Middlesex University

UNIVERSITY OF MIAMI
ETHICS PROGRAMS

**Proceedings of the MDX 2025 Health IT Workshop on**

# Emerging Technologies in Healthcare: Legal, Ethical, Social & Governance Aspects

## 26th & 27th June 2025
## Middlesex University, London, UK

**Proceedings of the MDX 2025 Health IT Workshop on**

***Emerging Technologies in Healthcare: Legal, Ethical, Social & Governance Aspects***

**26th & 27th June 2025**
**Middlesex University, London, UK**

**Editors:**
Carlisle GEORGE (*Middlesex University, UK*),
Diane WHITEHOUSE (*EHTEL, Belgium*),
Penny DUQUENOY (*Middlesex University, UK)*

https://tinyurl.com/MDX-WKS-2025

ISBN 979-8-89965-689-7


Proceedings DOI: https://doi.org/10.60528/2q6192

## Foreword

Before Artificial Intelligence changed the world, before Covid, before some countries abandoned if not attacked ancient scientific traditions, there was a clear need for scholars and others to identify and make sense of the many ethical issues raised by the use of computers in health care.

That need was anticipated by few – two of which were, however, Middlesex University and the University of Miami. Their subsequent decade-long collaboration has made significant contributions to education and research on precisely those issues.

The 2025 Emerging Technologies in Healthcare workshop is the 7th in a series hosted at Middlesex University. It has come to be one of the world's leading venues for sharing innovative scholarship on ethics and biomedical informatics.

This workshop continues a tradition of featuring intellectually and geographically diverse scholars and others. Their areas of expertise include computer science, medicine, ethics and philosophy, law and behavioural science; they come from Belgium, Egypt, the United Kingdom and the United States of America; they work for universities, nongovernmental organizations and industry. And the topics they address are as compelling and challenging as ever: computational intelligence, integrity and accountability, data and science, mobile health technology and others.

Perhaps most significantly, this multidisciplinary and interprofessional approach has afforded unique opportunities for students.

The workshop occurs at an extraordinary time in international science, education, commerce and communication. It is a period in which longstanding customs are under attack, in which falsehoods are re-imagined as "misinformation" and in which legitimate trust is being wilfully undermined. All of this is all the more reason for fostering international collaborations and open exchanges of the sort represented by this important workshop.

**Prof. Kenneth W. Goodman**
Professor Emeritus of Medicine and jointly of Philosophy
University of Miami Miller School of Medicine Institute for Bioethics and Health Policy, USA.

**Contents**

## Welcome from the Pro Vice Chancellor for Global Research and Knowledge Exchange

Emerging technologies such as artificial intelligence (AI), mHealth, telemedicine, and the Internet of Things are transforming healthcare delivery worldwide. While these innovations offer immense benefits to patients and healthcare professionals alike, they also present complex challenges that demand our careful attention, in particular relating to legal, ethical, social and governance issues. Understanding these challenges and collaboratively developing solutions to mitigate undesirable or potentially harmful effects is vital for the well-being of society.

Universities like Middlesex play a pivotal role in advancing research at the intersection of healthcare and technology. We serve as a vibrant hub of innovation, bringing together multidisciplinary teams of researchers, clinicians, and other stakeholders to explore cutting-edge solutions to some of healthcare's most pressing challenges.

Here at Middlesex, we take pride in our longstanding commitment to educating healthcare practitioners and driving innovative research in healthcare technologies. This workshop is an important platform for fostering interdisciplinary dialogue and collaboration. It enables us to examine the legal frameworks shaping emerging technologies, to navigate the ethical considerations guiding our decisions, to assess the social impact of integrating these technologies into healthcare systems, and to develop governance strategies that ensure responsible and equitable implementation.

I would like to extend special thanks to Dr Carlisle George, Ms Diane Whitehouse, Dr Penny Duquenoy and Prof Kenneth Goodman for their outstanding work in organising this significant event, which brings together diverse voices and expertise from across disciplines.

On behalf of Middlesex University, I warmly welcome you all and encourage you to actively engage throughout the workshop, sharing your perspectives, challenging ideas, and collaborating to build a comprehensive understanding of these important topics.

**Janet Jones**
Professor and Pro Vice-Chancellor, Global Research and Knowledge Exchange
Middlesex University, UK

## Welcome from the Dean of the Faculty of Science and Technology

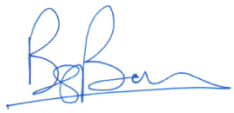
Healthcare has always been a driver for innovation and advances in technology often, leading to wholesale improvements in health outcomes across the demographic divide. The healthcare domain has always been the focus of interest for government, the private sector and universities and captures perfectly the tensions around ethical governance, innovation, creation of new knowledge for humankind, public policy and private profit. Of course, the latest advances in Artificial Intelligence (AI), in particular the use of generative-AI brings this tension to the fore.

The UK is seeking to play a leading role in the uptake of generative AI in the healthcare sector. The AI Opportunities Plan[1] published in January 2025 cites the importance of strategic collection of data to enhance research in areas such as disease and prediction of health outcomes through the UK Biobank. The Industrial Strategy paper[2] notes the plan for a National Data Library which will be the source of innovation for private enterprise. Meanwhile, the Centre for Data Ethics and Innovation, a government unit launched in 2018, became the Responsible Technology Adoption Unit in February 2024 and then in January 2025 its name was quietly modified and its functions embedded across UK.Gov.

Into this space, then, Universities play a central role in facilitating platforms for honest discussion and exchange: to ensure inter-disciplinary and transdisciplinary needs are met and to allow for insightful and meaningful critique of technologies and their impact on public policy. The focus of this workshop on *legal, ethical, social and governance aspects of emerging technologies in healthcare* is of critical importance, since effective human oversight is essential for the successful implementation, use, acceptance and further development of these technologies.

The Faculty of Science and Technology, Middlesex University, is pleased to host this event at our London campus. Our Faculty mission *of empowering our graduates to become 'change-makers' who deploy their scientific knowledge for social good* resonates with the themes and presentations from expert speakers. Some of the research in our faculty, especially in our ALERT (Aspects of Law and Ethics Relating to Technology) research group, which focuses on interdisciplinary work exploring legal/ethical issues raised by emerging technologies, is also an excellent reason for why we are delighted to host this workshop.

I welcome you to Middlesex University and hope that you leave after having shared an enjoyable, stimulating and challenging workshop in London at one of the UK's most international universities.



**Balbir S, Barn**
Professor of Software Engineering
Dean, Faculty of Science and Technology
Middlesex University, UK.

[1] https://www.gov.uk/government/publications/ai-opportunities-action-plan/ai-opportunities-action-plan
[2] https://www.gov.uk/government/consultations/invest-2035-the-uks-modern-industrial-strategy/invest-2035-the-uks-modern-industrial-strategy

# Workshop Introduction

## 1. Overview

Emerging technologies (e.g., artificial intelligence (AI), blockchain, mobile apps, social media, social robots, virtual reality) continue to play an integral role in the provision of healthcare spanning fields such as digital health, mHealth, telemedicine, telecare and telehealth. International institutions, like the European Commission and World Health Organization, have steadfastly promoted the use of technologies to transform the healthcare sector and to facilitate cross-border services. Healthcare technologies are useful in providing direct access to healthcare and in facilitating and supporting healthcare-related functions. Examples include medical data storage (electronic health or medical records, cloud computing), data sharing (locally, nationally, and across borders), data analytics, and access to healthcare services. Policy makers and businesses continue to seek ways to use healthcare technologies to provide benefits to stakeholders including patients, healthcare professionals, clinics and hospitals, governments, research initiatives, and business ventures. These and other developments in healthcare technologies bring challenges that raise questions spanning legal, ethical, social, and governance concerns, particularly for the scale-up and implementation of health IT. This workshop primarily focuses on legal, ethical, social and governance aspects of new and emerging technologies in healthcare as well as developments regarding regulatory and ethical frameworks.

## 2. History

The current 2025 workshop is the seventh in a series of workshops led by Dr Carlisle George, Convenor of the ALERT (Aspects of Law and Ethics Related to Technology) Research Group at Middlesex University with co-organisers: Ms Diane Whitehouse (*European Health Telematics Association*); Dr Penny Duquenoy (*Middlesex University*); and Prof Kenneth Goodman (*University of Miami*). The previous six workshops were held in 2011, 2013, 2014, 2016, 2018, and 2019. These workshops have provided a unique forum for anyone interested in legal, ethical, social and governance aspects of Health IT. They have facilitated networking and collaborations, many publications, attracted doctoral students to the ALERT Research Group, and have also led to the external recognition and engagement of the expertise of ALERT Group members. To date, the seven workshops have collectively attracted participants (including students, academics, researchers, medical professionals, lawyers, ethicists, and policy makers) from over 60 institutions/organisations from around the world as shown in the list given on the following page.

All workshops have been free for participants to attend, and have been funded jointly by Middlesex University and the University of Miami Miller School of Medicine. They are known affectionately as MDX Health IT Workshops.

## 3. List of Institutions represented at MDX Health IT workshops (2011-2025)

1. ANEC (European consumer voice in standardisation), Brussels.
2. American University in Cairo, Egypt.
3. Aristotelean University of Thessaloniki, Greece.
4. Beecham Research, London, UK.
5. Brunel University, UK.
6. Cambridge Health Informatics Limited, UK.
7. Cairo University, UK.
8. Charles Sturt University, Australia.
9. Coventry University, UK.
10. De Montfort University, UK.
11. East Tennessee State University, USA.
12. East & North Hertfordshire, NHS Trust, UK.
13. Eindhoven University of Technology, Netherlands.
14. European Health Telematics Association, Belgium, Brussels.
15. European Commission, Directorate-General Research & Innovation, Brussels.
16. European Commission, Directorate-General Information Society, ICT & Health Unit, Brussels.
17. EXUS Ltd, London, UK.
18. Fraunhofer Institute for Biomedical Engineering, Germany.
19. Galala University, Egypt.
20. Glyndŵr University, UK.
21. Guy's and St. Thomas' NHS Foundation Trust, London, UK.
22. Health & Social Care Information Centre, UK.
23. Health on the Net (HON) Foundation, Geneva, Switzerland.
24. IESE Business School, Spain.
25. Independent Centre for Privacy Protection Schleswig Holstein, Germany.
26. Information Catalyst, Manchester, UK.
27. International Electrotechnical Commission's (IEC) Standardization Evaluation Group
28. International Baccalaureate, UK.
29. Johannes-Gutenberg University, Germany.
30. Katten Muchin Rosenman LLP, Texas, USA.
31. King's College London, UK
32. Lister Hospital Stevenage, UK
33. Leeds University, UK.
34. London School of Economics & Political Science, UK.
35. Middlesex University, UK.
36. Netherlands Organisation for Applied Scientific Research, Netherlands.
37. Nelson Mullings Law Firm, USA
38. NHS Digital England, UK.
39. Nottingham University, UK.
40. NorthShore University Health System, USA.
41. Nuffield Council on Bioethics, UK.
42. Padua University, Italy.
43. Rostock University, Germany.
44. Royal Free NHS Trust, London, UK.
45. Swansea University, Wales, UK.
46. The Castlegate Consultancy, UK.
47. The European Institute for Innovation for Health Data, Ghent, Belgium.
48. Tilburg Institute of Law, Netherlands.
49. University of Bedfordshire, UK.
50. University College London, UK.
51. University of Delhi, Delhi, India.
52. University of Hull, UK.
53. University of Leuven, Belgium.
54. University of Miami, Miller School of Medicine, USA.
55. University of Oxford, UK.
56. University of Turku, Finland.
57. University of Wales Trinity Saint David, UK.
58. University of Westminster, UK.
59. University of Winchester, UK.
60. University of Piemonte Orientate Alessandria Area, Italy.
61. Uppsala University, Sweden.
62. UK Council for Health Informatics Professions, UK.

## Organisation and Sponsorship

### Workshop Organising Committee

- **Dr Carlisle George**: Associate Professor & Barrister, Convenor ALERT Research Group, Middlesex University, UK. (https://www.eis.mdx.ac.uk/research/groups/Alert/).

- **Ms Diane Whitehouse**: Principal eHealth Policy Consultant, EHTEL, Brussels, Belgium.

- **Dr Penny Duquenoy**: Visiting Researcher, Middlesex University, UK.

- **Prof Kenneth Goodman**: Emeritus Professor of Medicine, Miller School of Medicine, University of Miami, USA.

### Workshop Sponsors

- **Faculty of Science and Technology,** Middlesex University, UK.
  http://www.mdx.ac.uk/about-us/our-faculties/faculty-of-science-and-technology

- **Institute for Bioethics and Health Policy,** Miller School of Medicine**,** University of Miami, USA.
  https://bioethics.miami.edu

**Workshop Programme Town Hall - Committee Room 1**

**Day 1 – Thursday, 26th June 2025**

| TIME | SESSION |
|---|---|
| 12.30–12.45 | **Registration, mix-and-mingle** |
| 12.45–13.00 | **Welcome, Middlesex University**<br>*Dr Carlisle George* - Assoc Professor & Barrister, Middlesex University, UK<br>*Dr Neelam Raina* - Director of Research, Middlesex University, UK |
| | ***Artificial Intelligence in Healthcare – Part 1*** |
| 13.00–13.40 | **What's Wrong with Computational Intelligence in Healthcare? (Keynote)**<br>*Professor Philip Scott* - Professor of Digital Health & Care, University of Wales Trinity Saint David,UK |
| 13.40–14.00 | **The ethics of AI chatbot information: Why botfo is a new and different language and why it matters**<br>*Mr Chris Zielinski* - President, World Association of Medical Editors (WAME). Centre for Global Health, University of Winchester, UK |
| 14.00 –14.40 | **Unbiased, Transparent and Reliable Artificial Intelligence in Healthcare: What Could Possibly Go Wrong? (keynote)**<br>*Professor Kenneth Goodman* - Emeritus Professor of Medicine, Miller School of Medicine, University of Miami, USA |
| 14.40–15.00 | **Coffee Break (20 mins)** |
| | ***Ensuring Integrity and Accountability of Health IT*** |
| 15.00–15.20 | **Accountability as a Primary Lens Through Which to Evaluate Healthcare Information Technology: Examples from the United States**<br>*Dr Paul DeMuro* - Counsel (Healthcare), Katten Muchin Rosenman LLP, Dallas, Texas, USA |
| 15.20–15.40 | **Emerging Digital Health Training and Skills Capacity-Building: What are the Workforce Needs?**<br>*Ms Diane Whitehouse* - Principal eHealth Policy Consultant, European Health Telematics Association |
| 15.40–16.30 | **Discussion Session –** ***The increasing use of technology in healthcare: Important benefits, concerns and challenges. (40 mins)*** |
| 16.30–16.45 | **Roundup of the day 1.** |

**Day 2 – Friday, 27th June 2025**

| TIME | SESSION |
|---|---|
| 09.30–10.00 | **Welcome and introduction to the day**<br>*Dr Carlisle George* - Assoc Professor & Barrister, Middlesex University, UK<br>*Prof Sally Priest* - Deputy Dean for Research & Knowledge Exchange, S&T, Middlesex University, UK |
| | ***Artificial Intelligence in Healthcare – Part 2*** |
| 10.00 –10.40 | **AI in health IT: Social and Ethical Considerations. (Keynote)**<br>*Professor Neil Gordon* - Professor in Computer Science, University of Hull, UK |
| 10.40–11.00 | **Medical AI: Are We Playing Russian Roulette with Patients' Lives?**<br>*Dr Uchenna Nnawuchi* - Researcher, ALERT Research Group, Middlesex University, London UK |
| 11.00–11.20 | **Decoding Memory States using Machine Learning: Ethical, Legal and Governance Implications**<br>*Mr Sailendra Ramana* - Solution Architect, Research & Innovation Ops, University College London,UK |
| 11.20–11.40 | **Artificial Intelligence in Biobanking: Potentials, Challenges and Future Trends in Oncology**<br>*Dr Radwa Hassan* - Director of Medical Informatics Unit, Galala University, Egypt<br>*Prof. Galal Hassan Galal-Edeen* - Professor of Information Systems Engineering, American University in Cairo, Egypt |
| 11.40–12.00 | **Coffee Break (30 mins)** |
| | ***Application of IT to Healthcare Related Challenges*** |
| 12.00–12.20 | **Using Surveillance Technologies to Prevent Suicides in Public Spaces: Ethical, Legal, Acceptable?**<br>*Dr Jay Mackenzie* - Reader in Psychology, University of Westminster, UK<br>*Prof Lisa Marzano* - Professor of Psychology, Middlesex University, UK |
| 12.20–12.40 | **Addressing Privacy and Security Challenges in Enabling Interoperability of Electronic Health Records: The TASIPPS Framework**<br>*Mr Adetunji Ademola* - Researcher, ALERT Research Group, Middlesex University, UK |
| 12.40–13.00 | **Revisiting the Lucy Letby Case: A Forensic Readiness Perspective for Healthcare Big Data Systems**<br>*Mr Cephas Mpungu* - ALERT Research Group, Middlesex University, UK |
| 13.00–14.00 | **LUNCH BREAK (1 hour)** |
| | ***Smartphone Apps & mHealth*** |
| 14.00–14.20 | **Personalised Care & Digital Applications in Maternity for Ethnic Minorities: Legal, Ethical, and Social Challenges**<br>*Dr Jasmine Leonce*- National Maternity Specialty Advisor NHSE. Consultant Obstetrician, East & North Hertfordshire NHS Trust<br>*Professor Trixie McAree* - Senior Midwife, Chief Midwife's Office NHS |
| 14.20–14.40 | **Non-Medical Device Apps for Good Health and Well-being: Regulatory Challenges**<br>*Dr Carlisle George* - Assoc Professor & Barrister, ALERT Research Group. Middlesex University, UK |
| 14.40–15.00 | **User Context-Centred Design of an mHealth System for a Developing Country**<br>*Prof. Galal Hassan Galal-Edeen* - Professor of Information Systems Engineering, American University in Cairo, Egypt. |
| 15.00–15.45 | **Discussion Session**<br>***Governance of health IT: Adequacy of current policies, regulatory (ethical/legal) frameworks and oversight bodies.*** |
| 15.45–16.00 | **Workshop Overview and Next Steps (Workshop Organisers)**<br>**Closing Remarks**<br>*Dr Aboubaker Lasebae* - Interim Head of Computer Science Department, Middlesex University, UK |

# What's Wrong With Computational Intelligence in Healthcare?

*Professor Philip Scott*
Professor of Digital Health & Care, University of Wales Trinity Saint David, UK
philip.scott@uwtsd.ac.uk

## 1. Introduction

The technology industry is once again pushing 'artificial intelligence' (AI) as a magical panacea, especially since the advent of generative AI that is purported to match human performance in language, art, and even software development. Big tech firms apparently regard their need for massive additional energy supplies to support the computing power for AI model creation as a 'natural right'. This corporate technology-driven hype has influenced government policy, as AI is seen as a driver of economic growth, and as a means of providing the public's expectations for improved services.

Healthcare is often cited as an 'industry' that will be transformed by AI. What is the reality? It is certainly true that healthcare is in great need of radical improvement in its use of data, information, and knowledge. Globally, it is estimated that, on average, only 60% of care is in line with evidence- or consensus-based guidance, 30% is wasteful or of low value (for example, it involves overdiagnosis or overtreatment), and 10% is harmful. This is in the context of a worldwide clinical workforce crisis, taking place especially in developing countries, and the broader quality improvement agenda summed up in the triple aim of decreasing per capita costs, improving the patient experience of care, and improving the health of populations.

This talk explores the meaning of intelligence from the human, organisational and computational perspectives. The presentation emphasises the criticality of knowledge as a concept distinct from data or information, and its importance in the paradigm of Learning Health Systems as a continuous cycle of quality improvement. We analyse the spectrum of computable knowledge in healthcare, from traditional clinical decision support using evidence-based guidelines through to the currently proliferating AI models for specific use cases. This focus turns our attention away from technology push to problem-based solutions where computable knowledge, whether expert-based or AI-powered, can genuinely contribute to transformational change.

## 2. Mobilising Computable Knowledge

In 2018, a multi-disciplinary movement known as Mobilising Computable Biomedical Knowledge (MCBK) was formed in the United States of America (USA), hosted by the University of Michigan. It soon spread to the United Kingdom (UK), hosted by the British Computer Society (BCS). The term Computable Biomedical Knowledge (CBK) is defined in the MCBK Manifesto as "the result of an analytic and/or deliberative process about human health, or affecting human health, that is explicit, and therefore can be represented and reasoned upon using logic, formal standards, and mathematical approaches."[1] One of principal aims of MCBK is the evolution of an open computable biomedical knowledge ecosystem dedicated to achieving the FAIR (Findable, Accessible, Interoperable, and Re-usable) principles: making CBK easily findable, universally accessible, highly interoperable, and readily reusable. MCBK has established chapters in the USA and the UK: chapters are also being formed in continental Europe and Australasia.

The presentation illustrates the practical application of MCBK activity by describing recent work with the National Institute for Health and Care Excellence (NICE) in the UK. NICE produces various kinds of clinical guidance such as evidence-based treatment and diagnostics guidelines and technology appraisals. As health policy is devolved in the UK to the four home nations (England, Scotland, Wales, and Northern Ireland), NICE guidance formally only applies in England. It does, however, have an international reputation, and Wales has

[1] MCBK Manifesto (2018).
https://mobilizecbk.med.umich.edu/about/manifesto

voluntarily chosen to adopt NICE guidance. NICE is already working on what it calls 'real-world evidence' (that is to say data from operational healthcare systems not just from randomised controlled trials). The NICE Five Year Strategy (2021-26)[2] included the goal to make "a shift from 'words' to 'data' and to a more modular approach to recommendations" and "guideline recommendations produced in an interactive, digitalised format". To support these aims, NICE established a Content Advisory Board that included the co-chair of MCBK-UK.

Hence, the talk discusses the progress made in the NICE Computable Implementation Guidance (NCIG) project, a first attempt to map the way forward from traditional narrative guidance to structures that can become computable and therefore be more rapidly implemented. NCIG contributed to NICE's recent decision to develop a new useable product strategy based on structured recommendations. Further work is now under discussion with NICE, with the purpose of improving the consistency of implementation of NICE products, by developing a standards-based approach for structured clinical recommendations in such a way that industry will readily and reliably adopt it.

**3. Conclusion**

The talk raises some broader ethical questions arising from AI developments, including new forms of research waste in healthcare AI and its environmental, behavioural, and philosophical consequences. We make a call to action to maintain a healthy scepticism of AI in healthcare and to drive the development of Learning Health Systems.

[2]NICE Strategy 2021 to 2026
https://www.nice.org.uk/Media/Default/Get-involved/Meetings-In-Public/Public-board-meetings/Mar-24-pbm-NICE-strategy-2021-2026.pdf

# AI Chatbot information:
# Why Botfo is a New and Different Language and Why it Matters

*Mr Chris Zielinski*
Visiting Fellow, University of Winchester, and President, World Association of Medical Editors
chris@chriszielinski.com

**Abstract:**
In this paper, the characteristics of chatbot information – known as "botfo" – are examined, particularly its lack of intentionality and desire to please, and the consequences stemming from this. Three types of bias typically found in botfo are reviewed, and the urgent need for reliable software to detect botfo is stressed.

## 1. Chatbots are not authors

In 2022, the World Association of Medical Editors (WAME) – a voluntary nonprofit association with free membership and global scope) found itself being asked questions relating to the use of generative artificial intelligence (AI) and chatbots to write papers on health topics published in academic journals. At that time, several papers had already been published in which one of the authors was listed as "ChatGPT".

As a result, in January 2023, WAME issued its widely quoted "Recommendations on Chatbots and Generative Artificial Intelligence in Relation to Scholarly Publications", revising and expanding them four months later in May [1]. Other bodies dealing with academic journals, including the Committee on Publication Ethics (COPE) [2] and the International Committee of Medical Journal Editors (ICMJE) [3] also published responses to questions on similar uses of chatbots.

While recognising the importance and potential value of generative AI in scientific communications, the central assertion in the WAME recommendations was that generative AI and chatbots could not be considered or listed as authors. The consensus was that, unlike human authors, chatbots "are unaccountable, and can't think, judge or be jailed" [4].

Moreover, because of the risk of various kinds of bias in AI-generated text, and the distinctive factual errors termed "hallucinations" (which include the invention of bogus reference sources and other facts), WAME recommended that the use of chatbots in preparing a given text should always be identified and declared.

The perspective adopted in the WAME recommendations mirrors all the others issued throughout the world of academic journal publishing by COPE, ICJME, as well as by leading individual journals, particularly in the health sector. The uniformity of outlook in the approaches and the widespread acceptance of and lack of disagreement about the recommendations are striking.

Thus, according to the various published recommendations, chatbots should not be considered as being on an evidentiary, legal, or indeed existential, level with human authors. Taking this argument one step further, it can be asserted, as a point of constructive classification, that the text produced by generative AI and chatbots should not be considered as equivalent to text produced by humans. The consequences of making this distinction are addressed below in the context of the intentionality of written and spoken communication, the development and ethics of chatbot language, and methods for detecting it.

## 2. The botfo language

Chatbots are designed to operate conversationally, and so their output is modelled on human speech. However, this generated output is fundamentally different from human language. To underscore this distinction, the portmanteau term "botfo" is used here as an abbreviation for "chatbot information" [4]. Although botfo often looks and sounds like what is written or said by a human being, there are at least three ways in which it is in fact a unique and distinct form of language:

- **Botfo is statistically generated.** Each word appears as the highest-probability next word to the previous one. The probabilities are based on the way these words have been used in the text(s) that the generative AI has been "trained on" or made to scan. Terabytes of sample text provide the statistical heritage for every word that emerges from a chatbot. While we humans also base much of what we say on what we have read or been told, the impulse to use a particular word may depend on how a grandparent or comedian or lyricist used it. There may even be a statistical basis for the way humans select their words from street slang, fashionable phrases, film dialogue or the punch lines of jokes. But humans are not machines, and, unlike botfo, we don't speak in algorithms.

- **Botfo lacks intentionality.** Botfo is always a response. It is initiated by a prompt, and it is designed to respond to the prompt as accurately and well as possible. Chatbots do not initiate conversations without some stimulus. In human discourse, when someone breaks the silence in speech or in writing, human listeners or readers assess whether it is worth trying to understand and interpret the words. This reaction to a broken silence is probably an evolutionary response developed in humans. Inevitably, humans ask the questions, Why are words being generated? Do they have a significance for me as a person and my world of interests and activities? Can the words be trusted and believed? Are they serious or a joke? Is the expression one of sarcasm, poetry, fantasy, or a statement of fact? Is it meaningless drivel or a life-altering command? Since such options exist in every human utterance made, listeners immediately try to judge the speaker's intention and to respond accordingly. Without such an intention, if the speaker is not intending to communicate anything, there is no reason for a human being to listen or read further. Botfo doesn't operate like that.

- **Botfo needs to satisfy both the prompt and the prompter**. When asked to prepare a text advocating for, or defending, some proposition, a chatbot will do its best to comply, to the extent of making up facts and providing bogus references. This apparently naïve desire to please is built into chatbots, and has been noted and criticised as "sycophancy" [5] but, nevertheless, this is what one would expect from a commercial product that has to satisfy the buyer/user. Chatbots are not designed to pursue facts to the bitter end, especially if the facts are uncomfortable for the user or if they don't provide a desired outcome. Botfo is not designed to look for the truth. Its primary objective is to satisfy the buyer/user. This is not a small failing: there is a clear distinction between the prompt-induced pseudo-intentionality of botfo and the academic approach to knowledge. When one considers the impact of generative artificial intelligence on the health sector, and its growing use in all fields of public health and research, particularly in counselling, mental health, and geriatric care, the risks of blind positivity are obvious. The road to health or good governance is not built on compromises with the truth.

## 3. The ethics of botfo

As the examination of botfo language indicates, while botfo is not the product of sentience and is thus not capable of being intentional, forms of simulated intentionality are nevertheless built into the software. This is the result of three phenomena: training biases: the distortions of reality contained in the materials used to train the Large Language Models (LLMs), residual biases: the conscious and unconscious, deliberate, and accidental biases of the programmers, and compliance programming: the built-in performative task of solving as many problems set by the user as possible (whether they are solvable or not).

- **Training biases** come from the texts the software is trained on. These biases include racism, sexism and every other -ism and bias known to humanity. They arise as a result of the almost desperate struggle to find and absorb as much training content as possible, irrespective of its adherence to ethical principles and norms.
- **Residual biases** stem from the unconscious sociopolitical attitudes of the programmers. These are biases of which the programmers themselves may be unaware. They are revealed when the software is break-tested to see if it can be made to blurt out racial, gender or other unwanted slurs.

- **Compliance programming biases** are the result of the programmers complying with explicit rules established by the AI company. For example, in May 2024, OpenAI made its first tentative attempt at drawing up ethical principles for chatbots [6]. The original lightweight outline was replaced by a much heavier duty model in February 2025 [7]. This evolving model has been derided by philosophers as being simplistic ("an engineer's approach to ethics"), but at least it is explicit. The problem is that, ultimately, ethics questions cannot be answered in a binary fashion: "A user can demand that ChatGPT answer 'yes' or 'no' — we've all met idiots — but it is also fundamentally idiotic for an AI to obey an order to give information it does not and cannot have." [8]. Based on three laudable concepts ("maximizing helpfulness and freedom for our users, … minimizing harm, …[and] choosing sensible defaults") [9], the model offers practical solutions to problems its programmers see as having an ethical basis. However, the main effect appears to be to reinforce chatbot sycophancy – the software's willingness to please and its reluctance to offend.

## 4. Detecting botfo

In human-to-human communications, when we really don't understand what someone means, we are free to say or write, "What do you mean?" With botfo, however, despite any explicit ethical rules programmed into the algorithms, the lack of intentionality in the software itself will always escape ethical regulation.

That is why botfo should always be identifiable as botfo, and not as a human artifact. Someone, not something, has to assume responsibility for every health procedure proposed or carried out, every legal agreement entered into, every student essay submitted, and journal paper published: essentially every public commitment entered into as a person.

When asked about future proposals for LLMs, Sam Altman, one of the leaders at OpenAI, said, "There are some minor short-term things that I think hopefully are very non-contentious. I think all generated content should have to be tagged as generated." [9]. In fact, however, this issue proves not to be minor, short-term, or non-contentious. "OpenAI has a method to reliably detect when someone uses ChatGPT to write an essay or research paper….[but] The project has been mired in internal debate at OpenAI for roughly two years…"

Why? It's a business question. "OpenAI employees have wavered between the startup's stated commitment to transparency and their desire to attract and retain users. One survey the company conducted of loyal ChatGPT users found nearly one third would be turned off by the anticheating technology" [10]. This is not just because there are so many cheats out there – or users with little understanding of such issues as copyright and plagiarism. It is possible that a chatbot use detection tool could also disproportionately affect groups such as non-native English speakers.

Nevertheless, "In April 2023, OpenAI commissioned a survey that showed people worldwide supported the idea of an AI detection tool by a margin of four to one." [10].

It should be emphasised that botfo detection tools are needed by everybody – the need for such tools is mentioned in the various public recommendations cited, such as those of WAME, COPE, and ICMJE. In the WAME Recommendations on Chatbots and Generative Artificial Intelligence in Relation to Scholarly Publications [1], WAME Recommendation 5 states, "Editors need appropriate tools to help them detect content generated or altered by AI. Such tools should be made available to editors regardless of ability to pay for them, for the good of science and the public, and to help ensure the integrity of healthcare information and reducing the risk of adverse health outcomes."

## 5. Conclusion

There is a real risk that intention-free, user-pleasing chatbot verbosity will swamp different forms of media at every level and promote doubts about the accuracy of what is being reported. The growing use of botfo for spreading disinformation and political subterfuge is inevitable, thanks to its lack of intentionality and affect.

Ultimately, the risks in failing to recognise botfo and to distinguish it from human communications are bad science, bad health, and a loss of human values.

But this doesn't need to happen. The web of scientific communication, including the world's medical journals and their editors, is based on the fundamental intention to produce and communicate genuine knowledge and research developed through an engagement with facts. This intention can only be achieved by the involvement of committed and critical human researchers (assisted by the judicious use of AI and other tools) who are willing to accept failures and the falsification of theories along with positive results. This is the future we humans need to achieve.

**References**


1. Zielinski C, Winker MA, Aggarwal R, Ferris LE, Heinemann M, Lapeña JF, Pai SA, Ing E, Citrome L, Alam M, Voight M, Habibzadeh F, for the WAME Board. Chatbots, Generative AI, and Scholarly Manuscripts. WAME Recommendations on Chatbots and Generative Artificial Intelligence in Relation to Scholarly Publications. WAME. May 31, 2023. https://wame.org/page3.php?id=106
2. COPE, 2024. Authorship and AI tools. https://publicationethics.org/cope-position-statements/ai-author
3. ICMJE, 2024. Defining the Role of Authors and Contributors. https://www.icmje.org/recommendations/browse/roles-and-responsibilities/defining-the-role-of-authors-and-contributors.html
4. Zielinski C. 2025. Why artificial intelligence is not an author. *European Science Editing* 2025,51: e142904. https://doi.org/10.3897/ese.2025.e142904
5. Gerken T. 2025. OpenAI has pulled a ChatGPT update after users pointed out the chatbot was showering them with praise regardless of what they said. Published 30 April 2025. https://www.bbc.co.uk/news/articles/cn4jnwdvg9qo
6. OpenAI, 2024. Desired behavior for our models in the OpenAI API and ChatGPT, Model Spec: May 08, 2024, https://cdn.openai.com/spec/model-spec-2024-05-08.html#overview
7. OpenAI, 2025. Desired behavior for our models in the OpenAI API and ChatGPT, OpenAI Model Spec: February 12, 2025 https://model-spec.openai.com/2025-02-12.html
8. Lopatto E. 2025. The questions ChatGPT shouldn't answer, *The Verge*, Mar 5, 2025 (https://www.theverge.com/openai/624209/chatgpt-ethics-specs-humanism).
9. Felsenthal E. 2023. Read TIME's Interview With OpenAI CEO Sam Altman. *TIME Magazine Leadership Brief* 27 June 2023 https://time.com/6288584/openai-sam-altman-full-interview/
10. Seetharaman D and Barnum M. 2024. There's a Tool to Catch Students Cheating With ChatGPT. OpenAI Hasn't Released It. *The Wall Street Journal* Aug. 4, 2024. https://www.wsj.com/tech/ai/openai-tool-chatgpt-cheating-writing-135b755a

# Unbiased, Transparent and Reliable Artificial Intelligence in Healthcare: What Could Possibly Go Wrong?

*Professor Kenneth W. Goodman*
Emeritus Professor of Medicine, Miller School of Medicine, University of Miami, USA
KGoodman@med.miami.edu

In an extraordinarily short time, many if not most educated laypeople have come to be able to itemize several "concerns" and "issues" raised by Artificial Intelligence (AI). For instance,

AI is biased
It hallucinates or confabulates
Privacy and confidentiality are imperilled
It is not transparent
It is under-tested or inadequately evaluated

These are good concerns to have, and they most certainly should be addressed. But this emphasis – on what has evolved into a rough and ready "pick list" – is missing something much larger: As we correct and improve AI in healthcare, we risk losing (or losing track of) a more fundamental challenge: what it means for people to have practical knowledge about health and medicine.

Consider an AI system that supports clinicians in offering diagnoses and prognoses. Imagine it to be blameless in that it is unbiased, does not hallucinate, and protects privacy. The system is reliable and consistently more accurate than humans. Because it then becomes irrational not to use such a system, the clinicians as users – "humans in the loop," if we like – then use it and, over time, accept it, come to rely on it, and employ it regularly to generate diagnoses and prognoses. These findings are then placed in patients' electronic health records, where they become data for the next generation of AI systems.

But this incremental storage is a problem: it constitutes the steady and increasingly rapid replacement of the record or of human discovery with the record of machine discovery. Indeed, the transition from human to machine composition of the scientific corpus has already begun, at least in some cases in electronic health records. Will we press on gladly? Or ought we to anticipate what will surely be a significant epistemic challenge, especially in the biosciences?

This reliance has been described as "harmful epistemic dependence" which "occurs when the impossibility of critically assessing the soundness of a system's output *in situ* implies a moral obligation to comply with its recommendation since a failure to do so constitutes a moral risk that cannot be justified then and there."[1]. Put differently, even a human in the loop becomes inadequate to the task. This circumstance has epistemic and practical implications.

Epistemically, a world in which knowledge is produced by machines and no longer by humans is a world in which humans have abandoned standard approaches to explanation, methodology, causation, inference and even truth … the foundations of epistemology. Though medics and their patients might care little about this, it should nevertheless be seen as an abandonment of millennia of effort to understand the world.

Practically, the reliance on AI tools in health and medicine will entail what in other contexts has been a source of some concern, that is, skill degradation. If a new technology renders obsolete what was once an essential human skill, we should weigh up whether that obsolescence matters or not. If it does matter, then we should take steps to prevent it. In an AI world, this prevention would entail a great deal of education which will consist of no little computer science; one may hypothesize that this new duty will not be gladly met. If it does not matter, then the role of human clinicians is fundamentally undermined.

This presentation argues that the role of human clinicians does matter, both epistemically and practically. The history of medical technology is in some respects the discovery and application of tools to improve seeing, hearing and touching. We are all better off for that. But AI is a tool to improve thinking and, in the context of the "blameless" systems imagined here, even replace thinking. If a human cannot confidently explain a diagnosis or prognosis or an action or intervention, then we will have surrendered a skill that is not only precious but which has for a very long time been an essential requirement for success in an ancient profession.

We already encounter patient-facing AI tools, with an inadequate understanding of how they shape behavior and influence efforts to access high quality health care. This situation is a challenge for health equity (no matter the protestations of the current malign and fatuous U.S. political regime).

None of this is to suggest we ought not continue to develop and use AI tools in health and medicine. Rather, it is to insist that we acquire far greater familiarity with these tools and, in consequence, be better equipped to identify appropriate uses and users of the tools. This is in fact the job of bioethics, which has, for centuries, evolved in parallel with the development of technologies to help us prevent and cure disease in individuals and in populations.

**References**

1. Pozzi G, Buijsman S, van den Hoven J. Harmful epistemic dependence on medical machine learning and its moral implications. J Med Ethics. 2025 Feb 20:jme-2024-110552. doi: 10.1136/jme-2024-110552.

# Accountability as a Primary Lens Through Which to Evaluate Healthcare Information Technology: Examples from the United States

*[1]Dr Brian R. Jackson, MD, MS and [2]Dr Paul R. DeMuro, JD, MBA, PhD*

[1]University of Utah School of Medicine, Utah, USA
brian.jackson@aruplab.com

[2]Epstein Becker Green, Texas, USA
PDeMuro@ebglaw.com

**Abstract**
All large complex systems require mechanisms to ensure that individuals, departments and organisations stay accountable to their stakeholders. Many of these mechanisms have over time been automated using information technology (IT), with variable effects on the nature of accountability. This position paper uses examples from the U.S. healthcare industry to illustrate how healthcare IT can modulate the accountability of healthcare providers and organisations toward the patients and communities they serve. It also proposes some ways in which that accountability can be enhanced through more thoughtful application and management of IT.

**Keywords**: Accountability, Information Technology, Sociotechnical

## 1. Introduction

Healthcare information technology (HIT) has long been promoted as a tool to achieve larger health policy goals. In the United States (U.S), for example, the 2009 Health Information Technology for Economic and Clinical Health (HITECH) Act promoted expanded use of electronic health records with a stated goal of improving the quality, safety, and efficiency of clinical care [1]. More recently, policymakers have expressed optimism about the potential for artificial intelligence (AI) within HIT to advance these same goals [2]. We believe, however, that a key policy goal has been largely missing from these discussions, namely HIT's potential to either increase or reduce the accountability of healthcare organizations toward their primary stakeholders.

## 2. IT's impact on accountability

Accountability refers to the procedures and processes through which individuals and/or organisations are held responsible for their actions and inactions by other interested parties [3]. Accountability can be enforced and reinforced through several modes, including social feedback, economic feedback, government regulation, and system architecture (including IT) [4]. Accountability by itself does not guarantee either high quality or improvement in quality, but by identifying problems and the responsible actors, it facilitates social and managerial feedback loops which can address those problems. In the absence of accountability, problems can simply be ignored or swept under the rug. Accountability has been recognized as a foundational principle in sociotechnical systems outside of healthcare [5]. As IT systems are increasingly used to automate functions that were historically performed by human beings, it is critical to consider the impact of these changes on accountability.

One interesting case study involves a lawsuit settled in 2009 between the New York Medical Society and UnitedHealth, a large for-profit medical insurer [6]. At the time, many insurance company contracts calculated physician reimbursement based on "usual and customary" fees. This calculation relied on a national database of physician rates maintained by Ingenix, a health analytics company. After UnitedHealth purchased Ingenix, there were allegations that UnitedHealth manipulated the Ingenix algorithm to exclude certain categories of physician fees, thus biasing the calculated values for its own benefit. As part of a court settlement, UnitedHealth agreed to stop using the Ingenix physician fee database for this purpose, and to finance the development of a new, independent fee database [6]. In theory, Ingenix's database, by aggregating

physician claims data across all major insurers and across the U.S., could have been implemented in a way that increased both public and physician visibility into physician charges, thus facilitating an increase in accountability versus previous, more fragmented approaches. Instead, though, it appears to have been used in a way that tilted the power dynamic toward insurers and thus reduced physicians' ability to hold insurers accountable for paying them fairly.

Artificial intelligence (AI) is an emerging IT domain deserving of special mention. Much has been written about the accountability of healthcare AI itself [7] but less about the impact of AI algorithms on the accountability of the larger systems in which they are implemented. For example, in 2023, a class action lawsuit was filed accusing UnitedHealth and its subsidiary NaviHealth of using an AI algorithm rather than human medical reviewers to assess the medical appropriateness of nursing home care [8]. This backlash has led to legislative proposals to restrict AI use in insurance coverage decisions, such as California's Physicians Make Decisions Act [9]. There is no obvious logic, however, to conclude that individual human reviewers would necessarily be more fair or accurate than automated reviewers might be. It is at least theoretically possible that within a robust framework of third-party audits and a straightforward appeals process, AI could be used to make coverage decisions that could provide more fairness and transparency than human reviewers.

There are many other ways in which AI systems could be implemented to improve accountability across healthcare systems. For example, AI could be used to analyse large data sets to identify areas of potential fraud, waste, or poor quality, so that these areas could then be subject to management intervention. On the other hand, the black-box nature of many AI systems creates a new set of risks. To the extent that AI is used to automate functions related to accountability, it will be important to set up separate, independent auditing and quality management systems to ensure that the AI is behaving as expected, not just from the perspective of the organization controlling the AI, but more importantly from the perspective of the public.

Table 1 illustrates some of the ways in which software might either improve or reduce accountability depending on how it is used.

**Table 1**: The Accountability Impact of IT on Healthcare Functions

| Function | Accountability dynamic | Improving accountability | Reducing accountability |
|---|---|---|---|
| **Pricing** | Consumers, the public, and purchasers such as employers holding hospitals and other providers accountable for fair pricing | As in retail shopping, IT could be used to track and display healthcare prices so that patients and purchasers of healthcare services could compare prices across providers | Software, including the category of Revenue Cycle Management (RCM), is widely used to maximise the amounts that can be charged under different insurance plans. Because this takes place after the clinical encounter, however, organisations are unable to tell patients what they might need to pay for a procedure |
| **Quality measurement and communication** | Consumers, the public, and purchasers such as employers holding hospitals and other providers accountable for high quality | Apps can allow consumers and purchasers to more easily compare key quality attributes across healthcare providers | Scoring systems can create the appearance of quality measurement even when their design and primary purpose are for marketing purposes |

| Function | Accountability dynamic | Improving accountability | Reducing accountability |
|---|---|---|---|
| **Appointment scheduling** | Consumers holding hospitals and other providers accountable for availability of services | In theory, online applications could make it easy for patients to schedule and re-schedule appointments themselves | Hospitals and health systems often centralise appointment scheduling, and then use software to manage the work of large bureaucratic scheduling departments. Because these departments are organisationally separated from the clinics they work with, both patients and clinicians are often frustrated with the lack of transparency and responsiveness to needs but have little recourse to other options |
| **Provider messaging** | Consumers holding physicians accountable for the quality of their communication | Online apps can make it easy for patients to message their providers and receive responses, thereby improving transparency and accountability between provider and patients | Overworked providers are increasingly turning to AI tools such as ChatGPT to respond to patient messages. While in theory providers are expected to review and approve all such messages, this creates new risks of communication failures |
| **Utilisation review** | Payers holding physicians and hospitals accountable for providing only medically appropriate services | Payers can use software to screen for utilisation patterns suggestive of fraud and abuse, and then investigate further | Payers in the U.S.A. have been known to use nontransparent proprietary algorithms to make coverage decisions. This bypasses human medical reviewers, thus insulating decisions from external scrutiny |
| **Other examples** | | To be discussed at workshop | To be discussed at workshop |

## 3. Summary

Accountability of healthcare providers and organisations toward their stakeholders is a key function which has been underappreciated in the health IT literature. IT systems have immense potential to either improve or degrade accountability, particularly when automation results in the partial or total replacement of processes involving human accountability. Emerging AI capabilities have the potential to magnify these effects. Careful attention is therefore required, including both auditing and quality management.

## References


1. Blumenthal, D. Launching HITECH. *New England Journal of Medicine*. 2010:362 (5): 382–385.
2. Matheny ME, Goldsack JC, Saria S, et al. Artificial intelligence in health and health care: Priorities for action. *Health Affairs*. 2025:44(2):163-170
3. Emanuel EJ, Emanuel LL. What is accountability in health care? *Ann Intern Med*. 1996: 124(2):229-239.
4. Lessig L. The New Chicago School. 27 *J Legal Stud*. 1998:27(S2: 661-691.
5. Chopra AK,Singh MP. Accountability as a foundation for requirements in sociotechnical systems. *IEEE Internet Computing*. 2021: 25(6):33-41.

6. Graybow M. UnitedHealth settles New York reimbursement probe. Reuters, 13 Jan 2009. https://www.reuters.com/article/markets/stocks/unitedhealth-settles-new-york-reimbursement-probe-idUSTRE50C5V8/. Accessed 21 April 2025.
7. Singhal A, Neveditsin N, Tanveer H, Mago V. Toward fairness, accountability, transparency, and ethics in AI for social media and healthcare: scoping review. *J Med Internet Research*. 2024;12:e50048.
8. Herman R. Judge rules lawsuit over UnitedHealth's AI care denials can move forward. StatNews. 13 Feb 2025. https://www.healthcarefinancenews.com/news/class-action-lawsuit-against-unitedhealths-ai-claim-denials-advances. Accessed 21 April 2025.
9. Metnick CV, Mitchel L, Sutton MD. Califiornia limits health plan use of AI in utilization management. Healthcare Law Blog. 7 Nov 2024. https://natlawreview.com/article/california-limits-health-plan-use-ai-utilization-management Accessed 21 April 2025.

# Emerging Digital Health Training and Skills Capacity-Building: What are the Workforce Needs?

*Ms Diane Whitehouse*
Principal eHealth Policy Consultant, EHTEL, Brussels, Belgium
diane.whitehouse@ehtel.eu

**Abstract**
Digital health is a core area of development in Europe. This focus will continue for at least the next 5-7 years but, as technologies carry on changing – often at even greater speeds – probably for far longer. This position paper outlines several critical issues that have come to the fore when discussions are held in the European Union about training and skills capacity-building in digital health. It focuses on the needs of two aspects of European workforces – health and care. It illustrates its work by using several examples of developments taking place in European Member States and/or regions. Concrete implementation of these developments is vital. More than simply research, the scaling-up and implementation of digital health training is critical, whether for health care professionals and health care providers or, alternatively, end-users such as citizens and patients.

## 1 Introduction

Europe is facing many current challenges, some very serious indeed and many existential in character. Among these are concerns which are legal and regulatory, ethical, or social in nature, and also related to governance. This position paper has as its orientation, workforce, educational, and skills needs in Europe, with a specific focus on digital health. First, the broad context facing health and care workforces in Europe is outlined. Second, the specific work of the BeWell project is highlighted, including several interesting case studies. What needs to happen next in order to scale up a focus on health and care workforce upskilling and reskilling receives specific focus. The paper ends by offering some insights into what can happen next vis-à-vis concrete implementation of digital skills development and expansion.

## 2 Policy context and concrete initiatives

There are at least two important policy threads that influence the field of digital health in Europe: the Union of Skills and the European Health Data Space (EHDS). Here, emphasis is placed on the Union of Skills.

Europe's Pact for Skills was ***the first of several pacts that were set up*** under Europe's skills agenda that began at the start of this decade. Launched in 2020, the Pact for Skills encouraged ***people and organisations*** to contribute to the further upskilling and reskilling of members of the European health workforce. Among the many types of organisations and people encouraged to become involved were: **companies**; employees ("workers"); ***national, regional and local authorities***; social partners; cross-industry and sectoral organisations; ***education and training providers***; chambers of commerce; and other forms of employment services.

Today, this aspiration for upskilling and reskilling is embedded in the ***European Union (EU) of Skills –*** a Europe-wide endeavour which is investing in people for a competitive union. The Skills Union is currently based around a wide-ranging European Commission-published Communication [1]. The document highlights problems such as the fact that one in five adults struggle with reading and writing; one in four 15-year olds fall short on skills like reading, maths, science; and four out of five small- and medium-sized enterprises cannot find the human resources ("talent") that they need. The highly comprehensive document draws attention to e.g., lifelong learning, micro-credentials, pledges made to upskill/reskill, the setting up of various academies, and the conduct of ongoing pilot actions. In the health sector, there are several professions lacking in numbers: generalist and specialist medical practitioners, nursing professionals and nursing associates, and health care assistants. On digitally-related skills especially, the Communication highlights the importance of fields such as artificial intelligence (AI), robotics, and cybersecurity.

Another supportive initiative is the EU Talent Pool [2], a European Commission proposal for a future regulation which encourages the European Member States to contribute voluntarily to an information technology (IT) platform. The platform's aim is to encourage appropriate jobseeker recruitment from third countries outside of the EU.

The World Health Organization continues to examine digital health needs. In terms of the health and care workforce, the current focus of one of its working groups is on workforce burnout and the mental health challenges faced by health and care employees [3].

## 3 The BeWell project and its activities

The BeWell project is supporting the EU by creating the foundations of many aspects of the critical journey towards strengthening of workforce training reskilling and upskilling. BeWell has two chief orientations: one is on digital skills; another is on green skills. Given the orientation of this position paper, here the focus is on the digital needs of the European health and care workforce.

Among BeWell's current five main activities are:

- A ***Policy brief*** on digital skills. Published by the European Observatory on Health Systems and Policies of the World Health Organization European Region, the policy brief was launched at a ***6 May 2025 event*** [4] ***on closing the digital skills gap*.**
- A **Skills monitor .** ***The monitor is a collection or repository*** of digital skills courses and programmes for health and care workforces. At present, more than 80 courses/programmes have been uploaded to the skills monitor site [5]. More examples will be added over the next 12 months until June 2026. Volunteers are encouraged to add more information to the repository about available courses/programmes in their own regions or countries through a short form which is available to them.
- A ***future skills strategy***. BeWell is running an ***Open Public Consultation*** on what developments should next be developed in terms of digital skills in Europe.
- Testing and piloting of an entirely ***new set of digital skills courses and programmes*** available in a range of European languages. The BeWell ***Pilot Training Platform*** contains 12 courses/programmes on both currently needed digital skills and eight emerging/future skills that are mostly digital in orientation. All the courses/programmes are available in English, and some can be followed in six further languages (Bulgarian, German, Greek, Italian, Romanian, and Norwegian).
- The creation and expansion of a (BeWell) ***Partnership***. The Partnership concentrates on health ecosystems. Its aim is to keep expanding its membership, and to share useful experiences and developments among its members. This partnership liaises with two other European partnerships with which it has synergies – one that has a focus on the health industry and another orientated towards long-term care.

## 4 Three interesting national and regional examples

Members of the health and care workforce in a variety of occupations need to have digital skills that fit with contemporary and future-oriented technologies. Both country-wide skills programme for digital skills and regional level programmes exist. Since there is likely to be no single best way for Member States and regions to develop their health and care workforce skills training, it is always worthwhile to examine the examples of others, including – ultimately – countries from outside Europe. Other countries' achievements can provide insights and directions that others may find useful to explore and, it is possible, eventually implement. Three examples follow. They draw on descriptions shared with EHTEL and with the BeWell project as a result of several EHTEL-related initiatives and events during 2023-2024.

### 4.1 France's digital health skills strategy

France is upgrading digital health skills for all health care learners and professionals [6]. The country's various legal acts enable progress to be made on digital skills and learning between 2021-2025, and thereafter. France was one of the first countries in the EU to make digital health a constituent part of its health training curricula. Digital training frameworks were developed either prior to 2023 or are now being developed. Their aim is to upgrade the digital health skills of all those professionals active in health and care delivery. A 122 million euro budget was allocated to implement these frameworks, the focus of which was on the use of digital tools, better collaboration and communication, more secure patient data, and preparation of future models of care.

A digital health training framework was set up. The French ***Digital Health Acceleration Strategy*** is part of the national France 2030 competitiveness investment plan, which has run from 2021-2025. Its first objective is centred on developing training in digital health, the confidence of stakeholders in digital health, and the professional attractiveness of the sector. In order to meet this ambition, the French Digital Health Delegation (DNS) has – as a priority – promoted a number of ***tools and guidelines*** for healthcare professionals, whether they are in full-time education and/or involved in life-long education.

Several frameworks were developed. Several full-time training courses for the following types of professionals are available: healthcare professionals – published in 2023; social workers, and administrative personnel – to be implemented around 2026. As a result of these deadlines, France anticipates that ***60% of healthcare learners*** will be familiar with digital health challenges by 2025, and – by 2027 – ***100% of learners*** in the sector will have been reached.

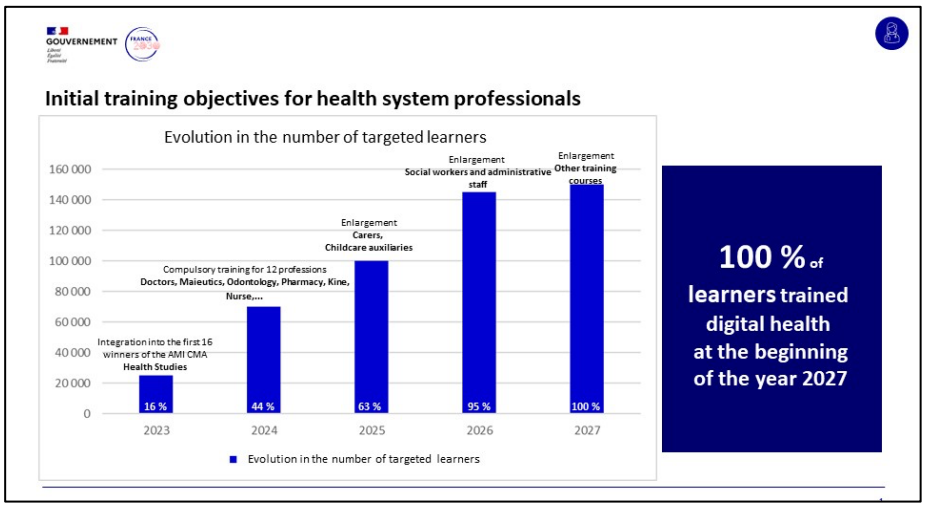



Figure 1: France – Initial training objectives for health system professionals

All these frameworks, and some relevant life-long training skills, are based around five key areas of competence: cybersecurity, health data, telehealth, communication, and digital tools.

France is determined to make ***digital skills part of every healthcare professional's training path***. The initiative is expected to evolve over time; keep up-to-date with ***new technologies and innovations***; and enable **cooperative involvement** with France's European neighbours.

### 4.2 NHS Education for Scotland

Scotland has a distinctive approach to the building of digital skills and leadership. Three Scottish strategies and plans for digital, data, and delivery aim at creating a health and social care system that supports safe, effective, and efficient care in the digital age. A workforce is foreseen that will have the capability and confidence to make effective use of data, information, and technology. Eight elements respond to a number

of challenges. They include the need to ensure senior leadership buy-in and the overcoming of the lack of a digital mindset. Specifically, the country's Building Digital Skills and Leadership Programme provides a response to the needs of the workforce. Based on an underpinning framework, the programme has identified six vital skills, knowledge, and behaviours essential to be applied at individual, team, or organisational levels. Members of the workforce are encouraged to transit learning pathways and work with a digital mindset, either as individuals or as teams. Indeed, for Scotland, it is multi-disciplinary, multi-professional teams that are able to drive change.

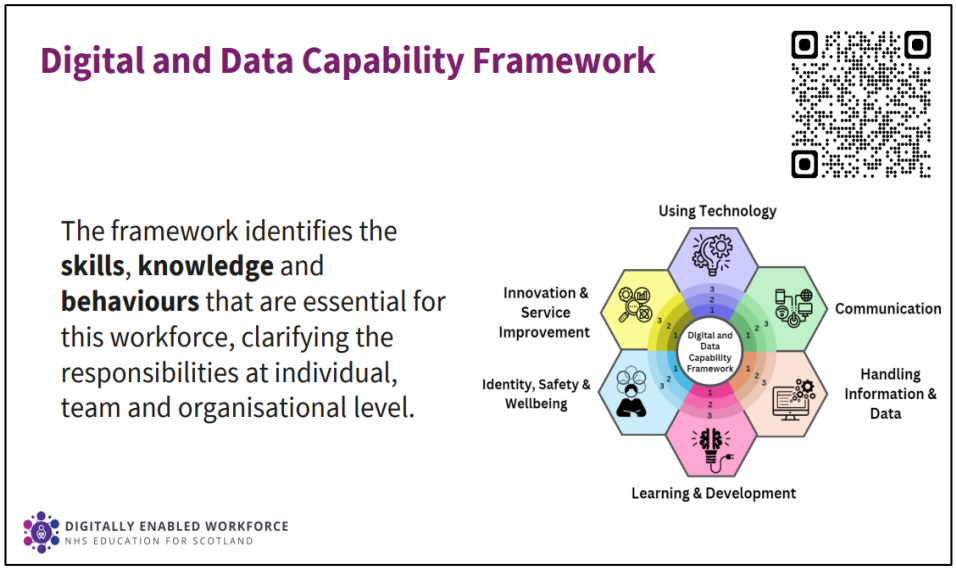


Figure 2: Scotland – Digital and data capability framework

Scotland's approach to developing a digitally-skilled workforce is both **structured and comprehensive**. The country is keen to share its knowledge with others. Scotland has many free-to-access national **resources** on offer. Among them are a learning network for 3,000+ members, a digital champions network, and a skills hub. The country's Microsoft M365 Skills Hub, which is full of webinars and guides, attracts some 18,000 users each month.

### 4.3 COMPDIG-Salut – digital competences of health professionals in Catalonia, Spain

**COMPDIG-SALUT** was an initiative promoted by the Catalan Department of Health and coordinated by the TIC Salut Social Foundation of Catalonia [7, 8], ***one of Spain's autonomous communities***. It focused on the need for digital competences for health professionals, including the development of a digital competence framework, an evaluation and accreditation model, and what actions are needed to train and quality health professionals in the relevant digital competences.

There were several key results of the initiative. One was the ***Digital Competences framework for Health Professionals***. It defines the specific digital skills that health professionals need to acquire in order to practice a healthcare profession or occupation from a safe, civic and critical perspective. Another was an **exploratory study** conducted in 2021 to assess the digital competence level of health professionals working in Catalonia. It observed the digital literacy of more than 800 healthcare professionals in Catalonia and their training needs. Nearly all had either a basic or an intermediate level of competence, indicating that there was still room for training improvements [9]. In 2022, ***an evaluation model*** was developed, and tested. The results showed the validity of the evaluation model while the need remained to ***reinforce training***. The study results served Catalonia as ***a guide*** to reformulate its ***accreditation and training strategy*** for healthcare professionals.

COMPDIG-Salut continued to apply the results of the initial study during 2023-2025 using funding supplied by the EU's Technical Support Instrument [10]. As a result, Catalonia expects to be able to define a strategic training plan to upskill its healthcare workforce, define a number of training pathways, and develop the relevant training content and materials. This is likely to lead to the ***standardisation, assessment, and improvement of*** the digital skills of healthcare professionals throughout the autonomous region.

## 5 Conclusion

Two key messages can be extracted from this position paper. When identifying emerging digital health skills for the health and care workforce, European Member States can draw on good (or best) practice examples from other countries. It will also be crucial to follow what is happening in the EU overall, especially at the umbrella level of concrete policy developments in skills strategies, and available training courses and programmes. As Europe's Union for Skills, its EU Talent Pool, its various projects and initiatives move forward, both trend-setting and fast-tracking countries and followers/beginners will no doubt increasingly come on board. Not only is it worthwhile to follow the specific outcomes of the BeWell project but also the directions that European workforce skills and training will take in the future.


## Acknowledgement

The orientation of the paper has emerged from the work of a four-year Erasmus+ Programme project called BeWell. The BeWell project has received funding from the ***European Union's ERASMUS Lump Sum Grant*** initiative under grant agreement ***N°825172*****.** Neither the project itself nor the European Commission should be held responsible for any opinions expressed in this position paper. They are entirely those of the author.


## References


1. Skills Union: https://commission.europa.eu/topics/eu-competitiveness/union-skills_en, last accessed 23 June 2025.
2. EU Talent Pool: https://ec.europa.eu/newsroom/home/items/807650/, last accessed 23 June 2025.
3. WHO Strategic Partners' Initiative for Data and Digital Health: https://iris.who.int/bitstream/handle/10665/380179/WHO-EURO-2025-11245-51017-77578-eng.pdf?sequence=1, last accessed 23 June 2025.
4. Launch of the WHO Observatory policy brief: https://eurohealthobservatory.who.int/news-room/events/item/2025/05/06/default-calendar/closing-the-digital-skills-gap-for-the-european-health-and-care-workforce , last accessed 23 June 2025.
5. BeWell Skills Monitor: https://bewell-project.eu/trainings/ , last accessed 23 Jue 2025.
6. Upgrading digital health skills for all healthcare learners and professionals: https://www.ehtel.eu/media-room/latest-news/249-upgrading-digital-health-skills-for-all-healthcare-learners-and-professionals-a-french-strategy-at-different-legal-speeds.html , last accessed 23 June 2025.
7. COMPDIG-Salut: https://www.ehtel.eu/media-room/latest-news/242-compdig-salut-enhancing-digital-competences-of-health-professionals-in-catalonia.html, last accessed 23 June 2025.
8. TIC Salut Social Foundation, an EHTEL member: https://www.ehtel.eu/18-articles/162-tic-salut-social.html , last accessed 23 June 2025.
9. COMPDIG-Salut results in the Journal of Medical Internet Research: https://www.ehtel.eu/media-room/latest-news/242-compdig-salut-enhancing-digital-competences-of-health-professionals-in-catalonia.html, last accessed 23 June 2025.
10. EU Technical Support Instrument: https://commission.europa.eu/funding-tenders/find-funding/eu-funding-programmes/technical-support-instrument/technical-support-instrument-tsi_en , last accessed 23 June 2025.

# AI in Health IT: Social and Ethical Considerations

*[1]Professor Neil Gordon and [2]Mr Hameed M. Aderemi*

[1,] University of Hull, Hull, HU6 7RX, UK
n.a.gordon@hull.ac.uk

[,2] University of Hull, Hull, HU6 7RX, UK
m.hameed-2021@hull.ac.uk

**Abstract.** The integration of Artificial Intelligence (AI) into Health IT is transforming patient care, decision-making, and medical research. However, these advances raise critical social and ethical issues. This paper will explore some of the impacts of AI-driven health technologies on privacy, data protection, and the freedom of access to medical information by highlighting tensions between innovation and individual rights. We examine shifts in the balance of power, as AI systems increasingly mediate interactions between patients, healthcare providers, and institutions, and potentially reinforce inequalities in access to care. We will consider the broader effects of AI adoption in both public and private healthcare organisations, addressing issues such as accountability, transparency, and bias. The paper will outline the need to educate both the public and health IT professionals about the societal implications of AI, ensuring that technological advances align with ethical principles and public trust. Overall, we will reflect on strategies for responsible AI deployment that uphold human rights and equitable healthcare outcomes, aligning the work with the United Nations' Sustainable Development Goal 3 for good health and well-being and where we are in terms of achieving this goal. The talk will provide examples from the United Kingdom, Nigeria and elsewhere of the sorts of practical approaches that illustrate good practice.

**Keywords:** eHealth, AI, mHealth, computer ethics

## 1. Introduction

AI is rapidly transforming health systems by enabling automated diagnostics, predictive analytics, robotic surgery, and personalised medicine [1]. The COVID-19 pandemic, for instance, accelerated AI adoption in contact tracing, symptom screening, and vaccine distribution [2]. This is illustrated in the Foresight AI system to help improve predictive healthcare in the United Kingdom (UK), using data to help predict health outcomes [3], While these technologies promise improved efficiency and clinical outcomes, they also raise critical ethical concerns and social challenges around privacy, data ownership, inequality, and the shifting locus of decision-making power in healthcare [4].

## 2. Privacy and Data Protection

AI requires access to large datasets, often including sensitive patient information. This raises urgent concerns regarding data security, consent, and surveillance. Although regulations like the General Data Protection Regulation (GDPR) in the European Union (EU) and Nigeria's Data Protection Regulation (NDPR) aim to safeguard personal health data, enforcement remains uneven, and many AI applications operate in regulatory grey zones [5].

A significant risk is the "re-identification" of anonymised datasets, where AI algorithms can infer identity from metadata or cross-referenced data sources [6]. As AI systems become more embedded in clinical decision-making, ensuring data protection and user autonomy is imperative.

### 3. Equity and Access to AI-Driven Health Technologies

Despite AI's potential to expand access to healthcare, especially in underserved regions, disparities persist. In low-resource settings, infrastructure limitations, lack of digital literacy, and algorithmic biases pose major obstacles from a societal perspective [7].

For example, in Nigeria, mobile health apps and remote diagnostics powered by AI are emerging as solutions to rural healthcare challenges. However, uneven internet access, underfunded health systems, and non-contextual AI models limit their effectiveness [8]. Conversely, in the UK, AI has been integrated into the National Health Service (NHS) in England through tools like the AI Triage System, but concerns remain about transparency and the explainability of AI-generated outcomes [9]. There are moves to look to regulate and improve access to AI, and this is explored later in section 8.

### 4. Power, Transparency, and Accountability

AI reshapes relationships between patients, providers, and institutions. As decision-making increasingly shifts from clinicians to AI systems, new power dynamics emerge. Patients may be excluded from understanding or challenging algorithmic decisions, especially in systems that lack interpretability [10].

Accountability is also obscured when errors occur. Determining liability—whether the fault lies with the AI developer, the healthcare provider, or the data—can be contentious. Transparent AI models, clear ethical guidelines, and regulatory oversight are essential for ensuring accountability and fostering public trust [11].

### 5. Bias in AI and its Health Implications

AI systems are only as fair as the data they are trained on. When datasets reflect historical biases—such as underrepresentation of certain populations—AI tools can perpetuate or amplify inequities in health outcomes. For example, one case involved a United States (U.S.) health risk algorithm that systematically underestimated the health needs of Black patients because it used healthcare spending as a proxy for health [12].

Correcting such bias requires inclusive data collection, community engagement, and the continuous auditing of AI systems for discriminatory outcomes [13].

### 6. Education and Ethical Literacy

Bridging the gap between innovation and ethical implementation requires educating both the public and Health IT professionals. Health workers need to understand not only how AI tools function, but also their implications for care quality, privacy, and informed consent [14]. Public engagement initiatives can build trust and promote informed consent, especially as patients interact with AI systems in clinical contexts.

### 7. Aligning AI with the SDGs: Progress and Shortfalls

The United Nations Sustainable Development Goal 3 (SDG 3)—Good Health and Well-being—calls for universal health coverage and access to quality healthcare services. AI can support this goal by optimising healthcare delivery and enabling targeted public health interventions [15]. Yet as of 2024, progress remains uneven. According to the UN's SDG Tracker, low-income countries still face significant barriers to achieving equitable healthcare due in part to technological divides. Efforts to align AI implementation with SDG 3 must account for systemic inequalities and support infrastructure development, equitable policy design, and participatory governance.

## 8. Toward Responsible AI in Healthcare

Ensuring that AI advances public health while safeguarding rights and improving society demands a multistakeholder approach [16]. Some states, such as the European Union, are implanting legislation to manage such safeguards [17]. These provide protection and frameworks within those states, but there is still a lack of global consensus. The following summarises guidance on key areas needed for responsible AI in health:

- **Policy Development**: Governments must establish enforceable ethical AI frameworks.
- **Inclusive Design**: AI tools should be developed with input from diverse populations.
- **Transparency**: Algorithms must be explainable and open to public scrutiny.
- **Monitoring and Evaluation**: Regular audits for fairness and accuracy are necessary.

By foregrounding ethical considerations in the development and deployment of AI systems, we can work toward more just and effective health technologies.

## 9. Conclusion

AI has the potential to reshape healthcare in unprecedented ways. However, its success must not be measured solely by efficiency or innovation, but by its alignment with human rights, public trust, and equitable outcomes. Drawing lessons from both developed and developing contexts, and grounding practices in global frameworks like the SDGs, we must pursue an ethical and inclusive vision for AI in Health IT.

## References


1. Singh, A. P., Saxena, R., Saxena, S., & Maurya, N. K. Artificial intelligence revolution in healthcare: Transforming diagnosis, treatment, and patient care. Asian Journal of Advances in Research, 7(1), 241-263, (2024).
2. Park, K. H., Qi L., and He L. UNESCO practices under the SDG4 and COVID-19 response framework: Topic modeling from 2003 to 2021, International Journal of Innovative Research and Scientific Studies 6.2, 322-329 (2023).
3. NHS, Foresight AI case study, https://digital.nhs.uk/data-and-information/research-powered-by-data/case-studies/foresight-ai (2025)
4. Floridi, L., Cowls, J., Beltrametti, M., and Chatilia, R., AI4People—an ethical framework for a good AI society: opportunities, risks, principles, and recommendations, Minds and machines 28, 689-707 (2018).
5. Mittelstadt B. Ethics of the health-related internet of things: a narrative review. Ethics and Information Technology. Sep;19(3):157-75, (2017).
6. Rocher L, Hendrickx JM, De Montjoye YA. Estimating the success of re-identifications in incomplete datasets using generative models. Nature communications. Jul 23;10(1):3069, (2019).
7. Gianfrancesco MA, Tamang S, Yazdany J, Schmajuk G. Potential biases in machine learning algorithms using electronic health record data. JAMA internal medicine. Nov 1;178(11):1544-7, (2018).
8. Agbeyangi AO, Lukose JM. Telemedicine Adoption and Prospects in Sub-Sahara Africa: A Systematic Review with a Focus on South Africa, Kenya, and Nigeria. InHealthcare, Mar 29 (Vol. 13, No. 7, p. 762). MDPI, (2025).
9. NHS, Artificial intelligence (AI) and machine learning, https://www.england.nhs.uk/long-read/artificial-intelligence-ai-and-machine-learning/, last accessed 2025/05/06. (2025)
10. Ananny M, Crawford K. Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. new media & society. Mar;20(3):973-89, (2018).
11. Leslie D. Understanding artificial intelligence ethics and safety. arXiv preprint arXiv:1906.05684, (2019).
12. Obermeyer Z, Powers B, Vogeli C, Mullainathan S. Dissecting racial bias in an algorithm used to manage the health of populations. Science. Oct 25;366(6464):447-53, (2019).
13. Veale M, Binns R. Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data. Big Data & Society. Nov;4(2):2053951717743530, (2017).

14. Morley J, Machado CC, Burr C, Cowls J, Joshi I, Taddeo M, Floridi L. The ethics of AI in health care: a mapping review. Social science & medicine. Sep 1;260:113172, (2020).
15. UN (United Nations)., The Sustainable Development Goals Report 2022. https://unstats.un.org/sdgs/report/2022/, (2022).
16. Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. AI in health and medicine. Nature medicine, 28(1), 31-38, (2022).
17. EU, The EU Artificial Intelligence Act, https://artificialintelligenceact.eu/ (2025)

# Medical AI: Are we playing Russian roulette with patient lives?

*Dr Uchenna Nnawuchi*
ALERT Research Group, Middlesex University, London, UK
UN055@live.mdx.ac.uk; uchennannawuchi@gmail.com

## 1. Introduction

Artificial Intelligence (AI) fuelled Machine Learning (ML) algorithms have championed the revolution of how various industries, including finance, transportation, and manufacturing, conduct their daily operations. One of the major advances of AI is in the healthcare sector. This convergence of technology and healthcare has ushered in a new era in how medical professionals diagnose, treat, and care for patients. As technology continues to evolve, the role of AI in healthcare expands, driven by computing power and cloud infrastructure. These advances have rendered AI tools increasingly efficient and capable of managing complex tasks across various healthcare domains, as they continue to witness some promising developments.[3] From diagnostics to drug discovery, personalised treatment plans, virtual assistants, and robotic surgery, AI is proving to be an indispensable tool in advancing healthcare efficiency and patient care.[4] It is no wonder why AI systems in healthcare are considered by some researchers to be more efficient, effective, and accurate than their human counterparts.[5]

Notably, AI is contributing to the reshaping of healthcare through its impact on diagnostic accuracy. By analysing medical images such as X-rays, MRIs, and CT scans, AI algorithms can detect subtle abnormalities that might be missed by the human eye. This increased precision helps doctors diagnose conditions earlier and tailor or personalise treatments to a particular patient, which is often more effective, leading to improved patient outcomes.[6] This move away from a one-size-fits-all approach to patient treatments ensures that patients receive the most effective treatments, minimising the risk of side effects and optimising recovery. Arguably, AI is paving the way for healthcare that is more efficient, effective and individualised. Furthermore, AI's revolution of the healthcare industry does not stop at diagnosis alone; it extends to aggregating data to identify patterns and predict potential health risks.[7] In this way, AI enables healthcare providers not only to react timeously to diseases but to anticipate them, allowing for earlier interventions and tailored treatment plans that improve overall patient care.[8] Thus, in resource-constrained environments, where healthcare professionals are often overburdened, AI's ability to comb and streamline processes and facilitate quicker diagnoses is especially crucial, helping to ensure that quality care reaches those who need it most.

In parallel, AI-powered robotic systems enhance surgical precision, enabling minimally invasive procedures that lead to faster recoveries and fewer complications,[9] and virtual health assistants offer 24/7 access to medical information and personalised care, improving patient engagement while reducing the workload on healthcare professionals.[10] Thus, these systems are fundamentally transforming healthcare, providing tools

---

[3] Ömer Aydin and Enis Karaarslan, 'OpenAI CHATGPT Interprets Radiological Images: GPT-4 as a Medical Doctor for a Fast Check-Up' (3 February 2025) arXiv.org https://arxiv.org/abs/2501.06269.

[4] Shuroug A Alowais and others, 'Revolutionizing Healthcare: The Role of Artificial Intelligence in Clinical Practice' (2023) 23 *BMC Medical Education* 689.

[5] S Holm, C Stanton and B Bartlett, "A New Argument for No-Fault Compensation in Health Care: The Introduction of Artificial Intelligence Systems" (2021) 29(3) Health Care Analysis 171, 180.

[6] Aydin n1.

[7] Rajkomar, A., Oren, E., Chen, K. et al. Scalable and accurate deep learning with electronic health records. npj Digital Med 1, 18 (2018). https://www.nature.com/articles/s41746-018-0029-1

[8] Pierce, R., Sterckx, S., & Van Biesen, W. (2022, February). *A riddle, wrapped in a mystery, inside an enigma: How semantic black boxes and opaque artificial intelligence confuse medical decision-making*. Bioethics. https://pmc.ncbi.nlm.nih.gov/articles/PMC9291279/ \

[9] Ashokkumar Joshi., 'Artificial Intelligence (AI) in Healthcare' (27 February 2024) SSRN https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4728304.

[10] Rajkomar, A., Dean, J., & Kohane, I. (2019). Machine learning in medicine. New England Journal of Medicine, 380 (14) 1347 - 1358.

that enhance precision, reduce costs, and streamline processes. Their ability to manage complex tasks with efficiency and accuracy is not only improving diagnostic and treatment capabilities but also offering faster, more personalised care. As these technologies continue to evolve, they promise a future where healthcare is more adaptive and effective, addressing the growing demands of both patients and healthcare professionals.[11]

**2. Potential Criticism of AI**

Despite the immense potential of AI in healthcare, several challenges persist that temper its promises. While AI's capacity to learn from vast amounts of data and enhance its performance is well-documented, it must be recognised that these systems are only as effective as the data they are trained on and the human expertise that moulds their algorithms.[12] *Certainly, those who pay the piper dictate its tune*. In other words, in practice, AI developers determine the type of data used, select relevant features, and define the intended function and precise goal of the algorithm. As a result, they shape how these systems operate, training them to perform in ways aligned with specific goals. This central role in the development of AI systems raises critical concerns: the risk of perpetuating bias. These systems often absorb biases inherent in their training data, which can perpetuate inequalities, especially when certain demographics are underrepresented. For instance, if a particular race, gender, or age group is inadequately represented in the training data, the AI's predictions and decisions may not accurately reflect the needs of these groups, potentially exacerbating disparities in healthcare.[13] Alongside the issue of bias, privacy concerns pose another major challenge. Healthcare data, crucial for training AI models, contains sensitive personal information.[14] The use of such data raises significant questions about data ethics, data security, data protection, patient consent, and the potential for misuse.[15]

**3. Lack of ML Inscrutability**

Germane to this paper is the issue of AI inscrutability. Some AI systems in healthcare operate as "black boxes," where the decision-making process is not transparent either to clinicians or patients.[16] This lack of transparency, coupled with the complexity of the underlying algorithms, raises significant concerns about the trustworthiness of AI-driven decisions.[17] Although AI can identify patterns and make predictions, the reasoning behind these decisions is often obscured, making it challenging for clinicians to explain or justify the role of AI in medical choices. More importantly, this opacity obstructs the process of obtaining informed consent from patients. If patients cannot understand how AI arrived at a particular diagnosis or treatment recommendation, they cannot make informed decisions regarding their care. Similarly, clinicians may struggle to adequately explain AI's role, undermining the principle of informed consent and creating a risk of automation bias, where clinicians may simply defer to the algorithm's decisions, assuming they are more efficient or knowledgeable.[18]

The issue of opacity extends beyond informed consent and trustworthiness to the question of accountability and liability when medical errors occur.[19] Traditional liability frameworks may fail to address the complexities introduced by AI-driven technology, leading to uncertainty about who is responsible for adverse

[11] Joshi n7.
[12] T Maddox, JS Rumsfeld and PR Payne, "Questions for Artificial Intelligence in Health Care" (2019) 321(1) Journal of the American Medical Association 31.
[13] Paul Nolan and Rita Matulionyte (n.d.). *Artificial Intelligence in medicine: Issues when determining negligence*. Journal of law and medicine. https://pubmed.ncbi.nlm.nih.gov/38332598/.
[14] Joshi n7.
[15] Dimple Patil. (n.d.). The Future of Artificial Intelligence in healthcare: Opportunities, challenges, and ethical considerations. https://www.researchgate.net/publication/385746632_The_future_of_artificial_intelligence_in_healthcare_Opportunities_challenges_and_ethical_considerations
[16] Sterckx n6.
[17] ibid.
[18] Nolan n11.
[19] ibid.

outcomes.[20] This creates potential violations of both human and digital rights, as patients may be unable to challenge AI decisions or seek accountability when harm occurs.[21] Consider circumstances when an AI system misdiagnoses a condition or provides an inaccurate treatment recommendation (false positive): who bears the responsibility? The clinician who relies on the AI's output, the manufacturer of the system, or perhaps the data scientists who developed the algorithms?[22] These questions are difficult to answer without a clear understanding or explanation of how the AI arrived at its conclusion. In such situations, traditional models of legal liability, which often rely on the ability to trace causation and foresee risks, struggle to accommodate the complexity and opacity of AI-driven decisions.[23] More troubling is the realisation that even the systems manufacturer does not understand how this decision was reached. This opacity problem led to the discourse on the right to explanation and explainability as a persuasive, appealing approach.[24]

## 4. The right to explanation

This lack of explainability not only complicates liability but also undermines trust at multiple levels. Clinicians, unsure of how an AI system arrived at its decision, may hesitate to fully trust its recommendations, which in turn diminishes their ability to explain these decisions to patients.[25] This lack of transparency creates a dangerous cycle where both clinicians and patients are left in the dark about critical health decisions, potentially jeopardising patient safety.[26] As patients enter treatment based on AI-driven recommendations, they may not have the opportunity to make informed choices because the rationale behind the AI's decision is inaccessible to them. This not only undermines the core principle of informed consent, which requires clinicians to provide clear and cogent explanations of diagnoses and treatment options but is also antithetical to the cornerstone of medical ethics that before a patient can fully consent to something, they should, at the very least, know all material matters.[27] The absence of this material knowledge may violate patients' rights to make autonomous decisions regarding their care.

Additionally, this lack of explanation also directly affects a patient in a court setting that is seeking to establish that a breach of duty has occurred.[28] Moreover, one might ask: if a risk cannot be identified, how can precautions be implemented to mitigate it? McDougall,[29] and Martin and Martin-Granel [30]argue that a patient

---

[20] ibid.
[21] ibid.
[22] ibid.
[23] ibid.
[24] Lilian Edwards and Michael Veale, 'Slave to the Algorithm? Why a "right to an Explanation" Is Probably Not the Remedy You Are Looking For' (SSRN, 24 May 2017); S Wachter, B Mittelstadt, and L Floridi, 'Why a Right to Explanation of Automated Decision-Making Does Not Exist in the General Data Protection Regulation' (2017) 7 International Data Privacy Law 76; ryce Goodman and Seth *Flaxman* Eu regulations on algorithmic decision-making and a right to explanation https://www.researchgate.net/publication/304548505_EU_regulations_on_algorithmic_decision-making_and_a_right_to_explanation; ae Wan Kim and Bryan R. Routledge, 'Why a Right to Algorithm decision-making should exist; A trust-based approach', Cambridge University Press https://www.cambridge.org/core/journals/business-ethics-quarterly/article/why-a-right-to-an-explanation-of-algorithmic-decisionmaking-should-exist-a-trustbased-approach/C620A6A5FCB781384D20E08BE4CD09BC; Maja Brkan, 'Do algorithm rule the world? Algorithmic decision-making and data protection in the framework of the GDPR and beyond,' https://techpolicyinstitute.org/wp-content/uploads/2018/02/Brkan_do-algorithms- rule.pdf#page=1&zoom=auto,-34,842
[25] Nolan n11.
[26] GI Cohen, "Informed Consent and Medical Artificial Intelligence: What to Tell the Patient?" (2019) 108(6) Georgetown Law Journal 1425; Julia Amann et al, "Explainability for Artificial Intelligence in Healthcare: A Multidisciplinary Perspective" (2020) 20(1) BMC Medical Informatics and Decision Making 1
[27] ibid.
[28] Nolan n11.
[29] RJ. McDougall, Computer knows best? The need for value-flexibility in medical AI J Med Ethics, 45 (3) (2019), pp. 156-160, 10.1136/medethics-2018-105118; C Bjerring, J. Busch Artificial intelligence and patient-centered decision-making Philos Technol, 34 (2) (2021), pp. 349-371, 10.1007/s13347-019-00391-6. Paul M Martin, Estelle Martin-Granel '2,500-year evolution of the term epidemic' Emerg Infect Dis, 12 (6) (2006), pp. 976-980, 10.3201/eid1206.051263.
[30] Paul M Martin, Estelle Martin-Granel '2,500-year evolution of the term epidemic' Emerg Infect Dis, 12 (6) (2006), pp. 976-980, 10.3201/eid1206.051263

is unlikely—or even unable—to desire or require a deep understanding of an AI system's inner workings. Instead, they will primarily seek an ordinary explanation of the system's efficacy and associated risks. However, while such an explanation may clarify the system's efficiency, it does not address the crucial question of whether the patient actually understands how the decision was reached, thereby enabling them to make an informed choice.[31] They posited that ML outputs warrant trust: they claimed that if we accept the trustworthiness of human decision-making in medicine despite its occasional opacity, it follows that similarly opaque, yet demonstrably accurate AI systems may also merit our trust. Furthermore, even if this reasoning holds water, what happens if an error occurs despite assurances regarding the system's efficiency? What recourse does the patient have in such a situation?

To address these concerns, explainability must be integrated as a core component of medical AI systems. Ensuring that AI algorithms are transparent and understandable—particularly in terms of how they make decisions—could bridge the gap between technological advances and patient safety. By making AI systems more interpretable, clinicians can better understand the reasoning behind AI-driven recommendations, which would enhance their ability to explain those decisions to patients. This, in turn, would restore patient trust, allowing patients to make fully informed choices regarding their treatment options.[32] Moreover, the ability to explain AI decisions could help mitigate risks related to liability, as it would enable healthcare providers to more effectively navigate the complexities of causation and responsibility in the event of an error.[33]

## 5. Conclusion

Incorporating explainability into AI healthcare systems is not without its challenges, particularly given the complexity of ML models and the vast datasets required for their training.[34] However, developing systems that can provide clear and functional explanations of their decision-making processes is critical to ensuring that AI remains a tool that enhances rather than endangers patient care.[35] As AI continues to play an increasingly prominent role in healthcare, regulators, developers, and healthcare providers must work collaboratively to prioritise explainability. Only by doing so can they mitigate the risks likened to a game of "Russian roulette", where AI-driven decisions directly impact patient lives. Thus, to ensure the protection of patient rights and the safe integration of AI into healthcare, AI systems developers and operatives, including manufacturers, must be accountable and transparent, empowering both clinicians and patients to trust the technology that is meant to improve their lives.

[31] McDougall n27 and Martin n28.
[32] Cohen n22.
[33] ibid.
[34] Nolan n10.
[35] Nnawuchi, U. and George, C. 2024. Orwellian odyssey: smart borders and the imperative for explainability. Farhaoui, Y., Herawan, T., Imoize, A.L. and El Allaoui, A. (ed.) 6th International Conference on Artificial Intelligence and Smart Environments. Errachidia, Morocco 07 - 09 Nov 2024 Springer

# Decoding memory states using machine learning: Ethical, legal and governance implications

*Mr Sailendra Ramanna*
Information Services Division (ISD), University College London, UK
s.ramanna@ucl.ac.uk

**Abstract** Brain-computer interface research has increasingly explored decoding cognitive states from neural signals for healthcare applications. This paper presents a deep learning pipeline for classifying "remember" vs. "forget" cognitive states from electroencephalography (EEG) data. The method integrates Dynamic Mode Decomposition (DMD) for spatiotemporal feature extraction with a convolutional neural network (CNN) classifier. The pipeline's stages are described and experimental results are summarised, indicating successful differentiation of memory recall states. Technical results suggest the model captures subsequent memory effects (with gamma-band EEG activity higher for remembered events). The focus of this work is also the broader context of ethical, legal and governance considerations. The paper emphasises that responsible innovation in emerging neurotechnology requires interdisciplinary collaboration, aligning technical advancement with ethical principles and legal frameworks to ensure safe and acceptable healthcare applications.

**Keywords** EEG · Psychology · DMD · Machine Learning · Ethics ·Governance

## 1.Introduction

Decoding human cognitive states from brain signals is a frontier of neurotechnology with significant potential in healthcare and neuroscience research. Whether a person remembers or forgets information from their brain signals can be detected. EEG recordings could aid memory training and early diagnosis of memory impairments. Prior studies have shown that EEG activity differs between subsequently remembered and forgotten items, especially in certain frequency bands [14].

Oscillations in the theta (4–8 Hz) and gamma (30–100 Hz) bands have been found to be stronger during encoding of items that are later remembered compared to those later forgotten [6]. These neural markers of memory encoding success provide a basis for computational approaches to predict memory performance. Deep learning has recently been applied to EEG data to capture complex, multivariate patterns that traditional analyses might miss [14]. Convolutional neural networks (CNNs) and other architectures can automatically learn features from raw or minimally processed EEG, enabling state-of-the-art classification of mental states and conditions [10]. However, training deep models on EEG requires careful consideration of feature extraction and dimensionality due to the high temporal resolution and noise in EEG signals.

The primary contribution of this paper is two-fold: (a) To describe a novel EEG classification pipeline using DMD and deep learning for decoding "remember" vs. "forget" cognitive states from gamma-band EEG activity; and (b) to discuss the ethical, legal and governance implications of such neurotechnology in healthcare. A 6-stage pipeline is proposed, where an EEG signal is fed and the signal goes through the different stages [8] as shown in Figure 1 below. The final output of this pipeline shows whether a person has remembered or forgotten the artefacts presented to them, in terms of degree of accuracy.

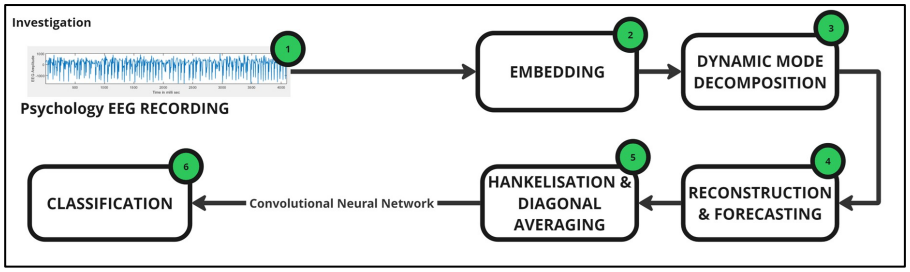



Figure 1: Different stages involved in the EEG classification pipeline

## 2.Data and Processing

This section details all the stages of the pipeline starting with the data capture, all the way through to the classification, as shown in Figure 1 above.

### 2.1 Data Capture

This investigation utilised an EEG dataset obtained from clinical physicians. The study was conducted by Middlesex University London in 2024. To record the original dataset, participants engaged in a memory encoding task. During the task, certain stimuli (words or images) were presented, and later a memory test identified which stimuli were remembered or forgotten by each participant. EEG was recorded from multiple scalp electrodes according to the standard 10-20 system (See Figure 2), with a focus on capturing high-frequency brain activity related to memory processing.

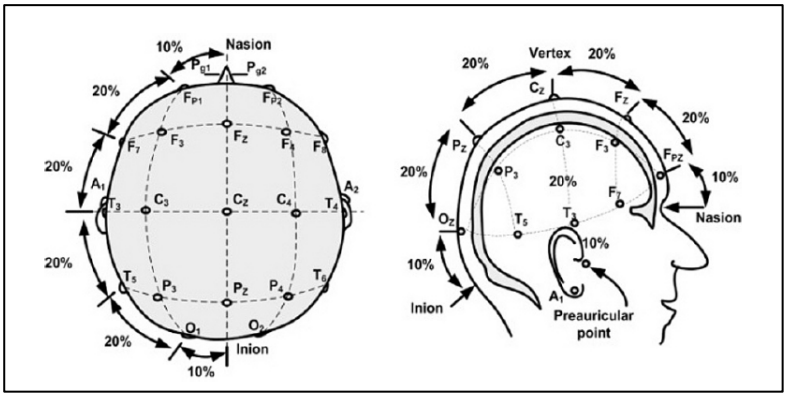



Figure 2: The international 10-20 system is a standardised method for positioning electrodes on the scalp during EEG recording. It divides the head into proportional distances (10% and 20%), ensuring consistent, reproducible placement to capture brain activity [4]

### 2.2 Embedding

Each EEG segment was standardised and projected onto its leading principal components (e.g., top ten), reducing dimensionality while preserving key spatial patterns across electrodes. This step filtered out noise and accelerated subsequent analyses by focusing on the most informative variance in each trial.

### 2.3 Dynamic Mode Decomposition (DMD) Feature Extraction

On the Principal Component Analysis (PCA)-reduced data, DMD computed a few dominant spatiotemporal modes and their eigenvalues, capturing oscillatory patterns (like gamma-band bursts) and their temporal behaviour. By limiting the Singular Value Decomposition (SVD) rank, DMD isolated the brain's primary rhythmic dynamics relevant to memory encoding. Prior work has shown that DMD can serve as a feature extraction or dimensionality reduction step for EEG, capturing information-rich patterns for subsequent classification [5,9,12,16].

### 2.4 Reconstruction & Forecasting

The extracted DMD modes were recombined to reconstruct each EEG epoch (a short, segmented interval of continuous EEG recording), verifying that the decomposition retains essential signal structure. Optionally, modes were propagated via their eigenvalues to forecast future signal states, offering insight into evolving neural dynamics.

### 2.5 Hankelisation & Diagonal Averaging

Hankelisation is a mathematical operation used in signal processing that transforms a one-dimensional time series signal into a structured matrix (Hankel matrix). Reconstructed time-series for each channel were converted into overlapping Hankel matrices (window = 50) to embed short-term temporal context. Diagonal averaging then collapsed each matrix into a smooth one-dimensional (1D) feature vector by averaging along its anti-diagonals, yielding compact, denoised representations for CNN input.

### 2.6 Deep Learning Classification (CNN Model)

The features derived from DMD (or alternatively, the time-frequency representations if DMD modes were assembled into images) were fed into a convolutional neural network. The CNN architecture was designed to handle the input feature dimensions and learn discriminative patterns corresponding to the two cognitive state classes. In the implementation, the CNN took as input either a two-dimensional (2D) representation of EEG features (e.g., concatenated features from multiple time snapshots) or sequences processed via 1D convolution across channels.

### 3.Training and Validation Procedure

The dataset was split into training and validation sets (and where possible, a separate test set or cross-subject folds for robustness). The model was trained for multiple epochs, and performance was monitored on validation data to detect convergence or overfitting. The training accuracy increased rapidly and eventually plateaued, while validation accuracy initially rose and then levelled off, a pattern suggestive of the model learning the underlying signal differences up to a point, after which generalisation became challenging. Figure 2 shows an example of training vs. validation accuracy over training epochs for the CNN classifier on the memory state EEG data.

### 4.Results

Section 4 discusses the EEG signal processing and classification pipeline employed in this research, emphasising its key computational steps. Sub-section 4.1 specifically elaborates on embedding, decomposition, and signal reconstruction techniques, fundamental to accurately decoding cognitive memory states from EEG signals. These steps collectively ensure the robustness and accuracy of the machine learning pipeline.

### 4.1 Pipeline Performance Overview

The DMD-CNN classification pipeline successfully learned to differentiate between EEG signals of subsequently remembered and forgotten items. Qualitatively, the model's performance indicated that there are discernible patterns in gamma-band EEG associated with whether a memory will be retained. During training, the CNN's accuracy on the training set quickly increased, often nearing perfect classification after sufficient epochs, reflecting the model's capacity to fit the data.

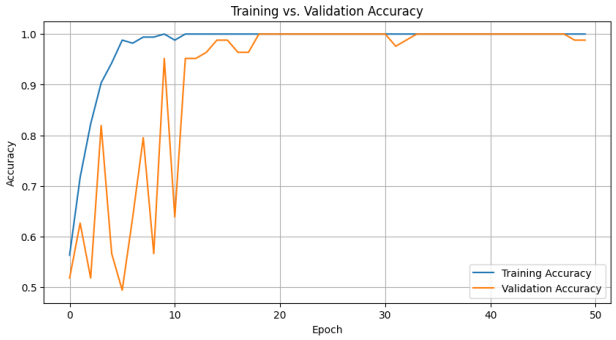



Figure 3: Training and validation accuracy trends over epochs for the EEG memory state CNN classifier. The training accuracy (blue) and validation accuracy (orange) both approach 100 suggesting the pipeline generalises relatively well.

After an initial rapid rise, validation performance tended to plateau, suggesting that the model extracted the most significant discriminative features early and that further training mainly improved fit to the training data rather than generalisation (as shown by the gap between the training and validation curves in Figure 2). This plateau around a certain accuracy level is consistent with the inherent difficulty of the task and the amount of informative signal present in the EEG data. In this study, the validation accuracy reached an average accuracy above 80%, demonstrating that the pipeline captured real differences related to memory encoding success. Previous EEG memory decoding studies (not involving DMD) reported moderate classification accuracy (often in the 70% range) for distinguishing between remembered and forgotten events [14]. This indicates the superiority of DMD for feature selection.

## 5. Ethical, legal, and governance implications

The ability to infer whether someone remembers or forgets information from brain signals touches on the concept of "mental privacy" or cognitive liberty [7]. Unlike observable behaviour, EEG-based decoding peers directly into brain activity, potentially revealing hidden states of mind. This raises the question of whether such mind-reading capabilities could infringe on an individual's right to privacy of thought. Ethicists have argued that neurotechnologies enabling access to one's mental states necessitate new "neurorights," such as the right to mental privacy and the right to cognitive freedom [13].

### 5.1 Data Governance: Consent and Data Handling in EEG Studies

Responsible data governance is essential at both the research and deployment stages of EEG-based cognitive state decoding. Under data protection regulations like the GDPR in Europe, EEG recordings can fall under the GDPR Article 9 (processing of special categories of personal data - e.g. biometric or health data), given their uniqueness and health relevance. The GDPR thus mandates a higher standard of protection: explicit consent for processing such data, which must be "freely given, specific, and informed," including an explanation of potential mental privacy risks [11].

In the context of the pipeline, the study was conducted in the United Kingdom and hence UK legislation applies. This means that any subject or patient whose EEG is recorded for memory decoding must be made aware that the data might reveal aspects of their memory performance or other unintended information, and they must agree to that knowledge being extracted. Additionally, data governance principles like purpose limitation and data minimisation are crucial [17].

Data minimisation would encourage only collecting and storing the data necessary for the intended function. Hence, storing derived features or summary metrics instead of raw EEG if feasible, to reduce the risk exposure in case of a data leak [15]. Secure handling of EEG data is imperative: datasets should be anonymised (stripped of direct identifiers) and stored encrypted to prevent unauthorised access [2]. Thus, strong cybersecurity measures are part of data governance for neurotech. Furthermore, individuals should

retain rights over their brain data: the right to withdraw and have data deleted is enshrined in laws like GDPR [1]. Implementing these rights means a participant could at any point request that their EEG recordings and analysis outcomes be erased from research records or device memory.

Continuous evaluation is needed to detect and mitigate any biases in training data or model that could lead to unequal outcomes, to ensure the tool benefits all segments of the population. Training users and professionals to appropriately calibrate trust in the system is part of ethical deployment [3].

## 6. Conclusion

This paper presented a deep learning-based EEG classification pipeline for distinguishing "remember" vs. "forget" cognitive states, integrated with a thorough analysis of the ethical, legal, and governance issues surrounding such neurotechnologies. The pipeline should be able to use different time series datasets. Technically, the combination of DMD and CNN modelling proved effective in capturing gamma-band EEG dynamics associated with memory encoding success, highlighting the potential for data-driven approaches to decode latent cognitive states. The pipeline's results, interpreted in light of existing neuroscience knowledge, demonstrate the promise of using EEG as a window into memory processes, which could have applications in educational tools, cognitive rehabilitation, or assessment of memory impairments (for example Alzheimer's patients). However, alongside these advances, responsible innovation is imperative. The ethical, legal and governance considerations discussed serve as a reminder that decoding the brain is not just a technical challenge, but a human one with profound implications.

## References


1. Ahluwalia, M.: Legal governance of brain data derived from artificial intelligence. Voices in bioethics **7** (2021).
2. Burwell, S., Sample, M., Racine, E.: Ethical aspects of brain computer interfaces: a scoping review. BMC medical ethics **18**(1), 60–60 (2017).
3. Ligman, D.: Ai and healthcare part 2: Risks, misuse, and ethical challenges (2025). URL https://dianeligman.medium.com/ ai-and-healthcare-part-2-risks-misuse-and-ethical-challenges-e36ce830145b. Accessed: 2025-05-08.
4. Ma, J.: Using nonlinear analysis and neural network to classify bipolar i disorder electroencephalogram signals from normal electroencephalograms. Network modeling and analysis in health informatics and bioinformatics (Wien) 12(1), 32 (2023).
5. Ogundile, O., Usman, A., Babalola, O., Versfeld, D.: Dynamic mode decomposition: A feature extraction technique based hidden markov model for detection of mysticetes' vocalisations. Ecological Informatics 63, 101306 (2021). DOI https://doi.org/10. 1016/j.ecoinf.2021.101306. URL https://www.sciencedirect.com/science/article/ pii/S1574954121000972.
6. Osipova, D., Takashima, A., Oostenveld, R., Fernandez, G., Maris, E., Jensen, O.: Theta and gamma oscillations predict encoding and retrieval of declarative memory. Journal of Neuroscience **26**(28), 7523–7531 (2006). DOI 10.1523/JNEUROSCI.1948-06.2006. URL https://www.jneurosci.org/content/26/28/7523.
7. Rainey, S.: Neurorights versus externalism about mental content: Characterizing the 'harm' of neurotechnological mind reading. Cambridge quarterly of healthcare ethics pp. 1– (2025).
8. Ramanna, S., Tirunagari, S., Windridge, D.: Epileptic seizure detection using constrained singular spectrum analysis and 1d-local binary patterns. Health and technology **10**(3), 699–709 (2020).
9. S, A., Gopalakrishnan, E.A., Soman, K.P.: Investigating the effectiveness of dmd and its variants for complex data analysis. In: 2019 International Conference on Intelligent Computing and Control Systems (ICCS), pp. 993–998 (2019).

10. Saeidi, M., Karwowski, W., Farahani, F.V., Fiok, K., Taiar, R., Hancock, P.A., Al-Juaid, A.: Neural decoding of eeg signals with machine learning: A systematic review. Brain Sciences **11**(11) (2021). DOI 10.3390/brainsci11111525. URL https://www.mdpi.com/2076-3425/11/11/1525.
11. SecurePrivacy: Neurodata consent frameworks: Managing eeg/brain-computer interface data under gdpr/ccpa (2025). URL https://secureprivacy.ai/blog/ neurodata-consent-eeg-brain-computer-interface-data-gdpr-ccpa. Accessed: 2025-05-08.
12. Solaija, M.S.J., Saleem, S., Khurshid, K., Hassan, S.A., Kamboh, A.M.: Dynamic mode decomposition based epileptic seizure detection from scalp eeg. IEEE Access **6**, 38683– 38692 (2018).
13. Strategies, P.: Neurorights foundation (2024). URL https://perseus-strategies. com/neurorights-foundation/#:~:text=Neurorights%20Foundation%20,serious%20concerns%20for%20mental. Accessed: 2024-05-08.
14. Sun, X., Qian, C., Chen, Z., Wu, Z., Luo, B., Pan, G.: Remembered or forgotten?—an eeg-based computational prediction approach. PLOS ONE **11**, e0167497 (2016). DOI 1371/journal.pone.0167497
15. Thakur, A., Zhu, T., Abrol, V., Armstrong, J., Wang, Y., Clifton, D.A.: Data encoding for healthcare data democratization and information leakage prevention. Nature communications **15**(1), 1582–1582 (2024).
16. Tirunagari, S., Kouchaki, S., Poh, N., Bober, M., Windridge, D.: Dynamic Mode Decomposition for Univariate Time Series: Analysing Trends and Forecasting (2017). URL https://hal.archives-ouvertes.fr/hal-01463744. Working paper or preprint.
17. Ward, A., Holmes, M.E., Ward, I., Lokmic-Tomkins, Z., East, L., Levett-Jones, T.: Environmental, social and governance principles in Australian publicly funded healthcare: an extension of value-based care. Australian health review 49(2), 1–3 (2025).

# Artificial Intelligence in Biobanking: Potentials, Challenges and Future Trends in Oncology

*[1]Dr Radwa Hassan, [2]Professor Fayek Elkhwsky and [3]Professor Galal H. Galal-Edeen*

[1]Department of Public Health and Community Medicine,
Faculty of Medicine, Galala University, Egypt
Radwa.ali@Gu.edu.eg

[2]Medical Research Institute, Alexandria University, Egypt.
Felkhwsky@alexu.edu.eg

[3]Heikal Department of Management, Onsi Sawiris School of Business
American University in Cairo, Egypt
Galaledeen@aucegypt.edu

**Abstract:**
***Background*:** Biobanks are vital for biomedical research, especially in oncology, and are key entry points for precision medicine. Digitalization and artificial intelligence (AI) advancements offer opportunities to enhance their functionality and contribute to advancements in research.

***Objectives*:** To investigate the potentials and challenges associated to the implementation of biobank digitalization with its new innovations in low- and middle-income countries (LMICs), especially in the field of oncology.

***Methods*:** The potentials and challenges of biobanks digitalization in LMICs are identified through a literature review. Authors highlight the oncology related domains, suggesting potential AI solutions to the emphasized challenges.

***Results*:** The study provides an overview of biobanking in LMICs, highlighting challenges such as data heterogeneity, privacy, security, standardization, algorithm interpretability, and ethical issues, while utilizing AI for data analysis and tracking.

***Conclusion*:** Digitalization of biobanks in LMICs can enhance research, healthcare, and tackle global health issues, particularly in oncology, using modern AI techniques.

**Key Words:** Biobanking, Digitalization, LMICs, Machine Learning, Oncology,

## 1. Introduction

The primary objective of biobanks is to gather, handle, preserve, and disperse biological specimens and related data in a controlled and structured way. Various biological sample types along with extensive databases containing clinicopathologic data are available through biobanks [1]. Tissues (normal and/or aberrant), biofluids (blood, serum, plasma, urine, saliva, liquor, effusion, bone marrow fluid part, sperm, cord blood), cells, peripheral blood cells (PBCs), and nucleotides (DNA, RNA, miRNA) are among the samples that biobanks can collect. Each of these biospecimens is combined with the related data (clinicopathologic, genetic, and personal data) to create databases [2].

Simply put, Machine Learning (ML) is a type of Artificial Intelligence (AI) that aims to enable computing machinery to learn from and identify patterns and relationships in sufficient representative data samples,

and to successfully apply this knowledge for making decisions on material that has not yet been seen [3]. ML uses a variety of learning algorithms (learning also has sub-types such as supervised, unsupervised, semi-supervised, and enforcement learning, depending on the existence and extent of labelled data and the evaluation of the results of learning). Typically, a repository of well-organized data is made available for a learning algorithm, it uses such data to build models that reflect the relationships between data items. The accuracy and quality if such models is evaluated, successively refined to be used in tasks such as clustering, classification, recognition of patterns and images, as well as the induction of rules and models that enable prediction using new, previously unseen data.

The aim of the current work is to investigate the potentials and challenges associated to the implementation of biobank digitalization with its new innovations in low- and middle-income countries (LMICs), especially in the field of oncology. Moreover, future trends that can cross the gaps and address the challenges of biobanks digitalization and artificial intelligence integration would be further highlighted.

## 2. Potential of AI in Biobanking

Oncology research has benefited greatly from the application of AI in biobanking, which has led to important breakthroughs in the field. In the area of oncology, the following are some developments and uses of AI in biobanking:

### *Data Management*

Biobanks house various omics data, including transcriptomics, proteomics, metabolomics, and genomes. AI is being used to integrate these multi-omics datasets, enabling a deeper understanding of molecular pathways. AI algorithms can identify novel biomarkers and therapeutic trajectories. This seamless data management helps standardize and harmonize disparate datasets, overcoming data variability and increasing research quality. AI also facilitates research networks, promoting cooperation and data exchange [4-7].

### *Image Analysis and Radiomics*

Radiomics is a research area that converts normal nursing photographs into quantitative image data, allowing for accurate analysis using AI or traditional biostatistics methods. This data can provide non-invasive biomarkers for cancer risk prediction, diagnosis, prognosis, treatment response monitoring, and tumor biology. Radiomics has great potential in cancer clinical research, aiding in diagnosis, differentiation, disease stage assessment, therapy response prediction, and improving prognostic insights [8,9].

### *Disease Prediction*

By analyzing large-scale datasets from biobanks, AI models can identify predictive markers and develop risk assessment models to assist in personalized treatment planning [10].

### *Personalised Medicine*

Personalized medicine creates novel therapeutic, prognostic, and diagnostic approaches by fusing sophisticated genomic profiling with well-established clinical-pathologic factors [11].

### *Drug Discovery*

AI is revolutionizing oncology drug discovery by utilizing biobanks for repurposing medications and identifying targets. AI models evaluate genetic and molecular data, accelerating medication development and predicting efficacy and toxicity [12].

The following approaches that have been emphasized were identified to enhance the applications of the analysed AI potentials in LMICs**:**

(1). Integration with Electronic Health Records (EHRs)

Integrating AI with EHRs can enhance the analysis of patient data, improve disease diagnosis, and predict patient outcomes. This can be particularly useful in LMICs, where access to healthcare services is limited.

(2). Capacity Building
AI can be used to build capacity in biobanking, particularly in LMICs, by providing training and education on AI applications in biobanking.

(3). Quality Control and Assurance
AI can be used to monitor and improve the quality of biobanked samples and data, ensuring that they meet the required standards for research and clinical applications.

## 3. Challenges

Several difficulties of using AI in biobanking are also identified and addressed.

### *Data Quality and Completeness*

Real-world data, including clinical characteristics, genomes, treatments, and outcomes, is crucial for personalized medicine implementation [1]. However, obtaining this data from large patient populations can be challenging, especially for uncommon illnesses with diverse backgrounds [11]. AI algorithms may face challenges with incomplete or missing data, affecting their accuracy and introducing biases. To ensure data precision, biobanks need strong quality control procedures, standardized procedures for sample processing, data collection, and annotation, and standardized annotation protocols and ontologies to ensure uniform and significant annotation among diverse biobanks [13, 15].

### *Privacy and Security*

Biobanks house sensitive patient data, raising concerns about privacy and security. AI techniques like federated learning and secure multi-party computation are explored to secure patient data and ensure privacy, enabling collaborative analysis across biobanks [16].

### *Ethical Issues*

The use of AI in biobanking brings up additional ethical issues including informed consent, data ownership, and data sharing protocols. In order to undertake AI-driven analyses of biobank data in ways that respect patient privacy and relevant data protection legislations, researchers and institutions must make sure that these procedures are conducted with sufficient regard to ethical considerations [17].

## 4. Future Trends and Promising Artificial Intelligence Paradigms

### *Virtual or Federated Biobanking*

Virtual biobanking, a blockchain-based movement, uses artificial intelligence to create a decentralized network of biobanks, promoting large-scale partnerships, improved data accessibility, and cross-institutional oncology research [18].

### *Adaptive Models and Continuous Learning*

Biobanks are dynamic resources that continuously gather data, allowing AI models to adapt and enhance performance through feedback loops and reinforcement learning techniques [19].

### *Synthetic data*

Synthetic data, created by a model extracting key features from real data, ensures authenticity and privacy without violating current laws for data sharing [11].

***Explainable AI***

Explainable AI aims to enhance understanding of AI algorithms' decision-making processes, aiding in biobanking and clinical practice, fostering trust and acceptance of AI-driven technologies [20].

## 5. Conclusion

The digitalization of biobanks in LMICs holds enormous promise for advancing research, enhancing healthcare, and tackling global health issues. Regarding oncology, by utilizing existing and upcoming technical advancements applying AI trends and techniques, LMICs can get beyond obstacles and fully utilize the potential of biobanks to advance scientific understanding of cancer and its determinants and enhance patient outcomes. To unlock such enormous potentials, approaches like improving scalability and computational infrastructure, data standardization and integration across biobanks, long-term data management, interdisciplinary and multidimensional collaboration must be emphasized.

## References


1. Frascarelli, C., Bonizzi, G., Musico, C. R., Mane, E., Cassi, C., Guerini Rocco, E., Farina, A., Scarpa, A., Lawlor, R., Reggiani Bonetti, L., Caramaschi, S., Eccher, A., Marletta, S., & Fusco, N. (2023). Revolutionizing Cancer Research: The Impact of Artificial Intelligence in Digital Biobanking. Journal of Personalized Medicine, 13(9), 1390. https://doi.org/10.3390/jpm13091390
2. Bonizzi, G., Zattoni, L., Fusco, N. (2022). Biobanking in the digital pathology era. Oncol Res. 2022 Aug 31; 29(4):229-233. Doi: 10.32604/or.2022.024892. PMID: 37303941; PMCID: PMC10207967.
3. Bhinder, B., Gilvary, C., Madhukar, N.S., Elemento, O. (2021). Artificial Intelligence in Cancer Research and Precision Medicine. Cancer Discov. 2021 Apr; 11(4):900-915. Doi: 10.1158/2159-8290.CD-21-0090. PMID: 33811123; PMCID: PMC8034385.
4. Babu, M., Snyder, M. (2023). Multi-Omics Profiling for Health. Molecular & Cellular Proteomics 22, 100561. https://doi.org/10.1016/j.mcpro.2023.100561
5. He, X., Liu, X., Zuo, F., Shi, H., & Jing, J. (2022). Artificial intelligence-based multi-omics analysis fuels cancer precision medicine. Seminars in Cancer Biology, 88, 187-200. https://doi.org/10.1016/j.semcancer.2022.12.009
6. Najjar, R. (2023). Redefining Radiology: A Review of Artificial Intelligence Integration in Medical Imaging. Diagnostics. 2023; 13(17):2760 https://doi.org/10.3390/diagnostics13172760
7. Holzinger, A., Keiblinger, K., Holub, P., Zatloukal, K., & Müller, H. (2023). AI for life: Trends in artificial intelligence for biotechnology. New Biotechnology, 74, 16-24. https://doi.org/10.1016/j.nbt.2023.02.001
8. Rana, M., Bhushan, M. (2023). Machine learning and deep learning approach for medical image analysis: diagnosis to detection. Multimed Tools Appl 82, 26731–26769. https://doi.org/10.1007/s11042-022-14305-w
9. Pan, F., Feng, L., Liu, B., Hu, Y., Wang, Q. (2023). Application of radiomics in diagnosis and treatment of lung cancer. Front Pharmacol. 2023 Nov 1; 14:1295511. Doi: 10.3389/fphar.2023.1295511. PMID: 38027000; PMCID: PMC10646419.
10. Battineni, G., Hossain, M.A., Chintalapudi, N., Amenta, F. (2022). A Survey on the Role of Artificial Intelligence in Biobanking Studies: A Systematic Review. Diagnostics (Basel). 2022 May 9; 12(5):1179. Doi: 10.3390/diagnostics12051179. PMID: 35626333; PMCID: PMC9140088.
11. D'Amico, S., Dall'Olio, D., Sala, C., Dall'Olio, L., Sauta, E., Zampini, M., Asti, G., Lanino, L., Maggioni, G., Campagna, A., Ubezio, M., Russo, A., Bicchieri, M.E., Riva, E., Tentori, C.A., Travaglino, E., Morandini, P., Savevski, V., Santoro, A., Prada-Luengo, I., Krogh, A., Santini, V., Kordasti, S., Platzbecker, U., Diez-Campelo, M., Fenaux, P., Haferlach, T., Castellani, G., Della Porta, M.G. (2023). Synthetic Data Generation by Artificial Intelligence to Accelerate Research and Precision Medicine in Hematology. JCO Clin Cancer Inform. 2023 Jun;7:e2300021. Doi: 10.1200/CCI.23.00021. PMID: 37390377; PMCID: PMC10569771.
12. Paul, D., Sanap, G., Shenoy, S., Kalyane, D., Kalia, K., & Tekade, R. K. (2021). Artificial intelligence in drug discovery and development. Drug discovery today, 26(1), 80–93. https://doi.org/10.1016/j.drudis.2020.10.010

13. Emmanuel, T., Maupong, T., Mpoeleng, D. et al. A survey on missing data in machine learning. J Big Data 8, 140 (2021). https://doi.org/10.1186/s40537-021-00516-9
14. Annaratone, L., De Palma, G., Bonizzi, G., Sapino, A., Botti, G., Berrino, E., Mannelli, C., Arcella, P., Di Martino, S., Steffan, A., Daidone, M. G., Canzonieri, V., Parodi, B., Paradiso, A. V., Barberis, M., Marchiò, C., & Alleanza Contro il Cancro (ACC) Pathology and Biobanking Working Group (2021). Basic principles of biobanking: from biological samples to precision medicine for patients. Virchows Archiv : an international journal of pathology, 479(2), 233–246. https://doi.org/10.1007/s00428-021-03151-0
15. Reer, A., Wiebe, A., Wang, X. and Rieger, J.W. (2023), FAIR human neuroscientific data sharing to advance AI driven research and applications: Legal frameworks and missing metadata standards.Front.Genet.14:1086802.
16. Heatherly, R. (2016) Privacy and Security within Biobanking: The Role of Information Technology. J Law Med Ethics. 2016 Mar;44(1):156-60. Doi: 10.1177/1073110516644206. PMID: 27256131.
17. Kargl, M., Plass, M., Müller, H. (2022). A Literature Review on Ethics for AI in Biomedical Research and Biobanking. Yearb Med Inform. 2022 Aug;31(1):152-160. Doi: 10.1055/s-0042-1742516. Epub 2022 Dec 4. PMID: 36463873; PMCID: PMC9719772.
18. Cremonesi, F., Vesin, M., Cansiz, S., Bouillard, Y., Balelli, I., Innocenti, L., & Lorenzi, M. (2025). Fed-biomed: Open, transparent and trusted federated learning for real-world healthcare applications. In Federated Learning Systems: Towards Privacy-Preserving Distributed AI (pp. 19-41). Cham: Springer Nature Switzerland.
19. Bajwa, J., Munir, U., Nori, A., & Williams, B. (2021). Artificial intelligence in healthcare: transforming the practice of medicine. Future healthcare journal, 8(2), e188–e194. https://doi.org/10.7861/fhj.2021-0095
20. Holzinger, A., Langs, G., Denk, H., Zatloukal, K., & Müller, H. (2019). Causability and explainability of artificial intelligence in medicine. Wiley interdisciplinary reviews. Data mining and knowledge discovery, 9(4), e1312. https://doi.org/10.1002/widm.1312

# Using Surveillance Technologies to Prevent Suicides in Public Spaces: Ethical, Legal, Acceptable?

*[1]Dr Jay-Marie Mackenzie, [2]Dr Ian Marsh, [1]Dr Beth Cliffe, [3]Dr Laura Joyner, [3]Dr Carlisle George, [4]Mr Alex Dark, [4]Dr Robin Pharoah[4], [3]Professor Lisa Marzano*

[1] University of Westminster, London
J.C.Mackenzie@westminster.ac.uk; B.Cliffe@westminster.ac.uk

[2] Canterbury Christchurch University
ian.marsh@canterbury.ac.uk

[3] Middlesex University, London
L.Joyner@mdx.ac.uk; c.george@mdx.ac.uk; L.Marzano@mdx.ac.uk

[4] Encounter
alex@thoughtscapes.co.uk; robin.pharoah@encounter.team

Suicide is a global health problem, so understanding what types of prevention and interventions are effective in reducing suicides is essential. The potential for technologies to help reduce suicides and increase interventions in public spaces is a rapidly expanding field (e.g. closed-circuit television (CCTV)) being activated when a person crosses a dangerous boundary). Yet, there is little evidence as to whether suicide surveillance technologies work in the way they are supposed to. Although surveillance in public spaces is not new, there are distinct challenges related to suicide prevention: for example, if smart surveillance identifies a person who may be suicidal, what happens next, who (if anyone) responds to them, and do potential responders have appropriate training? An added challenge is that some technologies are controversial (e.g. those that could be used to track people's movements). Hence, understanding their ethics, legality and acceptability are important. However, to date this is an under-explored topic. This position paper situates some of the ethical, legal and acceptability issues which specifically relate to the use of surveillance technologies in the field of suicide prevention.

**Keywords:** Suicide Prevention, Smart Surveillance Technologies, Suicide Surveillance Technologies.

## 1. Suicide Prevention in Public Spaces

Suicide is a global public health issue, with the World Health Organization estimating that 720,000 suicides occur worldwide each year [1]. In the United Kingdom (UK), approximately 6,000 suicides are recorded annually [2], of which around a third occur in public places [3]. These deaths have devastating consequences for families, potential witnesses, staff dealing with their aftermath [4], and substantial financial implications for affected industries: the cost of suicides to the British Rail Industry is estimated at over £70 million annually [5]. In addition, suicides in public locations can attract considerable media attention [6], potentially introducing a degree of notoriety about specific locations, and thus increasing the risk of imitative behaviours and suicide clusters [6].

Efforts to prevent suicides in public places are therefore high on the agenda of national prevention strategies. This is particularly the case at 'high risk locations' (i.e., public locations associated with a high number of suicide-related incidents and/or which offer access to highly lethal means of suicide). In England, the National Institute for Health and Care Excellence (NICE) recommends that suicide prevention measures are vital at high-risk locations [7], and – by 2019 – over 60% of local authorities delivered interventions at such locations

[8]. These measures include restriction of access to means, promotion of responsible media reporting, encouragement of help-seeking and intervention by a third-party [3] – and, increasingly, the placing of surveillance technologies to detect, deter and disrupt suicidal behaviour. These may then initiate a police response when a sensor or camera detects someone on a parapet, or trigger a pre-recorded announcement encouraging the person to move to a safe location and/or or seek help [9-11].

### 1.1 The Role of Smart Surveillance Technologies in Suicide Prevention

Going beyond the role of standard CCTV, smart surveillance technology (SST) systems enhance or fully automate the surveillance process by extracting and interpreting information received by devices and/or sensors [12]. The use of such systems involves an input/data source (e.g., CCTV cameras, thermal imaging cameras, proximity sensors), data processing/analytics (e.g., motion/intrusion/linger detection, behaviour prediction, facial recognition), and a response/data use (e.g., audio alert to person at risk and/or responders).

Examples of these technologies are currently deployed in a variety of contexts, often covertly. They include automated camera systems at metro stations to alert staff when human movement is detected on the tracks [10]; 'virtual alarmed fencing' in the CCTV systems used at coastal locations [13]; monitoring systems to detect when someone has jumped from a bridge [9] or to identify behaviours that might precede a suicide [13]. The emerging evidence suggests promising results for accurate surveillance and detection of suicidal behaviour via automated CCTV monitoring and/or minimal staff training [14, 15, 16]. However, there is currently insufficient evidence to conclude that these technologies can safely and effectively reduce suicides – an important gap which our research is seeking to address (https://fundingawards.nihr.ac.uk/award/NIHR151521). For instance, some of these systems may only identify a suicide attempt immediately before, during or after it occurs, limiting possible intervention. The acceptability and financial viability of using smart surveillance for suicide prevention is also poorly understood; so too is the potential for the technologies to cause adverse effects (e.g. displacement of suicidal behaviours to other locations, or the possibility of an unexpected auditory alarm exacerbating or hastening suicidal behaviour). Further considerations concern ethical and legal principles such as transparency, consent, privacy and data protection, accessibility, and accountability [17].

## 2. Ethics, Legality and Acceptability; The Unknowns

In the field of suicide prevention, the focus is often on whether interventions are effective – and cost-effective – in reducing suicides. However, what is sometimes neglected is whether or not interventions are considered ethical and acceptable to the people who experience them [18]. Furthermore, and of especial importance in the case of surveillance technology, is how members of the public who are also using these locations feel about these types of technology, and what legal and ethical implications there are for the various actors operating in and around these locations (e.g. frontline staff/emergency first responders, CCTV control room operators, land owners and public health leads). There are already discussions around the ethics of artificial intelligence (AI)-based surveillance technology for crime prevention [19, 20]. In addition, although suicide remains a crime in some countries it is not a crime in the majority of countries in the world, and hence people should not be criminalised as a result of their suicidal behaviour.

Therefore, it is essential that the ethics, legality, and acceptability of such technologies are considered through the lens of suicide prevention and not just through a crime reduction lens. As Marks et al [17] state; "*Though AI-based suicide prediction has the potential to improve our understanding of suicide while saving lives, it raises many risks that have been underexplored. The risks include stigmatisation of people with mental illness, the transfer of sensitive personal data to third-parties such as advertisers and data brokers, unnecessary involuntary confinement, violent confrontations with police, exacerbation of mental health conditions, and paradoxical increases in suicide risk*" .

As part of our current wider project in which we are evaluating a range of SSTs being used in public spaces, we are exploring issues relating to the ethics, legality and acceptability of SST implementation. Examples of issues we have come across so far include:

- To what extent are the personal rights and freedoms of the public affected by the use of suicide prevention surveillance technologies (e.g., effects on the right to privacy)?
- If an SST were able to detect a person in distress, what intervention would the person receive and who would be adequately trained to deliver it?
- What does the introduction of technology mean for human interventions? Would they decrease human presence or could an SST increase the chance of human intervention?
- How can we ensure that an SST-assisted intervention is safe, acceptable and proportionate?
- If technology is removed for some reason in the future, what could the implications and unintended consequences of this be?
- What liability arises with implementing these technologies (examples include: having the duty to intervene, introducing staff training, making sure that the technology works well, mitigating bias/discrimination)?
- What about unintended consequences of implementing the technology – could people move to spaces where there is less chance of intervention or could they change their behaviour to avoid detection by an SST (e.g. bypassing a camera if it is visible, or by leaving their phone at home to avoid being tracked)?
- If potential responders are able to detect a suicidal person, do they have enough time to respond?
- Could such systems unintentionally create stigma? Could this stigma lead to criminalisation?
- Are there any risks of bias / discrimination being introduced? What could this discrimination mean for people who fit certain profiles?
- In terms of equity and accessibility, will some groups of people not be detected by SSTs?
- Data processing / sharing – who can data be shared with and what is the legal basis for sharing? For how long can personal identifiable data be stored? What kinds of personal data are being processed?

During the workshop we will share preliminary findings from our project, and encourage a wider discussion about the ethics, accessibility and legality of this topic.

**Funding:** This project is funded by the National Institute for Health Research (NIHR) PHR programme NIHR151521. The views expressed are those of the authors and not necessarily those of the NIHR or the Department of Health and Social Care.

**Acknowledgments:** We would firstly like to thank the group of NSPA (National Suicide Prevention Alliance) Lived Experience Influencers who formed an advisory group which continues to provide invaluable feedback across this project. We are also grateful for the advice and support from our colleagues from the wider project team (consisting of Professor Peter Craig, Anthony Purvis, Dr Manuela Deidda, Dr Philip Worrall, Dr Giorgio Ciminata and Amina Mahmood) and steering group (consisting of Professor Rachel Aldred, Professor Amy Chandler, Professor Frank De Vocht, Professor Keith Hawton, Fiona Malpass, Professor Emma McIntosh and Dr Hilary Norman). 

**References**


1. World Health Organization. Suicide 2021 [cited 2022 15/08]. Available from: https://www.who.int/news-room/fact-sheets/detail/suicide.
2. Office of National Statistics. Suicides in England and Wales: 2023 registrations 2023 [cited 2025 15/08]. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/birthsdeathsandmarriages/deaths/bulletins/suicidesintheunitedkingdom/2023
3. Frost, P. Persons who die by suicide outside their home address. A statistical analysis and exploration of the relationship between age, gender and ethnicity on location type, method of death and distance travelled. Royal College of Policing, 2020, https://library.college.police.uk/docs/theses/FROST_Suicide_prevention_outside_the_home_2020.pdf
4. Giupponi, G., H. Thoma, D. Lamis, A. Forte, M. Pompili, and H. P. Kapfhammer. Posttraumatic Stress Reactions of Underground Drivers After Suicides by Jumping to Arriving Trains; Feasibility of an Early Stepped Care Outpatient Intervention. *Journal of Trauma & Dissociation* 2019; 20 (5): 495–510. https:// doi. org/ 10. 1080/ 15299 732. 2019.1597810.
5. Tyers, R., Stewart, I. Suicide prevention: Transport House of Commons Library, Research Briefing. 2024
6. Marzano, L., Spence, R., Marsh, I., Barbin, A., & Kruger, I. How does news coverage of suicide affect suicidal behaviour at a high-frequency location? A 7-year time series analysis. *BMJ Public Health* 2025; 3,(1) .
7. Jones-Berry S. NICE guidance on preventing suicide in community and custodial settings. 2018.
8. Chadwick T, Owens C, Morrisey J. Local Suicide Prevention Planning in England: An Independent Progress. 2019.
9. Lee J, Lee C-M, Park N-K. Application of sensor network system to prevent suicide from the bridge. Multimedia Tools and Applications. 2016;75(22):14557-68.
10. Zhang T, Aftab W, Mihaylova L, Langran-Wheeler C, Rigby S, Fletcher D, et al. Recent Advances in Video Analytics for Rail Network Surveillance for Security, Trespass and Suicide Prevention—A Survey. Sensors. 2022;22(12):4324.
11. Bouachir W, Gouiaa R, Li B, Noumeir R. Intelligent video surveillance for real-time detection of suicide attempts. Pattern Recognition Letters. 2018;110:1-7.
12. Langheinrich, M., Finn, R., Coroama, V. and Wright, D., Quo vadis smart surveillance? How smart technologies combine and challenge democratic oversight. In *Reloading Data Protection: Multidisciplinary Insights and Contemporary Challenges* 2013. (pp. 151-182). Dordrecht: Springer Netherlands.
13. Onie S, Li X, Liang M, Sowmya A, Larsen ME. The Use of Closed-Circuit Television and Video in Suicide Prevention: Narrative Review and Future Directions. JMIR Ment Health. 2021;8(5):e27663.
14. Mishara BL, Bardon C, Dupont S. Can CCTV identify people in public transit stations who are at risk of attempting suicide? An analysis of CCTV video recordings of attempters and a comparative investigation. BMC Public Health. 2016;16(1):1245.
15. Li X, Onie S, Liang M, Larsen M, Sowmya A. Towards Building a Visual Behaviour Analysis Pipeline for Suicide Detection and Prevention. Sensors. 2022;22(12):4488. PubMed PMID: doi:10.3390/s22124488.
16. Liang M, Li X, Onie S, Larsen M, Sowmya A, editors. Improved Spatio-temporal Action Localization for Surveillance Videos. 2021 International Conference on Digita; Image Computing: Techoqies and Applications (DICTA); 2021; Gold Coast, Australia.
17. Marks, M., 2019. Artificial intelligence-based suicide prediction. *Yale JL & Tech.*, *21*, p.98.
18. Mackenzie JM, Marsh I, Fields B, Kruger I, Katsampa D, Crivatu I, Marzano L. Preventing suicides on the railways: learning from lived and living experiences. BMC Public Health. 2025 May 2;25(1):1625.
19. Ţical, G.M., 2024. Artificial intelligence technologies: a new era for crime prevention. *Annals–Series on Military Sciences*, *16*(2), pp.115-128.
20. Blount, K., 2024. Using artificial intelligence to prevent crime: implications for due process and criminal justice. *AI & Society*, *39*(1), pp.359-368.

# Addressing Privacy and Security Challenges in Enabling Interoperability of Electronic Health Records: The TASIPPS Framework

*Mr Adetunji Ademola*
ALERT Research Group, Middlesex University, London UK
AA4802@live.mdx.ac.uk

## 1. Introduction

Interoperability has become a cornerstone of modern information systems, especially in healthcare, where seamless data exchange can lead to better patient outcomes, reduced costs, and improved system efficiency [1]. Interoperability of Electronic Health Records (EHRs) enables healthcare systems to share patient data across platforms, enhancing care coordination and reducing duplication [2]. Interoperability involves various stakeholders—hospitals, laboratories, insurers, and patients - all interacting through digital platforms that are often web-enabled [3]. However, with increased data sharing comes heightened concerns over privacy and security [4]. The sensitive nature of medical data demands robust privacy protections and security protocols [5].

## 2. Need for Privacy and Security in Interoperability

Technical barriers such as legacy systems, inconsistent standards, and different organizational policies have hindered the achievement of interoperability. Data integrity during transmission and the secure sharing of information, without violating privacy, remain major concerns. Ensuring interoperability thus requires a holistic approach balancing technical capability with stringent security and privacy requirements.

Medical data is among the most sensitive categories of personal information, encompassing diagnoses, treatment histories, genetic profiles, and mental health records. A breach of such data can result in discrimination, stigma, or identity theft [6]. Unlike financial data, which can often be replaced (e.g., credit cards), medical data is inherently personal and permanent. This makes privacy protections not just a legal requirement (under the General Data Protection Regulation (GDPR) incorporated in the UK Data Protection Act 2018) but an ethical necessity [7]. Data privacy, which involves the proper handling, processing, and storage of personal information [8] is necessary to ensure that individuals' privacy rights are protected. It is essential for interoperable systems to ensure data privacy, among other measures, by minimizing data collection and sharing only what is necessary [9].

Effective implementation of privacy measures begins with robust consent mechanisms where patients must provide informed, explicit consent before their data is processed or shared [10]. The principle of data minimization ensures that only essential information is collected [11], which is vital in the preservation of privacy to ensure minimal information collection on patients. In relation to security, encryption of data while the data is both ‘at rest’ and in transit is critical. Multi-factor authentication (MFA) and role-based access control (RBAC) help prevent unauthorized access [12]. Access control which is a security mechanism that is fundamental in the preservation of security, is typically based on user roles and it further helps restrict data access to authorized personnel [13]. Monitoring and logging user activity support early detection of security threats which is crucial in the fight against system breaches [14]. Together, these strategies form a foundation for ethical and legally compliant data exchange in interoperable systems [15].

Web technologies, such as application programming interfaces (APIs) and cloud platforms, have advanced healthcare interoperability by allowing real-time data exchange [16]. Standards like Fast Healthcare Interoperability Resources (FHIR) facilitate structured sharing, yet they also introduce new vulnerabilities [2]. Risks like cross-site scripting (XSS), insecure APIs, and man-in-the-middle (MITM) attacks threaten data integrity during the implementation of interoperability [17]. Secure communication protocols (e.g., HTTPS), authentication methods like OAuth 2.0, and regular security audits are absolutely necessary countermeasures.

## 3. Addressing Privacy and Security in Interoperability: The TASIPPS Framework

The author proposed the Technical and Semantic Interoperability, Preserving Privacy and Security (TASIPPS) framework. It offers a comprehensive solution for addressing privacy and security issues in interoperable EHR systems [18] as seen in Figure 1.

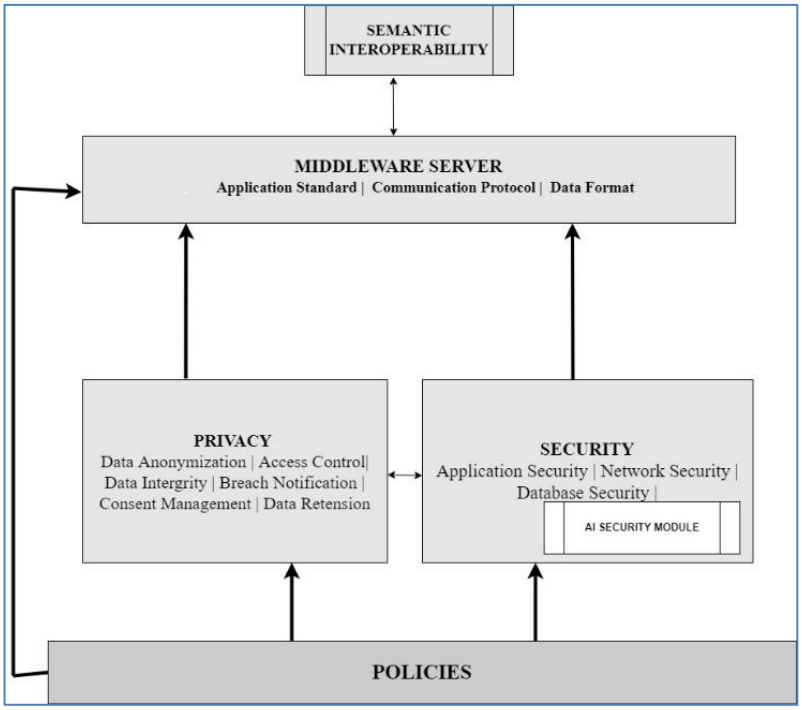



Figure 1- The TASIPPS Framework

The TASIPPS framework consists of the following five modules:

- The *Middleware Server* Module: positioned as an intermediary between disparate electronic health record (EHR) systems, avoiding a central point of access that would be vulnerable to attacks.
- The *Semantic Interoperability* Module: ensures data flows securely between EHR systems without creating a single point of failure, enhancing both security and reliability.
- The *Privacy* Module: safeguards the confidentiality and individual rights of users' personal information.
- The *Security* Module: manages the security aspects of the framework which is critical in the sharing and exchanging of EHRs.
- The *Policy* Module: dedicated to governing and regulating various facets of system operations and data management.

The TASIPPS framework integrates both technical and semantic interoperability mechanisms with dedicated privacy and security modules. It utilizes advanced encryption (AES-256), hashing algorithms (SHA-512), and OAuth 2.0 for authentication. It also incorporates data masking, access control, and consent management to comply with data protection standards required by the UK GDPR. The TASIPPS framework not only ensures

accurate and meaningful data exchange but also empowers patients to have greater control over their data, making it suitable for those real-world healthcare settings where trust, security, and compliance are crucial.

The UK GDPR mandates "privacy by design and by default," which requires systems to embed privacy measures from the outset. The same legislation also states that data should not be stored longer than necessary and that it be secured against breaches. These requirements are implemented in the TASIPPS framework, which places a major emphasis on security and privacy preservation. The TASIPPS framework has been implemented in a working prototype to support testing and research aimed at validating its suitability and effectiveness in achieving interoperability.

## 4. Conclusion

In conclusion, privacy and security are integral to the success of interoperable healthcare systems. As data becomes increasingly digitized and shared across web-enabled platforms, the risk of data breaches and data misuse escalates. Effective privacy and security measures—ranging from technical safeguards like encryption and access control to legal compliance—are essential. This is especially the case with regard to healthcare data. The TASIPPS framework presents a compelling model that holistically addresses the complexities of interoperability while ensuring patient privacy and system security. Its alignment with legal mandates and integration of advanced technical features makes it a viable solution for the legal, ethical and secure exchange of health information in the digital age.

## References


1. Adegoke, K.; Adegoke, A.; Dawodu, D.; Bayowa, A.; Adekoya, A.(2025). Interoperability in Digital Healthcare: Enhancing Consumer Health and Transforming Care Systems. *Preprints* **2025**, 2025021774. https://doi.org/10.20944/preprints202502.1774.v1
2. Saini, V., Reddy, S.G., Kumar, D. and Ahmad, T., (2021). Evaluating FHIR's impact on Health Data Interoperability. *IoT and Edge Comp. J*, *1*(1), pp.28-63.
3. Mohammad-Noor, N., (2017). The adoption of ICT in Malaysian public hospitals: the interoperability of electronic health records and health information systems (Doctoral dissertation, Loughborough University).
4. Lockhart, E.N.S.(2025). Privacy and Data Security in Digital Health: Implications for Family Therapy. Contemp Fam Ther (2025). https://doi.org/10.1007/s10591-024-09731-x
5. Jin, H., Luo, Y., Li, P. and Mathew, J., (2019). A review of secure and privacy-preserving medical data sharing. IEEE access, 7, pp.61656-61669.
6. Kumar, K.P., Prathap, B.R., Thiruthuvanathan, M.M., Murthy, H. and Pillai, V.J., (2024). Secure approach to sharing digitized medical data in a cloud environment. Data Science and Management, 7(2), pp.108-118.
7. Renuka, O., RadhaKrishnan, N., Priya, B.S., Jhansy, A. and Ezekiel, S., (2025). Data Privacy and Protection: Legal and Ethical Challenges. *Emerging Threats and Countermeasures in Cybersecurity*, pp.433-465.
8. Nowrozy, R., Ahmed, K., Kayes, A.S.M., Wang, H. and McIntosh, T.R., (2024). Privacy preservation of electronic health records in the modern era: A systematic survey. *ACM Computing Surveys*, *56*(8), pp.1-37.
9. Tang, X., Shi, L., Wang, X., Charbonnet, K., Tang, S. and Sun, S., (2024). Zero-knowledge proof vulnerability analysis and security auditing. *Cryptology ePrint Archive*.
10. Verreydt, S., Yskout, K. and Joosen, W., (2021). Security and privacy requirements for electronic consent: a systematic literature review. ACM Transactions on Computing for Healthcare, 2(2), pp.1-24
11. Shojaei, P., Vlahu-Gjorgievska, E. and Chow, Y.W., (2024). Security and privacy of technologies in health information systems: A systematic literature review. *Computers*, *13*(2), p.41.
12. Swanzy, P.N., Abukari, A.M. and Ansong, E.D., (2024). Data security framework for protecting data in transit and data at rest in the cloud. *Current Journal of Applied Science and Technology*, *43*(6), pp.61-77.

13. Sookhak, M., Jabbarpour, M.R., Safa, N.S. and Yu, F.R., (2021). Blockchain and smart contract for access control in healthcare: A survey, issues and challenges, and open issues. *Journal of Network and Computer Applications*, *178*, p.102950.
14. Kanaparthi, Y., Abdellatif, T.M., Seyam, A.A. and Satrya, G.B., (2024). November. Identifying Security Threats in the System Using Automated Security Logs. In *2024 Third International Conference on Sustainable Mobility Applications, Renewables and Technology (SMART)* (pp. 1-5). IEEE.
15. Nookala, G., Gade, K.R., Dulam, N. and Thumburu, S.K.R., 2024. Governance for Data Ecosystems: Managing Compliance, Privacy, and Interoperability. MZ Journal of Artificial Intelligence, 1(2).
16. Chitta, S., Sadhu, A.K.R., Gudala, L. and Reddy, S.G., (2022). Unlocking Health Data: API-Driven Solutions for Interoperability Challenges. *Journal of Informatics Education and Research*, *2*, p.45.
17. Asadi, F., (2025). A Comparative Analysis of Subresource Integrity (SRI) vs. Content Security Policy (CSP) for Resource Integrity Verification: A Comparative Study of SRI and CSP Effectiveness.
18. Ademola, A.; George, C.; Mapp, G (2024). Addressing the Interoperability of Electronic Health Records: The Technical and Semantic Interoperability, Preserving Privacy and Security Framework. *Appl. Syst. Innov.* **2024**, *7*, 116. https://doi.org/10.3390/asi7060116

# Revisiting the Lucy Letby Case: A Forensic Readiness Perspective for Healthcare Big Data Networks

*Mr Cephas Mpungu*
ALERT Research Group, Middlesex University, UK
CM1677@live.mdx.ac.uk

**Abstract**
Healthcare systems today generate vast volumes of structured, semi-structured, and unstructured data across distributed networks. While this data is essential for delivering patient care, it also presents a valuable opportunity to support digital forensic investigations when adverse incidents occur. This paper presents a case for proactive Digital Forensics Readiness (DFR) within healthcare big data environments, using the Lucy Letby case as a detailed and contextual illustration. This incident represents one of the most serious recent healthcare-related criminal cases in the United Kingdom. Building on the author's previously published decentralised DFR framework, this paper examines how the framework's Hadoop-based storage architecture, when integrated with User and Entity Behaviour Analytics (UEBA), can significantly improve forensic readiness. The paper focuses on three key areas. The first is the alignment of DFR implementation with data governance legislation and secure logging policies. The second is the application of forensic principles, specifically the identification, collection, and preservation of data across a wide range of healthcare information sources. The third demonstrates how the proposed framework supports incident detection, investigation, and evidence preservation, using a systems-level analysis of the Lucy Letby case.

**Keywords:** big data, digital forensics, digital forensics readiness, healthcare.

## 1. Introduction

The increasing volume and complexity of data in healthcare, particularly within big data networks, have introduced significant challenges in governance, security, and accountability [1]. According to [2], hospitals and clinical environments generate large volumes of data daily, including logs from electronic health records (EHRs), access control systems, closed-circuit television (CCTV) footage, Internet of Medical Things (IoMT) devices, and other structured and unstructured digital communication. While this information is critical to clinical care and effective network administration, it also holds substantial evidential value in cases involving malpractice, insider threats, and criminal activity.

In this context, big data encompasses not only sheer volume but also the crucial aspect of variety, spanning structured, semi-structured (e.g., logs, XML), and unstructured formats (e.g., video, free-text notes). Harnessing this diverse range of data is paramount for digital forensics, as critical contextual clues often reside within the less conventional, unstructured datasets, enabling more comprehensive investigations.

The Lucy Letby case, in which a neonatal nurse was convicted of multiple infant murders at the Countess of Chester Hospital, tragically revealed shortcomings in the hospital's ability to detect and respond to behavioural anomalies [3]. An effective big data Digital Forensics Readiness (DFR) framework, which facilitates the detection and proactive collection of critical evidence to facilitate digital investigations, could have possibly prevented further harm to some of the babies involved.

This paper builds upon a previously published DFR framework developed for big data wireless medical networks [4]. This framework was specifically chosen for its robust capabilities in handling diverse big data sources and detecting anomalous behaviours, making it highly pertinent to addressing the systemic failures identified in the Lucy Letby case. The framework is designed to capture and manage structured, semi-structured, and unstructured evidential data (for digital forensics investigations) using Hadoop Distributed File Systems (HDFS). The system is designed to store very large data sets, and runs on commodity hardware. It integrates a User and Entity Behaviour Analytics (UEBA) system for continuous monitoring and employs

decentralised, tamper-resistant storage to support evidential data availability, scalability, and integrity throughout the forensic process.

This work presents three key contributions. First, it examines how DFR stages were implemented within the framework in accordance with data protection and retention laws, including UK General Data Protection Regulation (GDPR) [5] and National Health Service (NHS) records management policies. Second, it discusses the application of core forensic principles of identification, collection, and preservation of data and metadata from diverse sources, ensuring adherence to legal standards for admissibility. Third, the utility of the proposed framework is demonstrated through a systems-level walkthrough of the Lucy Letby case, illustrating how the framework supports incident detection, investigation, and evidence preservation.

Ultimately, this paper illustrates how healthcare organisations can strengthen their operational resilience, regulatory compliance, incident response capabilities, and business continuity by embedding forensic readiness as a foundational element of their big data infrastructures.

The remainder of this paper begins with an analysis of legal and compliance considerations in Section 2, establishing the governance framework for digital forensics readiness in healthcare. Section 3 then details the application of digital forensic principles within the proposed framework and demonstrates its utility through a systems-level walkthrough of the Lucy Letby case. The paper concludes in Section 4 with key insights and future considerations for healthcare forensic preparedness.

## 2. Legal and Governance Context

To ensure legal admissibility and regulatory compliance, the proposed DFR framework in [4] is architected in accordance with key legislation and standards governing data handling, privacy, and forensic integrity within healthcare environments.

The UK GDPR incorporated in the Data Protection Act (DPA) 2018 forms the core legislative foundation, particularly GDPR Articles 5 and 32, which mandate lawful processing, data minimisation, secure storage, and appropriate retention of personal and sensitive data. This compliance is operationalised in the framework during the collection and logging of structured, semi-structured, and unstructured data from various hospital systems, including network traffic logs, keycard login data, electronic health records (EHRs), radiography reports, and unstructured data such as CCTV footage and clinical notes.

Table 1 provides a summary of how specific components of the framework map to relevant standards, laws, and best practices.

Table 1. Compliance Mapping of the DFR Framework to Legal, Regulatory, and Technical Standards [4].

| **Framework Component** | **Compliance Objective** | **Mapped Law / Standard** |
|---|---|---|
| *Security Information and Event Management* (SIEM) and *Intrusion Detection System* (IDS) for structured log aggregation | Real-time log correlation and security incident detection | ISO/IEC 27001:2022, [6]<br><br>ISO/IEC 27035 [7] |
| Big data DFR (BdDFR) Hadoop-based DFR1/DFR2 for decentralised storage | Tamper-resistant, decentralised, scalable forensic data storage | ISO/IEC 27043:2015 [8] |
| *Transport Layer Security* (TLS) and filesystem-level encryption via *Red Hat Enterprise Linux* (RHEL 9) | Protection of data in transit and at rest | NHS Digital Security Standards [9]<br><br>UK GDPR (Article 32 ) [5] |

| | | |
|---|---|---|
| Pseudonymisation and anonymisation techniques | Privacy protection for PII and special category data | UK GDPR Recital 26, Article 89 [5];<br><br>NHS Records Management Code of Practice [10] |
| Retention policy automation using RHEL log rotation. | Data lifecycle management and legal hold enforcement | NHS data retention policies [10] |
| *Identity and Access Management* (IAM) with blockchain-based *Multifactor Authentication* (MFA) | Enforces secure and auditable access to evidentiary logs | ISO/IEC 27002 [11] |
| *User and Entity Behaviour Analytics* (UEBA) integration | Insider threat detection and anomaly-based monitoring | NIST SP 800-53 Rev. 5 (AU-6: Audit Review)[12] |

This framework ensures that every forensic action, from data collection and logging to investigation and reporting, is executed in a secure and compliant manner. It also reinforces the admissibility of evidence and supports internal and external audits, ultimately enhancing the accountability and resilience of the healthcare environment.

In Section 3, the framework is applied to a real-world scenario in line with digital forensics principles, demonstrating how it could have supported alerting, detection, and investigation in the Lucy Letby case.

## 3. Application of Digital Forensics Principles and Case Demonstration

The effectiveness of a robust DFR framework hinges on its alignment with core forensic principles: identification, collection, and preservation of potential evidential data. These principles are operationalised through various components of the proposed framework, illustrated in Figure 1. Lucy Letby is referred to simply as Lucy in Figure 1.

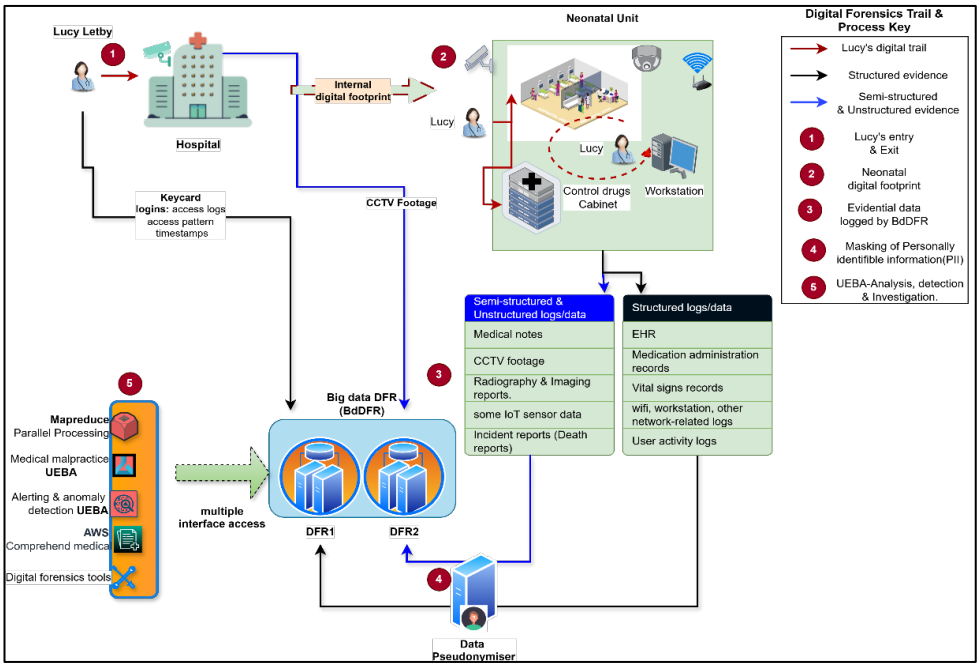


Fig. 1. Early detection of Lucy Letby's harm to babies using the DFR Framework [4]

Figure 1 presents a five-stage architecture aligned with the *Digital Forensics Trail & Process Key,* illustrating the sequential progression of identification, collection, and preservation of digital evidence. Stage 1 and 2 track Lucy's digital trail (red arrows) through the hospital and neonatal activity, logging access and internal movement. Stage 3 collects structured, semi-structured, and unstructured data (black and blue arrows) into a decentralised big data forensics storage system (BdDFR). Stage 4 handles pseudonymisation of personally identifiable information (PII). Finally, Stage 5 enables forensic analysis and detection using multiple tools and interfaces, including UEBA, MapReduce, and anomaly detection.

The identification of potential evidential data is enabled by integrating Security Information and Event Management (SIEM) storage, as discussed in [4] with UEBA. This integration continuously monitors behavioural patterns across diverse sources, including EHRs, access control logs, medication systems, and IoMT devices. Such a setup facilitates the early detection of anomalies like unusual access times, abnormal system usage, or behavioural deviations from baseline activity.

Collection of data is performed through structured, semi-structured, and unstructured data streams. Structured data from firewalls, IDS, EHR systems, and SIEM platforms is forwarded to DFR1 nodes within HDFS, while semi-structured logs (e.g., XML-based system messages) and unstructured data (e.g., CCTV footage, medical notes, radiography reports) are captured and routed to DFR2 nodes. This collection process ensures time-synchronised logging and cryptographic hashing, thereby preserving the integrity and traceability of each log entry. Data collection also includes automated log rotation and retention control, ensuring that logs are available for both short-term alerts and long-term investigation.

Preservation is enforced through decentralised storage and cryptographic protection. Role-based access control, blockchain-secured multi-factor authentication, and filesystem encryption are implemented to restrict access to the data by authorised forensic analysts and third parties. Additionally, anonymisation and pseudonymisation techniques are applied to datasets in compliance with UK GDPR and NHS Records Management guidelines, allowing investigators to conduct analysis without violating patient privacy.

Figure 1 specifically models how the framework would have operated in the context of the Lucy Letby case to enable early detection through behavioural analytics that would have linked the nurse's system activity with physical access data and shift schedules. The systematic logging of keycard access events, electronic health record interactions, and patient data activity, securely and immutably stored within the DFR1 and DFR2 nodes, would have enabled the reconstruction of a detailed forensic timeline. As illustrated in the figure, correlating these logs with CCTV footage and physical access records would have allowed investigators to verify whether her system access coincided with actual presence on-site. This cross-analysis could have exposed inconsistencies between Letby's assigned duties and her recorded activities, enabling earlier intervention and reducing the likelihood of further potential harm to any of the babies.

Furthermore, forensic readiness would have ensured that data necessary for internal disciplinary review or criminal prosecution was preserved and packaged in accordance with legal and evidentiary standards. Access to evidential logs would have been governed, auditable, and legally admissible, thereby avoiding the risk of evidential gaps due to data loss or non-compliance.

By embedding digital forensic principles directly into the design of healthcare big data environments, the proposed framework offers healthcare organisations the ability to proactively respond to threats, meet regulatory demands, and ensure transparency and accountability in clinical settings.

**4. Conclusion**

This paper has emphasised the critical role of proactive DFR in healthcare big data environments. By integrating a decentralised DFR framework with real-time monitoring and alerting tools such as UEBA, in line with relevant legislation and standards, healthcare organisations can strengthen their capacity to detect,

investigate, and preserve evidence related to adverse incidents. A systems-level analysis of the Lucy Letby case illustrated how such a framework, when aligned with legal obligations and forensic best practices, could have supported earlier detection of behavioural anomalies and enabled timely intervention, thereby reducing the potential for further harm to the vulnerable infants in the hospital concerned.

Future work will focus on empirical validation of the framework's performance metrics and explore its adaptability to emerging healthcare technologies and regulatory landscapes.

**References**


1. Price, W.N. and Cohen, I.G. (2019). Privacy in the Age of Medical Big Data. *Nature Medicine*, 25(1), pp.37–43. doi: https://doi.org/10.1038/s41591-018-0272-7
2. TrueNorthVictor. (2022). *Structured vs Unstructured Data in Healthcare (With Examples)*. TrueNorth ITG. Available at: https://www.truenorthitg.com/healthcare-unstructured-data/
3. Turner, D. (2024) 'The Lucy Letby case: What does it mean for the United Kingdom's healthcare system?', *QMUL School of Law Blog*. [Online] Available at: https://www.qmul.ac.uk/lac/our-legal-blog/blogs/the-lucy-letby-case-what-does-it-mean-for-the-united-kingdoms-healthcare-system.html
4. Mpungu, C., George, C. and Mapp, G., (2024). Enhancing Digital Forensics Readiness In Big Data Wireless Medical Networks: A Secure Decentralised Framework. *arXiv preprint arXiv:2408.00568*.
5. GDPR (2018). Data Protection Regulation (GDPR). Available at: https://gdpr-info.eu/art-32-gdpr/.
6. ISO (2022). *ISO/IEC 27001 standard – information security management systems*. [online] ISO. Available at: https://www.iso.org/standard/27001.
7. SecAware (2016). *ISO/IEC 27035 Security incident management*. [online] Iso27001security.com. Available at: https://www.iso27001security.com/html/27035.html.
8. ISO (2015). *ISO/IEC 27043:2015*. [online] ISO. Available at: https://www.iso.org/standard/44407.html
9. NHS Digital. (2025.). *Data Security and Protection Toolkit assessment guides*. [online] Available at: https://digital.nhs.uk/cyber-and-data-security/guidance-and-assurance/data-security-and-protection-toolkit-assessment-guides
10. NHS (2023). *Records Management Code of Practice*. [online] NHS Transformation Directorate. Available at: https://transform.england.nhs.uk/information-governance/guidance/records-management-code/.
11. IsecT (2013). *ISO/IEC 27002 code of practice*. [online] Iso27001security.com. Available at: https://www.iso27001security.com/html/27002.html.
12. National Institute of Standards and Technology (2020). Security and Privacy Controls for Information Systems and Organizations. *Security and Privacy Controls for Information Systems and Organizations*, [online] 5(5). doi:https://doi.org/10.6028/nist.sp.800-53r5.

# Personalised Care in Maternity addressing: Leveraging digital applications to address disparities considering Legal, Ethical, and Social Challenges

*[1]Dr Jasmine Leonce and [2]Professor Trixie McAree*

[1]National Maternity Speciality and Improvement Advisor NHSE
jasmine.leonce@nhs.net

[2]Senior Midwife, Chief Midwife's Office NHSE
trixie.mcaree2@nhs.net

**Abstract**

Personalised maternity care ensures service users have access to unbiased, evidence-based information in multiple languages and formats, empowering them to make informed decisions throughout pregnancy and birth. Digital apps play a crucial role in enhancing this care, this is enabled by the widespread use of smartphones. This is especially relevant as one of the three core drivers for change in the NHS is from analogue to digital. However, although these new approaches supplement care, their use raises legal, ethical, and social concerns, as highlighted in a meta-analysis by Paccoud et al. [1]. This is particularly relevant for women who are likely to experience the worst outcomes. For example, ethnic minority women often experience unconscious bias in maternity care [2,3], leading to less personalised support. Digital apps designed specifically for these communities may help address this issue by providing tailored guidance and resources.

This paper examines the key principles for deploying digital tools in personalised maternity care, explores the challenges they present, and proposes potential solutions to ensure equitable and effective implementation.

## 1. Introduction

Personalised maternity care empowers women to take control of decisions about their care, throughout their pregnancy and childbirth experience. Tailored to individual needs, such as religion, social deprivation, mental health, and ethnicity, this approach has been shown to reduce health inequalities, improve clinical outcomes and enhance experience of care [4]. Therefore, recurrent reports such as National Maternity Review (known as Better Births) [5] committed to providing maternity services that centre on women, their babies, and families, informed by unbiased information. Subsequent reports, including the NHS Providers Bold Action: Tackling Inequalities in Maternity Care [6] and NHS England's Three-year Maternity and Neonatal plan [7] reinforce that personalised care must align with safety, trust, and improved data systems to enhance accessibility and engagement.

Maternal mortality data in the United Kingdom (UK) show that Black women are three times more likely to die during pregnancy than white women, [8] with unconscious bias contributing to disparities [2,3]. The reports outlined above, amongst others, all highlight the lack of personalised care as a factor in poor outcomes, particularly when women feel unheard.

Personalised care is, therefore, undeniably crucial to improving maternity outcomes, and equipping healthcare staff to deliver it is equally essential. Comprehensive training must prioritise psychological safety and trust, fostering meaningful conversations that empower women to make informed decisions about their care.

Digital technologies play a pivotal role in this approach, enhancing access to information, supporting risk prediction, and facilitating seamless communication between service users and providers. When integrated effectively, these tools help create a maternity care system that is more responsive, equitable, and tailored to individual needs.

## 2. Digital technologies

Digital technologies are transforming maternity care by providing personalised support throughout pregnancy, birth, and postpartum period. With nearly half of pregnancies in England unplanned, women are often unaware of the choices open to them. Digital tools help service users understand their care options, manage risk assessments, and engage actively in their health journey. There are three distinct purposes for the use of digital technologies:

- Several commercially available systems serve as Electronic Patient Records (EPR), allowing women and healthcare professionals to access information such as blood results or care plans from any location.
- Some provide information about choices; others are there to motivate behaviours change such as stopping smoking.
- Others provide information such as options for pain management in labour.

Digital applications may enhance communication, predict risks, and facilitate multidisciplinary care. Specific platforms aim to address cultural and accessibility challenges, ensuring tailored support for Black, Brown, and South Asian mothers. These innovations empower users, streamline care processes, and contribute to safer maternal outcomes. These are available in different languages and may help to mitigate information overload. Some of these apps can also be accessed by health care professionals, which enables joint access to information and multidisciplinary planning of care.

## 3. Challenges: legal, ethical, and social

While maternity apps can support culturally sensitive, equitable care, they also pose legal, ethical, and social challenges. These are "vast and complex" [9]. Key concerns include trust in artificial intelligence (AI), accountability and reponsibility[7,] data security risks, [10,11] regulatory gaps, unclear consent processes, and potential exclusion of users with limited digital literacy or access. AI-driven apps may perpetuate bias, particularly affecting Black and Asian women [12] while private developers' monetisation of data raises conflicts of interest. Ensuring transparency, regulation, and user-friendly design is crucial for safe and effective implementation. It is important to consider what is the desired outcome from a person using the app, for example if a health behaviour is sort, it is important to understand that there is no strong evidence supporting the effectiveness of apps in improving health behaviours [13]. Having established that use of digital technologies which mostly incorporate some form of AI, are indeed beneficial to improving health care it is important to establish principles by which they should function [14]. We discuss two of them in this paper.

### *Uptake of digital technology*

Van Kassel et al [16] mapped barriers to uptake, devising the first comprehensive map of 85 factors influencing the uptake of digital therapeutics across health systems, bridging technological and human perspectives. This highlights the complexity of using digital therapeutics but also underlines the importance of involving all stake holders. A key finding in this work is that uptake requires strategic alignment with system intra intra-operability funding models, policy readiness, and patient awareness to drive equitable access and effective implementation.

Furthermore, Ramasawmy et al [16] highlights that while digital health interventions are increasingly used, frameworks addressing their implementation often overlook health inequalities, particularly ethnic disparities. Of 58 papers analysed, only 38% referenced health inequalities, and just 5 frameworks considered all relevant levels—society, system, intervention, and individual. Ethnicity was frequently embedded within broader social determinants rather than explicitly examined. The findings emphasize the need for a more integrated approach to digital health equity, suggesting that future frameworks should address systemic barriers and culturally tailored interventions to improve accessibility. This is

further highlighted by the introduction of a maternity app in a multicultural city, but where uptake was poor by the very women who may have benefitted the most [17].

***Quality Assurance and Benchmarking***

App developers aiming to benchmark against NICE's Evidence Standards Framework (ESF) [18] for digital health and care technologies must ensure their solutions meet standardized clinical and economic evaluation criteria. This involves aligning their evidence base with NHS England's requirements, incorporating rigorous stakeholder engagement, and adhering to agile, iterative development methodologies. Developers should focus on demonstrating measurable health outcomes, regulatory compliance, and cost-effectiveness to support adoption within health systems. Continuous engagement with NICE updates and evolving digital health standards is crucial for maintaining relevance in an innovative and fast-moving field.

The need for health professionals to take care is highlighted in the concern that current commercially available peripartum apps range widely in quality. A 2021 American systematic study of 301 peripartum apps found that only 25 met evaluation criteria, with significant gaps in maternal health content, inclusivity of people of colour, and usability. Apps with better maternal health information and inclusive imagery correlated with higher usability scores, yet this did not translate into greater user adoption. Overall, most existing apps lack sufficient maternal health information and fail to adequately serve diverse populations, highlighting the need for improved evidence-based, inclusive, and user-friendly designs [19].

***Management of data***

The number of medical devices and the enormous amounts of data they gather can be a problem for an organisations IT/care teams to oversee [20]. This could be in terms of inputs or understanding on acting on information read within the app.

## 4. Conclusion

Digital maternity apps have the potential to enhance personalised care but should be approached with caution, to eliminate risks. These include the risk digital exclusion. Equitable access requires offline capabilities, multilingual support, and community-based digital education. Regulation and stakeholder collaboration are key to ensuring informed decision-making and trusted digital solutions. Endorsed apps from reputable health bodies may provide assurance on data security and privacy. Addressing maternal health inequalities also demands a positive workplace culture, community-based care, and systemic reforms to integrate digital technology effectively.

## References


1. Paccoud I, Leist AK, Schwaninger I, van Kessel R, Klucken J. (2024). Socio-ethical challenges and opportunities for advancing diversity, equity, and inclusion in digital medicine. *DIGITAL HEALTH*. 2024;10. doi:10.1177/20552076241277705
2. Siden, J.Y, Carver, A.R, Mmeje, O.O and Townsel, C. D, (2022). Reducing Implicit Bias in Maternity Care: A Framework for Action. Women's Health Issues, Volume 32, Issue 1, 3 – 8, January 2022. DOI: 10.1016/j.whi.2021.10.008
3. Church, E., (2024) Student midwife develops app to combat skin tone bias. Nursing Times, 20th September 2024. https://www.nursingtimes.net/research-and-innovation/student-midwife-develops-app-to-combat-skin-tone-bias-20-09-2024/
4. Winfield, S. and Booker, M. (2021). Personalised care in maternity. *BMJ*, 02 August 2021. https://www.britishjournalofmidwifery.com/content/comment/personalised-care-in-maternity
5. NHS England (2016). Better Births: Improving outcomes of maternity services in England – A Five Year Forward View for maternity care.

https://www.england.nhs.uk/publication/better-births-improving-outcomes-of-maternity-services-in-england-a-five-year-forward-view-for-maternity-care/
6. NHSProviders (2025). Bold action: tackling inequalities in maternity care. https://nhsproviders.org/resources/bold-action-tackling-inequalities-in-maternity-care
7. NHS England (2023). Three-year delivery plan for maternity and neonatal services https://www.england.nhs.uk/publication/three-year-delivery-plan-for-maternity-and-neonatal-services/
8. MBRRACE-UK (2024), Saving Lives, Improving Mothers' Care: Lessons learned to inform maternity care from the UK and Ireland Confidential Enquiries into Maternal Deaths and Morbidity 2020-22.https://www.npeu.ox.ac.uk/assets/downloads/mbrrace-uk/reports/maternal-report-2024/MBRRACE-UK%20Maternal%20MAIN%20Report%202024%20V2.0%20ONLINE.pdf
9. Murphy, K., Di Ruggiero, E., Upshur, R. *et al.* Artificial intelligence for good health: a scoping review of the ethics literature. *BMC Med Ethics* **22**, 14 (2021). https://doi.org/10.1186/s12910-021-00577-8
10. Seh AH, Zarour M, Alenezi M, Sarkar AK, Agrawal A, Kumar R, Khan RA.(2020). Healthcare Data Breaches: Insights and Implications. *Healthcare (Basel).* 2020 May 13;8(2):133. doi: 10.3390/healthcare8020133. PMID: 32414183; PMCID: PMC7349636.
11. Plante, O. (2023). What Are The Security Risks in Healthcare Apps? The Expert's Guide https://www.fleksy.com/blog/cybersecurity-risks-healthcare-apps/
12. Banjoo, O. (2018). Bias In Maternal AI Could Hurt Expectant Black Mothers. https://www.vice.com/en/article/bias-in-maternal-ai-could-hurt-expecting-black-mothers/
13. Milne-Ives M, Lam C, De Cock C, Van Velthoven MH, Meinert E. (2020). Mobile Apps for Health Behavior Change in Physical Activity, Diet, Drug and Alcohol Use, and Mental Health: Systematic Review *JMIR Mhealth Uhealth* 2020;8(3):e17046 doi: 10.2196/17046 PMID: 32186518 PMCID: 7113799
14. HL (2018). AI in the UK: ready, willing and able? House of Lords, Select Committee on Artificial Intelligence. Report of Session 2017-19 https://publications.parliament.uk/pa/ld201719/ldselect/ldai/100/100.pdf
15. van Kessel R, Roman-Urrestarazu A, Anderson M, Kyriopoulos I, Field S, Monti G, Reed SD, Pavlova M, Wharton G, Mossialos E (2023). Mapping Factors That Affect the Uptake of Digital Therapeutics Within Health Systems: Scoping Review. *J Med Internet Res* 2023;25:e48000. doi: 10.2196/48000 PMID: 37490322 PMCID: 10410406
16. Ramasawmy et al (2022). Frameworks for Implementation, Uptake, and Use of Cardiometabolic Disease–Related Digital Health Interventions in Ethnic Minority Populations: Scoping Review. *JMIR Cardio* 2022, vol. 6, issue 2. https://discovery.ucl.ac.uk/id/eprint/10147744/1/Murray_PDF.pdf
17. Pierce P, Whitten M, Hillman S. (2023). The impact of digital healthcare on vulnerable pregnant women: A review of the use of the MyCare app in the maternity department at a central London tertiary unit. *Front Digit Health.* 2023 Apr 21;5:1155708. doi: 10.3389/fdgth.2023.1155708. PMID: 37153515; PMCID: PMC10161794.
18. Unsworth, H, et al (2021). The NICE Evidence Standards Framework for digital health and care technologies – Developing and maintaining an innovative evidence framework with global impact. *Digital Health*, Volume 7: 1–20.
19. Tucker, L., Villagomez, A.C. & Krishnamurti, T. (2021). Comprehensively addressing postpartum maternal health: a content and image review of commercially available mobile health apps. *BMC Pregnancy Childbirth* **21**, 311 (2021). https://doi.org/10.1186/s12884-021-03785-7
20. Choudhuri, A., JChatterjee, J.M., and Garg, S. (2019). Chapter 6 - Internet of Things in Healthcare: A Brief Overview, In Editor(s): Valentina E. Balas, Le Hoang Son, Sudan Jha, Manju Khari, Raghvendra Kumar, *Internet of Things in Biomedical Engineering*, Academic Press, 2019: 131-160, ISBN 9780128173565, https://doi.org/10.1016/B978-0-12-817356-5.00008-5.

# Non-Medical Device Apps for Good Health and Well-being: Regulatory Challenges

*Dr Carlisle George*
Associate Professor & Barrister - ALERT Research Group, Middlesex University, UK
c.george@mdx.ac.uk

## 1. Introduction

General healthcare/well-being awareness and technological innovations have resulted in numerous smartphone apps (especially powered by artificial intelligence (AI)) being developed and marketed to the public. Due to the widespread use of smartphones, apps can be easily downloaded and are hence widely available. Revenue from digital fitness and well-being apps (e.g., to address anxiety, depression, stress, panic attacks, mindfulness, meditation, mood tracking and self-care) in the United Kingdom (UK) is projected to reach £.1.65 billion in 2025 and £2.16 billion by 2029 [1]. Most fitness and well-being apps are not categorised as medical devices, and are therefore not regulated by the Medicines and Healthcare products Regulatory Agency (MHRA). This paper discusses medical device apps versus non-medical device apps with a focus on the regulatory challenges of the latter especially in the UK context.

## 2. Apps as medical devices versus non-medical devices

There are distinct differences between apps defined as medical devices as opposed to those classified as non-medical devices.

A medical device is defined under Part 1 of the Medical Devices Regulations 2002 (SI 2002 No 618, as amended). It includes any "*any instrument, apparatus, appliance, software, material, or other article... intended by its manufacturer to be used specifically for diagnosis or therapeutic purposes or both...*". Medical devices are intended for human use including: to perform a medical purpose such as diagnosing, preventing, monitoring, treating, or alleviating a medical condition; and controlling conception [2]. In the context of a smartphone app, if an app is marketed to be used for a medical purpose, then it is classed as a medical device. Determination of a "medical purpose" can be made by statements and instructions for use (including in any promotional material) given by an app developer or app store description/category [3]. Further guidance on how to determine whether an app qualifies as a medical device is provided in [3]. Examples of medical device apps include those used for therapy, remote patient consultation, medication management, and glucose monitoring. The MHRA regulates these medical device apps.

All apps that do not meet the requirements of a "medical device" (i.e., are not intended for a medical purpose) are classed as "non-medical" devices and are therefore not subject to regulatory oversight by the MHRA. Many fitness and "well-being" apps are classified as non-medical devices. These apps are often used for activities such as collecting and communicating health-related data (without altering the data), simple data processing (e.g., storage, search, and matching), providing advice on nutrition and fitness, and monitoring vital signs, including heart rate or blood pressure. Many health and well-being apps incorporate AI, especially Generative AI, to provide a variety of functionalities including personalised plans, adaptive programs, intelligent coaching, real-time feedback, mood tracking, and chatbots ([4,5]).

## 3. Regulation

Apps are regulated by a variety of different regulatory instruments. Although apps classified as non-medical devices are not regulated under the MHRA, there are several UK-based and international regulatory frameworks relevant to both medical and non-medical devices. These include: (i) ISO/IEC standards created by the International Organization for Standardization (ISO) and International Electrotechnical Commission (IEC); (ii) ethical and governance standards, and (iii) other examples of UK legislation.

### 3.1 MHRA regulation

The regulation of medical devices (including apps under this category) involves a very complex and comprehensive oversight/regulatory regime (under the MHRA), which includes the following [6]:

- *Registration*: The requirement to register the device with the MHRA before it is placed on the market.
- *Conformity assessment*: A process to prove that the device meets requirements under the Medical Devices Regulations 2002 (SI 2002 No 618, as amended) and other relevant legislation, resulting in a UK Conformity Assessed (UKCA) mark or a European Conformity (CE) mark (to demonstrate compliance with European Union legislation and safety standards).
- *Risk management assessment*: A process to identify, monitor and prevent risks associated with the medical device.
- *Data protection compliance*: The need to comply with data protection law.
- *Labelling requirements*: In addition to a UKCA or CE conformity mark, the need to properly display instructions for safe use.
- *Post market surveillance and vigilance*: This covers various obligations including collecting surveillance data (e.g., for improvement of usability and safety), submitting vigilance reports to the MHRA when certain incidents involving the device occur in the UK, and taking appropriate safety action when required.

### 3.2 ISO/IEC standards

ISO/IEC standards are guidelines and specifications (developed by industry experts) aimed at defining products, processes or services to ensure fitness for purpose [7]. They cover important topics such as quality, safety, and security. Examples of such standards relevant to app development are ISO 9001 (quality management), ISO 27001 (information security management), ISO/IEC 27034 (application security), and **ISO/IEC 29134 (privacy by design)**. App developers can gain ISO certification via third-party auditing. For non-medical devices, implementing ISO standards or ISO certification is not mandatory; however, such standards/certification help with an organisation's reputation and user trust among others.

### 3.3 Ethical and governance standards

By 2025, ethical and governance standards have become particularly relevant, especially due to the widespread use of AI in most industries (including app development) and the accompanying ethical concerns they bring. Some ethical concerns of AI include [8]: informational privacy; security, transparency and accountability; bias and discrimination; and safety. Several organisations have published guidance focusing on ethical/governance standards for AI, including:

- EU Ethics Guidelines for Trustworthy AI [9].
- The World Health Organization (WHO) guidance on Ethics and Governance of AI of Large Language Models (LLMs) [10].
- OECD AI principles for Trustworthy AI [11].
- UK Government's five principles for guiding/informing responsible development and use of AI [12].

These guidance documents include several common ethical requirements such as the need for: transparency and explainability; privacy and data protection; accountability; safety, security and robustness; diversity, non-discrimination and fairness; and human agency and oversight.

Regarding app development, in 2023, the UK published a voluntary code of practice for app store operators and app/platform developers [13]. The code contains eight requirements to protect users, including: app baseline privacy and security requirements; app store baseline privacy and security requirements; vulnerability disclosure; the need to keep apps updated; privacy and security information for users; developer guidance on security and privacy; developer feedback on security and privacy; and handling personal data breach notifications.

### 3.4 Legislation

In addition to the Medical Devices Regulations 2002 (SI 2002 No 618, as amended), there are several other forms of legislation in the UK impacting the development and marketing of all apps. Some important frameworks include:

- *Data Protection*: The Data Protection Act 2018 (incorporating the UK GDPR) focuses on the protection of personal data. It mandates various requirements, including the need for: informed consent, data security, data breach notification, data accuracy, data minimisation, transparency, storage limitation, data protection by design and default, and subjects' rights. App developers need to ensure compliance with data protection obligations since the use of apps will involve the processing of personal data.

- *Consumer Protection*: The Consumer Rights Act 2015 applies to both physical and digital goods (including software) and requires that goods (apps) meet various requirements, such as the need to be of satisfactory quality, fit for purpose, and as described to the consumer.

- *Security*: The Product Security and Telecommunications Infrastructure (PSTI) Act 2022 contains provisions for the security of internet-connectable products, including: security by design; ongoing security; and user transparency requirements. App developers are impacted when an app is needed to operate a device (e.g., a wearable). In the UK in 2024, a new Cyber Security and Resilience Bill was announced that will add additional cybersecurity obligations to app developers, such as mandating security by design and default. These additional obligations reflect similar EU legislation, such as the EU Cyber Resilience Act 2024, which enhances cybersecurity standards of products sold in the EU that contain a digital component.

- *Product Safety*: The General Product Safety Regulations 2005 place obligations on businesses to ensure that consumer products (including software/apps) meet safety requirements (i.e. they pose no risks or minimal risks to consumers). These obligations apply to any business in a product's supply chain (e.g. app developers and distributors). App developers in the UK should be aware of some important product regulations if they intend to market their products in the EU. The EU Artificial Intelligence Act (EU AI Act) 2024 contains rules for the development, marketing, and use of AI systems in the EU. It also classifies AI systems according to risk levels and places accompanying obligations on the providers/deployers of these systems. The EU Regulation on General Product Safety 2023/988 (updated to cover emerging digital technologies) outlines various safety obligations (including for cybersecurity).

- *Liability*: The Consumer Protection Act 1987 addresses product liability in the UK. The producer of a product is strictly liable if a defect in the product causes harm/damage. Under the Act, the definition of "product" only extends to software provided on a physical medium (e.g. a USB disk) and does not extend to software provided electronically (such as downloaded apps). The Act is based on a 1985 EU Product Liability Directive 85/374/EEC 1985. However, in the EU, the 1985 Directive has been replaced by the new EU Product Liability Directive *2024/2853,* which explicitly includes digital products (software, AI systems) and applies the same liability standards as for physical goods. This liability includes manufacturer liability for defective digital and AI products that cause damage (such as bodily injury, property damage, destruction or corruption of data). While, post-Brexit, the new European directive does not apply in the UK, app developers in the UK need to be aware of it if they intend to market their apps in the EU.

## 4. Discussion

The discussion which follows highlights important aspects of the regulatory landscape in the UK which impact on smartphone apps.

Non-medical device apps for good health and well-being are not subject to the stricter regulatory regime of the MHRA and Medical Devices Regulations 2002 (SI 2002 No 618, as amended). Although non-medical device apps are regulated under various other legislations, ISO/IEC and ethical standards, the enforceability and adequacy of these regulatory frameworks are debatable. Non-medical device apps for good health and well-being are easy to develop by anyone using automated tools (often with little oversight or enforced governance processes). They are simple to market worldwide, since no registration or formal process is required before they are deployed. Furthermore, there is no compulsory conformity mark requirement, and post-market surveillance is not mandated. There is also a lack of sufficiently standardised and enforceable regional and international regulations. For example, the UK Code of Practice for app store operators and app developers is voluntary.

Ensuring adequate data privacy and security continues to be a concern/challenge for apps processing personal data related to health. Despite the existence of legislation focusing on data protection and data security, studies conducted in the last five years have found that apps have inadequate data privacy and security protections [14,15].

In addition, the widespread use of non-medical device apps for good health and well-being has raised various other concerns and regulatory challenges. A 2025 research study found that the use of some fitness apps resulted in negative body image problems and compulsive exercising [16]. This study concluded that some people who use diet apps were more likely to have specific behaviours and attitudes related to eating, such as food preoccupation and dietary restraint. Previous related studies [17] found that some apps offer inaccurate data (e.g. calorie expenditure data), leading to users acting on this erroneous data (e.g., increased exercising). In addition, the apps can cause users to experience anxiety due to constant monitoring.

Regarding apps for mental well-being, especially those using AI-driven emotional support chatbots, concerns have been raised regarding users' overreliance on them, resulting in people not seeking professional help and hence delaying evidence-based treatment [18]. These apps may offer users a false sense of security while, however, they receive inadequate care, delaying proper treatment and worsening their condition. A 2024 study found that some AI companion apps using chatbots generally failed to provide help in times of a mental health crisis [4]. Most responses to crisis messages did not express any empathy, almost half were unhelpful, and more than half were risky. The consequences of using some apps can be fatal, as was the case of a man who committed suicide after being encouraged to do so by an AI chatbot (Eliza) on an app called Chai [19]. The man had an obsession with climate change and consulted the AI chatbot. An investigation of his six-week conversation with the chatbot found that the bot encouraged him to take his life after he proposed sacrificing himself to save the planet. The Chai app is available for anyone to download on Google Play and other app stores. Reviews of the app indicate a 4.1 out of 5 rating, and user comments such as the following confirm its popularity: "*I enjoy the app, it cures my boredom since we can choose how things go and have endless conversations",* and *"I think this is my fav chat app so far."*[20].

These examples, arguably, demonstrate that apps for good health and well-being may not be sufficiently regulated. Many of these apps continue to pose potential risks or harm to consumers, and therefore, there is a need for greater regulatory oversight of them to ensure public safety and trust.

Moving ahead, greater regulatory oversight for apps used for good health and well-being can perhaps be achieved by considering a multiplicity of approaches, including the following:

- Widening the definition of a "medical device" to include all apps used for any aspect of human health and well-being, hence making them subject to MHRA oversight.

- Developing better compliance mechanisms for existing legislations related to privacy, security, and safety.
- Improving industry codes of practice for app development.
- Implementing mandatory requirements for app stores.
- Improving regulations on digital dependency to address problems identified with apps associated with mental well-being.
- Implementing a common global regulatory approach headed by international organisations.
- Improving educational outreach and awareness of potential risks/harms related to apps.

## 5. Conclusion

Non-medical device apps (for health and well-being) continue to be easily developed, marketed, and made readily available to consumers, especially due to the widespread use of smartphones. Many of these apps pose potential risks and harm to consumers that are often not immediately apparent. It is therefore necessary to examine carefully how the many concerns of these apps can be adequately addressed by improving the current UK regulatory landscape. As we move into the future, when society will become more dependent on digital technologies, the need for effective and globalised regulatory mechanisms for health and well-being apps is imperative.

## References


1. Statista (2025). Digital Fitness & Well-Being https://www.statista.com/outlook/hmo/digital-health/digital-fitness-well-being/united-kingdom
2. NHSE (2025). Medical devices and digital tools. Version 1.3, 3 March 2025. https://www.england.nhs.uk/long-read/medical-devices-and-digital-tools/
3. MHRA (2023). Guidance: Medical device stand-alone software including apps (including IVDMDs).https://assets.publishing.service.gov.uk/media/64a7d22d7a4c230013bba33c/Medical_device_stand-alone_software_including_apps__including_IVDMDs_.pdf
4. De Freitas, J., Cohen, I.G. The health risks of generative AI-based wellness apps. *Nat Med* **30**, 1269–1275 (2024). https://doi.org/10.1038/s41591-024-02943-6
5. Kuru, H (2023). Identifying Behavior Change Techniques in an Artificial Intelligence-Based Fitness App: A Content Analysis. *Health Education & Behavior*. 2023;51(4):636-647. doi:10.1177/10901981231213586
6. MHRA(2025). Guidance: Regulating medical devices in the UK https://www.gov.uk/guidance/regulating-medical-devices-in-the-uk
7. Van Velthoven, M.H., Wyatt, J.C., Meinert, E., Brindley, D., and Wells, G. (2018). How standards and user involvement can improve app quality: A lifecycle approach. International Journal of Medical Informatics, Volume 118, 2018, Pages 54-57, ISSN 1386-5056, https://doi.org/10.1016/j.ijmedinf.2018.08.003.
8. Stahl, B.C. (2021). Ethical Issues of AI. In: Artificial Intelligence for a Better Future. SpringerBriefs in Research and Innovation Governance. Springer, Cham. https://doi.org/10.1007/978-3-030-69978-9_4
9. EU (2019). Ethics Guidelines for Trustworthy AI https://www.aepd.es/sites/default/files/2019-12/ai-ethics-guidelines.pdf
10. WHO (2025). Ethics and governance of artificial intelligence for health: Guidance on large multi-modal models. https://www.who.int/publications/i/item/9789240084759
11. OECD (2024). OECD AI Principles Review https://oecd.ai/en/ai-principles
12. Gov.UK (2023a). A pro-innovation approach to AI regulation https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper
13. Gov.UK (2023b) Code of practice for app store operators and app developers https://www.gov.uk/government/publications/code-of-practice-for-app-store-operators-and-app-developers

14. Tangari G, Ikram M, Ijaz K, Kaafar M A, Berkovsky S. (2021). Mobile health and privacy: cross sectional study BMJ 2021; 373 :n1248 doi:10.1136/bmj.n1248
15. Biswas, A.A., Zulfiker, M.S., Rahman, M.M, Jani, M. R., and Anwar, M. M., (2025) "Data Privacy and Security Analysis for Mental Health Chatbot Applications," *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, Melbourne, Australia, 2025, pp. 1245-1249. Doi: 10.1109/HRI61500.2025.10973904.
16. Koster, A (2025). Millions rely on fitness apps for motivation and guidance. They may do more harm than good. https://www.sbs.com.au/news/article/do-fitness-apps-cause-problematic-food-and-exercise-habits/ydtjhgou3
17. Davis, A (2018), Are health apps actually bad for your health? https://www.bbc.co.uk/bbcthree/article/9fe47476-ad1f-414c-a925-cf078a2145a8
18. Babu A and Joseph AP (2025) Digital wellness or digital dependency? a critical examination of mental health apps and their implications. Front. Psychiatry 16:1581779. Doi: 10.3389/fpsyt.2025.1581779
19. El Atillah, I (2023). Man ends his life after an AI chatbot 'encouraged' him to sacrifice himself to stop climate change. https://www.euronews.com/next/2023/03/31/man-ends-his-life-after-an-ai-chatbot-encouraged-him-to-sacrifice-himself-to-stop-climate-
20. Google (2025). Chai: Chat AI Platform https://play.google.com/store/apps/details?id=com.Beauchamp.Messenger.external&hl=en_GB

# User Context-Centred Design of an mHealth System for a Developing Country

*Professor Galal H. Galal-Edeen*
Heikal Department of Management, Onsi Sawiris School of Business
American University in Cairo, Egypt.
Galaledeen@aucegypt.edu

## 1. Introduction

Egypt is a developing country of more than 107 million inhabitants in 2024 [1]. It has one of the highest out-of-pocket (OOP) spend rates on healthcare in the world, estimated to exceed 60% of total healthcare spending [2]. This position paper outlines a field project supported by the European Union, that aimed to leverage modern mobile communication technology to support strategic public healthcare improvement and strategy formulation by the Ministry of Health and Population (MoHP) in Egypt. The scope is limited to Non-Communicable Diseases (NCDs), such as coronary diseases, diabetes and hypertension. This paper highlights the contextual imperatives that drove the research process, leading to a design that is grounded in the contextual conditions revealed, in way that increases its success chances when ultimately deployed.

Given the relatively recent diffusion of smart mobile phone technology across multiple socio-economic classes, there is a clear opportunity to exploit such technology to support healthcare strategic planning aims. The project outlined here aimed at designing a usable, community-supported and easily accessible mobile phone application to collect the vital statistics needed to guide healthcare policy formulation and facility distribution strategies.

This project was triggered by a request from the head of the Egyptian Curative Organisation, one of the organisations that comprise the Egyptian state healthcare provision landscape, managed by the Egyptian Ministry of Health and Population (MoHP). That ministry was the main client of this design project. A special effort was made to frame the problem by identifying the need and motivation of the ministry's priorities, this resulted in a request for a mobile communication-based tools, possibly a mobile app, to help track vital health readings like blood pressure, blood glucose levels, and temperature.

Ultimately, the collected data are to be used to support health policy makers with their strategic resource deployment decisions. The success of the planned intervention required a high level of buy-in and mass use so that the collected statistics are meaningful at the national level.

## 2. The Cultural Context

The design of a product/ service targeted towards mass audience, requires a careful study of the culture, limitations, the users and usage context. Thorough field research, therefore, forms an integral part of the design cycle.

The target population include both urban and rural area dwellers, typically characterised by:

1. An illiteracy rate of around 25.8% (age 10+) in 2017, leaving a literacy rate of around 74.5% for adults aged 15 years and over [3].
2. Low-income levels: the overall average monthly wage for public and private sector employees was 5,005 EGP in 2023. This equates to roughly $160–170 USD per month (using 2023 exchange rates) [4].
3. Low access to conventional computing technology (e.g. personal computers). However, Egypt has rapidly expanded mobile and Internet access. ITU data show about 92.0 mobile-cellular subscriptions per 100 inhabitants in 2022, meaning roughly 0.92 SIM cards per person. Internet use is also widespread: an ITU/UN estimate puts the share of individuals using the Internet at 71.9% in 2022 [5].

4. Preventive health check-ups are underutilized, due e.g., to high cost and poor awareness. So, most people do *not* get regular screenings, which has been identified as a key obstacle [6] which surely affects negatively the management and avoidance of NCDs.
5. Cardiovascular diseases are the leading cause of death in Egypt, with NCDs accounting for a significant portion of the disease burden, with significant geographical disparities in healthcare access [7].

The first phase of research was explorative in nature, aiming to map out the peculiarities of the research space. Egypt is a high Power Distance Index (PDI) country [8], so several official permissions had to be obtained, after which various initial visits to hospitals and interviews with several physicians and healthcare workers, public and private, as well as officials from the MoHP, were conducted and the team was able to finalise the following set of research questions:

1. *What kind of lifestyle do people (at the Base of the Pyramid (BoP)) have?*
2. *What is the impact of a Non-Communicable Disease (NCD) on a patient's life?*
3. *How does the current healthcare system of Egypt work?*
4. *What technologies are available to the BoP?*
5. *What is the professional opinion of the healthcare system and NCDs in particular?*
6. *What are the needs, values and future hopes for and of the BoP?*

To address the above questions, a mix of literature reviews, field observations, interviews, meetings in which respondents were utilised to build a picture of the topic under investigation.

**3. The Intervention**

The group targeted by this intervention are BoP members, typically characterised by being at the lower spectrum of the national household income levels, with a level of education that is a little more above the literacy level, which impacts their knowledge of their diseases and ability to access related information sources in general. The field research revealed the following stakeholders:

- Ministry of Health and Population (MoHP)
- Doctors
- Patients
- Raeda Refiya (Health Visitor: State trained and appointed)
- Pharmacies

It was decided not to consider the patients (stakeholders at the BoP) to be the main users of the application. This may sound counter-intuitive; however, the following factors were behind that decision:

1. The issue of patients' illiteracy: low literacy reduced the patient's ability to take the initiative in feeding the necessary data into the application.

2. Cultural and lifestyle factors:
- Low personal responsibility for one's health.
- Priority in life is more about survival than improving health.
- Challenges to integrate new ideas or modalities into the current systems and culture.
- The need for trustable sources to convince patients to use a (relatively) new technology and following an unusual modality.

3. The patients themselves would be an unreliable source for NCD data.

4. Limited sources of (mobile) technology, such as mobile Internet connectivity, because of the added financial burden.

It was decided that the best configuration for this system was to integrate a number of stakeholders, each type with a relatively low burden of responsibility towards operating and sustaining the application, and yet some perceived value that each can derive from it. Instead of relying on one major user or beneficiary, a constellation of stakeholders was considered to drive the necessary functionality.

Figure 1 below illustrates this:

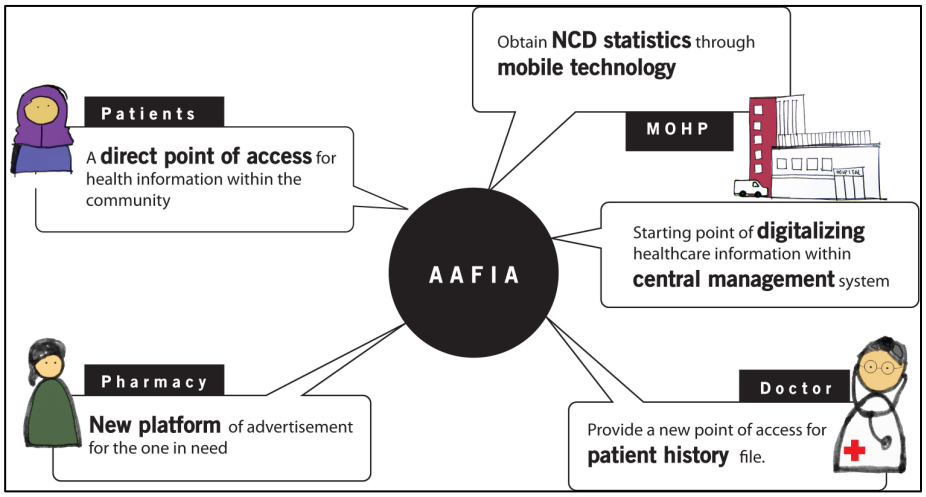



Figure 1: A virtuous configuration of stakeholders' roles and interests.

This stakeholders' multi-win configuration proved very promising to overcome the issues identified above. Following this high-level design of the system's main functions, configuration and stakeholders' involvement, a detailed design of the interaction was developed and a low-fidelity prototype built (using wireframes and screen mock-ups) and forwarded for user testing. The design was modified and finalised based on the feedback obtained.

**4. Future Expansion**

The proposed design was not deployed in the field due to funding and operational obstacles. However, it was important to show potential expansion in the future within the core functions established in the first (design) phase, the Aafia system can be extended to benefit more potential users, as follows:

- *The Raeda Rafeya (RR) can be integrated within the system (similar to one of the design directions previously identified), creating a new service system within the already built-in database. RRs can benefit from the personal information on patients as well as access to health and medical information database currently unavailable to them. This would give them direct and easy access to information and can be a more efficient way to transfer messages to patients in rural areas.*

- *The Doctor will get some basic information about their patients and also a platform to keep data to monitor their NCD patients. As the system expands, doctors will be able to also add the health database comprising their personal advice by paying a small fee which can be a way to promote their capability within this competitive market.*

***A vision for sustainability***

This system can provide long-term benefits as it can reduce costs incurred by healthcare providers and patients, especially for tertiary care by trying to shift the focus towards better primary healthcare. Since the system is designed to support monitoring and prevention phases, it can reduce patient's risk of progressing to more serious stages or prolonging their current health status. If the currently designed core system functions well when deployed, it can then be expanded towards more profitable services that generate revenue, e.g. through paid users like private doctors and their patients.

## 5. Conclusions and Lessons Learned

Designing a healthcare intervention for use by the public, many of whom maybe illiterate or poorly educated, is a complex undertaking that requires careful consideration of the users' wider context. From this project, we can draw a number of important lessons:

1) Thorough research of the context, supported by appropriate research tools and "probes", which have a high but simple visual content that can be easily assimilated by the stakeholders, was found to be vital for reliable fact finding.
2) The use of culturally sensitive entry methods, language and tools is very important, and some adjustment of the research tools is often necessary.
3) A careful investigation of the context can bring up important design features that are critical to the field success of the proposed intervention.
4) It's important for such an intervention to provide perceived value for all the stakeholders, who basically collaborate - perhaps indirectly - to "lift" the intervention and make it operationally successful and economically viable.
5) It's important to consider the service ecosystem design, since a wide release of such a system in a large country like Egypt, would critically rely on it.
6) Using a suitable number of prototypes, on suitable platforms, to conduct user testing is vital for removing initial design problems and to gain vital user feedback.
7) The designers of similar healthcare systems need to adopt flexible attitudes to design - for example, by shifting primary user focus when needed - and pay close attention to the broader usage contexts

## References


1. CAPMAS (2025a).Central Agency for Public Mobilization and Statistics. https://www.sis.gov.eg/Story/197754/CAPMAS-Egypt's-population-officially-reaches-107mn?lang=en-us. Accessed 6th June 2025
2. Fasseeh et al. Journal of the Egyptian Public Health Association (2022) 97:1 https://doi.org/10.1186/s42506-021-00089-8
3. CAPMAS (2025b) https://www.capmas.gov.eg/Admin/News/PressRelease/202198111315_999-e.pdf**.** Accessed 8th June 2025
4. CAPMAS (2025c) https://censusinfo.capmas.gov.eg/metadata-ar-v4.2/index.php/catalog/1920/download/6926**.** Accessed 8th June 2025
5. UNData (2025) http://data.un.org/Data.aspx?q=Egypt&d=ITU&f=ind1Code%3AI911%3BcountryCode%3AEGY#:~:text=Country%20or%20Area%20Year%20Value,0031. Accessed 8th June 2025
6. Gamaa, ARA.; Khedr, MA; Fatah, NKAE (2024). From insight to action: Exploring knowledge and practices about periodic medical check-up among community dwelling older adults in Egypt. (2024) Geriatr Nurs. 2024 Nov-Dec;60:620-627. doi: 10.1016/j.gerinurse.2024.10.040. Epub 2024 Nov 4. PMID: 39500026.
7. Radwan G, Adawy A. The Egyptian health map: a guide for evidence-based decision making. East Mediterr Health J. 2019 Jul 24;25(5):350-361. doi: 10.26719/emhj.18.048. PMID: 31364760.
8. Hofstede, G. (2011). Dimensionalizing cultures: The Hofstede model in context. Online readings in psychology and culture, 2(1), 8.

## Final Message from the Interim Head of Department of Computer Science

The numerous contributions to this workshop collectively represent a rich and diverse array of insights into the legal, ethical, social, and governance challenges facing Health Information Technology (Health IT). This workshop highlighted the complexity and urgency of these issues, as well as their growing significance in an era increasingly shaped by artificial intelligence (AI) and other emerging technologies.

From discussions on AI accountability to the equitable deployment of digital health tools, the contributions underscore the multifaceted nature of Health IT. Addressing these dimensions requires thoughtful and adaptive regulatory frameworks, as well as inclusive design principles and a steadfast commitment to ethics and responsibility.

This workshop serves as a vital forum for interdisciplinary engagement, bringing together scholars, students, researchers, practitioners, policymakers, technologists, and other stakeholders. Such a diversity of perspectives is indispensable in navigating the dynamic and rapidly evolving digital health landscape in a manner that is both responsible and sustainable.

The Department of Computer Science at Middlesex University is honoured to support and host this event. We envision the future of Health IT not just as a journey of technological progress, but as one equally grounded in core human values—transparency, trust, privacy, fairness, inclusivity, and accountability.

I wish to extend my sincere congratulations to the members of our ALERT (Aspects of Law and Ethics Related to Technology) research group for their exemplary work in organising the seventh edition of this workshop series, which began in 2011. Their dedication, intellectual rigour, and spirit of collaboration have made significant contributions not only to academic discourse but also to real-world impact.

Finally, I offer my deepest gratitude to all who have contributed to the success of this workshop — our co-sponsor (the University of Miami), co-organisers, speakers, and attendees alike. Your engagement and collaborative spirit have been both meaningful and inspiring.

**Dr Aboubaker Lasebae**
Interim Head of Department of Computer Science

**Proceedings of the MDX 2025 Health IT Workshop**

**on**

***Emerging Technologies in Healthcare:***
***Legal, Ethical, Social & Governance Aspects***

**26th & 27th June 2025**
**Middlesex University, London, UK**

https://tinyurl.com/MDX-WKS-2025

ALERT (Aspects of Law and Ethics Related to Technology) Research Group
Faculty of Science and Technology

https://www.eis.mdx.ac.uk/research/groups/Alert/

ISBN 979-8-89965-689-7


Proceedings DOI: https://doi.org/10.60528/2q6192