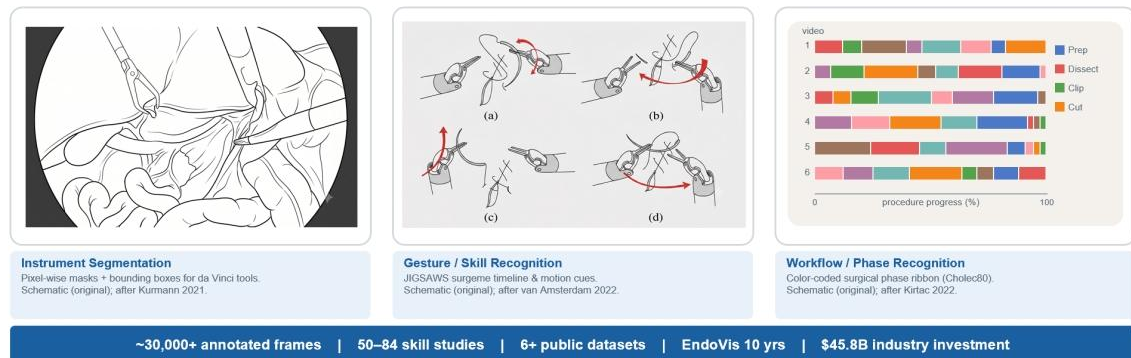


Surgery: Mature Kinematic AI Ecosystem



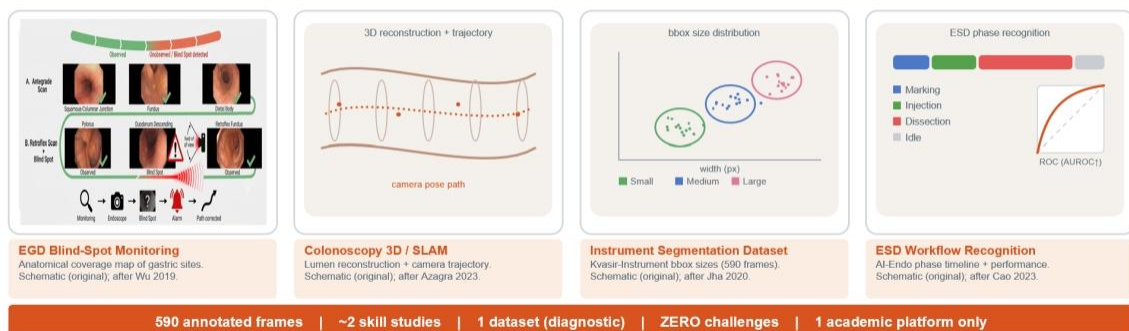
Supplementary Figure 1 The structural ecosystem gap: Artificial intelligence-driven kinematics in surgery vs gastrointestinal endoscopy (surgery: Mature kinematic artificial intelligence ecosystem). AI: Artificial intelligence; Each figure was redrawn schematically based on the original text.

(Left) **Kurmann T**, Márquez-Neila P, Allan M, Wolf S, Sznitman R. Mask then classify: multi-instance segmentation for surgical instruments. *Int J Comput Assist Radiol Surg* 2021; **16**: 1227-1236 [PMID: 34143374 DOI: 10.1007/s11548-021-02404-2]

(Middle) **van Amsterdam B**, Funke I, Edwards E, et al. Gesture Recognition in Robotic Surgery With Multimodal Attention. *IEEE Trans Med Imaging* 2022; **41**: 1677-1687 [PMID: 35108200 DOI: 10.1109/TMI.2022.3147640]

(Right) **Kirtac K**, Aydin N, Lavanchy JL, Beldi G, Smit M, Woods MS, Aspart F. *Appl Sci* 2022; **12**: 8746 [DOI: 10.3390/app12178746]

GIE: Fragmented Kinematic AI



Supplementary Figure 2 The structural ecosystem gap: Artificial intelligence-

driven kinematics in surgery vs gastrointestinal endoscopy (gastrointestinal endoscopy: Fragmented kinematic artificial intelligence). AI: Artificial intelligence;

GIE: Gastrointestinal endoscopy; SLAM: Simultaneous localization and mapping; SfM:

Structure from motion; GI: Gastrointestinal; ESD: Endoscopic submucosal dissection.

Each figure was redrawn schematically based on the original text.

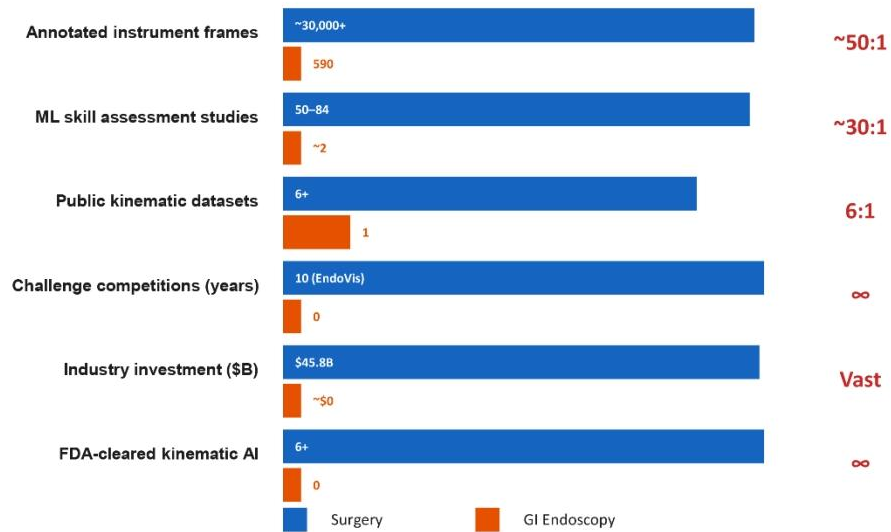
(Left) **Wu L**, Zhang J, Zhou W, An P, Shen L, Liu J, Jiang X, Huang X, Mu G, Wan X, Lv X, Gao J, Cui N, Hu S, Chen Y, Hu X, Li J, Chen D, Gong D, He X, Ding Q, Zhu X, Li S, Wei X, Li X, Wang X, Zhou J, Zhang M, Yu HG. Randomised controlled trial of WISENSE, a real-time quality improving system for monitoring blind spots during esophagogastroduodenoscopy. *Gut* 2019; **68**: 2161-2169 [PMID: 30858305 DOI: 10.1136/gutjnl-2018-317366]

(Middle 1) **Azagra P**, Sostres C, Ferrández Á, Riazuelo L, Tomasini C, Barbed OL, Morlana J, Recasens D, Batlle VM, Gómez-Rodríguez JJ, Elvira R, López J, Oriol C, Civera J, Tardós JD, Murillo AC, Lanás A, Montiel JMM. Endomapper dataset of complete calibrated endoscopy procedures. *Sci Data* 2023; 10: 671 [PMID: 37789003 DOI: 10.1038/s41597-023-02564-7]

(Middle 2) **Jha D**, Ali S, Emanuelsen K, Hicks SA, Thambawita V, Garcia-Ceja E, Riegler MA, de Lange T, Schmidt PT, Johansen HD, Johansen D, Halvorsen P. Kvasir-Instrument: Diagnostic and therapeutic tool segmentation dataset in gastrointestinal endoscopy. 2020. Available from: <https://arxiv.org/abs/2011.08065>

(Right) **Cao J**, Yip HC, Chen Y, Scheppach M, Luo X, Yang H, Cheng MK, Long Y, Jin Y, Chiu PW, Yam Y, Meng HM, Dou Q. Intelligent surgical workflow recognition for endoscopic submucosal dissection with real-time animal study. *Nat Commun* 2023; **14**: 6676 [PMID: 37865629 DOI: 10.1038/s41467-023-42451-8]

Quantified Gap: Surgery vs GI Endoscopy Kinematic AI



Supplementary Figure 3 The structural ecosystem gap: Artificial intelligence-driven kinematics in surgery *vs* gastrointestinal endoscopy (quantified gap: Surgery *vs* gastrointestinal endoscopy kinematic artificial intelligence). AI: Artificial intelligence; GI: Gastrointestinal.

Supplementary Table 1 PROBAST + artificial intelligence risk of bias assessment for 38 technical studies

Study	Domain	D1: Participants and data	D2: Predictors/features	D3: Outcome/ground truth	D4: Analysis	Overall
Lux (2023) ⁵⁶	WS	Unclear but sample)	(multicenter convenience specified)	Low (DL features pre-specified)	Low (documentation quality pre-defined)	High (no external validation; no calibration)
Freedman (2020) ³²	Cov	High selected videos only)	(single-center; Low features pre-defined)	Unclear consensus reference; no protocol)	(expert as small test set)	High (no external validation; no formal)
Wu (2019) ⁵⁹	Cov	Low (large dataset 170K images; multiple sources)	Low (DNN features automated)	Low annotation verification)	(expert with only; no calibration reported)	Unclear (internal validation)
Chen (2025) ³⁶	WF	Unclear dataset)	(single-center Low features automated)	Low (Transformer level annotation; multi-reviewer)	Low (expert frame-annotation; multiple validation)	Unclear (public benchmark; cross-baselines; cross-validation)
Biffi (2025) ³⁷	WF	Low frames; multi-	(large-scale 2.7M Low (TCN features automated)	Unclear protocol)	(annotation not fully comprehensive)	Low (public benchmark; metrics)

		procedure)			described)		
Zhang (2026) ⁷⁸	WF	High (very small dataset 33 videos)	Low (Mamba architecture specified)	Low (expert pre-annotation)	High (no external validation; High overfitting risk with small n)		
Wittbrodt (2024) ⁶³	Skill	Unclear (single-center retrospective)	Low (random forest features pre-specified)	Low (expert assessment reference)	High (no external validation; High as small sample)		
Cold (2024) ⁷⁹	Skill	Low (systematic review of 17 studies)	Low (N/A (systematic review))	Low (N/A (systematic review))	Low (systematic search; Low quality assessment included)		
Martin (2020) ⁴¹	Nav	High (animal model only; 2 pigs)	Low (CNN features automated)	Unclear (successful navigation formally defined)	High (no external validation; High not very small n)		
Hwang (2026) ⁴²	Nav	High (simulation only; no clinical data)	Low (supervised DL features pre-specified)	Unclear (simulation success metrics; clinical correlation)	Unclear (multiple environments tested but all simulated)		
Corsi (2023) ⁴³	Nav	High (simulation only)	Low (constrained RL features automated)	Low (formal safety verification pre-defined)	Unclear (simulation only; formal methods applied)		

Prendergast (2018) ⁶⁴	Nav	High (phantom only)	Low (ML localization features pre-specified)	Low (known phantom geometry as ground truth)	High (no external validation; High phantom only)
Huang (2021) ⁶⁵	Nav	High (simulation only)	Low (magnetic control features automated)	Unclear (navigation success in simulation)	High (no external validation)
Lazo (2022) ⁶⁶	Nav	High (simulation only)	Low (DL visual servoing features automated)	Low (target reaching criteria pre-defined)	High (simulation only; no external validation)
Pore (2022) ⁶⁷	Nav	High (simulation only)	Low (end-to-end features automated)	RL Unclear (simulation completion metrics)	High (simulation only; no external validation)
Tan (2025) ⁶⁸	Nav	High (simulation only)	Low (human-intervention features automated)	Low (safety metrics RL formally defined)	Unclear (simulation with human-in-loop; no clinical data)
Ma (2021) ⁴⁴	SLAM	Unclear (small dataset 18 videos)	Low (RNN-SLAM features automated)	Low (camera pose ground truth available)	High (no external validation; High small dataset)
Ozyoruk (2021) ⁴⁵	SLAM	Low (public dataset 35K frames; diverse)	Low (self-supervised features automated)	Low (synthetic + real ground truth depth)	Low (public benchmark; Low multiple baselines provided)

Bonilla (2024) ⁴⁶	SLAM	Unclear (limited endoscopy videos)	Low (3DGS features automated)	Unclear (reconstruction quality metrics only)	High (no external validation; High limited comparisons)
Wang (2024) ⁴⁷	SLAM	Unclear (endoscopy videos; source not fully described)	Low (3DGS+SLAM features automated)	Low (reconstruction metrics pre-defined)	Unclear (real-time demo but limited validation)
Recasens (2021) ⁶⁹	SLAM	Unclear (limited endoscopy videos)	Low (DL depth+motion features automated)	Low (ego-motion ground truth)	High (no external validation) High
Shao (2022) ⁷⁰	SLAM	Unclear (endoscopy videos; limited diversity)	Low (self-supervised features automated)	Low (depth ground truth available)	High (no external validation) High
Hayoz (2023) ⁷¹	SLAM	Unclear (endoscopy videos; limited diversity)	Low (DL pose estimation automated)	Low (camera pose ground truth)	Unclear (internal validation with ablation)
Azagra (2023) ⁷²	SLAM	Low (40 procedures; multi-source)	Low (SLAM baseline features automated)	Low (calibrated camera poses ground truth)	Low (public dataset with Low as baselines)

Bobrow (2023) ⁷³	SLAM	Unclear (38 videos; single source)	Low (NeRF features automated)	Low (paired depth ground truth from SfM)	Low (public dataset; multiple baselines)	Unclear
Shi (2023) ⁷⁴	SLAM	Unclear (colonoscopy videos; limited diversity)	Low (ColonNeRF features automated)	Low (reconstruction metrics pre-defined)	High (no external validation)	High
Elvira (2024) ⁷⁵	SLAM	Unclear (full procedures; limited n)	Low (SLAM automated)	Low (CUDA-SIFT-features completeness metrics)	Unclear (mapping demonstrated on full procedures but no external validation)	Unclear
Guo (2025) ⁸⁰	SLAM	Unclear (sparse endoscopic views)	Low (UC-NeRF features automated)	Low (reconstruction quality metrics)	High (no external validation)	High
Kaletka (2026) ⁸¹	SLAM	Unclear (endoscopy videos; limited n)	Low (3DGS features automated)	Low (rendering quality metrics)	High (no external validation)	High
Trovato (2010) ⁴⁸	Rob	High (simulation only)	Low (Q-learning features automated)	Unclear (locomotion success not formally defined)	High (simulation only; no validation)	High
Turan (2019) ⁴⁹	Rob	High (simulation only)	Low (deep RL features automated)	Low (navigation metrics pre-defined)	High (simulation only; no external validation)	High

Incetan (2021) ⁵⁰	Rob	High (virtual environment only)	Low (deep RL features automated)	Low (VR environment metrics pre-defined)	Unclear (virtual environment; scenarios tested)	High (multiple scenarios tested)
Jha (2020) ⁵¹	Rob	Low (590 clinical images; verified)	Low (expert segmentation automated)	Low (DL features annotated bounding boxes)	Low (expert-masks + provided)	Low (public dataset; baseline)
Ali (2021) ⁵²	Rob	Low (challenge dataset; multi-source)	Low (Transformer/CNN features automated)	Low (expert annotation; reviewer)	Low (challenge multi-multiple teams validated)	Low (format; Low)
Band (2022) ⁵³	Rob	Unclear (multicenter retrospective; convenience sample)	Low (DL automated)	Low (features annotation)	High (expert validation; small test set)	High (limited external High)
Jha (2025) ⁵⁴	Rob	Low (multiple benchmark datasets)	Low (various features automated)	Low (DL benchmark annotations)	Low (standardized benchmark comprehensive comparison)	Low (challenge format; Low)
Zhang (2022) ⁷⁶	Rob	High (simulation only)	Low (deep RL features automated)	Low (coverage metrics pre-defined)	High (simulation external validation)	High (only; no High)

Ng (2024) ⁷⁷	Rob	High only)	(simulation Low (deep RL features Low (navigation High (simulation only; no High
		only)	automated) metrics pre-defined) external validation)

D1: Participants and data sources; D2: Predictors/features; D3: Outcome/ground truth; D4: Analysis. WS: Withdrawal speed monitoring; Cov: Coverage/blind spot mapping; WF: Workflow recognition; Nav: Autonomous navigation; SLAM: Simultaneous localization and mapping/3D reconstruction; Rob: Robotic control/instrument tracking; AI: Artificial intelligence; DL: Deep-learning; DNN: Deep neural network; RL: Reinforcement learning; NeRF: Neural radiance field; SLAM: Simultaneous localization and mapping; ML: Machine learning; TCN: Temporal Convolutional Network; CNN: Convolutional neural network; 3DGS: 3D Gaussian Splatting; VR: Virtual reality; SfM: Structure from Motion.

Supplementary Table 2 The structural gap: Artificial intelligence-driven kinematics in surgery *vs* gastrointestinal endoscopy

Dimension		Surgery (AI kinematics)		GIE (AI kinematics)		Gap
First kinematic dataset		JIGSAWS, 2014 (76 variables, 103 trials) ¹⁰		No public kinematic dataset		10+ years
Instrument segmentation frames		Approximately 30000+ (EndoVis 2015-2024) ¹¹		590 (Kvasir-Instrument) ⁵¹		Approximately 50:1
ML skill assessment studies		50–84 (per systematic reviews) ^{17,18}		Approximately 2		Approximately 30:1
Workflow/phase recognition		Cholec80, 2016 (> 900 citations, > 91% accuracy) ¹²		ESD workflow began 2021 ³³		5 years
First autonomous procedure (animal)		STAR in vivo pig suturing, 2016 ¹³		Simulation only (2025)		10+ years
First autonomous laparoscopic surgery		STAR laparoscopic 2022 ¹⁴		pig, No equivalent		N/A
Remote robotic procedure		Operation Lindbergh, 2001; 73+ human cases ¹⁶		No remote GIE		20+ years
Annual challenge competitions		EndoVis, 9 editions (2015-2024), approximately 40 sub-challenges		Zero kinematic challenges		9 editions
FDA-cleared intraoperative AI		Moon Surgical Maestro, 2022 ¹⁵		None (diagnostic CAdE only)		Structural
Public datasets (kinematic)		6+ (JIGSAWS, DESK, ROSMA, HeiChole, <i>etc.</i>)		1 (Kvasir-Instrument, diagnostic)		6:1
Industry		\$45.8B cumulative		Single academic		Orders of

localization and mapping"[tiab] OR "3D reconstruction"[tiab] OR "instrument tracking"[tiab] OR "tool segmentation"[tiab] OR "instrument segmentation"[tiab] OR "skill assessment"[tiab] OR "workflow recognition"[tiab] OR "phase recognition"[tiab] OR "surgical phase"[tiab] OR "quality monitoring"[tiab] OR "motion analysis"[tiab] OR "scope tracking"[tiab] OR "depth estimation"[tiab] OR "visual odometry"[tiab] OR "robotic endoscop*" [tiab] OR "autonomous colonoscop*" [tiab] OR "loop detection"[tiab] OR "pose estimation"[tiab]): 186900

#4 #1 AND #2 AND #3: 677

#5 #4; Filters: Date – From 2010/01/01 to 2026/01/31

Results: 574 records

EMBASE (Ovid)

Interface: Ovid (Wolters Kluwer)

Search date: January 31, 2026

#1 exp artificial intelligence/ OR exp machine learning/ OR exp deep learning/ OR exp neural network/ OR exp convolutional neural network/ OR exp reinforcement learning/ OR exp computer vision/ OR (artificial intelligence OR machine learning OR deep learning OR convolutional neural network* OR recurrent neural network* OR reinforcement learning OR computer vision OR transfer learning OR generative adversarial OR transformer OR U-Net OR YOLO OR neural radiance field* OR NeRF OR Gaussian splatting).ti,ab,kw. : 944741

#2 exp endoscopy/ OR exp colonoscopy/ OR exp esophagogastroduodenoscopy/ OR exp capsule endoscopy/ OR exp endoscopic retrograde cholangiopancreatography/ OR (colonoscop* OR gastroscop* OR endoscop* OR esophagogastroduodenoscop* OR EGD OR endoscopic submucosal dissection OR ESD OR ERCP OR capsule endoscop* OR enteroscop*).ti,ab,kw. : 1195251

#3 (kinematic* OR withdrawal time OR withdrawal speed OR blind spot* OR mucosal coverage OR coverage mapping OR navigation OR autonomous OR SLAM OR simultaneous localization and mapping OR 3D reconstruction OR instrument tracking OR tool segmentation OR instrument segmentation OR skill assessment OR

workflow recognition OR phase recognition OR surgical phase OR quality monitoring OR motion analysis OR scope tracking OR depth estimation OR visual odometry OR robotic endoscop* OR autonomous colonoscop* OR loop detection OR pose estimation).ti,ab,kw. : 42684

#4 #1 AND #2 AND #3 : 519

#5 Limit #4 to yr="2010-2026": 516

Results: 516 records

Cochrane Central Register of Controlled Trials (CENTRAL)

Interface: Cochrane Library (Wiley)

Search date: January 31, 2026

#1 [mh "artificial intelligence"] OR [mh "machine learning"] OR [mh "deep learning"] OR [mh "neural networks, computer"] OR (artificial intelligence OR machine learning OR deep learning OR convolutional neural network* OR recurrent neural network* OR reinforcement learning OR computer vision OR transfer learning OR generative adversarial OR transformer OR U-Net OR YOLO OR neural radiance field* OR NeRF OR Gaussian splatting):ti,ab,kw: 13766

#2 [mh "endoscopy, gastrointestinal"] OR [mh colonoscopy] OR [mh esophagogastroduodenoscopy] OR [mh "capsule endoscopy"] OR [mh "cholangiopancreatography, endoscopic retrograde"] OR (colonoscop* OR gastroscop* OR endoscop* OR esophagogastroduodenoscop* OR EGD OR endoscopic submucosal dissection OR ESD OR ERCP OR capsule endoscop* OR enteroscop*):ti,ab,kw: 47515

#3 (kinematic* OR withdrawal time OR withdrawal speed OR blind spot* OR mucosal coverage OR coverage mapping OR navigation OR autonomous OR SLAM OR simultaneous localization and mapping OR 3D reconstruction OR instrument tracking OR tool segmentation OR instrument segmentation OR skill assessment OR workflow recognition OR phase recognition OR surgical phase OR quality monitoring OR motion analysis OR scope tracking OR depth estimation OR visual odometry OR robotic endoscop* OR autonomous colonoscop* OR loop detection OR pose

estimation):ti,ab,kw: 69521

#4 #1 AND #2 AND #3: 220

#5 #4 with Cochrane Library publication date from Jan 2010 to Jan 2026, in Trials: 203

Results: 203 records

IEEE Xplore

Interface: IEEE Xplore Digital Library (IEEE)

Search date: January 31, 2026

#1 ("All Metadata":"artificial intelligence" OR "All Metadata":"machine learning" OR "All Metadata":"deep learning" OR "All Metadata":"convolutional neural network" OR "All Metadata":"recurrent neural network" OR "All Metadata":"reinforcement learning" OR "All Metadata":"computer vision" OR "All Metadata":"transfer learning" OR "All Metadata":"generative adversarial" OR "All Metadata":"transformer" OR "All Metadata":"U-Net" OR "All Metadata":"YOLO" OR "All Metadata":"neural radiance field" OR "All Metadata":"NeRF" OR "All Metadata":"Gaussian splatting"): 1553859

#2 ("All Metadata":"endoscopy" OR "All Metadata":"colonoscopy" OR "All Metadata":"gastroscopy" OR "All Metadata":"esophagogastroduodenoscopy" OR "All Metadata":"capsule endoscopy" OR "All Metadata":"endoscopic submucosal dissection" OR "All Metadata":"ESD" OR "All Metadata":"ERCP"): 14388

#3 ("All Metadata":"kinematics" OR "All Metadata":"withdrawal time" OR "All Metadata":"blind spot" OR "All Metadata":"coverage" OR "All Metadata":"navigation" OR "All Metadata":"autonomous" OR "All Metadata":"SLAM" OR "All Metadata":"3D reconstruction" OR "All Metadata":"instrument tracking" OR "All Metadata":"tool segmentation" OR "All Metadata":"skill assessment" OR "All Metadata":"workflow recognition" OR "All Metadata":"phase recognition" OR "All Metadata":"quality monitoring" OR "All Metadata":"motion analysis" OR "All Metadata":"depth estimation" OR "All Metadata":"visual odometry" OR "All Metadata":"robotic endoscopy" OR "All Metadata":"pose estimation"): 531759

#4 #1 AND #2 AND #3: 252

#5 #4; Filters: Year Range: 2010–2026: 231

Results: 231 records

Supplementary sources

Reference list searching: Reference lists of all included studies and relevant systematic reviews were manually screened for additional eligible studies.

Forward citation tracking: Google Scholar was used for forward citation tracking of key included studies.

Preprint servers: arXiv (cs.CV, cs.RO, eess.IV) and medRxiv were searched using combinations of key terms (endoscopy, colonoscopy, SLAM, navigation, kinematics, workflow recognition, deep learning).

Conference proceedings: Proceedings of MICCAI (Medical Image Computing and Computer Assisted Intervention), IROS (IEEE/RSJ International Conference on Intelligent Robots and Systems), and ICRA (IEEE International Conference on Robotics and Automation) from 2010 to 2025 were hand-searched for relevant studies.

Notes

The search strategy was developed by the review author (C.S.B.) with reference to existing systematic reviews of AI in gastrointestinal endoscopy. MeSH terms (PubMed, Cochrane) and Emtree terms (Embase) were exploded to capture all relevant narrower terms. Free-text synonyms were included to maximize sensitivity. The kinematic concept block (Block 3) was intentionally broad to capture the diverse terminology used across gastroenterology, computer science, and robotics literature for endoscope motion, navigation, coverage, workflow, skill, and spatial reconstruction applications.

Deduplication was performed using Endnote 21 (Clarivate Analytics) with a two-step approach: automated detection followed by manual verification of remaining potential duplicates.