Course Glossary: AI Ethics

- **AI Ethics:** This field focuses on ensuring that artificial intelligence is developed and used in ways that align with human values, fairness, and accountability.
- **Bias in AI:** Unfair influence on AI decisions caused by biased data or assumptions, leading to discriminatory outcomes.
- **Fairness:** An ethical principle requiring AI systems to treat individuals equitably, without favoritism or discrimination based on attributes like race, gender, or location.
- **Privacy-Personalization Paradox:** The tension between offering personalized AI experiences and protecting individual privacy.
- **Autonomy-Control Dilemma:** The ethical trade-off between allowing AI to act independently and maintaining human oversight and control.
- **Black Box:** A term for AI systems where the internal decision-making is not visible or understandable, even if inputs and outputs are known.
- **Explainable AI (XAI):** A branch of AI that emphasizes making AI models understandable to humans by revealing how decisions are made.
- **AI Accountability:** The process of ensuring that humans remain responsible for AI decisions and outcomes—AI is a tool, not a scapegoat.
- **SHAP (SHapley Additive exPlanations):** An explainability method that quantifies the contribution of each input feature to a model's prediction.
- **LIME (Local Interpretable Model-agnostic Explanations):** A technique that explains individual predictions by approximating complex models locally with simpler, interpretable ones.
- **Ethical Framework:** A structured set of principles that guide the development and deployment of AI systems to ensure they align with values like sustainability, equity, and transparency.
- **Data Ethics:** A subset of AI ethics focused on collecting, storing, and using data in a way that respects privacy and fairness.
- **Continual Improvement:** Iterating and refining AI systems to enhance ethical performance as new data, challenges, and technologies arise.