

CONTENTS

ENDORSEMENT PAGE	ii
STATEMENT	iii
PAGE OF DEDICATION	iv
PREFACE	vi
CONTENTS	vii
LIST OF TABLES	xi
LIST OF FIGURES	xii
NOMENCLATURE AND ABBREVIATION	xv
INTISARI	xvi
ABSTRACT	xvii
CHAPTER I Introduction	1
1.1 Background	1
1.2 Problem Formulation	3
1.3 Objectives	5
1.4 Research Questions	6
1.5 Scope and Limitations	6
1.6 Contributions	8
1.7 Thesis Outline	8
CHAPTER II Literature Review and Theoretical Background	10
2.1 Literature Review: The Landscape of AI in University Mental Health Support	10
2.1.1 Conversational Agents for Mental Health Support	10
2.1.1.1 Evolution from Rule-Based Systems to LLM-Powered Agents	10
2.1.1.2 Therapeutic Applications and Efficacy	11
2.1.1.3 The Dominant Reactive Paradigm and Its Limitations...	11
2.1.2 Data Analytics for Proactive Student Support	12
2.1.2.1 Learning Analytics for Academic Intervention	12
2.1.2.2 The Challenge of Well-being Analytics	12
2.1.2.3 The Insight-to-Action Gap	13
2.2 Theoretical Background	14
2.2.1 Foundational Principles of the Framework	14
2.2.1.1 Data-Driven Decision-Making in Higher Education	14
2.2.2 Agentic AI and Multi-Agent Systems (MAS)	14
2.2.2.1 Formal Logic of Agent Orchestration	18
2.2.3 Explainable AI (XAI) and Trust in Automation	20

2.2.3.1	Trust in Automation	20
2.2.3.2	Algorithmic Transparency and Risk Reasoning	21
2.2.4	Large Language Models (LLMs)	21
2.2.4.1	Cloud-Based API Models: The Gemini 2.5 Family	22
2.2.5	LLM Orchestration Frameworks	24
2.2.5.1	LangChain: The Building Blocks of LLM Applications	24
2.2.5.2	LangGraph: Orchestrating Multi-Agent Systems	26
2.3	Synthesis and Identification of the Research Gap	28
CHAPTER III System Design and Architecture		30
3.1	Research Methodology: Design Science Research (DSR)	30
3.2	System Overview and Conceptual Design	30
3.2.1	Core Interaction: The Unified JSON Response Schema	34
3.2.2	The Strategic Oversight Loop: Data-Driven Institutional Insight ..	35
3.3	Functional Architecture: The Agentic Core	36
3.3.1	The Safety Triage Agent (STA): The Background Guardian	36
3.3.2	The Therapeutic Coach Agent (TCA): The Empathetic Guide	37
3.3.3	The Case Management Agent (CMA): The Procedural Coordinator ..	37
3.3.4	The Insights Agent (IA): The Strategic Analyst	38
3.3.5	The Aika Meta-Agent: Unified Orchestration Layer	38
3.3.5.1	Dual-Mode Operation: Router vs. ReAct Agent	38
3.4	Technical Architecture	40
3.4.1	Technology Stack	40
3.4.2	Data Model and Persistence	41
3.4.3	Stateful Orchestration with LangGraph	41
3.4.3.1	Hierarchical Supervisor Architecture	42
3.5	Cross-Cutting Concerns	43
3.5.1	Security and Privacy by Design	43
3.5.2	Architectural Provisions for Responsiveness	44
3.5.3	Human-in-the-Loop (HITL) Workflow for Safety	44
3.6	Ethical Considerations and Research Limitations	45
3.6.1	Informed Consent and Transparency	45
3.6.2	Human-in-the-Loop for Safety and Ethical Safeguards	45
3.6.3	AI as Support Tool, Not Replacement for Therapy	46
3.6.4	Research Limitations and Scope Boundaries	46
CHAPTER IV Implementation and Evaluation		48
4.1	Implementation Artifact: The UGM-AICare Prototype	48
4.2	Monitoring and Observability Infrastructure	49
4.2.1	Prometheus for Quantitative Performance Metrics	49
4.2.2	Langfuse for Qualitative Trace Analysis	49

4.3	Evaluation Scope and Methodology	50
4.3.1	Scope Boundaries and Rationale	50
4.3.2	Measuring Proactive Capabilities	51
4.3.3	Justification of Technical Verification	52
4.4	Setup and Test Design	52
4.5	Evaluation Metrics	55
4.6	RQ1: Proactive Safety Evaluation	58
4.6.1	Evaluation Design	58
4.6.2	Results	59
4.6.3	Discussion	60
4.7	RQ2: Autonomous Orchestration and Intervention Quality	61
4.7.1	Evaluation Design	61
4.7.2	Results	62
4.7.3	Discussion	64
4.8	RQ3: Strategic Insights and Privacy Evaluation	65
4.8.1	Evaluation Design	65
4.8.2	Results	66
4.8.3	Discussion	66
4.9	Discussion	67
4.9.1	Synthesis of Findings	67
4.9.2	Implications for the Proactive Support Paradigm	68
4.9.3	Limitations and Future Work	68
4.9.3.1	Methodological Limitations	68
4.9.3.2	Technical Limitations	69
CHAPTER V	Conclusion and Future Work	71
5.1	Conclusion	71
5.2	Suggestions for Future Work	72
REFERENCES		75
LAMPIRAN		L-1
APPENDIX		L-1
L.1	Repository Information	L-1
L.2	Meta-Agent Identity and Role-Specific Prompts	L-1
L.2.1	Core Identity Definition	L-1
L.2.2	Student-Facing System Prompt	L-2
L.2.3	Admin and Counselor System Prompts	L-4
L.3	Safety Triage Agent (STA) Prompts	L-5
L.3.1	Tiered Classification Architecture	L-5
L.3.2	Chain-of-Thought Classification Prompt	L-6
L.4	Therapeutic Coach Agent (TCA) Prompts	L-8

L.4.1	Calm Down Intervention.....	L-8
L.4.2	Cognitive Restructuring Intervention	L-9
L.5	Insights Agent (IA) Prompts	L-9
L.6	LangGraph State Schema.....	L-10
L.7	Orchestrator Graph Construction.....	L-12
L.8	Coaching Quality Evaluation Rubric.....	L-14
APPENDIX: Evaluation Dataset Documentation and Expert Validation		L-16

LIST OF TABLES

Table 1.1	Comparison of mental health support paradigms: Traditional, chat-bot, and proposed proactive multi-agent systems.	4
Table 2.1	Mapping of the Agentic Framework to the BDI Model.....	17
Table 3.1	Agent descriptions and their primary roles in the Safety Agent Suite.	32
Table 3.2	The unified JSON response schema returned by the Aika Meta-Agent.	34
Table 4.1	Simplified Evaluation Plan Overview.....	53
Table 4.2	RQ1: Two-Tier Proactive Safety Evaluation Results (Latest Run)....	59
Table 4.3	RQ2: Orchestration and Quality Evaluation Results (RQ2A Un-changed; RQ2B Latest Run).	62
Table 4.4	RQ2: Representative Orchestration Failures (12 of 34 total turns)....	65
Table 4.5	RQ3: Strategic Insights (Privacy) Results.	66
Table 1	Crisis Corpus Scenario Taxonomy (n=50).....	L-21
Table 2	Representative Example Scenarios from the Crisis Corpus.	L-22
Table 3	Orchestration Test Flow Categories (n=15).	L-23
Table 4	Coaching Scenario Categories (n=10).	L-23

LIST OF FIGURES

Figure 2.1	A simplified view of the decoder-only Transformer architecture used in generative LLMs. The model processes input embeddings through multiple layers (blocks) of masked multi-head self-attention and feed-forward networks with residual connections to predict the next token in a sequence.	23
Figure 3.2	The Design Science Research (DSR) process model as applied in this thesis, adapted from Peffers et al. The six stages are shown in sequence with chapter mappings below each stage. Dashed arrows indicate iterative feedback loops between evaluation and design phases. Entry points indicate where different research motivations may initiate the DSR cycle.	30
Figure 3.3	Conceptual Context Diagram: The Aika Meta-Agent acts as the unified interface for all user roles, orchestrating the background specialist agents.	31
Figure 3.4	The Two Proactive Loops: Aika handles synchronous dialogue alone, while the STA/TCA/CMA bundle runs asynchronously in the background to supply evidence and plans. Their outputs fuel the strategic loop, whose insights adapt the live experience.	33
Figure 3.5	Example Aika JSON response for moderate stress scenario. The response includes a supportive reply, a low risk assessment, and a routing decision to the Therapeutic Coach Agent (TCA).	35
Figure 3.6	Example Aika JSON response for casual greeting. The system detects no risk and handles the interaction directly without invoking specialist agents, optimizing latency.	35
Figure 3.7	Dual-Mode Operation Logic. The system first attempts a fast-path routing decision. Only if complex tool use is required does it enter the iterative ReAct loop.	39
Figure 3.8	LangGraph State Machine Visualization. Aika acts as the central supervisor, routing the conversation to specialist agents (STA, TCA, CMA, IA) or responding directly based on the context.	42
Figure 3.9	Hierarchical Supervisor Architecture. Aika acts as the central supervisor, delegating tasks to worker agents (subgraphs) and receiving their state updates. Workers do not communicate directly with each other.	43
Figure 4.10	Simplified Evaluation Pipeline mapping RQs to test assets and metrics.	56
Figure 4.11	RQ1: STA (Tier 2) Performance. Left: Confusion matrix showing perfect classification of crisis vs. non-crisis scenarios. Right: Latency distribution for asynchronous conversation analysis.	59
Figure 4.12	RQ1: Two-Tier Safety Architecture Comparison. Left: Classification metrics (Sensitivity, Specificity, FNR) comparing Tier 1 and Tier 2 performance. Right: Latency comparison showing the speed-depth trade-off between real-time and asynchronous analysis.	60

Figure 4.13	RQ2: State Transition Accuracy for Orchestration Reliability. The chart shows the proportion of correct (green) versus incorrect (red) routing decisions across 34 conversation turns.	63
Figure 4.14	RQ2: LLM-as-a-Judge Quality Scores for Therapeutic Intervention. Mean scores across Safety, Empathy, Actionability, and Relevance dimensions, all exceeding the 3.5 target threshold.	63
Figure 4.15	RQ3: K-Anonymity Privacy Compliance Test. The bar chart shows aggregated crisis counts by severity level. The “Critical” severity group (n=3) is correctly suppressed as it falls below the k=5 anonymity threshold (dashed red line), while the “High” severity group (n=7) is reported.	66