Regulatory-Ready AI in Pharmaceutical Sciences: An Umbrella Review of Explainability, Reproducibility, Bias, Validation, and Governance Readiness

Daniel Kim[1]*, Jihoon Park[1], Min-seo Kang[2]

[1]Department of AI for Red Ginseng and Lignan Bioactivity, College of Pharmacy, Kyung Hee University, Seoul, South Korea.
[2]Department of Pharmacoinformatics for Viral Protease Inhibitors, College of Pharmacy, Chung-Ang University, Seoul, South Korea.

ABSTRACT

Artificial intelligence is increasingly applied across drug discovery, pharmaceutical development, clinical pharmacology, pharmaceutics, pharmacovigilance, and related decision-support contexts, yet high predictive performance alone does not establish that an AI system is suitable for regulated or safety-relevant use. This umbrella review synthesized review-level evidence concerning the explainability, reproducibility, bias control, validation, documentation, lifecycle oversight, and governance conditions that may support regulatory-readiness assessment in pharmaceutical sciences. PubMed/MEDLINE, Scopus-indexed and Web of Science-indexed records, ScienceDirect, SpringerLink, Wiley Online Library, Oxford Academic journals, IEEE Xplore, Nature Portfolio journals, Frontiers journals, and MDPI journals were searched for publications dated from 1 January 2017 to 11 July 2026. After database searching, citation chasing, deduplication, title-and-abstract screening, and full-text eligibility assessment, the final evidence corpus contained 37 publications, comprising 31 review-level syntheses and six essential methodological or conceptual anchors. Evidence was extracted by regulatory-readiness domain and synthesized thematically, with methodological limitations appraised narratively and review overlap considered qualitatively. Review-level evidence converged on the need for purpose-specific explainability, transparent data provenance, reproducible workflows, explicit documentation, representative datasets, bias and fairness assessment, leakage-resistant internal validation, independent external validation, prospective evaluation where relevant, uncertainty reporting, human oversight, auditability, lifecycle monitoring, and controlled model updating. However, terminology, evaluation methods, reporting practices, validation depth, and governance implementation were inconsistent. External, prospective, clinical, and long-term lifecycle evidence remained less developed than retrospective performance evidence. Regulatory-ready pharmaceutical AI requires an integrated evidence and governance architecture rather than any single technical feature. Explainability, reproducibility, bias awareness, validation, documentation, oversight, and continuous monitoring may strengthen readiness assessment, but review-level evidence does not itself establish regulatory approval, clinical utility, autonomous decision-making suitability, or deployment readiness.

Key Words: *Regulatory-ready AI, Pharmaceutical sciences, Umbrella review, Explainability, Reproducibility, Bias*

eIJPPR 2026; 16(1):19-34

HOW TO CITE THIS ARTICLE: Kim D, Park J, Kang MS. Regulatory-Ready AI in Pharmaceutical Sciences: An Umbrella Review of Explainability, Reproducibility, Bias, Validation, and Governance Readiness. Int J Pharm Phytopharmacol Res. 2026;16(1):19-34. https://doi.org/10.51847/wpEOcJWd45

INTRODUCTION

Artificial intelligence and machine-learning methods are now used across target identification, molecular representation, virtual screening, property prediction, lead optimization, clinical-development planning, trial design, and evidence generation. This expansion has created opportunities to process complex chemical, biological, pharmacological, and clinical data at scales that are

Corresponding author: Daniel Kim
Address: Department of AI for Red Ginseng and Lignan Bioactivity, College of Pharmacy, Kyung Hee University, Seoul, South Korea.
E-mail: ✉ daniel.kim@khu.ac.kr
Received: 10 October 2025; **Revised:** 12 January 2026; **Accepted:** 16 January 2026

difficult to manage through conventional analytical approaches. At the same time, the breadth of these applications means that pharmaceutical AI cannot be evaluated through a single performance criterion or a uniform readiness standard. Drug-discovery models, clinical-trial tools, dosing systems, manufacturing applications, and pharmacovigilance algorithms operate within different evidentiary contexts and produce different potential consequences when their outputs are inaccurate, unstable, non-generalizable, or misunderstood [1-3].

The pharmaceutical AI literature has frequently emphasized improvements in computational efficiency, pattern recognition, predictive ranking, and hypothesis generation. Nevertheless, review-level analyses also identify persistent limitations involving heterogeneous datasets, incomplete reporting, uncertain data provenance, limited interpretability, insufficient independent validation, and weak translation from retrospective computational evaluation to operational evidence. These limitations are especially important when models are expected to influence candidate prioritization, trial design, dose selection, safety surveillance, manufacturing decisions, or other activities that may affect the quality of pharmaceutical evidence. Consequently, technological advancement should not be interpreted as equivalent to methodological maturity, and methodological maturity should not be interpreted as equivalent to regulatory acceptance [4].

Regulatory readiness is therefore better understood as a staged evidentiary and organizational condition than as a binary characteristic of an algorithm. An AI system may be technically functional while remaining inadequately documented, difficult to reproduce, vulnerable to data leakage, insufficiently evaluated across relevant populations or settings, or unsupported by an effective governance structure. Explainability may improve the inspectability of model behavior, but it does not establish that predictions are valid. Reproducibility may demonstrate that an analysis can be repeated, but it does not prove that the model generalizes. Bias assessment can reveal risks without demonstrating that all inequities have been removed. Similarly, governance principles may describe desirable responsibilities without showing that those responsibilities have been implemented, resourced, monitored, and enforced [1].

These distinctions create a need to synthesize evidence above the level of individual applications. The objective of this umbrella review was to examine how existing reviews characterize the evidence requirements and unresolved limitations associated with regulatory-ready AI in pharmaceutical sciences. The review focused on explainability, interpretability, transparency, reproducibility, replicability, data provenance, documentation, bias, fairness, representativeness, validation, uncertainty, lifecycle monitoring, auditability, human oversight, risk management, governance readiness, and regulatory alignment. It further aimed to distinguish review-level convergence from implementation evidence and to define boundaries preventing readiness concepts from being misrepresented as regulatory approval or deployment authorization.

Regulatory-ready AI in pharmaceutical sciences

Regulatory-ready AI in pharmaceutical sciences can be defined as AI for which the intended use, evidence basis, data lineage, model behavior, performance limitations, validation strategy, accountability structure, and lifecycle controls are sufficiently characterized to permit a structured readiness assessment. This definition does not imply that a model has been reviewed, accepted, authorized, or approved by a regulatory authority. It instead describes the condition in which evidence is organized, traceable, and proportionate to the consequences of the proposed use. In pharmaceutics and drug-delivery design, for example, AI models may be used to support formulation selection, process optimization, material characterization, release prediction, manufacturing control, or delivery-system design. Readiness in these settings depends not merely on prediction accuracy but also on formulation-data quality, reproducible preprocessing, technically meaningful validation, version control, documentation, and integration with pharmaceutical expertise [5].

Clinical pharmacology illustrates why readiness requirements must be aligned with decision consequences. Machine-learning models developed for dose individualization may combine laboratory data, patient characteristics, treatment history, and dynamic response measurements, but apparent performance within a development dataset does not establish transportability to another institution, population, clinical workflow, or treatment protocol. A systematic review of personalized heparin-dosing models found a small and heterogeneous evidence base, with limited independent validation and substantial variation in modeling and reporting practices. Such findings support a distinction between a model that predicts within a retrospective dataset and a model supported by evidence adequate for prospective clinical decision support [6].

Pharmacovigilance presents a related but operationally distinct readiness challenge. Machine-learning systems may assist with case intake, coding, duplicate detection, literature screening, signal prioritization, causality-related classification, or adverse-event identification. However, these activities occur within safety systems in which missed cases, false prioritization, automation bias, or

poorly controlled model updates can affect the interpretation and management of safety information. Review-level evidence indicates that much pharmacovigilance AI research remains retrospective, task-specific, and dependent on heterogeneous datasets. Explainability, sustained workflow evaluation, human-review boundaries, error escalation, lifecycle monitoring, and evidence from continued operational use are less consistently reported than technical performance [7].

Across pharmaceutical regulatory affairs, readiness consequently requires an evidence chain connecting data governance, model development, validation, intended use, documented limitations, human accountability, change management, and continuing oversight. Transparent data sources and preprocessing steps are needed to reconstruct how outputs were produced. Model documentation must clarify users, inputs, outputs, exclusions, known failure modes, and conditions requiring escalation. Validation should reflect the proposed context of use rather than an abstract benchmark alone, while auditability should permit reconstruction of model versions, decisions, interventions, and updates. These components may support regulatory alignment and evidence review, but none independently constitutes regulatory approval [8].

Umbrella review scope

The umbrella review was designed to answer six questions: what review-level evidence exists on explainability and interpretability; how reviews address reproducibility, transparency, data provenance, and documentation; how bias, fairness, representativeness, and generalizability are treated; which internal, external, prospective, clinical, technical, and lifecycle validation approaches are reported; how governance readiness, human oversight, auditability, model drift, and regulatory alignment are addressed; and which evidence gaps prevent readiness features from supporting stronger implementation claims. These questions were framed to capture evidence across drug discovery, computational pharmacology, pharmaceutics, clinical pharmacology, pharmacovigilance, biomedical AI, and pharmaceutical regulatory science while maintaining the review-level focus required for an umbrella synthesis [9].

Searches were completed on 11 July 2026 and covered publications from 1 January 2017 through the search date. Information sources included PubMed/MEDLINE; Scopus-indexed and Web of Science-indexed records accessible through bibliographic, publisher, and cited-reference interfaces; ScienceDirect; SpringerLink; Wiley Online Library; Oxford Academic journals; IEEE Xplore; Nature Portfolio journals; Frontiers journals; and MDPI journals. Search strings combined terms for artificial intelligence, machine learning, and deep learning with pharmaceutical sciences, drug discovery, pharmacology, pharmaceutics, pharmacovigilance, clinical pharmacology, or regulatory science and with review-design and readiness concepts such as explainability, reproducibility, bias, validation, transparency, governance, auditability, monitoring, and regulatory readiness. Focused searches were also conducted for explainable drug-discovery AI, pharmacological validation, external validation, data leakage, algorithmic bias, pharmacovigilance governance, responsible AI, model drift, and pharmaceutical quality-system applications. Methodological guidance for overview synthesis emphasizes that heterogeneous review questions and evidence types require explicit search, appraisal, and synthesis decisions rather than an assumption that all included reviews are methodologically interchangeable [10].

Eligible publications were peer-reviewed journal articles with a DOI, published within the defined period, and classified as systematic reviews, scoping reviews, mapping reviews, critical reviews, methodological reviews, high-quality narrative reviews, umbrella-method articles, or regulatory-science syntheses. Publications had to address an eligible pharmaceutical or transferable biomedical AI domain and contribute evidence to at least one regulatory-readiness dimension. Primary empirical studies were excluded unless they served an essential methodological or conceptual function that could not be adequately supported through review literature alone. Books, theses, websites, software documentation, conference-only papers, preprints, product materials, regulatory webpages, policy documents, guidance documents, white papers, records without DOIs, application inventories lacking readiness relevance, and articles with insufficient methodological detail were excluded. Review overlap was considered by examining domain, publication period, recurring landmark studies, datasets, and likely repetition of influential primary evidence, but no quantitative overlap statistic was calculated [11].

The screening ledger contained 68 records identified through database and journal-platform searching and eight records identified through backward or forward citation chasing, producing 76 records before deduplication. Twenty-two duplicate instances were removed, leaving 54 unique records for title-and-abstract screening. Six records were excluded at this stage, and 48 full-text articles were assessed for eligibility. Eleven full-text articles were excluded: three were not review-level publications, two lacked sufficient pharmaceutical or transferable biomedical readiness relevance, two lacked a substantive explainability, reproducibility, bias, validation, or governance focus, one was conference-only or preprint-only, one had no eligible DOI-bearing journal version, one

provided insufficient methodological detail, and one substantially overlapped with a stronger eligible report. The resulting corpus contained 37 publications: 31 review-level syntheses and six essential methodological or conceptual anchors.

For each eligible publication, information was charted on review aim, review type, pharmaceutical or AI domain, described evidence base, explainability findings, reproducibility and transparency findings, bias and fairness findings, validation findings, governance implications, limitations, likely overlap, and contribution to the umbrella synthesis. Review quality was appraised narratively across review-type clarity, search transparency, eligibility criteria, extraction methods, quality appraisal, reporting completeness, domain relevance, recency, validation depth, governance depth, regulatory-readiness specificity, conflict-of-interest reporting, and limitations reporting. Formal AMSTAR 2 or ROBIS scores were not generated. Evidence was synthesized thematically by readiness domain, and convergence was interpreted as consistency across review-level conclusions rather than as a count of statistically independent findings. **Table 1** summarises the umbrella review scope, search strategy, eligibility criteria, screening numbers, review-level extraction domains, appraisal logic, and synthesis approach.

Table 1. Umbrella Review Scope for Regulatory-Ready AI in Pharmaceutical Sciences: Search Strategy, Eligibility Criteria, PRISMA-Style Screening Numbers, Review-Level Evidence Domains, Quality and Overlap Appraisal, and Synthesis Boundaries

Review component	Specification	Exact Part 1 result	Purpose	Evidence captured	Interpretation boundary
Review design	Umbrella review of review-level and essential methodological evidence	Original umbrella review; not a systematic review of primary studies, scoping review, or meta-analysis	Synthesize evidence above the level of individual AI applications	Review-level convergence, divergence, and gaps	Does not estimate pooled model effects or treatment outcomes
Review questions	Six questions covering explanation, reproducibility, bias, validation, governance, and evidence gaps	Questions addressed explainability; transparency, provenance, and documentation; bias and fairness; validation; governance and monitoring; and readiness limitations	Maintain alignment between search, extraction, and synthesis	Regulatory-readiness dimensions across pharmaceutical AI	Questions do not determine regulatory acceptance
Search date	Final search date	11 July 2026	Fix the temporal boundary of the evidence search	Publications available by the search date	Later publications are not represented
Search period	Eligible publication interval	1 January 2017–11 July 2026	Capture contemporary pharmaceutical and biomedical AI review evidence	Peer-reviewed journal publications in the specified period	Publication date does not establish methodological quality
Information sources	Bibliographic databases, indexes, and journal platforms	PubMed/MEDLINE; Scopus-indexed and Web of Science-indexed records; ScienceDirect; SpringerLink; Wiley Online Library; Oxford Academic journals; IEEE Xplore; Nature Portfolio; Frontiers; and MDPI journals	Maximize multidisciplinary coverage	Pharmaceutical, biomedical, informatics, methodological, and governance literature	No claim is made that inaccessible proprietary bulk exports were obtained
Core search logic	AI terms combined with pharmaceutical	Search combinations included AI, machine learning, deep learning,	Identify review-level literature	Broad and focused readiness evidence	Search relevance does not

ISSN (Online) 2249-6084 (Print) 2250-1029

	domains, review terms, and readiness concepts	pharmaceutical sciences, drug discovery, pharmacology, pharmaceutics, pharmacovigilance, clinical pharmacology, regulatory science, review, explainability, reproducibility, bias, validation, transparency, governance, auditability, and monitoring	directly relevant to readiness		guarantee eligibility
Focused searches	Supplementary topic combinations	Explainable drug-discovery AI; pharmacological validation; external validation; data leakage; AI bias; pharmacovigilance governance; responsible AI; model drift; auditability; and pharmaceutical quality systems	Improve retrieval of narrow readiness topics	Specialized methodological and governance evidence	Focused searching may identify overlapping records
Inclusion criteria	Peer-reviewed DOI-bearing journal publication within period and relevant to at least one readiness domain	Systematic, scoping, mapping, critical, methodological, high-quality narrative, umbrella-method, and regulatory-science publications were eligible	Preserve review-level emphasis while retaining essential anchors	Review scope, findings, limitations, validation, and governance implications	Inclusion does not certify the quality of the underlying primary studies
Exclusion criteria	Ineligible date, type, evidence level, DOI status, domain, or readiness relevance	Excluded books, theses, websites, software pages, conference-only papers, preprints, product materials, regulatory webpages, policy documents, guidance documents, white papers, unsuitable primary studies, and weak-fit reviews	Prevent source-type and scope drift	Records capable of supporting mapped manuscript claims	Exclusion is methodological, not a judgment that the topic lacks value
Database/platform records	Records identified through structured searching	68	Establish the primary identification count	Bibliographic and journal-platform records	Count refers to the working screening ledger, not unfiltered search-engine hits
Other-source records	Records identified through citation chasing and related-record searches	8	Supplement database retrieval	Backward and forward citation-linked publications	Citation chasing can increase overlap

Total before deduplication	Combined identification total	76	Provide the starting PRISMA-style count	All entered record instances	Includes duplicate instances
Duplicate removal	Duplicate record instances removed	22	Prevent repeated screening of the same publication	Duplicate titles, DOIs, and bibliographic records	Duplicate count does not represent excluded unique studies
Records after deduplication	Unique records entering title-and-abstract screening	54	Define the screening denominator	Unique potentially eligible records	Uniqueness was assessed bibliographically
Title-and-abstract screening	Records screened and excluded	54 screened; 6 excluded	Remove clearly irrelevant records before full-text assessment	Domain, article type, and readiness relevance	Abstract-level exclusion was limited to clear ineligibility
Full-text eligibility	Full texts assessed	48	Apply complete eligibility criteria	Review design, DOI, methods, domain, and readiness contribution	Full-text access does not ensure inclusion
Full-text exclusions	Excluded full texts with counted reasons	11 total: not review level, 3; insufficient domain relevance, 2; no substantive readiness focus, 2; conference/preprint only, 1; no eligible DOI-bearing version, 1; insufficient methodological detail, 1; overlapping report, 1	Provide transparent reasons for final exclusions	Specific eligibility failures	Counts were fixed before manuscript drafting
Final evidence corpus	Publications retained after eligibility assessment	37 publications: 31 review-level syntheses and 6 essential methodological or conceptual anchors	Define the evidence available for synthesis	Pharmaceutical, biomedical, methodological, validation, and governance literature	The anchors do not convert the review into a primary-study systematic review
Extraction domains	Structured evidence-charting fields	Aim, review type, domain, evidence base, explanation, reproducibility, bias, validation, governance, limitations, overlap, and synthesis contribution	Ensure claim-to-source traceability	Findings and limitations actually reported in included publications	Unreported categories were recorded as not clearly reported
Quality appraisal	Narrative methodological appraisal	No formal AMSTAR 2 or ROBIS scores generated	Evaluate confidence without inventing scores	Search, eligibility, extraction, quality appraisal, reporting, relevance, recency, validation, governance, conflicts, and limitations	Narrative appraisal is not equivalent to a validated numerical rating

Overlap appraisal	Qualitative assessment of likely repeated primary evidence	Considered by domain, period, recurring studies, datasets, and landmark examples	Avoid treating repeated review conclusions as fully independent evidence	High-overlap clusters in drug discovery, XAI, validation, bias, and governance	No corrected covered area or quantitative overlap statistic was calculated
Synthesis method	Thematic umbrella synthesis by regulatory-readiness domain	Convergent findings, divergent interpretations, gaps, governance needs, research priorities, and confidence boundaries were mapped	Integrate heterogeneous review evidence without inappropriate pooling	Domain-level evidence patterns and interpretation boundaries	Convergence does not prove implementation effectiveness or approval
Registration and quantitative synthesis	Registration and meta-analysis status	No registration number claimed; no meta-analysis performed	Prevent unsupported methodological claims	Transparent statement of review limits	No pooled effect estimates or fabricated registration details
Regulatory interpretation	Readiness treated as staged evidence and governance condition	Regulatory alignment, readiness, implementation, deployment, clinical utility, and approval were kept distinct	Prevent overstatement	Evidence supporting reviewability, validation, oversight, and monitoring	No included synthesis establishes authority endorsement or regulatory approval

Explainability and reproducibility evidence

The explainability literature consistently indicates that explanation is not a single technical property and should not be evaluated independently of its intended purpose. In drug discovery, explanations may be used to inspect learned molecular features, evaluate whether a prediction is chemically or biologically plausible, identify potential shortcuts, communicate model reasoning to domain specialists, or support hypothesis generation. Broader medical-AI reviews similarly distinguish between explanations required by developers, scientists, clinicians, quality personnel, affected individuals, auditors, and governance bodies. These audiences do not necessarily require the same information, level of detail, or explanation method. Technical surveys also demonstrate substantial heterogeneity among feature-attribution, rule-based, example-based, counterfactual, visualization, attention, and intrinsically interpretable approaches. Accordingly, review-level evidence supports purpose-specific explanation requirements rather than a universal explainability threshold [12-14].

Explanation quality depends on more than visual coherence or intuitive appeal. Technical reviews emphasize fidelity to actual model behavior, stability under small input changes, robustness across plausible perturbations, sensitivity to relevant features, consistency with task structure, and usefulness for the intended evaluator. In pharmaceutical applications, these properties should be interpreted in relation to chemical representations, assay conditions, molecular scaffolds, pharmacological mechanisms, formulation variables, or patient-level covariates. An explanation that appears scientifically plausible can still be unfaithful to the model, while a faithful explanation may expose that the model relies on an undesirable artifact. Explanation methods therefore require evaluation as analytical objects rather than presentation layers added after predictive performance has been established [15].

Critical review evidence further cautions that post-hoc explanation can create false reassurance when it is treated as a substitute for validation. A heat map, feature ranking, molecular substructure highlight, or counterfactual example may help users explore a model, but it cannot demonstrate that the underlying dataset was representative, that preprocessing was leakage-free, that the prediction is calibrated, that performance transfers to an independent setting, or that the model improves an intended pharmaceutical or clinical process. Explanations may also be unstable, incomplete, sensitive to modeling choices, or susceptible to persuasive presentation. Explainability should therefore be positioned as one component of reviewability and human oversight, while claims regarding reliability, generalizability, safety, or utility require separate evidence [16].

Reproducibility evidence reveals parallel weaknesses in the availability and reporting of data, code, model configurations, preprocessing procedures, random seeds, software dependencies, computational environments, hyperparameter selection, and evaluation logic. A systematic assessment of health-related machine-learning research found that the materials needed to independently reproduce analyses were frequently incomplete or

inaccessible, although privacy, intellectual-property, and data-governance restrictions may legitimately limit unrestricted sharing [17]. Methodological guidance consequently recommends a reproducibility package that describes data provenance, inclusion and exclusion logic, preprocessing order, model versions, training procedures, evaluation splits, uncertainty measures, and the conditions under which an analysis can be rerun or independently replicated [18]. Particular attention is required for data leakage because information transferred improperly between training, tuning, and testing stages can inflate apparent performance while preserving the superficial appearance of rigorous validation. Leakage may arise through duplicated observations, preprocessing performed before dataset separation, temporally inappropriate variables, non-independent molecular structures, outcome-derived features, repeated patients, or benchmark contamination. Detecting and documenting these risks is therefore a prerequisite for credible internal validation and subsequent regulatory-readiness assessment [19].

Bias, validation, and governance readiness

Review-level evidence characterizes bias as a lifecycle and socio-technical problem rather than an isolated defect that can be eliminated through a single fairness metric. Bias may arise from historical data-generating processes, sampling decisions, eligibility rules, measurement practices, outcome definitions, missingness, labeling, molecular or patient representation, model optimization, performance evaluation, workflow integration, and feedback after implementation. Open and transparent scientific practices may improve scrutiny of these sources, while subgroup analysis can reveal unequal behavior across relevant populations or contexts. Nevertheless, transparency alone does not remove bias, and a model that satisfies one statistical fairness criterion may remain unsuitable for another pharmaceutical or clinical context. Bias assessment must therefore define the affected groups, plausible harms, intended use, data limitations, and consequences of false-positive and false-negative outputs [20-22].

Validation evidence showed a recurrent disparity between strong retrospective performance claims and the depth of evidence required for intended-use conclusions. Reviews of comparative clinical AI studies identified weaknesses in study design, reporting completeness, comparator selection, and the language used to describe model superiority. Validation was sometimes reported without a clear distinction between resampling within a development dataset and testing on an independent population, site, time period, assay environment, compound series, or clinical workflow. Such ambiguity can support inflated interpretations because apparent accuracy does not reveal whether data leakage was prevented, hyperparameters were selected independently, calibration was adequate, relevant subgroups were evaluated, or the comparator reflected actual practice. Claims of readiness should therefore state exactly which validation stage has been completed and which stages remain unsupported [23].

External validation was less common and less consistently independent than internal evaluation. Systematic evidence from clinical AI comparisons showed that studies reporting high performance frequently lacked prospective designs, independently sourced validation datasets, transparent participant selection, or reporting adequate to judge generalizability. Although much of this evidence arose from medical imaging, the methodological lesson transfers cautiously to pharmaceutical AI: a model tested on data derived from the same institution, database, assay family, molecular scaffold distribution, or development pipeline may not demonstrate transportability to materially different conditions. External validation can strengthen evidence for generalizability, but it does not independently establish workflow benefit, clinical utility, manufacturing suitability, or acceptable risk [24].

Governance readiness begins where model evaluation becomes connected to organizational responsibility. Reviews of implementation barriers emphasize that real-world impact depends on workflow fit, prospective assessment, monitoring, calibration, human–AI interaction, and mechanisms for responding when performance deteriorates [25]. Responsible-AI syntheses further indicate that technical controls must be coordinated with professional, organizational, ethical, and societal responsibilities rather than managed as a software-only exercise [26]. Governance models consequently encompass data stewardship, development controls, validation review, user competence, decision authority, auditability, incident response, change management, and continuing oversight [27]. Recent framework syntheses show substantial convergence around transparency, accountability, equity, safety, and monitoring, but they also reveal differences in operational detail, enforcement mechanisms, resources, and implementation maturity. Governance principles may therefore support readiness mapping, while evidence of governance implementation requires documented responsibilities, functioning controls, review processes, and lifecycle records [28].

Review-level evidence synthesis

Across drug-discovery reviews, the dominant pattern was rapid methodological expansion accompanied by slower development of standardized evaluation and translational evidence. Deep-learning approaches have been applied to molecular representation, target prediction, virtual screening, property estimation, lead optimization, and

related discovery tasks, but reviews repeatedly identify inconsistent benchmarks, limited comparability, variable data quality, narrow training distributions, and insufficient experimental or independent validation. These findings indicate that innovation in model architecture and predictive performance has advanced more quickly than shared evidence standards for determining whether a model is reliable outside its original development conditions [29].

Comprehensive reviews of computer-assisted drug discovery reinforce the interdependence of data, representation, modeling, and validation. Model outputs depend on how chemical and biological entities are encoded, how datasets are curated, which negative and positive examples are included, how missing or uncertain labels are handled, and whether evaluation partitions reflect the intended generalization problem. Random splitting may be inadequate when a model is expected to generalize across scaffolds, targets, assay platforms, laboratories, or development programs. Data provenance, benchmark design, uncertainty characterization, and chemically meaningful interpretation are therefore not secondary reporting details; they are part of the evidence needed to understand what a model has learned and where its predictions may fail [30].

The cross-review synthesis also revealed that interpretability, documentation, reproducibility, and validation gaps are not confined to one AI method or pharmaceutical task. Reviews covering classical machine learning, deep learning, and hybrid computational approaches describe recurring dependence on small, imbalanced, proprietary, heterogeneous, or weakly annotated datasets. Comparisons may be complicated by inconsistent endpoints, preprocessing pipelines, performance measures, and validation designs. Furthermore, the availability of an explanation method does not guarantee faithful scientific reasoning, and the availability of code does not guarantee that data or computational environments can be reconstructed. Regulatory-readiness assessment must therefore examine the integrity of the complete evidence chain rather than use one favorable feature as a proxy for overall credibility [31].

Clinical-pharmacology and pharmacokinetic applications demonstrate the need to align validation with the consequences of use. Predictive models may support exposure estimation, dose optimization, treatment-response assessment, trial simulation, or personalized pharmacotherapy, but review-level evidence emphasizes that biological plausibility, calibration, representative populations, independent testing, and prospective evaluation remain essential when outputs could influence patient-level decisions [32]. Broader clinical-pharmacology reviews similarly stress equitable evaluation, transparent reporting, integration with pharmacological reasoning, and clear boundaries between computational decision support and accountable professional judgment [33]. In pharmacovigilance, systematic evidence remains methodologically heterogeneous and predominantly retrospective, limiting conclusions about sustained operational performance, drift, error recovery, or the effectiveness of human oversight during routine safety surveillance [34].

Taken together, the included reviews support a staged interpretation of regulatory readiness. Evidence foundations include representative data, provenance, documentation, version control, and reproducible workflows. Model-level reviewability includes interpretability, explainability, uncertainty communication, bias analysis, and defined limitations. Validation progresses from leakage-resistant internal testing toward independent external, prospective, clinical, analytical, or operational evaluation as appropriate to the intended use. Governance then connects these technical elements to accountability, auditability, human oversight, risk management, controlled updating, incident response, and lifecycle monitoring. These stages are mutually reinforcing, but none alone establishes regulatory approval. **Table 2** presents the review-level synthesis of regulatory-readiness evidence across explainability, reproducibility, bias, validation, and governance domains.

Table 2. Review-Level Evidence Synthesis for Regulatory-Ready AI in Pharmaceutical Sciences: Explainability, Reproducibility, Bias, Validation, Governance Readiness, Lifecycle Monitoring, Human Oversight, Regulatory Alignment, Evidence Gaps, and Research Priorities

Regulatory-readiness domain	Review-level evidence pattern	Commonly reported requirement	Commonly reported gap	Regulatory relevance	Governance need	Research priority	Confidence boundary
Explainability	Widely recommended, but definitions, purposes, audiences, and	Define who needs an explanation, for which decision, at what level, and with	Post-hoc explanations may be unstable, persuasive, or weakly evaluated	Can improve reviewability and scientific scrutiny but does not	Assign responsibility for selecting, testing, documenting, and reviewing	Develop pharmaceutical task-specific explanation benchmarks and user-	Qualified review-level confidence; limited implementation evidence

	evaluation methods vary	which fidelity criteria		establish validity	explanation methods	evaluation methods	
Interpretability	Simpler or domain-aligned models may facilitate inspection and scientific reasoning	Distinguish intrinsic interpretability from post-hoc explanation and relate requirements to intended use	Terminology is inconsistent, and complexity–performance trade-offs are often assumed	May support inspectability but cannot replace independent validation	Define when interpretable modeling is required and who evaluates adequacy	Compare interpretable and complex models under matched pharmaceutical tasks and evidence standards	Qualified confidence
Transparency	Reviews converge on reporting data, methods, intended use, limitations, and changes	Disclose data sources, preprocessing, architecture, training logic, performance, users, exclusions, and modifications	Proprietary restrictions and incomplete reporting remain common	Supports evidence reconstruction, peer scrutiny, and inspection readiness	Establish disclosure rules, access controls, documentation ownership, and review cycles	Define minimum transparency packages by pharmaceutical risk and use context	High principle-level confidence; operational standards remain variable
Reproducibility and replicability	Code, data, software, seeds, environments, and preprocessing details are often incomplete	Preserve executable workflows, model versions, dependency information, evaluation code, and reproducibility records	Independent reproduction is uncommon, and data access may be constrained	Strengthens evidence credibility but does not independently establish validity or utility	Maintain controlled repositories, access procedures, version records, and independent review	Conduct independent multi-team reproduction and replication studies	High confidence that reporting gaps exist; lower confidence in universal implementation models
Data provenance	Provenance is repeatedly connected to traceability, bias analysis, and change control	Record data origin, authority, collection context, curation, preprocessing, linkage, exclusions, and versions	Lineage information is often fragmented or detached from model records	Foundational for auditability and defensible reconstruction of evidence	Assign data stewardship, retention, lineage, access, and change responsibilities	Develop interoperable provenance standards for pharmaceutical datasets	Moderate confidence; limited evaluated implementation evidence
Documentation	Documentation is consistently treated as a prerequisite for accountable review	Define intended use, inputs, outputs, users, assumptions, limitations, failure modes, validation status, and changes	Minimum documentation content and formats vary	Supports readiness assessment and controlled review but does not prove system quality	Assign document owners, approval authority, update cycles, and archival controls	Test standardized model and data documentation packages in pharmaceutical settings	High principle-level confidence

ISSN (Online) 2249-6084 (Print) 2250-1029

Bias assessment	Bias may originate throughout data generation, model development, evaluation, and use	Predefine relevant groups, data limitations, potential harms, subgroup analyses, and remediation procedures	Assessments are often narrow, post hoc, or restricted by missing demographic and contextual data	Supports risk identification; cannot demonstrate that all bias has been eliminated	Establish recurring bias review, escalation thresholds, corrective action, and documentation	Evaluate bias longitudinally across sites, populations, compounds, assays, and model updates	High confidence that bias is lifecycle-based; limited evidence on remedy effectiveness
Fairness and representativeness	Fairness is context-dependent; narrow datasets limit transportability	Define fairness objectives and characterize demographic, geographic, temporal, molecular, assay, and practice coverage	No universal fairness metric; underrepresented groups and rare events remain poorly assessed	Necessary for defining the valid scope of performance claims	Include domain experts and affected stakeholders in fairness and representativeness decisions	Compare fairness definitions against patient-safety and pharmaceutical decision consequences	Qualified confidence
Internal validation	Commonly reported but vulnerable to leakage, optimistic tuning, and inappropriate partitioning	Use nested tuning, leakage-resistant preprocessing, calibration, and splits aligned with intended generalization	Random splits may fail to test site, time, scaffold, population, or workflow transfer	Supports development-stage evidence only	Require independent review of preprocessing, split logic, and evaluation code	Compare validation designs that reproduce realistic pharmaceutical use conditions	High confidence
External validation	Less frequent and not always materially independent	Test on independent populations, sites, time periods, compounds, assays, laboratories, or workflows	Independent datasets may be unavailable, narrow, or poorly characterized	Necessary for stronger generalizability claims but not sufficient for clinical or operational utility	Define independence criteria, acceptance thresholds, and failure-analysis procedures	Conduct multi-site, temporally separated, and cross-platform validation studies	High confidence in the requirement; limited completed evidence
Prospective and clinical validation	Prospective evaluation is uncommon, and predictive performance is often conflated with utility	Predefine intended-use outcomes, workflow consequences, human interaction, benefit, harm, and stopping criteria	Retrospective studies dominate; few reviews identify mature prospective evidence	Stronger support for intended-use performance but not equivalent to approval	Establish study oversight, monitoring, incident handling, and protocol governance	Conduct silent prospective studies, pragmatic evaluations, and use-specific clinical validation where appropriate	High confidence that evidence is sparse; applicability varies by use case
Uncertainty reporting	Less consistently reported than	Report calibration, confidence or	Point estimates may be	Supports risk-sensitive	Define uncertainty thresholds,	Evaluate uncertainty under	Qualified confidence

	discrimination or ranking performance	prediction intervals, out-of-distribution behavior, and abstention conditions	presented without actionable uncertainty	decisions and escalation but does not guarantee correct predictions	review pathways, and conditions for non-use	distribution shift, rare events, and incomplete data	
Model drift and lifecycle monitoring	Widely recognized as necessary but rarely supported by mature long-term studies	Monitor input distributions, calibration, subgroup behavior, errors, workflow changes, and outcomes	Drift thresholds, monitoring intervals, and corrective procedures are not standardized	Pre-deployment validation alone is insufficient for repeatedly used or updated models	Assign monitoring ownership, alert review, retraining criteria, rollback authority, and periodic reassessment	Conduct longitudinal studies of drift detection and lifecycle assurance in operational settings	High principle-level confidence; low implementation confidence
Auditability and traceability	Reviews emphasize reconstructable data, model, decision, and change histories	Maintain versions, access logs, approvals, validation records, overrides, incidents, and update histories	Audit packages are usually described conceptually rather than evaluated empirically	Supports accountability and inspection but does not independently establish quality	Implement secure records, review procedures, retention policies, and escalation mechanisms	Define and test risk-proportionate audit packages for pharmaceutical AI	Moderate confidence
Human oversight	Frequently recommended, although responsibilities are often vague	Define authority, competence, review triggers, override rights, escalation, and accountability	Human involvement may be nominal and vulnerable to automation bias or workload pressures	Supports controlled decision assistance but does not guarantee safe judgment	Provide training, protected override mechanisms, workload design, and responsibility allocation	Evaluate which oversight models detect errors and support recovery in real workflows	High requirement confidence; limited effectiveness evidence
Risk management	Reviews support proportional controls based on intended use and potential consequences	Identify hazards, affected stakeholders, failure modes, mitigations, residual risks, and acceptance criteria	Risk taxonomies and thresholds vary across settings	Enables structured readiness assessment rather than approval determination	Assign risk ownership, multidisciplinary review, incident response, and corrective action	Develop pharmaceutical use-case risk tiers linked to evidence and monitoring expectations	Moderate-to-high confidence
Governance readiness	Frameworks converge on accountability, transparency, fairness, safety, oversight, and monitoring	Establish enforceable governance across data, models, users, workflows, and lifecycle changes	Operational authority, resources, enforcement, and maturity differ substantially	Represents an organizational capability rather than a model property	Define committees, decision rights, competencies, records, review cycles, and	Compare governance implementation models and their effects on safety, quality, and	High principle convergence; limited implementation evidence

					enforcement mechanisms	accountability	
Regulatory alignment	Pharmaceutical reviews emphasize intended use, traceability, validation, controlled change, and monitoring	Assemble risk-proportionate evidence within the relevant pharmaceutical quality and governance system	Mature precedents are limited, and terminology is fragmented	Supports readiness assessment but does not establish regulatory approval	Integrate regulatory-science expertise, quality systems, documentation, and inspection readiness	Study use-specific evidence packages without presuming authority acceptance	Qualified confidence; no approval inference

Figure 1 presents a review-level regulatory-readiness framework linking explainability, reproducibility, bias control, validation, governance, lifecycle monitoring, human oversight, and regulatory-alignment boundaries for AI in pharmaceutical sciences.

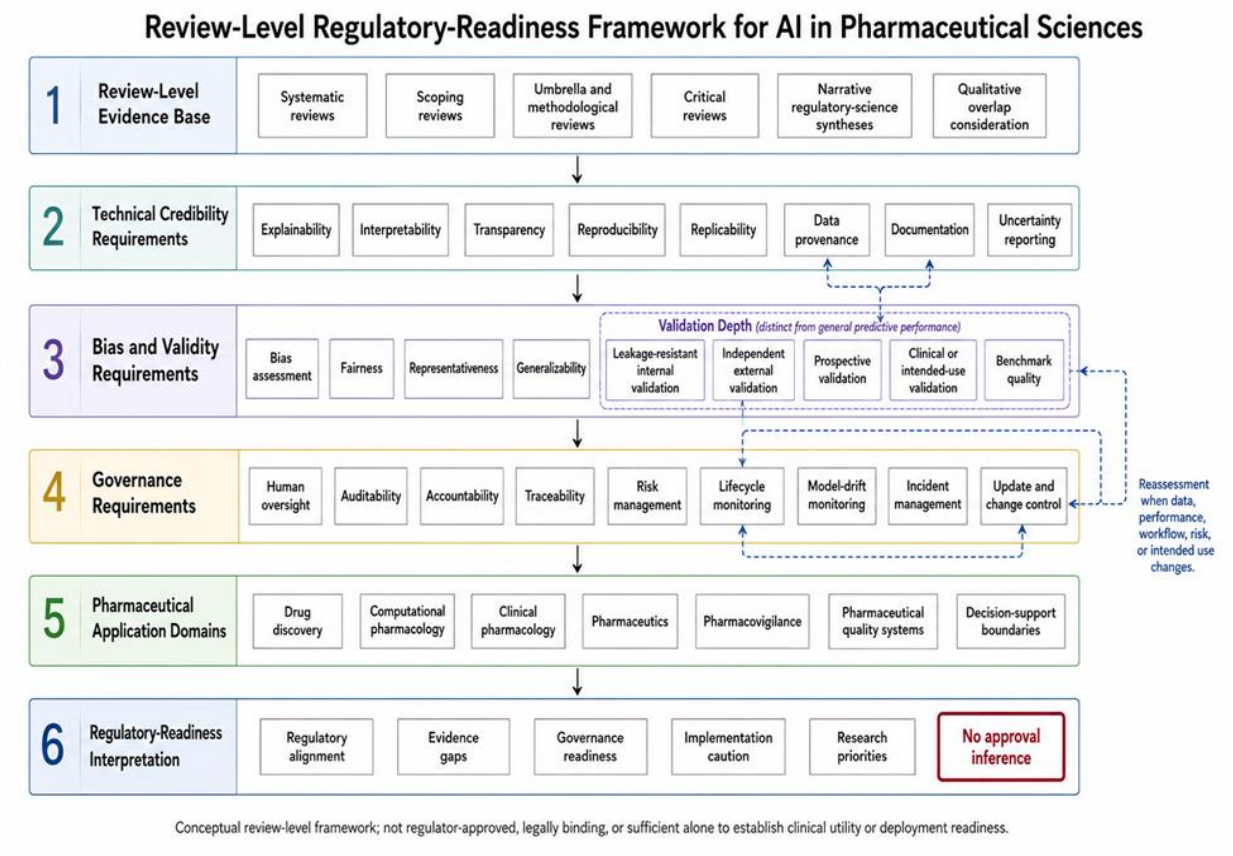

Figure 1. Review-Level Regulatory-Readiness Framework for AI in Pharmaceutical Sciences

Alt-text description

A conceptual regulatory-readiness framework showing review-level evidence flowing into explainability, reproducibility, bias assessment, validation, data provenance, documentation, governance, lifecycle monitoring, human oversight, risk management, regulatory alignment, and evidence gaps.

Regulatory and research implications

For pharmaceutical researchers, AI developers, clinical pharmacologists, pharmacovigilance teams, quality scientists, and journal reviewers, the review findings support a shift from performance-centered reporting toward complete evidence-chain reporting. Model-development publications should identify the intended use, data-generating context, inclusion and exclusion logic, preprocessing order, partitioning strategy, benchmark rationale, calibration, uncertainty, subgroup behavior, failure modes, validation status, and conditions under

ISSN (Online) 2249-6084 (Print) 2250-1029

which the model should not be used. Ethical mapping evidence also indicates that accountability, human agency, privacy, fairness, transparency, and organizational responsibility should be considered throughout development and use rather than added after technical evaluation. Research review and peer review should therefore assess whether evidence claims remain proportionate to the design, data, and validation actually reported [35].

Governance bodies should distinguish agreement on principles from evidence that controls operate effectively. International ethics guidance shows broad convergence around transparency, fairness, non-maleficence, responsibility, and privacy, but substantial variation remains in definitions, prioritization, implementation procedures, enforcement, and accountability. Pharmaceutical governance should translate high-level principles into assigned decision rights, documentation requirements, risk classifications, validation expectations, review triggers, monitoring responsibilities, incident procedures, and controlled update processes. It should also specify how disagreements between model outputs and professional judgment are documented and resolved. Principle-level convergence may justify a shared vocabulary for governance, but it does not demonstrate that an organization, workflow, or AI system is compliant, effective, or accepted by a regulatory authority [36].

The most important research priorities are standardized reporting structures, reproducibility packages compatible with controlled data access, use-specific external-validation protocols, prospective evaluation designs, bias-audit frameworks, uncertainty and calibration standards, lifecycle monitoring strategies, model-drift response procedures, post-implementation surveillance where relevant, auditable human-oversight models, and interoperable documentation and provenance standards. Within pharmaceutical manufacturing and quality systems, regulatory-science literature further emphasizes risk-based lifecycle validation, data integrity, traceable software and model versions, controlled change, continued verification, deviation handling, and inspection-ready records. Future research should evaluate how these requirements function in real pharmaceutical settings and how evidence expectations should vary with intended use and potential consequence. Such work may strengthen regulatory alignment, but it should not presume that a generalized framework determines the evidentiary expectations or decisions of any authority [37].

CONCLUSION

This umbrella review indicates that regulatory-ready AI in pharmaceutical sciences depends on an integrated combination of explainability, reproducibility, transparent data provenance, documentation, bias and fairness assessment, representative evidence, fit-for-purpose validation, uncertainty communication, auditability, human oversight, risk management, governance readiness, model-drift control, and lifecycle monitoring. The review-level literature converges on these domains as mutually dependent evidence requirements while also revealing inconsistent terminology, uneven reporting, limited independent validation, sparse prospective and long-term operational evidence, and a persistent gap between governance principles and demonstrated governance implementation. These findings support a staged readiness assessment in which technical credibility, validation depth, organizational accountability, and continued oversight are evaluated together. They do not establish that any AI model, workflow, platform, or pharmaceutical application has achieved regulatory approval, clinical utility, autonomous decision-making suitability, or deployment readiness.

Acknowledgments: None

Conflict of interest: None

Financial support: None

Ethics statement: None

REFERENCES

[1] Vamathevan J, Clark D, Czodrowski P, Dunham I, Ferran E, Lee G, et al. Applications of machine learning in drug discovery and development. Nat Rev Drug Discov. 2019;18(6):463-77. doi:10.1038/s41573-019-0024-5

[2] Mak KK, Pichika MR. Artificial intelligence in drug development: Present status and future prospects. Drug Discov Today. 2019;24(3):773-80. doi:10.1016/j.drudis.2018.11.014

[3] Harrer S, Shah P, Antony B, Hu J. Artificial intelligence for clinical trial design. Trends Pharmacol Sci. 2019;40(8):577-91. doi:10.1016/j.tips.2019.05.005

[4] Paul D, Sanap G, Shenoy S, Kalyane D, Kalia K, Tekade RK. Artificial intelligence in drug discovery and development. Drug Discov Today. 2021;26(1):80-93. doi:10.1016/j.drudis.2020.10.010

[5] Vora LK, Gholap AD, Jetha K, Thakur RRS, Solanki HK, Chavda VP. Artificial intelligence in pharmaceutical technology and drug delivery design. Pharmaceutics. 2023;15(7):1916. doi:10.3390/pharmaceutics15071916

[6] Falconer N, Abdel-Hafez A, Scott IA, Marxen S, Canaris S, Barras M. Systematic review of machine learning models for personalised dosing of heparin. Br J Clin Pharmacol. 2021;87(11):4124-39. doi:10.1111/bcp.14852

[7] Kompa B, Hakim JB, Palepu A, Kompa KG, Smith M, Bain PA, et al. Artificial intelligence based on machine learning in pharmacovigilance: A scoping review. Drug Saf. 2022;45(5):477-91. doi:10.1007/s40264-022-01176-1

[8] Patil RS, Kulkarni SB, Gaikwad VL. Artificial intelligence in pharmaceutical regulatory affairs. Drug Discov Today. 2023;28(9):103700. doi:10.1016/j.drudis.2023.103700

[9] Hunt H, Pollock A, Campbell P, Estcourt L, Brunton G. An introduction to overviews of reviews: Planning a relevant research question and objective for an overview. Syst Rev. 2018;7(1):39. doi:10.1186/s13643-018-0695-8

[10] Lunny C, Brennan SE, McDonald S, McKenzie JE. Toward a comprehensive evidence map of overview of systematic review methods: Paper 2—risk of bias assessment; synthesis, presentation and summary of the findings; and assessment of the certainty of the evidence. Syst Rev. 2018;7(1):159. doi:10.1186/s13643-018-0784-8

[11] Pollock A, Campbell P, Brunton G, Hunt H, Estcourt L. Selecting and implementing overview methods: Implications from five exemplar overviews. Syst Rev. 2017;6(1):145. doi:10.1186/s13643-017-0534-3

[12] Jiménez-Luna J, Grisoni F, Schneider G. Drug discovery with explainable artificial intelligence. Nat Mach Intell. 2020;2(10):573-84. doi:10.1038/s42256-020-00236-4

[13] Amann J, Blasimme A, Vayena E, Frey D, Madai VI; Precise4Q Consortium. Explainability for artificial intelligence in healthcare: A multidisciplinary perspective. BMC Med Inform Decis Mak. 2020;20(1):310. doi:10.1186/s12911-020-01332-6

[14] Tjoa E, Guan C. A survey on explainable artificial intelligence (XAI): Toward medical XAI. IEEE Trans Neural Netw Learn Syst. 2021;32(11):4793-813. doi:10.1109/TNNLS.2020.3027314

[15] Samek W, Montavon G, Lapuschkin S, Anders CJ, Müller KR. Explaining deep neural networks and beyond: A review of methods and applications. Proc IEEE. 2021;109(3):247-78. doi:10.1109/JPROC.2021.3060483

[16] Ghassemi M, Oakden-Rayner L, Beam AL. The false hope of current approaches to explainable artificial intelligence in health care. Lancet Digit Health. 2021;3(11). doi:10.1016/S2589-7500(21)00208-9

[17] McDermott MBA, Wang S, Marinsek N, Ranganath R, Foschini L, Ghassemi M. Reproducibility in machine learning for health research: Still a ways to go. Sci Transl Med. 2021;13(586). doi:10.1126/scitranslmed.abb1655

[18] Vollmer S, Mateen BA, Bohner G, Király FJ, Ghani R, Jonsson P, et al. Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. BMJ. 2020;368. doi:10.1136/bmj.l6927

[19] Kapoor S, Narayanan A. Leakage and the reproducibility crisis in machine-learning-based science. Patterns (N Y). 2023;4(9):100804. doi:10.1016/j.patter.2023.100804

[20] Norori N, Hu Q, Aellen FM, Faraci FD, Tzovara A. Addressing bias in big data and AI for health care: A call for open science. Patterns (N Y). 2021;2(10):100347. doi:10.1016/j.patter.2021.100347

[21] Challen R, Denny J, Pitt M, Gompels L, Edwards T, Tsaneva-Atanasova K. Artificial intelligence, bias and clinical safety. BMJ Qual Saf. 2019;28(3):231-7. doi:10.1136/bmjqs-2018-008370

[22] Panch T, Mattie H, Atun R. Artificial intelligence and algorithmic bias: Implications for health systems. J Glob Health. 2019;9(2):020318. doi:10.7189/jogh.09.020318

[23] Nagendran M, Chen Y, Lovejoy CA, Gordon AC, Komorowski M, Harvey H, et al. Artificial intelligence versus clinicians: Systematic review of design, reporting standards, and claims of deep learning studies. BMJ. 2020;368. doi:10.1136/bmj.m689

[24] Liu X, Faes L, Kale AU, Wagner SK, Fu DJ, Bruynseels A, et al. A comparison of deep learning performance against health-care professionals in detecting diseases from medical imaging: A systematic review and meta-analysis. Lancet Digit Health. 2019;1(6). doi:10.1016/S2589-7500(19)30123-2

[25] Kelly CJ, Karthikesalingam A, Suleyman M, Corrado G, King D. Key challenges for delivering clinical impact with artificial intelligence. BMC Med. 2019;17(1):195. doi:10.1186/s12916-019-1426-2

[26] Siala H, Wang Y. SHIFTing artificial intelligence to be responsible in healthcare: A systematic review. Soc Sci Med. 2022;296:114782. doi:10.1016/j.socscimed.2022.114782

[27] Reddy S, Allan S, Coghlan S, Cooper P. A governance model for the application of AI in health care. J Am Med Inform Assoc. 2020;27(3):491-7. doi:10.1093/jamia/ocz192

[28] Alami H, Pozelli Sabio R, Pérez EJ, Gagnon MP, Langlois L, Denis JL, et al. Artificial intelligence

governance in health systems: Systematic review of frameworks and integrative model proposal. J Med Internet Res. 2026;28. doi:10.2196/87448

[29] Chen H, Engkvist O, Wang Y, Olivecrona M, Blaschke T. The rise of deep learning in drug discovery. Drug Discov Today. 2018;23(6):1241-50. doi:10.1016/j.drudis.2018.01.039

[30] Yang X, Wang Y, Byrne R, Schneider G, Yang S. Concepts of artificial intelligence for computer-assisted drug discovery. Chem Rev. 2019;119(18):10520-94. doi:10.1021/acs.chemrev.8b00728

[31] Dara S, Dhamercherla S, Jadav SS, Babu CM, Ahsan MJ. Machine learning in drug discovery: A review. Artif Intell Rev. 2022;55(3):1947-99. doi:10.1007/s10462-021-10058-4

[32] Paliwal A, Jain S, Kumar S, Wal P, Khandai M, Khandige PS, et al. Predictive modelling in pharmacokinetics: From in-silico simulations to personalized medicine. Expert Opin Drug Metab Toxicol. 2024;20(4):181-95. doi:10.1080/17425255.2024.2330666

[33] Ryan DK, Maclean RH, Balston A, Scourfield A, Shah AD, Ross J. Artificial intelligence and machine learning for clinical pharmacology. Br J Clin Pharmacol. 2024;90(3):629-39. doi:10.1111/bcp.15930

[34] Pilipiec P, Liwicki M, Bota A. Using machine learning for pharmacovigilance: A systematic review. Pharmaceutics. 2022;14(2):266. doi:10.3390/pharmaceutics14020266

[35] Morley J, Machado CCV, Burr C, Cowls J, Joshi I, Taddeo M, et al. The ethics of AI in health care: A mapping review. Soc Sci Med. 2020;260:113172. doi:10.1016/j.socscimed.2020.113172

[36] Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. Nat Mach Intell. 2019;1(9):389-99. doi:10.1038/s42256-019-0088-2

[37] Niazi SK. Regulatory perspectives for AI/ML implementation in pharmaceutical GMP environments. Pharmaceuticals (Basel). 2025;18(6):901.