AI Governance & Risk Checklist

Use this checklist to ensure responsible AI deployment, covering key principles like transparency, fairness, accountability, security, and privacy.

1. Transparency & Explainability

- [ ] Have we defined the level of explainability required for this use case?
- [ ] Can the model's decisions be traced back to the data that influenced them (e.g., via RAG)?
- [ ] Is there a process for communicating the model's capabilities and limitations to end-users?

2. Fairness & Bias

- [ ] Have we audited the training and source data for potential biases?
- [ ] Have we implemented tools to detect and measure bias in the model's outputs?
- [ ] Is there a diverse team overseeing the project to identify potential blind spots?

3. Accountability & Human Oversight

- [ ] Is there a clear owner for the AI model and its outcomes?
- [ ] For high-stakes decisions, is there a "human-in-the-loop" for final approval?
- [ ] Are all model predictions, user prompts, and outputs logged for auditing?

4. Security & Privacy

- [ ] Does the solution comply with all relevant data privacy regulations (e.g., GDPR, CCPA)?
- [ ] Is sensitive data (PII) properly anonymized or masked before use?
- [ ] Has the system been tested for AI-specific vulnerabilities like prompt injection?

5. Reliability & Safety

- [ ] Has the model been rigorously tested in a simulated environment before deployment?
- [ ] Are there fail-safe mechanisms in place if the model produces harmful or nonsensical outputs?
- [ ] Is there a continuous monitoring system to detect model drift or performance degradation?