ELEC0139
Final Blog Submission

Student Number : 23048826

May 9, 2025

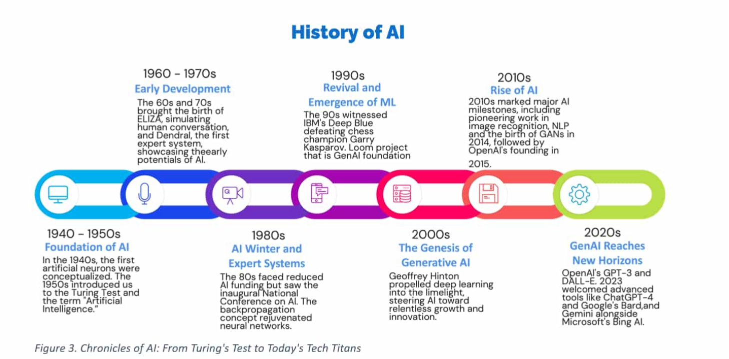
Abstract

This report contains the PDF version of the final blog. It is verified that the total word count of the blog is 2510 Words. The online version hosted on a website can be seen by clicking HERE or using this link : https://ucl-elec0139-blog.netlify.app . The online blog consists of 4 posts and this report version is divided into 3 parts as mentioned in the moodle guide.

Part 1 of the Blog

Evolution of AI from Theory to Reality

In the 1950s, Alan Turing had posed his famous question “Can Machines Think ?”. Back then, he considered the question to be vague & ambiguous and instead proposed the Turing Test as a way to measure the intelligence of machines [1]. It had human evaluators judge conversation transcripts between a machine and a human and identify which was the machine. Seven decades later, we are closer than ever to hitting this benchmark with the latest large language models (LLMs) like GPT o4, Deepseek R1 easily passing the Turing Test [2]. They are capable of complex thought & reasoning, fluent dialogue, and even image generation. AI/ML models have been theorized for and even implemented such as the M.P. neuron or the Perceptron for over half a century. From these theoretical concepts, it has evolved massively with the potential to be embedded in all aspects of modern life as seen below.



History of AI
1940 – 1950s
Foundation of AI
1960 – 1970s
Early Development
1980s
AI Winter and Expert Systems
1990s
Revival and Emergence of ML
2000s
The Genesis of Generative AI
2010s
Rise of AI
2020s
GenAI Reaches New Horizons

In the 1940s, the first artificial neurons were conceptualized. The 1950s introduced us to the Turing Test and the term "Artificial Intelligence."

The 60s and 70s brought the birth of ELIZA, simulating human conversation, and Dendral, the first expert system, showcasing theearly potentials of AI.

The 80s faced reduced AI funding but saw the inaugural National Conference on AI. The backpropagation concept rejuvenated neural networks.

The 90s witnessed IBM's Deep Blue defeating chess champion Garry Kasparov. Loom project that is GenAI foundation

Geoffrey Hinton propelled deep learning into the limelight, steering AI toward relentless growth and innovation.

2010s marked major AI milestones, including pioneering work in image recognition, NLP and the birth of GANs in 2014, followed by OpenAI's founding in 2015.

OpenAI's GPT-3 and DALL-E. 2023 welcomed advanced tools like ChatGPT-4 and Google's Bard,and Gemini alongside Microsoft's Bing AI.

Figure 3. Chronicles of AI: From Turing's Test to Today's Tech Titans

Figure 1: History of AI

While its capabilities are real and growing, so is the hype surrounding it with many arguing that we are currently in an AI bubble where companies are adding AI features that weren't even asked for such as smart AI enabled toothbrushes or another AI startup whose entire business model is just a wrapper around ChatGPT. But while gimmicky AI products dominate headlines, the healthcare sector shows where AI might actually be revolutionary, benefitting significantly from recent advancements in AI.

The Global Healthcare Crisis

Currently, healthcare systems around the world are in unprecedented strain due to multiple challenges. One of the main issues is the growing gap between patient demand and availability of medical facilities. The World Health Organisation (WHO) estimates a shortage of 10 million health sector workers by 2030, with it disproportionately affecting low and middle income countries [3] (as seen in this map below made by the WHO). Even in developed countries like the UK, the National Health Service (NHS) is currently facing a record backlog with millions of patients still awaiting treatment with delays in surgery, appointments and even diagnostics [4] as seen in this graph.

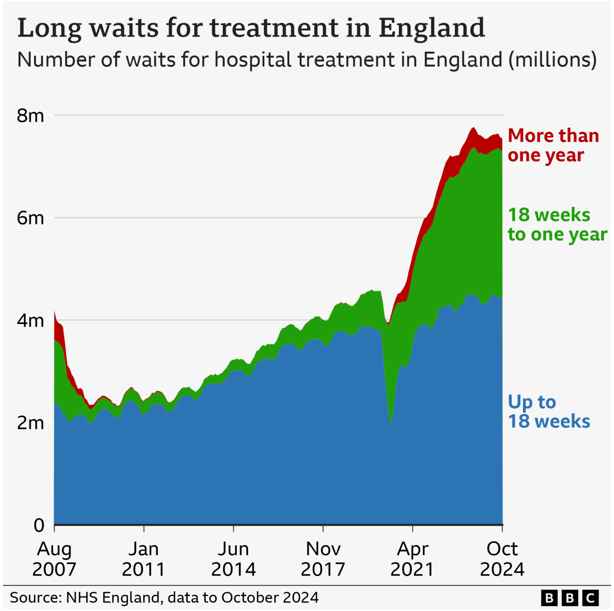
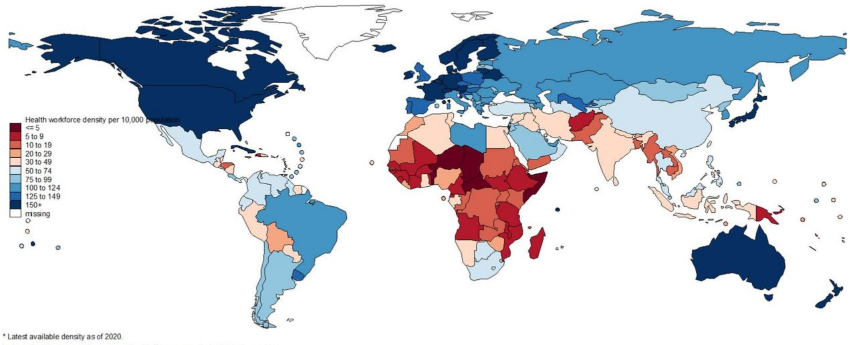


Figure 2: NHS Backlog



* Latest available density as of 2020.
incl. medical doctors, nursing personnel, midwifery personnel, dentists, pharmacists

Figure 3: WHO map showing current density of healthcare workers

The issue is made even worse due to aging populations in developed countries and negative population growth (as seen in this figure of Italy's pop. pyramid), growth in life expectancy has led to issues like diabetes, cancer and heart disease more common. The demographic issues put extra pressure on a system that's already stretched thin. Healthcare worker burnout has reached critical levels made worse by the Covid-19 pandemic [5]. Studies have shown that higher levels of stress & fatigue among medical workers have caused lower productivity and increased medical errors.

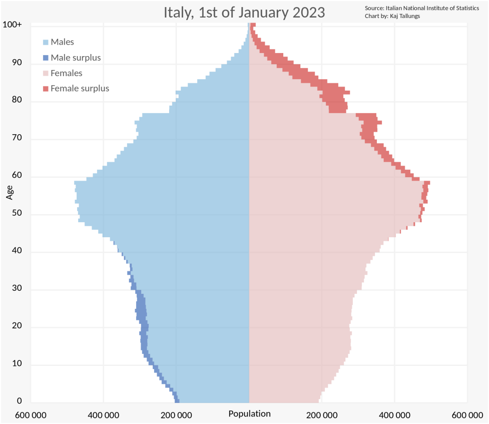


Figure 4: Population Pyramid of Italy (2023)

Another major issue facing the sector is inefficiencies in the healthcare delivery process. Manual documentation and using outdated systems often lead to poor coordination leading to misdiagnosis and delayed treatment. Studies have shown that this affects 1 in 20 patients in the USA, a figure that is extremely high when considering the exorbitant costs of healthcare in the country [6]. Access to healthcare is also another issue. Rural areas often suffer from limited access to specialist staff, testing equipment, or even having no hospitals itself. Medical problems deemed less important like mental health services are usually underfunded and understaffed in such rural areas [7].

How AI/ML can Increase Accuracy/Efficiency and Improve Access

AI and ML techniques can offer a viable and much needed augmentation to the entire domain. Advancements in computer vision and deep learning have created previously unimaginable diagnostic capabilities. Studies have shown that AI based systems utilizing computer vision can match or even outperform expert radiologists in tasks like detecting pneumonia from chest X-rays or even identifying early signs of cardiovascular disease. DeepMind by Google developed an AI model that outperformed human specialists in breast cancer detection from mammograms. There are already startups like Viz.ai and Aidoc [8], that help hospitals integrate AI image analysis tools into their workflow. Furthermore, AI systems can also analyze patient records and real-time health data to create personal treatment plans unique to each patient. It has the potential to dramatically expand access to healthcare in lower resource regions of the world, supporting healthcare workers by automating triage and reducing burden on the already overworked staff.

Figure 5: Futuristic AI integration in pediatric healthcare

AI chatbots and virtual assistants have already been deployed by the NHS (North West London NHS Trust) during the Covid-19 global pandemic which is a real example of AI expanding access to healthcare and reducing the burden on healthcare staff. The chatbots were available 24/7, offer basic advice regarding symptoms, triage non-emergency queries and even followed up previous messages like a human conversation. This automation helped reduce the burden on clinical staff and freed up doctors to focus on more complex cases [9]. Early preventative care is another area where AI shows significant potential. ML models can identify patients at risk well in advance of traditional screening methods by analysing data such as lifestyle choices, genetic data and health records. By having earlier interventions, such systems can reduce hospital visits which lowers healthcare costs and can save lives. The integration of AI with wearable health devices like fitness trackers, is further reshaping the healthcare landscape. They allow patients to monitor various vital health metrics themselves without needing to visit clinics.

However, the integration of AI into healthcare comes with its own issues regarding bias, ethical concerns, lack of transparency, and questions of accountability. Thus, it is essential that fairness, trust and safety are prioritized considering healthcare is a field where accuracy and speed could be the difference between recovery and death. The next section of the blog will cover how different AI/ML technologies address the aforementioned challenges in the healthcare domain in much more detail and is complemented with a brief description of such ML/AI technologies.

Part 2 of the Blog

Part 3 of the Blog

The previous sections of the blog have shown the effect AI/ML can have on the healthcare domain but while these technologies offer unprecedented capabilities in treatment and diagnostics, they also raise serious ethical, legal and societal issues. Problems like bias, lack of accountability & transparency, data privacy concern cause an overall lack of trust in such AI systems. In the UK, the National Health Service (NHS) has actively tried to integrate AI so as to increase efficiency and clinical outcomes. However, they have encountered multiple issues while doing so which will be described in more detail in this section of the blog as well as the risk if AI systems are left unchecked and what the UK and EU are doing to ensure that such technologies are used fairly and ethically.

Bias & Fairness of AI in Healthcare

One of the main ethical issues in the usage of AI for healthcare is algorithmic bias. Bias happens when an AI model unintentionally favors some groups over others because of imbalances or flaws in the dataset used to train the ML models. In healthcare, such biases can even lead to clinical harm and erode people's trust in the system. A recent example of this was pulse oximeters, which were found to overestimate blood oxygen levels in humans with darker skin tones. Although no AI was involved here, it highlights how even commonly used medical devices can have systemic flaws and biases [10]. Imagine the outcome if similar issues happened in AI systems due to flaws in the data used.

Studies have shown that ML models trained to detect melanoma performed much worse on test subjects with darker skin once again, it was attributed to the underrepresentation of such patients in the training dataset. ML models for chest X-rays and cardiovascular risk have also performed more poorly for minority classes like women and ethnic minorities. These cases are not coincidence, they are caused by using training datasets that disproportionately represent white, male, and affluent patients, this reinforces bias and inequalities in clinical outcomes. Also, when AI models are evaluated only on aggregate accuracy, their underperformance on smaller and underserved classes often goes unnoticed.

Bias can also arise from historical data that reflects disparities in procedure. Studies have shown that black patients in the UK have historically been prescribed less pain medication than white patients for the same conditions, due to racially biased assumptions about black people having a higher pain tolerance. If an AI system is trained on prescription records that reflect these disparities, it may think that certain ethnic groups need less pain medication. In practice, this could lead the model to recommend less aggressive pain treatment for black patients.

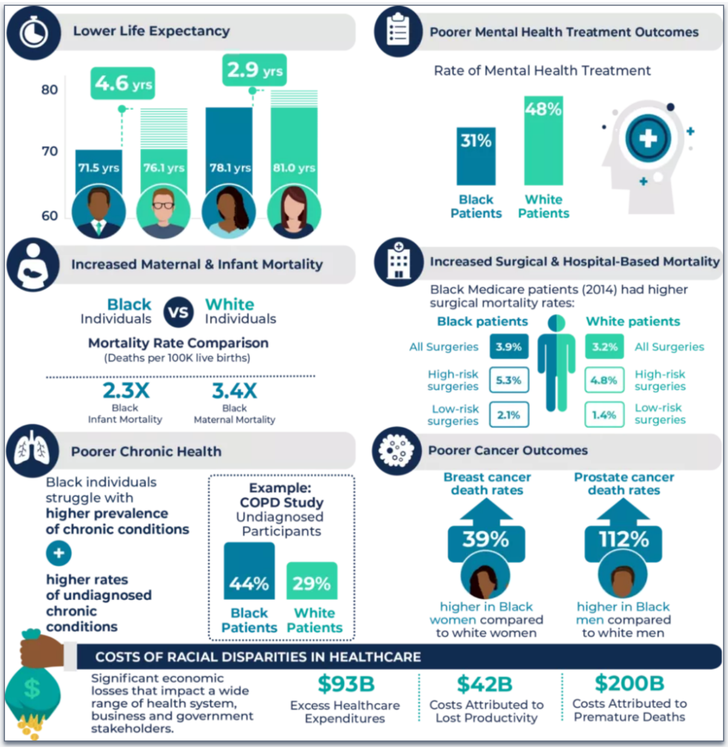


Figure 6: Racial Bias in healthcare in the USA

This creates a positive feedback loop where patients receive less care which is used to train the AI further resulting in future patients receiving even less care because the AI deems from the training data that they don't need as much medication. This is extremely dangerous and is effectively automated systemic discrimination [11]. Some methods can be used to reduce such bias resulting in more fair and equitable AI models such as:

- Diverse data sourcing to reflect real-world demographics.
- Subgroup performance analysis instead of just overall accuracy.
- Fairness-aware model selection, using metrics like equalized odds or demographic parity.

A UK government review found that many medical devices, including AI ones show performance gaps across ethnic and gender lines. It recommended actions to consider fairness at every stage of the AI lifecycle from how data is collected to how systems are monitored after deployment [12]. Bias in Healthcare is not just a matter of ethics, it is a clinical issue which if left unaddressed risks making existing inequalities even worse than it already is.

The Need for Explainability & Trust

A major issue with many AI systems (especially complex deep learning models) is their lack of interpretability. In the case of healthcare, when workers aren't able to interpret the working and understand how the AI model came up with the conclusion it reached. This opacity can cause serious problems, this is known as the Black Box Problem. There are tools for explaining how AI's came to a decision like saliency maps and feature attribution methods which shows what part of an input influenced the models decision. But, these explanations are often misleading or unstable, highlighting areas that affect the output but don't mean anything clinically. This causes healthcare workers to not clearly trust the model's reasoning. Research on trust in explainable AI (XAI) shows mixed results. When the explanations are clear and clinically relevant, they can improve confidence and lead to better diagnosing of patients, but when the explanations are vague/misleading, it can backfire, with workers relying overly on AI to the point that they trust the output, even when it's wrong or not sensible [13]. Development of Interpretable by design AI models can solve this issue, some methods they use are:

- Generalized additive models (GAMs), they weigh individual risk factors in a transparent manner to show how they come to some prediction.
- Rule-based systems, they back up their predictions with if-then logic that healthcare workers can use to verify the results.

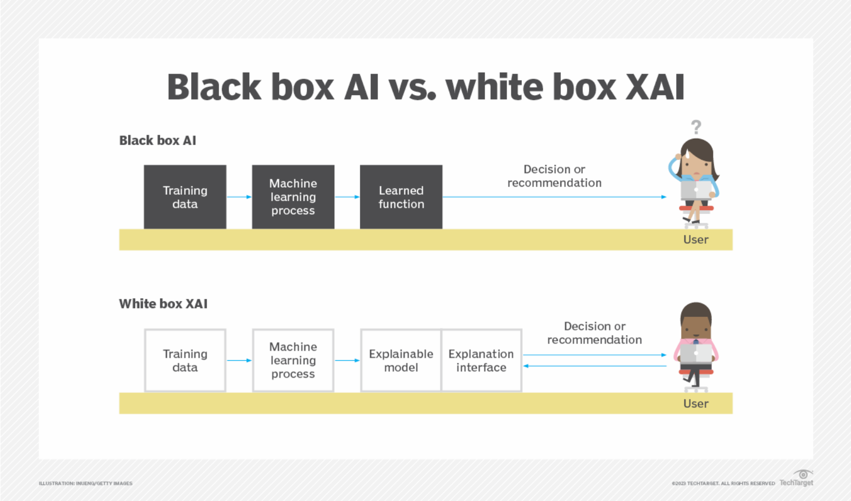


Figure 7

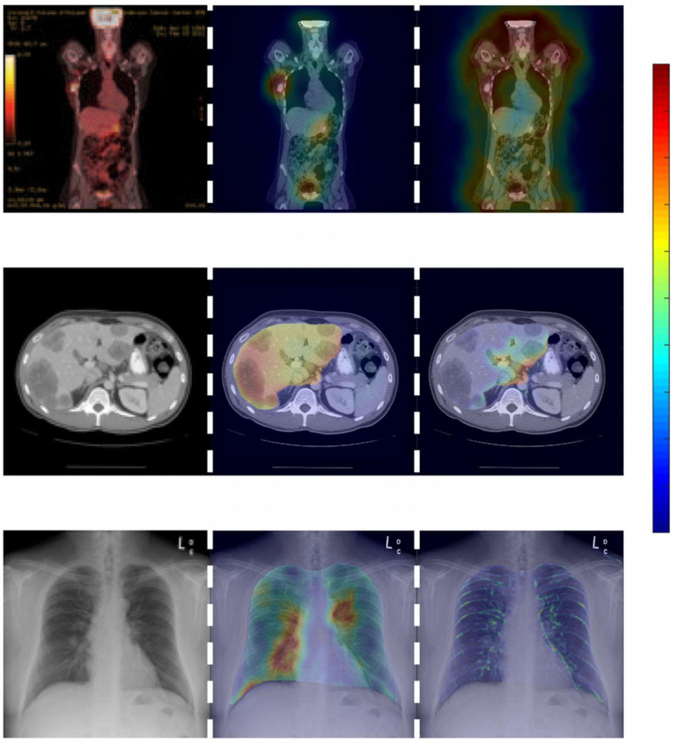


Figure 8: Saliency Maps

Data Privacy, Consent, and Governance

Another major issue is the need of sensitive health data for training Healthcare AI models. Most developed countries have strict privacy regulations regarding medical data like electronic health records (EHRs), genomic data, and diagnostic image results. In the UK, there is the Data Protection Act 2018 which is aligned with the General Data Protection Regulation (GDPR) [14]. AI research faces multiple real world barriers, Patient data is often siloed across multiple NHS trusts making it hard to train robust AI models using all the data. Even if the trusts collaborate, there are still complex approval processes and consent needed to gain access to the patient data. Deidentification techniques are usually used to anonymize the data but studies have shown that it is possible to reidentify over 80% of individuals in anonymized datasets using ML models trained on public data which is a huge privacy breach [15]. Furthermore, some medical data like facial structure in cranial MRI's are impossible to anonymize without losing out vital information. To mitigate these issues, the NHS have used the following privacy preserving methods: - Federated Learning (FL), models are trained using decentralized data sources without actually moving the patient data off-site. - Differential Privacy (DP), adds statistical noise to ML model outputs which prevents reverse engineering of individual records making it harder to reidentify individuals. - Trusted Research Environments (TRE), are secure digital workspaces where researchers can analyze sensitive data without being able to download, copy, or misuse it thus minimizing the risk of data exposure [16].

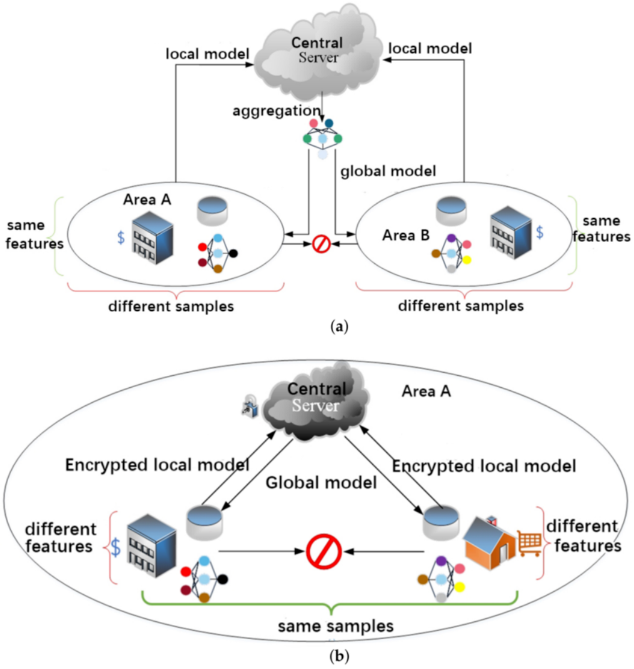


Figure 9: Federated Learning

Legal Accountability and Shared Liability

As AI gets integrated into the medical field, it raises a lot of questions about accountability. If one of these models misdiagnoses a patient or outright suggests the wrong treatment plan, who is liable for it ? Is it the doctor who used the AI tool or the hospital, or the original developer that created said tool. Currently, the UK does not have any AI specific liability ruling, instead general medical rules can be applied. In this case:

- The healthcare worker can be held responsible for over relying on the AI tool
- The developer company can be held for creating software deemed defective.

- The hospital can be held responsible for failure in properly implementing the software without oversight.

But the question of Liability gets messy if we're considering more complex AI systems with autonomy. There are AI systems that make decisions autonomously without any human override or transparency. If such systems misdiagnose, the responsibility will likely fall purely on the developers since the healthcare staff had no role in it. However, if AI systems are only used as a suggestion tool with the hospital staff having the final say in all decisions then the responsibility could be on the person who made the decision entirely [17]. The Medicines and Healthcare products Regulatory Agency (MHRA) maintains that AI tools are only to be used for assistance and that human staff should remain the final decision makers. However, as AI grows more complex, a more detailed framework is needed to determine responsibility in such cases.

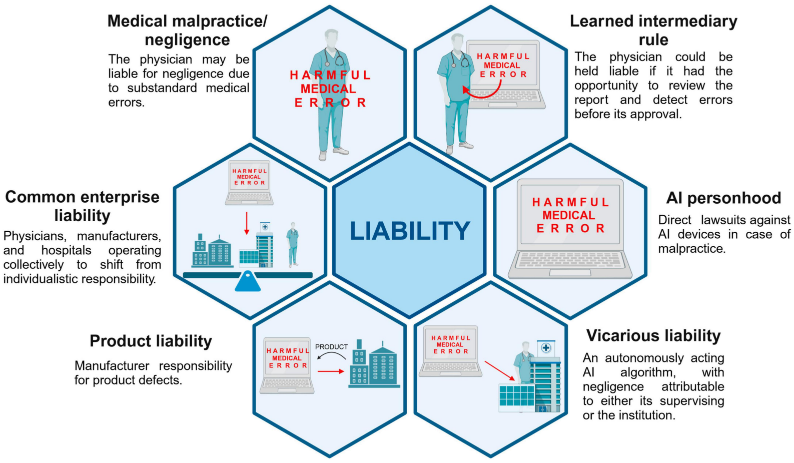

Medical malpractice/ negligence
The physician may be liable for negligence due to substandard medical errors.
HARMFUL MEDICAL ERROR
Learned intermediary rule
The physician could be held liable if it had the opportunity to review the report and detect errors before its approval.
Common enterprise liability
Physicians, manufacturers, and hospitals operating collectively to shift from individualistic responsibility.
LIABILITY
AI personhood
Direct lawsuits against AI devices in case of malpractice.
Product liability
Manufacturer responsibility for product defects.
PRODUCT
Vicarious liability
An autonomously acting AI algorithm, with negligence attributable to either its supervising or the institution.

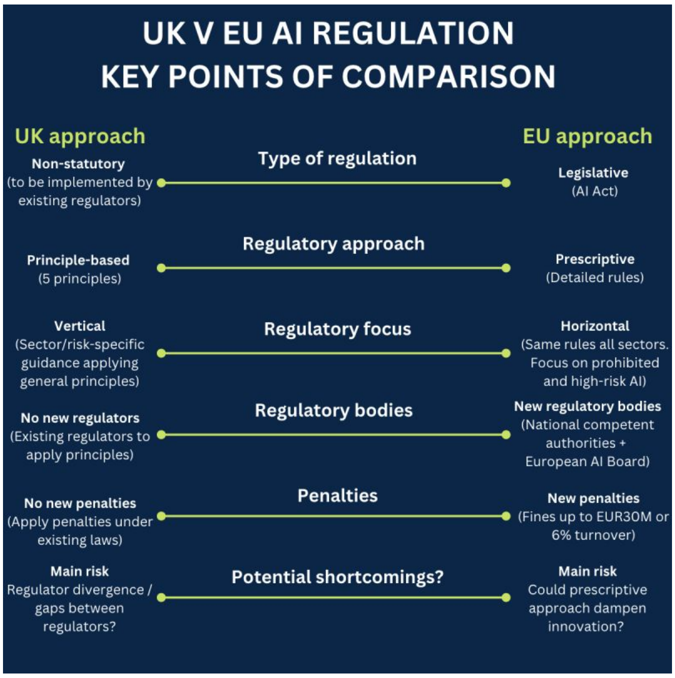
Figure 10: More Examples of Legal Accountability

Regulation: UK and EU Approaches

Regulators in various countries are solving these ethical issues with different strategies. The EU's AI Act, passed in 2023, is the world's most comprehensive framework for high-risk AI applications such as those used in the healthcare domain [18]. The EU AI Act contains sections on mandatory risk assessments, bias mitigation, requirements for human oversight and even mandates on transparency such as disclosing limitations. The UK meanwhile has opted for a more flexible approach to AI regulation. Instead of adopting a standalone AI bill, it has issued regulation white papers and asked existing bodies like the MHRA to adapt their ruling and to consider the role of AI [19]. They have also introduced several practices to support safe and ethical AI such as:

- The AI Airlock Sandbox, which allows for real-world testing of new AI technologies in a safe environment.
- Revised Good Machine Learning Practice (GMLP) guidelines.
- Updated medical device classification rules to reflect the risks posed by AI
- Collaboration with international regulators like Health Canada to equalize AI safety standards

Ultimately, the goal is not just to build intelligent AI systems, we need to also ensure that the systems are operating justly and supporting healthcare delivery in a way that protects everyone including all minorities.



UK V EU AI REGULATION
KEY POINTS OF COMPARISON

UK approach		EU approach
Non-statutory (to be implemented by existing regulators)	Type of regulation	**Legislative** (AI Act)
Principle-based (5 principles)	Regulatory approach	**Prescriptive** (Detailed rules)
Vertical (Sector/risk-specific guidance applying general principles)	Regulatory focus	**Horizontal** (Same rules all sectors. Focus on prohibited and high-risk AI)
No new regulators (Existing regulators to apply principles)	Regulatory bodies	**New regulatory bodies** (National competent authorities + European AI Board)
No new penalties (Apply penalties under existing laws)	Penalties	**New penalties** (Fines up to EUR30M or 6% turnover)
Main risk Regulator divergence / gaps between regulators?	Potential shortcomings?	**Main risk** Could prescriptive approach dampen innovation?

Figure 11: UK vs EU on AI Regulation

Conclusions to the Blog

References

[1] A. M. Turing, "Computing machinery and intelligence," *Mind*, vol. 59, no. 236, pp. 433–460, 1950.

[2] C. R. Jones and B. K. Bergen, "People cannot distinguish gpt-4 from a human in a turing test," *arXiv preprint arXiv:2405.08007*, 2024.

[3] W. H. Organization, "Health workforce: Overview," 2023.

[4] B. M. Association, "Nhs backlog data analysis," 2023.

[5] W. H. Organization, "Mental health and psychosocial considerations during the covid-19 outbreak," 2020.

[6] H. S. et al., "The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving us adult populations," *BMJ Quality & Safety*, vol. 23, no. 9, pp. 727–731, 2014.

[7] Philips, "How ai can expand access to care in low-resource areas," *https://www.usa.philips.com/healthcare/article/ai-expands-care-low-resource-areas*, 2024.

[8] Viz.ai, "Viz.ai one – ai-powered care coordination," *https://www.viz.ai/vizai-one*, 2024.

[9] H. Communications, "Virtual assistants and chatbots," *https://healthcare-communications.com/solutions/virtual-assistants-and-chatbots/*, 2024.

[10] D. of Health and S. Care, "Equity in medical devices: Independent review," 2024.

[11] C. Stetler, "Ai algorithms used in healthcare can perpetuate bias," *Rutgers University News*, 2024.

[12] D. of Health and S. Care, "Independent review on equity in medical devices," *https://www.gov.uk/government/publications/independent-review-on-equity-in-medical-devices*, 2024.

[13] E. Tjoa and C. Guan, "Explainability for artificial intelligence in healthcare," *BMC Medical Informatics and Decision Making*, vol. 20, no. 1, pp. 1–9, 2020.

[14] I. C. Office, "Guidance on ai and data protection," 2023.

[15] W. Price and I. Cohen, "Privacy and artificial intelligence: challenges for protecting health information in a new era," *BMC Medical Ethics*, vol. 22, no. 1, pp. 1–6, 2021.

[16] U. H. D. R. Alliance, "Trusted research environments: A strategy to build public trust and meet changing health data science needs," 2020.

[17] J. Morley and L. Floridi, "Artificial intelligence in health care: accountability and safety," *Bulletin of the World Health Organization*, vol. 98, no. 4, pp. 251–256, 2020.

[18] E. Parliament, "Eu ai act: first regulation on artificial intelligence," 2023.

[19] Medicines and H. products Regulatory Agency, "Mhra's ai regulatory strategy ensures patient safety and industry innovation into 2030," 2024.