Cureus
Part of SPRINGER NATURE

Open Access Review Article

Governing Generative Artificial Intelligence in Anatomy Education: A Proposed Risk-Tiered Framework for Donor-Derived Material, Anatomical Fidelity, and Assessment

Review began 08/15/2026
Review ended 08/24/2026
Published 09/02/2026

© Copyright 2026
Mahgoub et al. This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

DOI: 10.7759/cureus.115656

Ahmed Mahgoub [1], Anaida Singh [2], Ayman Mahgoub [3]

1. Anatomy, University of Khartoum, Khartoum, SDN 2. Anatomical Sciences, St. George's University, St. George's, GRD 3. Research and Development, Anova Health, Riyadh, SAU

Corresponding author: Ahmed Mahgoub, mahgoub310@gmail.com

Abstract

Generative artificial intelligence (GenAI) is now routinely used in anatomy education to draft explanations, generate and interpret images, produce assessment items, and support feedback. Broad responsible-AI principles are well established, but they do not resolve the questions anatomy educators actually face: whether donor-derived material may be entered into an external system, what counts as adequate verification of a generated anatomical structure, and which assessment decisions must remain human. Anatomy faculty themselves report unclear governance in these areas and ask for discipline-specific policy. This article proposes an operational governance framework for GenAI in anatomy education, developed from a structured synthesis of 71 sources spanning direct anatomy studies, transferable health professions education (HPE) evidence, professional guidance, and ethical scholarship. Five discipline-specific pressure points motivate the proposal: donor-derived inputs retain stewardship obligations after digitization; anatomical fidelity is relational and structural rather than stylistic, so fluency and realism are not evidence of correctness; representation and clinically relevant variation form part of anatomical truth; assessment converts content error into consequential judgment about learners; and provenance, intellectual property, and disclosure support traceability and accountability. The framework governs the defined use case rather than the AI product. Purpose, audience, input and data class, consequence, scale, and degree of AI autonomy determine one of four proposed tiers (Low, Moderate, High, and Restricted/Avoid) under a conservative rule in which the highest triggered criterion sets the tier. Permitted uses then pass five control gates covering inputs and provenance, anatomical fidelity, variation and representation, assessment validity where applicable, and disclosure and accountability. Four decisions remain non-delegable to a machine, and approval is treated as a revisable state maintained through a monitoring, incident-response, and revalidation loop rather than a one-time authorization. The framework is an evidence-informed proposal for local adaptation, not a validated instrument; it has not been prospectively evaluated, its tier thresholds are not empirically calibrated, and local law, donor-program policy, and institutional governance remain controlling.

Categories: Anatomy, Healthcare Technology
Keywords: anatomy education, artificial intelligence governance, assessment, body donation, educational technology, generative artificial intelligence, human oversight, risk classification

Introduction And Background

Generative artificial intelligence (GenAI) is increasingly being studied and used in anatomy education. Large language and multimodal models can generate explanations and assessment items, create or interpret visual content, support feedback, and assist educational and research tasks [1]. Recent anatomy-education literature has examined both the potential utility of these systems and concerns regarding factual accuracy, reliability, transparency, bias, over-reliance, and the preservation of human judgment [1-3].

In this article, GenAI refers to artificial intelligence systems capable of generating new content, including text or images, in response to user inputs. This distinction matters because not all evidence relevant to GenAI governance comes from generative systems. This review focuses on GenAI; evidence from non-generative AI is included selectively when it provides directly transferable evidence for governance questions such as automated assessment, human oversight, or accountability, and such evidence is identified as transferable where it is used.

The need for discipline-specific governance is not merely theoretical. In a cross-sectional survey of 30 anatomy faculty at medical and health sciences colleges in the United Arab Emirates, 28 participants (93.3%) reported using GenAI. The most frequently reported educational uses were generating quiz or examination questions, summarizing complex topics, and drafting instructional materials; image or diagram generation was less common [4]. Faculty-identified needs included clearer governance and disclosure policies and safeguards for assessment [4]. The small sample limits generalizability, but the findings provide direct evidence that governance uncertainty is being reported by anatomy educators themselves rather than inferred on their behalf.

Anatomy presents a combination of governance questions that take distinctive forms in the discipline. Five pressure points recur. Donor-derived material raises questions about whether particular images or data may legitimately enter an external GenAI system. Generated explanations and images require verification of anatomical accuracy in their structural and relational detail rather than evaluation based on fluency or visual plausibility alone. Learner-facing resources raise related but distinct questions about clinically relevant anatomical variation and demographic or phenotypic representation. GenAI involvement in assessment can introduce errors into processes that carry consequences for learners. Finally, provenance, permissions, material AI contribution, and accountable human ownership must remain sufficiently traceable to support audit and accountability. These five concerns - donor-derived material, anatomical fidelity, variation and representation, assessment, and provenance and disclosure - form the discipline-specific pressure points developed in this article.

Existing responsible-AI frameworks provide substantial foundations for addressing these questions. Health CARE-AI offers consensus-based guidance for responsible AI use across health professions education (HPE), research, and care, emphasizing professional and institutional accountability, human judgment, appropriate information use, equity, and transparency [5]. The National Institute of Standards and Technology (NIST) Generative Artificial Intelligence Profile provides cross-sector guidance for identifying and managing GenAI risks across the AI lifecycle [6]. FUTURE-AI provides lifecycle-oriented recommendations for trustworthy AI in healthcare, spanning design, validation, deployment, and monitoring [7]. The United Nations Educational, Scientific and Cultural Organization (UNESCO) provides human-centered guidance specifically for GenAI in education and research, including privacy, human agency, validation, and institutional policy [8], while the Organisation for Economic Co-operation and Development (OECD) AI Principles address human-centered values, transparency and explainability, robustness and safety, traceability, and accountability [9].

These instruments establish substantial cross-cutting principles, risk-management processes, lifecycle expectations, and accountability structures. They do not, however, resolve several discipline-specific determinations required in anatomy education: which donor-derived inputs may legitimately enter a GenAI system, what constitutes adequate verification of generated anatomy, how variation and representation should be evaluated, which assessment uses require heightened control, and which decisions should remain human-accountable. An anatomy-specific operational layer is therefore needed to translate cross-cutting guidance into discipline-specific decisions.

Relevant operational approaches also exist within HPE. Masters' AMEE Guide No. 158 addresses the ethical use of AI in HPE [10], while Gin et al. adapt the established concept of entrustment to AI and propose graduated levels of AI participation and human supervision for specific educational tasks [11]. These are important antecedents to the present proposal and mean that task-specific governance, graduated oversight, and retained human responsibility should not be presented as conceptually unique to the proposed framework.

The contribution proposed here is narrower and discipline-specific. Among the 71 sources retained for framework development, we did not identify a single framework integrating all five anatomy-specific pressure points into an operational approach linking use-case classification, authorization, verification, human accountability, and continuing monitoring. This observation is bounded by the review's search strategy and retained evidence set and should not be interpreted as evidence that no other relevant framework exists.

This article therefore proposes an anatomy-specific operational layer that builds on, rather than replaces, existing responsible-AI and HPE guidance. The defined educational use case - not the AI product in the abstract - is characterized by its purpose, audience, input and data class, consequence, scale, and degree of AI autonomy. These characteristics inform a proposed risk tier, after which relevant control gates, retained human decisions, and lifecycle monitoring are applied. The framework addresses GenAI uses initiated or authorized by educators and governed at course, departmental, or institutional level. Learner-originated uses outside those governed workflows, including independent upload of donor-derived material to consumer GenAI systems, fall outside the operational scope of the proposed framework, although institutions remain responsible for establishing policies governing such conduct. Learner-originated GenAI use is therefore identified as a priority for further governance research. The proposed framework is intended for local adaptation by anatomy programs and institutions. It has not been prospectively evaluated, and its tier thresholds have not been empirically calibrated. It is not a substitute for applicable law, donor-program policy, institutional governance, or regulatory compliance.

Review

Evidence identification and framework development

Design and Rationale

This article uses a structured, evidence-informed state-of-the-art review with framework development. It was not designed as a systematic or scoping review, and no claim of compliance with the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) is made. The evidence base spans empirical anatomy studies, reviews, transferable HPE research, professional and policy guidance, and conceptual or ethical scholarship. These sources serve different evidentiary functions and cannot appropriately be pooled or appraised using a single methodological instrument without obscuring those differences [12,13].

The purpose of the review was therefore not to estimate a pooled effect or provide an exhaustive map of all potentially relevant literature. Rather, it was to identify evidence relevant to recurrent governance problems in anatomy education, characterize the directness and limitations of that evidence, and use the resulting synthesis to develop an operational framework while distinguishing the evidence from the author-developed governance components informed by it.

Evidence Sources, Search Scope, and Dates

Evidence identification was organized through six prespecified streams. PubMed/MEDLINE provided the structured bibliographic component. Google Scholar was used as a supplementary discovery and verification source. Professional, institutional, publisher, and DOI sources were consulted where relevant. The original documentation preserves the purpose, search date, concept blocks, and discovery or verification sources for each stream. The historical record does not preserve a separate formal backward- or forward-citation-chasing procedure or source-specific yield; none is reconstructed here. The six prespecified evidence streams, their search dates, purposes, concept blocks, and discovery or verification sources are summarized in Table *1*.

Stream	Search date	Prespecified purpose	Concept blocks used	Discovery and verification sources
S1 - Core anatomy and GenAI	Aug 7, 2026	Identify anatomy-specific GenAI and AI sources directly addressing or materially informing governance, ethics, oversight, reliability, and implementation.	Anatomy education; generative AI/large language models; ethics; governance; reliability; oversight; implementation	PubMed/MEDLINE; Google Scholar; publisher and DOI verification; professional and institutional sources where relevant
S2 - Donor-derived materials and digital governance	Aug 7, 2026	Identify anatomy and human-remains evidence on donor consent, provenance, digital images, derivatives, privacy, ownership, and external dissemination.	Body donation; cadaveric/human-tissue images; consent; dignity; provenance; digitization; privacy; ownership; dissemination	PubMed/MEDLINE; Google Scholar; IFAA; Human Tissue Authority; publisher and DOI verification
S3 - Anatomical accuracy and verification	Aug 7, 2026	Identify direct anatomy evidence on factual accuracy, hallucination, multimodal recognition, citation reliability, model comparison, and verification.	Anatomy; factual accuracy; hallucination; image recognition; multimodal AI; citations/references; model comparison; verification	PubMed/MEDLINE; Google Scholar; publisher and DOI verification
S4 - Visual fidelity and variation/representation	Aug 7, 2026	Determine whether visual fidelity and representation require separate but linked controls and identify evidence concerning anatomical variation, representation, and bias.	AI-generated images; medical/anatomical illustration; anatomical fidelity; variation; representation; demographic bias	PubMed/MEDLINE; Google Scholar; publisher and DOI verification
S5 - Assessment	Aug 7, 2026	Define controls for formative versus summative use, AI-generated items, automated scoring, practical/OSPE applications, psychometrics, feedback, and human oversight.	Anatomy assessment; MCQ/item generation; OSPE/practical assessment; automated scoring; psychometrics; feedback; human oversight	PubMed/MEDLINE; Google Scholar; publisher and DOI verification
S6 - Governance and accountability	Aug 8, 2026	Define accountability, authorization, documentation/disclosure, monitoring, re-evaluation, incident response, withdrawal, and revalidation.	Responsible AI; governance; accountability; authorization; monitoring; incident response; revalidation; disclosure; institutional policy	PubMed/MEDLINE; Google Scholar; NIST; UNESCO; AAMC; OECD; publisher and DOI verification

TABLE 1: Search-stream documentation

The table provides a structured, evidence-informed state-of-the-art review with framework development. The final substantive evidence set comprised 71 retained sources. GenAI-focused literature was emphasized from 2022 through August 8, 2026; older foundational anatomy and donor-ethics sources were retained where relevant. Exact historical query strings and complete per-stream retrieval, deduplication, screening, and exclusion counts were not prospectively archived and have not been retrospectively reconstructed. Accordingly, this table documents the historical search process at stream level and should not be interpreted as a PRISMA-style record flow. No separate formal citation-chasing stream or yield is reconstructed from the historical record. The organizational sources named in the table are those preserved in the historical stream documentation. WHO and ICMJE documents are retained in the 71-source evidence matrix, but the surviving record does not preserve a source-specific retrieval route for those documents; no such route is reconstructed here.

AI: artificial intelligence; MCQ: multiple-choice question; OECD: Organisation for Economic Co-operation and Development; OSPE: objective structured practical examination; NIST: National Institute of Standards and Technology; UNESCO: United Nations Educational, Scientific and Cultural Organization; PRISMA: Preferred Reporting Items for Systematic Reviews and Meta-Analyses; WHO: World Health Organization; ICMJE: International Committee of Medical Journal Editors; IFAA: International Federation of Associations of Anatomists; AAMC: Association of American Medical Colleges

GenAI-focused literature was emphasized from 2022 through 8 August 2026. Older sources were retained where required to address foundational donor ethics, human-tissue imagery, anatomy education, or professional responsibilities that predate contemporary GenAI.

The primary structured bibliographic search was limited principally to PubMed/MEDLINE. A prospectively defined broader database set including Scopus, Web of Science, Embase, ERIC, IEEE Xplore, or ACM Digital Library was not searched. This limitation is acknowledged explicitly below.

Exact historical search strings were not prospectively archived, and complete retrieval, deduplication, screening, and exclusion counts cannot be reconstructed reliably. Those values are therefore not estimated or inserted retrospectively. The evidence-identification process is auditable at the level of the six documented streams and retained source matrix, but it is not independently reproducible as a systematic search.

Regulatory instruments discussed later in the article were verified against authoritative legal sources during manuscript revision and are treated as contextual legal references rather than being retrospectively inserted into the historical 71-source framework-development set.

Eligibility and Evidence Inclusion

Sources were retained when they made a substantive contribution to an anatomy-relevant governance question. Eligible evidence therefore included direct anatomy empirical research and reviews; transferable HPE evidence where anatomy-specific evidence was limited; professional, consensus, policy, regulatory, or standards-based guidance; and conceptual or ethical scholarship relevant to donor stewardship, provenance, accountability, or related governance boundaries.

The review focuses on generative AI. Non-generative AI evidence was retained selectively where it addressed a directly transferable governance problem, particularly automated assessment, human oversight, supervision, or accountability. Such evidence is identified as transferable rather than presented as evidence of GenAI performance.

Older literature concerning body donation, cadaveric imagery, consent, dignity, provenance, privacy, and stewardship remained relevant because these responsibilities predate contemporary GenAI.

The final eight governance domains were not used as a retrospective eligibility criterion. Sources were first retained because they substantively informed the review question and were subsequently coded to a principal governance domain or as cross-cutting. This distinction avoids defining eligibility by the framework that the retained evidence was itself being used to develop.

To clarify the scope of the final evidence set without retrospectively reconstructing a PRISMA-style screening history, the revised synthesis treats as outside scope: reports concerned only with clinical-care or research AI when no transferable educational-governance implication was identified; conference abstracts; and preprints subsequently superseded by a peer-reviewed report, for which the published version is used. Non-generative AI was not excluded categorically when it provided directly transferable governance evidence. These statements describe the scope of the final synthesis and are not presented as prospectively archived exclusion rules or as a basis for reconstructing exclusion counts.

Methodological, definitional, background, and regulatory-context citations used to explain the review design or contemporary legal environment are also distinguished from the 71 substantive framework-development sources.

Selection, Coding, and Evidence Classification

The final retained evidence base comprised 71 substantive sources: 25 direct anatomy sources, 24 transferable HPE sources, 13 professional or normative sources, and 9 conceptual or ethical works. Each retained source was assigned an evidence-basis category and a principal governance domain or was coded as cross-cutting. The complete source-level coding is provided in the Appendices.

Screening and source coding were performed by one author (AM). No independent double-screening or duplicate source coding was undertaken, and no formal inter-rater reliability assessment was performed. The categorizations should therefore be interpreted as a transparent single-rater synthesis rather than independently replicated classifications.

The working evidence matrix contains fields for relevance, source credibility, and framework utility. Their exact historical operational role cannot now be reconstructed with sufficient confidence. They are therefore not used in the revised article as inclusion or exclusion criteria, evidence weights, certainty ratings, or determinants of recommendation strength.

Appraisal and Evidentiary Strength

No single formal risk-of-bias instrument was applied across the retained set, and no GRADE-style certainty assessment was undertaken. The heterogeneous source types were instead appraised descriptively according to their design, setting, task, sample or output set where applicable, and relevance to the governance question.

Among the 25 direct anatomy sources, 20 were original empirical studies, two systematic reviews, one scoping review, one critical review, and one other educational source. The empirical literature included bounded model-output evaluations, expert-rating studies, authentic learner cohorts, psychometric studies, content analyses, automated assessment applications, and repeated comparisons across model versions.

Study scale varied substantially, ranging from bounded model-output evaluations to authentic learner cohorts of several hundred students. Preregistration status was not systematically captured or verified in the evidence matrix and therefore was not used as an appraisal criterion.

The evidence is unevenly distributed across governance domains. Direct anatomy evidence is concentrated most heavily in anatomical fidelity and assessment; the representation domain relies on transferable HPE evidence, while human oversight and institutional lifecycle governance are also supported predominantly by transferable and professional sources.

These evidence profiles are descriptive rather than quantitative certainty judgments. Proposed controls should not be interpreted as having been prospectively validated simply because the failure mode they address is documented empirically.

Five discipline-specific pressure points

Donor-Derived Inputs Retain Stewardship Obligations After Digitization

Photographs, videos, scans, and derived representations of donated bodies remain connected to the consent arrangements, institutional responsibilities, and stewardship expectations under which the material was obtained. Professional guidance represented within the retained evidence emphasizes appropriate consent or permission, protection against identification, controlled distribution, secure storage, provenance, dignity, and institutional oversight [14-18].

The literature does not establish one universal rule governing every use of donor-derived imagery. The framework therefore does not create a universal consent rule. Before donor-derived material is processed through GenAI, the responsible program should establish its provenance, applicable consent or permission scope, intended purpose, relevant law and institutional policy, and the nature of the proposed platform exposure.

De-identification also requires more than removal of names or electronic metadata. Potentially identifying features of donor-derived images may include facial structure, tattoos, scars, surgical hardware, dentition, unusual pathology, visible labels, or other distinctive physical characteristics. The proposed review of these features is an author-developed operational safeguard, not a validated donor-image de-identification instrument.

Where removal of potentially identifying characteristics would materially compromise educational value, de-identification should not simply be assumed to resolve the governance problem; the use should instead be escalated to the appropriate donor-program or institutional authority.

A conservative default is also proposed for external or commercial platforms: donor-derived imagery should not be uploaded to or processed by an external commercial GenAI system unless institutional authorization has been obtained and relevant contractual and technical safeguards have been reviewed. Zero-retention or no-training commitments may reduce particular risks but do not independently establish permission for donor-image processing.

Relevant platform review may include data retention, use for model training or service improvement, secondary processing, subcontractor or third-party access, storage arrangements, deletion provisions, security controls, and contractual responsibility. Such review should involve the relevant anatomy or donor-program authority together with appropriate institutional privacy, information-security, legal, compliance, or equivalent expertise rather than an individual educator acting alone.

Independent learner use of consumer GenAI outside educator- or institution-authorized workflows is outside the operational tiering scope of this framework. Institutions should nevertheless address learner photography, upload, transformation, or circulation of donor-derived material through applicable laboratory, donor, professionalism, privacy, and AI-use policies. Learner-originated use remains an important research and policy priority.

Anatomical Fidelity Is Structural and Relational, Not Stylistic

Anatomical correctness depends on structure identity, position, course, attachment, branching, laterality, orientation, and spatial relationships. A fluent explanation or visually persuasive image may therefore remain anatomically incorrect.

Direct anatomy studies demonstrate task-, model-, modality-, version-, and workflow-dependent limitations in generated anatomical content, image interpretation, and question answering [19-25]. The governance implication is that the specific output requires verification rather than reliance on model reputation, aggregate benchmark performance, fluency, or photorealism.

Source reliability is a separate fidelity problem. Anatomy-specific evaluation has documented hallucinated, inaccurate, or unsupported references [23]. A bibliographically plausible citation supplied by a GenAI system should therefore not be accepted without locating the underlying publication and confirming that it supports the associated claim.

Recognition should also be distinguished from subsequent reasoning. Where a task depends on identification of an anatomical image or structure before interpretation or explanation, an upstream recognition error should not be obscured by plausible downstream reasoning [24].

Model performance can also change over time. Comparisons across successive model versions show that aggregate improvement does not establish equal reliability across topics or cognitive levels [25]. Validation of one provider, model version, modality, prompt structure, retrieval system, or workflow should therefore not be generalized automatically to a materially different configuration.

The proposed control is qualified anatomical verification before authoritative learner-facing use. Relevant structural claims should be checked against appropriate independent authoritative sources. Visual content should be examined for identity, laterality, orientation, spatial relationships, omissions, additions, labels, and clinically meaningful departures from accepted anatomy. Final acceptance of authoritative anatomy remains human-accountable.

Anatomical Variation and Demographic Representation Are Related but Distinct

Two related but different issues should be distinguished: clinically relevant anatomical variation and demographic or phenotypic representation.

Clinically relevant anatomical variation concerns meaningful departures from the most commonly taught arrangement that may affect anatomical understanding or clinical reasoning, including vascular, biliary, urinary, neural, and other variations. Direct evidence evaluating whether GenAI-generated educational anatomy reliably represents such variants remains very limited. The framework should therefore not claim that GenAI systematically omits clinically relevant variants. Instead, the absence of direct evidence itself constitutes an important research gap.

Demographic and phenotypic representation concerns whether generated visual material unduly narrows the range of bodies represented or reproduces stereotyped patterns. The retained evidence supporting this concern is principally transferable from medical-education and clinical image-generation studies. Alon et al.'s study, for example, is a systematic literature review rather than an individual image-generation experiment [26].

Learner-facing visual resources may therefore require two different forms of review. Individual images should satisfy anatomical-fidelity requirements, while sets of images may require review for unjustifiably narrow demographic or phenotypic representation. Where clinically relevant anatomical variation is part of the educational objective, important variants should be considered explicitly rather than treating a single configuration as universal.

The representation component of this proposed gate is informed predominantly by transferable empirical evidence. The anatomical-variation component is an evidence-informed precautionary safeguard in an area where direct anatomy-specific GenAI evidence remains sparse.

Assessment Converts Content Error Into Consequential Judgment

GenAI can assist assessment-item development, but item generation does not itself establish validity. Appropriate evaluation may include anatomical and factual accuracy, alignment with learning objectives and blueprint, intended cognitive level, item-writing quality, fairness, and post-use psychometric performance.

Kıyak et al. evaluated expert-reviewed, department-approved AI-generated anatomy questions administered to 502 second-year medical students, reporting usable item difficulty and discrimination characteristics [27].

Lopiano et al. evaluated ChatGPT-4 in gross anatomy and embryology. Initial prompting generated 300 questions, with 80.0% judged accurate and 7.7% higher order; modified prompting generated a further 210 questions, with 70.5% judged accurate and 41.0% higher order [28]. This represents a trade-off observed in one study, rather than a general law that increasing higher-order generation necessarily reduces accuracy.

Tekin et al. generated 240 anatomy questions using GPT-4, selected 30 for administration to 280 students, and reported acceptable internal consistency, while expert agreement with GPT-4's cognitive-level classification was only 26.6% for the administered questions [29]. Such findings illustrate why satisfactory psychometrics cannot substitute for content and cognitive-level review.

Performance on image-dependent practical tasks is also variable. Meo et al. evaluated Google Bard on morphological, histopathological, and radiological image-identification tasks using an OSPE-type format, providing transferable evidence that image-dependent assessment performance should not be inferred from text-only performance [30].

Automated assessment evidence also requires modality-specific interpretation. Bernard et al. evaluated decision-tree methods rather than GenAI for grading a 54-question anatomy OSPE using data from 368 learners and reported high local grading accuracy [31]. This is relevant anatomy evidence concerning automated assessment governance but should not be represented as direct GenAI evidence.

The proposed framework therefore allows GenAI to assist with question drafting, formative material, feedback, and defined components of assessment workflows where appropriate human review and task-specific validation are present. Summative or consequential uses warrant stronger content and blueprint review, quality assurance (QA), psychometric or performance monitoring where appropriate, and a named accountable human decision-maker.

The proposed boundary that final grading, progression, pass/fail, disciplinary, or contested high-stakes judgments remain human-accountable is an evidence-informed precautionary governance position, not an empirically calibrated GenAI threshold.

Provenance, Intellectual Property, and Disclosure Support Traceability and Accountability

A fifth pressure point concerns whether the origin, transformation history, permission status, material AI contribution, and accountable ownership of educational material remain sufficiently traceable.

Provenance concerns the origin and history of material. Permission and intellectual-property considerations concern whether content may appropriately or lawfully be reused, transformed, or redistributed. Disclosure concerns whether material GenAI involvement is communicated proportionately and whether accountable human ownership remains identifiable.

Ho et al. examined 239 English-language YouTube videos relevant to digital anatomy education and identified a substantial gap between ethical expectations and observed reporting of provenance and consent [32]. That finding should not be generalized to all anatomy resources, but it demonstrates that loss of traceability in digital educational practice is an observed problem.

Cornwall et al. discuss legal and ethical questions associated with using existing anatomical illustrations to create new illustrations, including permission, attribution, and derivative use [33]. Similar questions arise when GenAI systems transform or build upon existing educational resources.

The framework therefore proposes four linked requirements: document source and provenance when material to the use case; verify applicable permissions and attribution; disclose meaningful AI involvement proportionately; and identify a responsible human approver.

These recommendations should not be converted into universal copyright conclusions. Copyright, licensing, derivative-use rights, ownership, and contractual requirements vary by jurisdiction, source material, platform terms, and publication context.

Disclosure should likewise be proportionate. Reusable learner-facing content, formal instructional exemplars, publications, or consequential assessment workflows may warrant substantially more explicit disclosure than private brainstorming or routine wording assistance.

Cureus
Part of SPRINGER NATURE

Framework development and derivation of the governance domains

The final framework contains eight governance domains, but the five anatomy-specific pressure points do not correspond to eight independent anatomy-specific failure modes.

Five domains arise directly from the discipline-specific pressure points: donor-derived material maps to Domain 2 (input and donor-derived material governance); anatomical fidelity to Domain 3 (anatomical fidelity and source verification); anatomical variation and representation to Domain 4 (variation, bias, and representation); assessment to Domain 5 (assessment validity and integrity); and provenance and disclosure to Domain 6 (provenance, intellectual property, and disclosure).

Three additional domains provide the operational governance architecture. Domain 1 - use-case and risk classification - is a cross-cutting author-developed classification layer. Only one retained source was coded principally to this domain, although several retained governance sources provide cross-cutting support. Domain 7 - human oversight and accountability - and Domain 8 - institutional authorization, monitoring, and incident response - provide cross-cutting governance mechanisms supported predominantly by transferable HPE and professional or normative evidence.

The resulting architecture is therefore best understood as five anatomy-specific problem domains plus one cross-cutting classification layer plus two cross-cutting governance layers. The evidence profile, synthesis rationale, minimum proposed control, status, and principal limitation for each of the eight governance domains are summarized in Table 2.

Governance domain	Evidence profile	Synthesis rationale	Minimum proposed control	Status	Principal limitation
1. Use-case and risk classification	One principal direct-anatomy source; additional cross-cutting governance support	Governance should be tied to the defined use case rather than to the AI product alone.	Define purpose, audience, input/data class, consequence, scale, and permitted autonomy before authorization.	P	Classification construct is author-developed and unvalidated.
2. Input and donor-derived material governance	Eight sources coded principally to this domain: four professional/normative, three conceptual/ethical, one direct anatomy	Digitization does not extinguish consent, dignity, provenance, privacy, or stewardship considerations.	Establish provenance and permission; review identifying features and platform authorization before processing.	G/P	Law, consent, and institutional requirements differ; operational safeguards are not validated instruments.
3. Anatomical fidelity and source verification	16 sources coded principally to this domain: 12 direct anatomy, 4 transferable HPE	Plausibility is not anatomical truth; performance varies by model, modality, version, and task.	Qualified verification of learner-facing/authoritative content and generated sources.	E/P	Most studies are bounded evaluations and should not be generalized across future systems.
4. Variation, bias, and representation	Seven sources coded principally to this domain: all transferable HPE	Demographic/phenotypic representation problems are documented; direct evidence concerning clinically relevant variants remains limited.	Review visual sets for representation; review important anatomical variants where educationally relevant.	E/P*	Empirical support is mainly transferable; representation and anatomical variation have different evidentiary bases.
5. Assessment validity and integrity	10 sources coded principally to this domain: 6 direct anatomy, 4 transferable HPE	GenAI may assist assessment, but satisfactory local psychometrics do not establish full validity or autonomous scoring safety.	Apply content/blueprint review, task validation, human approval, and appropriate post-use monitoring.	E/P	Autonomous consequential GenAI assessment remains insufficiently evaluated.
6. Provenance, intellectual property, and disclosure	Eight sources coded principally to this domain: three conceptual/ethical, three professional/normative, two direct anatomy	AI-assisted resources raise linked questions of provenance, permission, attribution, transformation, disclosure, and accountability.	Document material provenance/permissions; proportionately disclose GenAI involvement; identify accountable approval.	E/G/P	Legal implications vary by jurisdiction, contract, platform, and source material.
7. Human oversight and accountability	Six sources coded principally to this domain: four transferable HPE, two professional/normative	Supervision and accountability should vary by task and consequence.	Assign accountable owner; define permitted autonomy; retain challenge/override and human responsibility for consequential decisions.	E/G/P	Anatomy-specific implementation evidence is sparse.
8. Institutional authorization, monitoring, and incident response	Seven sources coded principally to this domain: four transferable HPE, three professional/normative	Authorization should remain conditional and lifecycle-based.	Maintain proportionate records, review triggers, monitoring, incident response, and revalidation.	E/G/P	Optimal monitoring frequency, burden, restart criteria, and outcome benefit remain unvalidated in anatomy.

TABLE 2: Evidence landscape across eight governance domains

"Principal" refers to the governance domain to which a retained source was primarily coded; eight additional sources were coded as cross-cutting. Status codes describe evidentiary provenance rather than certainty: E = empirical evidence documents the relevant failure mode or rationale; G = professional or normative guidance supports the principle; P = the operational rule, threshold, or boundary is author-proposed. Combined codes indicate mixed provenance. E does not mean that the proposed control itself has been prospectively validated. *For Domain 4, the empirical evidence is predominantly transferable rather than anatomy-specific.

HPE: health professions education

The framework also builds on existing HPE scholarship concerning responsible AI and graduated supervision. Masters' AMEE Guide No. 158 provides an established ethical framework for AI use in HPE [10], while Gin et al. apply entrustment concepts to AI and emphasize graduated responsibility and supervision [11]. The present framework therefore does not claim novelty for graduated supervision, entrustment, or human accountability themselves. Its proposed contribution is to embed such oversight within an anatomy-specific architecture that also addresses donor-derived inputs, anatomical fidelity, variation and representation, assessment, provenance, and lifecycle governance.

A proposed risk-tiered framework for anatomy education

Risk Attaches to the Defined Use Case, Not the AI Product

The proposed unit of governance is the defined educational use case, not an AI product in the abstract. The same general-purpose model may be used for private brainstorming, development of a reusable learner handout, creation of an anatomical illustration, processing of donor-derived material, or contribution to summative assessment. Those uses do not present equivalent governance requirements.

Each defined use case is characterized according to six dimensions: purpose; audience; input and data class; consequence; scale; and permitted AI autonomy.

These dimensions form an author-developed, evidence-informed classification construct. They do not constitute a validated anatomy risk score or empirically established taxonomy.

Four Proposed Levels of Authorization and Control

Four tiers are proposed: Low, Moderate, High, and Restricted/Avoid. The highest triggered criterion determines the provisional tier as a conservative design rule. The rule is intended to prevent a high-consequence feature from being diluted by otherwise low-risk aspects of a workflow, but it is entirely author-proposed and may result in over-classification. Its inter-rater reliability, sensitivity to actual harm, and administrative burden require prospective evaluation. The four proposed risk tiers, their triggers, minimum controls, authorization requirements, monitoring and escalation provisions, and evidentiary status are summarized in Table 3.

Risk tier	Primary triggers and typical uses	Minimum proposed controls	Authorization and permitted AI role	Monitoring/escalation	Status
Low	Private, nonsensitive, nonconsequential support. No donor-derived/confidential/restricted input and no direct learner consequence. Examples: brainstorming, outlining, wording alternatives, nonconsequential administrative support.	Follow applicable institutional AI rules; educator reviews any material before it becomes educational content.	Individual educator within policy. AI may assist or draft.	Minimal documentation where locally required. Reclassify if material becomes learner-facing/reusable, restricted data are introduced, or consequence increases.	P
Moderate	Learner-facing or reusable material from nonsensitive inputs with low consequence. Routine course/cohort distribution and normal LMS placement may remain Moderate when properly reviewed.	Qualified anatomy review; source verification; structural fidelity checks; visual-set representation review; variation review where educationally relevant; proportionate disclosure.	Course/module lead or locally defined process. A0/A1 autonomy ordinarily appropriate; each learner-facing output approved before release.	Record responsible owner, use case, model/provider where known, verifier, limitations, and relevant approvals. Reassess after material change, recurrent error, restricted input, increased consequence, High scale, or High autonomy.	E/P
High	Authorized donor-derived/restricted processing; summative or otherwise consequential assessment; formal AI-generated anatomical exemplar; High scale; or High autonomy.	Task-specific validation; structured expert QA; donor/privacy/security review where applicable; assessment review/psychometrics where relevant; named accountability.	Department plus relevant assessment, donor, privacy/security, data-governance, or institutional authority. Strong supervision required.	Maintain formal use record and use-case-relevant monitoring. Reassess after material change, incident, recurrent error, or deterioration in validated performance.	E/G/P
Restricted/Avoid	Unauthorized restricted-input processing; A3 autonomous consequential learner decisions; autonomous/unverified release of generated anatomy as authoritative ground truth; use outside mandatory authorization conditions.	Correct the authorization, data, validation, or autonomy problem before proceeding; post-hoc checking alone is insufficient.	Do not proceed under ordinary authorization. Any exception requires appropriate institutional/legal/ethical authority.	Contain breaches; preserve relevant records; correct/withdraw outputs or decisions; investigate control failure; revalidate before any authorized restart.	E/G/P

TABLE 3: Proposed risk-tier authorization matrix for generative AI use in anatomy education

The four tiers and their boundaries are author-developed and have not been empirically calibrated. The highest triggered criterion determines the provisional tier as a conservative design rule. Status codes are defined in the Table 2 legend. Local law, donor-program policy, institutional rules, assessment regulations, procurement requirements, and platform terms remain controlling.

AI: artificial intelligence; QA: quality assurance; LMS: learning management system

Technical integration is not an independent classification dimension or High-tier trigger. It matters only when it changes an existing dimension, for example by introducing restricted inputs, increasing autonomy or scale, automatically releasing content, or linking GenAI directly to a consequential decision.

Operational Definitions of Key Threshold Terms

To support consistent classification, Table 4 defines the threshold terms used in the tiering framework and provides locally adjustable operational anchors.

Term	Proposed operational definition	Operational anchor	Status	Principal limitation
High scale	Deployment in which breadth, persistence, or volume materially increases potential error impact or makes effective oversight more difficult.	Examples include repeated reuse across multiple cohorts/programs without renewed review; interinstitutional/public dissemination; substantial-volume automated generation/release. Normal single-course/cohort or routine LMS use does not independently constitute a High scale.	P	Qualitative threshold is intentionally local and uncalibrated.
AI autonomy	Extent to which AI output reaches learners or affects educational processes without qualified human review of that specific instance before effect.	A0: assistive. A1: supervised generation with instance-level prerelease approval. A2: sampled/exception-based oversight; some outputs lack instance-level prerelease review. A3: autonomous action without instance-level prerelease review. A2/A3 = High autonomy. A3 is Restricted/Avoid when releasing authoritative anatomy or making consequential learner decisions.	E/G/P	A0-A3 ladder is author-developed but conceptually aligned with HPE supervision/entrustment; consequential A3 boundary is precautionary.
Material model or workflow change	Any change not covered by previous authorization/validation that could plausibly alter quality, safety, validity, or governance assumptions.	Provider/model/version, modality/tool access, retrieval corpus, system configuration, fine-tuning, data-handling terms, or changes to audience, inputs, consequence, scale, or autonomy.	E/P	Evidence supports version/task dependence; trigger definition is proposed.
Significant incident	Event materially breaching authorization conditions or causing/credibly risking consequential harm to learners, donor interests, privacy/confidentiality, assessment integrity, or another protected interest.	Examples: restricted data leaving scope; incorrect authoritative anatomy reaching learners; consequential assessment effect; material consent/privacy/security/permission breach. Error intercepted before effect may be treated as a near miss.	G/P	Operational incident boundary is proposed and informed by lifecycle-governance principles.
Qualified anatomical verification	Review by a person competent to teach, examine, or make expert judgments about the anatomy at the intended level, using independent authoritative sources.	Check identity/nomenclature; course/attachments/relations; branching/segmentation/count; laterality/orientation; spatial relationships; labels; relevant variation; generated references. Fluency, photorealism, model reputation, or memory alone are insufficient.	E/P	Need is supported by direct anatomical evidence; the specific checklist is proposed.

TABLE 4: Proposed operational definitions of threshold terms

Operational anchors are intended for local adaptation and are not validated cut-points. Status codes are defined in the Table 2 legend.

HPE: health professions education; LMS: learning management system

At authorization, each Moderate- or High-risk use should have a documented planned review date or review interval, together with defined events that trigger earlier reassessment. The interval should reflect consequence, persistence, model stability, and the institution's ability to detect material changes. No universal numerical interval is proposed because the available evidence does not establish an empirically justified frequency.

Five control gates and four proposed non-delegable human decisions

Tier assignment determines governance intensity but does not replace the substantive checks required by the use case. Before a Moderate- or High-risk use is released or operated, it should pass each applicable control gate. A failed gate should lead to correction, suspension, or escalation rather than reassignment to a lower tier.

Gate 1: Inputs, provenance, and platform suitability. Establish provenance, applicable permission or consent, relevant de-identification considerations, and platform authorization before donor-derived, confidential, identifiable, or otherwise restricted material enters a workflow.

Gate 2: Anatomical fidelity and source verification. Verify authoritative or learner-facing anatomy, including relevant structural relationships, labels, recognition, and generated references. The need for verification has substantial direct anatomy support; the exact proposed checklist is author-developed.

Gate 3: Variation and representation. Review demographic and phenotypic representation at image-set level where applicable. Where clinically relevant anatomical variation is part of the educational objective, assess important variants separately. Evidence for the representation component is mainly transferable; direct evidence for GenAI representation of anatomical variants remains limited.

Gate 4: Assessment validity and integrity. Where GenAI contributes to assessment, apply content and blueprint review, cognitive-level review where relevant, task-specific validation, quality and fairness checks, human approval, and proportionate post-use monitoring.

Gate 5: Disclosure and accountability. Document meaningful AI involvement proportionately, maintain appropriate provenance and permissions, identify an accountable human owner or approver, and preserve suitable routes for challenge, correction, or override.

Four decisions are proposed as non-delegable human decisions: (1) final acceptance of authoritative learner-facing anatomical claims, labels, images, or formal exemplars, (2) authorization to process donor-derived, confidential, identifiable, or otherwise restricted material, particularly beyond established permission, (3) final summative, progression, pass/fail, disciplinary, or contested consequential learner judgments, and (4) authorization to restart a suspended High-risk use following a significant incident. These are governance proposals informed to differing degrees by direct evidence, professional guidance, transferable evidence, and precautionary reasoning; they should not be described as empirically validated universal rules. Table *5* maps the five pressure points and major framework components to their primary governance domains, the nature of the supporting evidence, evidentiary status, and principal limitations.

Framework component	Operational statement	Primary domain(s)	Nature of evidence	Status	Principal limitation
Donor pressure point	Donor material retains provenance, consent/permission, dignity, privacy, and stewardship obligations after digitization.	2	Professional/ethical + limited direct anatomy	G/P	Operational controls partly proposed.
Fidelity pressure point	Generated anatomy requires structural and source-level verification.	3	Strong direct anatomy evidence	E/P	Failure mode documented; exact verification procedure proposed.
Variation/representation pressure point	Anatomical variation and demographic/phenotypic representation are distinct but related concerns.	4	Representation evidence transferable; variation evidence sparse	E/P	Variation safeguard precautionary.
Assessment pressure point	Governance intensity rises when AI output contributes to consequential assessment.	5	Direct anatomy + transferable HPE	E/P	Exact nondelegation threshold precautionary.
Provenance/disclosure pressure point	Origin, transformation, permission, AI involvement, and accountability should remain traceable.	6	Direct + professional + conceptual/ethical	E/G/P	Legal requirements jurisdiction-dependent.
Use-case classification	Classify purpose, audience, input/data class, consequence, scale, and autonomy.	1 + cross-cutting	Predominantly transferable/normative	P	Author-developed and unvalidated.
Human oversight	Assign human ownership and match supervision to consequence/autonomy.	7	Transferable HPE + professional	E/G/P	Direct anatomy implementation evidence limited.
Lifecycle governance	Authorization, monitoring, incident response, correction, and revalidation continue after initial approval.	8	Transferable HPE + professional	E/G/P	Effectiveness/burden unvalidated in anatomy.
Highest-trigger rule	Highest triggered criterion sets provisional tier.	1, 7, 8	Conservative author design	P	Not calibrated; possible over-classification.
Low tier	Private nonsensitive nonconsequential support remains educator-managed.	1, 7, 8	Normative/transferable	P	Boundary proposed.
Moderate tier	Reusable/learner-facing nonsensitive low-consequence content requires prerelease qualified review.	3, 4, 6, 7	Direct + transferable	E/P	Tier boundary proposed.
High tier	Donor/restricted processing, consequential assessment, formal anatomical exemplars, High scale, or High autonomy trigger stronger governance.	2, 3, 5, 7, 8	Mixed empirical/professional/transferable	E/G/P	Tier boundary proposed.
Restricted/Avoid tier	Unauthorized restricted processing, autonomous consequential decisions, or autonomous/unverified authoritative anatomy should not proceed ordinarily.	2, 3, 5, 7, 8	Normative/precautionary + empirical rationale	E/G/P	Not a universal legal prohibition.
Gate 1	Confirm provenance, permission, identifying features, and platform suitability.	2, 6	Professional/ethical	G/P	Principle guidance-supported; details proposed.
Gate 2	Verify structural anatomy, claims, images, labels, and references.	3	Direct anatomy	E/P	Empirical rationale; checklist proposed.
Gate 3	Review representation and relevant anatomical variation.	4	Mainly transferable	E/P	Variation component remains under-evidenced.
Gate 4	Apply assessment content/blueprint review, validation, human approval, and monitoring.	5	Direct + transferable	E/P	Exact workflow proposed.
Gate 5	Document AI contribution and named accountability; preserve challenge/override.	6, 7, 8	Professional/normative + transferable	E/G/P	Implementation context-dependent.
Proposed human decision 1	Final authoritative anatomy acceptance remains human.	3, 7	Empirical rationale + governance	E/P	Absolute rule proposed.
Proposed human decision 2	Restricted-input authorization remains human/institutional.	2, 7	Professional/ethical	G/P	Scope depends on consent/law/policy.
Proposed human decision 3	Final consequential learner judgments remain human-accountable.	5, 7	GenAI limitations + transferable assessment + governance	E/G/P	Precautionary boundary; not directly validated.
Proposed human decision 4	Restart after significant High-risk incident requires governance approval.	7, 8	Lifecycle governance	G/P	Restart criteria unvalidated.
Lifecycle loop	Govern inputs -> verify -> validate where applicable -> deploy -> monitor -> respond -> revalidate/reclassify.	2-8	Cross-cutting synthesis	E/G/P	Frequency, burden, and outcome benefit unknown.

TABLE 5: Framework-to-evidence crosswalk

The crosswalk distinguishes documented failure modes and professional principles from operational rules proposed in this framework. The four tiers, highest-trigger rule, A0-A3 autonomy ladder, control-gate procedures, and four proposed non-delegable decisions have not been prospectively validated. Status codes are defined in the Table 2 legend.

HPE: health professions education; AI: artificial intelligence

Reclassification, monitoring, incident response, and revalidation

Authorization is conditional and revisable rather than permanent. Reassessment should occur when one or more characteristics supporting the original tier or authorization materially change.

Potential triggers include introduction of donor-derived, confidential, identifiable, or otherwise restricted inputs; change in purpose or audience; transition from formative to summative use; material increase in scale; increased AI autonomy; changes in provider, model, modality, retrieval system, workflow, or data-handling conditions; recurrent material errors or near misses; deterioration of assessment performance where relevant; new relevant safety evidence; or a significant incident.

A material change does not automatically require prohibition or elevation to High risk. The affected characteristics and controls should be reassessed and the tier, validation requirements, and authorization reconsidered accordingly.

Monitoring should be proportionate to the actual use case. Moderate- and High-risk uses should ordinarily maintain enough information to reconstruct the approved use, accountable owner, relevant provider/model where known, tier, applicable gates, authorization, verification or validation, important limitations, material changes, incidents or near misses, and corrective action.

Monitoring should focus on relevant failure modes rather than generic surveillance. Depending on the use, these may include anatomical errors, reference failures, assessment performance problems, unauthorized exposure of restricted material, recurrent representational problems, or unexpected changes after model or workflow modification.

Scheduled review can complement but should not replace event-triggered review. Where providers do not make model updates sufficiently transparent, programs should use the locally documented review interval established at authorization, but no universal calendar interval is prescribed.

Incident response should also be proportionate. Where continuing operation may create material risk, actions may include temporary containment or suspension, preservation of relevant records, assessment of affected learners or donor interests, correction or withdrawal of affected material or decisions, review of the failed or absent control, modification of the workflow, and targeted revalidation.

An error intercepted before release may be treated as a near miss where documenting it provides useful information about control performance. Routine editing corrections need not automatically become formal incidents.

Revalidation should address what changed or failed rather than requiring repetition of every control automatically. A model change affecting anatomical images may require renewed fidelity testing, whereas a change in data-retention terms may instead require renewed privacy or donor-governance review. If a significant incident results in suspension of a High-risk use, restart should require documented approval by the responsible governance authority. Figure 1 presents the resulting anatomy-specific GenAI governance decision algorithm. Figure 2 summarizes the conditional lifecycle of authorization, input governance, output verification, assessment validation where applicable, deployment, monitoring, incident response, and revalidation or reclassification.

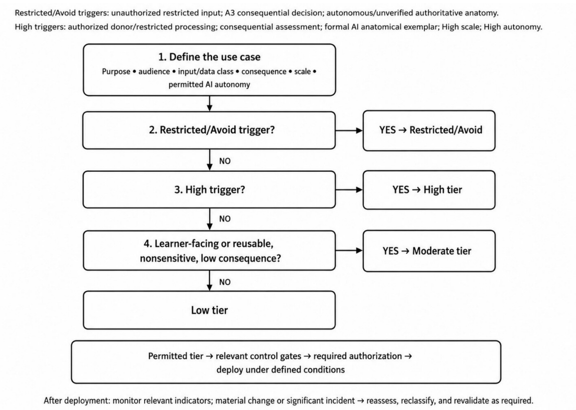

FIGURE 1: Anatomy-specific generative artificial intelligence governance decision algorithm

The defined educational use case is characterized by purpose, audience, input/data class, consequence, scale, and permitted AI autonomy. Restricted/Avoid triggers are considered first, followed by High-tier triggers; remaining learner-facing or reusable low-consequence uses are classified as Moderate, while private, nonsensitive, nonconsequential uses may remain Low. The highest triggered criterion determines the provisional tier. Relevant control gates, named human or institutional authorization, proportionate monitoring, incident response, and revalidation follow. The algorithm is an author-developed, evidence-informed proposal and has not been empirically calibrated or prospectively validated. Original schematic prepared using Matplotlib (version 3.10.8; Matplotlib Development Team, via NumFOCUS, Inc., Austin, TX, USA).

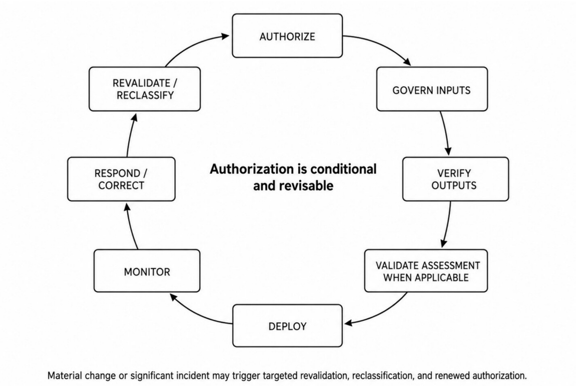

FIGURE 2: Anatomy-AI governance lifecycle control loop

Authorization is conditional and continuing rather than permanent. Approved uses remain subject to applicable input governance, output verification, assessment validation where relevant, proportionate monitoring, incident response, correction or withdrawal, and revalidation. A material change or significant incident may trigger reclassification and renewed authorization. No universal calendar-based revalidation interval is prescribed because the optimal review frequency for anatomy-education GenAI use has not been empirically established. Original schematic prepared using Matplotlib (version 3.10.8; Matplotlib Development Team, via NumFOCUS, Inc., Austin, TX, USA).

GenAI: generative artificial intelligence

Regulatory context: partial convergence with the EU AI Act

The proposed framework was not derived from a regulatory classification system. Nevertheless, the use of consequence as a reason for heightened governance shows partial convergence with the European Union Artificial Intelligence Act.

Annex III of Regulation (EU) 2024/1689 identifies several education-related AI systems as high-risk, including systems intended to determine access or admission, evaluate learning outcomes, assess the appropriate level of education a person receives or may access, and monitor or detect prohibited behavior during tests [34]. The statutory classification is subject to the Act's qualifications and should not be generalized to every educational AI use. The 2026 amending regulation did not remove education and vocational-training systems from the Annex III classification list.

This provides relevant external regulatory context for placing consequential assessment under heightened control, but it does not validate the proposed High tier. The EU Act regulates AI systems according to statutory criteria, whereas the present framework classifies defined educational use cases by purpose, audience, input/data class, consequence, scale, and autonomy. The two classification systems should therefore not be presented as equivalent.

Regulation (EU) 2026/1744 amended the application timetable. Under amended Article 113, Chapter III, Sections 1-3, with the exception specified in the Regulation, apply from December 2, 2027, for systems classified as high-risk pursuant to Article 6(2) and Annex III and from August 2, 2028, for systems classified pursuant to Article 6(1) and Annex I [35]. Regulation (EU) 2026/1744 was published in the Official Journal on July 24, 2026, and entered into force on July 27, 2026.

Article 50 should be considered separately. It establishes role- and use-specific transparency obligations rather than a universal disclosure requirement applying identically to every learner-facing educational GenAI use. The regulation's general application date is August 2, 2026, while Regulation (EU) 2026/1744 introduced a specific transition for providers of synthetic-content systems already placed on the market before that date: those providers must comply with Article 50(2) by December 2, 2026 [35]. Institutions must therefore determine applicable obligations according to their role, the AI system, intended purpose, type of generated content, and jurisdiction.

The convergence is consequently limited but relevant. EU law independently recognizes that certain AI uses affecting educational access, learning outcomes, and examination processes can warrant heightened governance. It does not resolve the anatomy-specific problems addressed by this framework, including donor-derived inputs and qualified verification of anatomical content. The proposed framework is not a legal-compliance instrument.

Putting the framework into practice

Implementation should be proportionate and should reuse existing educational and institutional structures wherever possible. The framework is not intended to create a new committee or extensive approval process for every interaction with GenAI.

A minimum viable implementation consists of a local inventory of material governed uses, named accountability, use-case classification and tier-matched authorization, concise documentation of applicable controls, and defined escalation, incident-response, and revalidation pathways.

Accountability should attach to the use case rather than merely the approved tool. Routine learner-facing content may be owned by a course or module lead; assessment use may fall within an existing assessment committee; donor-derived processing may require anatomy or body-donation authority; and restricted data or external systems may require privacy, information-security, procurement, legal, compliance, or institutional AI-governance involvement.

The purpose of the use record is accountability rather than exhaustive surveillance. It should contain enough information to reconstruct why a material governed use was permitted and what safeguards applied. Exhaustive prompt archiving is not required by this framework.

Lower-risk activity should ordinarily proceed under standing institutional rules rather than universal case-by-case preapproval. Escalation should occur when a use introduces restricted inputs, becomes consequential, reaches the High-scale threshold, moves to High autonomy, creates unresolved consent or permission issues, or experiences a material change or significant incident.

Technical integration requires escalation only when it changes another risk dimension. For example, putting a reviewed handout into an LMS does not independently create High risk; connecting a GenAI workflow directly to a grading system so that outputs affect scores without instance-level human review changes consequence and autonomy and therefore does.

Analyses of institutional GenAI guidance also suggest why a discipline-specific operational layer may be useful. McDonald et al. found substantial variation in institutional guidance and cautioned that faculty-facing requirements can become burdensome [36], while Jin et al. identified persistent needs for clearer institutional roles, communication, and implementation frameworks [37]. The purpose of the anatomy-specific layer is therefore to reduce local ambiguity while avoiding unnecessary duplication of existing governance.

Existing course leadership, assessment committees, donor-program governance, educational-technology teams, privacy and security offices, procurement processes, ethics or compliance mechanisms, and institutional AI-governance structures should be reused where they already address the relevant control.

The proposed implementation model is intended to make governance proportionate rather than unnecessarily burdensome, but that benefit remains untested. Administrative workload, classification consistency, feasibility across differently resourced programs, and effects on educational or safety outcomes require prospective evaluation.

Applying the framework to this article

A framework that emphasizes disclosure, verification, and human accountability should make its own use of GenAI transparent.

Under the proposed framework, this manuscript-preparation workflow is classified as Moderate risk because it produced reusable scholarly material from published or publicly accessible information, did not require donor-derived, learner-specific, or confidential inputs, and did not itself make a consequential decision concerning a learner.

The authors retained responsibility for source selection and interpretation, framework decisions, manuscript content, and conclusions. GenAI was used for language checking rather than as an author, evidentiary source, or autonomous scholarly decision-maker. A detailed artificial intelligence disclosure is provided in the Acknowledgments.

Limitations

Several limitations constrain both the evidence synthesis and the proposed framework.

First, this was a structured, evidence-informed state-of-the-art review with framework development rather than a systematic or scoping review. Although six search streams have documented purposes, concept blocks, dates, and discovery or verification sources, exact historical query strings and complete retrieval, deduplication, screening, and exclusion counts were not prospectively archived. Those data have not been reconstructed retrospectively. The retained evidence and stream-level process are therefore auditable, but the historical search cannot be reproduced as a systematic search.

Second, structured bibliographic searching was limited primarily to PubMed/MEDLINE, supplemented by

Google Scholar and targeted professional, institutional, publisher, and organizational sources. A broader prospectively defined set of databases such as Scopus, Web of Science, Embase, ERIC, IEEE Xplore, or ACM Digital Library was not searched. Relevant literature may therefore have been missed, particularly across education, computer science, ethics, law, policy, and rapidly emerging or non-indexed sources.

Third, no single formal risk-of-bias or certainty-of-evidence instrument was applied across the heterogeneous retained evidence. Descriptive appraisal and evidence classification were used instead. This limits standardized comparison of methodological quality. Preregistration status was not systematically captured or verified and was not used as an appraisal criterion.

Fourth, screening and source coding were performed by one author (AM), without independent double-screening, duplicate coding, or formal inter-rater reliability assessment. The domain and evidence-basis classifications therefore represent a transparent single-rater synthesis. Historical working relevance, credibility, and framework-utility fields are not used as evidence weights or certainty ratings because their exact historical operational role cannot now be reconstructed confidently.

Fifth, the evidence base is heterogeneous and unevenly distributed. Direct anatomy evidence is concentrated in anatomical fidelity and assessment. Donor governance relies mainly on professional guidance and ethical scholarship, whereas the evidence for representation, human oversight, and lifecycle governance is drawn predominantly from transferable HPE and professional sources. Many anatomy studies evaluate specific models, prompts, outputs, institutions, or time points rather than governance interventions. Their results should not be generalized automatically across models, versions, modalities, tasks, institutions, or future releases.

Sixth, important direct evidence gaps remain. Evidence concerning GenAI representation of clinically important anatomical variants, long-term representational effects, implementation of human-accountability models, anatomy-specific incident-response systems, and autonomous consequential GenAI assessment remains limited. Several proposed safeguards therefore rest on transferable evidence, professional guidance, ethical reasoning, or precautionary governance.

Seventh, the framework itself is author-developed and unvalidated. The six classification dimensions, four tiers, A0-A3 autonomy ladder, highest-trigger rule, operational definitions, five control gates, and four proposed non-delegable human decisions have not been prospectively tested as an integrated governance instrument. Tier boundaries are not empirically calibrated; inter-rater reliability is unknown; over- and under-classification have not been measured; and administrative burden, feasibility, educational impact, and reduction of governance failures remain unquantified.

Eighth, the framework addresses primarily educator- and institution-authorized workflows. Independent learner use of consumer GenAI outside such workflows is not operationally classified by the proposed tiers. Learner upload or transformation of donor-derived or otherwise restricted material nevertheless remains an important institutional-policy and research concern.

Finally, donor law, privacy requirements, intellectual property rules, assessment regulation, procurement requirements, contracts, platform terms, and AI regulation vary across jurisdictions and evolve over time. This framework is not a legal-compliance instrument and does not replace applicable law, donor agreements, institutional policy, professional standards, contractual requirements, or regulatory classifications.

Future research should therefore test whether independent educators classify identical use cases consistently; quantify administrative burden and feasibility; evaluate whether the proposed gates detect or prevent their targeted failures; examine the appropriateness of the highest-trigger and autonomy rules; and determine whether tier-matched governance measurably changes the quality, safety, fairness, or educational value of learner-facing material. The complete condensed source-level coding for the 71 retained substantive sources is presented in Table 6.

Ref.	Source (first authors, year)	Source type	Evidence basis	Principal governance domain
[2]	Lazarus et al. (2024)	Critical review	Conceptual/ethical scholarship	Cross-cutting/multiple
[3]	Cornwall et al. (2024)	Ethics analysis	Conceptual/ethical scholarship	2. Input and donor-derived material governance
[1]	Yi et al. (2026)	Critical review	Direct anatomy evidence	Cross-cutting/multiple
[5]	Sonnenberg et al. (2026)	Consensus statement	Transferable HPE evidence	Cross-cutting/multiple
[14]	International Federation of Associations of Anatomists (2023)	Professional guideline	Professional/normative guidance	2. Input and donor-derived material governance
[15]	Jones et al. (2026)	Professional guideline	Professional/normative guidance	Cross-cutting/multiple
[17]	Cornwall et al. (2016)	Ethics analysis	Conceptual/ethical scholarship	2. Input and donor-derived material governance
[16]	Iwanaga et al. (2025)	Editorial/perspective	Conceptual/ethical scholarship	6. Provenance, IP and disclosure
[19]	Eldesoqui et al. (2025)	Original empirical research	Direct anatomy evidence	3. Anatomical fidelity and source verification
[26]	Alon et al. (2026)	Systematic review	Transferable HPE evidence	4. Variation, bias and representation
[38]	Gupta et al. (2026)	Scoping review	Transferable HPE evidence	3. Anatomical fidelity and source verification
[39]	Agarwal et al. (2026)	Scoping review	Transferable HPE evidence	5. Assessment validity and integrity
[40]	Che and Liu (2026)	Narrative review	Transferable HPE evidence	3. Anatomical fidelity and source verification
[41]	World Health Organization (2024)	Professional guideline	Professional/normative guidance	7. Human oversight and accountability
[42]	Abdellatif et al. (2022)	Narrative review	Conceptual/ethical scholarship	Cross-cutting/multiple
[43]	Joseph et al. (2025)	Systematic review	Direct anatomy evidence	Cross-cutting/multiple
[44]	Chytas et al. (2025)	Scoping review	Direct anatomy evidence	3. Anatomical fidelity and source verification
[45]	Erbek et al. (2026)	Systematic review	Direct anatomy evidence	3. Anatomical fidelity and source verification
[46]	Collins et al. (2024)	Original empirical research	Direct anatomy evidence	1. Use-case and risk classification

[47]	Temizsoy Korkmaz et al. (2025)	Original empirical research	Direct anatomy evidence	3. Anatomical fidelity and source verification
[48]	Topcu et al. (2026)	Original empirical research	Direct anatomy evidence	2. Input and donor-derived material governance
[49]	Morgan et al. (2026)	Other	Direct anatomy evidence	5. Assessment validity and integrity
[50]	Almamari et al. (2026)	Original empirical research	Direct anatomy evidence	Cross-cutting/multiple
[51]	Bolgova and Mavrych (2025)	Original empirical research	Direct anatomy evidence	3. Anatomical fidelity and source verification
[52]	Hennessy et al. (2020)	Professional guideline	Professional/normative guidance	6. Provenance, IP and disclosure
[18]	Human Tissue Authority (2023)	Professional guideline	Professional/normative guidance	2. Input and donor-derived material governance
[53]	Human Tissue Authority (2025)	Policy/framework document	Professional/normative guidance	2. Input and donor-derived material governance
[54]	Passalacqua et al. (2025)	Consensus statement	Professional/normative guidance	2. Input and donor-derived material governance
[55]	Keet and Kramer (2022)	Ethics analysis	Conceptual/ethical scholarship	Cross-cutting/multiple
[56]	Lottering et al. (2022)	Original empirical research	Direct anatomy evidence	6. Provenance, IP and disclosure
[33]	Cornwall et al. (2025)	Ethics analysis	Conceptual/ethical scholarship	6. Provenance, IP and disclosure
[32]	Ho et al. (2026)	Original empirical research	Direct anatomy evidence	6. Provenance, IP and disclosure
[57]	Rossetti et al. (2026)	Systematic review	Conceptual/ethical scholarship	6. Provenance, IP and disclosure
[58]	Kramer and Moxham (2026)	Editorial/perspective	Conceptual/ethical scholarship	2. Input & donor-derived material governance
[59]	Iwanaga et al. (2022)	Consensus statement	Professional/normative guidance	6. Provenance, IP and disclosure
[20]	Hyeamang et al. (2025)	Original empirical research	Direct anatomy evidence	3. Anatomical fidelity and source verification
[21]	Mavrych et al. (2025)	Original empirical research	Direct anatomy evidence	3. Anatomical fidelity and source verification
[60]	Al-Khater (2025)	Original empirical research	Direct anatomy evidence	3. Anatomical fidelity and source verification
[61]	Sivri et al. (2026)	Original empirical research	Direct anatomy evidence	3. Anatomical fidelity and source verification
[25]	Paslı et al. (2026)	Original empirical research	Direct anatomy evidence	3. Anatomical fidelity and source verification
[23]	Ülkir and Paslı (2026)	Original empirical research	Direct anatomy evidence	3. Anatomical fidelity and source verification
[24]	Sarantopoulos et al. (2026)	Original empirical research	Transferable HPE evidence	3. Anatomical fidelity and source verification
[22]	Bae et al. (2025)	Original empirical research	Direct anatomy evidence	3. Anatomical fidelity and source verification
[62]	Kumar et al. (2024)	Original empirical research	Transferable HPE evidence	3. Anatomical fidelity and source verification
[63]	Currie et al. (2024)	Original empirical research	Transferable HPE evidence	4. Variation, bias and representation
[64]	Bui et al. (2026)	Original empirical research	Transferable HPE evidence	4. Variation, bias and representation
[65]	Shaari et al. (2025)	Original empirical research	Transferable HPE evidence	4. Variation, bias and representation
[66]	Currie et al. (2025)	Original empirical research	Transferable HPE evidence	4. Variation, bias and representation
[67]	Ali et al. (2024)	Original empirical research	Transferable HPE evidence	4. Variation, bias and representation
[68]	Senft Everson et al. (2026)	Original empirical research	Transferable HPE evidence	4. Variation, bias and representation
[28]	Lopiano et al. (2026)	Original empirical research	Direct anatomy evidence	5. Assessment validity and integrity
[27]	Kıyak et al. (2025)	Original empirical research	Direct anatomy evidence	5. Assessment validity and integrity
[29]	Tekin et al. (2025)	Original empirical research	Direct anatomy evidence	5. Assessment validity and integrity
[31]	Bernard et al. (2024)	Original empirical research (non-generative/conventional AI: decision-tree automated grading)	Direct anatomy evidence; transferable to GenAI governance	5. Assessment validity and integrity
[69]	Mezil et al. (2026)	Original empirical research (non-generative/conventional AI)	Direct anatomy evidence; transferable to GenAI governance where applicable	5. Assessment validity and integrity
[70]	Al-Mashhadani et al. (2026)	Systematic review	Transferable HPE evidence	5. Assessment validity and integrity
[71]	Ahmed et al. (2025)	Original empirical research	Transferable HPE evidence	5. Assessment validity and integrity

[30]	Meo et al. (2024)	Original empirical research	Transferable HPE evidence	5. Assessment validity and integrity
[72]	Association of American Medical Colleges (2025)	Professional guideline	Professional/normative guidance	8. Institutional authorization, monitoring and incident response
[10]	Masters (2023)	Professional guideline	Transferable HPE evidence	7. Human oversight and accountability
[6]	Autio et al. (2024)	Policy/framework document	Professional/normative guidance	8. Institutional authorization, monitoring and incident response
[7]	Lekadir et al. (2025)	Consensus statement	Transferable HPE evidence	8. Institutional authorization, monitoring and incident response
[11]	Gin et al. (2025)	Policy/framework document	Transferable HPE evidence	7. Human oversight and accountability
[73]	Izquierdo-Condoy et al. (2026)	Narrative review	Transferable HPE evidence	8. Institutional authorization, monitoring and incident response
[9]	Organisation for Economic Co-operation and Development (2024)	Policy/framework document	Professional/normative guidance	7. Human oversight and accountability
[8]	Miao and Holmes (2023)	Professional guideline	Professional/normative guidance	8. Institutional authorization, monitoring and incident response
[37]	Jin et al. (2025)	Original empirical research	Transferable HPE evidence	7. Human oversight and accountability
[74]	Ganguly et al. (2025)	Original empirical research	Transferable HPE evidence	7. Human oversight and accountability
[36]	McDonald et al. (2025)	Original empirical research	Transferable HPE evidence	8. Institutional authorization, monitoring and incident response
[75]	International Committee of Medical Journal Editors (2026)	Professional guideline	Professional/normative guidance	6. Provenance, IP and disclosure
[76]	Alfredo et al. (2024)	Systematic review	Transferable HPE evidence	8. Institutional authorization, monitoring and incident response

TABLE 6: Condensed evidence matrix of the 71 retained sources

The final retained evidence set comprises 71 substantive sources. The Ref. column gives the formal manuscript reference number. Evidence-basis categories describe the role and disciplinary setting of each source: direct anatomy evidence; transferable HPE evidence; professional/normative guidance; or conceptual/ethical scholarship. "Direct anatomy evidence" does not imply that the AI modality was necessarily generative; Bernard et al. and Mezil et al. concern non-generative/conventional AI and are used only where their findings are transferable to governance questions such as automated assessment and human oversight. Historical working fields for relevance, source credibility, and framework utility are retained in the audit file, but their precise operational role cannot now be reconstructed with sufficient confidence; they are therefore not used in this revised article as inclusion/exclusion criteria, evidence weights, certainty ratings, or determinants of recommendation strength. Zilundu et al. [4] is cited in the Introduction as motivating contextual evidence concerning faculty-reported governance needs and was not included in the 71-source coded framework-development set. Methodological review-design references [12,13] and EU regulatory instruments [34,35] are likewise outside the retained evidence set and are used only for methodological or contemporary legal context.

HPE: health professions education; EU: European Union; AI: artificial intelligence

Conclusions

Anatomy education needs an operational layer between general responsible AI principles and daily practice. This article proposes classifying the defined use case rather than the AI product. The highest triggered criterion sets a provisional tier, with stronger controls for donor-derived inputs, authoritative learner-facing anatomy, and consequential assessment, while human accountability is retained. Authorization should remain conditional on monitoring, incident response, and revalidation.

The framework is an evidence-informed proposal for local adaptation, not a validated instrument or legal-compliance tool. Prospective evaluation should test classification consistency, feasibility, administrative burden, and effects on educational quality and safety. Applicable law, donor-program policy, assessment rules, and institutional governance remain controlling.

Appendices

ID	Citation key	Full citation	Year	Country/region	DOI/URL	Source type	Publication status	Peer reviewed?	Discipline/context	Anatomy-context source?	Evidence basis	AI modality	Educational use	Inpu
1	Lazarus, 2024	Lazarus MD, Truong M, Douglas P, Selwyn N. Artificial intelligence and clinical anatomical education: Promises and perils. Anatomical Sciences Education. 2024;17(2):249-262. doi:10.1002/ase.2221.	2024	Australia	https://doi.org/10.1002/ase.2221	Critical review	Published	Yes	Anatomy education	Yes	Conceptual/ethical scholarship	Conventional AI (foundational only)	Learning support	No s input
2	Cornwall, 2024	Cornwall J, Hildebrandt S, Champney TH, Goodman K. Ethical concerns surrounding artificial intelligence in anatomy education: Should AI human body simulations replace donors in the dissection room? Anatomical Sciences Education. 2024;17(5):937-943. doi:10.1002/ase.2335.	2024	New Zealand/United States	https://doi.org/10.1002/ase.2335	Ethics analysis	Published	Yes	Anatomy education/body donation ethics	Yes	Conceptual/ethical scholarship	Conventional AI (foundational only)	Learning support	Cada imag
3	Yi, 2026	Yi KH, Wan J, Rosellini I, Lau KH, Lee W, Haykal D. Artificial Intelligence in Anatomic Education: Educational Utility, Safety Boundaries, and Implementation Considerations.	2026	International	https://doi.org/10.1097/SCS.0000000000012573	Critical review	Online first	Yes	Anatomic education	Yes	Direct anatomy evidence	General AI/mixed modalities	Learning support	No s input

		Journal of Craniofacial Surgery. 2026 Mar 17. Online ahead of print. doi:10.1097/SCS.0000000000012573.												
4	Sonnenberg, 2026	Sonnenberg LK, Wiljer D, Mamdani M, et al. Principles for Responsible AI in Health Professions Education, Research, and Care: Health CARE-AI (Contextual, Accountable, Responsible, and Equitable Artificial Intelligence) Framework Delphi Consensus Study. JMIR Medical Education. 2026;12:e91626. doi:10.2196/91626.	2026	International	https://doi.org/10.2196/91626	Consensus statement	Published	Yes	Health professions education/health care	No	Transferable HPE evidence	General generative AI	Institutional governance	No s input
5	IFAA 2023 Images	International Federation of Associations of Anatomists (IFAA), Federative International Committee for Ethics and Medical Humanities. Recommendations for Good Practice Around Human Tissue Image Acquisition and Use in Anatomy Education and Research. 2023.	2023	International	https://ifaa.net/committees/ethics-and-medical-humanities-ficem/recommendations-for-good-practice-around-human-tissue-image-acquisition-and-use-in-anatomy-education-and-research/	Professional guideline	Guidance/policy	No	Anatomy ethics/human tissue imaging	Yes	Professional/normative guidance	Not applicable (non-AI source)	Resource generation	Hum imag
6	Jones, 2026 IFAA	Jones DG, Champney TH, Hildebrandt S, et al. IFAA Recommendations for good practice for the donation and anatomical study of human remains (revised 2026). Anatomical Sciences Education. 2026 Jul 1. Online ahead of print. doi:10.1002/ase.70289.	2026	International	https://doi.org/10.1002/ase.70289	Professional guideline	Online first	Yes	Anatomy ethics/body donation governance	Yes	Professional/normative guidance	Not applicable (non-AI source)	Institutional governance	Cada imag
7	Cornwall, 2016	Cornwall J, Callahan D, Wee R. Ethical issues surrounding the use of images from donated cadavers in the anatomical sciences. Clinical Anatomy. 2016;29(1):30-36. doi:10.1002/ca.22644.	2016	New Zealand/Portugal	https://doi.org/10.1002/ca.22644	Ethics analysis	Published	Yes	Anatomy ethics/cadaveric images	Yes	Conceptual/ethical scholarship	Not applicable (non-AI source)	Resource generation	Cada imag
8	Iwanaga, 2025	Iwanaga J, Kim HJ, Akita K, et al. Ethical Use of Cadaveric Images in Anatomical Textbooks, Atlases, and Journals: A Consensus Response From Authors and Editors. Clinical Anatomy. 2025;38(2):222-225. doi:10.1002/ca.24258.	2025	International	https://doi.org/10.1002/ca.24258	Editorial/perspective	Published	Yes	Anatomy ethics/publishing	Yes	Conceptual/ethical scholarship	Not applicable (non-AI source)	Resource generation	Cada imag
9	Eldesoqui, 2025	Eldesoqui M, Albadawi EA, AlQumaizi KI, Radwan MNM, Ebrahim HA, Elsaid MAE. Assessing the Anatomical Accuracy of AI-Generated Medical Illustrations: A Comparative Study of Text-to-Image Generator Tools in Anatomy Education. Clinical Anatomy. 2025;38(6):712-717. doi:10.1002/ca.70002.	2025	Saudi Arabia/Egypt	https://doi.org/10.1002/ca.70002	Original empirical research	Published	Yes	Anatomy education	Yes	Direct anatomy evidence	Text-to-image	Anatomical image generation	AI-ge imag
10	Alon, 2026	Alon L, Shoval DH, Levkovich I. Bias, representation, and clinical fidelity in AI-generated images for medical education: a systematic literature review. npj Digital Medicine. 2026;9:469. doi:10.1038/s41746-026-02608-3.	2026	Israel	https://doi.org/10.1038/s41746-026-02608-3	Systematic review	Published	Yes	Medical education/AI-generated images	No	Transferable HPE evidence	Text-to-image	Anatomical image generation	AI-ge imag
11	Gupta, 2026	Gupta K, Ferri Latinovich M, Ferri Latinovich M, Singh K, Patlas MN, Jajodia A. Generative Artificial Intelligence for Medical Image Creation in Health Professions Education: a Scoping Review. Journal of Medical Systems. 2026;50(1):19. doi:10.1007/s10916-026-02350-z.	2026	International	https://doi.org/10.1007/s10916-026-02350-z	Scoping review	Published	Yes	Health professions education/medical imaging	No	Transferable HPE evidence	Text-to-image	Resource generation	AI-ge imag
12	Agarwal, 2026	Agarwal P, Agarwal R, Iezhitsa I. AI for Assessment in Medical Education in Post LLM Era: A Scoping Review. Medical Science Educator. 2026;36(2):1027-1039. doi:10.1007/s40670-026-02653-3.	2026	Malaysia	https://doi.org/10.1007/s40670-026-02653-3	Scoping review	Published	Yes	Medical education assessment	No	Transferable HPE evidence	Generative assessment tool	Question generation	Asse conte
13	Che Liu, 2026	Che X, Liu Y. Guiding student use of generative AI in undergraduate pharmacology: a narrative review and proposed source-checking process framework. Frontiers in Medicine. 2026;13:1910543. doi:10.3389/fmed.2026.1910543.	2026	China	https://doi.org/10.3389/fmed.2026.1910543	Narrative review	Published	Yes	Undergraduate pharmacology/health professions education	No	Transferable HPE evidence	Large language model	Learning support	Publi
14	WHO 2024 LMM	World Health Organization. Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models. Geneva: World Health Organization; 2024.	2024	Global	https://www.who.int/publications/i/item/9789240084759	Professional guideline	Guidance/policy	No	Health AI governance	No	Professional/normative guidance	Multimodal generative model	Institutional governance	Othe

16	Abdellatif, 2022	Abdellatif H, Al Mushaiqri M, Albalushi H, Al-Zaabi AA, Roychoudhury S, Das S. Teaching, Learning and Assessing Anatomy with Artificial Intelligence: The Road to a Better Future. Int J Environ Res Public Health. 2022;19(21):14209. doi:10.3390/ijerph192114209.	2022	Oman	https://doi.org/10.3390/ijerph192114209	Narrative review	Published	Yes	Anatomy education	Yes	Conceptual/ethical scholarship	Conventional AI (foundational only)	Other	No s input
17	Joseph, 2025	Joseph TS, Gowrie S, Montalbano MJ, Bandelow S, Clunes M, Dumont AS, Iwanaga J, Tubbs RS, Loukas M. The Roles of Artificial Intelligence in Teaching Anatomy: A Systematic Review. Clinical Anatomy. 2025;38(5):552-567. doi:10.1002/ca.24272.	2025	Grenada/international collaboration	https://doi.org/10.1002/ca.24272	Systematic review	Published	Yes	Gross anatomy education	Yes	Direct anatomy evidence	General AI/mixed modalities	Learning support	No s input
18	Chytas, 2025	Chytas D, Noussios G, Paraskevas G, Vasiliadis AV, Giovanidis G, Troupis T. Can ChatGPT play a significant role in anatomy education? A scoping review. Morphologie. 2025;109(365):100949. doi:10.1016/j.morpho.2025.100949.	2025		https://doi.org/10.1016/j.morpho.2025.100949	Scoping review	Published	Yes	Anatomy education	Yes	Direct anatomy evidence	Large language model	Learning support	Publi
19	Erbek, 2026	Erbek E, Çalışkan Pala S, Bolatlı G, Çavuş F. Artificial intelligence in anatomy education: a systematic review of ChatGPT's effectiveness as a learning tool. Proc (Bayl Univ Med Cent). 2026;39(2):315-323. doi:10.1080/08998280.2026.2612683.	2026	Türkiye	https://doi.org/10.1080/08998280.2026.2612683	Systematic review	Published	Yes	Anatomy education	Yes	Direct anatomy evidence	Large language model	Learning support	Publi
20	Collins, 2024 Anatomy GPT	Collins BR, Black EW, Rarey KE. Introducing AnatomyGPT: A customized artificial intelligence application for anatomical sciences education. Clinical Anatomy. 2024;37(6):661-669. doi:10.1002/ca.24178.	2024	United States	https://doi.org/10.1002/ca.24178	Original empirical research	Published	Yes	Anatomical sciences education	Yes	Direct anatomy evidence	Large language model	Learning support	Publi
21	Temizsoy Korkmaz, 2025	Temizsoy Korkmaz F, Ok F, Karip B, Keleş P. A structured evaluation of LLM-generated step-by-step instructions in cadaveric brachial plexus dissection. BMC Medical Education. 2025;25:903. doi:10.1186/s12909-025-07493-0.	2025	Türkiye	https://doi.org/10.1186/s12909-025-07493-0	Original empirical research	Published	Yes	Cadaveric anatomy education/surgical anatomy	Yes	Direct anatomy evidence	Large language model	Teaching explanation	Publi
22	Topcu, 2026 Cadaver AI	Topcu I, Karip B, Demir P, et al. Exploring the moral status of cadavers through artificial intelligence: a thematic analysis of AI-based perspectives in anatomy education. Philosophy, Ethics, and Humanities in Medicine. 2026;21:12. doi:10.1186/s13010-026-00222-5.	2026	Türkiye	https://doi.org/10.1186/s13010-026-00222-5	Original empirical research	Published	Yes	Anatomy education/medical ethics	Yes	Direct anatomy evidence	Large language model	Learning support	Publi
23	Morgan, 2026, Assessment Literacy	Morgan A, Crabtree M, Sachsenmeier C, Troy A. Harnessing Generative Artificial Intelligence to Teach Assessment Literacy in Anatomy. Medical Science Educator. 2026;36:3-4. doi:10.1007/s40670-026-02659-x.	2026	United States	https://doi.org/10.1007/s40670-026-02659-x	Other	Published	Yes	Anatomy education/assessment literacy	Yes	Direct anatomy evidence	Generative assessment tool	Question generation	Asse conte
24	Almamari, 2026	Almamari A, Al-Farsi M, Almamari A-R, Abdellatif H. The Perceived Effectiveness of AI-Powered Tools in Undergraduate Anatomy Education: A Cross-Sectional Multifaceted Evaluation. Clinical Anatomy. 2026. Online ahead of print. doi:10.1002/ca.70188.	2026	Oman	https://doi.org/10.1002/ca.70188	Original empirical research	Online first	Yes	Undergraduate anatomy education	Yes	Direct anatomy evidence	General generative AI	Learning support	Asse conte
25	Bolgova Mavrych, 2025	Bolgova O, Mavrych V. Evolution of AI in anatomy education study based on comparison of current large language models against historical ChatGPT performance. Scientific Reports. 2025;15:37545.	2025	Saudi Arabia	https://www.nature.com/articles/s41598-025-22437-w	Original empirical research	Published	Yes	Anatomy education	Yes	Direct anatomy evidence	Large language model	Learning support	Asse conte
26	Hennessy, 2020	Hennessy CM, Royer DF, Meyer AJ, Smith CF. Social Media Guidelines for Anatomists. Anatomical Sciences Education. 2020;13(4):527-539. doi:10.1002/ase.1948.	2020	United Kingdom/United States/Australia	https://doi.org/10.1002/ase.1948	Professional guideline	Published	Yes	Anatomy education/social media	Yes	Professional/normative guidance	Not applicable (non-AI source)	Resource generation	Cada imag
		Human Tissue Authority. Filming or										Not		

27	HTA 2023 Filming	photographing of bodies or body parts that have been donated. Last updated 11 October 2023. United Kingdom.	2023	United Kingdom	https://www.hta.gov.uk/anatomy/filming	Professional guideline	Guidance/policy	No	Anatomy regulation/body donation	Yes	Professional/normative guidance	applicable (non-AI source)	Resource generation	Cada imag
28	HTA 2024 Code A	Human Tissue Authority. Code A: Guiding principles and the fundamental principle of consent — Section three: Consent requirements. Updated 2024. United Kingdom.	2024	United Kingdom	https://www.hta.gov.uk/guidance-professionals/codes-practice-standards-and-legislation/codes-practice/code-guiding-principles-and-fundamental-principle-consent/code-section-three-consent-requirements	Policy/framework document	Guidance/policy	No	Human tissue regulation/anatomy	Yes	Professional/normative guidance	Not applicable (non-AI source)	Resource generation	Hum imag
29	Passalacqua, 2025	Passalacqua NV, Bartelink E, McQuade WEP, Steadman D, Boyd D, Spradley K, Sauerwein K, Ho R. The Development of Standards for the Ethical Use of Human Skeletal Remains for Education, Research, and Training in Forensic Anthropology. American Journal of Biological Anthropology. 2025;186(3):e70022. doi:10.1002/ajpa.70022.	2025	United States	https://doi.org/10.1002/ajpa.70022	Consensus statement	Published	Yes	Forensic anthropology/human remains education	Partial/indirect	Professional/normative guidance	Not applicable (non-AI source)	Other	Hum imag
30	Keet and Kramer, 2022	Keet K, Kramer B. Advances in Digital Technology in Teaching Human Anatomy: Ethical Predicaments. Advances in Experimental Medicine and Biology. 2022;1388:173-191. doi:10.1007/978-3-031-10889-1_8.	2022	South Africa	https://doi.org/10.1007/978-3-031-10889-1_8	Ethics analysis	Published	Unclear	Anatomy education/digital technology	Yes	Conceptual/ethical scholarship	Not applicable (non-AI source)	Resource generation	Cada imag
31	Lottering, 2022	Lottering T, Billings B, Brits D, Hutchinson E, Kramer B. The ethical use of digital technology in teaching anatomy: A southern African perspective. Annals of Anatomy. 2022;244:151990. doi:10.1016/j.aanat.2022.151990.	2022	Southern Africa	https://doi.org/10.1016/j.aanat.2022.151990	Original empirical research	Published	Yes	Anatomy education/digital resources	Yes	Direct anatomy evidence	Not applicable (non-AI source)	Resource generation	Hum imag
32	Cornwall, 2025 Illustrations	Cornwall J, White R, Pennefather P, Hildebrandt S, Gregory J, Smith HF, Organ J, Krebs C. Legal and ethical considerations around the use of existing illustrations to generate new illustrations in the anatomical sciences. Anatomical Sciences Education. 2025;18(3):289-300. doi:10.1002/ase.70002.	2025	New Zealand/Canada/United States	https://doi.org/10.1002/ase.70002	Ethics analysis	Published	Yes	Anatomical sciences/illustration/GenAI	Yes	Conceptual/ethical scholarship	Text-to-image	Anatomical image generation	Educ imag
33	Ho, 2026 YouTube	Ho A, Hulme A, Prvan T, Štrkalj G. Evidence for an ethics-practice gap in digital anatomy education: Provenance and consent disclosure in English-language YouTube videos. Annals of Anatomy. 2026;266:152856. doi:10.1016/j.aanat.2026.152856.	2026	Australia	https://doi.org/10.1016/j.aanat.2026.152856	Original empirical research	Published	Yes	Digital anatomy education/YouTube	Yes	Direct anatomy evidence	Not applicable (non-AI source)	Resource generation	Hum imag
34	Rossetti, 2026 Body Data	Rossetti N, Licata M, Fusco R, Vanni A, Picozzi M. The Body as Data: from human remains to digital files. A review of current literature. Digital Applications in Archaeology and Cultural Heritage. 2026;41:e00543. doi:10.1016/j.daach.2026.e00543.	2026	Italy	https://doi.org/10.1016/j.daach.2026.e00543	Systematic review	Published	Yes	Digital human remains/anthropology/heritage	Partial/indirect	Conceptual/ethical scholarship	Not applicable (non-AI source)	Resource generation	Hum imag
35	Kramer Moxham, 2026	Kramer B, Moxham B. Beyond the grave: Do the dead have rights? Anatomical Sciences Education. 2026. doi:10.1002/ase.70259.	2026	South Africa/United Kingdom	https://doi.org/10.1002/ase.70259	Editorial/perspective	Online first	Yes	Anatomical sciences ethics	Yes	Conceptual/ethical scholarship	Not applicable (non-AI source)	Resource generation	Cada imag
36	Iwanaga, 2022 Statement	Iwanaga J, Singh V, Takeda S, et al. Standardized statement for the ethical use of human cadaveric tissues in anatomy research papers: Recommendations from Anatomical Journal Editors-in-Chief. Clinical Anatomy. 2022;35(4):526-528. doi:10.1002/ca.23849.	2022	International	https://doi.org/10.1002/ca.23849	Consensus statement	Published	Yes	Anatomical research/donor acknowledgement	Yes	Professional/normative guidance	Not applicable (non-AI source)	Other	Cada imag
37	Hyeamang, 2025	Hyeamang LJ, Sekhar TC, Rush E, Beresheim AC, Cheverko CM, Brooks WS, et al. AI's ability to interpret unlabeled anatomy images and supplement educational research as an AI rater. Anatomical Sciences Education. 2025;18(10):1102-1113. doi:10.1002/ase.70074.	2025	United States/Australia	https://doi.org/10.1002/ase.70074	Original empirical research	Published	Yes	Anatomy education/image interpretation	Yes	Direct anatomy evidence	Multimodal generative model	Learning support	Educ imag
38	Mavrych, 2025 Gross Anatomy	Mavrych V, Ganguly P, Bolgova O. Using large language models (ChatGPT, Copilot, PaLM, Bard, and Gemini) in Gross Anatomy course: Comparative analysis. Clinical	2025	Saudi Arabia	https://doi.org/10.1002/ca.24244	Original empirical research	Published	Yes	Gross anatomy education	Yes	Direct anatomy evidence	Large language model	Teaching explanation	Asse conte

		Anatomy. 2025;38(2):200-210. doi:10.1002/ca.24244.												
39	AlKhater, 2025	Al-Khater KMK. Comparative assessment of three AI platforms in answering USMLE Step 1 anatomy questions or identifying anatomical structures on radiographs. Clinical Anatomy. 2025;38(2):186-199. doi:10.1002/ca.24243.	2025	Saudi Arabia	https://doi.org/10.1002/ca.24243	Original empirical research	Published	Yes	Anatomy education/radiographic anatomy	Yes	Direct anatomy evidence	Multimodal generative model	Imaging/radiological anatomy	Medi
40	Sivri, 2026 Radiological	Sivri I, Ozden FM, Celik H, Gokturk O, Colak T. Comparative evaluation of radiological anatomy knowledge and accuracy of ChatGPT-5, Gemini 2.5, and Grok 4 across normal and thinking modes. Anatomical Sciences Education. 2026 Jun 26. Online ahead of print. doi:10.1002/ase.70292.	2026	Türkiye	https://doi.org/10.1002/ase.70292	Original empirical research	Online first	Yes	Radiological anatomy education	Yes	Direct anatomy evidence	Multimodal generative model	Imaging/radiological anatomy	Medi
41	Pasli Beser, 2026	Paslı B, Günenç Beşer C. Is artificial intelligence getting better at anatomy? A two-year review of ChatGPT's free public versions. Anatomical Sciences Education. 2026 May 14. Online ahead of print. doi:10.1002/ase.70255.	2026	Türkiye	https://doi.org/10.1002/ase.70255	Original empirical research	Online first	Yes	Anatomy education	Yes	Direct anatomy evidence	Large language model	Learning support	Asse conte
42	Ulkir Pasli 2026	Ülkir M, Paslı B. Reference Hallucination, Citation Reliability, and Readability of Large Language Models in Anatomy-Related Question Answering. Clinical Anatomy. 2026 Jul 15. Online ahead of print. doi:10.1002/ca.70187.	2026	Türkiye	https://doi.org/10.1002/ca.70187	Original empirical research	Online first	Yes	Anatomy education/scholarly reference use	Yes	Direct anatomy evidence	Large language model	Teaching explanation	Publi
43	Sarantopoulos, 2026	Sarantopoulos A, Pana ZD, Larentzakis A, Kondylis S, Maina A, Ziogas N, Ntourakis D. ChatGPT-4.0 and Medical Students: A Recognition-Gated Comparative Evaluation on Image-Based Medical Examinations. Journal of Medical Education and Curricular Development. 2026;13:23821205261449387. doi:10.1177/23821205261449387.	2026	Cyprus	https://doi.org/10.1177/23821205261449387	Original empirical research	Published	Yes	Medical education/image-based anatomy assessment	Partial/indirect	Transferable HPE evidence	Multimodal generative model	Imaging/radiological anatomy	Medi
45	Bae, 2025 Anatomy Images	Bae JS, Kim GY, Kim HJ, Han SH, Youn KH. Comparison of anatomy image generation capability in AI image generation models. J Vis Commun Med. 2025;48(2):44-51. doi:10.1080/17453054.2025.2504491.	2025	Republic of Korea	https://doi.org/10.1080/17453054.2025.2504491	Original empirical research	Published	Yes	Anatomy/biomedical illustration	Yes	Direct anatomy evidence	Text-to-image	Anatomical image generation	AI-ge imag
46	Kumar, 2024 Medical Illustrations	Kumar A, Burr P, Young TM. Using AI Text-to-Image Generation to Create Novel Illustrations for Medical Education: Current Limitations as Illustrated by Hypothyroidism and Horner Syndrome. JMIR Med Educ. 2024;10:e52155. doi:10.2196/52155.	2024	United Kingdom	https://doi.org/10.2196/52155	Original empirical research	Published	Yes	Medical education/medical illustration	Partial/indirect	Transferable HPE evidence	Text-to-image	Resource generation	AI-ge imag
47	Currie, 2024 Medical Students	Currie G, Currie J, Anderson S, Hewis J. Gender bias in generative artificial intelligence text-to-image depiction of medical students. Health Educ J. 2024;83(7):732-746. doi:10.1177/00178969241274621.	2024	Australia	https://doi.org/10.1177/00178969241274621	Original empirical research	Published	Yes	Medical education/representation	No	Transferable HPE evidence	Text-to-image	Resource generation	AI-ge imag
48	Bui, 2026 Medical Students	Bui M, Yi PH, Onuh I, Kuckelman I, Ross AB. Portrayal of Medical Students in Artificial Intelligence-Generated Images. WMJ. 2026;125(1):158-161. PMID:41980156.	2026	United States	https://pubmed.ncbi.nlm.nih.gov/41980156/	Original empirical research	Published	Yes	Medical education/representation	No	Transferable HPE evidence	Text-to-image	Resource generation	AI-ge imag
49	Shaari, 2025 Otolaryngology	Shaari AL, Saad AM, Patel AM, Filimonov A. Representation of Demographics in Otolaryngology by Artificial Intelligence Text-to-Image Platforms. Laryngoscope Investig Otolaryngol. 2025;10(3):e70152. doi:10.1002/lio2.70152.	2025	United States	https://doi.org/10.1002/lio2.70152	Original empirical research	Published	Yes	Surgical education/professional representation	No	Transferable HPE evidence	Text-to-image	Resource generation	AI-ge imag
50	Currie, 2025 Imaging Bias	Currie G, Hewis J, Hawk E, Rohren E. Gender and Ethnicity Bias of Text-to-Image Generative Artificial Intelligence in Medical Imaging, Part 2: Analysis of	2025	Australia/United States	https://doi.org/10.2967/jnmt.124.268359	Original empirical research	Published	Yes	Medical imaging professions/representation	No	Transferable HPE evidence	Text-to-image	Resource generation	AI-ge imag

		DALL-E 3. J Nucl Med Technol. 2025;53(2):162-168. doi:10.2967/jnmt.124.268359.												
51	Ali, 2024 Surgeon Representation	Ali R, Tang OY, Connolly ID, et al. Demographic Representation in 3 Leading Artificial Intelligence Text-to-Image Generators. JAMA Surg. 2024;159(1):87-95. doi:10.1001/jamasurg.2023.5695.	2024	United States	https://doi.org/10.1001/jamasurg.2023.5695	Original empirical research	Published	Yes	Surgery/professional representation	No	Transferable HPE evidence	Text-to-image	Resource generation	AI-ge imag
52	Senft Everson, 2025 Cancer Survivor	Senft Everson N, Gaysynsky A, Iles IA, Schrader KE, Chou W-YS. What does an AI-generated 'cancer survivor' look like? An analysis of images generated by text-to-image tools. J Cancer Surviv. 2025. doi:10.1007/s11764-025-01760-1.	2025	United States	https://doi.org/10.1007/s11764-025-01760-1	Original empirical research	Published	Yes	Patient-facing medical imagery/representation	No	Transferable HPE evidence	Text-to-image	Resource generation	AI-ge imag
53	Lopiano, 2026	Lopiano S, Cohen ES, Holan G, Grachan JJ. Prompting the Question: Evaluating ChatGPT-Generated Gross Anatomy and Embryology Questions for Accuracy and NBME Item-Writing Style. Cureus. 2026;18(4):e107397. doi:10.7759/cureus.107397.	2026	United States	https://doi.org/10.7759/cureus.107397	Original empirical research	Published	Yes	Gross anatomy/embryology assessment	Yes	Direct anatomy evidence	Large language model	Question generation	Asse conte
54	Kiyak Peker, 2025	Kıyak YS, Soylu A, Coşkun Ö, Budakoğlu II, Peker TV. Can ChatGPT Generate Acceptable Case-Based Multiple-Choice Questions for Medical School Anatomy Exams? A Pilot Study on Item Difficulty and Discrimination. Clinical Anatomy. 2025;38:505-510. doi:10.1002/ca.24271.	2025	Türkiye	https://doi.org/10.1002/ca.24271	Original empirical research	Published	Yes	Medical school anatomy assessment	Yes	Direct anatomy evidence	Large language model	Question generation	Asse conte
55	Tekin, 2025	Tekin A, Karamus NF, Çolak T. Anatomy exam model for the circulatory and respiratory systems using GPT-4: a medical school study. Surg Radiol Anat. 2025;47:158. doi:10.1007/s00276-025-03667-z.	2025	Türkiye	https://doi.org/10.1007/s00276-025-03667-z	Original empirical research	Published	Yes	Anatomy assessment/medical school	Yes	Direct anatomy evidence	Large language model	Question generation	Asse conte
56	Bernard, 2024 OSPE	Bernard J, Sonnadara R, Saraco AN, Mitchell JP, Bak AB, Bayer I, Wainman BC. Automated grading of anatomical objective structured practical examinations using decision trees: An artificial intelligence approach. Anat Sci Educ. 2024;17:967-978. doi:10.1002/ase.2305.	2024	Canada	https://doi.org/10.1002/ase.2305	Original empirical research	Published	Yes	Anatomy OSPE grading	Yes	Direct anatomy evidence	Conventional AI (foundational only)	Scoring/grading	Stud work
57	Mezil, 2026 AI OSPE	Mezil Y, Yu P, Rahimpour L, Wang T, Rajyam S, Varambally M, et al. Harnessing artificial intelligence for automatic feedback in a virtual anatomy study tool: A Q-methodology study. Anat Sci Educ. 2026;19(5):684-694. doi:10.1002/ase.70206.	2026	Saudi Arabia/Canada	https://doi.org/10.1002/ase.70206	Original empirical research	Published	Yes	Anatomy OSPE practice/feedback	Yes	Direct anatomy evidence	Conventional AI (foundational only)	Feedback	Stud work
58	Al Mashhadani, 2026 Feedback	Al-Mashhadani MM, Ajaz F, Guraya SS, Ennab F. Use of large language models for providing automated feedback in medical imaging education: a systematic review. Front Med (Lausanne). 2026;13:1803921. doi:10.3389/fmed.2026.1803921.	2026	United Arab Emirates	https://doi.org/10.3389/fmed.2026.1803921	Systematic review	Published	Yes	Medical imaging education/automated feedback	No	Transferable HPE evidence	Large language model	Feedback	Stud work
	Ahmed, 2025	Ahmed A, Kerr E, O'Malley A. Quality assurance and validity of AI-generated				Original empirical			Medical education		Transferable HPE	Large	Question	Asse

59	SBA Quality	single best answer questions. BMC Med Educ. 2025;25:300. doi:10.1186/s12909-025-06881-w.	2025	United Kingdom	https://doi.org/10.1186/s12909-025-06881-w	research	Published	Yes	assessment	No	evidence	language model	generation	conte
60	Meo, 2024 OSPE Bard	Meo SA, AbuKhalaf AA, Meo MZS, Meo MO, Ayub R, ElToukhy RA, Usmani AM, Hajjar WM. Role of artificial intelligence (Google Bard) in morphological, histopathological, and radiological image identifications: Objective Structured Practical Examination (OSPE) type-based performance. Saudi Med J. 2024;45(5):531-536. doi:10.15537/smj.2024.45.5.20240141.	2024	Saudi Arabia	https://doi.org/10.15537/smj.2024.45.5.20240141	Original empirical research	Published	Yes	OSPE-type image identification/medical sciences	Partial/indirect	Transferable HPE evidence	Multimodal generative model	Practical/OSPE	Medi
61	AAMC, 2025 Responsible AI	Association of American Medical Colleges. Principles for the Responsible Use of Artificial Intelligence in and for Medical Education. Version 2.0 completed July 31, 2025.	2025	United States	https://www.aamc.org/about-us/mission-areas/medical-education/principles-ai-use	Professional guideline	Guidance/policy	No	Medical education governance	No	Professional/normative guidance	General generative AI	Institutional governance	Instit conte
62	Masters, 2023 AMEE158	Masters K. Ethical use of Artificial Intelligence in Health Professions Education: AMEE Guide No. 158. Med Teach. 2023;45(6):574-584. doi:10.1080/0142159X.2023.2186203.	2023	Oman	https://doi.org/10.1080/0142159X.2023.2186203	Professional guideline	Published	Yes	Health professions education ethics	No	Transferable HPE evidence	General generative AI	Institutional governance	Stud work
63	NIST, 2024 GenAIRMF	Autio C, Schwartz R, Dunietz J, et al. Artificial Intelligence Risk Management Framework: Generative Artificial Intelligence Profile. NIST AI 600-1. National Institute of Standards and Technology; 2024. doi:10.6028/NIST.AI.600-1.	2024	United States	https://doi.org/10.6028/NIST.AI.600-1	Policy/framework document	Published	No	Cross-sector generative-AI risk governance	No	Professional/normative guidance	General generative AI	Institutional governance	Instit conte
64	Lekadir, 2025 FUTUREAI	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:10.1136/bmj-2024-081554.	2025	International	https://doi.org/10.1136/bmj-2024-081554	Consensus statement	Published	Yes	Healthcare AI governance	No	Transferable HPE evidence	Conventional AI (foundational only)	Institutional governance	Instit conte
65	Gin, 2025 Entrustment	Gin BC, O'Sullivan PS, Hauer KE, Abdulnour RE, Mackenzie M, ten Cate O, Boscardin CK. Entrustment and EPAs for Artificial Intelligence (AI): A Framework to Safeguard the Use of AI in Health Professions Education. Acad Med. 2025;100(3):264-272. doi:10.1097/ACM.0000000000005930.	2025	United States/The Netherlands	https://doi.org/10.1097/ACM.0000000000005930	Policy/framework document	Published	Yes	Health professions education governance	No	Transferable HPE evidence	General generative AI	Institutional governance	Instit conte
66	Izquierdo Condoy, 2026 FACETS	Izquierdo-Condoy JS, Arias-Intriago M, Montero Corrales L, Ortiz-Prado E. Artificial Intelligence in Medical Education: Transformative Potential, Current Applications, and Future Implications. JMIR Medical Education. 2026;12:e77127. doi:10.2196/77127.	2026	International	https://doi.org/10.2196/77127	Narrative review	Published	Yes	Medical education implementation	No	Transferable HPE evidence	General generative AI	Institutional governance	Instit conte
67	OECD, 2024 AI Principles	Organisation for Economic Co-operation and Development. OECD AI Principles. Adopted 2019; updated 2024.	2024	International	https://www.oecd.org/en/topics/ai-principles.html	Policy/framework document	Guidance/policy	No	Cross-sector AI governance	No	Professional/normative guidance	General generative AI	Institutional governance	Instit conte

68	UNESCO, 2023 GenAI	Miao F, Holmes W. Guidance for Generative AI in Education and Research. Paris: UNESCO; 2023 (updated online 2026).	2023	Global	https://www.unesco.org/en/articles/guidance-generative-ai-education-and-research	Professional guideline	Guidance/policy	No	Education/research generative-AI governance	No	Professional/normative guidance	General generative AI	Institutional governance	Stud work
69	Jin, 2025 Global Policies	Jin Y, Yan L, Echeverria V, Gašević D, Martinez-Maldonado R. Generative AI in higher education: A global perspective of institutional adoption policies and guidelines. Computers and Education: Artificial Intelligence. 2025;8:100348. doi:10.1016/j.caeai.2024.100348.	2025	International	https://doi.org/10.1016/j.caeai.2024.100348	Original empirical research	Published	Yes	Global higher-education GenAI adoption policy	No	Transferable HPE evidence	General generative AI	Institutional governance	No s input
70	Ganguly, 2025 Research Guidance	Ganguly A, Johri A, Ali A, McDonald N. Generative artificial intelligence for academic research: evidence from guidance issued for researchers by higher education institutions in the United States. AI and Ethics. 2025;5:3917-3933. doi:10.1007/s43681-025-00688-7.	2025	United States	https://doi.org/10.1007/s43681-025-00688-7	Original empirical research	Published	Yes	Higher-education GenAI research guidance	No	Transferable HPE evidence	General generative AI	Institutional governance	Publi
71	McDonald, 2025 Policies	McDonald N, Johri A, Ali A, Collier AH. Generative artificial intelligence in higher education: Evidence from an analysis of institutional policies and guidelines. Computers in Human Behavior: Artificial Humans. 2025;3:100121. doi:10.1016/j.chbah.2025.100121.	2025	United States	https://doi.org/10.1016/j.chbah.2025.100121	Original empirical research	Published	Yes	Higher-education institutional GenAI policy	No	Transferable HPE evidence	General generative AI	Institutional governance	No s input
72	ICMJE, 2026 AI	International Committee of Medical Journal Editors. Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals: Use of Artificial Intelligence in Publishing. Updated January 2026.	2026	International	https://www.icmje.org/recommendations/browse/artificial-intelligence/	Professional guideline	Guidance/policy	No	Medical publishing/AI transparency and accountability	No	Professional/normative guidance	General generative AI	Resource generation	Publi
73	Alfredo, 2024 Human Centred	Alfredo R, Echeverria V, Jin Y, Yan L, Swiecki Z, Gašević D, Martinez-Maldonado R. Human-centred learning analytics and AI in education: A systematic literature review. Computers and Education: Artificial Intelligence. 2024;6:100215. doi:10.1016/j.caeai.2024.100215.	2024	International	https://doi.org/10.1016/j.caeai.2024.100215	Systematic review	Published	Yes	AI in education/human-centred system design	No	Transferable HPE evidence	Conventional AI (foundational only)	Institutional governance	Stud work

TABLE 7: Source-level evidence matrix for the 71 retained substantive sources

This appendix contains the complete source-level coding for the 71 substantive sources retained in the evidence synthesis. Evidence-basis categories are direct anatomy evidence, transferable health professions education (HPE) evidence, professional or normative guidance, and conceptual or ethical scholarship. "Anatomy-context source?" indicates whether a source concerns anatomy, anatomical material, anatomy education, or anatomy-related governance and is distinct from the narrower "direct anatomy evidence" category. AI modality is classified as GenAI, conventional AI (foundational/transferable), general AI/mixed modalities, or not applicable (non-AI source). "Illustrative tier relevance" is a historical working field and does not classify the source itself. Relevance, source credibility, framework utility, and triage-priority fields are retained as historical working fields; their precise operational role cannot now be reconstructed and they are not used in the revised manuscript as eligibility criteria, evidence weights, certainty ratings, recommendation-strength measures, or determinants of framework inclusion. This appendix contains retained sources only; historical record-level exclusion data were not preserved and are not retrospectively reconstructed. Principal governance-domain coding corresponds to the eight domains presented in main-text Table 2. DOI or stable URL fields are provided where available.

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Ahmed Mahgoub, Ayman Mahgoub

Acquisition, analysis, or interpretation of data: Ahmed Mahgoub, Anaida Singh, Ayman Mahgoub

Drafting of the manuscript: Ahmed Mahgoub, Ayman Mahgoub

Critical review of the manuscript for important intellectual content: Ahmed Mahgoub, Anaida Singh, Ayman Mahgoub

Disclosures

Conflicts of interest: In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Acknowledgements

ChatGPT (OpenAI) was used to check this manuscript for English language, grammar, spelling, punctuation, and readability issues. All substantive content, source selection, interpretations, framework decisions, and conclusions were developed and verified by the authors.

References

1. Yi KH, Wan J, Rosellini I, Hong Lau K, Lee W, Haykal D: Artificial intelligence in anatomic education: educational utility, safety boundaries, and implementation considerations. J Craniofac Surg. 2026, 10.1097/SCS.0000000000012573
2. Lazarus MD, Truong M, Douglas P, Selwyn N: Artificial intelligence and clinical anatomical education: promises and perils. Anat Sci Educ. 2024, 17:249-262. 10.1002/ase.2221
3. Cornwall J, Hildebrandt S, Champney TH, Goodman K: Ethical concerns surrounding artificial intelligence in anatomy education: should AI human body simulations replace donors in the dissection room?. Anat Sci Educ. 2024, 17:937-943. 10.1002/ase.2335
4. Zilundu PLM, Mazengenya P, Narayanan JK, Zhou L: Faculty perspectives on AI integration in anatomy education in the United Arab Emirates: cross-sectional survey. JMIR Med Educ. 2026, 12:e87418. 10.2196/87418
5. Sonnenberg LK, Wiljer D, Mamdani M, et al.: Principles for responsible AI in health professions education, research, and care: Health CARE-AI (contextual, accountable, responsible, and equitable artificial intelligence) Framework Delphi consensus study. JMIR Med Educ. 2026, 12:e91626. 10.2196/91626
6. Autio C, Schwartz R, Dunietz J, et al.: Artificial intelligence risk management framework: generative artificial intelligence profile (NIST AI 600-1). NIST Trustworthy and Responsible AI. National Institute of Standards and Technology, Gaithersburg, MD; 2024. 10.6028/NIST.AI.600-1
7. 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. 10.1136/bmj-2024-081554
8. Miao F, Holmes W: Guidance for generative AI in education and research . 2023 (ed): UNESCO, Paris, France; 10.54675/EWZM9535
9. Organisation for Economic Co-operation and Development: recommendation of the Council on Artificial Intelligence (OECD/LEGAL/0449). (2024). Accessed: August 24, 2026: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.
10. Masters K: Ethical use of artificial intelligence in health professions education: AMEE guide no. 158 . Med Teach. 2023, 45:574-584. 10.1080/0142159X.2023.2186203
11. Gin BC, O'Sullivan PS, Hauer KE, Abdulnour RE, Mackenzie M, Ten Cate O, Boscardin CK: Entrustment and EPAs for artificial intelligence (AI): a framework to safeguard the use of AI in health professions education. Acad Med. 2025, 100:264-272. 10.1097/ACM.0000000000005930
12. Grant MJ, Booth A: A typology of reviews: an analysis of 14 review types and associated methodologies . Health Info Libr J. 2009, 26:91-108. 10.1111/j.1471-1842.2009.00848.x
13. Barry ES, Merkebu J, Varpio L: Understanding state-of-the-art literature reviews . J Grad Med Educ. 2022, 14:659-662. 10.4300/JGME-D-22-00705.1
14. International Federation of Associations of Anatomists, Federative International Committee for Ethics and Medical Humanities: recommendations for good practice around human tissue image acquisition and use in anatomy education and research. (2023). Accessed: August 24, 2026: https://ifaa.net/committees/ethics-and-medical-humanities-ficem/recommendations-for-good-practice-around-human-tissue....
15. Jones DG, Champney TH, Hildebrandt S, et al.: IFAA recommendations for good practice for the donation and anatomical study of human remains (revised 2026). Anat Sci Educ. 2026, 1-6. 10.1002/ase.70289
16. Iwanaga J, Kim HJ, Akita K, et al.: Ethical use of cadaveric images in anatomical textbooks, atlases, and journals: a consensus response from authors and editors. Clin Anat. 2025, 38:222-225. 10.1002/ca.24258
17. Cornwall J, Callahan D, Wee R: Ethical issues surrounding the use of images from donated cadavers in the anatomical sciences. Clin Anat. 2016, 29:30-36. 10.1002/ca.22644
18. Human Tissue Authority: filming or photographing of bodies or body parts that have been donated . (2023). Accessed: August 24, 2026: https://www.hta.gov.uk/anatomy/filming.
19. Eldesoqui M, Albadawi EA, AlQumaizi KI, Radwan MNM, Ebrahim HA, Elsaid MAE: Assessing the anatomical accuracy of AI-generated medical illustrations: a comparative study of text-to-image generator tools in anatomy education. Clin Anat. 2025, 38:712-717. 10.1002/ca.70002
20. Hyeamang LJ, Sekhar TC, Rush E, et al.: AI's ability to interpret unlabeled anatomy images and supplement educational research as an AI rater. Anat Sci Educ. 2025, 18:1102-1113. 10.1002/ase.70074
21. Mavrych V, Ganguly P, Bolgova O: Using large language models (ChatGPT, Copilot, PaLM, Bard, and Gemini) in gross anatomy course: comparative analysis. Clin Anat. 2025, 38:200-210. 10.1002/ca.24244
22. Bae JS, Kim GY, Kim HJ, Han SH, Youn KH: Comparison of anatomy image generation capability in AI image generation models. J Vis Commun Med. 2025, 48:44-51. 10.1080/17453054.2025.2504491
23. Ülkir M, Paslı B: Reference hallucination, citation reliability, and readability of large language models in anatomy-related question answering. Clin Anat. 2026, 10.1002/ca.70187
24. Sarantopoulos A, Pana ZD, Larentzakis A, Kondylis S, Maina A, Ziogas N, Ntourakis D: ChatGPT-4.0 and medical students: a recognition-gated comparative evaluation on image-based medical examinations. J Med Educ Curric Dev. 2026, 13:23821205261449387. 10.1177/23821205261449387
25. Paslı B, Günenç Beşer C: Is artificial intelligence getting better at anatomy? A two-year review of ChatGPT's free public versions. Anat Sci Educ. 2026, 19:1279-1287. 10.1002/ase.70255
26. Alon L, Shoval DH, Levkovich I: Bias, representation, and clinical fidelity in AI-generated images for medical education: a systematic literature review. NPJ Digit Med. 2026, 9:469. 10.1038/s41746-026-02608-3
27. Kıyak YS, Soylu A, Coşkun Ö, Budakoğlu Iİ, Peker TV: Can ChatGPT generate acceptable case-based multiple-choice questions for medical school anatomy exams? A pilot study on item difficulty and discrimination. Clin Anat. 2025, 38:505-510. 10.1002/ca.24271
28. Lopiano S, Cohen ES, Holan G, Grachan JJ: Prompting the question: evaluating ChatGPT-generated gross anatomy and embryology questions for accuracy and NBME (National Board of Medical Examiners) item-writing style. Cureus. 2026, 18:e107397. 10.7759/cureus.107397
29. Tekin A, Karamus NF, Çolak T: Anatomy exam model for the circulatory and respiratory systems using GPT-4: a medical school study. Surg Radiol Anat. 2025, 47:158. 10.1007/s00276-025-03667-z
30. Meo SA, AbuKhalaf AA, Meo MZS, et al.: Role of artificial intelligence (Google Bard) in morphological, histopathological, and radiological image identifications: Objective Structured Practical Examination (OSPE) type-based performance. Saudi Med J. 2024, 45:531-536. 10.15537/smj.2024.45.5.20240141
31. Bernard J, Sonnadara R, Saraco AN, Mitchell JP, Bak AB, Bayer I, Wainman BC: Automated grading of anatomical objective structured practical examinations using decision trees: an artificial intelligence approach. Anat Sci Educ. 2024, 17:967-978. 10.1002/ase.2305
32. Ho A, Hulme A, Prvan T, Štrkalj G: Evidence for an ethics-practice gap in digital anatomy education: provenance and consent disclosure in English-language YouTube videos. Ann Anat. 2026, 266:152856. 10.1016/j.aanat.2026.152856
33. Cornwall J, White R, Pennefather P, et al.: Legal and ethical considerations around the use of existing illustrations to generate new illustrations in the anatomical sciences. Anat Sci Educ. 2025, 18:289-300. 10.1002/ase.70002
34. European Parliament, Council of the European Union: regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending regulations (EC) no 300/2008, (EU) no 167/2013, (EU) no 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act). (2024). Accessed: August 27, 2026: https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng.
35. European Parliament, Council of the European Union: regulation (EU) 2026/1744 of the European Parliament and of the Council of 8 July 2026 amending regulations (EU) 2024/1689, (EU) 2018/1139 and (EU) 2023/1230 as regards the simplification of the implementation of harmonised rules on artificial intelligence (Digital Omnibus on AI). (2026). Accessed: August 27, 2026: https://eur-lex.europa.eu/eli/reg/2026/1744/oj/eng.
36. McDonald N, Johri A, Ali A, Collier AH: Generative artificial intelligence in higher education: evidence from an analysis of institutional policies and guidelines. Comput Human Behav Artif Humans. 2025, 3:100121. 10.1016/j.chbah.2025.100121
37. Jin Y, Yan L, Echeverria V, Gašević D, Martinez-Maldonado R: Generative AI in higher education: a global perspective of institutional adoption policies and guidelines. Comput Educ Artif Intell. 2025, 8:100348. 10.1016/j.caeai.2024.100348
38. Gupta K, Latinovich MF, Latinovich MF, Singh K, Patlas MN, Jajodia A: Generative artificial intelligence for medical image creation in health professions education: a scoping review. J Med Syst. 2026, 50:19. 10.1007/s10916-026-02350-z
39. Agarwal P, Agarwal R, Iezhitsa I: AI for assessment in medical education in post LLM era: a scoping review . Med Sci Educ. 2026, 36:1027-1039. 10.1007/s40670-026-02653-3
40. Che X, Liu Y: Guiding student use of generative AI in undergraduate pharmacology: a narrative review and proposed source-checking process framework. Front Med (Lausanne). 2026, 13:1910543. 10.3389/fmed.2026.1910543
41. World Health Organization: ethics and governance of artificial intelligence for health: guidance on large multi-modal models. (2025). Accessed: August 24, 2026: https://www.who.int/publications/i/item/9789240084759.
42. Abdellatif H, Al Mushaiqri M, Albalushi H, Al-Zaabi AA, Roychoudhury S, Das S: Teaching, learning and

assessing anatomy with artificial intelligence: the road to a better future. Int J Environ Res Public Health. 2022, 19:14209. 10.3390/ijerph192114209
43. Joseph TS, Gowrie S, Montalbano MJ, et al.: The roles of artificial intelligence in teaching anatomy: a systematic review. Clin Anat. 2025, 38:552-567. 10.1002/ca.24272
44. Chytas D, Noussios G, Paraskevas G, Vasiliadis AV, Giovanidis G, Troupis T: Can ChatGPT play a significant role in anatomy education? A scoping review. Morphologie. 2025, 109:100949. 10.1016/j.morpho.2025.100949
45. Erbek E, Çalışkan Pala S, Bolatlı G, Çavuş F: Artificial intelligence in anatomy education: a systematic review of ChatGPT's effectiveness as a learning tool. Proc (Bayl Univ Med Cent). 2026, 39:315-323. 10.1080/08998280.2026.2612683
46. Collins BR, Black EW, Rarey KE: Introducing AnatomyGPT: a customized artificial intelligence application for anatomical sciences education. Clin Anat. 2024, 37:661-669. 10.1002/ca.24178
47. Temizsoy Korkmaz F, Ok F, Karip B, Keleş P: A structured evaluation of LLM-generated step-by-step instructions in cadaveric brachial plexus dissection. BMC Med Educ. 2025, 25:903. 10.1186/s12909-025-07493-0
48. Topcu I, Karip B, Demir P, Kac B: Exploring the moral status of cadavers through artificial intelligence: a thematic analysis of AI-based perspectives in anatomy education. Philos Ethics Humanit Med. 2026, 21:12. 10.1186/s13010-026-00222-5
49. Morgan A, Crabtree M, Sachsenmeier C, Troy A: Harnessing generative artificial intelligence to teach assessment literacy in anatomy. Med Sci Educ. 2026, 36:3-4. 10.1007/s40670-026-02659-x
50. Almamari A, Al-Farsi M, Almamari AR, Abdellatif H: The perceived effectiveness of AI-powered tools in undergraduate anatomy education: a cross-sectional multifaceted evaluation. Clin Anat. 2026, 10.1002/ca.70188
51. Bolgova O, Mavrych V: Evolution of AI in anatomy education study based on comparison of current large language models against historical ChatGPT performance. Sci Rep. 2025, 15:37545. 10.1038/s41598-025-22437-w
52. Hennessy CM, Royer DF, Meyer AJ, Smith CF: Social media guidelines for anatomists. Anat Sci Educ. 2020, 13:527-539. 10.1002/ase.1948
53. Human Tissue Authority: code A: guiding principles and the fundamental principle of consent - section three: consent requirements. (2025). Accessed: August 24, 2026: https://www.hta.gov.uk/guidance-professionals/codes-practice-standards-and-legislation/codes-practice/code-guiding-pr....
54. Passalacqua NV, Bartelink E, McQuade WEP, et al.: The development of standards for the ethical use of human skeletal remains for education, research, and training in forensic anthropology. Am J Biol Anthropol. 2025, 186:e70022. 10.1002/ajpa.70022
55. Keet K, Kramer B: Advances in digital technology in teaching human anatomy: ethical predicaments . Adv Exp Med Biol. 2022, 1388:173-191. 10.1007/978-3-031-10889-1_8
56. Lottering T, Billings B, Brits D, Hutchinson E, Kramer B: The ethical use of digital technology in teaching anatomy: a southern African perspective. Ann Anat. 2022, 244:151990. 10.1016/j.aanat.2022.151990
57. Rossetti N, Licata M, Fusco R, Vanni A, Picozzi M: The body as data: from human remains to digital files. A review of current literature. Digit Appl Archaeol Cult Heritage. 2026, 41:543. 10.1016/j.daach.2026.e00543
58. Kramer B, Moxham B: Beyond the grave: do the dead have rights? . Anat Sci Educ. 2026, 19:1360-1366. 10.1002/ase.70259
59. Iwanaga J, Singh V, Takeda S, et al.: Standardized statement for the ethical use of human cadaveric tissues in anatomy research papers: recommendations from anatomical journal Editors-in-Chief. Clin Anat. 2022, 35:526-528. 10.1002/ca.23849
60. Al-Khater KMK: Comparative assessment of three AI platforms in answering USMLE Step 1 anatomy questions or identifying anatomical structures on radiographs. Clin Anat. 2025, 38:186-199. 10.1002/ca.24243
61. Sivri I, Ozden FM, Celik H, Gokturk O, Colak T: Comparative evaluation of radiological anatomy knowledge and accuracy of ChatGPT-5, Gemini 2.5, and Grok 4 across normal and thinking modes. Anat Sci Educ. 2026, 10.1002/ase.70292
62. Kumar A, Burr P, Young TM: Using AI text-to-image generation to create novel illustrations for medical education: current limitations as illustrated by hypothyroidism and Horner syndrome. JMIR Med Educ. 2024, 10:e52155. 10.2196/52155
63. Currie G, Currie J, Anderson S, Hewis J: Gender bias in generative artificial intelligence text-to-image depiction of medical students. Health Educ J. 2024, 83:732-746. 10.1177/00178969241274621
64. Bui M, Yi PH, Onuh I, Kuckelman I, Ross AB: Portrayal of medical students in artificial intelligence-generated images. WMJ. 2026, 125:158-161.
65. Shaari AL, Saad AM, Patel AM, Filimonov A: Representation of demographics in otolaryngology by artificial intelligence text-to-image platforms. Laryngoscope Investig Otolaryngol. 2025, 10:e70152. 10.1002/lio2.70152
66. Currie G, Hewis J, Hawk E, Rohren E: Gender and ethnicity bias of text-to-image generative artificial intelligence in medical imaging, part 2: analysis of DALL-E 3. J Nucl Med Technol. 2025, 53:162-168. 10.2967/jnmt.124.268359
67. Ali R, Tang OY, Connolly ID, et al.: Demographic representation in 3 leading artificial intelligence text-to-image generators. JAMA Surg. 2024, 159:87-95. 10.1001/jamasurg.2023.5695
68. Senft Everson N, Gaysynsky A, Iles IA, Schrader KE, Chou WS: What does an AI-generated "cancer survivor" look like? An analysis of images generated by text-to-image tools. J Cancer Surviv. 2026, 20:1612-1621. 10.1007/s11764-025-01760-1
69. Mezil Y, Yu P, Rahimpour L, et al.: Harnessing artificial intelligence for automatic feedback in a virtual anatomy study tool: a Q-methodology study. Anat Sci Educ. 2026, 19:684-694. 10.1002/ase.70206
70. Al-Mashhadani MM, Ajaz F, Guraya SS, Ennab F: Use of large language models for providing automated feedback in medical imaging education: a systematic review. Front Med (Lausanne). 2026, 13:1803921. 10.3389/fmed.2026.1803921
71. Ahmed A, Kerr E, O'Malley A: Quality assurance and validity of AI-generated single best answer questions . BMC Med Educ. 2025, 25:300. 10.1186/s12909-025-06881-w
72. Association of American Medical Colleges: principles for the responsible use of artificial intelligence in and for medical education. (2025). Accessed: August 24, 2026: https://www.aamc.org/about-us/mission-areas/medical-education/principles-ai-use.
73. Izquierdo-Condoy JS, Arias-Intriago M, Montero Corrales L, Ortiz-Prado E: Artificial intelligence in medical education: transformative potential, current applications, and future implications. JMIR Med Educ. 2026, 12:e77127. 10.2196/77127
74. Ganguly A, Johri A, Ali A, McDonald N: Generative artificial intelligence for academic research: evidence from guidance issued for researchers by higher education institutions in the United States. AI Ethics. 2025, 5:3917-3933. 10.1007/s43681-025-00688-7
75. ICMJE: Use of artificial intelligence in publishing . (2026). Accessed: August 24, 2026: https://www.icmje.org/recommendations/browse/artificial-intelligence/.
76. Alfredo R, Echeverria V, Jin Y, Yan L, Swiecki Z, Gašević D, Martinez-Maldonado R: Human-centred learning analytics and AI in education: a systematic literature review. Comput Educ Artif Intell. 2024, 6:100215. 10.1016/j.caeai.2024.100215