# Project: AI Mental Health Support Chatbot

*Goal: Provide accessible, low-stigma emotional support and coping strategies*

## Ethical Principles & Mitigation Strategies

1. **Fairness & Bias Mitigation**
   - **Risk:** Training data may underrepresent marginalized groups
   - **Action:**

```
# Strategy implementation
dataset = curate_data(sources=[
    "inclusive_mental_health_forums",
    "cross_cultural_clinical_studies"
])
audit_results = bias_detection_tool(model,
                                    protected_attribu
professional_review = engage_experts(
    domains=["LGBTQ+_health", "cultural_psychology"]
)
```

2. **Transparency & Explainability**
   - **Risk:** Users might overtrust the bot's advice
   - **Action:**

```
# System message template
DISCLAIMER = """
⚠️ I am not a licensed therapist.
My suggestions are based on public resources up to 20
For clinical help, contact a professional.
```

```
"""

# Explainable response pattern
def generate_response(user_input):
    explanation = f"I suggested deep breathing becaus
    return f"{DISCLAIMER}\n\n{explanation}"
```

3. **Privacy & Data Security**

- **Risk:** Exposure of sensitive user data
- **Action:**

```
# Data handling protocol
def process_chat(user_input):
    anonymized_input = remove_identifiers(user_input)
    encrypted_data = aes_encrypt(anonymized_input)
    if not user_consent:
        schedule_deletion(encrypted_data, delay=HOURS
```

4. **Accountability & Human Oversight**

- **Risk:** Failure to escalate crises
- **Action:**

```
# Emergency protocol
CRISIS_KEYWORDS = ["kill myself", "end it all", "suic

def check_crisis(user_input):
    if any(keyword in user_input for keyword in CRISI
        connect_human_support()
        show_crisis_resources(local=True)
        log_incident(severity=CRITICAL)
```

5. **Beneficence & Societal Impact**

- **Risk:** Normalizing avoidance of human therapists
- **Action:**

```
# Usage boundaries
MAX_SESSIONS = 3

def check_usage(user_id):
    if session_count(user_id) > MAX_SESSIONS:
        return "Regular chats with me aren't a substi
```

## Implementation Roadmap