Part 4: Ethical Reflection

Reflection on My Project: "SafePing" – An AI Safety Alert Tool for Women

"SafePing" is a safety alert system designed to protect women at risk of violence or abduction by using AI to send discreet emergency alerts and location data. As I continue developing this project, I will ensure it adheres to ethical AI principles in the following ways:

1. **Privacy and Consent:** Users will explicitly consent to data collection (e.g., GPS) and be informed how their data will be used and stored. Location data will be encrypted and only shared with authorized emergency contacts during alerts.

2. **Bias and Fairness:** I will test the model with diverse user data (e.g., race, language, location) to ensure it works equitably across all demographics and regions. I will use open datasets and fairness auditing tools (like AI Fairness 360) to detect and reduce any algorithmic bias.

3. **Transparency:** Users will be clearly informed how the AI system functions, especially how and when it triggers alerts. I will publish documentation and ethical impact assessments.

4. **Accountability:** I will ensure human oversight in the emergency response chain and create a reporting mechanism for users to flag false alerts or issues.

By integrating these principles from the start, I aim to build SafePing as a responsible, inclusive, and safe AI tool that serves all women fairly.

Bonus Task : Ethical AI Use in Healthcare – Policy Guideline

Ethical AI Use in Healthcare: Internal Guideline Document

Purpose:
To ensure responsible, transparent, and fair use of AI in clinical and operational healthcare environments.

---

1. Patient Consent Protocols

- **Informed Consent Required:** Patients must be informed in plain language whenever their data is used to train or power AI systems.
- **Opt-In Model:** AI use for diagnostics, treatment recommendations, or administrative decisions requires active patient consent.
- **Withdrawal Rights:** Patients may opt out of AI-assisted care or data sharing at any point without penalty.

---

2. Bias Mitigation Strategies

- **Representative Datasets:** Ensure AI systems are trained on diverse data covering all ages, genders, ethnicities, and conditions.
- **Routine Fairness Audits:** Conduct audits (e.g., with AIF360 or Fairlearn) every 6 months to evaluate bias in predictions or treatments.
- **Error Disparity Tracking:** Monitor misdiagnosis rates across demographic groups and adjust models to reduce disparities.

---

3. Transparency Requirements

- **Explainability:** AI systems must provide clear, understandable explanations for any diagnosis or recommendation made.
- **Documentation:** Maintain technical documentation including training data sources, algorithm logic, and evaluation metrics.
- **Disclosure to Patients:** Inform patients when AI is used and how it impacts their care, including any known limitations.

Compliance and Oversight

- **Ethics Board Approval:** All AI projects must undergo review by a multidisciplinary ethics committee.

- **Incident Reporting:** Create a system for reporting unintended harms or algorithmic errors for investigation and correction.