Position: Ophthalmology as a Lens for Trustworthy GenAI in Europe—Uncertainty-Aware AI under the EU AI Act

Mariya Erokhina[1,2] **Achref Doula**[1,3] **Lukas Bisorca-Gassendorf**[1,4] **Alejandro Sanchez Guinea**[1,3]
[1]EyeTrustAI* [2]UPF Barcelona School of Management [3]TU Darmstadt [4]Augenklinik Petrisberg
[erokhina.mariya, achref.doula, dr.bisorca, ale.sanchez.guinea]@gmail.com

Abstract

Artificial intelligence is rapidly advancing in ophthalmology, offering data-driven models for early detection, triage, and decision support. In Europe, adoption is shaped by an unusually strict regulatory landscape defined by the parallel application of the Medical Device Regulation (MDR) and the Artificial Intelligence Act (AI Act). The MDR emphasizes safety and performance, while the AI Act imposes horizontal obligations for high-risk AI systems, including transparency, robustness, and human oversight. Although adopted in 2024, the AI Act's provisions will phase in through 2026–27, creating a moving target for compliance. This dual environment risks raising entry barriers and slowing innovation but also opens opportunities: uncertainty-aware methods, such as calibrated confidence estimates, out-of-distribution detection, and risk communication, can directly address transparency and reliability requirements while aligning with clinical trust needs. Using ophthalmology as a lens, we argue that uncertainty quantification can turn Europe's regulatory strictness from bottleneck into enabler for trustworthy GenAI in healthcare.

1 Introduction

Ophthalmology has long been a testbed for data-driven medicine. Imaging modalities such as fundus photography and OCT have enabled early detection of diabetic retinopathy, glaucoma, and age-related macular degeneration [Gulshan et al., 2016, Ting et al., 2017, De Fauw et al., 2018]. Generative AI (GenAI) now extends this trajectory, supporting synthetic data generation [Yu et al., 2018], automated triage [Esteva et al., 2021], and multimodal decision support [Avramidis et al., 2022]. It is important to distinguish that synthetic data may carry different regulatory implications depending on its use: while training-time augmentation can support robustness and privacy, operational use in clinical decision-making may trigger distinct MDR and AI Act classifications. Yet integration into European healthcare faces a uniquely demanding landscape: the parallel application of the Medical Device Regulation (MDR) [MDR, 2017] and the Artificial Intelligence Act (AI Act) [EUA, 2024]. The MDR focuses on demonstrating medical device safety and performance, while the AI Act introduces horizontal obligations for high-risk AI systems, including risk management, record-keeping, technical documentation, interpretability, and human oversight [Veale and Zuiderveen Borgesius, 2023]. Crucially, the AI Act uses "transparency" in a narrow legal sense: it refers not to explainability in general, but to concrete duties such as providing documentation for regulators, clear information for users, and controls that enable meaningful oversight [Smuha, 2021, Wachter and Mittelstadt, 2021].

For ophthalmology, this dual regime is a particular hurdle. GenAI models often fall short not because they are unsafe, but because they are not yet designed to evidence compliance in conformity

*https://www.eyetrustai.com

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: The Second Workshop on GenAI for Health: Potential, Trust, and Policy Compliance.

assessment [He et al., 2019, Faes et al., 2020]. Evidence gaps include lack of calibration, insufficient uncertainty reporting, and limited documentation of distributional validity. Uncertainty-aware methods address these gaps: calibrated confidence, out-of-distribution detection, and risk communication can indicate when predictions are reliable and when they should be escalated to human experts [Kompa et al., 2021, Abdar et al., 2021]. These methods align directly with the MDR's safety rationale and the AI Act's oversight mandate. In this paper, we use ophthalmology as a lens to argue how uncertainty-aware design can transform Europe's regulatory strictness from bottleneck into enabler for trustworthy GenAI in healthcare.

2 The Regulatory Landscape in Europe

The European Union provides one of the most stringent environments for deploying AI in healthcare, shaped by the parallel application of the Medical Device Regulation (MDR) [MDR, 2017] and the Artificial Intelligence Act (AI Act) [EUA, 2024]. Ophthalmic AI systems used for diagnosis, triage, or treatment are typically classified as medical devices, often Class IIa or IIb, reflecting their influence on patient management [Faes et al., 2020, He et al., 2019]. Conformity assessment is therefore central to adoption. The AI Act, adopted in 2024, extends scrutiny by designating most medical AI as "high-risk," requiring risk management, technical documentation, accuracy, robustness, and human oversight [Veale and Zuiderveen Borgesius, 2023]. While calibration or uncertainty estimation are not mandated, they can operationalize obligations such as reliability monitoring or risk communication [Kompa et al., 2021, Abdar et al., 2021]. Still, legal compliance demands more than technical fixes: hazard analysis, mitigation, and lifecycle monitoring remain essential.

Oversight requirements illustrate this interplay. Article 14 of the AI Act obliges that high-risk AI remain subject to "human oversight," which may involve clinicians, technicians, or other trained personnel, supported by design features that clarify system purpose and limitations [Smuha, 2021, Wachter and Mittelstadt, 2021]. Uncertainty signals can help, but only if paired with appropriate training and interfaces. Together, the MDR and AI Act form a dual pathway where safety, performance, and oversight must be evidenced under overlapping but distinct regimes. For ophthalmic GenAI, the dual challenge is to meet clinical standards while simultaneously producing regulatory evidence [He et al., 2019, Faes et al., 2020]. Yet this landscape can also become an enabler: uncertainty-aware design aligns technical reliability with legal obligations, offering a pathway to compliance and clinical trust [Kompa et al., 2021, Abdar et al., 2021, Gawlikowski et al., 2023].

3 Challenges in Deploying GenAI for Ophthalmology

Clinical Challenges. Ophthalmology workflows already rely on escalation and role sharing. Introducing GenAI heightens concerns about accountability. Clinicians often perceive liability as personal, even though under EU product liability rules responsibility may be shared with manufacturers [PLD, 1985, 2022]. This gap between perceived personal liability and formal legal allocation can itself discourage adoption, amplifying the trust deficit clinicians report when facing opaque AI systems. Oversight may be exercised by clinicians or other trained staff, so interfaces and alerts must direct the right cases to the right human. Uncertainty signals (i.e., calibrated confidence, abstentions, triage thresholds) serve as *oversight affordances*, supporting Article 14 of the AI Act while matching ophthalmic workflows [Kompa et al., 2021, Gawlikowski et al., 2023]. Effective oversight should also address system-level uncertainty such as pipeline failures or sensor degradation.

Technical Challenges. GenAI faces risks of unreliable predictions under distribution shift and hallucinations in generated outputs. Calibration is fragile when deployment diverges from training data, undermining escalation logic [Guo et al., 2017, Ovadia et al., 2019]. Methods such as ensembles, evidential approaches [Sensoy et al., 2018], and conformal prediction [Vovk et al., 2005, Angelopoulos and Bates, 2021] can improve reliability, providing explicit abstentions and coverage guarantees. Guardrails such as selective generation, constrained decoding, and uncertainty-aware triage should therefore be treated as *oversight affordances*, translating risk management requirements into operational behaviors clinicians can act upon.

System-Level Challenges. At scale, uncertainty-aware GenAI must navigate the General Data Protection Regulation (GDPR) and reimbursement hurdles. GDPR requires lawful bases, safeguards, and robust anonymization, constraining cross-border pooling of ophthalmic data and raising questions

about re-identification even from synthetic data [GDP, 2016, WP2, 2014, Miotto et al., 2018]. Beyond data governance, adoption also depends on Health Technology Assessment (HTA) and reimbursement: EU-wide joint clinical assessments will interact with national payer decisions [HTA, 2021]. Without demonstrated outcome gains and cost-effectiveness, tools risk exclusion from reimbursement and limited impact [He et al., 2019]. Here, uncertainty evidence (e.g., subgroup calibration or shift detection) can improve comparative-effectiveness models and strengthen payer confidence, thereby increasing the likelihood of integration into routine care [Topol, 2019, Carvalho et al., 2019].

Ophthalmology as Microcosm. These barriers converge acutely in ophthalmology. Its high imaging volume, early AI adoption, and medico-legal sensitivity make it an ideal testbed for uncertainty-aware design under MDR and AI Act. Benchmark datasets such as EyePACS, UK Biobank, and REFUGE [EyePACS, 2015, Sudlow et al., 2015, Orlando et al., 2020] have provided fertile ground for uncertainty quantification research in ophthalmic imaging. Together, these developments position ophthalmology as an influential testbed for advancing uncertainty-aware methods. Lessons generalize across imaging-intensive fields (e.g., radiology, dermatology, pathology), where uncertainty-aware GenAI can similarly bridge clinical trust and regulatory obligations. Importantly, this generalization is *conceptual rather than methodological*: it concerns policy design and compliance templates more than direct transferability of UQ pipelines across domains.

4 Opportunities through Uncertainty-Aware GenAI

Europe's dual regulatory framework creates barriers but also unique opportunities: uncertainty-aware GenAI offers concrete mechanisms to bridge regulatory obligations and clinical adoption. By embedding calibrated confidence, distribution shift monitoring, and oversight affordances, compliance can be reframed as an enabler of safe deployment. Table 1 summarizes how these methods can map onto specific regulatory duties and clinical benefits.

Regulatory Alignment. The AI Act does not mandate confidence measures but requires risk management, documentation, and oversight [EUA, 2024, Veale and Zuiderveen Borgesius, 2023]. Uncertainty Quantification (UQ) methods operationalize these duties: calibrated thresholds for risk management, OOD detection for data governance, uncertainty summaries for documentation, coverage guarantees for transparency, triage interfaces for oversight, and ensemble-based checks for robustness [Kompa et al., 2021, Angelopoulos and Bates, 2021, Ovadia et al., 2019, Gawlikowski et al., 2023]. It is important to stress that "transparency" in the AI Act (Art. 13) refers to concrete legal duties (i.e., documentation, user information, and oversight controls) rather than the broader research concept of "explainability" [Smuha, 2021, Wachter and Mittelstadt, 2021]. In this paper we use "explainability" in the medical AI literature sense (interpretability for clinicians), and not as a legal category.

Prioritizing practical methods. Among available UQ techniques, *calibration* remains the most directly auditable and interpretable for regulators, providing quantifiable reliability evidence for the MDR technical file. *Shift detection* ranks next in importance by enabling lifecycle monitoring and dataset quality assurance under Articles 9–10 of the AI Act. *Conformal prediction* offers formal coverage guarantees valuable for transparency but currently lacks large-scale clinical validation. *Ensemble and evidential approaches* complement these but may be resource-intensive. Hence, calibration and shift monitoring form the most practical near-term priorities for regulatory alignment.

As Table 1 highlights, uncertainty quantification provides a bridge between technical reliability and legal compliance.

From prototypes to compliance. Bridging research prototypes and compliance-ready systems requires explicit validation, auditing, and documentation pipelines. Calibration metrics and uncertainty summaries should appear in the *technical documentation* required by the MDR, supported by traceable logging and periodic audits mandated by the AI Act. Model cards, version control of calibration data, and structured uncertainty reports can facilitate post-market monitoring and harmonized conformity assessment without fundamentally changing clinical workflows.

Trustworthiness and Clinical Adoption. Clinician acceptance depends on actionable uncertainty communication [Gawlikowski et al., 2023]. Studies show that calibrated intervals and interface cues

[2]Transparency is a legal duty under Art. 13; explainability here refers to interpretability in the medical AI literature [EUA, 2024, Veale and Zuiderveen Borgesius, 2023].

Table 1: Illustrative mappings between EU regulatory requirements, example uncertainty-aware approaches, and potential clinical benefits in ophthalmology. The approaches shown are *examples only*, i.e., they are representative of possible techniques, not prescriptive recommendations. Other methods may fulfill these requirements in different or complementary ways.

Regulatory Requirement (EU AI Act / MDR)	**Example Approaches (non-prescriptive)**	**Potential Clinical Benefit (illustrative)**
Transparency (AI Act, Art. 13, legal duty) + Explainability (technical/research term)[2]	Conformal prediction, calibrated scores [Angelopoulos and Bates, 2021, Vovk et al., 2005]	Provide reliability cues; may support clinician trust [Kompa et al., 2021, Topol, 2019]
Risk management	Bayesian methods, OOD detection [Gal and Ghahramani, 2016, Ovadia et al., 2019]	Flag ambiguous or novel cases; may reduce silent failures [He et al., 2019]
Human oversight	Triage thresholds, clinician-in-the-loop with audit trails [Kompa et al., 2021, Topol, 2019]	Enable meaningful intervention and accountability [EUA, 2024, MDR, 2017]
Robustness (MDR + AI Act, Art. 15)	Ensembles, shift monitoring [Ovadia et al., 2019, Gawlikowski et al., 2023]	Support safety across diverse devices/populations [Faes et al., 2020, Ting et al., 2017]

improve trust and decision-making [Tonekaboni et al., 2019, Carvalho et al., 2019, Topol, 2019]. In practice, uncertainty-aware workflows can manifest in several forms: (a) *screening clinics* where AI auto-grades retinal images and defers ambiguous cases to ophthalmologists; (b) *teleophthalmology triage* systems that flag uncertain referrals for in-person review; and (c) *intra-hospital routing* tools that dynamically assign high-risk images to senior specialists. Each scenario demonstrates how calibrated confidence and oversight affordances can integrate into everyday ophthalmic care while maintaining regulatory traceability.

Leveraging European Heterogeneity. Diversity across populations, devices, and workflows can be reframed as an asset. The AI Act requires representative datasets and documented biases [EUA, 2024, EUAIACT Art. 10]; MDR emphasizes evidence across populations and devices [MDR, 2017]. UQ enables subgroup calibration, conformal coverage, and drift detection to produce auditable fairness and robustness evidence [Angelopoulos and Bates, 2021, Vovk et al., 2005]. For ophthalmology, stratified reliability claims strengthen safety files and HTA evaluations [HTA, 2021, Topol, 2019]. Moreover, subgroup-specific calibration and tail-risk detection directly address demographic fairness concerns by ensuring that performance and uncertainty estimates remain stable across under-represented populations and long-tail cases—a core expectation of both the MDR and AI Act fairness provisions.

Ophthalmology as Proof-of-Concept. Ophthalmology combines tractability (ocular imaging) with high clinical stakes, making it an ideal pilot for uncertainty-aware pipelines. Benchmark datasets (EyePACS, UK Biobank, REFUGE) already support UQ research [EyePACS, 2015, Sudlow et al., 2015, Orlando et al., 2020]. Success here could establish templates for radiology, dermatology, and other imaging specialties, positioning Europe as a leader in trustworthy GenAI. Its relatively contained scope (ocular images) makes it tractable for piloting uncertainty-aware pipelines, while lessons on trust, liability, and workflow integration can generalize to other fields.

5 Broader Implications and Recommendations

Although framed through ophthalmology, our analysis extends to other imaging-intensive specialties. The strategies discussed (i.e., uncertainty quantification for transparency, hybrid human-AI workflows, and regulatory alignment) generalize to radiology, dermatology, and pathology. Ophthalmology's role here is illustrative: it shows how uncertainty-aware GenAI can meet Europe's strict demands while providing transferable insights. The same evidence base (stratified performance, subgroup calibration, shift monitoring) also supports downstream adoption decisions by HTA bodies and payers beyond ophthalmology [HTA, 2021, He et al., 2019, Topol, 2019]. Europe's demanding regime may thus serve as a proving ground: demonstrating compliance in ophthalmology creates benchmarks for broader trustworthy GenAI.

Table 2: Illustrative stakeholder-specific challenges and possible recommendations for deploying uncertainty-aware GenAI in European healthcare. These mappings are representative examples, not prescriptive instructions.[3]

Stakeholder	Current Challenge	Illustrative Recommendation
Developers	Limited interpretability complicates compliance with AI Act duties for transparency, documentation, and oversight [EUA, 2024, Veale and Zuiderveen Borgesius, 2023].	Embed UQ and trustworthy design early; use calibrated thresholds, OOD/shift monitoring, clinician-in-the-loop overrides, and logging to support conformity assessment [Kompa et al., 2021, Angelopoulos and Bates, 2021, Ovadia et al., 2019, Gawlikowski et al., 2023].
Regulators	Strict requirements risk slowing innovation; uncertainty reporting, subgroup evidence, and logging remain under-specified [Smuha, 2021, Veale and Zuiderveen Borgesius, 2023].	Provide clear guidance and leverage AI regulatory sandboxes (Arts. 57–58) with ophthalmology pilots to co-develop practical controls and documentation [EUA, 2024, He et al., 2019, Topol, 2019].
Clinicians & Health Systems	Trust deficit and medico-legal liability; need to integrate AI Act and GDPR logging/record-keeping into workflows [GDP, 2016, EUA, 2024, Holzinger et al., 2022].	Adopt hybrid workflows; link outputs and uncertainty to decisions/outcomes in EHRs; provide training and UI to mitigate automation bias; support standards for interoperable uncertainty logging (e.g., HL7 FHIR) [Kompa et al., 2021, He et al., 2019, Topol, 2019].

Developers should embed uncertainty quantification and trustworthy design from the outset: calibrated thresholds, OOD/shift monitoring, clinician-in-the-loop overrides, and comprehensive logging can operationalize risk management and populate the technical file [Kompa et al., 2021, Angelopoulos and Bates, 2021, Ovadia et al., 2019, Gawlikowski et al., 2023]. Robust evaluation protocols (e.g., covering calibration, subgroup reliability, and shift resilience) should accompany these design choices to ensure claims are both clinically and regulatorily defensible.

Regulators should provide clear guidance and actively use AI regulatory sandboxes (Articles 57–58 of the EU AI Act) to co-develop uncertainty reporting, subgroup evidence, and logging practices with ophthalmology pilots [EUA, 2024, Veale and Zuiderveen Borgesius, 2023]. Such sandboxes can also shape international practice, positioning European pilots as templates for global regulatory innovation.

Clinicians and health systems should adopt hybrid workflows where routine cases are accelerated and ambiguous ones escalated; integrate logging of outputs, uncertainty summaries, decisions, and outcomes into the EHR; and support research on interoperable standards such as HL7 FHIR extensions [Kompa et al., 2021, Holzinger et al., 2022]. These cross-cutting perspectives are summarized in Table 2, which outlines representative challenges and illustrative recommendations for each stakeholder group.

Limitations and future work. Despite their promise, uncertainty-aware methods face several technical and clinical limitations. Calibration often degrades under distribution shifts, ensembles and conformal predictors can add computational overhead, and reliable uncertainty visualization in clinical interfaces remains an open challenge. Clinical validation studies are still limited in scale and diversity, and prospective audits of uncertainty communication are rare. Most current evidence addresses diagnostic imaging rather than generative use cases, where evaluation protocols for hallucination control are still emerging. Future work should combine multi-center calibration benchmarks, clinician-in-the-loop usability testing, and regulatory sandbox pilots to mature these methods from research proof-of-concepts to compliance-ready systems.

[3]Legal note: The EU AI Act does not ban "black-box" models. High-risk systems must demonstrate compliance through documentation, testing, logging, transparency, and human oversight. We use "black-box" to denote limited interpretability; compliance can still be shown via evidence and controls [EUA, 2024, Veale and Zuiderveen Borgesius, 2023].

References

Council directive 85/374/eec of 25 July 1985 on the Approximation of the laws, regulations and administrative provisions of the member states concerning liability for defective products. Official Journal of the European Communities L 210/29, 1985. URL https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A31985L0374.

Opinion 05/2014 on anonymisation techniques. Technical Report WP216, Article 29 Data Protection Working Party, 2014.

Regulation (eu) 2016/679 of the european parliament and of the council (general data protection regulation), 2016. OJ L 119, 4.5.2016, p. 1–88.

Regulation (eu) 2017/745 of the european parliament and of the council on medical devices, 2017. Official Journal of the European Union, L117, 1–175.

Regulation (eu) 2021/2282 on health technology assessment and amending directive 2011/24/eu, 2021. OJ L 458, 22.12.2021, p. 1–82.

Proposal for a directive on liability for defective products (recast). COM(2022) 495 final, 28 September 2022, 2022. URL https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX:52022PC0495.

Regulation (eu) 2024/1689 laying down harmonised rules on artificial intelligence (ai act), 2024. Official Journal of the European Union, L2024.

Moloud Abdar, Farhad Pourpanah, Saeid Hussain, Dana Rezazadegan, Li Liu, Mohammad Ghavamzadeh, Paul Fieguth, Xiaochun Cao, Abbas Khosravi, U. Rajendra Acharya, Vladimir Makarenkov, and Saeid Nahavandi. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Information Fusion*, 76:243–297, 2021. doi: 10.1016/j.inffus.2021.05.008.

Anastasios N Angelopoulos and Stephen Bates. Conformal prediction: A unified review of theory and new challenges. *arXiv preprint arXiv:2107.07511*, 2021.

Georgios Avramidis, Stefano Bini, Tobias Kersten, and et al. Multimodal deep learning for medical imaging and healthcare: Review, opportunities and challenges. *Artificial Intelligence in Medicine*, 133:102394, 2022. doi: 10.1016/j.artmed.2022.102394.

Diogo V Carvalho, Eduardo M Pereira, and Jaime S Cardoso. Machine learning in clinical practice: prospects and pitfalls. *Computers in Biology and Medicine*, 113:103672, 2019.

Jeffrey De Fauw, Joseph R Ledsam, Bernardino Romera-Paredes, et al. Clinically applicable deep learning for diagnosis and referral in retinal disease. *Nature Medicine*, 24(9):1342–1350, 2018.

Andre Esteva, Eric Topol, et al. Deep learning-enabled medical computer vision. *npj Digital Medicine*, 4(1):5, 2021.

EyePACS. Diabetic retinopathy detection. https://www.kaggle.com/c/diabetic-retinopathy-detection, 2015. Accessed: 2025-09-05.

Livia Faes, Saskia Wagner, Darren J Fu, et al. Artificial intelligence and the eu medical device regulation: the dawn of a new era for medical ai? *Eye*, 34:2213–2216, 2020.

Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model uncertainty in deep learning. In *International Conference on Machine Learning (ICML)*, pages 1050–1059, 2016.

Jakob Gawlikowski, Crina R Tassi, Mohammad Ali, et al. A survey of uncertainty in deep neural networks. *Artificial Intelligence Review*, 56:1513–1589, 2023.

Varun Gulshan, Lily Peng, Marc Coram, Martin C. Stumpe, Derek Wu, Arunachalam Narayanaswamy, Subhashini Venugopalan, Katrina Widner, Travis Madams, Jorge Cuadros, and et al. Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs. *JAMA*, 316(22):2402–2410, 2016. doi: 10.1001/jama.2016.17216.

Chuan Guo, Geoff Pleiss, Yu Sun, and Kilian Q Weinberger. On calibration of modern neural networks. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, pages 1321–1330, 2017.

Jing He, Sally L. Baxter, Jieying Xu, Jimeng Xu, Xiaoqian Zhou, and Kang Zhang. Practical barriers to deploying machine learning in healthcare. *Nature Medicine*, 25(1):30–36, 2019.

Andreas Holzinger, Chris Biemann, Constantinos S Pattichis, and Douglas B Kell. Toward trustworthy ai in medicine. *ACM Computing Surveys*, 55(6):1–38, 2022.

Benjamin Kompa, Jasper Snoek, and Andrew L Beam. Second opinion needed: communicating uncertainty in medical machine learning. *NPJ Digital Medicine*, 4(1):4, 2021.

Riccardo Miotto, Fei Wang, Shuang Wang, Xiaoqian Jiang, and Joel T Dudley. Deep learning for healthcare: review, opportunities and challenges. *Briefings in Bioinformatics*, 19(6):1236–1246, 2018.

José Ignacio Orlando, Huazhu Fu, Joana Barbosa Breda, Karel van Keer, Deepak Rajan Bathula, Andres Diaz-Pinto, Rui Fang, Pheng-Ann Heng, Jeong Hun Kim, Jong Chul Lee, et al. Refuge challenge: A unified framework for evaluating automated methods for glaucoma assessment from fundus photographs. In *Medical Image Analysis*, volume 59, page 101570. Elsevier, 2020.

Yaniv Ovadia, Emily Fertig, Jianfeng Ren, Zachary Nado, D Sculley, Sebastian Nowozin, Joshua V Dillon, Balaji Lakshminarayanan, Jasper Snoek, Dustin Tran, et al. Can you trust your model's uncertainty? evaluating predictive uncertainty under dataset shift. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 32, 2019.

Murat Sensoy, Lance Kaplan, and Melih C. Kandemir. Evidential deep learning to quantify classification uncertainty. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 31. Curran Associates, Inc., 2018.

Nathalie A. Smuha. Beyond a human rights-based approach to ai governance: Promise and limits of the eu ai act. *Law, Innovation and Technology*, 13(3):336–375, 2021. doi: 10.1080/17579961.2021.1953036.

Cathie Sudlow, John Gallacher, Naomi Allen, Valerie Beral, Paul Burton, John Danesh, Paul Downey, Paul Elliott, Jane Green, Martin Landray, et al. Uk biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Medicine*, 12(3):e1001779, 2015.

Daniel Shu Wei Ting, Carol Yim-Lui Cheung, Geoffery Lim, Gavin Siew Wei Tan, Niyazi Det Quang, Adeline Gan, Husain Hamzah, Rocio Garcia-Franco, Ie Keat Yeo, Sze Hwee Lee, and et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA*, 318(22):2211–2223, 2017. doi: 10.1001/jama.2017.18152.

Sana Tonekaboni, Shalmali Joshi, Melissa D. McCradden, and Anna Goldenberg. Clinicians and medical artificial intelligence: The need for clarity in evaluation and regulation. *Nature Medicine*, 25:1467–1468, 2019.

Eric J Topol. High-performance medicine: the convergence of human and artificial intelligence. *Nature Medicine*, 25(1):44–56, 2019.

Michael Veale and Frederik Zuiderveen Borgesius. Demystifying the eu artificial intelligence act. *Computer Law Review International*, 24(3):97–112, 2023.

Vladimir Vovk, Alex Gammerman, and Glenn Shafer. *Algorithmic Learning in a Random World*. Springer, 2005.

Sandra Wachter and Brent Mittelstadt. Why fairness cannot be automated: Bridging the gap between eu non-discrimination law and ai. *Computer Law & Security Review*, 41:105567, 2021. doi: 10.1016/j.clsr.2021.105567.

Lequan Yu, Hao Chen, Qi Dou, Jing Qin, and Pheng-Ann Heng. Generative adversarial networks for medical image analysis: A review. In *Proceedings of the IEEE International Symposium on Biomedical Imaging (ISBI)*, pages 201–204, 2018. doi: 10.1109/ISBI.2018.8363567.