MACHINE LEARNING IN MEDICAL DIAGNOSTICS – INADEQUACY OF EXISTING LEGAL REGIMES

Julianna Chan Lok Yin

Postgraduate Student, University of Hong Kong, School of Law Pokfulam, Hong Kong julianna.cly@gmail.com

Keywords: Machine Learning; Machine Learning Technology; Artificial Intelligence; Healthcare; BlackBox; Medical Negligence; Diagnostic Error; Liability; Litigation; Legal System; Legal Regimes

Abstract: Machine learning and artificial intelligence encompass huge potentials and benefits to society as a whole, especially in the healthcare industry. However, with such potentials also come the question of allocation of liability where things go wrong. As we continue to venture into the unchartered territories of machine learning technologies in healthcare, the urgency to create a suitable regulatory regime becomes ever so pressing. This paper analyses the non-suitability of existing legal regimes in regulating machine learning advancements in the healthcare sector, specifically in terms of the attribution of responsibilities when incorrect decisions are made using such technologies. It deals with four main issues: (1) the incompatibility between the tort of negligence and the nature of machine learning; (2) the possibility of regulating artificial intelligence advancements by imposing strict liability; (3) issues anticipated when trying to subject self-learning machines to a legal regime designed for human; and (4) potential options in regulating machine learning technologies in the healthcare sector.

1. Regulatory Landscape of Machine Learning in Healthcare

Machine learning («ML») encompasses huge potentials and benefits to society as a whole, and the last few years have seen a tremendous amount of attention at the intersection of ML and healthcare. As we continue to venture into the unchartered territories of ML technologies in healthcare, the urgency to create a suitable regulatory regime becomes ever so pressing.

According to a recent study published by Johns Hopkins University, medical diagnostic error is currently the third leading cause of death in the United States, and accounts for more than 250,000 deaths every year in the U.S. alone.¹ ML technologies and developments have the potential to reduce this alarming number, increase patient safety and improve clinical reliability. ML research within the healthcare sector is already growing rapidly and attracts a tremendous amount of attention from investors and start-ups alike.² However, the legal liability involving apportionment of responsibility which surrounds these novel developments remains to be an area of huge concerns. The high level of AI-human interaction inherent in these developments calls for adequate protection of the public from harm. However, if we are to benefit from the valuable potentials of ML in healthcare, a good balance ought to be struck in order to prevent the unnecessary hindering of developments.

Although the regulation of ML in healthcare is still a relatively unchartered territory, this does not mean that we should simply sit back and wait for developments. This paper aims to analyze the non-suitability of existing legal regimes in regulating machine learning advancements in the healthcare sector, specifically in terms of the

¹ MAKARY / DANIEL, Medical error – the third leading cause of death in the US, BMJ 2016;353;i2139.

² Some examples of this include Google's artificial intelligence laboratory DeepMind's partnership with the National Health Service of the United Kingdom, and IBM Watson's continued development in cognitive computing for healthcare.

attribution of responsibilities when incorrect decisions are made using these novel technologies. Our existing legal system is centered around the activities of human beings. As we move towards the new age where decisions are increasingly being made by intelligent machines, would the creation of a new legal system be a necessary step in order to regulate these novel, non-human intelligent entities?

2. The Role of Machine Learning in Medical Diagnostic

The information traditionally used by clinicians in making medical diagnosis are typically collected from patients' past medical history and results of their physical examinations. Most medical errors tend to occur during the process of integrating and interpreting information. In a study by the Institute of Medicine at the National Academics of Science, Engineering and Medicine, it was reported that diagnostic errors occur due to a variety of factors, including inadequate collaboration and communication, inefficient collaboration and integration of health information technologies, and the lack of a healthcare system which adequately support the diagnostic process.³ Another study by the Agency of Healthcare Research and Quality demonstrated that some common reasons for diagnostic errors include cognitive bias and incorrect applications of heuristics.⁴ Machine learning, in the most general term, is a field of artificial intelligence which uses statistical techniques to allow machines to «learn». They have the ability to access data, automatically learn and improve from information derived from the data without being explicitly programmed to do so. Machine learning technology has the capability to enhance evidence-based medicines and hence reduce the inefficiency in collaboration and integration of health information. This is achieved through characterizing and analyzing patterns within data to uncover associations that cannot simply be reduced into equations. ML algorithms can determine health risks and conditions much faster and much more accurately than human, capturing patterns unseen by the human eye from vast amount of data within a matter of seconds. There are already a number of existing applications of machine learning technologies to medical diagnosis. Some examples include the use of AI Chat-bots with speech recognition capability to identify patterns in symptoms to form potential diagnoses, algorithms to diagnose cancers through deep learning, and facial recognition software with machine learning elements to aid clinicians in diagnosing rare diseases.⁵ As MIT's principal scientist ANDREW MCAFEE has expressed, «if [machine learning] is not already the world's best diagnostician, it will be soon».⁶

Medical professionals are constantly required to make value judgments and decisions concerning the life of others, and the stakes are high. One misdiagnosis by clinicians could well lead to a life or death situation for the patients concerned. Hence, the law makes available the claims of professional malpractice to those who rely on the decision making of their qualified, skilled and trained healthcare professionals. The promises of artificial intelligence and machine learning in healthcare sounds good and well, but how would the questions of liability be answered when a doctor makes a decision by relying on faulty information and diagnosis provided by machine learning systems, and that decision subsequently result in loss or damage to a third-party patient?

There are a number of obstacles in the way of any attempts to impose liability on machine learning systems. Focusing on English law, this paper will first consider the incompatibility between the tort of negligence and nature of machine learning. Thereafter, the possibility of regulating these AI advancements by imposing strict liability will be addressed. Subsequently, the paper will move on to consider the issues anticipated when trying to subject self-learning machines to a legal system and regime designed for human.

³ Improving Diagnosis in Health Care, Institute of Medicine, The National Academics of Science, Engineering and Medicine, September 2015; see also KUMBA SENNAAR, Machine Learning for Medical Diagnostics – 4 Current Applications, January 2011, https://www.techemergence.com/machine-learning-medical-diagnostics-4-current-applications/ (all websites last accessed on January 23th, 2018).

⁴ Agency for Healthcare Research and Quality, Diagnostic Errors, available at https://psnet.ahrq.gov/primers/primer/12.

⁵ Kuumba Sennaar, n4.

⁶ Competing in the AI Economy: An Interview with Andrew McAfee, MIT Initiative on the Digital Economy, http://ide.mit.edu/news-blog/news/competing-ai-economy-interview-andrew-mcafee./.

Lastly, the paper will ponder on a number of potential options we might have in endeavoring to regulate machine learning technologies in the healthcare sector. It is worth pointing out that the focus of the paper is on machine learning developments for the more immediate future, rather than the stronger forms of artificial intelligence, namely superintelligence.

3. The Incompatibility between Tort of Negligence and Machine Learning

In most cases of medical malpractice and medical negligence, private claims of personal injury would be brought via the tort of negligence. However, the legality issues become much more complicated where the human doctor or physician cannot be held fully liable for a decision made using ML. The classic definition of negligence is as such – «the omission to do something which a reasonable man, guided upon those considerations which ordinarily regulate the conduct of human affairs, would do, or doing something which a prudent and reasonable man would not do».⁷ Visible from the statement is the emphasis on the human concept of reasonableness. This implies that negligence seeks the failure on part of human, as opposed to machines. The analysis below will demonstrate the inherent incompatibility between the concept of reasonable foreseeability of harm, and the nature of machine learning technologies.

A research paper published by the Oueen Mary University of London provides a very detailed and insightful analysis into the question of legal liability concerning ML technologies.⁸ It is the essence of this paper which forms the basis of the discussion which now follows. The four key elements which make up a claim of negligence are: (1) the existence of a duty of care, (2) breach of that duty, (3) causation, and (4) remoteness of damages. The doctor-patient relationship is one that automatically gives rise to a presumed legal duty of care. Therefore, the establishment of a duty of care is not an issue even when ML systems are used by doctor as part of the diagnostic process. A duty of care can equally be established between the producer of the ML technology and the doctor who relies on the information it provides, due to the fact that the producers would have undertaken to the users of technology the responsibility to ensure that no flaws exist in the algorithm. However, where a patient suffers from harm due to a decision made by a doctor relying on information from a ML system, and that doctor is able to discharge liability by proving that his reliance is reasonable, a gap of liability opens up. It was clearly held by the House of Lords in the famous English law case of Caparo v Dickman⁹, that those who simply give advice will only owe a duty of care to the subset of the society that is entitled to rely on the advice¹⁰ - for the purpose of our discussion, that would imply that a duty would only be owed by the producers to the medical professionals to whom the information is supplied to, but not to the patients who suffer harm from the faulty information indirectly. Hence, when incorporating ML into the diagnostic process, there will be instances where harm does transpire, but yet liability remains non-attributable.

The incompatibility becomes more profound as we go on to analyze the requirement for a breach of that duty of care, and this is where the «black-box» nature of ML technology becomes an issue. In cases of medical negligence, a doctor is deemed negligent where his conduct falls below the objective standard of care of an ordinarily competent person in his profession.¹¹ However, it is rather difficult to determine where this «objective standard of care» lies when ML systems are involved and opaque computational models are used to make health care related decisions. Assuming that a doctor relies on information generated from a ML system in making a certain diagnosis and hence owes a duty of care to the patient in question, he will only be in breach of that duty if it can be shown that his reliance on the information was not reasonable – and this is where the

⁷ Blyth v Birmingham Waterworks Co (1856) 11 EX 781, 784 per Alderson B.

REED CHRIS / KENNEDY ELIZABETH / SILVA, SARA, Responsibility, Autonomy and Accountability: Legal Liability for Machine Learning (October 17, 2016), Queen Mary School of Law Legal Studies Research Paper No. 243/2016.
U105112 KD 164

⁹ [1951] 2 KB 164.

¹⁰ N8, at 12.

¹¹ Bolam v Friern Hospital Management Company [1957] 1 WLR 582.

major incompatibility comes into play.¹² Some academics refer to the use of machine learning techniques in healthcare as «black-box medicine» due to the non-transparent nature of such algorithms.¹³ Essentially, once a ML algorithm is developed, the patterns that it generates and the decisions that it arrives at cannot be explicitly understood or stated, not even by the programmers of the algorithm.¹⁴ This therefore makes it arguable whether a doctor relying on information arrived at by a ML system can ever be said to have any meaningful understanding of the information at all, and whether he can in any way assess the reliability of such information.¹⁵ What implication does this have on the long standing legal concept of «reasonableness»? Would it be considered «reasonable» for a doctor to rely wholly on information generated from ML technologies because doctors or healthcare professionals do not have the technological expertise required to challenge and cast doubts on the reasoning of ML systems? If this is the case, would it mean that doctors' reliance on such information would always be «reasonable»? How the «black-box» nature of ML would and should affect the concept of standard of care is a question that will have to be left in the hands of judges and legislative law-makers.

On the other hand, the «black-box» nature of ML also makes matters very complicated when deciding whether the producers of the technology have failed to take reasonable care. Before going into the issue of opacity, would the requisite standard of care merely be that of an ordinarily competent software engineer/programmer? This would be the case under the current law as established in the case of Bolam.¹⁶ However, this does not seem to be a satisfactory standard given that the systems would have been designed for the purpose of being used in healthcare. With this in mind, it does make one wonder whether the standard of care ought to be raised higher, perhaps to one closer to the standard of a reasonably competent doctor.¹⁷ Whilst remaining at the requirement of the standard of a reasonably competent software engineer/programmer, the only way a court could assess whether the requisite standard has been reached would be to examine if a technological failure is due to insufficient pre-production trial and testing which failed to uncover flaws in the algorithm.¹⁸ This would require an understanding of the reason for the technological failure. For a court to even attempt to familiarize itself with such complex logical reasoning of a ML algorithm would require time and money that not many claimant would be able to afford. In addition, the opacity of ML systems means it will not be possible to pin-point when and what went wrong.

We therefore have a legal framework where doctors can escape liability by simply demonstrating that reliance on the ML technology is not unreasonable, and producers will be able to get off the hook if they can demonstrate that the algorithm learned how to make a decision through exposure to data, and that sufficient tests were carried out. Very heavy burden therefore rests on the claimants who wish to establish a breach of duty, both against the doctor who makes clinical decision, and against the producer of the ML technology.

Evidently, a claimant's claim would probably fail even before arriving at the issues of causation and remoteness of damages. If they do fortunately get there, they would find themselves in the face of yet more obstacles. For the purposes of establishing negligence, it is necessary for the claimant to prove that «but for» the relevant breach of duty, the injury would not have occurred.¹⁹ The remoteness test then requires the type of damage

¹² Ibid.

¹³ PRICE, Black Box Medicine, Harvard Journal of Law&Technology Volume 28, Number 2 Spring 2015, at 421. Available at SSRN: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2499885/, 2015.

¹⁴ For more details concerning the opacity of such algorithms, see JENNA BURREL, «How the machine «thinks»: Understanding opacity in machine learning algorithms» (2016) Big Data Society 1,2.

¹⁵ HART, When artificial intelligence botches your medical diagnosis, who's to blame? https://qz.com/989137/when-a-robot-ai-doctormisdiagnoses-you-whos-to-blame/, 2017.

¹⁶ N11.

¹⁷ N8 at p.17.

¹⁸ N8 at 13.

¹⁹ Cork v Kirby MacLean [1952] 2 ALL ER 402.

sustained by the claimant to be a reasonably foreseeable one.²⁰ The requirement of the «but-for» test can create problems – if the doctor is making his own judgment whilst at the same time relying on the ML algorithm's recommended diagnosis, it may not be possible to prove causation on balance of probability.

All in all, it is therefore evident that ML technologies present very difficult challenges for the law to impose liability on doctors and technological producers through the current tort of negligence. Court's current approach in establishing a breach of duty does not allow the black-box nature and complex interaction of different elements within ML technologies to be taken into account when determining liability.

4. Possibility of Imposing Strict Liability

Under UK law, it may be possible for the strict liability regime under the Consumer Protection Act 1987 to apply to ML technologies.²¹ This consumer protection regime imposes liability on producers or suppliers of products acting in the course of business²² where any damage is caused wholly or partly by a defect in a product, irrespective of any element of fault.²³ Two main difficulties in establishing such liability regarding ML algorithms lie in the legislation's definition of «product», and the «state of the art» defence available. If our diagnostic algorithm forms part of a medical equipment, then it would qualify as a product for the purpose of $s_1(2)$ of the CPA. But if the algorithm exists as an online service that healthcare professionals can subscribe to and use for example, it would not qualify as a «product», and hence the CPA regime will not apply.²⁴ On top of the fact that it would be difficult to demonstrate that the product is «defective» (as it can only be done by proving that the level of safety people are generally entitled to expect has not been reached), the «state of the art» defence in the CPA allows the producer to escape liability by showing that in the current state of the art of the industry, a reasonable producer would not have discovered the defect.²⁵ As aforementioned, the opacity of ML algorithms makes it impossible for even the programmer to understand or explain the logical reasoning involved in the generation of outputs. With this in mind, if the harm in question is caused by the decisionmaking process inherent in the ML algorithm which, conveniently for the producer, cannot be explained, then the producer will always escape liability under the CPA, which would once again leave us with the tort of negligence as the only available alternative.

5. Potential Options

If we simply sit back and wait for the law of negligence to evolve over time through case laws to deal with these issues, before we get there, the population would be exposed to the foreseeable risks and gaps of liability that come with the advancement of machine learning technologies. Even if courts do slowly respond to such need for change in the tort of negligence, other issues are still inherent when we try to subject self-learning machines to a legal system and regime designed for human. First of all, without a legal regime specifically designed to regulate these novel AI technologies, every time harm is alleged to be caused by AI systems, courts would be required to determine liability by unravelling and attempting to «understand» novel technologies whilst applying ill-fitting case law decided without having such technologies in mind.²⁶ This in turn creates uncertainty and unpredictability, things which ideally should not exist in a legal system. Secondly, the protection provided by court is remedial, not preventative. With the high level of AI-human interaction involved in ML

²⁰ The Wagon Mound (No 1) [1967] 1 AC 617, Privy Council (Australia).

²¹ N8 at p.5.

²² Consumer Protection Act 1987 s.4(1)(c).

²³ CPA 1987 s.2(1).

²⁴ N8, at p.6.

²⁵ CPA 1987 s.4(1)(e); see also ibid.

²⁶ Quinn Emanuel Urquhart & Sullivan, LLP, Artificial Intelligence Litigation: Can the Law Keep Pace with The Rise of the Machines?, https://www.quinnemanuel.com/the-firm/publications/article-december-2016-artificial-intelligence-litigation-can-the-lawkeep-pace-with-the-rise-of-the-machines/, 2016.

diagnostic systems, legislative regulations ought to be created to enhance protection of the public. Thirdly, liability could arise even when the ML system makes a «correct» decision. There are social policy factors embedded in the law that might not be considered as wholly objectively relevant but would be taken into account when a decision is made by a human. Where such a decision is made by a ML system untainted by such social policy concerns or unconscious biases, a decision that is correct from the algorithmic perspective could still nonetheless be incorrect in the eyes of the law. Finally, can justice be served under the existing legal regime when ML systems are involved? As aforementioned, under the existing law, there will be instances where a claimant harmed indirectly by ML systems is unable to bring a claim at all due to the inability to establish a duty of care. These are all problems that cannot be dealt with under the existing law.

Apart from waiting for the tort of negligence to evolve over time, there are a number of other possibilities for the gap in liability to be filled. The first and most novel would be to hold the ML system responsible. But the problem is, no matter how sophisticated these systems and technology may be, as long as the human involvement in the decision-making process remain evident and they remain as tools «used» by human, they are simply tools with no legal personhood.²⁷ To impose liability on ML algorithms and technologies would controversially require the creation of a new legal personhood for these AI entities. Before the day comes when we see fully autonomous superintelligent machines, there would not be sufficient justifications for the creation of such legal personhood. As a result, we are unlikely to see such drastic and unconventional move by the legal profession in the foreseeable future. Even when the day does come when we are faced with superintelligent machines, any potential justifications for the creation of a new legal personhood would still face the obstacle that «justice» in the way we know it would not be served. In cases of medical negligence, patient compensation can achieve two objectives – to transfer cost of the harm away from the patient, and to act as deterrent to give incentive to the healthcare professional to avoid mistakes in future.²⁸ It is apparent that it would not be possible to seek monetary compensation from ML algorithms nor to achieve any deterrent effect on them.

Another possibility might be to devise a liability regime specifically for ML systems. This may be done by holding the programmer/producer responsible, or alternatively by holding the organization which runs the system accountable. Considering first the difficulties with imposing strict liability on the programmer/producer, due to the fact that such algorithms are often designed and produced by not just one but a group of individuals, difficulties would arise in attempting to pin down the individual responsible for the part of the algorithm which caused damage. Also, the temporal and physical distance between research, design, and implementation can often preclude any awareness of the ultimate usage of such an algorithm.²⁹ Take IBM Watson for example – despite the fact that it is now a pioneer in the field of AI related clinical decision making, its original design was meant for participation in a quiz show. For this reason, the injustice that could result from the imposition of liability on programmers/producers who may have designed the algorithms without any clues as to its final purpose and usage which ultimately causes the damage, may just outweigh the justification for imposing liability on them. In addition, an obvious problem with such allocation of responsibility would be its deterrent effect on future developments.

It may also be an option to impose liability on the organizations that run the systems. For the purpose of our discussion this would be the hospital or clinic. A potential way this may be achieved is through the creation of a new and unique variant of vicarious liability and agency law. This would have the benefit of providing a clear target for retribution. However, the hindering effect on future developments would again be strong as organizations may cease to invest in such developments or to use them due to the fear of liability.

²⁹ N26.

²⁷ VLADECK, Machines without principals: liability rules and artificial intelligence, Washington Law Review Vol. 89, 117-150, at 121.

²⁸ FENN PAUL / GRAY ALASTAIR (2002), Deterrence and Liability for Medical Negligence: Theory and Evidence.

The vast benefits that ML technologies can bring to the field of medical diagnostic is something that the society needs and demands, as the increasing adoption of ML technologies in the field will certainly help to reduce the alarming number of deaths caused by medical diagnostic errors. The importance of a good and well-balanced regulatory regime for ML technologies cannot be stressed enough. Law-makers ought to put the creation of such regulatory regime at the forefront of their minds, and to devise a way of filling the existing gaps in liability without hindering the developments of the potentially life-saving technologies.

6. References

Blyth v Birmingham Waterworks Co (1856) 11 EX 781.

Bolam v Friern Hospital Management Company [1957] 1 WLR 582.

Caparo v Dickman [1951] 2 KB 164.

CASEY C., BENNETT / KRIS, HAUSER, Artificial Intelligence framework for simulating clinical decision-making: A Markov decision process approach, Artificial Intelligence in Medicine 57 (2013), 9–19.

CHEN, JONATHAN, Machine Learning and Prediction in Medicine – Beyond the Peak of Inflated Expectations, New England Journal of Medicine, June 2017.

Competing in the AI Economy: An Interview with Andrew McAfee, MIT Initiative on the Digital Economy, http://ide.mit.edu/news-blog/news/competing-ai-economy-interview-andrew-mcafee/, 2018.

Consumer Protection Act 1987.

Cork v Kirby MacLean [1952] 2 ALL ER 402.

Diagnostic Errors, Agency for Healthcare Research and Quality, https://psnet.ahrq.gov/primers/primer/12/, 2019.

FENN, PAUL / GRAY, ALASTAIR, (2002), Deterrence and Liability for Medical Negligence: Theory and Evidence.

Institute of Medicine, The National Academics of Science, Engineering and Medicine, Improving Diagnosis in Health Care, September 2015.

SINGH, JATINDER / WALDEN, IAN / CROWCROFT, JON / BACON, JEAN, Responsibility & Machine Learning: Part of a Process, https://ssrn.com/abstract=2860048 or http://dx.doi.org/10.2139/ssrn.2860048/, 2016.

BURREL, JENNA, «How the machine «thinks»: Understanding opacity in machine learning algorithms» (2016), Big Data Society 1,2.

KUMBA, SENNAAR, Machine Learning for Medical Diagnostics – 4 Current Applications, Jan 2011, available at https://www.techemergence.com/machine-learning-medical-diagnostics-4-current-applications/, 2011.

MARUTHAPPU, MAHIBEN, Debate and Analysis, Artificial Intelligence in medicine: current trends and future possibilities, British Journal and General Practice, March 2018.

MAKARY, MARTIN/ MICHAEL, DANIEL, Medical error – the third leading cause of death in the US, BMJ 2016;353;i2139.

REED, CHRIS / KENNEDY, ELIZABETH / SILVA, SARA, Responsibility, Autonomy and Accountability: Legal Liability for Machine Learning, Queen Mary School of Law Legal Studies Research Paper No. 243/2016, https://ssrn.com/abstract=2853462/, 2016.

Nuffield Council on Bioethics, The Collection, Linking and Use of Data in Biomedical Research and Health Care: Ethical Issues, http://nuffieldbioethics.org/wp-content/uploads/Biological_and_health_data_web.pdf/, 2015.

HART, ROBERT, When artificial intelligence botches your medical diagnosis, who's to blame?, https://qz.com/ 989137/when-a-robot-ai-doctor-misdiagnoses-you-whos-to-blame/, 2017.

SHERIFF, KATHERINE, Defining Autonomy in the Context of Tort Liability: Is Machine Learning Indicative of Robotic Responsibility?, https://ssrn.com/abstract=2735945/ or http://dx.doi.org/10.2139/ssrn.2735945/, 2015.

Taming the Machine?, The SMU Blog. Available at: http://blog.smu.edu.sg/masters/llm/taming-the-machine-liability-responsibility/, 2016.

The Wagon Mound (No 1) [1967] 1 AC 617, Privy Council (Australia).

VLADECK, Machines without principals: liability rules and artificial intelligence, Washington Law Review Vol. 89, 117-150

W. NICHOLSON PRICE II, Black Box Medicine, Harvard Journal of Law&Technology Volume 28, Number 2 Spring 2015, at 421, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2499885/, 2015.