Supporting information

Synthetic data in electron microscopy. AI co-expert-assisted review

Daria M. Arkhipova,[a] Alexey S. Kashin,[a] Artem P. Vorozhtsov,[a] Valentine P. Ananikov[a,b]*

[a] *N.D.Zelinsky Institute of Organic Chemistry, Russian Academy of Sciences, 119991 Moscow, Russian Federation*
[b] *RUDN University, 117198 Moscow, Russian Federation*

Review preparation checklist

Table S1. Checklist describing basic info for the review article preparation.

Stage	Checklist field to fill	Information completed by authors
I. Define topic, scope and review concept	Review title	Synthetic data in electron microscopy
	Review type	Systematic review based on focused literature selection
	Central message/key hypothesis of the review	Strategically designed synthetic data can augment or replace manual annotation in electron microscopy, overcoming data scarcity and bridging the simulation-to-reality gap for deep learning applications.
	Review scope (what is included)	Synthetic data generation, processing and usage in scanning electron microscopy (SEM), transmission (TEM), scanning transmission electron microscopy (STEM), cryo-electron microscopy/tomography (cryo-EM/ET), and volume electron microscopy.
	Out-of-scope topics (what is excluded)	Optical microscopy; non-image synthetic data (numerical, tabular)
	Years covered by the review	2020–2026 (primary contemporary focus)
	Document types included	Articles, reviews, conference papers, preprints
	Description of AI systems utilized in the preparation of this review	Qwen3.5-27B-FP8 (self-hosted *via* vLLM) Deepseek-V4 (*via* UI on chat.deepseek.com)
	AI usage at this stage	Not used
II. Plan review strategy and workflow	Initial search keywords, keyword blocks	Synthetic data, simulated data, electron microscopy (SEM, TEM, STEM)
	Exact search strings (with Boolean logic if used)	'Synthetic data' AND ('SEM' OR 'TEM' OR 'STEM') ('electron microscopy' OR 'TEM' OR 'SEM') AND ('synthetic data' OR 'simulated data') and manual search
	Authors and areas of expertise	D.M.Arkhipova – synthetic data, chemistry A.S.Kashin – electron microscopy, microscopy data A.P.Vorozhtsov – AI development and applications V.P.Ananikov – chemistry, biology, microscopy, nanotechnology, AI in science, digital technologies
	AI usage at this stage	Not used
III. Perform broad literature search	Databases searched for abstracts and metadata	Scopus, SciFinder, OpenAlex, CoLab (https://colab.ws)
	Date of search	March 2026
	Search period covered	2004–2026
	Total number of records found before deduplication	2153
	Number of records found per database (if available)	Colab – 649 Scopus – 238 SciFinder – 252 OpenAlex – 1014
	Total number of unique records (after deduplication)	1348
	AI usage at this stage	Not used

© 2026 Uspekhi Khimii, ZIOC RAS, Russian Academy of Sciences

Table S1 (continued).

Stage	Checklist field to fill	Information completed by authors
IV. Screen and select relevant studies	Number of selected records after matching to the topic	151
	Inclusion criteria	Synthetic image data generation and application; learning and validation on synthetic data in electron microscopy (SEM, TEM/STEM, cryo-EM/ET, and volume-EM) articles, preprints, reviews, conference materials
	Exclusion criteria	Optical microscopy; non-relevant to the subject
	Screening level used (title/abstract/keyword/manual full screening)	Title, abstract, keyword screening; full-text review when eligibility remains unclear
	AI usage at this stage	Qwen LLM and API for literature screening; human verified
V. Retrieve full-text articles	Number of full-text articles analyzed	151
	Number of articles excluded from coverage at full-text stage	0
	AI usage at this stage	Not used
VI. Expand and complete literature collection	Number of additional studies included	24
	Final number of studies in the literature set	175
	AI usage at this stage	Not used
VII. Analyze literature and build structure	Main conceptual themes identified	Contents: Main headings – 18 Subheadings – 77
	Basis for section structure	Hybrid section structure
	AI usage at this stage	QWEN and Deepseek in chat mode were used for topics sorting, content analysis, build table of contents; human verified
VIII. Assign studies to sections	Method used for assigning studies to sections	Full-text articles consideration in respect to complete table of contents structure
	Whether studies were allowed to appear in multiple sections	Yes
	AI usage at this stage	Qwen and Deepseek were used to assist with assigning articles to sections; human verified
IX. Write individual sections	Sections preparation strategy	Each section was written individually, with the text and figures prepared together. Complete table of contents and review structure was taken into account upon preparation of each section
	AI usage at this stage	QWEN and Deepseek were used to prepare drafts of the text; human verified
X. Integrate sections and evaluate	How consistency across sections was checked	Manual cross-checking by authors
	AI usage at this stage	Not used
XI. Verify facts, references and originality	How factual claims were checked against source literature	Manual checking
	Tool used for reference integrity verification	Manual verification by human
	Tools and thresholds used for originality control	iThenticate
	How AI-generated errors or hallucinations were checked	Human-expert checking
	AI usage at this stage	Not used
XII. Revise, polish and update	Article processing made	Complete read of the review by all co-authors, corrections and revisions (multiple times)
	Final literature update performed on (date)	23.07.2026
	Number of late-added references included during final update	23
	AI usage at this stage	English and style polishing at some places of the text; human verified
Overall transparency summary	Total number of databases searched	4

Table S1 (continued).

Stage	Checklist field to fill	Information completed by authors
Overall transparency summary	Total number of unique records in raw search for screening	1348
	Total number of selected studies	151
	Additional studies located manually	47
	Total number of studies finally included in the review	198
	Total number of pages in the review	56 (without references)
	Total of figures and tables	Figures – 30 Tables – 6