Code of Ethics for Machine Learning

This is a code of ethics from a newly created ML association. This ML code of ethics applies to all members of the association and can be seen below. Each point is then emphasized in paragraph format.

Article 1
Any model that influences human decisions must be explainable.

Article 2
No model will discriminate with intentional bias or harm.

Article 3
A model must obey the law of the land where it is deployed.

Article 4
All data used to build a model must be collected with at-minimum implied consent.

Article 5
Deployed models must remain robust through consistent human involvement.

Article 1
Machine learning models are inherently complex. Since models often handle sensitive decision-making tasks, they should be built in a way that uncovers the ‘reasoning’ behind their decisions. Itransitions *Explainable AI and the Future of Machine Learning* article explains the importance of this concept with a focus on benefits from these explanations, with the primary beneficiary being businesses. They are able to make more-informed decisions and provide transparency for regulatory oversight.

Article 2
Bias perpetuates injustices currently present in the world. This occurs through the use of either biased data or biased parameters used to model the data. In *Patterns, Predictions, and Actions: A story about machine learning*, the book demonstrates how common learning techniques like the perceptron algorithm can only make generalizations from data it sees. In order for a model to prevent discrimination, it must knowingly be fed data that does not already have these inherent biases. If those biases are present, the model parameters should be changed in order to weight against the bias for a more equitable model.

Article 3
Obeying the law of the land is necessary in order to deploy a well-recieved model. Different cultures and countries have different ethical gauges by which to measure models. Considering these differences, it is best to approach compliance through moral relativism. Moral relativism goes against Kant’s universal moral approach through his categorical imperative, as mentioned in *Anscombe’s Summary of Kantian Ethics*. However, remaining morally relative is a utilitarian ideal that maximizes happiness by conforming to the norms of the culture in question.

Article 4

Data is required to train models. This data often comes from humans and how they interact with the world around them. Ferryman's talk *Reframing Data as a Gift* provides insight into how human consent is necessary. It is possible to use human data in a way that is adverse to the individuals providing the data. Considering this, it is important that humans be allowed to provide or revoke consent. When consent is implied, that consent is obtained through public data from public spaces.

Article 5

Human interaction is necessary to correct the course of a model's self-learning. Vallor and Bekey's article *Artificial Intelligence and the Ethics of Self-Learning Robots* describes situations where models can be left in a perpetuating automation bias. Without human involvement, biases obtained by the model could propagate indefinitely. In order to keep models robust to biased influence, human involvement remains necessary well after deployment. This involvement is also necessary to keep models opaque and explainable as they learn from new information.

Code Does Not Address

Anonymity should not be mandated by this code of ethics in ML. There are situations where anonymity cannot be maintained or guaranteed in order to conform to local laws or customs which contradicts article 3. In addition, useful and relevant data should not be automatically deleted or destroyed after a given period of time. Data is necessary to train models. Without an abundant amount of reliable data models will not improve over time, contradicting article 5.