# **Fawzooz AI - AI Impact Assessment (AIA) Report**

Report ID: [Unique Report ID, e.g., AIA-AIS-001]

AI System Name: [Name of the AI System]

System Version: [Version of the System Assessed]

Date of Assessment: [Date]

Assessment Team: [List of names and roles of people who conducted the assessment]

### 1. Executive Summary

*(This section should be completed last. It provides a high-level overview of the assessment's key findings and recommendations for senior management and the AI Governance Committee.)*

A brief summary of the AI system's purpose, the main potential benefits and risks identified, and the overall conclusion of the assessment (e.g., "Proceed with recommendations," "Requires significant redesign," "Halt project").

### 2. AI System Description

#### 2.1. Purpose and Intended Use

* **What is the primary goal of this AI system?** (e.g., To provide users with personalized coping strategies for mild anxiety.)
* **Who are the intended users?** (e.g., Adults over 18 experiencing symptoms of stress and anxiety, not intended for clinical diagnosis.)
* **What specific decisions or outputs does the AI generate?** (e.g., Recommends specific meditation exercises, suggests journaling prompts, provides educational content.)
* **In what context will it be used?** (e.g., Via a mobile application, as a tool to supplement, not replace, professional therapy.)

#### 2.2. System Functionality and Technology

* **High-level description of the AI model(s) used:** (e.g., A transformer-based language model for chat interaction, a collaborative filtering model for content recommendation.)
* **Key data sources for training and operation:** (e.g., Anonymized user interaction data, curated clinical knowledge base, user-provided mood logs.)
* **Degree of Autonomy:** (e.g., The system operates autonomously but includes clear pathways for users to escalate to a human support agent.)

### 3. Impact Assessment and Analysis

*(This is the core of the assessment. Evaluate the system's potential impact across the following domains. Be specific and consider both positive and negative impacts.)*

#### 3.1. Impact on Individuals and Users

* **Fundamental Rights and Dignity:**
  + **Autonomy:** Does the system enhance or diminish user autonomy? Could it create dependency?
  + **Privacy:** What are the risks of re-identification? How is sensitive mental health data protected?
  + **Well-being (Psychological Safety):** What is the risk of the AI providing harmful, incorrect, or triggering advice? What safeguards are in place to prevent this?
* **Fairness and Non-discrimination:**
  + **Potential Biases:** Could the system perform differently for users based on demographics (age, gender, ethnicity), cultural background, or disability?
  + **Data Bias:** Are the training datasets representative of the intended user population? What steps were taken to identify and mitigate data bias?
  + **Accessibility:** Is the system accessible to people with disabilities (e.g., visual, auditory, cognitive)?

#### 3.2. Impact on Society

* **Societal Well-being:**
  + Could this technology change how people approach mental healthcare (positively or negatively)?
  + Is there a risk of over-reliance on AI for mental health, potentially preventing people from seeking necessary professional care?
* **Economic Impact:**
  + Does this system impact the roles of mental health professionals?

#### 3.3. Impact on the Organization (Fawzooz AI)

* **Reputational Risk:** What is the potential damage to user trust and the Fawzooz AI brand if the system fails or causes harm?
* **Legal and Compliance Risk:** Does the system comply with all relevant regulations (e.g., HIPAA, GDPR)?

### 4. Risk Mitigation and Control Measures

*(For each significant negative impact identified above, detail the specific measures that will be taken.)*

| **Impact/Risk Identified** | **Proposed Mitigation / Control** | **Owner** | **Status** |
| --- | --- | --- | --- |
| **Example:** Risk of chatbot providing harmful advice in a crisis. | 1. Implement a high-accuracy crisis detection model. <br> 2. Create a "hard-stop" protocol that immediately disengages the AI and provides crisis hotline numbers. <br> 3. Implement a "Human-in-the-Loop" alert for clinician review. | Head of Clinical AI | Planned |
| **Example:** Risk of gender bias in content recommendations. | 1. Conduct a fairness audit on the recommendation model. <br> 2. Augment training data to ensure balanced representation. <br> 3. Add a feedback mechanism for users to report irrelevant or biased content. | Lead Data Scientist | In Progress |
|  |  |  |  |

### 5. Consultation and Stakeholder Engagement

* **Who was consulted during this assessment?** (e.g., Licensed therapists, clinical advisors, ethicists, potential users from diverse backgrounds.)
* **What was the key feedback received?**
* **How was this feedback incorporated into the system design or this assessment?**

### 6. Conclusion and Recommendations

* **Overall Assessment:** (e.g., The assessment concludes that while significant risks exist, the proposed mitigation measures are sufficient to manage them to an acceptable level.)
* **Final Recommendation:**
  + [ ] **Proceed:** The project is approved to proceed to the next SDLC phase.
  + [ ] **Proceed with Conditions:** The project is approved, contingent on the implementation of all mandatory recommendations in this report.
  + [ ] **Requires Redesign:** The identified risks are too high. The project must return to the design phase to address fundamental issues.
  + [ ] **Halt:** The project is deemed too high-risk and is not approved to move forward.

### 7. Approvals

**Assessment Lead:**

[Name], [Title]

**AI Governance Committee Review:**

[Name], [Chair of AIGC]