

Chapter 10: Ethical challenges and governance of intelligent systems in sensitive clinical environments

10.1. Introduction

Advances in artificial intelligence are rapidly transforming healthcare and enabling new smart healthcare systems. An informed and considered approach to addressing the ethical challenges associated with the adoption of these new intelligent systems is critically important.



Fig 10.1: Ethical and regulatory challenges of AI technologies in healthcare

There are various sets of state of the art best practice guidelines developed by interdisciplinary teams; however, the feasibility of putting these guidelines into practice remains to be validated by actually implementing a system. This paper contributes to that end by considering the development, deployment, and governance of intelligent systems focusing on security and privacy in smart healthcare environments. Lessons learned and practical advice for future intelligent system projects are drawn from the experience of developing a real-time sanitization monitoring system for a healthcare setting, aimed at preventing the spread of disease-carrying pathogens. In order to distribute and then discuss this practical advice, ethical challenges for intelligent systems reported in the literature are reviewed, and then considered specifically for smart healthcare environments.

10.1.1.Background and Significance

In sensitive clinical environments the clinical effectiveness of any health solution has strong impact, being this impact relative to the society level of a global issue, or to a city as well as to an individual. This chapter is based on a broad review of literature to discuss and outline some of the key challenges which are of ethical and governance nature, and are most relevant considering the use of intelligent solutions, beyond the AI, in the management of sensitive clinical environments. The use of AI and other intelligent technologies in the management of sensitive clinical environments has gained considerable attention due to the hope associated with the possibility of significant benefits for societal, city, institutional and individual levels. Smart Hospitals use intelligent systems to automate the processes, to monitor activities and to create personalized treatments. Some of that automation uses elements that apply deep learning. Although there are many advantages offered by the use of these advanced models, there are unique ethical and governance challenges that come up together with them, such as interpretability and identifying the basis of a clinical decision. Nonetheless, Intelligent Systems are digitally based and are themselves data sources. The use of such systems creates new data types, volumes and linkages. This creates new risks of data breaches, which are realistic, as Independent Ethical Hackers performed penetration testing of various United States hospitals and found out that these systems had weak external perimeter security.

10.2. Understanding Intelligent Systems

In the last decades, technology has made remarkable progress, making our world almost unrecognizable. Due to the computer science expansion and the technology development, the society has involuntarily stepped into what we call the era of the SMART components, i.e. the time when the systems that surround us are expected not just to fulfil their basic functions but to possess also sophisticated skills and to behave in a purposeful way. Intelligent systems (IS) are widely used in many areas, from forecasting to decision making, mainly because with the development of IT, an increasing amount of data sources could be accessed easily.

An intelligent clinical support (ICS) has been developed that can help the physicians make decisions and hence improve the quality of the medical services. There are two sets of tools that support analytical processing: one tailored for the clinical cognition which is built on a Case Based Reasoning (CBR) system model, and the other one is tailored for the medical act, consisting of a quite efficient ensemble of clustering algorithms and statistical testing aimed at establishing significant difference in large data sets for variables that are relevant for the patient. The intelligent patient management (iPM) is a software product that combines home care services with the traditional healthcare services offered by a hospital. It enables remote monitoring for the patients and improves the management of the medical resources of a hospital, at the same time. In order to identify the relevant aspects from the data flow respective for the patient's case, several methods of data mining are presented. With these models, the prediction of the patient's evolution can be realized with as many details as the initial data stock can be used.

10.2.1. Definition and Scope

Intelligent systems and their applications have seen increased development in general over recent years. Moreover, particularly after the pandemic, there have been calls to utilize them in the healthcare domain too. On the one hand, ensuring the highest possible safety standards and protecting critical health system infrastructure is paramount. On the other hand, such systems have the increased potential to affect individual care. This latter aspect multiplies the importance to guarantee that these systems are both effective and ethically sound. There are known risks connected to the introduction of autonomous systems into sensitive environments. These include issues like the lack of awareness of the decisions taken, the fact that humans shift their responsibility to the system, or even the incentivizing of system misuse. This gives rise to the problem that it can be hard to distinguish whether certain actions were taken either primarily with the intent

to harm the patient or out of negligent disregard. However, besides the ones named beforehand, there are several ethical challenges posed either by the specific use case of such systems or the environment in which they are applied. This text intends to review the topic in the context of applications in sensitive clinical environments.

Before specifically starting to address the challenges and different aspects of ethics and governance in such systems, one should acknowledge what is meant by "intelligent." There is a lack of a single universal definition, and the term can be used to describe a broad range of functionalities or capabilities of the system. In fact, in the context of clinical decision support systems, this can range from simple personification of statistical alerts to systems running complex machine learning models, visualizing and summarizing data into actionable insights for clinicians, to even fully autonomous treatment. Similarly, the definition of "ethical" can span multiple disciplines with a vastly different approach to the subject.

10.2.2. Types of Intelligent Systems in Healthcare

There are a broad and diverse array of intelligent systems for healthcare. Those treating and diagnosing illnesses could include an automated system for healthcare triage or a virtual assistant for carers to monitor the medical condition of their patients. Some types of AI, such as certain forms of analytics or prediction, are more easily standardised and regulated in what they do. Some AI applications in healthcare are still human-directed, with the AI being used to supplement the human decision-making process rather than make decisions itself. Often here, the 'black box' issue is not viewed as a concern because while an AI-powered tool might learn new data between uses, each time it is asked to predict the outcome of a treatment or analyse an image it is asked the same question in how it presents its output. This section will therefore focus on the coordination and integration of AI-driven intelligent systems for healthcare whose use might push the sector in a new direction and how this currently relates to the current governance landscape. Moreover, assuming increasing integration of intelligent systems and automation within the information technology infrastructures of healthcare environments, this creates a tower-of-babel scenario with heterogeneous technologies from multiple vendors interacting in complex ways.

Amidst an array of ethical challenges, technoscientific governance, which encompasses a focus on the material substrate and informational elements of science and technology policy and regulation, is receiving increasing attention in relation to healthcare artificial intelligence. However, less attention has been paid to the coordination and integration of these technologies in complex sociotechnical infrastructures. Sensitive environments see technologies interacting with human agents and the environment in intricate ways. Within wider coordination and integration, the harmonisation of regulation and standards, both at the material substrate and information levels, in order to foster interoperability between these systems and components may offer a novel perspective through which to examine ethical challenges. Informed by a nascent project at a multimodal neurosurgical intervention suite, this article will focus on the ethical challenges of the unique environments in which intelligent systems for healthcare are deployed and the implications for intelligent systems research and design.

Refocusing science and technology policy and governance on techno-scientific elements elucidates the role of the 'black box' material substrate in enabling and constraining the development, use, and governance of science and technology, and the information inherent in it. Assembly of a science and technology is far from a single event or agent, but rather a continually evolving sociotechnical process of inscriptions and retranslations between stakeholders of different speech communities. Such textual mobility and the experimental practices and norms associated with these technologies are often hard-won technological achievements over decades. Furthermore, the consumption of science and technology in the form of products and services relies on informational elements which often leak beyond its intended circulated presentation through skilled users, bodily sensations, and insights from a variety of social contexts.

10.3. Ethical Frameworks in Healthcare

Intelligent systems driven by artificial intelligence (AI) have gained increasing interest in the healthcare domain due to their enhancement of medical diagnostic and treatment processes in terms of accuracy and efficiency (Chen & Hao, 2020; Lee & Kim, 2020; Kumar & Singh, 2021). AI-based systems are able to provide healthcare by independent learning through adaptive machine learning algorithms, providing medical solutions from treatment planning to assistive care.

There is ongoing research on general ethical guidelines in new AI-driven sensitive applications that seek to provide formal and rigorous analysis of the ethical considerations of such systems. The present study is the first to go beyond general ethical guidelines and offer a formal ethical analysis focusing on the healthcare domain. To provide further rigor to this analysis, it was applied using a novel framework inspired by the 4P ("People, Process, Product, Policy") framework.

An intelligent system in a healthcare domain is a special AI-empowered technology applied for a specific healthcare solution; sensitivity is referred to as the intangible quality of the system that drives it to have a non-negligible effect on the healthcare provision. While AI is typically seen as a technological innovation providing opportunities not only in healthcare but also on a broader societal and industry level, it may also raise ethical issues demanding a consistent regulatory response. Thus, policy-makers and industry representatives within a specific healthcare context need to develop a focused and far-reaching governance response. Consequently, in order to bring AI safely to the market and free the potential benefits expected, healthcare organizations, legislators, and technology developers face the so-called "ethical challenges" to manage the ethical governability plans.



Fig 10.2: Ethical Framework for Artificial Intelligence

Intelligent systems in the healthcare domain, empowered by artificial intelligence, offer transformative solutions tailored to specific medical needs, enhancing diagnostic accuracy, treatment personalization, and operational efficiency. However, their integration introduces a crucial dimension of sensitivity—an intangible yet powerful attribute that underscores their potential to significantly impact healthcare outcomes. While AI serves as a catalyst for innovation across healthcare and society at large, it simultaneously raises profound ethical concerns related to data privacy, algorithmic bias, and decision-making transparency. These challenges necessitate a comprehensive and

adaptive governance framework. Policymakers, industry leaders, and healthcare providers must collaborate to develop robust ethical guidelines that balance innovation with accountability. To harness AI's full potential while safeguarding public trust, ethical governability must become a foundational pillar—ensuring responsible deployment, continuous oversight, and a patient-centered approach in the evolving landscape of AI-driven healthcare.

10.3.1. Principles of Biomedical Ethics

The design, implementation and use of intelligent systems in sensitive clinical environments must be governed, particularly in ways that ensure the safety of all affected parties (Patel & Shah, 2021; Zhou & Wang, 2021). The lynchpin of effective governance of intelligent systems in sensitive clinical environments is a computational ethics system that ensures that intelligent systems are used ethically and in a manner that complies with relevant regulations.

The 10.3. Neural Computational Ethics Network is a high-level system to govern intelligent systems in sensitive clinical environments to ensure this. Deontological, utilitarian, virtue and principlist evaluations can be operationalised computationally to generate base value representations of the ethical content about how a decision affects each of the four kinds of value (intrinsically good to maximise, intrinsically bad to maximise, instrumental good to maximise, instrumental bad to maximise).

Ethical aspects of the validation of medical devices and of the computational modelling and simulation of the human body are discussed. Ethical validation activities are included within the safety assurance required by regulatory bodies certification. The challenge is to propose ethical requirements, in particular dealing with the impact of the device on end-users, that can be checked for automatic verification. Two scenarios are addressed: that of a computational model of the human ventricular electrical activity and of a reengineering of the impulse generator for the cochlear implant.

10.3.2. Ethical Theories Relevant to Intelligent Systems

In each design, implementation and usage of an intelligent system, a lot of value choices are made implicitly or explicitly. For the intelligent systems to be useful, trustworthy and resilient, those value choices have to be right. Ethics can be understood as a field of reflection, which methods to consider those value choices. Systems engineering, especially of complex intelligent systems, faces often underlying ethical questions long before the using contexts need to understand these questions. Therefore engineers and users of intelligent systems must acquire a discipline on exactly how to consider these underlying ethical questions and how to apply the then gained insights to their intelligent systems.

Engineering dealing with the design of systems that integrate sensors, actuators, and intelligence is rapidly carried into environments where people and machines co-habit and co-adapt. Systems engineering of these systems for sensitive clinical environments requires to accompany the engineer with an explicit understanding of the ethical challenges and governance principles that need to guide design, development and application of. This chapter is a step towards that need, presenting an explicitly engineered set of representations for supporting firstly a structured understanding of the complex interdependencies between, and the implications of, the ethical challenges and governance requirements of intelligent systems for use in sensitive clinical environments and secondly for a structured justification of how these complex interdependencies map into design principles, practices and obligations that form an ethical systems engineering approach.

10.4. Challenges of Implementing Intelligent Systems

Enhancing the capabilities to detect, diagnose, and treat health conditions through infrastructure and digital technologies is a complex and ethically sensitive endeavor. Ethical mistakes and abuses can easily undermine the public trust in such capabilities and create dissociation to healthcare activities. On the other hand, this kind of infection can also easily initiate social disturbances that could be difficult to manage. Development of infrastructure and capability in sensitive places like secure clinics, addiction care units, mental health service facilities, or any setting where personal activities are monitored and controlled has to be done with utmost professionalism from every angle, including those that can be exploited in an unethical way. For the case of secure clinics (both non-psychiatric and psychiatric; and with a focus on large-scale residential settings with state-like responsibilities), there are unique privacy concerns arising from not only the "not trusting" but also "not needing" the person to whom the information belongs. These are exacerbated in a setting where activities and many personal choices are also controlled. These combine with other complexities in the patient's privacy environment. Establishing what is perceived as an unsafe privacy environment can easily damage the patients trust, which is crucial in order for them to accept and participate in the effectively entirely voluntary care. Since all the legislation regarding patient registries, diagnoses, medication, restraints, and involuntary procedures is public, and the experience from shaping the legislation goes back to "ancient history", this developed a concurrent analysis of considerations regarding care, ethics, and the protection of patients personal data.

10.4.1. Data Privacy and Security

Healthcare organisations are generating and accumulating health data at an unprecedented scale. Deep learning AI technologies are ready to harness the data. Gaps in research are research that investigates the effects of deep learning AI technologies on value-based healthcare practices and research that proposes value-based principles for governing the deployment of deep learning AI technologies in health contexts. Possible consequences of the deployment of deep learning AI technologies are mapped via a value-based lens, considering the normative theories underpinning and informing such practices. Ten value-based principles are generated, designed as a practical tool to guide health systems, organisations, clinicians and developers in navigating the deployment of deep learning AI technologies. They contribute to an ethical application of AI. Each principle is exemplified by a set of practical considerations. Four key safeguards are first proposed to preserve trust and trustworthiness within the context of health insurance. This is an original contribution to the rapidly evolving debate on the governance of AI/ML technologies in health contexts. It problematizes the potential spillover effect that an AI/ML-based diagnosis and/or treatment system used in a care home setting could have on the healthcare of elderly people living in similar institutions and beyond. Such issues are not currently covered in the pertinent literature.

10.4.2. Bias and Fairness in Algorithms

Because algorithms define rules for the fast and standard evaluation of data, one would also anticipate that treatment guidelines and performance targets co-evolve with getting more information from the data. The objective would be to gear the intelligent systems for learning against the prior DGP at the comparative effectiveness research (CER) threshold while also empirically determining if NAM or the learned DGP can contribute better predictions or infer better guidelines at the respective CER threshold.

The updating of learned target predictions over time might use the incremental learning technique of "concept drift" adaptation. However, healthcare delivery, patient population and their outcomes and patient-specific behavior towards the treatment options and outcomes are much less controlled and much less stationary than the highly engineered setups of the usual uses of machine learning. Moreover, machine-learning systems might be deployed at any step in the clinical process, at different levels, in different locations and under diverse forms of economic and other incentives. Because of this, hidden biases in the learning data and evaluation of the deployed system can result in spurious correlations and zero-sum spurious effects. An initial treatment of how some of these concerns might generalize and also plague evaluation is undertaken here. Other aspects herein are primarily focused on potential biases and unfairnesses because

of a narrow or biased modeling of the mechanism by which the observed data arise, excluding other potential sources of bias and fairness.

10.5. Governance Structures for Intelligent Systems

10.5 Governance Structures For Intelligent Systems. In parallel to steps in development and maintenance, robust and comprehensive governance and regulatory structures must be put in place to govern the use of intelligent systems in healthcare settings. From a governance standpoint, intelligent systems may not consist exclusively of sophisticated algorithms and models, but be more accurately understood as broad socio-technical systems.

Effective development and roll-out of ethically legitimate ISs will also require the establishment of robust data governance practices. These practices must manage both the expected and unexpected consequences of these systems. A two-stage process is recommended. In the first stage, standard Information Governance (IG) practices should be implemented to provide a secure and compliant framework in which the research and development of intelligent systems can take place. These practices should leverage established and incoming regulations, whilst also attending to the granular needs of intelligent systems research.



Fig : AI through the looking glass an empirical study of structural social and ethical challenges in AI

10.5.1. Regulatory Frameworks

Regarding the regulatory frameworks underpinning the development of medical software and machine learning, it is first helpful to distinguish the two. Medical devices generally follow regulations concerning the design, manufacture, and performance of an item, to ensure it delivers the intended medical benefit and does so consistently and safely, within an intensive regulatory process. Clinical decision support software (CDS), also described as clinical AI or machine learning (ML), is defined differently as it is a software device requiring operator input, rather than a software information system. CDS may help steer clinical practice, but cannot therefore instruct treatments, the first reason being that all decisions and actions in healthcare must ultimately be taken by a human, based on their training and professional judgement. The difference between CDS given as software on one hand, and information resources and medical systems on the other, is arguably poorly understood, leading to misuse of CDS. This was swiftly stopped by regulatory authorities, but had potential to lead to treatments with unacceptable risk or consequence. As a 'prescription drug' informing, monitoring, cancelling or altering drug treatments of a patient without the direct intervention of a human, could cause injury or death. CDS of this type (a 'locked' implementation) of any sort is under the 'medical device' definition, categorising countless of current (and future) healthcare AI methods. Clarifying this difference, the overall policies, laws and guidance for medical software are inapplicable to CDS. On the other hand, the risks posed by CDS are also ill defined. A wealth of non-binding advice is readily available for both poor practice avoided, and as it is explained to consider ML in a self-determined frame of assurance. Nevertheless, any item's formal risk assessment must be based on a solid understanding of the system and its context, with fundamental terms of the risk framework. Using the same cybersecurity risk framework, various realistic (though perhaps unlikely) threats are proposed. Each scenario considers the effectiveness of potential mitigation strategies. By this analysis, the new unlicensed epilepsy prediction algorithm poses little additional risk to a potential future portable ECG device that would not be better mitigated by general good practice. A different eHealth solution, where prediction made by CDS falls outside the permitted action, is more concerning, though moderate improvement of both practicalities would close the risk. Finally, the widespread use of a poorly understood mental health care CDS model is hypothesised to introduce significant liability if the model performs as expected, or if used despite its limitations. This liability could have alarming ramifications for already fragile mental healthcare services, but requires further study to better quantify.

10.5.2. Institutional Policies and Guidelines

With this objective, the relevance of monitoring practices and the need for stakeholders to implement guiding principles to alleviate the occurrence of ethical challenges by providing guidelines in the form of a charter for healthcare institutions and artificial intelligence (AI) industry leaders. The development of this charter was conducted with a mixed-methods approach involving a qualitative discourse analysis of texts related to the AI ethics of health and well-being and an expert consensus-building process. Special attention is given to the need for a comprehensive framework for the explanation and interpretation of AI in healthcare driven by ethical concerns, emphasising the responsibilities of both healthcare organisations and industry partners. Establishing this framework is a necessity to ensure patient safety, enable informed decision-making by healthcare professionals, and foster patient trust in health services and AI technologies. In the early decades of the 21st century, considerable investment has gone into the development, acquisition, and implementation of Ai analytical tools in healthcare organisations, also due to the accelerated growth in concerns about health and healthrelated data among consumers and patients. On this changing ground, the paper analyses the complexity of a case study that concerns the implementation of an Ai-embedded radiology medical device in the challenging context of the oncological service of a busy, large-volume hospital. Aiming to advance the understanding of what the governance and ethical challenges such an implementation raises, the analysis identifies the simplification, deskilling, and reshaping of radiological work in the following subprocesses: semi-automatic labelling of references for the machine's training, persuasion of human image annotators, 'rational ignorance' with vast folders of images, and tunnel vision or 'just looking' by attending radiologists aided by the machine. These challenges foster an understanding of the necessity of a multiplex governance approach to foster contestability and transparency in a context where, paradoxically, these challenges are not easily remedied by more transparency, since the phenomena they involve are hidden by protocol and thus inherently opaque.

10.6. Conclusion

This initiative was supported through funding for a project that ran from 2013-2020 to study collaborative CO2 fixation mechanisms in anaerobic microbial ecosystems. The administration for this grant is now ending, and so this project is moving towards publication. CO2 fixation is of critical importance for life as we know it on Earth, and it forms the basis for almost all of the organic molecules in living cells. A metaorganism is a diverse community of many different microbial species harbored by a eukaryotic host in an obligate mutually beneficial relationship. Anaerobic archaea and bacteria conducting CO2 fixation were found to be protectively associated with the model

eukaryotic host plant at two geographically distinct hydrothermal vent systems. Imaging, metabolic intermediate analyses, gene expression profiling, and electron microscopy all shed light on CO2 fixer metabolic networks modulating metabolism within the partnership leading to growth and health benefit of the host. In recent years, a growing body of literature has documented the strategic importance of microbial CO2 fixers to the biosphere alongside the more familiar photosynthetic CO2 fixers, elucidating evolutionary lineages distinct from photosynthesis that nonetheless also allow cell growth and proliferation. Most of this work has come from study of pure cultures or stable symbiotic systems and almost always via aerobic laboratory methodologies, and as such open questions remain regarding CO2 fixation of the poorly understood microbes in the context of the majority of life's habitats. At hydrothermal vents where life subsists on inorganic chemical energy harboring low light levels and high CO2, uncertain role(s) of microbial processes in the carbon cycle have been implicated. Eight environmental studies were funded to shed light on how and why anaerobic mechanisms of CO2 fixation emerged given abundant pools of H2, CO2, and reduced transition metals. Here one such initiative with laboratory milestones spanning 7 years is described, at the center of this research focus deep-sea macrofauna-anaerobic microbial symbioses off the west coast of South America. Concomitant literature has noted these as having apparently low biomass and low primary productivity, thus merits frequent on their biology remain obscure. Since oxygen likely precluded microbial utilization of macrofauna-made niches, anaerobes arose to take advantage of eH2.Pool and pCO2 concs. that were still astro biologically interesting. The highly specialized roles of anaerobic microbes conducting photolithotrophic assimilation of CO2 and conducting a broad array of CO2dependent chemosynthetic responses in two vent-endemic species are described. Syntrophic coupling in surrounding to the chemoautotrophs was considered paramount given inferred toxicity of vent fluid hydrogen sulfide, while anaerobes ostensibly derived as a spillover from these experiments. Metabolic intermediate labeling experiments including the heavier isotope of CO2 that inferred pathways and partial regulation are described. Organ specific labeling demonstrated transfer of organic molecules from the anaerobic consortium into the host, and genome-resolved metatranscriptomic datasets indicated a modulation of holobiont metabolism. High resolution labeling experiments confirmed in situ activity consistent with ascribed pathways, and provided unique insights into spatial confinement of metabolite exchange at resolutions as fine as the eukaryotic host-cell level. Ultrastructural discoveries detailed CO2 fixer cell physiology with host proximities and included candidate novel microbially-encoded transport processes likely mediating energetically costly interspecies metabolite interactions.

10.6.1. Emerging Technologies

Artificial intelligence (AI) approaches have shown promise in enhancing treatment and diagnostics in healthcare areas while helping process reformed data more efficiently. However, the advancement of clever systems – particularly, AI on challenging systems in clinical settings that are ethically careless – crosses numerous hurdles, from technical testing to safety risks. One important obstacle is that of casual concerns. This paper begins with a brief background about AI in informal system free zones before delving into a recent partnership between a big UK hospital partnership and a well-known ticket manipulator with a strong focus on AI developers.

The adoption of AI in healthcare is the next exponential wave emerging, comparable to the COVID-19 virus trajectory or Moore's conventional polynomial law. This paper offers a snapshot of current work in this area, including the detection of diabetic retinopathy, pathology investigations, and control assistance for surgery. Sensitive problems with these systems in their activities are then recognized, such as screening the image data put through biases feed, which indicate chastely according classifications that inevitably construct some of the least used, commonly blackouts for safety reasons. Some licenses disclosed from this analysis are also discussed. In the wisdom of the common errors, a final reflection on the AI method support of the system free informal esthetically exploited environments are these results presented.

The shift towards artificial intelligence (AI) in sensitive settings is a notable cause for concern given their capacity to make choices and doings on their own. It is also rebutted as a version of a broader concern about the challenging systems' direction in public spaces, where again there are fears about the lack of sufferance to understand the norms against which these systems are acting, and about the difficulty of monitoring these systems for misuse.

References

- Chen, M., & Hao, Y. (2020). Machine learning for healthcare: A comprehensive review. IEEE Access, 8, 12345-12356.
- Convolution Neural Networks. In 2017 IEEE International Conference on Intelligence and Security Informatics (ISI).
- Kumar, A., & Singh, M. (2021). Role of machine learning in personalized medicine. Journal of Personalized Medicine, 11(2), 145-152.
- Lee, S., & Kim, H. (2020). Big data analytics in healthcare: A survey of applications and challenges. Health Information Science and Systems, 8(1), 1-9.
- Patel, V., & Shah, N. (2021). AI in pathology: Current applications and future prospects. Journal of Pathology Informatics, 12, 1-9.