Part 3: Ethical Reflection

Deploying a predictive model in a company raises important ethical considerations, particularly around potential biases in the dataset. Bias can occur if certain teams, departments, or demographic groups are underrepresented in the historical data used to train the model. For instance, if past promotions favored a particular group or some teams had fewer recorded performance reviews, the model may inherit these patterns and make predictions that unintentionally disadvantage underrepresented employees. This can lead to unfair outcomes, reduced trust in AI systems, and even legal or reputational risks for the company.

Tools like IBM AI Fairness 360 can help identify and mitigate such biases. They provide fairness metrics to detect disparities in predictions across groups and offer mitigation techniques, including reweighting data, adjusting sampling methods, or applying algorithmic interventions to reduce discriminatory effects. By integrating these tools, organizations can proactively monitor and adjust their models, ensuring that AI-driven decisions are more equitable. Ultimately, ethical reflection and bias mitigation improve both the reliability of predictive models and fairness in workplace decision-making, promoting a more inclusive and responsible use of AI in business environments.