Week 9 Glossary

Algorithmic accountability

A concept where model owners and developers are accountable for the decisions that their machine learning systems make.

Allocative harm

A discriminatory system that withholds certain opportunities, freedoms, or resources from specific groups.

Ethical AI

The subfield of AI that studies model fairness and accountability

Fairness

The various attempts at correcting algorithmic bias in automated decision processes based on machine learning models.

Model deployment

The process of using a machine learning model in a production environment where it can be used for its intended purpose

Representational harm

A type of harm where a system reinforces negative stereotypes along the lines of identity and protected class.

Societal failures

Failures which happen when a machine learning model produces unintended discrimination or disparate impact and lacks accountability. It is where failure exists between the model's owner and the social context in which the model operates.