclusion log.
23d	Discuss implications for practice, policy, and research.	Complete	Sections 8-10.
24a	Provide registration information or state that the review was not registered.	Complete	Abstract and Section 2.1: not registered.
24b	Provide access to the protocol or state that none was prepared.	Complete	Abstract and Section 2.1: no formal protocol.
24c	Describe protocol/registration amendments.	Not applicable	No protocol or registration existed.
25	Describe financial and non-financial support and the funder's role.	Complete	Abstract and Funding statement: no external funding.
26	Declare competing interests.	Complete	Conflicts of Interest statement.
27	State availability of forms, extracted data, analytic data/code, and other materials.	Complete	Data Availability Statement and Tables S1-S5. All data supporting the review are contained in the article and its Supplementary Materials; the full-text exclusion log is not available, as stated in Section 8.6.

Table S5a. Aggregate full-text exclusion categories

Primary exclusion reason	Reports
Ineligible archaeological-heritage or urban context	82
No eligible immersive, XR, digital-twin, or reconstructive method	58
Ineligible publication type or insufficient methodological detail	36
Total full-text reports excluded	176

Note on PRISMA item 16b: The study-level log of reports excluded after full-text assessment was not retained, so the aggregate

counts in Table S5a cannot be resolved into individual records. To address PRISMA item 16b as fully as the surviving records permit, Table S5b lists reports that fall within the scope of the documented search strategy and could reasonably appear eligible to a reader, together with the primary reason each was not included in the corpus. The list combines reports confirmed by the reviewer as assessed at full text and excluded, and reports cited elsewhere in this article that were considered and not included; in both cases the reason was assigned by applying the eligibility criteria stated in Section 2.3. The list is illustrative rather than exhaustive and does not enumerate all 176 full-text exclusions.

Table S5b. Reports that could reasonably appear eligible and the primary reason for their exclusion

Bracketed numbers correspond to the reference list of the main manuscript; reports without a bracketed number are not cited elsewhere in the article. Each report is assigned a single primary reason, using the same three categories as Table S5a.

Report	Primary reason for exclusion	Clarifying note
[49] Cecere, L., Colace, F., Lorusso, A., & Santaniello, D. (2024). Predictive maintenance of an archaeological park: an IoT and digital twin based approach. Book chapter, pp. 323-341. doi:10.1007/978-3-031-60611-3_23	No eligible immersive, XR, digital-twin, or reconstructive method	IoT and digital-twin framework for predictive maintenance of an archaeological park; the system supports monitoring and management but delivers no immersive or reconstructive experience.
[46]Kyropoulou, D., Karalis, P., Dotsika, E., Rizou, F., et al. (2025). A 3D geospatial platform and AI tour application for visualizing archaeological, analytical, and conservation data: the case of the Vryokastro archaeological ensemble, Kythnos Island. Journal of Archaeological Science: Reports, 68, 105480. doi:10.1016/j.jasrep.2025.105480	No eligible immersive, XR, digital-twin, or reconstructive method	Web-based 3D geospatial platform and AI-guided virtual tour for archaeological, analytical, and conservation data; no AR, VR, or MR modality is implemented and no user evaluation is reported.
[25] Gabellone, F. (2023). Development of an immersive VR experience using integrated survey technologies and hybrid scenarios. Heritage, 6(2), 1169-1186. doi:10.3390/heritage6020065	Ineligible archaeological-heritage or urban context	Immersive VR experience built on integrated survey data for a single monument and a hybrid indoor scenario; the site is not treated in explicit relation to a street, block, or district, so the single-monument boundary rule in Section 2.3 applies.
[47] Cicalese, F., Grimaldi, M., & Fasolino, I. (2024). The applicability of the urban digital twin in the detailed choices of the urban plan. Conference proceedings, pp. 211-220. doi:10.1007/978-3-031-54118-6_20	No eligible immersive, XR, digital-twin, or reconstructive method	Urban digital twin applied to detailed planning choices through conventional two-dimensional and GIS interfaces, with no immersive delivery layer.
[50] Tortora, A., Amaro, I., Della Greca, A., & Barra, P. (2025). Exploring the role of generative artificial intelligence in virtual reality: opportunities and future perspectives. Book chapter, pp. 125-142. doi:10.1007/978-3-031-93700-2_9	Ineligible publication type or insufficient methodological detail	Perspective chapter; no implemented system, case study, or evaluation is reported.

Report	Primary reason for exclusion	Clarifying note
[58] Xu, W. (2026). Cultural narratives through augmented reality (AR): hidden stories in the Chinatown Storytelling Centre. Digital Applications in Archaeology and Cultural Heritage, 40, e00476. doi:10.1016/j.daach.2025.e00476	Ineligible archaeological-heritage or urban context	AR storytelling delivered inside an interpretive centre; the experience has no spatial relationship to the surrounding historic district.
[54] White, G., Zink, A., Codecá, L., & Clarke, S. (2021). A digital twin smart city for citizen feedback. Cities, 110, 103064. doi:10.1016/j.cities.2020.103064	No eligible immersive, XR, digital-twin, or reconstructive method	City-scale digital twin for citizen feedback delivered through a web interface; no immersive modality is implemented or evaluated.
[59] Pinto, I., & Huertas, A. (2025). A comparative study of VR and AR heritage applications on visitor emotional experiences. Virtual Reality, 29(1), 36. doi:10.1007/s10055-025-01109-0	Ineligible archaeological-heritage or urban context	Comparative evaluation of VR and AR heritage applications experienced at attraction and museum level, below the district-scale threshold set in Section 2.3.
[62]Franczuk, J., & Koszewski, K. (2025). Algorithmic design in architectural heritage: innovation in virtual reconstruction of the Roman Forum Transitorium in Musti, Tunisia. Heritage, 8(9), 362. doi:10.3390/heritage8090362	Ineligible archaeological-heritage or urban context	Algorithmic and HBIM reconstruction of a single monumental complex, with interactive visualisation of competing reconstruction hypotheses; the complex is not analysed in relation to the surrounding street network or urban fabric.
[48]Shojaei, D., Sabri, S., & Rajabifard, A. (2026). Integrating emerging technologies to streamline urban planning, cadastral registration, and building approvals. In Digital Transformation of Building Permitting (pp. 169-185). CRC Press. doi:10.1201/9781003650287-13	No eligible immersive, XR, digital-twin, or reconstructive method	Overview of enabling technologies for planning, cadastral, and building-approval workflows; the chapter addresses administrative and regulatory processes without an immersive or reconstructive system.
[55] Lercari, N. (2017). 3D visualization and reflexive archaeology: a virtual reconstruction of Çatalhöyük history houses. Digital Applications in Archaeology and Cultural Heritage, 6, 10-17. doi:10.1016/j.daach.2017.03.001	No eligible immersive, XR, digital-twin, or reconstructive method	Virtual reconstruction presented through desktop 3D visualization rather than an immersive system.
[51] Clarizia, F., Colace, F., De Santo, M., Lorusso, A., Marongiu, F., & Santaniello, D. (2022). Augmented reality and gamification technics for visit enhancement in archaeological parks. 2022 IEEE 2nd IoT Vertical and Topical Summit for Tourism (IoTT), pp. 1-4. doi:10.1109/IoTT56174.2022.9925901	Ineligible publication type or insufficient methodological detail	Four-page summit contribution; the report does not provide sufficient architectural or evaluation detail for extraction under the coding scheme in Section 2.5.
[52] Casillo, M., Cecere, L., Colace, F., Lombardi, M., Lorusso, A., & Santoriello, A. (2024). Serious gaming for cultural heritage: the case study of the Domus of Abellinum. Book chapter, pp. 203-214. doi:10.1007/978-981-97-3305-7_16	No eligible immersive, XR, digital-twin, or reconstructive method	Serious game for a single domus delivered on a non-immersive platform, without an AR, VR, or MR component.
[6]Mursal, U., Ustaoglu, A. O., Baskose, Y., Yalcin, I., Kocaman, S., &	No eligible immersive, XR,	Photogrammetric digital twin used for geotechnical rockfall-hazard analysis at a

Report	Primary reason for exclusion	Clarifying note
Gokceoglu, C. (2025). Rockfall hazard assessment for natural and cultural heritage site: close vicinity of Rumkale (Gaziantep, Türkiye) using digital twins. Heritage, 8(7), 270. doi:10.3390/heritage8070270	digital-twin, or reconstructive method	heritage site; the model supports risk assessment and does not deliver an immersive, interpretive, or reconstructive experience.