

Review

Ethical Considerations for Machine Learning Research Using Free-Text Electronic Medical Records: Challenges, Evidence, and Best Practices

Guosong Wu ^{1,2,*}  and Fengjuan Yang ²¹ Department of Health Care Analytics, Shannon School of Business, Cape Breton University, Sydney, NS B1M 1A2, Canada² The Centre for Health Informatics, Cumming School of Medicine, University of Calgary, Calgary, AB T2N 1N4, Canada

* Correspondence: guosong_wu@cbu.ca; Tel.: +1-(902)-574-8277

Abstract

The increasing availability of free-text components in electronic medical records (EMRs) offers unprecedented opportunities for machine learning research, enabling improved disease phenotyping, risk prediction, and patient stratification. However, the use of narrative clinical data raises distinct ethical challenges that are not fully addressed by conventional frameworks for structured data. We conducted a narrative review synthesizing conceptual and empirical literature on ethical issues in free-text EMR research, focusing on privacy, fairness, autonomy, interpretability, and governance. We examined technical methods, including de-identification, differential privacy, bias mitigation, and explainable AI, alongside normative approaches, such as participatory design, dynamic consent models, and multi-stakeholder governance. Our analysis highlights persistent risks, including re-identification, algorithmic bias, and inequitable access, as well as limitations in current regulatory guidance across jurisdictions. We propose ethics-by-design principles that integrate ethical reflection into all stages of machine learning research, emphasize relational accountability to patients and stakeholders, and support global harmonization in governance and stewardship. Implementing these principles can enhance transparency, trust, and social value while maintaining scientific rigor. Ethical integration is therefore not optional but essential to ensure that machine learning research using free-text EMRs aligns with both clinical relevance and societal expectations.

Keywords: machine learning; electronic medical records; free-text data; ethics; privacy; fairness; governance; transparency; consent; interpretability



Academic Editor: Francisco Epelde

Received: 9 October 2025

Revised: 16 November 2025

Accepted: 1 December 2025

Published: 6 December 2025

Citation: Wu, G.; Yang, F. Ethical Considerations for Machine Learning Research Using Free-Text Electronic Medical Records: Challenges, Evidence, and Best Practices. *Hospitals* **2025**, *2*, 29. <https://doi.org/10.3390/hospitals2040029>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The growing availability of large-scale electronic medical records (EMRs) has catalyzed a new generation of machine learning (ML)-driven health research [1]. While structured fields such as diagnosis codes, medication lists, and laboratory values provide valuable information, they give only a partial view of the clinical narrative. Free-text components of EMRs capture rich contextual details about symptoms, clinical reasoning, psychosocial factors, and patient preferences, which are often absent from structured data [2,3]. Incorporating unstructured text has been shown to improve disease phenotype identification, prediction of adverse events, and patient stratification for preventive care [4–9].

Yet the use of free-text data introduces a distinct ethical landscape. Because clinical narratives contain deeply personal health information that can be misused to affect an individual's social standing, employment, or insurability, ethical safeguards are especially critical in the use of free-text EMR data. Textual records are inherently personal and idiosyncratic, reflecting both patient disclosures and clinician interpretations. They often contain direct identifiers or rare contextual clues that make re-identification possible even after anonymization [8,9]. Unlike structured data, free-text notes are challenging to govern under existing privacy frameworks because identifiers are embedded in linguistic context rather than predefined fields [10,11]. ML systems trained on clinical narratives risk amplifying social and clinical biases present in the source documentation [12,13].

A growing body of literature from informatics, medical ethics, and data protection law highlights the urgency of establishing explicit ethical standards for free-text EMR research. Position statements from the International Medical Informatics Association (IMIA) and the UK Health Data Research Alliance emphasize that traditional anonymization and consent models may be inadequate for unstructured text [14,15]. Despite this, there remains no widely accepted framework to reconcile data utility, patient autonomy, and institutional responsibility.

This article advances the discussion by articulating five core ethical questions that recur across international jurisdictions. Each question is explored through current evidence, normative reasoning, and emerging best practices. The aim is not to provide prescriptive rules but to encourage reflective, “ethics-by-design” approaches that integrate ethical analysis into every stage of ML research using EMR text. Key ethical risks, their ML lifecycle stages, and recommended mitigation strategies are summarized.

2. Methods

2.1. Search Strategy

We developed a comprehensive search strategy (see Appendix A) and applied it to PubMed, Web of Science, IEEE Xplore, EMBASE, and CINAHL from inception through September 2025. The strategy combined terms for electronic medical records (EMRs), electronic health records (EHRs), artificial intelligence (AI), machine learning (ML), natural language processing (NLP), and relevant ethical concepts including privacy, consent, bias, fairness, accountability, and governance. Boolean operators (AND, OR) were employed to combine synonyms within and across concepts to ensure broad coverage while maintaining specificity.

2.2. Inclusion and Exclusion Criteria

Articles were eligible for inclusion if they addressed ethical, legal, or governance issues related to the use of free-text EMRs in ML research, including conceptual frameworks, empirical studies, or methodological evaluations. Both original research and reviews were considered. Studies were excluded if they did not focus on EMRs or AI/ML applications, addressed only structured data, or were commentaries, editorials, letters, protocols, or non-human studies.

2.3. Study Selection

All retrieved records were imported into EndNote 20 (Clarivate, Philadelphia, PA, USA) for management and deduplication. Two reviewers (GW and FY) independently screened titles and abstracts for relevance. Articles deemed potentially eligible were subjected to full-text review by the same reviewers. Discrepancies were resolved through discussion. A PRISMA flow diagram was used to document the number of records identified, screened, excluded, and ultimately included.

2.4. Data Extraction and Synthesis

From each included study, we extracted key information regarding ethical domains, ML lifecycle stage, technical and procedural controls, and contextual considerations. Data were organized into a structured table to summarize recurring themes and illustrate the relationships between ethical risks, technical mitigation strategies, and governance practices. Given the narrative nature of this review, results were synthesized descriptively, emphasizing the convergence and divergence of ethical considerations across studies.

2.5. Results

This narrative review incorporates key principles of the PRISMA framework to ensure transparency in study selection and reporting (Figure 1). We screened a total of 2746 articles from PubMed, Web of Science, IEEE Xplore, EMBASE, and CINAHL. After title and abstract review, 187 full-text articles were assessed for eligibility, yielding 19 studies (Table 1) examining ethical considerations in using free-text EMRs for machine learning research. Of them, 7 were empirical studies and 12 were conceptual studies. Across the literature, five core ethical domains were consistently addressed: privacy and confidentiality, bias and fairness, consent and autonomy, validity and accountability, and governance and stewardship. The studies highlighted ethical risks spanning the ML lifecycle, from data collection through model development and deployment, and described corresponding technical and procedural mitigation strategies. Commonly reported approaches included de-identification, differential privacy, explainable AI, dynamic consent models, federated learning, and multi-stakeholder governance frameworks, reflecting both practical and normative efforts to promote responsible ML use in healthcare.

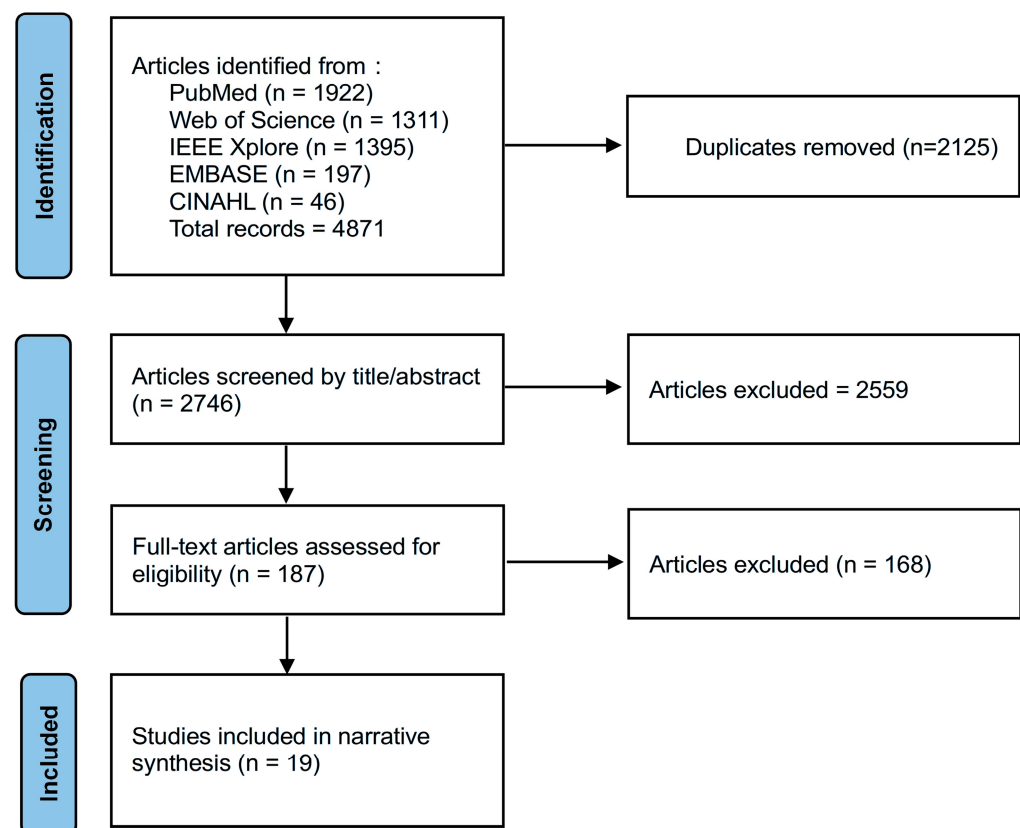


Figure 1. PRISMA Flow Diagram of Study Selection.

Table 1. Ethical Domains, Risks and Mitigation Strategies in ML Using Free-Text EMR Data.

Ethical Domain	Key Risk	ML Lifecycle Stage	Technical Controls	Procedural/Governance Controls
Privacy & Confidentiality [4,5,10–12,14,16–18]	Re-identification via narrative clues; incomplete de-identification	Data Collection, Deployment	Hybrid rule + ML de-identification; differential privacy; federated learning; synthetic data	Trusted Research Environments; continuous risk audits; access logging; ethics review
Bias & Fairness [19–25]	Stigmatizing or unequal documentation; under-representation of subgroups	Preprocessing, Model Development, Deployment	Fairness-aware learning; subgroup performance evaluation	Participatory design with clinicians/patients; bias audits; transparent model cards
Consent & Autonomy [26,27]	Lack of patient awareness or control over text use	Data Collection	–	Dynamic/meta-consent models; public registries; lay communication of data use; consent-to-contact mechanisms
Validity & Accountability [21,25]	Model opacity; poor generalizability; missing documentation	Model Development, Validation, Deployment	Explainable AI methods (SHAP, counterfactuals)	TRIPOD-AI/CONSORT-AI reporting; audit trails; external validation; algorithmic impact assessments
Governance & Stewardship [28,29]	Fragmented regulation; inequitable benefit distribution	All stages	–	Multi-stakeholder oversight; harmonized global standards (OECD, GDPR, HIPAA, TCPS2); ethics-by-design; capacity-building for LMICs

3. Ethical Dimensions of Machine Learning in Health Research

3.1. How Can Privacy and Confidentiality Be Preserved in Free-Text EMR Data?

Privacy and confidentiality considerations should be addressed throughout the ML lifecycle, particularly during data collection, preprocessing, and deployment, to prevent unauthorized access or inadvertent disclosure of sensitive patient information. Tu et al. demonstrated that automated de-identification of primary care free-text EMR data can effectively remove most direct identifiers, but some patient health information may still be retained, highlighting the trade-off between privacy protection and preserving clinical content [10].

Free-text EMR data present unique challenges to privacy and confidentiality. Unlike structured fields, clinical narratives may contain direct identifiers such as names, addresses, and institutional references, as well as indirect identifiers including rare disease mentions, occupational details, or geographic clues that can make individuals re-identifiable even after de-identification [16,17]. Studies have shown that de-identification algorithms for clinical text achieve imperfect performance, with recall rates ranging from 85–98% but sometimes missing subtle identifiers embedded in narrative context [30,31]. These limitations raise the risk of “data leakage”, particularly when datasets are shared across institutions or integrated with external sources like social determinants, as highlighted by Cui [18].

Privacy protection in text-based ML research thus depends on multilayered strategies rather than a single anonymization process. State-of-the-art methods include hybrid de-identification pipelines that combine rule-based pattern matching with transformer-based named-entity recognition [19], differential privacy mechanisms that add statistical noise to outputs [20], and “privacy-preserving learning” paradigms such as federated learning and homomorphic encryption [21]. Each offers partial protection but also introduces trade-offs in model performance and reproducibility. For example, federated learning enables models

to be trained on distributed data without centralizing identifiable text, but it can still leak sensitive information through gradient inversion attacks [22].

From an ethical standpoint, the principle of respect for persons and the duty of nonmaleficence demand that researchers minimize foreseeable harms arising from data misuse. The General Data Protection Regulation (GDPR) and Canada's Tri-Council Policy Statement (TCPS2) both classify identifiable or potentially re-identifiable text data as personal information, requiring proportionate safeguards [32]. The HIPAA Privacy Rule in the United States mandates either removal of 18 specified identifiers or expert assessment of minimal re-identification risk. However, these frameworks were not designed for the linguistic complexity of clinical text. Ford and colleagues note that de-identification alone cannot ensure privacy and advocate for contextual governance, including access auditing, dynamic consent, and secure research environments, rather than relying solely on anonymization [23].

Best practices emerging internationally emphasize a layered governance model:

1. Technical protections such as de-identification, encryption, and controlled vocabularies [16,17,30,31].
2. Organizational controls, including restricted access within secure trusted research environments [28,33].
3. Ethical oversight, requiring institutional review boards to assess text-specific risks and mitigation [24].

Future research should also explore synthetic text generation as an ethical alternative. Recent evidence suggests that synthetic data-trained open-source language models can achieve performance comparable to proprietary models for tasks such as radiology reporting, providing a promising approach to preserve privacy while enabling high-quality machine learning research [25]. However, this approach raises epistemic questions about data authenticity and the potential loss of contextual nuance that underpins clinical reasoning.

In summary, preserving privacy in machine learning using EMR text requires moving from a static concept of confidentiality to a contextual model that integrates technical robustness with continuous ethical assessment.

3.2. How Should Bias and Fairness Be Managed in Text-Based ML Models?

Bias and fairness considerations should be integrated across preprocessing, model development, and evaluation stages to ensure equitable performance and prevent systemic discrimination. Liu et al. offer a users' guide to evaluating clinical ML models, highlighting how methodological rigor (e.g., validation strategy, reference-standard selection) is critical to uncover subtle biases [34]. Obermeyer et al. empirically demonstrate how an algorithm that uses healthcare cost as a proxy for health needs systematically underestimates risk in Black patients, illustrating a real-world fairness failure rooted in proxy choice [26].

Bias in EMR text arises from both linguistic and systemic origins. Clinicians' documentation patterns reflect sociocultural norms, diagnostic heuristics, and implicit biases. Empirical research shows that narrative language differs across patient groups, with stigmatizing descriptors more frequently applied to racialized or low-income patients [35,36]. ML models trained on such notes risk learning and perpetuating inequitable patterns in downstream predictions, raising ethical concerns directly linked to the principles of justice and equity.

Bias manifests in several forms. Representation bias occurs when certain populations are underrepresented in the training corpus, leading to degraded model performance for minority groups. Annotation bias emerges when labeling or data curation reflect the subjective interpretations of coders. Outcome bias arises when historical data encode

inequitable clinical practices, such as differential access to care. These biases can converge, producing algorithmic disparities that compound existing inequities [26,27].

Mitigating such bias requires methodological and ethical interventions. Technically, researchers can employ fairness-aware learning strategies, such as reweighting samples, adversarial debiasing, or post hoc calibration to equalize error rates across subgroups [37]. However, fairness cannot be reduced to a statistical property. Ethical fairness also entails recognition of context: why certain disparities exist and whether correction may inadvertently obscure systemic problems [38]. For instance, “equalizing” documentation patterns without addressing underlying disparities in healthcare delivery risks masking structural inequities rather than rectifying them.

Interpretability and stakeholder engagement play central roles. Ethical ML development should include participatory design involving clinicians, ethicists, and patient representatives who can identify how textual features might encode bias. Transparent reporting frameworks such as Model Cards for Clinical natural language processing (NLP) and the CONSORT-AI extension help standardize documentation of dataset composition and subgroup performance [39,40]. Journals and funders increasingly require bias assessment as part of responsible research practices, aligning with WHO’s 2023 “Guidance on Ethics and Governance of AI in Health” [33].

Fairness in text-based machine learning is not a one-time adjustment but an ongoing ethical commitment. Researchers should integrate bias analysis into a broader social responsibility to examine how language reflects and can transform medical culture. Future ethical frameworks may benefit from cross-disciplinary training in sociolinguistics, critical race theory, and clinical ethics to sensitize machine learning researchers to the moral implications of language data.

3.3. *What Are the Obligations Around Consent, Autonomy, and Transparency?*

Ethical obligations related to consent, autonomy, and transparency should be addressed during data collection and ongoing participant engagement, ensuring patients understand and control how their EMR data are used in ML research. For example, Obermeyer et al. highlight the importance of clear patient communication to prevent misuse of algorithmic outputs [26], and Rajkomar et al. discuss implementing fairness-aware mechanisms alongside transparent consent procedures to uphold patient autonomy [27].

In most EMR-based ML studies, patient consent is waived under the assumption that de-identified data entail minimal risk. However, free-text data challenge this rationale. Because text can embed intimate personal narratives, including mental health experiences, family dynamics, or clinician impressions, the potential for dignitary harm persists even after technical anonymization [41]. The principle of respect for autonomy thus requires a re-examination of what meaningful consent entails when patients’ words are used in research.

Several consent models have been proposed. Broad consent allows reuse of data for unspecified future research. Dynamic consent enables participants to manage permissions over time through digital platforms. Meta-consent allows individuals to predefine preferences based on research type or sensitivity level [42,43]. Among these, dynamic consent has been promoted as most compatible with evolving data-intensive research, but its feasibility remains limited for retrospective EMR text where millions of records are already stored.

Transparency is the cornerstone of ethical legitimacy. Surveys show that patients generally support secondary use of their data when governance is transparent and benefits are clear [44]. Conversely, secrecy or commercial data-sharing arrangements erode trust. In the context of free-text EMR data, transparency involves clear communication about data types used, de-identification procedures, and intended ML applications. Ethical

transparency goes beyond legal disclosure to relational transparency, emphasizing ongoing engagement that treats patients as moral stakeholders rather than passive data sources [45].

Institutional Review Boards (IRBs) and data custodians thus bear a dual obligation: to protect individual autonomy through oversight and to ensure social value by enabling responsible data use. Emerging practice in the UK and Canada now mandates public registries for ML projects using identifiable health data (Health Data Research UK, London, UK). In parallel, several jurisdictions are exploring “consent-to-contact” models, where patients opt in for recontact if their narratives are to be used for model development. This approach balances feasibility with respect for autonomy [46].

In summary, while broad consent remains dominant, future ethical governance should move toward participatory and transparent frameworks. ML research using free-text EMR data must acknowledge that data are extensions of persons, and ethical legitimacy requires both procedural consent and substantive respect.

3.4. How Can Validity, Interpretability, and Accountability Be Maintained in Text-Based ML Research?

Validity, interpretability, and accountability should be embedded into model development, validation, and deployment stages to promote reproducibility, transparent reporting, and responsible decision-making. For example, Xu et al. review how federated learning can help preserve data privacy by training models across institutions without centralizing sensitive patient data [21], while Pandita et al. empirically demonstrate that open-source LLMs fine-tuned on synthetic clinical narratives can achieve performance comparable to proprietary models, offering a privacy-preserving and reproducible alternative [25].

Ensuring scientific validity and interpretability of ML models trained on EMR text is an ethical obligation grounded in the principle of beneficence. Invalid or opaque models can lead to harmful clinical or policy decisions, violating the duty to do good and avoid harm [47]. Free-text EMR data are particularly challenging because clinical narratives are noisy, inconsistent, and context-dependent. Abbreviations, colloquialisms, and idiosyncratic clinician language complicate preprocessing and model generalization [48]. Furthermore, the lack of external validation across institutions amplifies the risk of dataset shift, where a model trained on one hospital’s documentation performs poorly elsewhere [49].

Model interpretability remains a central ethical issue. Deep learning models, such as transformer-based language models (e.g., BERT, GPT, or BioClinicalBERT), achieve state-of-the-art performance in clinical NLP but often operate as “black boxes” [50]. Without transparency in how predictions are generated, clinicians and patients cannot meaningfully evaluate algorithmic recommendations. The European Commission’s High-Level Expert Group on AI explicitly identifies explainability as a core requirement for trustworthy AI [51]. Interpretability tools, including attention visualization, Shapley additive explanations (SHAP, which quantify the contribution of each feature to a model’s prediction), and counterfactual reasoning, offer partial solutions but have limitations in clinical reliability [29]. Ethical accountability thus extends beyond technical explainability to include epistemic accountability: ensuring that models’ representations correspond to valid clinical reasoning and are reviewed by domain experts [52].

Reproducibility and documentation also underpin ethical validity. Studies often fail to report dataset preprocessing steps or hyperparameter tuning, undermining replicability [53]. To address this, international guidelines such as TRIPOD-AI and CONSORT-AI call for transparent reporting of data sources, model architectures, and validation procedures [34,53]. Journals increasingly require model-sharing statements or code repositories as a condition of publication. Accountability mechanisms further include algorithmic impact assessments, structured evaluations of potential harms, benefits, and bias before deployment [54].

The distribution of ethical concerns, risks, technical controls, accountability, and feedback across ML stages is summarized in Figure 2. The heatmap, scored 0–5 (0 = not relevant; 5 = very high), illustrates that privacy and autonomy concerns are most prominent during data collection and deployment, while bias and fairness risks peak in preprocessing, model development, and deployment. Technical safeguards such as de-identification, bias detection, and monitoring are applied selectively according to the ML stage, and accountability responsibilities are shared across researchers, data scientists, institutions, and ethics boards. Feedback mechanisms, including continuous review, bias re-evaluation, model updates, and post-deployment monitoring, are emphasized particularly in later stages to ensure ongoing ethical oversight. This framework visually integrates multiple dimensions of ethical responsibility, highlighting critical points for intervention to maintain validity, interpretability, and accountability throughout the ML lifecycle.

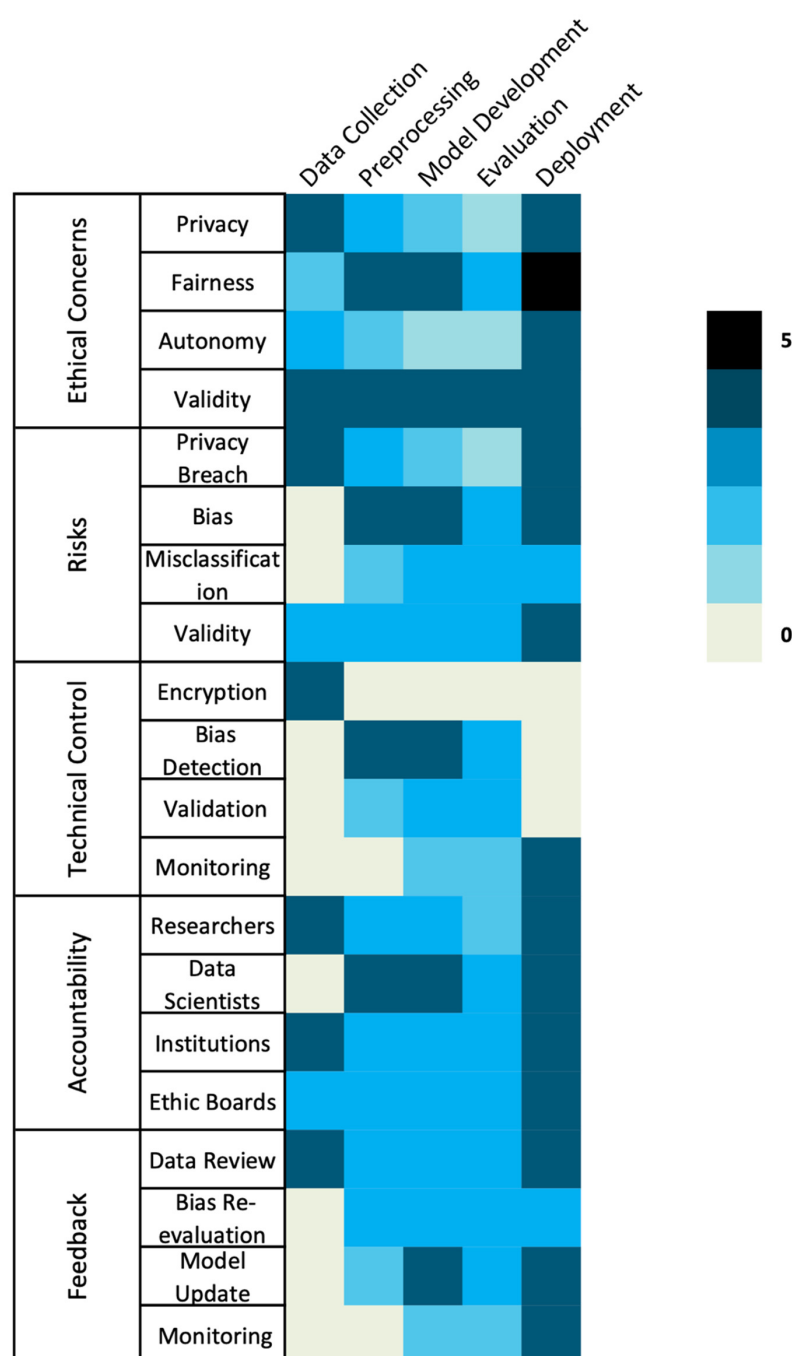


Figure 2. Ethical Risk Heatmap Across the Machine Learning Lifecycle.

Responsibility must be distributed across the research ecosystem. Clinicians provide contextual judgment, data scientists ensure technical robustness, and institutions bear legal and moral accountability for data governance [55]. Emerging frameworks propose “audit trails” that document every major analytic decision from data extraction to publication [56]. Such traceability supports retrospective accountability in case of harm or error.

Finally, epistemic humility is essential. No ML model can capture the full complexity of clinical reasoning. A responsible research culture acknowledges uncertainty and resists overstating predictive validity. Ethical ML research should thus commit to external validation, continuous performance monitoring, and open publication of negative findings [57].

3.5. How Should Governance, Stewardship, and Global Ethical Harmonization Be Structured?

Governance, stewardship, and harmonization efforts should span all stages of the ML lifecycle, aligning legal, ethical, and societal expectations while promoting equitable benefits and international ethical standards. For example, Sun et al. review how EMR systems’ data processing pipelines must grapple with the inherent diversity, incompleteness, and privacy concerns of clinical data [48].

Governance frameworks determine who controls data, under what conditions, and for whose benefit. For free-text EMR data, governance must reconcile individual rights with collective social value [58]. The challenge is amplified in international collaborations where datasets cross legal and cultural boundaries. Privacy regulations such as the EU GDPR, Canada’s Personal Information Protection and Electronic Documents Act (PIPEDA), and the U.S. HIPAA Privacy Rule differ in definitions of identifiability and permissible secondary uses [51,59,60]. Textual data often fall into gray zones: they may be technically de-identified but remain “personal” in the sense of being expressive of human experience [61].

Practical implementation in low- and middle-income countries (LMICs) illustrates these cross-jurisdictional challenges. For example, Ghosheh and colleagues synthesized ICU EHR data from Vietnam to develop predictive models for hospital-acquired infections. Their work highlights constraints such as limited data availability, infrastructure gaps, and differing local regulations [62]. Ethical governance must therefore be adaptable, balancing local requirements with global standards to ensure equitable access, accountability, and trust in AI systems across diverse settings.

Ethical stewardship requires both procedural justice, which involves transparent and accountable decision-making, and substantive justice, which ensures equitable distribution of benefits and burdens [63]. Multi-stakeholder governance bodies can help ensure that patients, clinicians, data scientists, and ethicists share responsibility for oversight [34]. Patients contribute perspectives on acceptable data use and potential impacts on autonomy; clinicians ensure that model development aligns with clinical realities; data scientists manage technical design, validation, and monitoring; and ethicists guide value-based decision-making and help navigate trade-offs between risk and utility. The UK’s Trusted Research Environments (TREs) and the Canadian Pan-Canadian Health Data Strategy exemplify institutional infrastructures balancing access with security through controlled enclaves, access logs, and public transparency reports [64].

International harmonization remains difficult. While the OECD Recommendation on Health Data Governance and the WHO 2023 Guidance on Ethics and Governance of AI in Health advocate common principles such as transparency, accountability, equity, but implementation varies widely. Low- and middle-income countries face particular challenges, as imported models may embed biases from high-income settings [52]. Ethical globalization therefore requires reciprocal capacity building: data partnerships that include shared governance training, technology transfer, and benefit sharing [55]. Selected interna-

tional frameworks relevant to these principles are summarized in Table 2, illustrating their applicability to EMR text and considerations for ML deployment.

Table 2. Overview of Selected International Guidelines and Frameworks Addressing Ethical, Governance, and Accountability Considerations for ML Using Health.

Framework/Guideline	Jurisdiction/Organization	Key Ethical Principles	Applicability to EMR Text	Relevant ML Considerations
IMIA Code of Ethics for Health Information Professionals (2021) [15]	International	Integrity, confidentiality, social responsibility	Stresses ethical literacy among informaticians	Guides training, ethical decision-making, and responsible reporting of ML outputs
WHO Guidance on Ethics and Governance of AI in Health (2023) [33]	World Health Organization	Accountability, inclusiveness, human oversight, sustainability	Global reference for AI/ML ethics; applicable to text analytics	Emphasizes ethics-by-design, participatory governance, human-in-the-loop oversight
OECD Recommendation on Health Data Governance (2017) [28]	OECD Members	Privacy, transparency, stewardship, interoperability	Promotes cross-national data-sharing ethics	Supports federated learning, standardized de-identification, multi-site validation
GDPR (2018) [46]	European Union	Lawfulness, fairness, transparency, purpose limitation, data minimization	Defines identifiable data broadly—applies to pseudonymized text	Requires careful de-identification, transparency, and lawful basis for ML training
TCPS 2 (2022) [32]	Canada	Respect for persons, concern for welfare, justice	Requires proportionate safeguards for re-identifiable EMR text	Dynamic/meta-consent models; ethics board oversight; risk-based de-identification
HIPAA Privacy Rule (1996) [48]	United States	Safe-harbor de-identification, expert determination, accountability	Limited guidance for narrative text—expert review required	Expert determination for free-text EMR; secure TREs; audit trails

Stewardship also entails sustainability. Data infrastructures must ensure long-term preservation, secure deletion policies, and responsible linkage with other datasets. Governance charters should specify conditions for data sharing, model ownership, and intellectual-property rights arising from publicly funded data. Without explicit provisions, the commercialization of models trained on public EMR text risks eroding public trust [54].

In practice, best governance combines technical safeguards with normative commitments. This includes ethics-by-design approaches that embed privacy, fairness, and accountability throughout the data lifecycle. Rather than treating ethics as a post hoc review, research institutions should integrate continuous ethical reflexivity, including committees with expertise in ML, mandatory ethics statements in publications, and cross-disciplinary education [57].

4. Discussion

The ethical use of free-text EMR data in ML research reveals overlapping challenges of privacy, fairness, autonomy, and accountability. Across these domains, a central tension emerges: how to balance innovation in medical AI with the obligation to protect individual rights and societal trust.

First, technical safeguards alone are insufficient. De-identification, bias correction, and explainable AI enhance compliance but cannot substitute transparent governance and ethical reflection [65,66]. Widely used de-identification tools, such as UCSF Philter, PhysioNet De-ID, and emerging LLM-based anonymization systems, exemplify technical strategies to protect patient privacy while enabling ML research, yet they remain insufficient without accompanying ethical oversight [67–69]. Ethical soundness must be built into study design,

validation, and interpretation, rather than applied retroactively after technical development. Emerging reporting standards, such as TRIPOD-LLM, are being developed to enhance transparency and reproducibility in LLM-driven clinical research. These frameworks address LLM-specific challenges such as prompt leakage, semantic re-identification, and contextual memorization, complementing existing guidance on ethical study design and oversight [70]. Global guidance, such as the WHO Ethics and Governance of Artificial Intelligence for Health and OECD Health Data Governance Framework, emphasize human oversight and institutional responsibility as cornerstones of trustworthy AI [28,33].

Second, beyond technical safeguards, ethical stewardship also requires clarity on data ownership and permissible use for both clinical care and research. Patients are the primary stakeholders and data originators, yet institutions often act as custodians, balancing privacy, utility, and regulatory obligations. Clear policies on access, secondary use, and data sharing are essential to align institutional practices with patient rights, enhance transparency, and support responsible research. Patients should be informed about the involvement of machine learning systems in their care, including how model outputs may inform clinical decisions.

Third, trust depends on relational accountability. Evidence shows that patients generally support secondary use of their data when safeguards and benefits are clearly explained [71,72]. Sustainable trust therefore requires openness about data flows, participatory governance, and clear mechanisms for redress. To further strengthen accountability and provenance in ML and LLM research, practical tools such as model cards, dataset versioning, and audit logging are increasingly recommended [40,73,74]. Model cards provide structured documentation of model purpose, performance, and limitations; dataset versioning ensures reproducibility by tracking changes in training data; and audit logs capture usage and outputs, enabling retrospective review and governance of model behavior. Beyond compliance with regulations like GDPR, ethical considerations in ML research also encompass algorithm design, bias mitigation strategies, and assessment of the downstream impact of models on clinical decision-making. Importantly, ethical responsibility is distinct from legal liability: accountability may involve multiple actors, including algorithm designers, those overseeing model training, and healthcare professionals applying the system in practice. Clarifying these roles supports ethics-by-design approaches and ensures that both technical and moral responsibilities are addressed alongside legal compliance.

Last but not least, ethical governance must operate across global contexts and be supported by robust institutional practices. Current ML and NLP systems are often developed using English-language data from high-income settings, risking systemic exclusion of underrepresented populations [75]. Scholars have called for equitable data partnerships and shared oversight to prevent “ethics dumping” and promote fairness in global health research [76,77]. Institutional safeguards such as Trusted Research Environments (TREs), AI ethics review boards, and IRB-level oversight help operationalize these principles, ensuring that both technical and procedural controls are effectively applied across diverse regulatory and cultural settings [78,79]. Governance is also shaped by differing political and regulatory environments: the EU AI Act adopts a risk-based framework for high-risk health applications, the United States relies on decentralized sectoral regulation, and China emphasizes centralized oversight and data localization [80,81]. These variations highlight the need for interoperable ethical principles that can guide responsible ML practice internationally.

Limitations

While our manuscript emphasizes regulatory and governance frameworks for ethical use of free-text EMR data, we do not extensively address complementary measures such as internal auditing, ethics training, or organizational procedures that reinforce researcher

responsibility. Additionally, although synthetic data can mitigate privacy risks and support ethical research practices, it does not replace the need for robust governance, oversight, and accountability mechanisms. Additionally, AI/ML methods, large language models (LLMs), and de-identification techniques are evolving rapidly; consequently, some findings and recommendations may become outdated as new tools, methodologies, or ethical standards emerge.

5. Conclusions

Free-text EMR data provide powerful insights into clinical reasoning and patient experience but raise profound ethical challenges. Addressing these challenges requires integrating ethics into every stage of machine learning research, from data extraction to model deployment.

Across privacy, fairness, autonomy, validity, and governance, one conclusion stands out: ethics is inseparable from scientific quality. ML tools that fail to address ethical design risk both social mistrust and scientific bias. Embedding “ethics-by-design” principles ensures models are transparent, equitable, and clinically meaningful [65].

Ultimately, the question is not whether EMR text data should be used, but how they can be used responsibly. Progress will depend on sustained collaboration among clinicians, data scientists, ethicists, and patients. In this sense, the future of artificial intelligence in medicine will depend as much on moral imagination as on algorithmic sophistication, ensuring that technological advancement aligns with the enduring goals of care, compassion, and justice.

6. Future Directions

Effective ethical governance and stewardship of ML in healthcare depend not only on policies and technical safeguards but also on sufficient financial and human resources to support infrastructure, oversight, and continuous evaluation. Future research could examine the impact of different national and regional AI policies on ML implementation in healthcare, providing guidance for harmonizing ethical standards across jurisdictions. Future work should also prioritize participatory ethics research exploring how patients and clinicians perceive risks and benefits; ethics-by-design methods that integrate fairness and transparency into model objectives [82]; and comparative analysis of how diverse regulatory systems, such as GDPR, HIPAA, and TCPS2, operationalize similar principles [83]. As Mittelstadt and Floridi suggested, ethical governance should act “not as a brake on innovation, but as its steering mechanism” [76].

Author Contributions: G.W. and F.Y. contributed equally as co-first authors. G.W. and F.Y. conceived and designed the study, conducted the literature review, and extracted data from the literature. G.W. and F.Y. prepared the tables and figures and jointly drafted the manuscript. All authors contributed to data interpretation, critically revised the manuscript for important intellectual content, and approved the final version. All authors accept full responsibility for the work. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: This study used only publicly available databases and did not involve human subjects; therefore, ethics approval was not required.

Informed Consent Statement: Not applicable.

Data Availability Statement: The dataset is available from the corresponding author upon reasonable request.

Acknowledgments: We are profoundly grateful for the love and support of our sunshine, Qingyang Wu, our angel, Hannah Jiayue Wu, and our princess, Aleena Yixing Wu, whose presence fills our lives with joy and love.

Conflicts of Interest: Wu is a funded Collaborative Evidence Informed Fellow at University of Oxford. The remaining author reports no conflicts of interest.

Appendix A. Search Strategy

Web of Science [1311]

electronic medical record (All Fields) and Electronic Health Records (OR—Search within topic) and Electronic Health Record (OR—Search within topic) and Electronic Medical Records (OR—Search within topic) and Electronic Medical Record (OR—Search within topic) and Ehr (OR—Search within topic) and Electronic Patient Record (OR—Search within topic) and Electronic Health Record Ehr (OR—Search within topic) and Electronic Medical Record Emr (OR—Search within topic) and Electronic Patient Records (OR—Search within topic)

AND

artificial intelligence (All Fields) and Artificial Intelligence (OR—Search within topic) and Artificial Intelligence Ai (OR—Search within topic) and Machine Learning (OR—Search within topic) and Nlp (OR—Search within topic)

AND

ethic (All Fields) and Privacy (OR—Search within topic) and Consent (OR—Search within topic) and Autonomy (OR—Search within topic) and Bias (OR—Search within topic) and Fairness (OR—Search within topic) and Accountability (OR—Search within topic) and Governance (OR—Search within topic)

PubMed [1922]

("electronic medical record*" OR EMR OR EHR OR "health record*") AND ("machine learning" OR "artificial intelligence" OR AI OR "natural language processing" OR NLP) AND (ethic* OR "privacy" OR "confidentiality" OR "consent" OR "autonomy" OR "bias" OR "fairness" OR "accountability" OR "governance")

IEEE [1395]

("electronic medical record*" OR EMR OR EHR OR "health record*") AND ("machine learning" OR "artificial intelligence" OR AI OR "natural language processing" OR NLP) AND (ethic* OR "privacy" OR "confidentiality" OR "consent" OR "autonomy" OR "bias" OR "fairness" OR "accountability" OR "governance")

EMBASE [197]

((‘electronic medical record*’ or EMR or ‘electronic health record*’ or EHR) and (‘free-text’ or ‘clinical text’ or narrative) and (‘machine learning’ or ‘artificial intelligence’ or ‘deep learning’ or ‘natural language processing’ or NLP) and (ethic* or privacy or autonomy or consent or bias or fairness or governance)).mp.

CINAHL [46]

("electronic medical record*" OR EMR OR "electronic health record*" OR EHR)
AND ("free-text" OR "clinical text" OR narrative)
AND ("machine learning" OR "artificial intelligence" OR "deep learning" OR "natural language processing" OR NLP)
AND (ethic* OR privacy OR autonomy OR consent OR bias OR fairness OR governance).

References

1. Wu, G.; Yang, F. Navigating the Transformative Impact of Artificial Intelligence in Health Services Research. *Health Sci. Rep.* **2025**, *8*, e70793. [CrossRef] [PubMed]
2. Wu, G.; Eastwood, C.; Zeng, Y.; Quan, H.; Long, Q.; Zhang, Z.; Ghali, W.A.; Bakal, J.; Boussat, B.; Flemons, W. Developing EMR-based algorithms to Identify hospital adverse events for health system performance evaluation and improvement: Study protocol. *PLoS ONE* **2022**, *17*, e0275250. [CrossRef] [PubMed]
3. Wu, G.; Soo, A.; Ronksley, P.; Holroyd-Leduc, J.; Bagshaw, S.M.; Wu, Q.; Quan, H.; Stelfox, H.T. A multicenter cohort study of falls among patients admitted to the ICU. *Crit. Care Med.* **2022**, *50*, 810–818. [CrossRef]
4. Hossain, E.; Rana, R.; Higgins, N.; Soar, J.; Barua, P.D.; Pisani, A.R.; Turner, K. Natural language processing in electronic health records in relation to healthcare decision-making: A systematic review. *Comput. Biol. Med.* **2023**, *155*, 106649. [CrossRef]
5. Kosowan, L.; Singer, A.; Zulkernine, F.; Zafari, H.; Nesca, M.; Muthumuni, D. Pan-Canadian Electronic Medical Record Diagnostic and Unstructured Text Data for Capturing PTSD: Retrospective Observational Study. *JMIR Med. Inform.* **2022**, *10*, e41312. [CrossRef]
6. Wu, G.; Cheliger, C.; Southern, D.A.; Martin, E.A.; Xu, Y.; Leal, J.; Ellison, J.; Bush, K.; Williamson, T.; Quan, H. Development of machine learning models for the detection of surgical site infections following total hip and knee arthroplasty: A multicenter cohort study. *Antimicrob. Resist. Infect. Control.* **2023**, *12*, 88. [CrossRef]
7. Wu, G.; Khair, S.; Yang, F.; Cheliger, C.; Southern, D.; Zhang, Z.; Feng, Y.; Xu, Y.; Quan, H.; Williamson, T. Performance of machine learning algorithms for surgical site infection case detection and prediction: A systematic review and meta-analysis. *Ann. Med. Surg.* **2022**, *84*, 104956. [CrossRef]
8. Cheliger, K.; Wu, G.; Laws, A.; Quan, M.L.; Li, A.; Brisson, A.-M.; Xie, J.; Xu, Y. Validation of large language models for detecting pathologic complete response in breast cancer using population-based pathology reports. *BMC Med. Inform. Decis. Mak.* **2024**, *24*, 283. [CrossRef]
9. Wu, G.; Cheliger, C.; Brisson, A.-M.; Quan, M.L.; Cheung, W.Y.; Brenner, D.; Lupichuk, S.; Teman, C.; Basmadjian, R.B.; Popwich, B. A new method of identifying pathologic complete response after neoadjuvant chemotherapy for breast cancer patients using a population-based electronic medical record system. *Ann. Surg. Oncol.* **2023**, *30*, 2095–2103. [CrossRef] [PubMed]
10. Tu, K.; Klein-Geltink, J.; Mitiku, T.F.; Mihai, C.; Martin, J. De-identification of primary care electronic medical records free-text data in Ontario, Canada. *BMC Med. Inform. Decis. Mak.* **2010**, *10*, 35. [CrossRef] [PubMed]
11. Jones, K.H.; Ford, E.M.; Lea, N.; Griffiths, L.J.; Hassan, L.; Heys, S.; Squires, E.; Nenadic, G. Toward the development of data governance standards for using clinical free-text data in health research: Position paper. *J. Med. Internet Res.* **2020**, *22*, e16760. [CrossRef]
12. Piasecki, J.; Walkiewicz-Żarek, E.; Figas-Skrzypulec, J.; Kordecka, A.; Dranseika, V. Ethical issues in biomedical research using electronic health records: A systematic review. *Med. Health Care Philos.* **2021**, *24*, 633–658. [CrossRef]
13. Wu, G.; Eastwood, C.; Sapiro, N.; Cheliger, C.; Southern, D.A.; Quan, H.; Xu, Y. Achieving high inter-rater reliability in establishing data labels: A retrospective chart review study. *BMJ Open Qual.* **2024**, *13*, e002722. [CrossRef] [PubMed]
14. Duong, T.A.; Lamé, G.; Zehou, O.; Skayem, C.; Monnet, P.; El Khemiri, M.; Boudjemil, S.; Hirsch, G.; Wolkenstein, P.; Jankovic, M. A process modelling approach to assess the impact of tele dermatology deployment onto the skin tumor care pathway. *Int. J. Med. Inform.* **2021**, *146*, 104361. [CrossRef]
15. International Medical Informatics Association. IMIA code of ethics for health information professionals. Retrieved April. **2002**, *14*, 2004. Available online: <https://imia-medinfo.org/wp/wp-content/uploads/2015/07/IMIA-Code-of-Ethics-2016.pdf> (accessed on 30 September 2025).
16. Neamatullah, I.; Douglass, M.M.; Lehman, L.-W.H.; Reisner, A.; Villarroel, M.; Long, W.J.; Szolovits, P.; Moody, G.B.; Mark, R.G.; Clifford, G.D. Automated de-identification of free-text medical records. *BMC Med. Inform. Decis. Mak.* **2008**, *8*, 32. [CrossRef] [PubMed]
17. Meystre, S.M.; Savova, G.K.; Kipper-Schuler, K.C.; Hurdle, J.F. Extracting information from textual documents in the electronic health record: A review of recent research. *Yearb. Med. Inform.* **2008**, *17*, 128–144.
18. Cui, Y. Digital pathways connecting social and biological factors to health outcomes and equity. *npj Digit. Med.* **2025**, *8*, 172. [CrossRef]
19. Liu, Y.; Chen, P.-H.C.; Krause, J.; Peng, L. How to read articles that use machine learning: Users' guides to the medical literature. *JAMA* **2019**, *322*, 1806–1816. [CrossRef] [PubMed]
20. Abadi, M.; Chu, A.; Goodfellow, I.; McMahan, H.B.; Mironov, I.; Talwar, K.; Zhang, L. Deep learning with differential privacy. In Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security, Vienna, Austria, 24–28 October 2016; pp. 308–318.
21. Xu, J.; Glicksberg, B.S.; Su, C.; Walker, P.; Bian, J.; Wang, F. Federated learning for healthcare informatics. *J. Healthc. Inform. Res.* **2021**, *5*, 1–19. [CrossRef]

22. Zhu, L.; Liu, Z.; Han, S. Deep leakage from gradients. In *Advances in Neural Information Processing Systems 32, Proceedings of the Annual Conference on Neural Information Processing Systems 2019, Vancouver, BC, Canada, 8–14 December 2019*; Curran Associates, Inc.: Red Hook, NY, USA, 2019; pp. 17–31.
23. Ford, E.; Carroll, J.A.; Smith, H.E.; Scott, D.; Cassell, J.A. Extracting information from the text of electronic medical records to improve case detection: A systematic review. *J. Am. Med. Inform. Assoc.* **2016**, *23*, 1007–1015. [\[CrossRef\]](#)
24. Panch, T.; Szolovits, P.; Atun, R. Artificial intelligence, machine learning and health systems. *J. Glob. Health* **2018**, *8*, 020303. [\[CrossRef\]](#)
25. Pandita, A.; Keniston, A.; Madhuripan, N. Synthetic data trained open-source language models are feasible alternatives to proprietary models for radiology reporting. *npj Digit. Med.* **2025**, *8*, 472. [\[CrossRef\]](#)
26. Obermeyer, Z.; Powers, B.; Vogeli, C.; Mullainathan, S. Dissecting racial bias in an algorithm used to manage the health of populations. *Science* **2019**, *366*, 447–453. [\[CrossRef\]](#)
27. Rajkomar, A.; Hardt, M.; Howell, M.D.; Corrado, G.; Chin, M.H. Ensuring fairness in machine learning to advance health equity. *Ann. Intern. Med.* **2018**, *169*, 866–872. [\[CrossRef\]](#)
28. OECD. *Recommendation of the Council on Health Data Governance*; OECD: Paris, France, 2016.
29. Sun, W.; Cai, Z.; Li, Y.; Liu, F.; Fang, S.; Wang, G. Data processing and text mining technologies on electronic medical records: A review. *J. Healthc. Eng.* **2018**, *2018*, 4302425. [\[CrossRef\]](#) [\[PubMed\]](#)
30. Uzuner, Ö.; South, B.R.; Shen, S.; DuVall, S.L. 2010 i2b2/VA challenge on concepts, assertions, and relations in clinical text. *J. Am. Med. Inform. Assoc.* **2011**, *18*, 552–556. [\[CrossRef\]](#) [\[PubMed\]](#)
31. Dernoncourt, F.; Lee, J.Y.; Uzuner, O.; Szolovits, P. De-identification of patient notes with recurrent neural networks. *J. Am. Med. Inform. Assoc.* **2017**, *24*, 596–606. [\[CrossRef\]](#)
32. TCPS 2: CORE-2022 (Course on Research Ethics). 2022. Available online: <https://tcps2core.ca/welcome> (accessed on 30 September 2025).
33. World Health Organization. *Ethics and Governance of Artificial Intelligence for Health*; WHO Guidance; World Health Organization: Geneva, Switzerland, 2021.
34. Liu, X.; Rivera, S.C.; Moher, D.; Calvert, M.J.; Denniston, A.K.; Chan, A.; Darzi, A.; Holmes, C.; Yau, C.; Ashrafian, H. Reporting guidelines for clinical trial reports for interventions involving artificial intelligence: The CONSORT-AI extension. *Sci. J. Newsl./Rev. Cientificas Y Boletines* **2024**, *48*, 1–15.
35. Wood, E.A.; Champion, T.R. Design and implementation of an integrated data model to support clinical and translational research administration. *J. Am. Med. Inform. Assoc.* **2022**, *29*, 1559–1566. [\[CrossRef\]](#)
36. Barcelona, V.; Scharp, D.; Idnay, B.R.; Moen, H.; Cato, K.; Topaz, M. Identifying stigmatizing language in clinical documentation: A scoping review of emerging literature. *PLoS ONE* **2024**, *19*, e0303653. [\[CrossRef\]](#) [\[PubMed\]](#)
37. Zhao, J.; Wang, T.; Yatskar, M.; Ordonez, V.; Chang, K.-W. Gender bias in coreference resolution: Evaluation and debiasing methods. *arXiv* **2018**, arXiv:1804.06876. [\[CrossRef\]](#)
38. Veinot, T.C.; Mitchell, H.; Ancker, J.S. Good intentions are not enough: How informatics interventions can worsen inequality. *J. Am. Med. Inform. Assoc.* **2018**, *25*, 1080–1088. [\[CrossRef\]](#)
39. Liu, X.; Faes, L.; Kale, A.U.; Wagner, S.K.; Fu, D.J.; Bruynseels, A.; Mahendiran, T.; Moraes, G.; Shamdas, M.; Kern, C. 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*, e271–e297. [\[CrossRef\]](#)
40. Mitchell, M.; Wu, S.; Zaldivar, A.; Barnes, P.; Vasserman, L.; Hutchinson, B.; Spitzer, E.; Raji, I.D.; Gebru, T. Model cards for model reporting. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, Atlanta, GA, USA, 29–31 January 2019; pp. 220–229.
41. Brady, R.-M.A.; Stettner, J.L.; York, L. Healthy spaces: Legal tools, innovations, and partnerships. *J. Law Med. Ethics* **2019**, *47*, 27–30. [\[CrossRef\]](#)
42. Cumyn, A.; Barton, A.; Dault, R.; Safa, N.; Cloutier, A.-M.; Ethier, J.-F. Meta-consent for the secondary use of health data within a learning health system: A qualitative study of the public’s perspective. *BMC Med. Ethics* **2021**, *22*, 81. [\[CrossRef\]](#)
43. Kaye, J.; Whitley, E.A.; Lund, D.; Morrison, M.; Teare, H.; Melham, K. Dynamic consent: A patient interface for twenty-first century research networks. *Eur. J. Hum. Genet.* **2015**, *23*, 141–146. [\[CrossRef\]](#) [\[PubMed\]](#)
44. Cumyn, A.; Ménard, J.-F.; Barton, A.; Dault, R.; Lévesque, F.; Ethier, J.-F. Patients’ and members of the public’s wishes regarding transparency in the context of secondary use of health data: Scoping review. *J. Med. Internet Res.* **2023**, *25*, e45002. [\[CrossRef\]](#)
45. Vayena, E.; Blasimme, A.; Cohen, I.G. Machine learning in medicine: Addressing ethical challenges. *PLoS Med.* **2018**, *15*, e1002689. [\[CrossRef\]](#)
46. Carrieri, D.; Howard, H.C.; Benjamin, C.; Clarke, A.J.; Dheensa, S.; Doheny, S.; Hawkins, N.; Halbersma-Konings, T.F.; Jackson, L.; Kayserili, H. Recontacting patients in clinical genetics services: Recommendations of the European Society of Human Genetics. *Eur. J. Hum. Genet.* **2019**, *27*, 169–182. [\[CrossRef\]](#)

47. Tilala, M.H.; Chenchala, P.K.; Choppadandi, A.; Kaur, J.; Naguri, S.; Saoji, R.; Devaguptapu, B.; Tilala, M. Ethical considerations in the use of artificial intelligence and machine learning in health care: A comprehensive review. *Cureus* **2024**, *16*, e62443. [\[CrossRef\]](#)
48. El-Hay, T.; Reps, J.M.; Yanover, C. Extensive benchmarking of a method that estimates external model performance from limited statistical characteristics. *npj Digit. Med.* **2025**, *8*, 59. [\[CrossRef\]](#)
49. Nerella, S.; Bandyopadhyay, S.; Zhang, J.; Contreras, M.; Siegel, S.; Bumin, A.; Silva, B.; Sena, J.; Shickel, B.; Bihorac, A. Transformers and large language models in healthcare: A review. *Artif. Intell. Med.* **2024**, *154*, 102900. [\[CrossRef\]](#)
50. AI HLEG. *High-Level Expert Group on Artificial Intelligence: Ethics Guidelines for Trustworthy AI*; European Commission: Luxembourg, 2019.
51. Ponce-Bobadilla, A.V.; Schmitt, V.; Maier, C.S.; Mensing, S.; Stodtmann, S. Practical guide to SHAP analysis: Explaining supervised machine learning model predictions in drug development. *Clin. Transl. Sci.* **2024**, *17*, e70056. [\[CrossRef\]](#)
52. Holm, S.; Ploug, T. Co-reasoning and epistemic inequality in AI supported medical decision-making. *Am. J. Bioeth.* **2024**, *24*, 79–80. [\[CrossRef\]](#) [\[PubMed\]](#)
53. Collins, G.S.; Moons, K.G.; Dhiman, P.; Riley, R.D.; Beam, A.L.; Van Calster, B.; Ghassemi, M.; Liu, X.; Reitsma, J.B.; Van Smeden, M. TRIPOD+ AI statement: Updated guidance for reporting clinical prediction models that use regression or machine learning methods. *BMJ* **2024**, *385*, e078378. [\[CrossRef\]](#)
54. Habli, I.; Lawton, T.; Porter, Z. Artificial intelligence in health care: Accountability and safety. *Bull. World Health Organ.* **2020**, *98*, 251. [\[CrossRef\]](#) [\[PubMed\]](#)
55. Chen, I.Y.; Pierson, E.; Rose, S.; Joshi, S.; Ferryman, K.; Ghassemi, M. Ethical machine learning in healthcare. *Annu. Rev. Biomed. Data Sci.* **2021**, *4*, 123–144. [\[CrossRef\]](#)
56. Otles, E.; Oh, J.; Li, B.; Bochinski, M.; Joo, H.; Ortwine, J.; Shenoy, E.; Washer, L.; Young, V.B.; Rao, K. Mind the performance gap: Examining dataset shift during prospective validation. In Proceedings of the Machine Learning for Healthcare Conference, Virtual, 6–7 August 2021; PMLR: pp. 506–534.
57. Cabitza, F.; Campagner, A.; Soares, F.; de Gadiana-Romualdo, L.G.; Challa, F.; Sulejmani, A.; Seghezzi, M.; Carobene, A. The importance of being external. methodological insights for the external validation of machine learning models in medicine. *Comput. Methods Programs Biomed.* **2021**, *208*, 106288. [\[CrossRef\]](#)
58. Casey, A.; Dunbar, S.; Gruber, F.; McInerney, S.; Falis, M.; Linksted, P.; Wilde, K.; Harrison, K.; Hamilton, A.; Cole, C. Privacy-Aware, Public-Aligned: Embedding Risk Detection and Public Values into Scalable Clinical Text De-Identification for Trusted Research Environments. *arXiv* **2025**, arXiv:2506.02063.
59. Panigutti, C.; Hamon, R.; Hupont, I.; Fernandez Llorca, D.; Fano Yela, D.; Junklewitz, H.; Scalzo, S.; Mazzini, G.; Sanchez, I.; Soler Garrido, J. The role of explainable AI in the context of the AI Act. In Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, Chicago, IL, USA, 12–15 June 2023; pp. 1139–1150.
60. Sokol, K.; Fackler, J.; Vogt, J.E. Artificial intelligence should genuinely support clinical reasoning and decision making to bridge the translational gap. *npj Digit. Med.* **2025**, *8*, 345. [\[CrossRef\]](#)
61. Dwivedi, R.; Dave, D.; Naik, H.; Singhal, S.; Omer, R.; Patel, P.; Qian, B.; Wen, Z.; Shah, T.; Morgan, G. Explainable AI (XAI): Core ideas, techniques, and solutions. *ACM Comput. Surv.* **2023**, *55*, 1–33. [\[CrossRef\]](#)
62. Ghosheh, G.O.; Thwaites, C.L.; Zhu, T. Synthesizing electronic health records for predictive models in low-middle-income countries (LMICs). *Biomedicine* **2023**, *11*, 1749. [\[CrossRef\]](#)
63. Adams, J. Introducing the ethical-epistemic matrix: A principle-based tool for evaluating artificial intelligence in medicine. *AI Ethics* **2025**, *5*, 2829–2837. [\[CrossRef\]](#)
64. Dankwa-Mullan, I. Health equity and ethical considerations in using artificial intelligence in public health and medicine. *Prev. Chronic Dis.* **2024**, *21*, E64. [\[CrossRef\]](#)
65. Radanliev, P. Privacy, ethics, transparency, and accountability in AI systems for wearable devices. *Front. Digit. Health* **2025**, *7*, 1431246. [\[CrossRef\]](#)
66. Papagiannidis, E.; Mikalef, P.; Conboy, K. Responsible artificial intelligence governance: A review and research framework. *J. Strateg. Inf. Syst.* **2025**, *34*, 101885. [\[CrossRef\]](#)
67. Boeschoten, L.; Voorvaart, R.; Van Den Goorbergh, R.; Kaandorp, C.; De Vos, M. Automatic de-identification of data download packages. *Data Sci.* **2021**, *4*, 101–120. [\[CrossRef\]](#)
68. Morris, J.X.; Champion, T.R.; Nutheti, S.L.; Peng, Y.; Raj, A.; Zabih, R.; Cole, C.L. Diri: Adversarial patient reidentification with large language models for evaluating clinical text anonymization. *AMIA Summits Transl. Sci. Proc.* **2025**, *2025*, 355. [\[PubMed\]](#)
69. Norgeot, B.; Muenzen, K.; Peterson, T.A.; Fan, X.; Glicksberg, B.S.; Schenk, G.; Rutenberg, E.; Oskotsky, B.; Sirota, M.; Yazdany, J. Protected Health Information filter (Philter): Accurately and securely de-identifying free-text clinical notes. *npj Digit. Med.* **2020**, *3*, 57. [\[CrossRef\]](#) [\[PubMed\]](#)
70. Gallifant, J.; Afshar, M.; Ameen, S.; Aphinyanaphongs, Y.; Chen, S.; Cacciamani, G.; Demner-Fushman, D.; Dligach, D.; Daneshjou, R.; Fernandes, C. The TRIPOD-LLM reporting guideline for studies using large language models. *Nat. Med.* **2025**, *31*, 60–69. [\[CrossRef\]](#)

71. Lucero, R.J.; Kearney, J.; Cortes, Y.; Arcia, A.; Appelbaum, P.; Fernández, R.L.; Luchsinger, J. Benefits and risks in secondary use of digitized clinical data: Views of community members living in a predominantly ethnic minority urban neighborhood. *AJOB Empir. Bioeth.* **2015**, *6*, 12–22. [[CrossRef](#)] [[PubMed](#)]
72. Safran, C.; Bloomrosen, M.; Hammond, W.E.; Labkoff, S.; Markel-Fox, S.; Tang, P.C.; Detmer, D.E. Toward a national framework for the secondary use of health data: An American Medical Informatics Association White Paper. *J. Am. Med. Inform. Assoc.* **2007**, *14*, 1–9. [[CrossRef](#)] [[PubMed](#)]
73. Bhattacharjee, S.; Chavan, A.; Huang, S.; Deshpande, A.; Parameswaran, A. Principles of dataset versioning: Exploring the recreation/storage tradeoff. *Proc. VLDB Endowment.* **2015**, *8*, 1346. [[CrossRef](#)]
74. Foale, P.L.; Silva, L.D.; Khomh, F.; Li, H.; Merlo, E. Logging requirement for continuous auditing of responsible machine learning-based applications. *Empir. Softw. Eng.* **2025**, *30*, 97. [[CrossRef](#)]
75. Hao, B.; Hu, Y.; Sotudian, S.; Zad, Z.; Adams, W.G.; Assoumou, S.A.; Hsu, H.; Mishuris, R.G.; Paschalidis, I.C. Development and validation of predictive models for COVID-19 outcomes in a safety-net hospital population. *J. Am. Med. Inform. Assoc.* **2022**, *29*, 1253–1262. [[CrossRef](#)]
76. Mittelstadt, B.D.; Floridi, L. The ethics of big data: Current and foreseeable issues in biomedical contexts. In *The Ethics of Biomedical Big Data*; Springer: Cham, Switzerland, 2016; pp. 445–480.
77. Leslie, D.; Mazumder, A.; Peppin, A.; Wolters, M.K.; Hagerty, A. Does “AI” stand for augmenting inequality in the era of COVID-19 healthcare? *BMJ* **2021**, *372*, n304. [[CrossRef](#)] [[PubMed](#)]
78. Kim, J.Y.; Hasan, A.; Kueper, J.; Tang, T.; Hayes, C.; Fine, B.; Balu, S.; Sendak, M. Establishing organizational AI governance in healthcare: A case study in Canada. *npj Digit. Med.* **2025**, *8*, 522. [[CrossRef](#)]
79. Kavianpour, S.; Sutherland, J.; Mansouri-Benssassi, E.; Coull, N.; Jefferson, E. Next-generation capabilities in trusted research environments: Interview study. *J. Med. Internet Res.* **2022**, *24*, e33720. [[CrossRef](#)]
80. Al-Maamari, A. Between Innovation and Oversight: A Cross-Regional Study of AI Risk Management Frameworks in the EU, US, UK, and China. *arXiv* **2025**, arXiv:2503.05773.
81. EU Artificial Intelligence Act. The Eu Artificial Intelligence Act. European Union. 2024. Available online: <https://artificialintelligenceact.eu/> (accessed on 1 September 2025).
82. Khan, U.S.; Khan, S.U.R. Ethics by Design: A Lifecycle Framework for Trustworthy AI in Medical Imaging From Transparent Data Governance to Clinically Validated Deployment. *arXiv* **2025**, arXiv:2507.04249. [[CrossRef](#)]
83. Brey, P.; Dainow, B. Ethics by design for artificial intelligence. *AI Ethics* **2024**, *4*, 1265–1277. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.