INTERNSHIP REPORT

MACHINE LEARNING IN HEALTHCARE: AN EXPLORATION OF THE POTENTIAL RISKS AND THE FORMS OF GOVERNANCE AND REGULATION MANAGING THESE RISKS IN EUROPE

Merijn van Leeuwen

STUDENT NUMBER
2557671

COMMISSIONING PARTY
Nictiz

ECTS + COURSE CODE SPECIALIZATION
27 + AM 1118

DATE
29-08-2019

VU SUPERVISOR
Teun Zuiderent-Jerak

ATHENA SCIENCE SHOP

VU VRIJE UNIVERSITEIT AMSTERDAM

Colophon

Title: Machine learning in healthcare: an exploration of the potential risks and the forms of governance and regulation managing these risks in Europe.

Key words: Artificial Intelligence in Medicine, Deep Learning, Computer-Aided Diagnosis, epistemic uncertainty, reflexive governance
Word count: 14984

Author: Merijn van Leeuwen
Commissioning organization: Nictiz
On-site supervisor: Henk Hutink
VU supervisor: Teun Zuiderent - Jerak

Februari 5, 2019 – June 29, 2019
Course code: AM_1219

This internship report was established in cooperation with the Athena Science Shop.

Athena Institute
Faculty of Science
VU University Amsterdam
De Boelelaan 1085
1081 HV Amsterdam
The Netherlands

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
ANI	Artificial Narrow Intelligence
AIME	Artificial Intelligence in Medicine (healthcare)
ANN	Artificial Neural Networks
CAD	Computer-aided diagnosis
CSMT	Covenant for Specialist Medical Technology
DL	Deep Learning
DL-CAD	DL-Based computer-aided diagnosis
EHR	Electronic Health Record
FDA	U.S Food and Drug Administration
HCP	Healthcare Professional
MDD	Medical Device Directive
ML	Machine Learning
ML-based SaMD	Machine learning-based software as a medical device
MDR	Medical Device Regulation
NB	Notified Body
RVS	Council for Health and Society
STS	Sociotechnical Systems
SSI	Semi-structured interview
TTM	Technology Triangle Model
USI	Unstructued interview
W.	Workshop

TABLE OF CONTENTS

SUMMARY

Introduction & Contextual background

Artificial Intelligence (AI) in medicine (AIME) may ameliorate the issues which healthcare is facing by "transforming healthcare from art to science". Central to the insurgence of AIME is the usage of self-learning Machine Learning (ML) or Deep Learning (DL) algorithms, applicable to many purposes within healthcare. Within the AIME, the field of radiology is a frontrunner, with novel computer aided diagnosis systems showing much potential. Despite these conceivable benefits of ML-based Software as a Medical Device (ML-based SaMD), there is a lack of knowledge regarding the risks and how these risks are being controlled by governance or regulation. Therefore, this study explores the risks, governance and regulation in Europe, aiming to provide relevant information for the establishment of suitable governance.

Conceptual framework

Risk management in healthcare aims to "increase the probabilities and impacts of positive events and to decrease the probabilities and impacts related to adverse events". However, in novel fields such as AIME there is a large degree of epistemic uncertainty, where both the risks and their probabilities are often unknown due to a lack of theoretically obtainable knowledge. Proactive risk management may facilitate safe operation under epistemic uncertainty, because it stimulates actors to communicate, produce knowledge and exert (pro)active oversight. Furthermore, reflexive governance theory shares core principles with proactive risk management and may help implement proactive risk management within society. Finally, a sociotechnical systems perspective was used to map risk factors from a perspective which recognizes the interrelatedness of technical, human and organisational factors in specific contexts.

Methods

A literature review was conducted in order to identify the risks and governance methods of ML-based SaMD. Furthermore, both semi-structured and unstructured interviews, brainstorm sessions and a workshop were held to elaborate on these results. Included stakeholder groups were healthcare professionals (HCP), developers and governance actors. Risks were mapped in our sociotechnical ML square model containing structure- and process elements, adapted from the technology triangle model for evaluating the safety of health IT. Finally, by using a fit-gap analysis, identified risk were compared with governance methods.

Results

Most of the risks factors which pertained to structure elements within the model were ML-specific risks such as reward hacking, gradual decay of ML performance, temporality problems, black-box decision making and algorithmic bias. Risks pertaining to process elements are strongly interrelated. For example, it was found that there is no ML-specific framework to accommodate continuous learning, potentially further increasing the risk of ML performance decay. Furthermore, a combination of black-box decision making processes and lack of relevant statistical and theory of mind knowledge among users may result in unsafe usage or out of sample input, since this knowledge may be required to gain insight into errors and avoid out of sample input. Furthermore, the black-box decision making may further enhance the opacity of defining liability in case of malpractice. The lack of

standardization of criteria to evaluate aspects such as data bias or algorithmic performance makes both development and evaluation of ML-based SaMD difficult. Further compounding this problem is the lack of data availability in Europe, leading developers to procure data from non-EU states, even though this data may contain biases for which evaluation standards do not exist. Regarding evaluation criteria, a recurrent these is a need for standardized protocols to evaluate the performance of ML algorithms and detect data bias. We identified three main data criteria: (1) data representativeness, (2) required size of datasets and (3) management of evaluation datasets.

Discussion

Given the degree of complexity, epistemic uncertainty and speed of development in the field of ML-based SaMD, a form of governance that promotes proactive risk management, increased communication and collaboration resulting in continuous reflexive development is recommended. We recommend the addition of a ML-specific framework to the MDR, while other EU member-states may need to adopt a regulatory body to regulate human-technology interactions. Furthermore, to enhance proactive risk management, the establishment of a frontrunner network containing influential frontrunners from all relevant stakeholders may be beneficial to aid in holistic and reflexive goal creation, agenda setting and experimenting as the technology advances. Integrating this perspective, we recommend that a ML-specific regulatory framework is added to software standards. Furthermore, standards to evaluate data bias, algorithmic bias and performance and algorithmic or output transparency should be established in an ongoing manner. Finally, medical education systems may need to teach relevant statistical and theory of mind knowledge, liability frameworks are clarified, and steps are taken to make European data available.