Cross-Modal Deep Hashing for Medical Image–Text Retrieval via Self-Supervised Asymmetric Semantic Excavation

Ronald Garker
Department of Electrical Engineering and Computer Science, University of Kansas, Lawrence, KS, USA.
parker1988@ku.edu

Viktor D. Lindberg
Department of Computer Science, Binghamton University, Binghamton, NY, USA.
viktor.d.lindberg@binghamton.edu

Lars Kelley
Department of Computer Science, University of New Hampshire, Durham, NH, USA.
larsmail@unh.edu

Abstract

The integration of medical imaging and unstructured clinical narratives has created a multimodal data landscape that promises transformative diagnostic and research capabilities, yet the sheer scale, heterogeneity, and privacy sensitivity of this data challenge conventional retrieval systems. This paper presents a systems-level analysis of cross-modal deep hashing frameworks that exploit self-supervised asymmetric semantic excavation to enable efficient medical image–text retrieval. We depart from traditional symmetric learning paradigms and examine how asymmetric network architectures, combined with margin-scalable semantic constraints, can excavate latent correspondences from unlabeled radiological archives without reliance on costly manual annotations. The discussion extends beyond algorithmic novelty to encompass structural trade-offs in system architecture, including the design of modality-specific encoders, hash code binarization pipelines, and distributed retrieval topologies suitable for hospital information systems. We scrutinize the interplay between retrieval precision and computational efficiency, highlighting how binary hash codes can reduce storage footprints by orders of magnitude while enabling sublinear nearest-neighbor search in high-dimensional joint embedding spaces. Critical attention is devoted to robustness under domain shift caused by varied imaging equipment and heterogeneous reporting styles, as well as to fairness concerns that arise when retrieval performance varies across demographic subgroups, a matter of acute importance in clinical decision support. Furthermore, we address infrastructure governance, sustainability of deep hashing model lifecycles, and the policy implications of deploying self-supervised retrieval tools within regulated healthcare environments. By situating cross-modal deep hashing within a broad socio-technical framework, the paper argues that self-supervised asymmetric semantic excavation offers a viable trajectory toward scalable, interpretable, and ethically grounded medical information access, provided that system design accounts for clinical workflows, regulatory compliance, and long-term maintainability.

Keywords

cross-modal retrieval; deep hashing; self-supervised learning; medical image–text; asymmetric semantic excavation; healthcare infrastructure; fairness; policy.

1. Introduction

The exponential growth of medical imaging repositories and the concomitant accumulation of textual records, ranging from radiology reports to surgical notes, have rendered multimodal data management a central concern for contemporary healthcare systems. Cross-modal retrieval, the task of identifying relevant images from textual queries or vice versa, holds immense promise for clinical decision support, case-based reasoning, and medical education. Yet realizing this promise at scale demands more than incremental algorithmic improvements; it requires a holistic interrogation of system architecture, data governance, and the ethical dimensions that attend to the deployment of artificial intelligence in health-critical domains. Deep hashing has emerged as a particularly attractive paradigm for large-scale retrieval because it transforms high-dimensional multimodal representations into compact binary codes, thereby enabling fast Hamming distance computation and dramatic reductions in memory consumption [1, 2]. This efficiency is indispensable in hospital networks where information systems must serve thousands of concurrent queries under strict latency budgets. Nevertheless, the medical domain introduces unique constraints that challenge off-the-shelf cross-modal hashing solutions, including the scarcity of paired image–text annotations due to the labor-intensive nature of clinical labeling, the heterogeneity of imaging protocols across vendors, the prevalence of longitudinal patient data governed by stringent privacy regulations, and the imperative of fairness across diverse patient populations.

Self-supervised learning has recently galvanized representation learning by demonstrating that rich semantic structures can be excavated from unlabeled data through well-designed pretext tasks [3, 4]. When transposed to cross-modal hashing, self-supervision obviates the need for exhaustive manual alignment between images and texts, a bottleneck that has historically limited the applicability of supervised cross-modal approaches in medicine [5]. The asymmetric semantic excavation framework, in particular, introduces a sophisticated inductive bias wherein the two modalities are processed by structurally dissimilar encoders that learn complementary views of shared semantics, while a margin-scalable constraint adaptively refines the separation between matching and non-matching pairs [6, 7, 8]. This asymmetry mirrors the intrinsic discrepancy between visual and linguistic modalities, where pixel-level features and narrative semantics occupy different manifolds that cannot be naïvely mapped through symmetric siamese architectures. A notable recent formulation integrates self-supervised asymmetric semantic excavation with margin-scalable constraints, achieving state-of-the-art retrieval performance on general domain benchmarks while maintaining code compactness [11]. The present analysis extends such foundational ideas to the medical domain, examining how architectural choices interact with the peculiarities of clinical data pipelines, and foregrounding systemic concerns that are often marginalized in purely algorithmic accounts.

Our work provides a comprehensive systems-oriented examination of cross-modal deep hashing for medical image–text retrieval under a self-supervised asymmetric semantic excavation paradigm. We do not propose a new model; instead, we dissect the design space, evaluate structural trade-offs, and articulate the infrastructural, governance, and policy considerations that must accompany the translation of such technology into real-world healthcare ecosystems. In doing so, we advocate for a viewpoint in which retrieval performance is not the sole optimization criterion, but rather one component within a larger constellation of system-level desiderata that include robustness, fairness, sustainability, and regulatory compliance.

2. Architectural Foundations and Modality-Specific Encoding

At the core of any cross-modal deep hashing system lies the design of modality-specific encoders that project images and texts into a shared semantic space where hashing is performed. In medical contexts, the image encoder must contend with high-resolution volumetric data, such as computed tomography stacks and magnetic resonance imaging sequences, which demand architectures capable of capturing both fine-grained anatomical detail and global contextual patterns. Convolutional neural networks and vision transformers pre-trained on large-scale natural image corpora have been successfully fine-tuned for medical imaging tasks [9, 10], yet domain-specific pretraining on radiologic datasets, utilizing contrastive objectives that align images with radiology reports, has been shown to yield representations that better preserve clinically meaningful semantics [3]. Similarly, the text encoder must process narratives characterized by specialized jargon, telegraphic syntax, and frequent use of negations and hedging expressions. Transformer-based language models fine-tuned on clinical notes, such as variants of BERT adapted to the biomedical domain, offer a promising pathway, although they introduce substantial computational overhead that must be balanced against the need for real-time inference in clinical settings [4].

The asymmetric encoding architecture departs from the symmetric weight-sharing designs common in earlier cross-modal hashing literature [2, 5]. By assigning structurally distinct networks to each modality, the system can tailor inductive biases: the image encoder may employ pooling strategies that emphasize spatial invariances, while the text encoder can exploit attention mechanisms that capture long-range dependencies across sentences. This asymmetry is further justified by the fact that the semantic granularity of radiology reports often does not map one-to-one onto the visual granularity of images; a single sentence may describe a global finding, while a region-of-interest in the image corresponds to a localized abnormality. During self-supervised training, the asymmetric architecture is regularized through a semantic excavation loss that mines informative negative pairs and adaptively adjusts margins based on the difficulty of each pair, as conceptualized in recent margin-scalable hashing frameworks [6, 11]. The system thereby learns to disentangle spurious correlations from genuine semantic correspondences without leveraging explicit pairing labels, a property that is invaluable in medical environments where paired data are limited but large corpora of unaligned images and reports exist.

The binarization stage introduces further architectural trade-offs. The continuous embeddings produced by the encoders must be transformed into binary hash codes through a quantization process, typically via a sign function or stochastic sampling. This non-differentiable operation is handled by approximations such as the straight-through estimator or by maintaining continuous relaxations during training. The choice of code length constitutes a critical hyperparameter that governs the trade-off between retrieval accuracy and storage efficiency. Shorter codes reduce the memory footprint and accelerate Hamming distance computation but risk collapsing semantically distinct concepts into identical hash buckets; longer codes preserve finer semantic distinctions at the expense of increased storage and computational cost. In medical retrieval, where fine-grained differentiation between pathologies is crucial, code lengths must be selected with careful clinical validation, and adaptive hashing schemes that learn variable-length codes conditioned on query complexity represent an important area for future infrastructure development.

3. Self-Supervised Semantic Excavation and Margin-Scalable Constraints

The semantic excavation mechanism underpinning self-supervised cross-modal hashing operates by constructing a latent space in which semantically similar image–text pairs are pulled together while dissimilar pairs are pushed apart, all without direct access to ground-truth alignment labels. This is typically implemented through a combination of inter-modal and intra-modal contrastive objectives. Inter-modal contrast encourages consistency between an image and its corresponding textual description, even when the mapping is one-to-many, as is common in medical records where multiple reports may describe the same imaging study from different clinical perspectives. Intra-modal contrast ensures that similar images or similar texts cluster together, thereby preserving the manifold structure inherent in each modality [7, 8]. The asymmetric formulation ensures that the gradient signals from each modality are modulated independently, permitting the system to prioritize one modality's learning signal when the other modality carries less discriminative information, a phenomenon frequently observed when reports contain boilerplate language that dilutes semantic distinctiveness.

The margin-scalable constraint refines this process by dynamically adjusting the separation margin for each training sample based on its relative difficulty. Hard negative pairs, those that are easily confused by the current model, receive larger margins, while easy negatives are assigned smaller margins. This adaptive curriculum relieves the optimization process from the brittleness of fixed-margin approaches that either converge too slowly or collapse the representation space. Drawing on the margin-scalable constraint formulation introduced in recent deep hashing research [11], the system can maintain a healthy gradient flow throughout training, which is particularly beneficial when learning from medical datasets that exhibit severe class imbalance and long-tailed distributions of findings. The capacity to scale margins also reduces the risk of overfitting to frequent but clinically trivial patterns, thereby enhancing the retrieval of rare pathologies, a robustness property that aligns with fairness goals in diagnostic support.

Semantic excavation in the medical domain must also contend with the phenomenon of semantic overlap across examinations. A chest radiograph showing cardiomegaly may share semantics with a report mentioning "enlarged cardiac silhouette," but also with a report describing pulmonary edema, because these conditions frequently co-occur. The hashing framework must therefore learn to disentangle such overlapping concepts to prevent false positives that could mislead clinicians. The margin-scalable mechanism can be augmented with soft similarity labels derived from structured ontologies such as RadLex or SNOMED CT, which provide a hierarchical understanding of medical concepts. Integrating such knowledge graphs into the self-supervised pipeline, while preserving the annotation-free philosophy, remains an open challenge that intersects with the design of governance-compliant data infrastructures.

4. System Infrastructure, Deployment, and Scalability

Translating a self-supervised asymmetric deep hashing model from a research prototype into a production-ready medical retrieval system necessitates a holistic infrastructure strategy that spans data ingestion, distributed indexing, query routing, and continuous monitoring. In a typical hospital information system, radiological images are stored in Picture Archiving and Communication Systems, while textual reports reside in Electronic Health Record databases. The deep hashing pipeline must interface with these disparate sources, performing on-the-fly feature extraction and hash code generation within latency budgets dictated by clinical workflows [12]. This demands careful orchestration of computational resources, often through a hybrid cloud–edge architecture where heavy encoding workloads are offloaded to high-performance computing clusters while lightweight hash lookup operations are executed on edge servers located within the clinical facility to reduce network delays and mitigate privacy exposure.

The hash code index itself must support high-throughput approximate nearest-neighbor search across potentially billions of entries as healthcare networks aggregate data across multiple hospital sites. Multi-index hashing and inverted multi-index structures can partition the code space to achieve sublinear search times, but they introduce additional complexity in index maintenance and load balancing [13]. In federated settings, where data cannot be centralized due to privacy regulations such as the Health Insurance Portability and Accountability Act in the United States or the General Data Protection Regulation in Europe, the hashing infrastructure must support decentralized index construction while preserving the ability to perform cross-modal queries across institutional boundaries [14]. Self-supervised asymmetric models, which can be trained collaboratively without sharing raw data through federated contrastive learning, present a natural fit for such distributed governance models, although the heterogeneity of local data distributions poses significant robustness challenges that will be discussed in the following section.

Monitoring and lifecycle management of the retrieval system introduce additional infrastructural demands. Medical knowledge evolves, imaging protocols change, and reporting styles shift over time. Consequently, the underlying data distribution drifts, a phenomenon known as concept drift, which can degrade retrieval relevance unless the hash model is periodically retrained. Automated drift detection mechanisms, coupled with human-in-the-loop validation, are essential to ensure that the system remains clinically safe. Versioned model registries, staged rollouts, and rollback capabilities become as critical to the infrastructure as the hashing algorithm itself. The sustainability of such systems also requires efficient use of computational resources; the adoption of binary representation already contributes to energy-efficient retrieval relative to floating-point embedding search, yet training large dual-encoder models remains carbon-intensive, motivating the development of more sample-efficient self-supervised objectives and model compression techniques [15].

5. Robustness, Fairness, and Clinical Trustworthiness

Robustness in cross-modal medical retrieval encompasses resilience against variations in image acquisition parameters, scanner manufacturers, patient positioning, and linguistic idiosyncrasies across radiologists. A hashing model trained predominantly on chest radiographs from a single institution may fail to generalize when deployed in a different hospital with older equipment or when confronted with reports written in a non-native language style. The asymmetric architecture, by decoupling modality processing, can partially mitigate such domain shifts because the textual encoder may remain stable while the image encoder adapts, or vice versa, but systematic robustness requires domain adaptation strategies that leverage unsupervised test-time normalization and data augmentation designed for medical imaging [16]. Margin-scalable constraints also contribute to robustness by preventing the model from anchoring on easy-to-learn yet domain-specific shortcuts, thereby encouraging the learning of more invariant semantic features.

Fairness emerges as a non-negotiable requirement when retrieval results influence clinical decisions. Cross-modal hashing systems may inadvertently exhibit performance disparities across demographic groups defined by sex, age, or ethnicity if the training data are unrepresentative or if certain findings are systematically underreported for minority populations [17]. For instance, pain assessment in radiographs may be described differently depending on cultural context, and an asymmetric semantic excavation process that relies on narrative patterns risks amplifying these biases. Addressing such disparities necessitates fairness-aware training objectives that penalize group-conditional retrieval variance and the integration of demographic annotations into the self-supervised pipeline in a privacy-preserving manner. Moreover, external auditing mechanisms, including independent evaluation on stratified test sets and the publication of disaggregated performance metrics, should be mandated as part of the deployment governance framework [18].

Trustworthiness further depends on the interpretability of retrieval results. When a clinician queries for “interstitial lung disease” and receives a set of ranked images, the system must provide visual explanations, such as saliency maps indicating which regions of the query text matched which image features, to enable verification and avoid automation bias. While binary hash codes are inherently opaque, explanation modules can be built orthogonally by projecting hash similarities back onto the original high-dimensional embedding space and generating attention visualizations. The adoption of such explainability layers, while adding computational overhead, is a structural feature that must be budgeted for in the overall system architecture from the outset, rather than retrofitted as an afterthought.

6. Governance, Policy, and Regulatory Alignment

The deployment of deep hashing retrieval systems within healthcare is inextricably linked to a dense web of regulatory and ethical obligations that extend far beyond technical performance metrics. In the United States, the Food and Drug Administration has increasingly classified software that provides diagnostic recommendations or prioritizes medical images as a medical device, subjecting it to premarket review and postmarket surveillance requirements [19]. A cross-modal retrieval engine that influences clinical workflow, for example by surfacing historical cases similar to a current patient’s presentation, could fall under such regulatory purview if its outputs are deemed to guide clinical decisions. Consequently, system designers must establish quality management systems, document intended use, and conduct rigorous clinical validation studies that quantify the impact of retrieval errors on patient outcomes. The self-supervised nature of the training process introduces additional regulatory complexity because the absence of explicit labels challenges traceability; documentation must clearly articulate the sources of training data, the pretext tasks employed, and the measures taken to prevent data leakage across patient cohorts.

On the international stage, the European Union’s Artificial Intelligence Act proposes to categorize certain AI applications in healthcare as high-risk, requiring conformity assessments that cover data governance, transparency, accuracy, and robustness [20]. The asymmetric semantic excavation framework can support compliance by providing a well-defined architecture whose modularity facilitates auditing; separate encoders for image and text can be evaluated independently for bias and vulnerability to adversarial perturbation before integration. Furthermore, the margin-scalable constraint, with its tunable sensitivity, offers a direct lever for controlling the trade-off between recall and precision, which can be adjusted according to clinical risk thresholds. Establishing accountability chains that link model design decisions to patient safety outcomes is a prerequisite for regulatory acceptance, and this necessitates interdisciplinary collaboration between engineers, clinicians, ethicists, and legal experts throughout the system lifecycle.

Policy instruments also govern data sharing and model reuse. Open-access medical datasets, while invaluable for training and benchmarking, must be carefully curated to prevent re-identification of patients, especially when high-dimensional image features and free-text reports are linked [21]. Self-supervised learning helps by reducing dependency on paired data, but even unpaired images and texts can leak sensitive information when combined with auxiliary data sources. Federated hashing frameworks that keep data in situ during training and only exchange model updates offer a pathway toward multi-institutional collaboration without centralizing sensitive records, yet they require robust governance protocols to prevent gradient inversion attacks that could reconstruct private data from shared model updates [22]. The tension between open science, model reproducibility, and patient privacy will remain a defining policy challenge for the foreseeable future, and cross-modal retrieval systems must be engineered with privacy-enhancing technologies such as differential privacy and secure aggregation from the ground up.

7. Sustainability, Long-Term Evolution, and System Integration

Sustainability in medical AI systems extends beyond environmental footprints to encompass the long-term viability of the technical solution within evolving healthcare ecosystems. Deep hashing models, once trained, offer significant operational efficiency because binary code storage and Hamming distance computation consume far less energy than floating-point similarity search engines [23]. However, the initial training phase—particularly the contrastive learning of large dual-encoder architectures on massive medical corpora—can have a substantial carbon impact if not optimized through efficient distributed training strategies and the use of renewable energy sources. Lifecycle assessments of retrieval systems should therefore report not only accuracy and speed but also energy and resource consumption metrics, enabling healthcare organizations to align technology choices with their sustainability commitments.

The integration of cross-modal hashing into existing clinical information systems demands standards-based interoperability. The Fast Healthcare Interoperability Resources standard and the Digital Imaging and Communications in Medicine protocol define how images and reports are exchanged, yet they do not yet incorporate native support for semantic hashing indices. Bridging this gap requires middleware that translates retrieval requests into hash code queries, manages index synchronization, and handles versioning as models are updated. This middleware must be designed with fault tolerance and graceful degradation in mind; if the deep hashing service becomes unavailable, the system should fall back to conventional keyword-based search rather than disrupting clinical operations entirely.

As medical knowledge expands, retrieval systems will need to accommodate emerging modalities such as histopathology whole-slide images, genomics reports, and wearable sensor data. The asymmetric architecture can be extended to multiple modalities by treating each as a distinct encoding pathway that converges in a shared Hamming space, a multi-modal generalization that raises its own set of optimization and governance challenges. Self-supervised objectives that align an arbitrary number of modalities through learned semantic correspondences will become increasingly important. The margin-scalable constraint's adaptability positions it as a versatile regularizer in these extended settings, provided that the semantic excavation process is carefully calibrated to prevent modality collapse, where one dominant modality overwhelms the hash code representation. Planning for such extensibility at the architectural level can future-proof investments in cross-modal retrieval infrastructure and reduce the total cost of ownership for healthcare networks.

8. Conclusion

This paper has provided a systematic examination of cross-modal deep hashing for medical image–text retrieval through the lens of self-supervised asymmetric semantic excavation, emphasizing structural design, infrastructural deployment, robustness, fairness, regulatory alignment, and sustainability. We have argued that the asymmetric encoding paradigm, combined with margin-scalable semantic constraints, offers a principled means of extracting clinically meaningful correspondences from unlabeled medical data while accommodating the profound asymmetries between visual and linguistic modalities. However, the translation of these algorithmic capabilities into trustworthy clinical tools demands that system architects attend to a far broader set of concerns than accuracy maximization alone. The choice of code length, encoding architecture, and training objective must be scrutinized for its implications on retrieval latency, storage costs, energy consumption, and resilience to domain shift. Fairness auditing and interpretability mechanisms must be woven into the system fabric, not bolted on retroactively. Governance frameworks must reconcile the self-supervised training paradigm with regulatory requirements for traceability, clinical validation, and data protection. Only through such comprehensive integration can cross-modal deep hashing fulfill its potential as a safe, equitable, and enduring component of the modern healthcare information infrastructure. Future research should pursue federated hashing protocols under formal privacy guarantees, develop standardized benchmarks that capture diverse patient populations and rare diseases, and pioneer lifecycle assessment methodologies tailored to medical AI retrieval systems.

References

1. Wang, K., Yin, Q., Yang, Y., & Wang, W. (2016). A comprehensive survey on cross-modal retrieval. arXiv preprint arXiv:1607.06215.

2. Jiang, Q.-Y., & Li, W.-J. (2017). Deep cross-modal hashing. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3232–3240).

3. Chen, T., Kornblith, S., Norouzi, M., & Hinton, G. (2020). A simple framework for contrastive learning of visual representations. In Proceedings of the International Conference on Machine Learning (pp. 1597–1607).

4. Lee, J., Yoon, W., Kim, S., Kim, D., Kim, S., So, C. H., & Kang, J. (2020). BioBERT: a pre-trained biomedical language representation model for biomedical text mining. Bioinformatics, 36(4), 1234–1240.

5. Shen, F., Xu, Y., Liu, L., Yang, Y., & Shen, H. T. (2019). Asymmetric deep cross-modal hashing. IEEE Transactions on Knowledge and Data Engineering, 32(3), 507–519.

6. Cao, Y., Long, M., Wang, J., & Liu, S. (2018). Deep Cauchy hashing for hamming space retrieval. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 1229–1237).

7. Hu, D., Nie, F., & Li, X. (2019). Deep binary reconstruction for cross-modal hashing. IEEE Transactions on Multimedia, 21(4), 973–985.

8. Lin, Z., Ding, G., Hu, M., & Wang, J. (2016). Semantics-preserving hashing for cross-view retrieval. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 3864–3872).

9. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., & Sánchez, C. I. (2017). A survey on deep learning in medical image analysis. Medical Image Analysis, 42, 60–88.

10. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., Uszkoreit, J., & Houlsby, N. (2021). An image is worth 16x16 words: Transformers for image recognition at scale. In International Conference on Learning Representations.

11. Yu, Z., Wu, S., Dou, Z., & Bakker, E. M. (2022). Deep hashing with self-supervised asymmetric semantic excavation and margin-scalable constraint. Neurocomputing, 483, 87-104.

12. Shivade, C., Raghavan, P., Fosler-Lussier, E., Embi, P. J., Elhadad, N., Johnson, S. B., & Lai, A. M. (2014). A review of approaches to identifying patient phenotype cohorts using electronic health records. Journal of the American Medical Informatics Association, 21(2), 221–230.

13. Jegou, H., Douze, M., & Schmid, C. (2011). Product quantization for nearest neighbor search. IEEE Transactions on Pattern Analysis and Machine Intelligence, 33(1), 117–128.

14. Rieke, N., Hancox, J., Li, W., Milletarì, F., Roth, H. R., Albarqouni, S., Bakas, S., Galtier, M. N., Landman, B. A., Maier-Hein, K., Ourselin, S., Sheller, M., Summers, R. M., Trask, A., Xu, D., Baust, M., & Cardoso, M. J. (2020). The future of digital health with federated learning. NPJ Digital Medicine, 3(1), 1–7.

15. Strubell, E., Ganesh, A., & McCallum, A. (2019). Energy and policy considerations for deep learning in NLP. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics (pp. 3645–3650).

16. Dou, Q., Coelho de Castro, D., Kamnitsas, K., & Glocker, B. (2019). Domain generalization via model-agnostic learning of semantic features. In Advances in Neural Information Processing Systems (Vol. 32).

17. Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447–453.

18. Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. In Proceedings of the Conference on Fairness, Accountability, and Transparency (pp. 220–229).

19. U.S. Food and Drug Administration. (2021). Proposed regulatory framework for modifications to artificial intelligence/machine learning-based software as a medical device. Discussion Paper.

20. European Commission. (2021). Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). COM(2021) 206 final.

21. Mandl, K. D., & Kohane, I. S. (2015). New federal patient privacy rules—HIPAA 2.0. New England Journal of Medicine, 373(2), 109–111.

22. Zhu, L., Liu, Z., & Han, S. (2019). Deep leakage from gradients. In Advances in Neural Information Processing Systems (Vol. 32).

23. Gong, Y., Lazebnik, S., Gordo, A., & Perronnin, F. (2013). Iterative quantization: A procrustean approach to learning binary codes for large-scale image retrieval. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(12), 2916–2929.