Review | Open Access

Artificial Intelligence Surgery

Niu *et al. Art Int Surg.* 2026;6:441-62 **DOI:**10.20517/ais.2026.48

Clinician-supervised multimodal AI orchestration in spine care: evidence, framework, and future directions

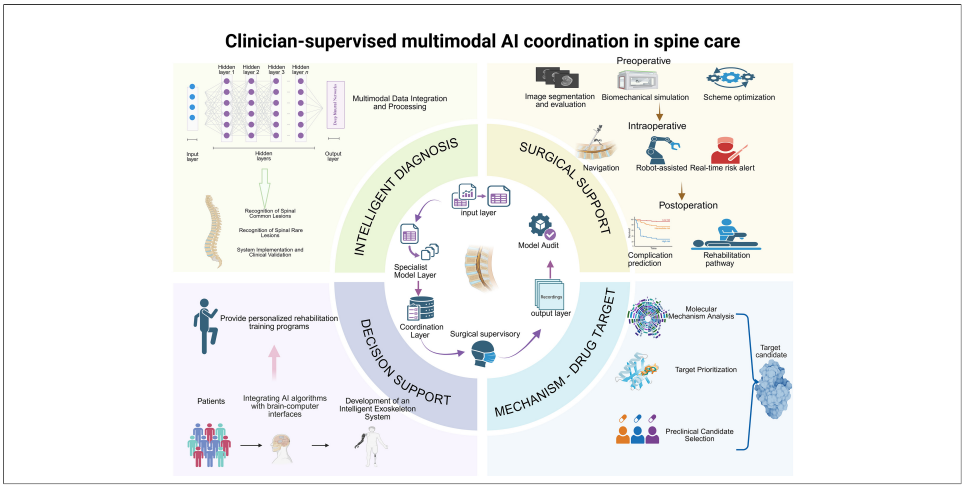
Jiayao Niu[1], Zheng Li[1], Yunpeng Zou[2], Zhipeng Wang[2], Mingjie Sun[3], Bin Ning[1,#], Wenzhao Wang[4,#], Ronghan Liu[1,5,#]

Keywords:
Artificial intelligence, spine care, AI orchestration systems, multimodal integration, clinical decision support, rehabilitation

Citation: Niu J, Li Z, Zou Y, Wang Z, Sun M, Ning B, Wang W, Liu R. Clinician-supervised multimodal AI orchestration in spine care: evidence, framework, and future directions. *Art Int Surg.* 2026;6:441-62. https://dx.doi.org/10.20517/ais.2026.48

Received: 28 May 2026
First Decision: 8 Jul 2026
Revised: 23 Aug 2026
Accepted: 9 Sep 2026
Published: 18 Sep 2026

Academic Editor:
Peter Passias
Copy Editor:
Tong Wang
Production Editor:
Tong Wang



Abstract

Degenerative, traumatic, deformity-related, and neoplastic spinal disorders place a growing burden on patients and health systems. Artificial intelligence (AI) has shown value in selected spine-care tasks, including imaging analysis, surgical-planning support, navigation assistance, risk prediction, rehabilitation monitoring, and early translational research. Most applications, however, remain task-specific tools rather than integrated clinical systems. This review therefore follows the routine spine-care pathway and focuses on clinician-supervised multimodal AI orchestration systems. In such systems, a DeepSeek-style large language model would be only one component. The broader clinical orchestration system would also require data governance, validated specialist modules, retrieval, uncertainty

[1]Central Hospital Affiliated to Shandong First Medical University, Shandong First Medical University & Shandong Academy of Medical Sciences, Jinan 250013, Shandong, China.
[2]School of Clinical Medicine, Shandong Second Medical University, Weifang 261053, Shandong, China.
[3]Center for Joint Surgery, Department of Orthopedic Surgery, The Second Affiliated Hospital of Chongqing Medical University, Chongqing 400000, China.
[4]School of Biomedical Sciences, The Chinese University of Hong Kong, Hong Kong 999077, China.
[5]Bone Biomechanics and Metabolism Laboratory, Central Hospital Affiliated to Shandong First Medical University, Jinan 250013, Shandong, China.

Correspondence to: Dr. Bin Ning, Dr. Ronghan Liu, Central Hospital Affiliated to Shandong First Medical University, Shandong First Medical University & Shandong Academy of Medical Sciences, Jinan 250013, Shandong, China. E-mail: ningbin@sdu.edu.cn; ronghanliu@email.sdfmu.edu.cn; Dr. Wenzhao Wang, School of Biomedical Sciences, The Chinese University of Hong Kong, Hong Kong 999077, China. E-mail: wenzhaowang@cuhk.edu.hk

estimation, safety filters, audit trails, and clinician oversight. We review evidence from admission and imaging assessment through preoperative planning, intraoperative support, postoperative monitoring, and rehabilitation follow-up. We distinguish direct spine-specific evidence from indirect technical analogies and future hypotheses. Multi-omics and drug-development studies are considered only as an outer-loop translational layer for mechanism generation, endotype discovery, biomarker development, and trial enrichment. Current evidence supports selected diagnostic, prognostic, rehabilitation-monitoring, and workflow-assistance tasks, but not a safe, end-to-end autonomous platform for spinal surgery. Near-term translation should prioritize external validation, prospective silent testing, calibration, evidence traceability, post-deployment surveillance, and explicit clinician control.

INTRODUCTION

Spinal disorders are among the leading causes of pain, disability, and healthcare use worldwide. They include degenerative conditions such as disc herniation and spinal stenosis, traumatic injuries such as vertebral fractures, spinal deformity, spinal tumors, and spinal cord injury (SCI). Their burden is increasing as populations age and more patients require complex, longitudinal care[1]. Recent reviews also show that spine AI remains fragmented across imaging, planning, prediction, and perioperative applications rather than integrated into a single care pathway[2,3]. Clinicians must therefore synthesize information distributed across imaging, operative records, clinical notes, functional assessments, and patient-reported outcomes, often across separate services and time points.

These limitations are the reason for the growing interest in artificial intelligence (AI) in both spine care and spinal surgery, and the scope of the exploration is now well beyond imaging only. Along the diagnostic path, in fact, deep learning tools are already used for the task of disc degeneration grading, for the detection of vertebral fractures, for the segmentation of spinal structures, and for stenosis assessment[4,5]. In perioperative care, AI is being evaluated mainly as an adjunct to surgical planning, navigation, robotic guidance, and biomechanical modeling[6]. Predictive models have also been developed for selected complications and recovery outcomes. Evidence is less mature in rehabilitation and translational research. Exoskeleton-assisted training and device-based neuromodulation have shown early promise in selected patients with SCI[7,8], whereas multi-omics studies in spinal degeneration and AI-assisted drug discovery remain largely exploratory or preclinical[9,10].

Yet most of these tools still operate in isolation. An imaging model may identify a disc herniation without considering the patient's symptoms, neurological examination, prior imaging, or the practical constraints of surgery. Navigation and robotic platforms provide technical assistance, but their outputs still require interpretation within the operative and clinical context. Biomechanical models and risk calculators are generally used to compare scenarios rather than determine treatment. Rehabilitation devices generate gait, pain, fatigue, adherence, and function data that are often reviewed separately. Molecular AI may generate mechanistic or therapeutic hypotheses but currently has little direct bearing on routine surgical care. The central question of this review is therefore whether outputs from task-specific tools can be coordinated in a transparent, uncertainty-aware, and clinician-supervised workflow.

The need to integrate information across modalities has drawn attention to foundation models as potential components of clinical decision-support systems. Here, the term DeepSeek-style model refers narrowly to a reasoning large language model (LLM) core, such as DeepSeek-V3 or DeepSeek-R1. Its potential functions include summarizing clinical text, synthesizing retrieved evidence, and formulating retrieval queries[11-13]. The model would be only one component of a deployable spine-care system. The surrounding clinical system would coordinate multimodal data and outputs from validated specialist models and would also require standardized data ingestion, multimodal encoders, evidence retrieval and traceability, uncertainty estimation,

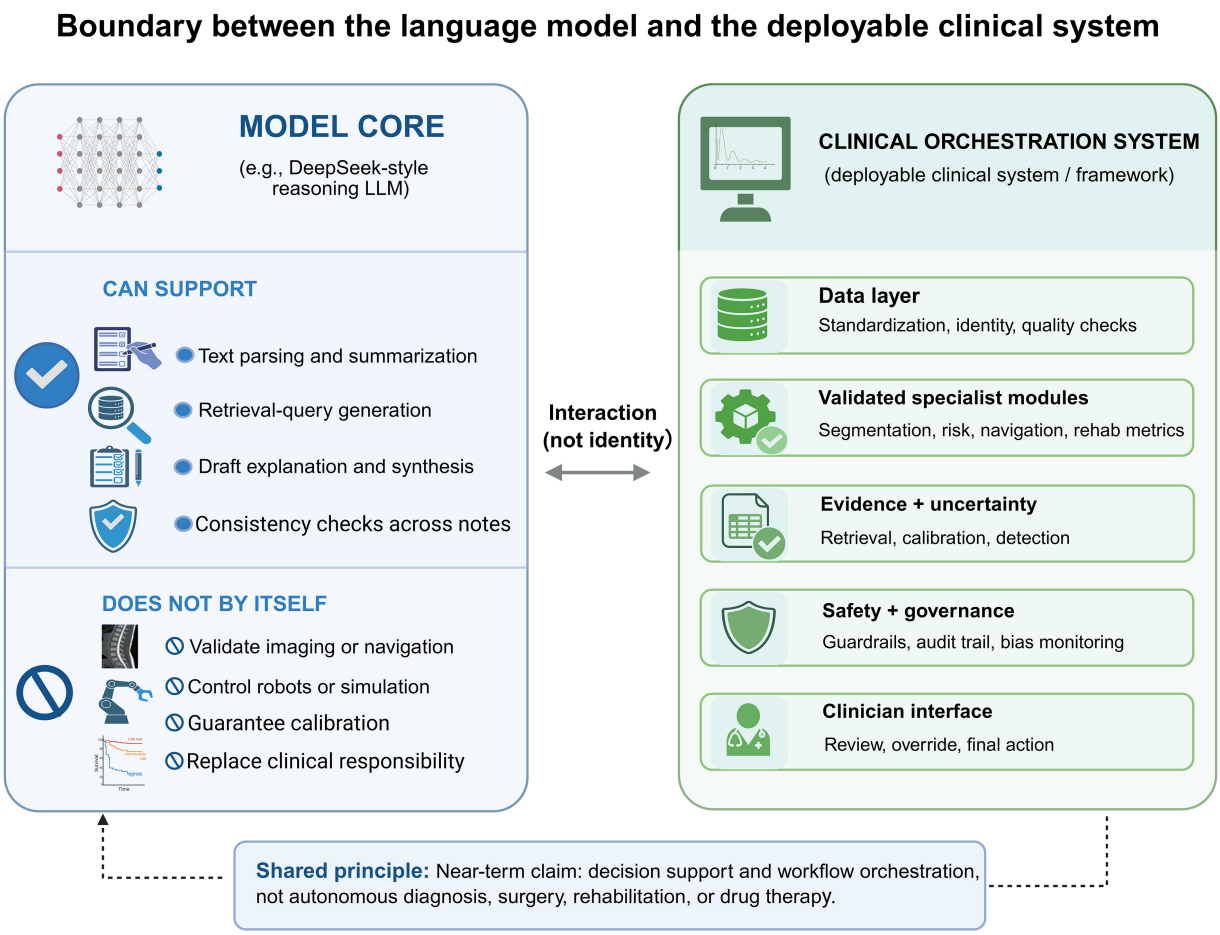


Figure 1. Boundary between the language model and the deployable clinical system. A DeepSeek-style LLM may support summarization, retrieval-query generation, and explanation drafting. A clinical system also requires data governance, validated specialist modules, uncertainty estimation, safety filters, audit trails, clinician-interface design, and post-deployment monitoring. Created in BioRender. Niu, J. (2026) https://BioRender.com/4cks0vq. LLM: Large language model.

safety checks, audit trails, post-deployment monitoring, and clinician oversight[14,15]. Figure 1 illustrates the distinction between the language-model component and the surrounding clinical system.

The role of this broader clinical orchestration system would be coordination. It would not replace specialist algorithms or clinicians. It could combine outputs from segmentation tools, risk calculators, biomechanical simulations, navigation platforms, rehabilitation metrics, and selected molecular pipelines, then present them in a traceable and clinically reviewable form. This review examines how clinician-supervised multimodal AI orchestration systems might support decision-making across the routine spine-care pathway. The discussion follows the patient journey from initial assessment and imaging through surgical planning, intraoperative support, postoperative monitoring, and rehabilitation [Figure 2]. Molecular and multi-omics research, together with AI-assisted drug discovery, are discussed separately as an outer-loop translational domain relevant to disease stratification, biomarker development, and future therapeutic studies, not as current routine-care modules.

Review methodology and evidence selection

For this narrative review, we searched PubMed/MEDLINE, Embase, Web of Science, Scopus, IEEE Xplore, and Google Scholar for articles published through April 2026. Search terms combined spine-related keywords, including “spine surgery”, “spinal disorders”, “lumbar disc herniation”, “lumbar spinal stenosis”,

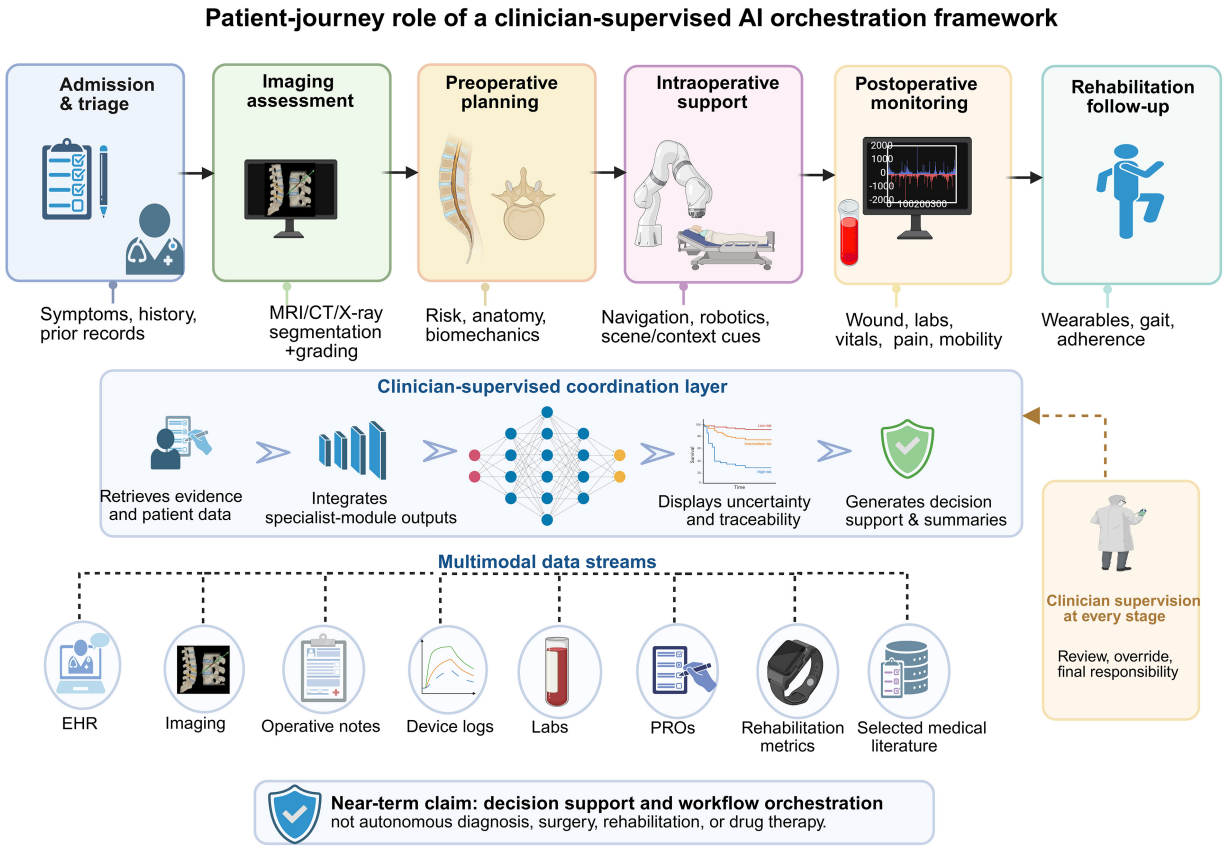


Figure 2. Patient-journey role of a clinician-supervised AI orchestration framework. The system coordinates data and specialist-module outputs from admission, imaging assessment, preoperative planning, intraoperative support, postoperative monitoring, and rehabilitation follow-up. At each stage, clinicians review outputs, resolve uncertainty, and retain final responsibility. Created in BioRender. Niu, J. (2026) https://BioRender.com/9i12e38. AI: Artificial intelligence; MRI: magnetic resonance imaging; CT: computed tomography; EHR: electronic health record; PROs: patient-reported outcomes.

"vertebral fracture", "spinal tumor", "spinal deformity", "spinal cord injury" and "spine rehabilitation" with AI-related terms, including "artificial intelligence", "machine learning", "deep learning", "foundation model", "large language model", "multimodal model", "clinical decision support", "surgical navigation", "robotic surgery", "augmented reality", "brain-computer interface" and "exoskeleton". Additional searches included translational terms such as "multi-omics", "drug-target prediction", "virtual screening", "absorption, distribution, metabolism, excretion, and toxicity (ADMET) prediction", and "organ-on-a-chip".

We included peer-reviewed studies evaluating AI applications across the spine-care continuum, including diagnosis, image segmentation and grading, surgical planning, navigation, robotics, intraoperative support, postoperative risk prediction, complication surveillance, rehabilitation, and translational research. Priority was given to clinical studies, external validation studies, multicenter evaluations, systematic reviews, meta-analyses, and methodologically relevant studies. Editorials, commentaries, conference abstracts, non-English articles, studies unrelated to spine care, and technical studies without a plausible clinical or translational connection were excluded. Isolated case reports were generally excluded, although landmark proof-of-concept studies in emerging fields were retained when more mature evidence was unavailable.

Database records and additional references identified through reference-list screening were deduplicated and assessed by title, abstract, and full text. Studies were retained when they informed the main clinical or translational themes of the Review. Direct evidence from spine-specific datasets was distinguished from

indirect evidence from other surgical, radiological, rehabilitation, or biomedical fields. Evidence from non-spine settings was used only to support technical analogy or future research directions and was not treated as proof of clinical effectiveness in spine surgery. The selected literature was organized into five categories: diagnostic and multimodal data integration; surgical planning, simulation, navigation, and intraoperative support; postoperative prognosis and rehabilitation; molecular and translational research; and safety, governance, and implementation. To improve transparency, the clinical readiness of each AI application was assessed using a qualitative framework adapted from technology readiness concepts used in healthcare AI evaluation. The assessment considered technical validation, external validation, clinical workflow integration, and evidence of patient-centered outcomes. Because most spine AI applications remain at different stages of translation and lack uniform prospective evaluation criteria, readiness categories were used as descriptive summaries of evidence maturity rather than formal regulatory classifications. The literature search and selection process is summarized in Figure 3. Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) 2020 and Preferred Reporting Items for Systematic Reviews and Meta-Analyses literature search extension (PRISMA-S) informed the reporting of the search process, although this article was conducted as a narrative review rather than a systematic review or meta-analysis. Because the search was iterative, Figure 3 is presented as a schematic of the identification and screening process and does not provide numerical counts for each intermediate stage. A total of 147 references were cited in the final narrative synthesis.

ROUTINE DIAGNOSIS AND LONGITUDINAL PHENOTYPING

Multimodal data integration and processing

Direct spine-specific evidence supports several narrowly defined imaging tasks, but integrated multimodal diagnosis remains less mature. Large multimodal models have been tested for lumbar foraminal stenosis, text-guided methods have linked magnetic resonance imaging (MRI) segmentation with abnormality identification, and external validation studies have assessed the generalizability of automated degeneration grading[16-18]. Overall, automated spine-image analysis currently performs best in anatomical segmentation and the grading of specific abnormalities. These studies represent an initial move beyond isolated lesion detection, but they do not establish reliable integration of imaging outputs with structured reports and the broader clinical record. This distinction is clinically important because a useful lumbar imaging assessment requires not only lesion detection, but also accurate localization and characterization and explicit consideration of whether the imaging findings correspond with the patient's symptoms and clinical history[19]. Broader experience in radiology further suggests that successful clinical translation depends on close collaboration among imaging specialists, clinicians, technical teams, and other implementation stakeholders[20].

Reliability also depends on data quality and compatibility. External validations of SpineNet in independent lumbar MRI cohorts show why performance should be assessed beyond the development dataset[21,22]. Variation in imaging acquisition and dataset characteristics, together with missing sequences and examinations obtained at different points in the clinical course, can complicate registration, segmentation, longitudinal comparison, and downstream analysis. Deformable image-registration methods provide one technical approach to image alignment[23]. Evidence from outside spine illustrates other possible strategies: domain adaptation has been used to address cross-device variation in thyroid ultrasound[24], while multi-sequence MRI studies in brain and abdominal imaging show how complementary sequences can be integrated for segmentation[25,26]. The applicability of these approaches to spine imaging remains to be established. General-purpose frameworks such as nnU-Net and an MRI study of dorsal root ganglion segmentation demonstrate the feasibility of automatically delineating selected anatomical structures[27,28]. However, performance may not generalize across anatomical targets, patient populations, or imaging protocols. The preprocessing, registration, and segmentation pipeline should therefore be validated in the intended clinical setting before its outputs are integrated with other clinical information.

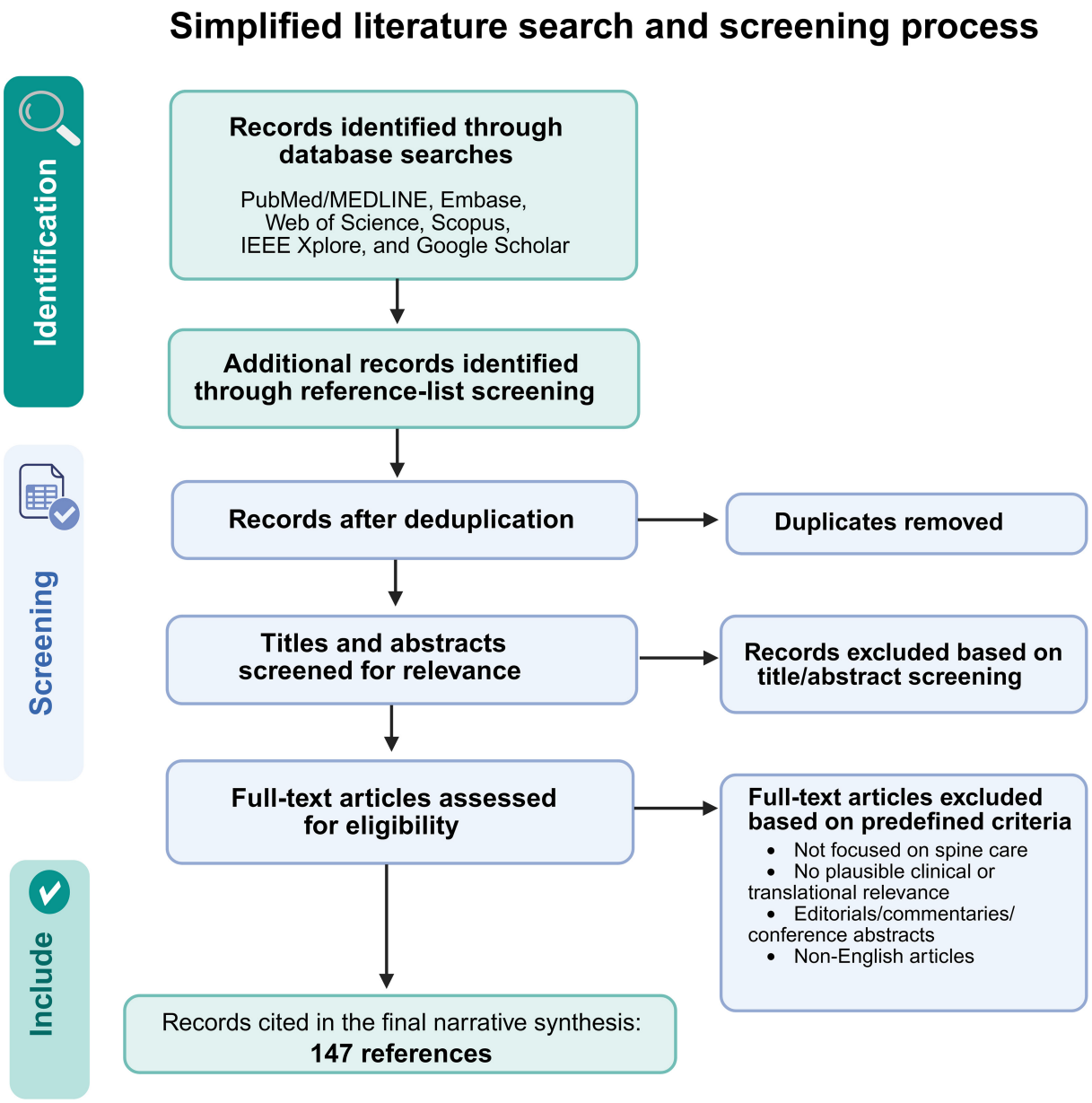


Figure 3. Simplified literature search and screening process. The flowchart summarizes the databases searched, reference-list screening, title and abstract screening, full-text eligibility assessment, priority inclusion criteria, and evidence-organization categories used in the revised manuscript. Created in BioRender. Niu, J. (2026) https://BioRender.com/42w4qf4.

Clinical records add another layer of complexity. The significance of an imaging finding depends on its spinal level, timing, disease course, and previous treatment, and the same finding may be described differently across reports and notes. Knowledge-graph-enhanced clinical models illustrate how structured concepts can be linked to electronic health record data[29]. Explainability work outside spine likewise illustrates the need to link model outputs to source features[30]. A language model could help organize radiology reports, progress notes, and earlier examinations, and retrieve the text supporting a summary. but such use requires task-specific testing for calibration and for unsupported or clinically inconsistent outputs[31,32]. Multimodal methods may then relate imaging findings to clinical text and structured patient

data. Some architectures perform this in stages, first aligning local imaging features with the relevant report text and then combining this evidence with structured clinical information or question-guided representations[33,34]. Within a clinician-supervised system, a foundation-model component could relate the outputs of specialist tools to the clinical record, standardize summaries, and flag uncertain or conflicting evidence for review. Lesion detection would still rely on dedicated imaging models, while interpretation and clinical decisions would remain with the clinical team.

Diagnosis of common spinal disorders

Among common spinal disorders, lumbar disc disease has received particular attention in AI-assisted imaging. For this condition, clinically useful analysis requires more than labeling a disc as normal or abnormal. The involved level must be identified correctly, and the extent and pattern of degeneration or herniation must be described, including any effect on the dural sac or adjacent neural structures[35]. Recent pipelines combine several of these tasks, including disc localization or segmentation with grading and abnormality classification[36-39]. Graph-based and boundary-constrained methods can also model the anatomical relationships between vertebrae and intervertebral discs to improve segmentation consistency[40]. Automated segmentation and measurement of the dural sac may provide additional quantitative information about canal narrowing[41]. Together, these methods provide a more detailed description of lumbar disc abnormalities than a single classification label.

Vertebral fracture and lumbar spinal stenosis pose different imaging tasks. Mild osteoporotic fractures can be difficult to distinguish from chronic or degenerative vertebral deformity, particularly when vertebral height loss is limited[42]. Recent fracture-detection pipelines commonly use a staged design in which vertebrae are first localized or segmented and then classified at the vertebral level[43,44]. Lumbar stenosis is less a problem of detecting a discrete lesion than of grading anatomical narrowing consistently. Deep-learning models have been developed to grade central canal stenosis across multiple lumbar levels, while segmentation-based methods can quantify dural sac narrowing[45]. Comparison across studies remains difficult because the anatomical targets, grading systems, datasets, and reference standards differ substantially[46].

Rare spinal disorders and longitudinal phenotyping

Rare spinal conditions are difficult to assess not simply because they are uncommon, but because their significance often emerges only when imaging is interpreted within a broader clinical history. Spinal tumors, congenital malformations, and syndromic deformities are heterogeneous, and relevant information may be distributed across institutions and different stages of follow-up. In syndromic deformity, systemic disease can directly alter surveillance and operative risk: *FBN1*-related disorders require attention to aortic disease and prior cardiovascular treatment[47,48]. This heterogeneity is not limited to spinal morphology; genetic background and systemic comorbidities can also change the clinical significance of the same deformity. *FBN1* variants have been linked to both Marfan syndrome and nonsyndromic scoliosis[49]. In Turner syndrome, cardiovascular risk may directly affect the timing and planning of deformity surgery[50]. A single spinal examination therefore provides only part of the information needed for multidisciplinary review.

Spinal tumors are a reasonable setting in which to evaluate imaging-based AI because radiology already plays a central role in lesion detection, localization, and treatment planning. Direct evidence from spine-specific studies, however, remains limited. AI methods have been used to localize poorly visualized tumors in radiotherapy settings, but their reliability in spinal oncology has not yet been established[51]. In metastatic spinal disease, vertebral destruction, marrow replacement, and epidural or paraspinal extension help define the extent of disease and narrow the differential diagnosis. The pattern of involvement may sometimes suggest a particular tumor type or primary site, but imaging alone is rarely conclusive. These findings need to be interpreted alongside the clinical history, systemic staging studies, and, when indicated, histopathology[52].

A multimodal system could make multidisciplinary review easier by bringing this information together and drawing attention to missing or conflicting evidence[53].

Congenital and syndromic spinal deformities are typically evaluated longitudinally because growth can alter curve severity, spinal balance, symptoms, and treatment risks. In children who undergo repeated imaging, three-dimensional ultrasound with automated landmark detection has been investigated as a means of measuring spinal morphology without repeated exposure to ionizing radiation[54]. These measurements still require interpretation in relation to vertebral development and changes in curve severity over time[55,56]. Studies in adolescent idiopathic scoliosis have used radiographic and clinical data to predict curve progression[57,58], providing a spine-specific precedent for longitudinal modeling. Time-dependent transformer models developed in oncology offer a broader methodological example for analyzing complex changes over time[59]. But their applicability to congenital or syndromic spinal deformity remains unproven. Such approaches could help compare serial findings and flag unexpected patterns of progression, although this role has not been established in routine care. For the foreseeable future, the main contribution of multimodal AI in rare spinal disorders is likely to be the organization of longitudinal and cross-specialty information for clinical review.

TREATMENT PLANNING AND SURGICAL SUPPORT WITHIN THE ROUTINE CARE PATHWAY

Biomechanical simulation and surgical plan optimization

Accurate preoperative assessment and procedurespecific planning remain central to spine surgery. Many of the decisions, however, still depend on the experience of the surgeon - in particular for anatomically complex or uncommon presentations. Patient-specific finite-element analysis can make biomechanical assumptions explicit, but conventional workflows depend on labor-intensive segmentation, meshing, and parameter assignment. Automated lumbar-spine pipelines now link deep-learning segmentation with finite-element modeling to reduce this burden[60,61]. Deep-learning methods can also segment vertebrae, intervertebral discs, and the spinal canal from lumbar MRI[62], while statistical shape models and related geometric methods support patient-specific mesh generation[63]. Estimates of bone quality derived from Hounsfield units, particularly when combined with patient-level factors, may make biomechanical simulations more representative of the individual patient[64]. These advances increase efficiency, but they do not by themselves establish that a simulated plan improves patient outcomes.

The clinical value of these tools lies less in producing a single "optimal" plan than in comparing reasonable surgical options under explicit assumptions. Risk calculators, registry-based models, and biomechanical simulations can help surgeons evaluate fixation levels, screw trajectories, implant choice, correction targets, and mechanical risk[65-68]. Multiscale modeling may add local information about screw size, angulation, fusion level, and bone-implant stress[69,70]. Orthopedic digital twins can be viewed as an emerging extension of patient-specific simulation, linking multiple data streams to support surgical simulation and prognostic modeling; however, the current evidence remains early and heterogeneous[71]. In this setting, multimodal coordination could present anatomy, simulated options, model assumptions, estimated risks, and uncertainty in one reviewable interface. It should not select the procedure, determine fusion levels, or replace surgeon judgment.

Simulation, navigation, and intraoperative support

Surgical simulation allows surgeons to examine patient-specific anatomy and rehearse difficult procedural steps before entering the operating room. Virtual reality, augmented reality, and mixed reality support different parts of this process: virtual reality provides a fully simulated environment, augmented reality overlays planned trajectories or structures on the operative view, and mixed reality anchors interactive three-dimensional models in the user's physical space[72-75]. Some platforms also incorporate haptic interfaces and

deformable or synthetic tissue models that reproduce selected aspects of bone and soft-tissue behavior[76,77]. In spine training, haptic systems have been used for pedicle screw placement and drilling, including models designed to distinguish cortical from cancellous bone[78]. Multilayered simulators can also reproduce selected aspects of soft-tissue manipulation, pressure on neural structures, and intraoperative bleeding[77,78]. Most published evidence for these systems comes from education and preoperative rehearsal; intraoperative applications have developed mainly through separate augmented-reality and image-guidance platforms[79,80].

In the operating room, navigation and robotic systems represent a major pathway through which AI-assisted technologies are being incorporated into spine surgery. Navigation platforms integrate preoperative computed tomography (CT)/MRI data, intraoperative imaging, and registration algorithms to provide real-time localization of instruments relative to patient anatomy. Robotic systems extend this workflow by translating a preoperatively defined trajectory into controlled instrument guidance. The strongest clinical evidence still concerns pedicle screw placement: compared with conventional fluoroscopic or freehand techniques, robot-assisted navigation has demonstrated improved screw-placement accuracy in comparative studies and meta-analyses, although the magnitude of benefit varies among platforms, procedures, and outcome definitions[81,82]. AI is increasingly embedded within these platforms. Machine learning enhances image segmentation, registration, and anatomical recognition from CT and fluoroscopy, enabling more accurate and efficient navigation[83-85]. In parallel, deep learning-based path planning and reinforcement learning frameworks are being developed to enable real-time trajectory optimization, allowing robotic systems to adjust guidance as anatomy or positioning changes during surgery[86-88].

Continued advances in navigation robotics are increasingly centered on the integration of multimodal data streams. AI-enabled planning systems may combine anatomical segmentation, deformity measurements, bone quality assessment, implant selection, and biomechanical simulation before surgery, then transfer these outputs into intraoperative navigation and robotic execution[89,90]. To translate these capabilities into clinical practice, emerging platforms incorporating machine vision, 3D reconstruction, automated registration, and real-time workflow monitoring aim to reduce dependence on manual interpretation and improve procedural consistency[91]. Key limitations persist: registration error, variability across institutions, equipment costs, learning curves, and insufficient prospective validation of patient-centered outcomes. Addressing these challenges, a future clinical orchestration system could coordinate planning algorithms, navigation data, robotic status, and uncertainty information in a unified interface while preserving surgeon control over every critical decision.

Postoperative risk, surveillance, and recovery

Postoperative risk is shaped by interacting baseline, procedural, and time-varying factors. Most published models have also been developed retrospectively, so their performance may decline in hospitals with different patient populations or postoperative care pathways. More refined outcome prediction could help clinicians explain risk more clearly, identify patients who may need closer surveillance, and adjust recovery plans before complications or delayed recovery become apparent. Multimodal AI may be useful in this setting because postoperative risk is usually shaped by several interacting factors rather than by a single measurement. For example, appropriately designed survival models can integrate continuous variables such as Cobb angle, categorical or ordinal factors such as the American Society of Anesthesiologists physical status classification (ASA) grade, and time-dependent intraoperative or postoperative physiological signals within a shared prognostic framework[92]. As new clinical information becomes available, these models could update the estimated risk rather than rely on a single assessment made soon after surgery. External validation and calibration would still be required before such estimates could guide care[93]. A recent comparative study in percutaneous kyphoplasty found that DeepSeek R1 performed similarly to conventional machine-learning models and modestly outperformed surgeon-only judgment in predicting bone-cement leakage, but it was

less reliable in predicting subsequent vertebral fracture[94]. Performance on one postoperative outcome therefore provides little assurance that the same model will perform well on another. Each outcome should be evaluated separately in the population and clinical setting in which the model is intended to be used.

Prediction and surveillance answer different questions: prediction estimates who may develop a complication, whereas surveillance asks whether current recovery is departing from the expected course. Continuous sensors and signal classifiers can detect physiological events[95,96]. However, their clinical value depends on prospectively validated thresholds, false-positive rates, and actionability. Multimodal models may place postoperative changes in context and distinguish expected fluctuations from early wound, thromboembolic, cardiopulmonary, infectious, or delayed-recovery signals[97]. Rehabilitation-specific reviews suggest that longitudinal summaries of strength, balance, pain, adherence, and function may support treatment review, while emphasizing inconsistent clinical effects and the need for real-world validation[98-100]. Such summaries may prompt reassessment, but investigation and treatment must remain grounded in the patient's examination and broader clinical context.

OUTER-LOOP TRANSLATIONAL EVIDENCE: FROM MECHANISM DISCOVERY TO PRECLINICAL VALIDATION

Molecular network analysis in spinal disorders

Multi-omics and computational molecular analyses belong to an outer-loop translational layer rather than the routine spine-care pathway. Their near-term purpose is to generate testable mechanisms, candidate biomarkers, and provisional endotypes for laboratory validation and future trial enrichment. Spine-specific studies have identified candidate pathways and network modules in intervertebral disc degeneration and adult degenerative scoliosis[9,101]. Mechanistic work on circRNA CDR1as after SCI goes a step further by linking a defined molecular signal to mothers against decapentaplegic homolog (SMAD)-related fibrosis in a disease-relevant model[102]. Broader biomedical and computational studies provide methodological precedents for integrating heterogeneous molecular data and modeling cellular, temporal, and subgroup variation[103-111], but they do not establish causality or clinical utility in spine care. Molecular findings should therefore remain outside routine diagnosis and surgical decision-making until they are confirmed in relevant spinal tissue, supported by perturbation experiments, and prospectively associated with meaningful phenotypes or outcomes. Within the proposed system, a foundation-model component could organize and trace this evidence, but it should not convert an association into a mechanism.

Target prioritization and preclinical candidate selection

Once a mechanism has adequate biological support, computational methods can help prioritize targets and candidates for experimental testing. Machine-learning-guided screening, generative design, and ADMET prediction can narrow chemical space and remove candidates with obvious liabilities[10,112-116], but favorable rankings are not evidence of therapeutic effect. Candidates should advance only after demonstrating target engagement, activity in disease-relevant spinal tissue, acceptable exposure and toxicity, and reproducible benefit in appropriate models. Work on osteogenic pathways illustrates the type of mechanistic confirmation required[117]. Organ-on-chip and perfused microenvironment systems may provide intermediate test platforms, although current examples remain general biomedical models rather than validated platforms for disc disease or spinal tumors[118-120]. The output of this outer-loop layer should therefore be a transparent, experimentally supported target, candidate, biomarker, or trial-enrichment hypothesis - not a treatment recommendation. A foundation-model component may support evidence retrieval, cross-study comparison, and auditability, while decisions to advance a candidate remain with the translational research team.

POSTOPERATIVE REHABILITATION AND LONGITUDINAL FOLLOW-UP

Brain-computer interfaces for motor intention decoding

Brain-computer interfaces (BCIs) address the loss of reliable communication between motor intention and movement after SCI. In a single-participant proof-of-concept study, a brain-spine interface decoded cortical activity and linked it to spinal stimulation to support voluntary standing and walking[121]. Because this evidence comes from a single highly selected participant receiving intensive experimental support, it demonstrates technical feasibility rather than an approach ready for routine rehabilitation. Even so, the findings suggest that intention-guided neuromodulation may restore selected components of functional mobility under closely supervised conditions. Broader clinical use will require robust decoding across sessions, safer and simpler hardware, standardized outcome measures, and evidence from larger studies. One important direction is to make the link between neural decoding and stimulation more adaptive. Rather than relying on fixed stimulation patterns, future systems may need to respond to changes in motor intention, fatigue, posture, task demands, residual motor output, and signal quality[122]. AI methods may improve signal classification, reduce the need for recalibration, and help track signal quality across repeated rehabilitation sessions. These real-time functions would be handled by task-specific decoders and device controllers rather than by an LLM.

The role of LLMs and related foundation models therefore should be positioned outside the real-time control loop. General BCI literature and an exploratory ChatGPT-BCI/virtual-reality proposal suggest possible higher-level roles in documentation, education, or information integration[123,124], but neither validates an LLM orchestration layer in SCI rehabilitation. Reviews of steady-state visual evoked potential (SSVEP) and lower-limb motor-imagery systems instead identify patient-specific signal processing, noise, and repeated recalibration as core technical challenges[125,126]. A clinician-supervised foundation-model layer would sit outside the real-time control loop, allowing device performance to be reviewed alongside pain, fatigue, and functional progress. This combined view may help the rehabilitation team notice when improved walking is accompanied by worsening pain or when declining signal quality precedes poorer performance. Translating BCIs into everyday clinical practice remains constrained by signal instability, repeated recalibration, latency, and the maintenance demands of implanted hardware. Gains in decoding accuracy or stimulation precision become clinically meaningful only when they reduce the burden on patients and rehabilitation teams and produce safer, more durable improvements in daily function. This proposed use of foundation models has not yet been validated in routine SCI rehabilitation. Future studies should therefore assess safety, usability, maintenance requirements, and the durability of functional benefits in real rehabilitation settings. Decisions about stimulation settings and rehabilitation progression would remain with the clinical team.

Exoskeletons and personalized rehabilitation management

Robotic exoskeletons allow people with SCI to practice repeated overground walking with adjustable mechanical support. Therapists can vary assistance according to residual motor function, balance, fatigue, and safety needs[127]. Current evidence suggests that exoskeleton-assisted training may improve walking capacity and selected functional outcomes in some patients with SCI, especially incomplete SCI. The size and durability of benefit vary[7,128]. Translation to daily life is less certain. Use may be limited by donning and doffing time, fatigue, discomfort, environmental barriers, device access, and the need for trained assistance[127].

Exoskeleton sessions generate clinically relevant data. These include kinematics, assistance level, walking distance, heart-rate response, pain, fatigue, adherence, the Walking Index for Spinal Cord Injury II (WISCI II), the 10-meter walk test, the 6-min walk test, and the Spinal Cord Independence Measure III (SCIM III)[129,130]. A foundation-model system could summarize these data across sessions. It could help clinicians identify improvement, plateau, compensation, declining tolerance, or mismatch between clinic performance

and daily function. Future trials should measure home and community use, caregiver burden, safety events, sustained adherence, and patient-reported value. Treatment decisions should remain with the rehabilitation team, consistent with expert guidance on exoskeleton prescription and therapist-in-the-loop control[131,132].

CLINICAL TRANSLATION, GOVERNANCE, AND RESPONSIBLE IMPLEMENTATION

Evaluation for clinical translation

When moving an AI tool into spine care, the starting question should be whether it improves practice, not whether its architecture is novel. A model is worth using only if, compared with current care, it improves the accuracy or consistency of clinical decisions, shortens review time, reduces risk, or leads to better patient outcomes. Evaluation should proceed from retrospective development and internal testing to external testing across institutions with different patient populations, scanners, imaging protocols, surgical practices, and rehabilitation pathways. Prospective silent deployment can then assess performance, calibration, failure modes, and workflow fit without allowing the system to influence care. Early live evaluation and, where justified, interventional trials are needed before outputs are used to guide clinical decisions[15,133]. Reporting and appraisal should match the study design: prediction-model studies should follow Transparent Reporting of a multivariable prediction model for Individual Prognosis Or Diagnosis (TRIPOD)+AI and be assessed with Prediction model Risk Of Bias Assessment Tool (PROBAST)+AI; imaging studies should follow Checklist for Artificial Intelligence in Medical Imaging (CLAIM); early live evaluations of decision-support systems should follow Developmental and Exploratory Clinical Investigations of Decision support systems driven by Artificial Intelligence (DECIDE-AI); and AI clinical trial protocols and reports should follow Standard Protocol Items: Recommendations for Interventional Trials (SPIRIT)-AI and Consolidated Standards of Reporting Trials (CONSORT)-AI, respectively[134-138].

For multimodal AI, integration itself should be treated as a testable claim. A system combining imaging, clinical records, surgical variables, rehabilitation measures, or wearable data should be compared with strong single-modality models, validated specialist tools, and existing clinician-led workflows[139]. Component analyses should determine whether retrieval, specialist modules, uncertainty estimation, and rule-based safety checks add measurable value or merely increase complexity. Studies should report calibration, performance across clinically relevant subgroups, failure patterns, safety during silent deployment, and effects on workload and decision-making. The same scrutiny applies to musculoskeletal digital twins: current work remains centered largely on patient-specific modeling and simulation, with limited evidence that these systems improve routine clinical care[140]. Table 1 summarizes the current evidence and clinical readiness of AI applications across the spine-care pathway.

Data governance, equity, and accountability

Responsible AI use depends on how data are collected, protected, and reused[141]. Spine care generates sensitive longitudinal information, including imaging, surgical records, rehabilitation outcomes, patient-reported measures, and data from wearable or home-monitoring devices. These data may be reused for model development, external validation, updating, and post-deployment monitoring, making clear arrangements for consent, access control, cybersecurity, retention, and ownership essential. Federated learning may reduce the transfer of raw patient data between centers, but it does not remove the risks of re-identification, information leakage, security breaches, or uneven data quality[143]. Patients should receive a practical explanation of how their data will be used, who may access them, and whether the data may be retained for future model updates or new applications[144].

Average performance can conceal important failures. A model may perform well overall yet remain unreliable in rare spinal disorders, underrepresented populations, patients with atypical anatomy or complex comorbidities, and hospitals with limited resources[145,146]. Subgroup evaluation should therefore be

Table 1. Evidence and readiness for clinician-supervised multimodal AI orchestration systems in spine care

Domain/Patient-journey use case	Evidence category	Current status/readiness	Main limitation and claim boundary
Data ingestion and multimodal coordination[14,20,139,141]	Conceptual system layer; indirect evidence from clinical AI frameworks	Low (conceptual framework; implementation and validation pending)	Requires governance, retrieval, calibration, out-of-distribution detection, audit, and clinician-interface validation; not a model-only function
Image segmentation and anatomical labeling[4,27,28,85]	Direct spine-specific task evidence	Moderate (technical performance demonstrated)	Mostly retrospective; scanner and protocol variability; limited prospective workflow testing
Disc degeneration, disc herniation, stenosis, and fracture assessment[21,22,43,45]	Direct spine-specific task evidence	Moderate (disease- and task-specific validation exist)	Disease- and task-specific; limited outcome impact and multimodal clinical correlation
Rare disorders, tumors, and deformity risk[47,55,58]	Limited direct evidence plus conceptual longitudinal modeling	Low (early feasibility; limited by small datasets and heterogeneous phenotypes)	Small datasets, heterogeneous phenotypes, and insufficient validated perioperative risk models
Preoperative planning and biomechanical simulation[61,63,67,68]	Direct technical evidence with limited clinical outcome evidence	Low-to-moderate (simulation feasible; outcome evidence and automated integration needed)	Useful for scenario comparison and patient-specific simulation; not validated as autonomous planning; large-scale outcome evidence remains sparse
Navigation, robotics, AR/VR, and simulation[73,81,83]	Assistive-technology evidence; AI coordination mostly indirect	Moderate (established assistive tools; patient-centered outcome evidence remains limited)	Must remain under surgeon command; feasibility studies should not be described as routine autonomous use
Intraoperative decision support[3,86,142]	Indirect and early-stage evidence	Low (early-stage; safety and real-time robustness unproven)	High-risk environment; real-time validation, latency, uncertainty handling, and failure modes remain unresolved
Postoperative risk prediction[92,94,97]	Direct spine-specific retrospective evidence; limited prospective evidence	Low to moderate (promising models; requires calibration and prospective workflow evaluation)	Calibration, subgroup performance, alert thresholds, and workflow impact require prospective testing
Complication surveillance and recovery benchmarking[65,98,99]	Mixed direct and indirect evidence	Low (exploratory; clinical utility and alert thresholds unvalidated)	False alarms and workload burden may undermine clinical value without actionability
Exoskeleton-assisted rehabilitation[7,127,128]	Early clinical evidence and systematic reviews in selected SCI populations	Low (clinical feasibility shown; real-world and home-use evidence limited)	Benefits vary; adherence, caregiver burden, pain, fatigue, and real-world transfer remain key
BCI and neuromodulation rehabilitation[121,122,124]	Proof-of-concept clinical evidence	Low (proof-of-concept; scalability and safety unresolved)	Small studies; invasive hardware, signal stability, recalibration, safety, and scalability remain barriers
Molecular mechanisms and multi-omics[9,101,102]	Translational/Preclinical evidence	Low (hypothesis-generating; experimental validation required)	Useful for mechanism and subtype hypotheses; weak linkage to routine spine surgical decisions
Drug target prediction and lead optimization[10,113]	Exploratory/Preclinical evidence	Low (computational exploration; biological and clinical validation needed)	Requires experimental validation, disease-specific biology, safety testing, and trials before clinical relevance
Foundation-model coordination system[11-13]	Conceptual/Prospective; no end-to-end spine-care validation	Low (prospective framework; no deployed system yet validated)	Best framed as clinician-supervised evidence synthesis and workflow orchestration, not autonomous care

Readiness levels represent a qualitative assessment of current clinical translation maturity rather than a formal regulatory classification. These judgments were based on the available evidence across four domains: (1) technical validation and reported performance; (2) external or multicenter validation; (3) integration into real clinical workflows; and (4) evidence of impact on clinical outcomes, safety, or decision-making. "Moderate" readiness indicates that clinical feasibility has been demonstrated, but evidence remains limited by factors such as single-center validation, retrospective evaluation, heterogeneous outcomes, or incomplete workflow integration. "Low-to-moderate" readiness indicates promising proof-of-concept or early clinical application, but substantial prospective validation, interoperability assessment, or outcome-based evidence is still required. "Low" readiness indicates that the technology is at an exploratory, conceptual, or early feasibility stage, with limited or no validation in relevant clinical environments. These classifications reflect the synthesis of the published literature and should not be interpreted as definitive implementation recommendations. AI: Artificial intelligence; SCI: spinal cord injury; AR: augmented reality; VR: virtual reality; BCI: brain-computer interface.

accompanied by continued monitoring after deployment, clear routes for reporting failures, and restrictions on use when performance is uncertain. A system that performs poorly in a defined population should be recalibrated, updated, restricted, or withdrawn from that use case. Responsibility for harm must also be agreed before deployment, because an adverse event may involve decisions made by the clinician, hospital, developer, and device vendor[147]. Unclear accountability may encourage either excessive reliance on the model or reluctance to use it at all. Practical constraints also matter. If reliable computing, connectivity, technical support, and time for workflow redesign are available only in well-resourced centers, AI may widen existing differences in access to and quality of spine care.

Safety architecture

The risks associated with AI-enabled systems partly depend on the degree of their integration with clinical operations. A language model used to summarize records or retrieve evidence has a different role from a navigation platform, robotic system, or neuromodulation device that interacts more directly with patient care. These functions should be evaluated separately. In high-risk settings, the broader clinical system may present information and outputs from validated specialist modules, but changes to an operative trajectory, instrument movement, or treatment delivery should remain under explicit clinician control[142]. Knowing when not to provide an answer is equally important. If the available data are incomplete, poor in quality, internally inconsistent, or outside the intended population, deferral may be safer than a confident recommendation. The system should make uncertainty visible and indicate why a result may be unreliable. High-risk outputs should also be traceable to the evidence or specialist-module result on which they are based. A fluent explanation is not a substitute for adequate support and may, in some circumstances, give an unsupported conclusion undue credibility.

Clinical use also requires an auditable record of how an output was produced and handled. Relevant logs should include the system version, the information used, retrieved evidence, generated outputs, and any clinical review or override. These records allow institutions to investigate incidents, identify changes in performance, and determine whether a function should be updated or restricted. Bias monitoring and post-deployment surveillance belong to the same process. The aim is not simply to keep a clinician nominally "in the loop", but to ensure that AI-supported decisions remain reviewable, contestable, and under clinical control. The proposed architecture is shown in Figure 4.

CONCLUSION

Multimodal AI orchestration systems could help coordinate information that is currently used separately across admission and triage, imaging, preoperative planning, intraoperative support, postoperative surveillance, and rehabilitation. Their credible near-term role is not autonomous diagnosis or treatment, but clinician-supervised evidence synthesis, workflow orchestration, and presentation of uncertainty. The language-model core should be distinct from the clinical system that governs data ingestion, specialist modules, retrieval, calibration, safety, audit, and human review. Molecular and drug-development applications sit outside routine care and should be judged by whether they generate experimentally validated mechanisms, biomarkers, endotypes, or trial-enrichment hypotheses. Progress toward clinical will depend on data quality, external validation, prospective testing, transparent uncertainty management, evidence traceability, governance, and continued clinician oversight. In the near term, the most credible role of these systems is to coordinate validated specialist modules, summarize evidence, expose uncertainty, and support clinician-led decisions.

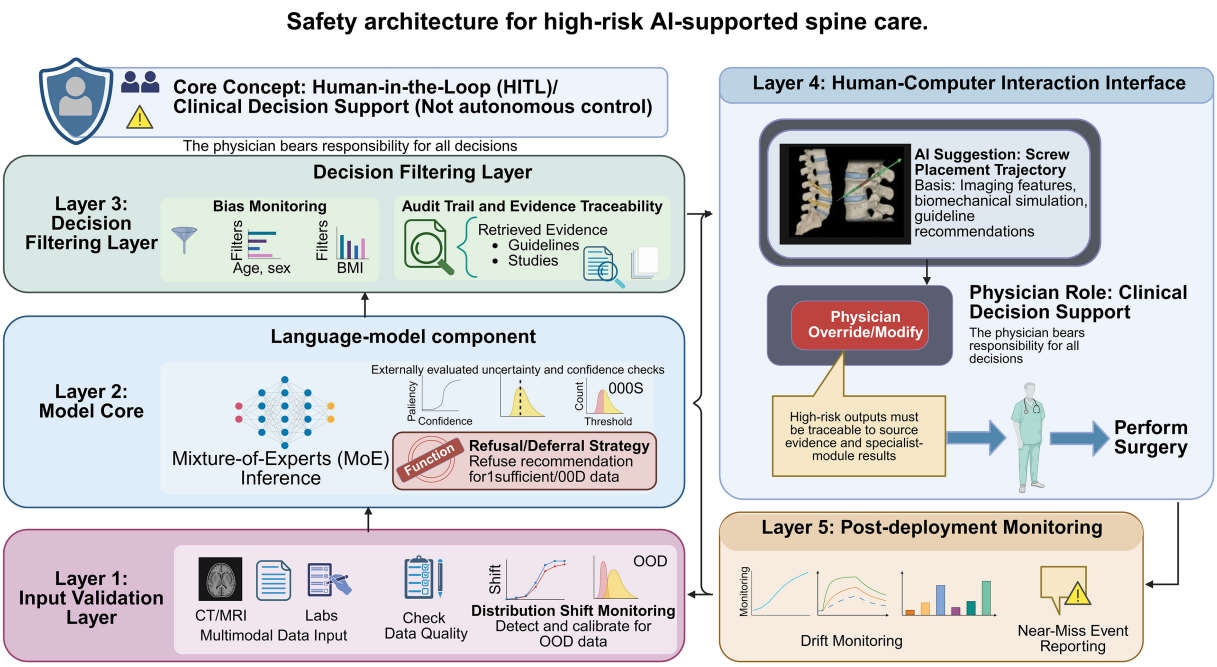


Figure 4. Safety architecture for high-risk AI-supported spine care. The framework includes input validation, model-core uncertainty quantification, refusal or deferral under insufficient or out-of-distribution data, decision filtering, bias monitoring, audit trails, evidence traceability, physician override, and post-deployment monitoring. The central principle is that AI should support clinical decision-making, while final responsibility and action remain with the physician. Created in BioRender. Niu, J. (2026) https://BioRender.com/agwy6nr. AI: Artificial intelligence; HITL: human-in-the-loop; BMI: body mass index; MoE: mixture-of-experts; CT: computed tomography; MRI: magnetic resonance imaging; OOD: out-of-distribution.

DECLARATIONS

Acknowledgments

The Graphical Abstract was created with BioRender.com [Created in BioRender. Niu, J. (2026) https://BioRender.com/b0qr600]. We acknowledge the support from the Institute of Philanthropy Medical and Health Scholarship-cum-Fellowship for Top Talent in Mainland.

Authors' contributions

Conceptualization, writing - review and editing, writing - original draft: Niu J
Writing - review and editing: Li Z, Zou Y, Wang Z, Sun M
Conceptualization, project administration, writing - original draft, writing - review and editing: Liu R, Wang W, Ning B

Availability of data and materials

Not applicable.

AI and AI-assisted tools statement

Not applicable.

Financial support and sponsorship

This work was supported by the National Natural Science Fund of China (82202750, 82071383, 82371392, 82300294, 82402801), Natural Science Foundation of Shandong Province (ZR2020KH007, ZR2021QH178, ZR2023QH321, ZR2024MH012), the "Taishan Scholar Distinguished Expert Program" of Shandong Province (tstp20231257), Shandong Province Youth Innovation Team Development Plan for Higher Education Institutions (2024KJJ010, 2024KJJ078); Shandong Province Medical and Health Science and Technology Project (202404071114); Jinan Clinical Medical Science and Technology Innovation Fund (202328035), the Beijing Natural Science Foundation (L252119), and Young Talent of Lifting engineering for Science and Technology in Shandong (SDAST2025QTA009).

Conflicts of interest

Wang W is the Guest Editor of the Special Issue "Artificial Intelligence and Digital Twins in Orthopedic Diseases and Bone Regeneration" of *Artificial Intelligence Surgery*. Wang W was not involved in any stage of the editorial process for this manuscript, including reviewer selection, manuscript handling, or decision-making. The other authors declare that there are no conflicts of interest related to this manuscript.

Ethical approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

REFERENCES

1. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet.* 2018;392:1789-858. DOI PubMed PMC
2. Muelbauer EJ, Alvi MA, Kennedy DJ, Fehlings MG. The future is now: how AI is reshaping spine care. *N Am Spine Soc J.* 2025;24:100825. DOI PubMed PMC
3. Yahanda AT, Joseph K, Bui T, et al. Current applications and future implications of artificial intelligence in spine surgery and research: a narrative review and commentary. *Global Spine J.* 2025;15:1445-54. DOI PubMed PMC
4. Lee S, Jung JY, Mahatthanatrakul A, Kim JS. Artificial intelligence in spinal imaging and patient care: a review of recent advances. *Neurospine.* 2024;21:474-86. DOI PubMed PMC
5. Kumar R, Dougherty C, Sporn K, et al. Intelligence architectures and machine learning applications in contemporary spine care. *Bioengineering.* 2025;12:967. DOI PubMed PMC
6. Macleod JS, Compton T, Bakaes Y, et al. Artificial intelligence in spine surgery: imaging-based applications for diagnosis and surgical techniques. *Curr Rev Musculoskelet Med.* 2025;18:398-405. DOI PubMed PMC
7. Guo S, Yang Y, Wang M, et al. Effects of an exoskeleton robot on motor function in patients with spinal cord injuries: a systematic review and meta-analysis. *Syst Rev.* 2025;14:218. DOI PubMed PMC
8. Li J, Chen T, Yan X, Luo L. The effect of device-based neuromodulation on the motor recovery of patients with spinal cord injury. *Spinal Cord.* 2025;63:621-32. DOI PubMed
9. Wang X, Chen N, Du Z, et al. Bioinformatics analysis integrating metabolomics of m^{6}A RNA microarray in intervertebral disc degeneration. *Epigenomics.* 2020;12:1419-41. DOI PubMed
10. Jusoh AS, Remli MA, Mohamad MS, Cazenave T, Fong CS. How generative Artificial Intelligence can transform drug discovery? *Eur J Med Chem.* 2025;295:117825. DOI PubMed
11. Zhang D, Li ZZ, Zhang ML, et al. From System 1 to System 2: a survey of reasoning large language models. *IEEE Trans Pattern Anal Mach Intell.* 2026;48:3335-54. DOI PubMed
12. Wang J, Redelmeier DA. Artificial intelligence chain-of-thought reasoning in nuanced medical scenarios: mitigation of cognitive biases through model intransigence. *BMJ Qual Saf.* 2026;35:456-63. DOI PubMed
13. Duan Z, Huang X, Lu R, et al. Multi-center benchmarking of large language models for clinical decision support in lung cancer screening. *Cell Rep Med.* 2025;6:102465. DOI PubMed PMC
14. Vasey B, Nagendran M, Campbell B, et al. Reporting guideline for the early stage clinical evaluation of decision support systems driven by artificial intelligence: DECIDE-AI. *BMJ.* 2022;377:e070904. DOI PubMed PMC
15. Lekadir K, Frangi AF, Porras AR, et al. FUTURE-AI: international consensus guideline for trustworthy and deployable artificial intelligence in healthcare. *BMJ.* 2025;388:e081554. DOI PubMed PMC
16. Yilihamu EE, Zeng FS, Shang J, Yang JT, Zhong H, Feng SQ. GPT4LFS (generative pretrained transformer 4 omni for lumbar foramina stenosis): enhancing lumbar foraminal stenosis image classification through large multimodal models. *Spine J.* 2025;25:2071-80. DOI PubMed
17. Liu J, Zhu X, Shi Z, et al. Text-guided multimodal deep learning in magnetic resonance imaging for spinal structures segmentation and lumbar abnormalities identification. *Quant Imaging Med Surg.* 2025;15:9710-28. DOI PubMed PMC

18. Gan L, Ma M, Liu Y, et al. A clinical-radiomics model for predicting axillary pathologic complete response in breast cancer with axillary lymph node metastases. *Front Oncol.* 2021;11:786346. DOI PubMed PMC
19. Balza R, Palmer WE. Symptom-imaging correlation in lumbar spine pain. *Skeletal Radiol.* 2023;52:1901-9. DOI PubMed
20. Bizzo BC, Dasegowda G, Bridge C, et al. Addressing the challenges of implementing artificial intelligence tools in clinical practice: principles from experience. *J Am Coll Radiol.* 2023;20:352-60. DOI PubMed
21. Mcsweeney TP, Tiulpin A, Saarakkala S, et al. External validation of spinenet, an open-source deep learning model for grading lumbar disk degeneration MRI features, Using the Northern Finland Birth Cohort 1966. *Spine.* 2023;48:484-91. DOI PubMed PMC
22. Grob A, Loibl M, Jamaludin A, et al. External validation of the deep learning system "SpineNet" for grading radiological features of degeneration on MRIs of the lumbar spine. *Eur Spine J.* 2022;31:2137-48. DOI PubMed
23. Zhu Z, Li Q, Wei Y, Song R. Hierarchical multi-level dynamic hyperparameter deformable image registration with convolutional neural network. *Phys Med Biol.* 2024;69:175007. DOI PubMed
24. Zhang K, Li Z, Cai C, et al. Semi-supervised graph convolutional networks for the domain adaptive recognition of thyroid nodules in cross-device ultrasound images. *Med Phys.* 2023;50:7806-21. DOI PubMed
25. Kuang Z, Yan Z, Abayazeed A, Wagner F, Yu L, Reyes M. ROXSI: robust cross-sequence semantic interaction for brain tumor segmentation on multi-sequence MR images. *IEEE J Biomed Health Inform.* 2025;29:2899-910. DOI PubMed
26. Amjad A, Xu J, Thill D, et al. Deep learning auto-segmentation on multi-sequence magnetic resonance images for upper abdominal organs. *Front Oncol.* 2023;13:1209558. DOI PubMed PMC
27. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods.* 2020;18:203-11. DOI PubMed
28. Nauroth-Kreß AC, Weiner S, Hölzli L, et al. Automated segmentation of the dorsal root ganglia in MRI. *Neuroimage.* 2025;311:121189. DOI PubMed
29. Xia Y, Li J, Deng B, et al. Knowledge graph-enhanced deep learning model (H-SYSTEM) for hypertensive intracerebral hemorrhage: model development and validation. *J Med Internet Res.* 2025;27:e66055. DOI PubMed PMC
30. Yousefzadeh N, Tran C, Ramirez-Zamora A, Chen J, Fang R, Thai MT. Neuron-level explainable AI for Alzheimer's Disease assessment from fundus images. *Sci Rep.* 2024;14:7710. DOI PubMed PMC
31. Huang W, Cao G, Xia J, Chen J, Wang H, Zhang J. H-calibration: rethinking classifier recalibration with probabilistic error-bounded objective. *IEEE Trans Pattern Anal Mach Intell.* 2025;47:9023-42. DOI PubMed
32. Huang Y, Yang G, Shen Y, et al. Application of large language models in complex clinical cases: cross-sectional evaluation study. *JMIR Med Inform.* 2025;13:e73941. DOI PubMed PMC
33. Khan R, Alzaben N, Daradkeh YI, Lee MY, Ullah I. Bilateral collaborative streams with multi-modal attention network for accurate polyp segmentation. *Sci Rep.* 2025;15:34182. DOI PubMed PMC
34. Abugabah A, Shukla PK, Shukla PK, Pandey A. An intelligent healthcare system for rare disease diagnosis utilizing electronic health records based on a knowledge-guided multimodal transformer framework. *BioData Min.* 2025;18:70. DOI PubMed PMC
35. Li Y, Fredrickson V, Resnick DK. How should we grade lumbar disc herniation and nerve root compression? A systematic review. *Clinical Orthopaedics & Related Research.* 2015;473:1896-902. DOI PubMed PMC
36. Su ZH, Liu J, Yang MS, et al. Automatic grading of disc herniation, central canal stenosis and nerve roots compression in lumbar magnetic resonance image diagnosis. *Front Endocrinol.* 2022;13:890371. DOI PubMed PMC
37. Zhang W, Chen Z, Su Z, et al. Deep learning-based detection and classification of lumbar disc herniation on magnetic resonance images. *JOR Spine.* 2023;6:e1276. DOI PubMed PMC
38. Xu Y, Zheng S, Tian Q, et al. Deep learning model for grading and localization of lumbar disc herniation on magnetic resonance imaging. *J Magn Reson Imaging.* 2024;61:364-75. DOI PubMed
39. Baur D, Bieck R, Berger J, et al. Automated three-dimensional imaging and pfirrmann classification of intervertebral disc using a graphical neural network in sagittal magnetic resonance imaging of the lumbar spine. *J Imaging Inform Med.* 2025;38:979-87. DOI PubMed PMC
40. Wang D, Yang Z, Huang Z, Gu L. Spine segmentation with multi-view GCN and boundary constraint. *Annu Int Conf IEEE Eng Med Biol Soc.* 2022;2022:2136-9. DOI PubMed
41. Ghobrial G, Roth C. Deep learning-based automated segmentation and quantification of the dural sac cross-sectional area in lumbar spine MRI. *Front Radiol.* 2025;5:1503625. DOI PubMed PMC
42. Wáng YXJ, Diacinti D, Aparisi Gómez MP, et al. Radiological diagnosis of prevalent osteoporotic vertebral fracture on radiographs: an interim consensus from a group of experts of the ESSR osteoporosis and metabolism subcommittee. *Skeletal Radiol.* 2024;53:2563-74. DOI PubMed PMC
43. Tian J, Wang K, Wu P, Li J, Zhang X, Wang X. Development of a deep learning model for detecting lumbar vertebral fractures on CT images: an external validation. *Eur J Radiol.* 2024;180:111685. DOI PubMed

44. Glessgen C, Cyriac J, Yang S, et al. A deep learning pipeline for systematic and accurate vertebral fracture reporting in computed tomography. *Clin Radiol.* 2025;83:106827. DOI PubMed
45. Won D, Lee HJ, Lee SJ, Park SH. Lumbar Spinal Stenosis Grading in Multiple Level Magnetic Resonance Imaging Using Deep Convolutional Neural Networks. *Global Spine J.* 2025;15:2309-17. DOI PubMed PMC
46. Wang T, Chen R, Fan N, et al. Machine learning and deep learning for diagnosis of lumbar spinal stenosis: systematic review and meta-analysis. *J Med Internet Res.* 2024;26:e54676. DOI PubMed PMC
47. Muiño-Mosquera L, Cervi E, De Groote K, et al. Management of aortic disease in children with *FBN1*-related Marfan syndrome. *Eur Heart J.* 2024;45:4156-69. DOI PubMed PMC
48. Le Huu A, Olive JK, Cekmecelioglu D, et al. Endovascular therapy for patients with heritable thoracic aortic disease. *Ann Cardiothorac Surg.* 2022;11:31-6. DOI PubMed PMC
49. Lin M, Liu Z, Liu G, et al. ; Deciphering Disorders Involving Scoliosis and COmorbidities (DISCO) study. Genetic and molecular mechanism for distinct clinical phenotypes conveyed by allelic truncating mutations implicated in *FBN1*. *Mol Genet Genomic Med.* 2020;8:e1023. DOI PubMed PMC
50. Bradley-Watson J, Glatzel H, Turner HE, Orchard E. Elective aortic surgery for prevention of aortic dissection in turner syndrome: the potential impact of updated european society of cardiology and international turner syndrome consensus group guidelines on referrals to the heart team. *Clin Endocrinol.* 2025;102:559-64. DOI PubMed
51. Xu D, Descovich M, Liu H, Sheng K. Robust localization of poorly visible tumor in fiducial free stereotactic body radiation therapy. *Radiother Oncol.* 2024;200:110514. DOI PubMed
52. Beaufort Q, Lefevre V, Carsuzaa T. Imaging of spinal metastases: modality performance, complications, and differential diagnosis. *Neurochirurgie.* 2026;72:101804. DOI PubMed
53. Bahouth SM, Yeboa DN, Ghia AJ, et al. Multidisciplinary management of spinal metastases: what the radiologist needs to know. *Br J Radiol.* 2022;95:20220266. DOI PubMed PMC
54. Huang Y, Jiao J, Yu J, Zheng Y, Wang Y. Anatomy-inspired model for critical landmark localization in 3D spinal ultrasound volume data. *Med Image Anal.* 2025;103:103610. DOI PubMed
55. Zhao S, Zhao H, Zhao L, et al. ; Deciphering disorders Involving Scoliosis and COmorbidities (DISCO) study. Unraveling the genetic architecture of congenital vertebral malformation with reference to the developing spine. *Nat Commun.* 2024;15:1125. DOI PubMed PMC
56. Wan HS, Wong DLL, To CS, Meng N, Zhang T, Cheung JPY. 3D prediction of curve progression in adolescent idiopathic scoliosis based on biplanar radio logical reconstruction: a systematic review. *Bone Jt Open.* 2024;5:243-51. DOI PubMed PMC
57. Tiede-Lewis LM, Dallas SL. Changes in the osteocyte lacunocanalicular network with aging. *Bone.* 2019;122:101-13. DOI PubMed PMC
58. Wang H, Zhang T, Zhang C, et al. An intelligent composite model incorporating global/regional X-rays and clinical parameters to predict progressive adolescent idiopathic scoliosis curvatures and facilitate population screening. *EBioMedicine.* 2023;95:104768. DOI PubMed PMC
59. Cortellini A, Santo V, Brunetti L, et al. Transformer-based AI approach to unravel long-term, time-dependent prognostic complexity in patients with advanced NSCLC and PD-L1 ≥ 50%: insights from the pembrolizumab 5-year global registry. *J Immunother Cancer.* 2025;13:e012423. DOI PubMed PMC
60. Wang Y, Hao R, Guo L, Zhou X. Biomechanical optimization study of posterior tilt extension stems in the repair of tibial plateau bone defects. *Front Bioeng Biotechnol.* 2025;13:1688915. DOI PubMed PMC
61. Ahmadi M, Zhang X, Lin M, et al. Automated finite element modeling of the lumbar spine: a biomechanical and clinical approach to spinal load distribution and stress analysis. *World Neurosurg.* 2025;201:124236. DOI PubMed
62. Silveira JT, S G, Kundapur PP. Automated lumbar spine segmentation in MRI using an enhanced U-Net with inception module and dual-output mechanism. *Sci Rep.* 2025;15:39215. DOI PubMed PMC
63. Caprara S, Carrillo F, Snedeker JG, Farshad M, Senteler M. Automated pipeline to generate anatomically accurate patient-specific biomechanical models of healthy and pathological FSUs. *Front Bioeng Biotechnol.* 2021;9:636953. DOI PubMed PMC
64. Zaidi Q, Danisa OA, Cheng W. Measurement techniques and utility of hounsfield unit values for assessment of bone quality prior to spinal instrumentation: a review of current literature. *Spine.* 2019;44:E239-44. DOI PubMed
65. McLean KA, Sgrò A, Brown LR, et al. ; TWIST Collaborators. Multimodal machine learning to predict surgical site infection with healthcare workload impact assessment. *NPJ Digit Med.* 2025;8:121. DOI PubMed PMC
66. Chang G, Rajapakse CS, Philipp TC, Madi R, Sheth NP, Protopsaltis TS. Preoperative CT-based finite element vertebral modulus analysis predicts bone quality-related complications after lumbar spine fusion. *Spine.* 2026; Epub ahead of print. DOI PubMed
67. Franceschini C, Ahmadi M, Zhang X, et al. Revolutionizing spine surgery with emerging AI-FEA integration. *J Robot Surg.* 2025;19:615. DOI PubMed PMC

68. Fritzell P, Mesterton J, Hagg O. Prediction of outcome after spinal surgery - using The Dialogue Support based on the Swedish national quality register. *Eur Spine J.* 2021;31:889-900. DOI PubMed
69. Nguyen D, Beauséjour MH, Solorzano Barrera C, Qiao N, Villemure I, Aubin CÉ. Finite element modeling of pedicle screw fixation considering patient-specific bone density. *Comput Methods Biomech Biomed Engin.* 2025:1-11. DOI PubMed
70. Sensale M, Vendeuvre T, Schilling C, Grupp T, Rochette M, Dall'Ara E. Patient-specific finite element models of posterior pedicle screw fixation: effect of screw's size and geometry. *Front Bioeng Biotechnol.* 2021;9:643154. DOI PubMed PMC
71. Wu H, Tang X, Wang W, Li R. The AI-powered orthopedic digital twin: a new paradigm for personalized diagnosis, surgical simulation, and prognostic management - a scoping review. *Art Int Surg.* 2026;6:320-39. DOI
72. Lungu AJ, Swinkels W, Claesen L, Tu P, Egger J, Chen X. A review on the applications of virtual reality, augmented reality and mixed reality in surgical simulation: an extension to different kinds of surgery. *Expert Rev Med Devices.* 2021;18:47-62. DOI PubMed
73. Ghaednia H, Fourman MS, Lans A, et al. Augmented and virtual reality in spine surgery, current applications and future potentials. *Spine J.* 2021;21:1617-25. DOI PubMed
74. Jung Y, Muddaluru V, Gandhi P, Pahuta M, Guha D. The development and applications of augmented and virtual reality technology in spine surgery training: a systematic review. *Can J Neurol Sci.* 2024;51:255-64. DOI PubMed
75. Bui T, Ruiz-Cardozo MA, Dave HS, et al. Virtual, augmented, and mixed reality applications for surgical rehearsal, operative execution, and patient education in spine surgery: a scoping review. *Medicina.* 2024;60:332. DOI PubMed PMC
76. Weiss MY, Melnyk R, Mix D, Ghazi A, Vates GE, Stone JJ. Design and validation of a cervical laminectomy simulator using 3d printing and hydrogel phantoms. *Oper Surg.* 2020;18:202-8. DOI PubMed
77. Adermann J, Geissler N, Bernal LE, Kotzsch S, Korb W. Development and validation of an artificial wetlab training system for the lumbar discectomy. *Eur Spine J.* 2014;23:1978-83. DOI PubMed
78. Luciano CJ, Banerjee PP, Bellotte B, et al. Learning retention of thoracic pedicle screw placement using a high-resolution augmented reality simulator with haptic feedback. *Neurosurgery.* 2011;69:ons14-9. DOI PubMed PMC
79. Pfandler M, Lazarovici M, Stefan P, Wucherer P, Weigl M. Virtual reality-based simulators for spine surgery: a systematic review. *Spine J.* 2017;17:1352-63. DOI PubMed
80. Sumdani H, Aguilar-Salinas P, Avila MJ, Barber SR, Dumont T. Utility of augmented reality and virtual reality in spine surgery: a systematic review of the literature. *World Neurosurg.* 2022;161:e8-17. DOI PubMed
81. Li HM, Zhang RJ, Shen CL. Accuracy of pedicle screw placement and clinical outcomes of robot-assisted technique versus conventional freehand technique in spine surgery from nine randomized controlled trials: a meta-analysis. *Spine.* 2020;45:E111-9. DOI PubMed
82. Riewruja K, Tanasansomboon T, Yingsakmongkol W, et al. A network meta-analysis comparing the efficacy and safety of pedicle screw placement techniques using intraoperative conventional, navigation, robot-assisted, and augmented reality guiding systems. *Int J Spine Surg.* 2024;18:551-70. DOI PubMed PMC
83. Girgis M, Kelliher A, Pheasant MS, Tang A, Badve S, Chen T. Artificial intelligence in intraoperative imaging and navigation for spine surgery: a narrative review. *J Clin Med.* 2026;15:2779. DOI PubMed PMC
84. Liebmann F, von Atzigen M, Stütz D, et al. Automatic registration with continuous pose updates for marker-less surgical navigation in spine surgery. *Med Image Anal.* 2024;91:103027. DOI PubMed
85. Lu H, Liu M, Yu K, Fang Y, Zhao J, Shi Y. A deep learning-based fully automated vertebra segmentation and labeling workflow. *Br J Hosp Med.* 2025;86:1-22. DOI PubMed
86. Ao Y, Esfandiari H, Carrillo F, et al. SafeRPlan: safe deep reinforcement learning for intraoperative planning of pedicle screw placement. *Med Image Anal.* 2025;99:103345. DOI PubMed
87. Hao L, Liu D, Du S, et al. An improved path planning algorithm based on artificial potential field and primal-dual neural network for surgical robot. *Comput Methods Programs Biomed.* 2022;227:107202. DOI PubMed
88. Klitzner F, Inigo Romillo B, Killeen BD, et al. Investigating robot control policy learning for autonomous x-ray-guided spine procedures. *Int J Comput Assist Radiol Surg.* 2026;21:1387-96. DOI PubMed
89. Ahmadi M, Chen H, Lin M, et al. Streamlined and efficient patient-specific modeling for lumbar spine segmentation and finite element analysis. *Sci Rep.* 2025;15:35619. DOI PubMed PMC
90. Götschi T, Maranta G, Cornaz F, Abel F, Farshad M, Widmer J. Automated multi-objective pedicle screw planning. *J Spine Surg.* 2026;15:51. DOI PubMed PMC
91. Luchmann D, Jecklin S, Cavalcanti NA, et al. Spinal navigation with AI-driven 3D-reconstruction of fluoroscopy images: an ex-vivo feasibility study. *BMC Musculoskelet Disord.* 2024;25:925. DOI PubMed PMC
92. Mesinovic M, Watkinson P, Zhu T. DySurv: dynamic deep learning model for survival analysis with conditional variational inference. *J Am Med Inform Assoc.* 2026;33:112-22. DOI PubMed PMC
93. Kim KB, Kim GB, Kim JH, Lee SM. Artificial intelligence in total knee arthroplasty: clinical applications and implications. *Knee Surg Relat Res.* 2025;37:44. DOI PubMed PMC

94. Wang T, Chen R, Liang M, et al. Comparative performance of LLMs and machine learning in predicting complications after percutaneous kyphoplasty for osteoporotic vertebral compression fractures. *NPJ Digit Med.* 2026;9:401. DOI PubMed PMC
95. Hoare D, Kingsmore D, Holsgrove M, et al. Realtime monitoring of thrombus formation in vivo using a self-reporting vascular access graft. *Commun Med.* 2024;4:15. DOI PubMed PMC
96. Sun Y, Meng K, Lang S, Li P, Wang W, Yang J. Multi-classification model for PPG signal arrhythmia based on time-frequency dual-domain attention fusion. *Sensors.* 2025;25:5985. DOI PubMed PMC
97. Shickel B, Loftus TJ, Ruppert M, et al. Dynamic predictions of postoperative complications from explainable, uncertainty-aware, and multi-task deep neural networks. *Sci Rep.* 2023;13:1224. DOI PubMed PMC
98. Sumner J, Lim HW, Chong LS, Bundele A, Mukhopadhyay A, Kayambu G. Artificial intelligence in physical rehabilitation: a systematic review. *Artif Intell Med.* 2023;146:102693. DOI PubMed
99. Lanotte F, O'brien MK, Jayaraman A. AI in rehabilitation medicine: opportunities and challenges. *Ann Rehabil Med.* 2023;47:444-58. DOI PubMed PMC
100. Khalid UB, Naeem M, Stasolla F, Syed MH, Abbas M, Coronato A. Impact of AI-powered solutions in rehabilitation process: recent improvements and future trends. *Int J Gen Med.* 2024;17:943-69. DOI PubMed PMC
101. Zhu A, Liu Y, Liu Y. Identification of key genes and regulatory mechanisms in adult degenerative scoliosis. *J Clin Neurosci.* 2024;119:170-9. DOI PubMed
102. Wang W, Liu C, He D, et al. CircRNA CDR1as affects functional repair after spinal cord injury and regulates fibrosis through the SMAD pathway. *Pharmacol Res.* 2024;204:107189. DOI PubMed
103. Mirza B, Wang W, Wang J, Choi H, Chung NC, Ping P. Machine learning and integrative analysis of biomedical big data. *Genes.* 2019;10:87. DOI PubMed PMC
104. Morello G, La Cognata V, Guarnaccia M, Gentile G, Cavallaro S. Artificial intelligence-driven multi-omics approaches in glioblastoma. *Int J Mol Sci.* 2025;26:9362. DOI PubMed PMC
105. Yang X, Mann KK, Wu H, Ding J. scCross: a deep generative model for unifying single-cell multi-omics with seamless integration, cross-modal generation, and in silico exploration. *Genome Biol.* 2024;25:198. DOI PubMed PMC
106. Wang R, Zheng Y, Zhang Z, et al. MATES: a deep learning-based model for locus-specific quantification of transposable elements in single cell. *Nat Commun.* 2024;15:8798. DOI PubMed PMC
107. Long W, Liu T, Xue L, Zhao H. spVelo: RNA velocity inference for multi-batch spatial transcriptomics data. *Genome Biol.* 2025;26:239. DOI PubMed PMC
108. Peng H, Chen P, Liu R, Chen L. Spatiotemporal information conversion machine for time-series forecasting. *Fundam Res.* 2024;4:1674-87. DOI PubMed PMC
109. Ramakrishnan S, Cortes-Gomez E, Athans SR, et al. Race-specific coregulatory and transcriptomic profiles associated with DNA methylation and androgen receptor in prostate cancer. *Genome Med.* 2024;16:52. DOI PubMed PMC
110. Qiu W, Dincer AB, Janizek JD, et al. Deep profiling of gene expression across 18 human cancers. *Nat Biomed Eng.* 2025;9:333-55. DOI PubMed
111. Pan W, Hathi D, Xu Z, Zhang Q, Li Y, Wang F. Identification of predictive subphenotypes for clinical outcomes using real world data and machine learning. *Nat Commun.* 2025;16:3797. DOI PubMed PMC
112. Luttens A, Cabeza de Vaca I, Sparring L, et al. Rapid traversal of vast chemical space using machine learning-guided docking screens. *Nat Comput Sci.* 2025;5:301-12. DOI PubMed PMC
113. Liu Z, Huang D, Zheng S, et al. Deep learning enables discovery of highly potent anti-osteoporosis natural products. *Eur J Med Chem.* 2021;210:112982. DOI PubMed
114. Sevgen E, Moller J, Lange A, et al. ProT-VAE: Protein Transformer Variational AutoEncoder for functional protein design. *Proc Natl Acad Sci U S A.* 2025;122:e2408737122. DOI PubMed PMC
115. Hostaš J, Ghaemi MS, Hu H, Lin J, Hu A, Ooi HK. VNFlow: integration of variational autoencoders and normalizing flows for novel molecular design. *J Cheminform.* 2025;17:161. DOI PubMed PMC
116. Li S, Fan J, He H, Zhou R, Liao J. MolP-PC: a multi-view fusion and multi-task learning framework for drug ADMET property prediction. *Chin J Nat Med.* 2025;23:1293-300. DOI PubMed
117. Yang Y, Zhou H, Wang Y, et al. Phillygenin improves the focal adhesion kinase/glycogen synthase kinase 3β/β-catenin axis to promote mesenchymal stem cell osteogenic differentiation. *Phytother Res.* 2025;39:4991-5005. DOI PubMed PMC
118. Li J, Chen J, Bai H, et al. An Overview of Organs-on-Chips Based on Deep Learning. *Research.* 2022;2022:9869518. DOI PubMed PMC
119. Xiao RR, Jing B, Yan L, Li J, Tu P, Ai X. Constant-rate perfused array chip for high-throughput screening of drug permeability through brain endothelium. *Lab Chip.* 2022;22:4481-92. DOI PubMed

120. Zhou L, Li G, Yao J, et al. Emulation and evaluation of tumor cell combined chemotherapy in isotropic/anisotropic collagen fiber microenvironments. *Lab Chip.* 2024;24:2999-3014. DOI PubMed
121. Lorach H, Galvez A, Spagnolo V, et al. Walking naturally after spinal cord injury using a brain-spine interface. *Nature.* 2023;618:126-33. DOI PubMed PMC
122. Hankov N, Caban M, Demesmaeker R, et al. Augmenting rehabilitation robotics with spinal cord neuromodulation: a proof of concept. *Sci Robot.* 2025;10:eadn5564. DOI PubMed
123. Yao Y, Hasan WZW, Jiao W, et al. ChatGPT and BCI-VR: a new integrated diagnostic and therapeutic perspective for the accurate diagnosis and personalized treatment of mild cognitive impairment. *Front Hum Neurosci.* 2024;18:1426055. DOI PubMed PMC
124. Awuah WA, Ahluwalia A, Darko K, et al. Bridging minds and machines: the recent advances of brain-computer interfaces in neurological and neurosurgical applications. *World Neurosurg.* 2024;189:138-53. DOI PubMed
125. Xu D, Tang F, Li Y, Zhang Q, Feng X. An analysis of deep learning models in SSVEP-based BCI: a survey. *Brain Sciences.* 2023;13:483. DOI PubMed PMC
126. Moreno-Castelblanco SR, Vélez-Guerrero MA, Callejas-Cuervo M. Artificial intelligence approaches for EEG signal acquisition and processing in lower-limb motor imagery: a systematic review. *Sensors.* 2025;25:5030. DOI PubMed PMC
127. Cha JM, Hong J, Yoo J, Rha DW. Wearable robots for rehabilitation and assistance of gait: a narrative review. *Ann Rehabil Med.* 2025;49:187-95. DOI PubMed PMC
128. Moriarty B, Jacob T, Sadlowski M, et al. The use of exoskeleton robotic training on lower extremity function in spinal cord injuries: a systematic review. *J Orthop.* 2025;65:1-7. DOI PubMed PMC
129. Kim RY, Biller OM, Mulcahey MJ. Evaluating therapeutic effects of exoskeletons and FES in SCI: integrative review of the literature. *Spinal Cord.* 2025;63:323-32. DOI PubMed PMC
130. White C, Doherty O, Smith E, et al. Exoskeleton training for spinal cord injury neuropathic pain (ExSCIP): protocol for a phase 2 feasibility randomised trial. *HRB Open Res.* 2024;7:55. DOI PubMed PMC
131. Onate D, Hogan C, Fitzgerald K, White KT, Tansey K. Recommendations for clinical decision-making when offering exoskeletons for community use in individuals with spinal cord injury. *Front Rehabil Sci.* 2024;5:1428708. DOI PubMed PMC
132. Hasson CJ, Manczurowsky J, Collins EC, Yarossi M. Neurorehabilitation robotics: how much control should therapists have? *Front Hum Neurosci.* 2023;17:1179418. DOI PubMed PMC
133. Collins GS, Dhiman P, Ma J, et al. Evaluation of clinical prediction models (part 1): from development to external validation. *BMJ.* 2024;384:e074819. DOI PubMed PMC
134. Collins GS, Moons KGM, Dhiman P, et al. TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ.* 2024;385:e078378. DOI PubMed PMC
135. Moons KGM, Damen JAA, Kaul T, et al. PROBAST+AI: an updated quality, risk of bias, and applicability assessment tool for prediction models using regression or artificial intelligence methods. *BMJ.* 2025;388:e082505. DOI PubMed PMC
136. Tejani AS, Klontzas ME, Gatti AA, et al. ; CLAIM 2024 Update Panel. Checklist for artificial intelligence in medical imaging (CLAIM): 2024 update. *Radiol Artif Intell.* 2024;6:e240300. DOI PubMed PMC
137. Cruz Rivera S, Liu X, Chan AW, et al. ; SPIRIT-AI and CONSORT-AI Consensus Group. Guidelines for clinical trial protocols for interventions involving artificial intelligence: the SPIRIT-AI extension. *Nat Med.* 2020;26:1351-63. DOI PubMed PMC
138. Liu X, Cruz Rivera S, Moher D, et al. ; SPIRIT-AI and CONSORT-AI Working Group. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: the CONSORT-AI extension. *Nat Med.* 2020;26:1364-74. DOI PubMed PMC
139. Acosta JN, Falcone GJ, Rajpurkar P, Topol EJ. Multimodal biomedical AI. *Nat Med.* 2022;28:1773-84. DOI PubMed
140. Diniz P, Grimm B, Garcia F, et al. Digital twin systems for musculoskeletal applications: a current concepts review. *Knee Surg Sports Traumatol Arthrosc.* 2025;33:1892-910. DOI PubMed
141. Ahmed MM, Okesanya OJ, Oweidat M, Othman ZK, Musa SS, Lucero-Prisno Iii DE. The ethics of data mining in healthcare: challenges, frameworks, and future directions. *BioData Min.* 2025;18:47. DOI PubMed PMC
142. Lim M, Labaran L, Vengsarkar VA, Jin L, Li X. The devil is in the details-near-miss wrong-level thoracic spine surgery despite robotic navigation: a case report. *JBJS Case Connect.* 2025;15. DOI PubMed
143. Giouroukou K, Marias K, Tsiknakis M, Klontzas ME. Rethinking privacy in medical imaging AI: from metadata and pixel-level identification risks to federated learning and synthetic data challenges. *Radiol Artif Intell.* 2026;8:e250273. DOI PubMed
144. Bhavani Sowndharya B, Mathan Muthu CM, Vickram AS, Saravanan A. Bio-ethical considerations in the application of artificial intelligence in spinal surgery. *Brain Spine.* 2025;5:104161. DOI PubMed PMC
145. Lawson McLean A. Towards precision medicine in spinal surgery: leveraging AI technologies. *Ann Biomed Eng.* 2024;52:735-7. DOI PubMed PMC
146. Hu L, Li D, Liu H, et al. Enhancing fairness in AI-enabled medical systems with the attribute neutral framework. *Nat Commun.* 2024;15:8767. DOI PubMed PMC

147. Lawton T, Morgan P, Porter Z, et al. Clinicians risk becoming "liability sinks" for artificial intelligence. *Future Healthc J.* 2024;11:100007. DOI PubMed PMC