

Ethical Considerations for the Saliford Motors Project

When using data to protect employee behaviour, we must address 3 primary ethical risks:

1. Privacy and Data Anonymisation

- **The Risk:** Survey data can be highly sensitive. If a manager can trace a “low satisfaction” score back to a specific employee, it could lead to retaliation.
- **The Solution:** All Personally Identifiable Information (PII) must be removed. We should only analyse data at the aggregate or departmental level unless a high-level intervention is strictly necessary.

2. Algorithmic Bias and Fairness

- **The Risk:** If the historical data contains biases (e.g., certain departments are unfairly given lower evaluations), the model will learn and repeat these biases. This could unfairly flag specific groups as “risky”.
- **The Solution:** We must regularly audit the model for **Fairness Metrics**. For example, we should check if the model predicts turnover at a higher rate for a specific salary level or department, regardless of their actual performance

3. The “Predictive Firing” Trap

- **The Risk:** Stakeholders might be tempted to use the model to “pre-emptively” fire employees who are flagged as likely to leave, rather than helping them.
- **The Solution:** The model’s purpose must be defined as **proactive support**. The goal is to identify systematic issues—like the “240-hour burnout wall”—and fix the environment, not to penalize the individual.

Key Ethical Question: “Does this analysis benefit the employee?” In this project, the answer is **yes**, as the goal is to reduce burnout and improve work-life balance through better workload management.