Appendix 1

Key terminology

Artificial Intelligence: 'an umbrella term covering a wide range of efforts to computationally reproduce complex human capabilities such as language use, vision and autonomous action.'(Caldwell et al. 2020). Most AI as we understand it describes a system which integrates a number of these capabilities into a single system which significantly outperforms that of human capabilities.

Machine learning (ML): what most would understand as a 'bottom-up' subset of AI. ML describes AI methods based on algorithms which discover and replicate patterns in data; decision criteria are therefore learned as more data is fed into them. This makes them by design, unpredictable and dynamic systems. Distinctions within ML:

- **Supervised and unsupervised learning:** systems which are, or are not explicitly told correct answers (labels)
- **Training data:** the data that an ML system learns its task and relevant patterns from- the purpose of training data is to ensure that an ML system can go on to generalise and interpret new, unseen data.
- **Deep learning:** ML that makes use of large complex, interconnected network of simple computational units connected together. Each unit usually performs a simple computational task which is then integrated with other units to create a complex system.
- **Reinforcement learning:** an ML approach whereby an autonomous software agent can perform actions on a system with the intention of maximising a particular goal or reward
- **Natural Language Processing:** AI applied to understanding, generating, and translating human language.

Bias: bias is a topic ethical issue in AI, whereby a system incorrectly learns based on flaws in real world data which may, for example, reflect systemic, social, or economic biases.