



Slides and notebooks: <https://ml4ns.github.io>

Notes by: *Nan Fletcher-Lloyd, Payam Barnaghi*

## **Lecture 9. Ethical considerations and responsible machine learning**

In lecture 9, we look at ethical considerations and the importance of responsible machine learning.

The goal of developing machine learning models and systems in healthcare is to deploy them in real-world settings to aid with patient care. These systems can be used as decision-support tools by keeping the human in the loop. Therefore, prior to deployment, it is crucial to investigate the models/systems with regard to trustworthiness, reliability and robustness. It is also important to understand user perception of these models/systems.

In this lecture, we discuss how the way in which a model is trained can affect the appropriateness of a model for a given problem, introducing the approaches of online and offline learning. We further expand on the use of continual learning methods by which to emulate the process of human learning, and we also introduce the forgetting problem of such models.

We consider the importance of investigating the suitability of a dataset for its planned analysis with regards to developing clinically applicable solutions (how and when the data was collected, and the purpose of the model such data will be used to train).

We then review ethical implications of machine learning methods with regards to algorithmic bias in which a model is trained on data likely to be influenced by many facets of social inequality, particularly those who contribute the most data. Here, we discuss the importance of having a robust evaluation of the models backed up by qualitative analyses. We further stress the importance of transparency in the reporting of new works and introduce the notion of explainability in our models.

We also note the importance of deploying and maintaining such models in an ethical, legal and morally responsible manner, including keeping up to date on regulatory requirements. Figure 9.1 illustrates a roadmap for developing and deploying effective machine learning systems in real-world settings.



Figure 9.1. A roadmap for deploying effective machine learning systems in healthcare applications

Source: Wiens, J., Saria, S., Sendak, M. *et al.* Do no harm: a roadmap for responsible machine learning for health care. *Nat Med* 25, 1337–1340 (2019).

Finally, we discuss the challenges posed by the collection and analysis of data, particularly regarding privacy issues.

This topic is assessed in conjunction with Topic 8 (Real-world Applications of Machine Learning in Neuroscience) and will evaluate your understanding of the importance of responsible machine learning.

Drawing on what you have learned across this module so far, we ask you to review a given paper (1 page maximum, font size 11, with 2cm margins) focusing on the different aspects of machine learning used in the presented work, particularly with regards to the limitations and potential ethical challenges posed using such methods and models, and exploring how such issues can be overcome, e.g., (can other models be used that would be more applicable and why or can different metrics be used to evaluate a model that might be more appropriate?).