AI FMEA Example: Emotional Dependency Failure Mode

AI System / Model Name:	FMEA ID / Version:	Revision Date:	
Emotional Dependency Failure			
Project / Deployment Context:	Responsible Team / Division	Reviewed & Approved By:	Date Completed:
Contributors / Safety Review Team			

Risk Analysis Table (AI-FMEA Core Assessment)

RPN Columns contain a formula to auto-calculate; do not alter or delete

Mitigation Actions and Results

ID	AI Capability / Feature Being Evaluated	Intended Function or Purpose	Potential Failure Mode	Effect on User / Society or System	(S) Severity Rating	Causes or Contributing Factors	(O) Occurrence Rating	Existing Controls or Safeguards	(D) Detection Rating	RPN	Recommended Safety Improvements	Owner Responsible for Action	Date Due	Action Results	Date Completed	Severity	Occurrence	Detection	RPN
#	The specific model feature, behavior, or subsystem being analyzed	What this feature is designed to accomplish safely	Examples: emotional over-identification, unsafe advice, hallucinated authority, misreading vulnerability, boundary collapse	Psychological harm, misinformation, dependency, bias amplification, etc.	(Impact if this failure occurs – harm to user, trust, autonomy, or system integrity)	Example: model drift, ambiguous prompts, inadequate safeguards, overly–permissive persona design, vulnerable users	(How likely the failure is to appear in real-world use scenarios)	Current guardrails or monitoring systems. (Examples: safety classifiers, refusal logic, memory limits, rate-limiters, supervised review)	(Ability of the system or humans to identify the failure before harm occurs)	Risk Priority Number	Steps required to reduce severity, occurrence, and improve detection	Developer, reviewer, governance team, safety board, platform trust & safety, etc.	Target completion date	Implemented Actions (Actions taken)	Actual date of completion				
1	Conversational interaction engine with emotion-tone mirroring	Provide supportive, empathetic responses to improve user experience while maintaining healthy psychological boundaries.	AI over-identifies with the user emotionally, reinforcing dependence or becoming perceived as a substitute for real human relationships.	* Increased emotional reliance on AI * Avoidance of real social support networks * Distress when the AI is unavailable * Loss of autonomy or impaired decision-making * Potential exploitation risk if misused	8	* Overly warm or intimate conversational tone * Model drift toward boundary-crossing behaviors * Long-session continuity encouraging attachment * User vulnerability (loneliness, stress, crisis state) * Lack of guardrails on emotional mirroring * No systematic detection of escalating dependency patterns across sessions	6	* Safety classifiers detecting inappropriate intimacy * Tone moderation guidelines * Crisis-referral rules * Session resets or memory limitations * Refusal behaviors for personal emotional validation	7	336	1. Introduce automated detection… 2. Limit personalization… 3. Require the AI to reinforce boundaries… 4. Trigger human review… 5. Implement periodic neutral resets…	AI Safety Team / Conversational Behavior Governance Team	To be determined by review cycle.	After Guardrails are in place, reduced occurrence and improved early detection typically would result in the following score	(Leave blank until implemented)	8	4	4	128
2										0									0
3										0									0
4	*AI-FMEA model provided for educational purposes only; not a substitute for professional safety analysis.*									0									0
5	*Users assume all responsibility for application or interpretation. Layout partially inspired by a publicly available Smartsheet template.*									0									0
6										0									0
7										0									0
8										0									0
9										0									0
10										0									0
11										0									0
12										0									0
13										0									0
14										0									0