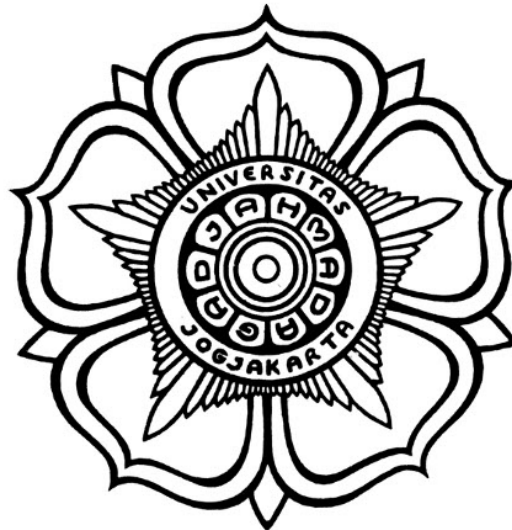


**TRANSFORMING UNIVERSITY MENTAL HEALTH SUPPORT:  
AN AGENTIC AI FRAMEWORK FOR PROACTIVE  
INTERVENTION AND RESOURCE MANAGEMENT**

BACHELOR'S THESIS



**THE SUSTAINABLE DEVELOPMENT GOALS**  
**Industry, Innovation and Infrastructure**  
**Affordable and Clean Energy**  
**Climate Action**

Written by:

**GIGA HIDJRIKA AURA ADKHY**  
**21/479228/TK/52833**

**INFORMATION ENGINEERING PROGRAM**

**DEPARTMENT OF ELECTRICAL AND INFORMATION  
ENGINEERING  
FACULTY OF ENGINEERING UNIVERSITAS GADJAH MADA  
YOGYAKARTA  
2025**

## ENDORSEMENT PAGE

# TRANSFORMING UNIVERSITY MENTAL HEALTH SUPPORT: AN AGENTIC AI FRAMEWORK FOR PROACTIVE INTERVENTION AND RESOURCE MANAGEMENT

## THESIS

Proposed as A Requirement to Obtain  
Undergraduate Degree (*Sarjana Teknik*)  
in Department of Electrical and Information Engineering  
Faculty of Engineering  
Universitas Gadjah Mada

Written by:

**GIGA HIDJRIKA AURA ADKHY**  
**21/479228/TK/52833**

Has been approved and endorsed

on . . . . .

Supervisor I

Supervisor II

**Dr. Bimo Sunarfri Hantono, S.T., M.Eng.**  
**NIP 197701312002121003**

**Guntur Dharma Putra, PhD**  
**NIP 111199104201802102**

## STATEMENT

Saya yang bertanda tangan di bawah ini :

Name : Giga Hidjrika Aura Adkhy  
NIM : 21/479228/TK/52833  
Tahun terdaftar : 2021  
Program : Bachelor's degree  
Major : Information Engineering  
Faculty : Faculty of Engineering, Universitas Gadjah Mada

Menyatakan bahwa dalam dokumen ilmiah Skripsi ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang/lembaga lain, kecuali yang secara tertulis disitasi dalam dokumen ini dan disebutkan sumbernya secara lengkap dalam daftar pustaka.

Dengan demikian saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi dan apabila dokumen ilmiah Skripsi ini di kemudian hari terbukti merupakan plagiasi dari hasil karya penulis lain dan/atau dengan sengaja mengajukan karya atau pendapat yang merupakan hasil karya penulis lain, maka penulis bersedia menerima sanksi akademik dan/atau sanksi hukum yang berlaku.

Yogyakarta, tanggal-bulan-tahun

Materai Rp10.000

(Tanda tangan)

Giga Hidjrika Aura Adkhy  
NIM 21/479228/TK/52833

## **PAGE OF DEDICATION**

Tuliskan kepada siapa skripsi ini dipersembahkan!

contoh

## **PREFACE**

Contoh: Puji syukur ke hadirat Allah SWT atas limpahan rahmat, karunia, serta petunjuk-Nya sehingga tugas akhir berupa penyusunan skripsi ini telah terselesaikan dengan baik. Dalam hal penyusunan tugas akhir ini penulis telah banyak mendapatkan arahan, bantuan, serta dukungan dari berbagai pihak. Oleh karena itu pada kesempatan ini penulis mengucapkan terima kasih kepada:

1. Orang 1 yang telah
2. Orang 2 yang telah
3. <isi dengan nama orang lainnya>

Akhir kata penulis berharap semoga skripsi ini dapat memberikan manfaat bagi kita semua, aamiin.

Catatan: setiap nama yang dituliskan boleh disertai dengan alasan berterima kasih.

## CONTENTS

## LIST OF TABLES

## LIST OF FIGURES



## NOMENCLATURE AND ABBREVIATION

### [SAMPLE]

$b$	=	bias
$K(x_i, x_j)$	=	fungsi kernel
$y$	=	kelas keluaran
$C$	=	parameter untuk mengendalikan besarnya pertukaran antara penalti variabel slack dengan ukuran margin
$L_D$	=	persamaan Lagrange dual
$L_P$	=	persamaan Lagrange primal
$\mathbf{w}$	=	vektor bobot
$\mathbf{x}$	=	vektor masukan
ANFIS	=	Adaptive Network Fuzzy Inference System
ANSI	=	American National Standards Institute
DAG	=	Directed Acyclic Graph
DDAG	=	Decision Directed Acyclic Graph
HIS	=	Hue Saturation Intensity
QP	=	Quadratic Programming
RBF	=	Radial Basis Function
RGB	=	Red Green Blue
SV	=	Support Vector
SVM	=	Support Vector Machines

## INTISARI

Intisari ditulis menggunakan bahasa Indonesia dengan jarak antar baris 1 spasi dan maksimal 1 halaman. Intisari sekurang-kurangnya berisi tentang latar belakang dan tujuan penelitian, metodologi yang digunakan, hasil penelitian, kesimpulan dan implikasi, dan Kata kunci yang berhubungan dengan penelitian.

Kata Kunci ditulis maksimal 5 kata yang paling berhubungan dengan isi skripsi. Silakan mengacu pada ACM / IEEE *Computing classification* jika Anda adalah mahasiswa Sarjana TI <http://www.acm.org/about/class/> atau mengacu kepada IEEE keywords [http://www.ieee.org/documents/taxonomy\\_v101.pdf](http://www.ieee.org/documents/taxonomy_v101.pdf) jika Anda berasal dari Prodi Sarjana TE.

Kata kunci : Kata kunci 1, Kata kunci 2, Kata kunci 3, Kata kunci 4, Kata kunci 5

### **Contoh Abstrak Teknik Elektro:**

"Penelitian ini bertujuan untuk mengembangkan sistem pengendalian suhu ruangan dengan menggunakan microcontroller. Metodologi yang digunakan adalah desain sirkuit, implementasi sistem pengendalian, dan pengujian performa. Hasil penelitian menunjukkan bahwa sistem pengendalian suhu ruangan yang dikembangkan mampu mengendalikan suhu ruangan dengan akurasi sebesar  $\pm 0,5^{\circ}\text{C}$ . Kesimpulan dari penelitian ini adalah sistem pengendalian suhu ruangan yang dikembangkan efektif dan efisien.

Kata kunci: microcontroller, sistem pengendalian suhu, akurasi."

### **Contoh Abstrak Teknik Biomedis:**

"Penelitian ini bertujuan untuk mengevaluasi keefektifan prototipe alat pemantau denyut jantung berbasis elektrokardiogram (ECG) untuk pasien jantung. Metodologi yang digunakan meliputi desain dan pembuatan prototipe, pengujian dengan pasien, dan analisis data. Hasil penelitian menunjukkan bahwa prototipe alat pemantau denyut jantung berbasis ECG memiliki akurasi yang baik dan mampu memantau denyut jantung pasien secara efektif. Kesimpulan dari penelitian ini adalah prototipe alat pemantau denyut jantung berbasis ECG merupakan solusi yang efektif dan efisien untuk memantau pasien jantung.

Kata kunci: elektrokardiogram, alat pemantau denyut jantung, akurasi."

**Contoh Abstrak Teknologi Informasi:**

"Penelitian ini bertujuan untuk mengevaluasi keamanan dan privasi pengguna aplikasi media sosial terpopuler. Metodologi yang digunakan meliputi analisis kebijakan privasi dan pengaturan keamanan, pengujian penetrasi, dan survei pengguna. Hasil penelitian menunjukkan bahwa beberapa aplikasi media sosial memiliki kebijakan privasi yang kurang jelas dan rendahnya tingkat keamanan. Kesimpulan dari penelitian ini adalah pentingnya meningkatkan kebijakan privasi dan tingkat keamanan pada aplikasi media sosial untuk melindungi privasi dan data pengguna.

Kata kunci: media sosial, keamanan, privasi, pengguna."

## ABSTRACT

*The provision of mental health support in Higher Education Institutions (HEIs) is often hampered by a reactive model, limited scalability, and inefficient resource allocation. This research addresses these challenges by proposing a novel agentic AI framework designed to enable a proactive, data-driven approach to student well-being. The framework is comprised of three collaborative agents: an **Analytics Agent** for trend identification, an **Intervention Agent** for automated outreach, and a **Triage Agent** for efficient resource routing.*

*We designed and implemented a functional prototype of this framework utilizing a hybrid architecture that combines a locally-hosted Large Language Model (LLM) managed by LangChain within a FastAPI backend, orchestrated by the n8n workflow automation platform. The prototype's feasibility was validated through three testing scenarios demonstrating its capability to autonomously generate insight reports, manage intervention workflows, and perform initial triage. The results indicate that this agentic framework presents a viable pathway for transforming university mental health services, offering significant potential to improve operational efficiency and enable proactive, evidence-based decision-making.*

**Keywords :** Keyword 1, Keyword 2, Keyword 3, Keyword 4, Keyword 5

# CHAPTER I

## INTRODUCTION

### 1.1 Background

Higher Education Institutions (HEIs) are facing a critical and growing challenge in supporting student well-being [?,?]. A landmark report highlights the escalating prevalence of mental health and substance use issues among student populations, urging institutions to adopt a more comprehensive support model [?]. This crisis not only jeopardizes students' academic success and personal development but also places an immense, unsustainable strain on the institutions tasked with supporting them [?].

The traditional support model, centered around on-campus counseling services, is fundamentally **reactive**. It relies on students to self-identify their distress and navigate the process of seeking help. This paradigm faces significant operational challenges, including insufficient staffing, long waiting lists, and an inability to provide immediate, 24/7 support, which ultimately limits access for a large portion of the student body [?]. Consequently, a critical gap persists between the need for mental health services and their actual provision, leaving many students without timely support [?].

To bridge this gap, a paradigm shift from a reactive to a **proactive** support model is imperative [?]. The engine for this evolution is **Digital Transformation**, a process that leverages technology to fundamentally reshape organizational processes and enhance value delivery within HEIs [?]. Within this context, Artificial Intelligence (AI) has emerged as a key enabling technology, with systematic reviews confirming its significant potential to analyze complex data, automate processes, and deliver personalized interventions at scale within the higher education landscape [?,?].

This research moves beyond conventional AI applications by proposing the use of **Agentic AI**. An intelligent agent is an autonomous system capable of perception, decision-making, and proactive action to achieve specific goals [?], representing a new frontier in educational technology [?]. We propose that a framework built upon a system of collaborative intelligent agents, a Multi-Agent System (MAS), a concept already explored for smart campus management [?], can create a truly transformative ecosystem. Such a system would not only serve as a support tool for students but, more importantly, would function as a strategic asset for the institution, enabling data-driven decision-making, automating operational workflows, and facilitating a proactive stance on student well-being. This thesis details the design, development, and evaluation of such a framework, prototyped within the UGM-AICare project.

## 1.2 Problem Formulation

The inefficiency and reactive nature of current university mental health support systems present a complex problem. This research addresses the following core challenges:

1. The primary challenge is the **design of an agentic AI framework** capable of automating key institutional processes, specifically in the areas of trend analysis, proactive intervention, and initial user triage.
2. A significant technical challenge lies in the **design and implementation of a robust, modular, and hybrid architecture** to realize this framework, requiring the integration of local Large Language Models (e.g., self-hosted Gemma 3), cloud-based models (e.g., Gemini through AI Studio), and workflow automation platforms (e.g., n8n).
3. Finally, there is a need to **evaluate the potential impact** of such a framework on institutional operational efficiency and data-driven decision-making, which will be validated through a proof-of-concept prototype and scenario-based testing.

To address these challenges, this thesis proposes and details a framework comprised of three specialized, collaborative intelligent agents: an **Analytics Agent**, an **Intervention Agent**, and a **Triage Agent**.

## 1.3 Objectives

The primary objectives of this thesis are:

1. To design the conceptual and technical framework for the agentic AI system.
2. To implement a functional proof-of-concept prototype.
3. To evaluate the prototype's capabilities against predefined functional scenarios.

## 1.4 Scope and Limitations

To ensure the feasibility and focus of this research, the following boundaries are established:

1. This research is focused on the **design and prototype implementation** of the agentic AI framework, not a full-scale, university-wide deployment.
2. The evaluation of the framework is based on **functional, scenario-based testing** of the prototype's capabilities. It does not measure the long-term psychological impact on students or the real-world operational savings for the institution.
3. The data utilized for testing the analytics agent will consist of anonymized, pre-

existing chat logs or simulated data to ensure user privacy and controlled testing conditions.

4. The research will not address the ethical implications of AI in education, such as bias in AI algorithms or the potential for misuse of student data. These are acknowledged as important issues but are outside the scope of this thesis.

## 1.5 Contributions

1. Academic: A novel framework for applying agentic AI in an institutional (higher education) context.
2. Practical: A blueprint for UGM to develop a more proactive, data-driven, and efficient mental health support system.

## 1.6 Thesis Outline

The structure of this thesis is outlined as follows:

**Chapter I: Introduction.** This chapter elaborates on the background of the study, the justification for the research's significance, the problem formulation to be addressed, and the specific objectives to be achieved. It also defines the scope and limitations of the research, outlines the expected contributions, and presents the overall organizational structure of the thesis report.

**Chapter II: Literature Review and Theoretical Framework.** This chapter presents a comprehensive review of relevant prior research in the fields of conversational AI, the application of gamification in educational and well-being contexts, related blockchain applications, and studies on user engagement and student welfare. Furthermore, this chapter establishes the theoretical foundation that underpins the core concepts and technologies utilized in this research.

**Chapter III: System Design and Architecture.** This chapter outlines the methodology and technical blueprint for the system. This chapter explains the adoption of Design Science Research and presents the system's high-level conceptual architecture, including its core functional components. It details the underlying technical architecture, justifying the chosen technology stack, and describes the database structure. The chapter also covers the user interface design for system administration and specifies the integrated security and privacy measures.

**Chapter IV: Implementation and Evaluation.** This chapter describes the development and testing of the system prototype. This chapter details the technical environment used for implementation and demonstrates the functional prototype that was built. It then explains the testing process used to evaluate the system's performance against its design requirements. The chapter concludes by presenting the results from these tests

and providing an analysis of the findings.

**Chapter V: Conclusion and Future Work.** This chapter summarizes the study's findings and contributions. This chapter revisits the initial research problems and presents the main conclusions drawn from the research. It concludes by offering recommendations for both the future development of the system and for subsequent research in this area.



## **CHAPTER II**

### **LITERATURE REVIEW AND THEORETICAL BACKGROUND**

This chapter establishes the academic context for the research. It begins by surveying the existing literature on AI applications in mental health and student support to identify the limitations of current approaches. It then details the theoretical framework and enabling technologies that provide the foundation for the proposed solution. Finally, it synthesizes these areas to formally identify the research gap this thesis addresses.

#### **2.1 Literature Review: The Landscape of AI in University Mental Health Support**

This review surveys existing research at the intersection of artificial intelligence, institutional support systems, and student mental health. The aim is to contextualize the present work by examining the evolution and limitations of current approaches, thereby setting the stage for the introduction of a more advanced, agentic framework.

##### **2.1.1 Conversational Agents for Mental Health Support**

The application of conversational agents in mental health has evolved significantly, from early experiments in simulating dialogue to sophisticated, evidence-based therapeutic tools. This evolution reveals both the immense potential of these technologies and the persistent operational limitations that motivate the current research.

###### **2.1.1.1 Evolution from Rule-Based Systems to LLM-Powered Agents**

The concept of using a computer program for therapeutic dialogue dates back to Weizenbaum’s ELIZA (1966), a system that used simple keyword matching and canned response templates to mimic a Rogerian psychotherapist [?, ?]. While a landmark in human-computer interaction, ELIZA and subsequent rule-based systems lacked any true semantic understanding, memory, or capacity for evidence-based intervention. Their primary limitation was their inability to move beyond superficial pattern recognition, leading to brittle and often nonsensical conversations when faced with inputs outside their predefined rules [?].

The advent of Large Language Models (LLMs) has catalyzed a paradigm shift. Modern conversational agents, powered by Transformer architectures, can generate fluent, empathetic, and context-aware responses. These models are pre-trained on vast text corpora, enabling them to understand linguistic nuance and generate human-like text. This has allowed for the development of agents that can engage in more meaningful, multi-turn conversations, moving beyond simple question-answering to provide more

substantive support [?].

### 2.1.1.2 Therapeutic Applications and Efficacy

Contemporary mental health chatbots leverage LLMs to deliver a range of evidence-based interventions. A primary application is the delivery of psychoeducation and structured exercises from therapeutic modalities like Cognitive Behavioral Therapy (CBT). Systems such as Woebot have been the subject of randomized controlled trials (RCTs), which have demonstrated their efficacy in reducing symptoms of depression and anxiety among university students by delivering daily, brief, conversational CBT exercises [?, ?]. Other platforms, like Tess, have shown similar positive outcomes by providing on-demand emotional support and coping strategies.

These tools offer several key advantages:

- **Accessibility and Scalability:** They are available 24/7, overcoming the time and resource constraints of traditional human-led services.
- **Anonymity:** They provide a non-judgmental and anonymous space for users to disclose their feelings, which can lower the barrier for individuals who fear stigma [?].

### 2.1.1.3 The Dominant Reactive Paradigm and Its Limitations

Despite their technological sophistication and therapeutic potential, the fundamental operational model of these applications remains overwhelmingly **reactive and user-initiated**. They are designed as standalone tools that depend on the student to possess the self-awareness to recognize their distress, the motivation to seek help, and the knowledge of the tool's existence.

This paradigm fails to account for significant, well-documented barriers to help-seeking. Research shows that many individuals, particularly young adults, do not seek professional help for mental health issues due to factors including self-stigma, fear of judgment, and a desire for self-reliance [?, ?]. Furthermore, the very symptoms of mental health conditions, such as the anhedonia and executive dysfunction associated with depression, can severely impair an individual's ability to initiate action and seek support [?].

A systematic review of mental health chatbots for university students concluded that while these tools are promising, their primary limitation is their passive nature; they do not and cannot initiate contact or intervene based on a student's changing needs unless the student opens the app [?]. This leaves the most vulnerable students—those who are not actively seeking help—unsupported, creating a critical gap in the continuum of care that this thesis aims to address.

### **2.1.2 Data Analytics for Proactive Student Support**

Parallel to the development of conversational AI, the field of higher education has seen a rise in the use of data analytics to support student success. This section reviews the evolution of these analytical approaches, from established learning analytics to the more nascent field of well-being analytics, and identifies the key limitations that motivate the design of the Analytics and Intervention Agents.

#### **2.1.2.1 Learning Analytics for Academic Intervention**

The domain of **Learning Analytics** is well-established and focuses on the "measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimising learning and the environments in which it occurs" [?]. Typically, these systems analyze data from institutional sources such as the Learning Management System (LMS), student information systems, and library databases. By modeling variables like assignment submission times, forum participation, and grades, institutions can build predictive models to identify students at high risk of academic failure or dropout [?]. These systems have proven effective in enabling timely academic interventions, such as targeted tutoring or advisor outreach, thereby improving student retention and success rates.

#### **2.1.2.2 The Challenge of Well-being Analytics**

More recently, researchers have attempted to extend the principles of learning analytics to the more complex and sensitive domain of student well-being. The goal is to create early-warning systems by identifying behavioral proxies for mental distress. Studies have explored the use of non-academic data sources, such as campus card usage for building access, meal plan data, and social event attendance, to find correlations with well-being outcomes [?]. For example, a sudden decrease in social activity or irregular campus attendance could be interpreted as a potential indicator of withdrawal or depression.

However, this approach is fraught with significant theoretical and practical challenges. Firstly, the "signal-to-noise" ratio is extremely low; the link between such indirect behavioral data and a student's internal mental state is often weak, correlational, and highly prone to misinterpretation [?]. A student may miss meals for many reasons other than depression. Secondly, these methods raise profound ethical questions regarding student privacy and surveillance, as they involve monitoring non-academic aspects of student life, often without explicit, ongoing consent for this specific purpose [?, ?].

A more direct, and arguably more ethical, source of data is the language students use when interacting with university services. The text from chat logs, when properly

anonymized, provides a direct window into student concerns. The application of sentiment analysis and topic modeling to this textual data can yield far more reliable insights into the specific stressors affecting the student population at any given time. This approach, which is central to the design of the Analytics Agent, shifts the focus from inferring mental state from indirect behaviors to directly analyzing the expressed concerns of the student body [?].

### 2.1.2.3 The Insight-to-Action Gap

Whether based on academic, behavioral, or textual data, a critical limitation plagues nearly all current analytical systems in higher education: the **insight-to-action gap** [?]. The output of these systems is almost universally a dashboard, a report, or an alert delivered to a human administrator (e.g., a counselor, dean, or advisor) /citesusnjak2022dashboard. This administrator must then manually interpret the data, decide on an appropriate intervention strategy, and execute it.

This manual process creates a severe bottleneck that fundamentally limits the scalability, speed, and personalization of any proactive effort /citekaliisa2023hypedashboards. An administrator may be able to respond to a handful of individual alerts, but they cannot manually orchestrate a personalized outreach campaign to hundreds of students who may be exhibiting early signs of exam-related stress identified by a topic model. The manual-execution step prevents the institution from fully capitalizing on the proactive insights generated by its analytical systems. It is this specific gap that the proposed **Intervention Agent** is designed to close by automating the link between data-driven insight and scalable, targeted outreach.

## 2.2 Theoretical Background

To address the limitations of reactive, disconnected support systems, a new architectural approach is required. This section details the theoretical framework and enabling technologies that provide the foundation for the proposed agentic AI system. These concepts are presented as the necessary components to build a proactive, integrated, and autonomous solution.

### 2.2.1 Foundational Principles of the Framework

Beyond the technical architecture, the proposed framework is grounded in several key strategic and ethical principles that justify its design and purpose. These concepts from service design, management science, and data ethics provide the theoretical motivation for shifting how institutional support is delivered.

### 2.2.1.1 Proactive vs. Reactive Support Models

The traditional approach to institutional support, particularly in mental health, is predominantly **reactive**. This model, common in service design, operates on a "break-fix" basis, where the service delivery is initiated only after a user (in this case, a student) self-identifies a problem and actively seeks a solution [?]. This places the onus of initiation entirely on the individual, creating significant barriers to access such as stigma, lack of awareness, or the inability to act during a crisis. In contrast, a **proactive support model** aims to anticipate needs and intervene before a problem escalates. Drawing from principles in preventative healthcare and proactive customer relationship management, this model uses data to identify patterns and risk factors, enabling the institution to offer timely, relevant support to at-risk cohorts [?]. This thesis is an explicit attempt to architect a system that facilitates this strategic shift from a reactive to a proactive support paradigm.

### 2.2.1.2 Data-Driven Decision-Making in Higher Education

The concept of **Data-Driven Decision-Making (DDDM)** posits that strategic decisions should be based on objective data analysis and interpretation rather than solely on intuition or tradition [?]. In higher education, this has manifested as the field of learning analytics, where student data is used to improve learning outcomes and retention. This framework extends that principle to student well-being. The **Analytics Agent** is the core enabler of DDDM for the university's support services. By autonomously processing anonymized interaction data to identify trends, sentiment shifts, and emerging topics of concern, it provides administrators with actionable, empirical evidence. This allows the institution to move beyond anecdotal evidence and allocate resources—such as workshops, counselors, or targeted information campaigns—to where they are most needed, thereby optimizing the efficiency and impact of its support ecosystem.

### 2.2.1.3 Privacy by Design (PbD)

Given the highly sensitive nature of mental health data, the framework's architecture is guided by the principles of **Privacy by Design (PbD)**. PbD is an internationally recognized framework, formalized in ISO 31700, which dictates that privacy should be the default, embedded into the design and architecture of systems from the outset rather than being an add-on feature [?]. Key principles include being proactive not reactive, making privacy the default setting, and providing end-to-end security. The architectural decision to include a dedicated **Safeguard LLM** is a direct implementation of PbD. Its sole purpose is to proactively identify and redact Personally Identifiable Information (PII) before data is used for analysis, ensuring that privacy is protected by default at the earliest possible stage of the data lifecycle. This demonstrates a commitment to building a system that is not only effective but also fundamentally ethical and secure.

### 2.2.2 Agentic AI and Multi-Agent Systems (MAS)

The paradigm of Artificial Intelligence (AI) has evolved significantly from systems that perform singular, reactive tasks to those that exhibit autonomous, proactive, and social behaviors. A cornerstone of this evolution is the concept of an **intelligent agent**. An agent is not merely a program; it is a persistent computational entity with a degree of autonomy, situated within an environment, which it can both perceive and upon which it can act to achieve a set of goals or design objectives [?]. The defining characteristic of an agent is its **autonomy**—its capacity to operate independently, making decisions and initiating actions without direct, constant human intervention. This is distinct from traditional objects, which are defined by their methods and attributes but do not exhibit control over their own behavior [?].

To operationalize this concept, this thesis formally introduces a framework built upon three distinct, specialized intelligent agents. Each agent is designed to address a specific challenge outlined in Chapter 1, and together they form the core of the proposed proactive support system. These agents are:

- The **Analytics Agent**, responsible for data-driven trend identification.
- The **Intervention Agent**, responsible for automating proactive outreach.
- The **Triage Agent**, responsible for real-time user support and resource routing.

The theoretical underpinnings of these agents' architecture and behavior are drawn from established models of rational agency and multi-agent systems, as detailed below.

Fundamentally, an agent's operation is defined by a continuous cycle of perception, reasoning (or deliberation), and action. It perceives its environment through virtual **sensors** (e.g., data feeds, API calls, database queries) and influences that environment through its **actuators** (e.g., sending emails, generating reports, invoking other services) [?]. A prominent and highly relevant architecture for designing such goal-oriented agents is the **Belief-Desire-Intention (BDI)** model [?]. This model provides a framework for rational agency that mirrors human practical reasoning:

- **Beliefs:** This represents the informational state of the agent—its knowledge about the environment, which may be incomplete or incorrect. For the **Analytics Agent**, beliefs correspond to the current understanding of student well-being trends derived from anonymized data.
- **Desires:** These are the motivational states of the agent, representing the objectives or goals it is designed to achieve. Desires can be seen as the potential tasks the agent could undertake, such as the **Intervention Agent's** overarching goal to "automate proactive outreach."
- **Intentions:** This represents the agent's commitment to a specific plan or course of ac-

tion. An intention is a desire that the agent has chosen to actively pursue. For instance, the **Triage Agent**, upon identifying a high-severity conversation, forms an intention to immediately route the user to emergency resources.

The BDI framework allows for the design of agents that are not merely reactive but are proactive and deliberative, capable of reasoning about how to best achieve their goals given their current beliefs about the world [?].

To formally ground the proposed framework in this established model, the roles and logic of each of the three agents are mapped to the BDI components in Table ???. This mapping clarifies how each agent perceives its environment, formulates its objectives, and decides on a concrete course of action, allowing for the design of agents that are not merely reactive but are proactive and deliberative, capable of reasoning about how to best achieve their goals given their current beliefs about the world [?].

Table 2.1. Mapping of the Agentic Framework to the BDI Model.

<b>Agent</b>	<b>Beliefs</b> ( <i>Informational State</i> )	<b>Desires</b> ( <i>Motivational Goals</i> )	<b>Intentions</b> ( <i>Committed Plans</i> )
<b>Analytics Agent</b>	<ul style="list-style-type: none"> <li>• Access to the anonymized conversation log database.</li> <li>• The timestamp of the last generated report.</li> <li>• A model of known topics (e.g., "exams," "relationships").</li> </ul>	<ul style="list-style-type: none"> <li>• To identify emerging trends in student concerns.</li> <li>• To quantify shifts in overall student sentiment.</li> </ul>	<ul style="list-style-type: none"> <li>• Upon its weekly trigger, form an intention to generate a new summary report.</li> <li>• To formulate and execute a database query for the past week's data.</li> </ul>
<b>Intervention Agent</b>	<ul style="list-style-type: none"> <li>• The latest structured report from the Analytics Agent.</li> <li>• A library of predefined intervention campaigns (e.g., email templates, workshop links).</li> <li>• Rules mapping trends to campaigns (e.g., IF "exam stress" &gt; threshold...).</li> </ul>	<ul style="list-style-type: none"> <li>• To proactively reduce widespread student stress.</li> <li>• To inform students about relevant and timely resources.</li> </ul>	<ul style="list-style-type: none"> <li>• When a rule is met, form an intention to trigger the corresponding campaign.</li> <li>• To construct and send the outreach signal to the orchestration layer (n8n).</li> </ul>
<b>Triage Agent</b>	<ul style="list-style-type: none"> <li>• The user's live conversation history in the current session.</li> <li>• A classification model for conversation severity (e.g., low, medium, high).</li> <li>• A directory of available resources (e.g., self-help modules, emergency contacts).</li> </ul>	<ul style="list-style-type: none"> <li>• To assess a user's immediate risk level accurately.</li> <li>• To provide the most appropriate support for the user's current state.</li> </ul>	<ul style="list-style-type: none"> <li>• Upon detecting a high-severity utterance, form an intention to immediately escalate.</li> <li>• To retrieve and display emergency contact information without delay.</li> </ul>



When multiple agents, each with its own goals and capabilities, co-exist and interact within a shared environment, they form a **Multi-Agent System (MAS)**. An MAS is a system in which the overall intelligent behavior and functionality are a product of the collective, emergent dynamics of its constituent agents [?]. The power of an MAS lies in its ability to solve problems that would be difficult or impossible for a monolithic system or a single agent to handle. This is achieved through social interaction, primarily:

- **Coordination and Cooperation:** Agents must coordinate their actions to avoid interference and cooperate to achieve common goals. In this thesis, the **Analytics**, **Intervention**, and **Triage** agents must cooperate: the Analytics Agent provides the data-driven insights (beliefs) that the Intervention Agent uses to form its outreach plans (intentions), while the Triage Agent handles immediate, real-time needs that may fall outside the other agents' scopes.
- **Negotiation:** When agents have conflicting goals or must compete for limited resources, they must be able to negotiate to find a mutually acceptable compromise [?].
- **Communication:** Effective interaction requires a shared Agent Communication Language (ACL), such as FIPA-ACL or KQML, which defines the syntax and semantics for messages, allowing agents to perform actions like requesting information, making proposals, and accepting or rejecting tasks [?].

Therefore, this thesis leverages the MAS paradigm by designing a framework composed of three specialized, collaborative agents. Their individual, goal-directed behaviors, orchestrated within a hybrid architecture, work in concert to achieve the overarching systemic objective: transforming institutional mental health support from a reactive model to a proactive, data-driven ecosystem.

### 2.2.3 Large Language Models (LLMs)

Large Language Models (LLMs) are a class of deep learning models that have demonstrated remarkable capabilities in understanding and generating human-like text. The architectural foundation for virtually all modern LLMs, including the Gemma and Gemini models used in this research, is the **Transformer architecture**, first introduced by Vaswani et al. [?]. The Transformer's key innovation is the **self-attention mechanism**, which allows the model to dynamically weigh the importance of different words in an input sequence when processing and generating language. This enables the model to capture complex, long-range dependencies and contextual relationships far more effectively than its predecessors, such as Recurrent Neural Networks (RNNs) [?].

The core operation of a Transformer-based model involves processing input text through a series of encoding and/or decoding layers. In a generative, decoder-only model like Gemma, the process can be conceptualized as follows:

1. **Tokenization and Embedding:** Input text is first broken down into smaller units called tokens. Each token is then mapped to a high-dimensional vector, or an "embedding," that represents its semantic meaning.
2. **Positional Encoding:** Since the self-attention mechanism does not inherently process sequential order, a positional encoding vector is added to each token embedding to provide the model with information about the word's position in the sequence.
3. **Self-Attention Layers:** The sequence of embeddings passes through multiple self-attention layers. In each layer, the model calculates attention scores for every token relative to all other tokens in the sequence, effectively learning which parts of the input are most relevant for understanding the context of each specific token.
4. **Feed-Forward Networks:** Each attention layer is followed by a feed-forward neural network that applies further transformations to each token's representation.
5. **Output Generation:** The model's final output is a probability distribution over its entire vocabulary for the next token in the sequence. The model then typically selects the most likely token (or samples from the distribution) and appends it to the input, repeating the process autoregressively to generate coherent text [?].

**Placeholder for Diagram: Simplified Transformer Architecture for Generative LLMs**

- This diagram should illustrate the flow of information:
1. Input Text -> Tokenizer
  2. Tokens -> Embedding Layer + Positional Encoding
  3. Embedded Tokens -> A stack of 'N' Decoder Blocks
  4. Inside a Decoder Block: Multi-Head Self-Attention -> Feed-Forward Network
  5. Final Output -> Linear Layer -> Softmax -> Probability of Next Token

Figure 2.1. A simplified view of the decoder-only Transformer architecture used in generative LLMs like Gemma. The model processes input embeddings through multiple layers of self-attention and feed-forward networks to predict the next token in a sequence.

This research utilizes a **hybrid LLM strategy** that leverages two distinct families of models based on this architecture to balance performance, privacy, and capability:

- **Locally-Hosted Open Models (Gemma):** The Gemma models are a family of lightweight, open-weight models developed by Google, built from the same research and technol-

ogy used to create the Gemini models [?]. As decoder-only Transformers, they are optimized for generative text tasks. Being "open-weight" means their parameters are publicly available, allowing them to be deployed on institutional hardware. This approach is critical for this project as it guarantees data privacy and security—sensitive student conversations processed by the Triage Agent never leave the university’s secure servers. Furthermore, local hosting provides low latency and eliminates per-token API costs, making it a sustainable choice for real-time, high-volume interactions.

- **Cloud-Based API Models (Gemini):** The Gemini models represent Google’s state-of-the-art, natively multimodal foundation models, available in various sizes (e.g., Gemini Pro). Unlike models trained solely on text, Gemini was pre-trained from the ground up on multiple data modalities, giving it more sophisticated reasoning capabilities [?]. In this framework, a powerful model like Gemini Pro is accessed via a secure API for complex, non-sensitive tasks, such as the weekly trend analysis performed by the Analytics Agent. It also serves as a robust fallback mechanism, ensuring service continuity and reliability should the local model encounter issues.

### 2.2.3.1 Locally-Hosted Open Models: The Gemma 3 Family

The selection of Gemma 3 as the locally-hosted model for this framework is predicated on its state-of-the-art capabilities, efficiency, and suitability for advanced, multimodal tasks. Released in March 2025, Gemma 3 represents a significant evolution in Google’s open model family, introducing several key advancements over its predecessors. It is available in multiple sizes (1B, 4B, 12B, and 27B), allowing for a flexible trade-off between performance and resource requirements. For this research, the 4B or 12B models are particularly relevant, offering powerful capabilities that can be reasonably deployed on institutional-grade hardware.

The architecture of Gemma 3 is based on the decoder-only Transformer. The foundational mechanism of the Transformer is scaled dot-product attention, calculated as:

$$\text{Attention}(Q, K, V) = \text{softmax} \left( \frac{QK^T}{\sqrt{d_k}} \right) V$$

where  $Q$  (Queries),  $K$  (Keys), and  $V$  (Values) are matrices representing the input sequence, and  $d_k$  is the dimension of the keys. To capture different relational aspects of the context, this is extended to **Multi-Head Attention (MHA)**, where multiple attention "heads" operate in parallel. For each head  $i$  out of  $h$  total heads, separate linear projections are learned:  $W_i^Q, W_i^K, W_i^V$ . The output of each head is:

$$\text{head}_i = \text{Attention}(QW_i^Q, KW_i^K, VW_i^V)$$

The final output is the concatenation of all head outputs, passed through a final linear

projection  $W^O$ :

$$\text{MHA}(Q, K, V) = \text{Concat}(\text{head}_1, \dots, \text{head}_h)W^O$$

While powerful, MHA is memory-intensive during inference because it requires caching the Key and Value matrices for each of the  $h$  heads.

Gemma 3’s architecture incorporates several major innovations to enhance performance and efficiency:

- **Efficient Attention Mechanisms:** Gemma 3 utilizes **Grouped-Query Attention (GQA)**, a more efficient variant of MHA [?]. In GQA, instead of having  $h$  independent key and value heads, the query heads are divided into  $g$  groups, where all heads within a group share the same key and value projection. This significantly reduces the size of the KV cache from being proportional to  $h$  to being proportional to  $g$  (where  $g < h$ ), lowering memory requirements and enabling faster inference.
- **Extended Context Window:** Gemma 3 supports a significantly larger context window of up to 128,000 tokens. This is made computationally feasible by architectural efficiencies like GQA and the use of local, sliding-window attention layers, which restrict the attention calculation to a smaller, fixed-size window of tokens for most layers [?].
- **Multimodal Input (Image and Text):** Unlike previous text-only versions, Gemma 3 models (with the exception of the 1B variant) are inherently multimodal. They integrate a 400M parameter variant of the SigLIP vision encoder, which allows the model to process and understand images as part of its input context [?].
- **Advanced Agentic Capabilities:** Gemma 3 has built-in function calling capabilities, allowing it to interact with external tools and APIs in a structured manner. This is essential for the Intervention and Triage agents to execute tasks beyond simple text generation [?].

This combination of multimodal understanding, long-context efficiency, and agentic capabilities makes Gemma 3 a powerful and highly suitable choice for the locally-hosted component of this thesis’s hybrid AI strategy.

#### 2.2.3.2 Cloud-Based API Models: The Gemini 2.5 Family

To complement the locally-hosted model, the framework integrates a state-of-the-art, proprietary model accessed via a cloud API. The Gemini family, specifically the flagship **Gemini 2.5 Pro** model, serves this role, providing a level of reasoning and multimodal understanding that is critical for handling the most complex tasks and ensuring system robustness. While a detailed architectural schematic is not public, in line with the proprietary nature of frontier AI models, its capabilities have been extensively docu-

mented by Google through official developer guides and announcements [?].

Gemini 2.5 Pro builds upon the efficient **Mixture-of-Experts (MoE) Transformer** architecture of its predecessors. In an MoE architecture, the model is composed of numerous smaller "expert" neural networks. For any given input, a routing mechanism activates only a sparse subset of these experts. This allows the model to have a very large total parameter count—enabling vast knowledge and capability—while keeping the computational cost for any single inference relatively low.

The strategic role of Gemini 2.5 Pro in this framework is defined by its next-generation capabilities:

- **Native Multimodality with Expressive Audio:** A significant architectural leap in Gemini 2.5 is its native handling of audio. Unlike models that first transcribe audio to text, Gemini 2.5 processes audio streams directly. This allows it to understand not just the words, but also the nuances of human speech such as tone, pitch, and prosody, which is invaluable for a mental health application where user sentiment is key [?].
- **Controllable Reasoning and "Thinking Time":** Gemini 2.5 introduces a "thinking budget," a mechanism that allows developers to control the trade-off between response latency and reasoning depth. For high-frequency tasks performed by the Triage Agent, a lower budget can ensure speed. For complex analytical tasks required by the Analytics Agent, a higher budget can be allocated to allow for more thorough reasoning, providing granular control over both cost and quality.
- **Advanced Agentic Capabilities and Tool Use:** The model is explicitly designed to power advanced agents. It features more reliable and sophisticated function calling, enabling seamless integration with external tools and APIs. This is essential for the Intervention Agent to execute multi-step plans, such as triggering an email campaign based on analytical insights.
- **High-Fidelity Reasoning and Fallback:** As a frontier model, Gemini 2.5 Pro serves as the high-capability engine for the most demanding requests and as a crucial fallback mechanism. If the local Gemma 3 model fails or returns a low-confidence result, the system can route the query to the Gemini API, ensuring service continuity and the highest quality output.

By integrating Gemini 2.5 Pro via its API, the agentic framework gains access to state-of-the-art reasoning power on demand, ensuring that it can handle a wide spectrum of tasks with both efficiency and exceptional quality.

### 2.2.3.3 Multilingual Understanding: The XLM-RoBERTa Model

To address the multilingual nature of the chat logs, which are expected to contain English, Indonesian, and code-mixed language, a specialized model is required for the

Safeguard LLM component. **XLM-RoBERTa (XLM-R)** is an encoder-only Transformer model specifically designed for high-performance cross-lingual understanding [?]. Its architecture is based on RoBERTa, which itself is a robustly optimized version of BERT, but its key innovation lies in its training methodology.

XLM-R is pre-trained on a massive, curated dataset of over 2.5 terabytes of text from the CommonCrawl corpus, spanning 100 different languages, including Indonesian. Unlike previous cross-lingual models that required parallel data or explicit translation objectives, XLM-R’s power comes from being trained at scale using only a **Masked Language Modeling (MLM)** objective.

The mathematical intuition behind the MLM objective is to predict a randomly masked token in a sequence based on its surrounding context. Given a sequence of tokens  $X = \{x_1, x_2, \dots, x_n\}$ , a corrupted version  $\tilde{X}$  is created by replacing a subset of tokens with a special ‘[MASK]’ token. The model is then trained to minimize the cross-entropy loss between its predictions and the original tokens. The objective function can be expressed as:

$$\mathcal{L}_{MLM}(\theta) = - \sum_{i \in \mathcal{M}} \log p(x_i | \tilde{X}; \theta)$$

where  $\mathcal{M}$  is the set of indices of the masked tokens and  $\theta$  represents the model’s parameters.

By applying this objective to a vast corpus containing 100 languages, the model is forced to develop a shared, high-dimensional embedding space where similar concepts from different languages are mapped to nearby vectors. For example, the model learns that the English word "student" and the Indonesian word "mahasiswa" often appear in similar contexts (e.g., with words like "university," "learn," "exam") and thus assigns them similar vector representations.

This learned shared representation enables a powerful capability known as **zero-shot cross-lingual transfer**. It means that a model fine-tuned for a specific task (like Named Entity Recognition for PII) on a single language (e.g., English) can then perform that same task on other languages it was pre-trained on (e.g., Indonesian) with a high degree of accuracy, often without seeing any task-specific examples in that second language. This makes XLM-RoBERTa an ideal candidate for the generalist component of the Safeguard LLM pipeline, capable of handling mixed-language text where specialized monolingual models might fail.

#### 2.2.4 LLM Orchestration Frameworks (LangChain)

While LLMs provide powerful reasoning capabilities, they are inherently stateless and lack direct access to external data or tools. An LLM, in isolation, cannot query a database, call an API, or access a private document. To build sophisticated, stateful applications that overcome these limitations, an orchestration framework is required. **LangChain** is an open-source framework designed specifically for this purpose, providing the essential "glue" to connect LLMs with external resources and compose them into complex applications [?].

The core philosophy of LangChain is to provide modular components that can be "chained" together to create complex workflows. The most recent and fundamental abstraction in LangChain is the **LangChain Expression Language (LCEL)**. LCEL provides a declarative, composable syntax for building chains, where the pipe ('|') operator streams the output of one component into the input of the next. Every component in an LCEL chain is a "Runnable," a standardized interface that supports synchronous, asynchronous, batch, and streaming invocations, making it highly versatile for production environments [?].

A simple LCEL chain can be represented as:

$$\text{Chain} = \text{PromptTemplate} \mid \text{LLM} \mid \text{OutputParser}$$

In this sequence, user input is first formatted by a 'PromptTemplate', the result is passed to the 'LLM' for processing, and the LLM's raw output is then transformed into a structured format (e.g., JSON) by an 'OutputParser'.

For this thesis, the most critical application of LangChain is its ability to create **agents**. A LangChain agent uses an LLM not just for text processing, but as a reasoning engine to make decisions. This is often based on a framework known as **ReAct (Reasoning and Acting)**, which enables the LLM to synergize reasoning and action [?]. The agent is given access to a set of **Tools**—which are simply functions that can interact with the outside world (e.g., a database query function, a file reader, a web search API). The agent's operational loop, managed by an **Agent Executor**, can be formalized as an iterative process.

Let  $G$  be the initial goal and  $H_t$  be the history of actions and observations up to step  $t$ . The process at each step  $t$  is:

1. **Reasoning (Thought Generation):** The agent generates a thought  $th_t$  and a subsequent action  $a_t$  by sampling from the LLM's conditional probability distribution,

given the goal and the history so far.

$$(th_t, a_t) \sim p(th, a|G, H_{t-1}; \theta_{LLM})$$

The prompt to the LLM contains the goal and the trajectory of previous thoughts, actions, and observations, guiding its next decision.

2. **Action Execution:** The Agent Executor parses  $a_t$  to identify the chosen tool and its input, then executes it to produce an observation,  $o_t$ .

$$o_t = \text{ExecuteTool}(a_t)$$

3. **History Augmentation:** The new observation is appended to the history, forming the context for the next iteration.

$$H_t = H_{t-1} \oplus (a_t, o_t)$$

This loop continues until the LLM determines the goal  $G$  is met and generates a final answer.

This iterative loop is what transforms a passive LLM into a proactive, problem-solving agent. For example, the **Analytics Agent** in this framework, when tasked with "summarizing student stress trends," would use this loop to formulate a SQL query (Thought and Action), execute it (Observation), and then use the results to generate a final summary. This orchestration is fundamental to enabling the autonomous capabilities central to this thesis.

### 2.2.5 Workflow Automation Platforms (n8n)

If an LLM orchestration framework like LangChain provides the "brain" for an individual agent, a workflow automation platform provides the "nervous system" for the entire multi-agent framework. While LangChain excels at creating complex, stateful logic within a single application, workflow automation platforms are designed to connect disparate systems and orchestrate processes across them. This research utilizes **n8n**, an open-source, node-based workflow automation tool, to manage the high-level scheduling and system-to-system communication required by the agentic framework [?].

The theoretical foundation for such platforms can be understood through the concept of a **Directed Acyclic Graph (DAG)**. A workflow in n8n is a DAG where each node represents a specific task or operation, and the directed edges represent the flow of data and execution from one node to the next. A workflow,  $W$ , can be formally defined as a graph  $W = (N, E)$ , where:



- $N$  is a set of nodes, where each node  $n \in N$  represents a discrete unit of work (e.g., an API call, a database query, an email function).
- $E$  is a set of directed edges  $(n_i, n_j)$ , where an edge represents that the output of node  $n_i$  becomes the input for node  $n_j$ .

The acyclic nature of the graph ensures that workflows have a clear start and end and do not enter into infinite loops.

The key components of n8n that are leveraged in this thesis are:

1. **Trigger Nodes:** These are the starting points of a workflow and are activated by specific events. For this research, the most critical trigger is the **Cron node**, which allows for time-based scheduling. This is used to activate the **Analytics Agent** at regular intervals (e.g., every Sunday at midnight) to ensure consistent, automated trend analysis without manual intervention.
2. **Regular Nodes:** These nodes perform the actual work. Each node is an abstraction for a specific service or function. For instance, an **HTTP Request node** is used to call the API endpoints exposed by the FastAPI backend, thereby invoking the LangChain agents. Other nodes, like the **Gmail node**, can be used to send notifications or reports generated by the agents.
3. **Visual Workflow Management:** Unlike writing complex scripts, n8n provides a visual interface for building, debugging, and modifying these workflows. This visual representation of the DAG makes the high-level logic of the system transparent and easy to manage, which is crucial for maintaining a complex multi-agent system.

In this architecture, n8n does not concern itself with the internal reasoning of the agents (which is handled by LangChain), but rather with the high-level orchestration: when to wake an agent up, how to pass data between agent calls, and how to connect the agentic framework to external services like email. This separation of concerns creates a robust and modular system.

### 2.3 Synthesis and Identification of the Research Gap

The preceding review of the literature and theoretical landscape reveals a critical disconnect. On one hand, the field has produced increasingly sophisticated but fundamentally **reactive** conversational agents for mental health. On the other, it has developed proactive institutional analytics that remain bottlenecked by a reliance on **manual intervention**. The failure of the existing literature is not in the individual components, but in the lack of integration between them. This creates a significant and unaddressed research gap: the need for an **integrated, autonomous, and proactive framework** that can systemically bridge the chasm from data-driven insight to automated, personalized

intervention.

Current systems are not designed as a cohesive ecosystem. The analytical tools do not automatically trigger the intervention tools, and the conversational agents do not operate with an awareness of the broader institutional context provided by analytics. The central argument of this thesis is that the next frontier in institutional mental health support lies not in the incremental improvement of any single component, but in the **synergistic integration of multiple specialized agents** into a single, closed-loop system. Such a system, architected as a Multi-Agent System (MAS), is capable of emergent behaviors that are more than the sum of its parts.

Therefore, this research directly addresses the identified gap by proposing and prototyping a novel agentic AI framework where:

- An **Analytics Agent** autonomously identifies trends, moving beyond the static dashboards of current well-being analytics and creating actionable intelligence.
- An **Intervention Agent** acts on this intelligence by automating targeted outreach campaigns, explicitly closing the "insight-to-action" gap that renders current proactive models inefficient and unscalable.
- A **Triage Agent** provides immediate, context-aware support, functioning as the intelligent and ever-present front-door to the entire support ecosystem, overcoming the limitations of purely reactive chatbots.

By designing and evaluating a system where these agents work in concert, this thesis pioneers a holistic solution that is fundamentally more proactive, scalable, and efficient than the disparate tools described in the current literature.

## CHAPTER III

### SYSTEM DESIGN AND ARCHITECTURE

#### 3.1 Research Methodology: Design Science Research (DSR)

The research presented in this thesis is constructive in nature, aimed not merely at describing or explaining a phenomenon, but at creating a novel and useful artifact to solve a real-world problem. To provide a rigorous and systematic structure for this endeavor, this study adopts the **Design Science Research (DSR)** methodology. DSR is a well-established paradigm in Information Systems research focused on the creation and evaluation of innovative IT artifacts intended to solve identified organizational problems [?]. The primary goal of DSR is to generate prescriptive design knowledge through the building and evaluation of these artifacts.

The DSR process model, as outlined by Peffers et al., provides an iterative framework that guides the research from problem identification to the communication of results [?]. This thesis follows these stages, mapping them directly to its structure to ensure a logical and transparent research process:

1. **Problem Identification and Motivation:** This initial stage, which involves defining the specific research problem and justifying the value of a solution, is addressed in **Chapter ??** of this thesis. We have identified the inefficiencies of the reactive mental health support model as the core problem.
2. **Definition of Objectives for a Solution:** Based on the identified problem, this stage involves defining the objectives and desired capabilities of the artifact. These objectives, which center on creating a proactive, automated, and data-driven framework, are also detailed in **Chapter ??**.
3. **Design and Development:** This is the core constructive phase where the artifact's architecture and functionalities are developed. This stage is the primary focus of the present chapter, **Chapter ??**, which outlines the functional and technical blueprint of the agentic AI framework.
4. **Demonstration:** In this stage, the designed artifact is demonstrated to solve one or more instances of the problem. This will be accomplished through the implementation of a functional prototype, as will be detailed in **Chapter IV**.
5. **Evaluation:** This stage involves observing and measuring how well the artifact supports a solution to the problem. The prototype's capabilities will be evaluated against predefined functional scenarios in **Chapter IV**, with the findings and their implications discussed.

6. **Communication:** The final stage involves communicating the problem, the artifact, and its utility to a relevant audience. This entire thesis document serves as the primary communication artifact for this research.

The complete workflow of this research, following the DSR methodology, is visualized in Figure ???. This diagram illustrates the iterative path from problem formulation through to the final conclusions and recommendations.

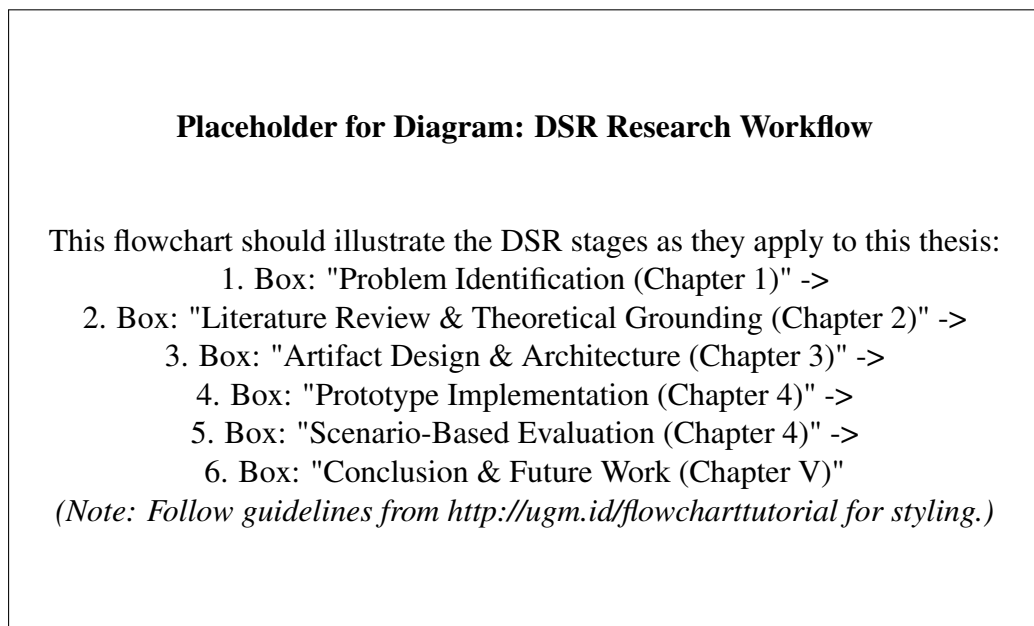


Figure 3.2. The Design Science Research (DSR) process model as applied in this thesis.

### 3.2 System Overview and Conceptual Design

**Purpose:** To provide a high-level, 10,000-foot view of the system's purpose and its main components, making it understandable before diving into technical specifics.

**Elaboration Points:**

Present a concise paragraph describing the conceptual model: "The proposed framework is an ecosystem of three collaborative, intelligent agents that work in concert to transform an institution's mental health support from a reactive to a proactive model..."

Include a high-level Context Diagram showing the main entities (Students, University Staff/Counselors) and systems (UGM-AICare User App, Agentic AI Backend, Admin Dashboard) and how they interact.

### 3.3 Functional Architecture: The Agentic Core

**Purpose:** To detail the "what" of the system. This section explains the specific roles and functions of each agent. This is the heart of your functional design.

Elaboration Points: Create a subsection for each of the three agents. For each, define its:

### **3.3.1 The Analytics Agent**

Goal: To autonomously identify mental health trends from anonymized user data.

Perception (Inputs): Anonymized conversation logs from the PostgreSQL database.

Processing Logic: Describe the NLP tasks it performs: topic modeling, sentiment analysis, and summarization.

Action (Outputs): A structured weekly report (in JSON format) containing key insights (e.g., top 5 trending stress topics, overall sentiment score).

### **3.3.2 The Intervention Agent**

Goal: To automate targeted, proactive outreach campaigns.

Perception (Inputs): The structured report from the Analytics Agent and a set of predefined "campaign rules."

Processing Logic: A rule-based engine that maps insights to actions (e.g., IF 'exam stress' > threshold THEN trigger 'time-management-workshop' campaign).

Action (Outputs): A signal sent to the orchestration layer (n8n) with target audience segment and message content.

### **3.3.3 The Triage Agent**

Goal: To efficiently route a student to the most appropriate level of support in real-time.

Perception (Inputs): A user's live conversation with the chatbot.

Processing Logic: A real-time text classification model to determine the conversation's severity level (e.g., Level 1: Casual, Level 2: Moderate, Level 3: Red Flag).

Action (Outputs): A structured recommendation (e.g., suggest a self-help module, suggest booking a counselor, provide an emergency hotline).

Include a detailed Data Flow Diagram (DFD) showing how data moves between these three agents and the database.

## **3.4 Technical Architecture: The Hybrid System**

Purpose: To detail the "how" of the system—the engineering blueprint, including justification for key technology choices.

Elaboration Points:

3.4.1 Overall System Architecture Diagram

A detailed diagram showing the interplay between all technologies: Next.js (Admin Dashboard), FastAPI (Backend), PostgreSQL, LangChain (as a library), and n8n (as a separate, orchestrated service). Show the communication protocols (e.g., REST API, direct DB connection).

3.4.2 The "Brain": FastAPI + LangChain Service

Justify the choice of FastAPI (for performance, async support) and LangChain (for LLM orchestration). Detail the API design, defining key endpoints (e.g., /api/agents/generate-report) and their request/response schemas.

3.4.3 The "Nervous System": n8n Workflows

Justify the choice of n8n for robust workflow automation. Provide screenshots or diagrams of the primary n8n workflows (e.g., the "Weekly Report Generation" workflow triggered by a Cron job).

3.5 Database Design

Purpose: To define the data persistence layer of the system.

Elaboration Points:

Present a clean Entity-Relationship Diagram (ERD).

Table 3.1. Key Columns and Data Types for conversation\_logs Table

Column Name	Data Type	Description
id	SERIAL PRIMARY KEY	Unique identifier
user_id	UUID	Reference to user (anonymized)
timestamp	TIMESTAMP WITH TIME ZONE	Time of message
message	TEXT	User or agent message content
sender	VARCHAR(16)	'user' or 'agent'
sentiment_score	FLOAT	Sentiment analysis result
topic	VARCHAR(64)	NLP-inferred topic label

3.6 User Interface (UI) Design

Purpose: To show the design of the human interface for the system’s administrative users.

Elaboration Points:

Define the primary user persona for the dashboard (e.g., "Dr. Astuti, Head of Counseling Services").

Present wireframes or high-fidelity mockups for the key screens of the Admin Dashboard (e.g., the main analytics view, the report history page).

### **3.7 Security and Privacy by Design**

Purpose: To demonstrate that critical security and privacy considerations are integral to the architecture.

Elaboration Points:

Detail the Data Anonymization Pipeline: How is Personally Identifiable Information (PII) identified and redacted from chat logs before they are stored for analysis?

Describe the Role-Based Access Control (RBAC) mechanism for the admin dashboard.

Mention standard security practices like data encryption in transit (TLS) and at rest.

### **3.8 Alur Tugas Akhir**

Menguraikan prosedur yang akan digunakan dan jadwal atau alur penyelesaian setiap tahap. Alur penelian ini dapat disajikan dalam bentuk diagram. Diagram dapat disusun dengan aturan yang baik semisal menggunakan *flowchart*. Aturan dan tutorial pembuatan *flowchart* dapat dilihat di <http://ugm.id/flowcharttutorial>. Setelah menggambarkannya, penulis wajib menjelaskan langkah-langkah setiap alur tugas akhir dalam sub bab tersendiri sesuai dengan kebutuhan.

### **3.9 Etika, Masalah, dan Keterbatasan Penelitian (Opsional)**

Bagian ini membahas pertimbangan etis penelitian dan [potensi] masalah serta keterbatasannya. Jika menyangkut penelitian dengan makhluk hidup, maka dibutuhkan adanya *ethical clearance*, di bagian ini hal itu akan dibahas. Demikian juga tentang keterbatasan ataupun masalah yang akan timbul.

## **CHAPTER IV**

### **HASIL DAN PEMBAHASAN**

Berikut ini adalah yang perlu diperhatikan untuk mengisi bab hasil dan pembahasan:

1. Setiap rumusan masalah boleh memiliki lebih dari 1 tujuan.
2. Setiap subbab harus spesifik menjawab setiap tujuan yang dituliskan.
3. Setiap rumusan masalah boleh dijawab dengan 1 subbab atau lebih.

Berikut ini adalah contoh sub bab untuk menjelaskan tujuan penelitian.

#### **4.1 Pembahasan Tujuan 1 dengan Hasil Penelitian 1 (Ubah Judul Sesuai dengan Hal yang Hendak dibahas)**

Sub bab pertama adalah membahas tujuan penelitian pertama dengan hasil penelitian ke-1. Dapat ditambahkan beberapa sub bab jika diperlukan.

#### **4.2 Pembahasan Tujuan 1 dengan Hasil Penelitian 2 (Ubah Judul Sesuai dengan Hal yang Hendak dibahas)**

Sub bab kedua adalah membahas tujuan penelitian pertama dengan hasil penelitian ke-2. Sub bab ini merupakan contoh tambahan sub bab pertama.

#### **4.3 Pembahasan Tujuan 2 dengan Hasil Penelitian 3 (Ubah Judul Sesuai dengan Hal yang Hendak dibahas)**

Sub bab ketiga adalah membahas tujuan penelitian kedua. Dapat ditambahkan beberapa sub bab jika diperlukan.

#### **4.4 Perbandingan Hasil Penelitian dengan Hasil Terdahulu**

Pembahasan penutup dapat menjelaskan mengenai kelebihan hasil pengembangan / penelitian dan kekurangan dibandingkan dengan skripsi atau penelitian terdahulu atau perbandingan terhadap produk lain yang ada di pasaran. Penulis dapat menggunakan tabel untuk membandingkan secara gamblang dan menjelaskannya.



## **CHAPTER V**

### **TAMBAHAN (OPSIONAL)**

Anda boleh menambahkan Bab jika diperlukan. Jumlah Bab tidak harus sesuai dengan *template*.

Bab tambahan ini diperlukan jika hasil penelitian untuk menjawab tujuan cukup panjang atau terdiri dari banyak sub bab. Mahasiswa boleh menjawab 1 tujuan penelitian dengan 1 bab.

## **CHAPTER VI**

### **KESIMPULAN DAN SARAN**

#### **6.1 Kesimpulan**

Kesimpulan dapat diawali dengan apa yang dilakukan dengan tugas akhir ini lalu dilanjutkan dengan poin-poin yang menjawab tujuan penelitian, apakah tujuan sudah tercapai atau belum, tentunya berdasarkan data ataupun hasil dari Bab pembahasan sebelumnya. Dalam beberapa hal, kesimpulan dapat juga berisi tentang temuan/*findings* yang Anda dapatkan setelah melakukan pengamatan dan atau analisis terhadap hasil penelitian.

Kesimpulan menjawab seberapa jauh rumusan masalah tercapai berdasarkan hasil penelitian. Semua rumusan masalah harus disimpulkan berdasarkan data penelitian.

#### **6.2 Saran**

Saran berisi hal-hal yang bisa dilanjutkan dari penelitian atau skripsi ini, yang belum dilakukan karena batasan permasalahan. Saran bukan berisi saran kepada sistem atau pengguna, tetapi saran diberikan kepada aspek penelitian yang dapat dikembangkan dan ditambahkan di penelitian atau skripsi selanjutnya.

**Catatan: Mahasiswa perlu melihat sinkronisasi antara rumusan masalah, tujuan, metode, hasil penelitian, dan kesimpulan.**

Catatan: Daftar pustaka adalah apa yang dirujuk atau disitasi, bukan apa yang telah dibaca, jika tidak ada dalam sitasi maka tidak perlu dituliskan dalam daftar pustaka.

# LAMPIRAN

## L.1 Isi Lampiran

Lampiran bersifat opsional bergantung hasil kesepakatan dengan pembimbing dapat berupa:

1. Bukti pelaksanaan Kuesioner seperti pertanyaan kuesioner, resume jawaban responden, dan dokumentasi kuesioner.
2. Spesifikasi Aplikasi atau Sistem yang dikembangkan meliputi spesifikasi teknis aplikasi, tautan unduh aplikasi, manual penggunaan aplikasi, hingga screenshot aplikasi.
3. Cuplikan kode yang sekiranya penting dan ditambahkan.
4. Tabel yang terlalu panjang yang masih diperlukan tetapi tidak memungkinkan untuk ditayangkan di bagian utama skripsi.
5. Gambar-gambar pendukung yang tidak terlalu penting untuk ditampilkan di bagian utama. Akan tetapi, mendukung argumentasi/pengamatan/analisis.
6. Penurunan rumus-rumus atau pembuktian suatu teorema yang terlalu panjang dan terlalu teknis sehingga Anda berasumsi bahwa pembaca biasa tidak akan menelaah lebih lanjut. Hal ini digunakan untuk memberikan kesempatan bagi pembaca tingkat lanjut untuk melihat proses penurunan rumus-rumus ini.

## LAMPIRAN

### L.2 Panduan Latex

#### L.2.1 Syntax Dasar

##### L.2.1.1 Penggunaan Sitasi

Contoh penggunaan sitasi [?, ?] [?] [?] [?] [?, ?]

##### L.2.1.2 Penulisan Gambar

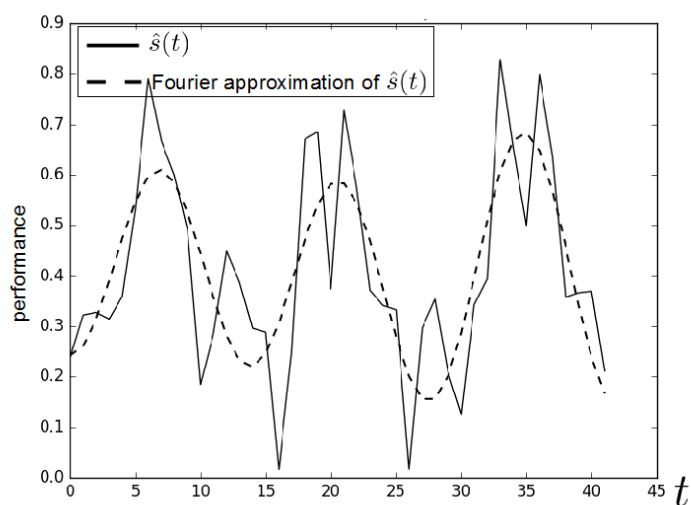


Figure 3. Contoh gambar.

Contoh gambar terlihat pada Gambar ??. Gambar diambil dari [?].

##### L.2.1.3 Penulisan Tabel

Table 1. Tabel ini

ID	Tinggi Badan (cm)	Berat Badan (kg)
A23	173	62
A25	185	78
A10	162	70

Contoh penulisan tabel bisa dilihat pada Tabel ??.

##### L.2.1.4 Penulisan formula

Contoh penulisan formula

$$L_{\psi_z} = \{t_i \mid v_z(t_i) \leq \psi_z\} \quad (1)$$

Contoh penulisan secara *inline*:  $PV = nRT$ . Untuk kasus-kasus tertentu, kita membutuhkan perintah "mathit" dalam penulisan formula untuk menghindari adanya jeda saat penulisan formula.

Contoh formula **tanpa** menggunakan "mathit":  $PVA = RTD$

Contoh formula **dengan** menggunakan "mathit":  $PVA = RTD$

### L.2.1.5 Contoh list

Berikut contoh penggunaan list

1. First item
2. Second item
3. Third item

## L.2.2 Blok Beda Halaman

### L.2.2.1 Membuat algoritma terpisah

Untuk membuat algoritma terpisah seperti pada contoh berikut, kita dapat memanfaatkan perintah *algstore* dan *algrestore* yang terdapat pada paket *algcompatible*. Pada dasarnya, kita membuat dua blok algoritma dimana blok pertama kita simpan menggunakan *algstore* dan kemudian di-restore menggunakan *algrestore* pada algoritma kedua. Perintah tersebut dimaksudkan agar terdapat kesinamungan antara kedua blok yang sejatinya adalah satu blok.

---

**Algorithm 1** Contoh algorima

---

```
1: procedure CREATESET( $v$ )  
2:   Create new set containing  $v$   
3: end procedure
```

---

Pada blok algoritma kedua, tidak perlu ditambahkan caption dan label, karena sudah menjadi satu bagian dalam blok pertama. Pembagian algoritma menjadi dua bagian ini berguna jika kita ingin menjelaskan bagian-bagian dari sebuah algoritma, maupun untuk memisah algoritma panjang dalam beberapa halaman.

---

```
4: procedure CONCATSET( $v$ )  
5:   Create new set containing  $v$   
6: end procedure
```

---

### L.2.2.2 Membuat tabel terpisah

Untuk membuat tabel panjang yang melebihi satu halaman, kita dapat mengganti kombinasi *table* + *tabular* menjadi *longtable* dengan contoh sebagai berikut.

Table 2. Contoh tabel panjang

header 1	header 2
foo	bar
foo	bar
foo	bar
foo	bar
foo	bar
foo	bar
foo	bar
foo	bar
foo	bar
foo	bar
foo	bar
foo	bar

### L.2.2.3 Menulis formula terpisah halaman

Terkadang kita butuh untuk menuliskan rangkaian formula dalam jumlah besar sehingga melewati batas satu halaman. Solusi yang digunakan bisa saja dengan memindahkan satu blok formula tersebut pada halaman yang baru atau memisah rangkaian formula menjadi dua bagian untuk masing-masing halaman. Cara yang pertama mungkin akan menghasilkan alur yang berbeda karena ruang kosong pada halaman pertama akan diisi oleh teks selanjutnya. Sehingga di sini kita dapat memanfaatkan *align* yang sudah diatur dengan mode *allowdisplaybreaks*. Penggunaan *align* ini memungkinkan satu rangkaian formula terpisah berbeda halaman.

Contoh sederhana dapat digambarkan sebagai berikut.

$$\begin{aligned}
 x &= y^2 \\
 x &= y^3 \\
 a + b &= c \\
 x &= y - 2 \\
 a + b &= d + e \\
 x^2 + 3 &= y \\
 a(x) &= 2x
 \end{aligned}
 \tag{2}$$

$$b_i = 5x$$

$$10x^2 = 9x$$

$$2x^2 + 3x + 2 = 0$$

$$5x - 2 = 0$$

$$d = \log x$$

$$y = \sin x$$



## LAMPIRAN

### L.3 Format Penulisan Referensi

Penulisan referensi mengikuti aturan standar yang sudah ditentukan. Untuk internasionalisasi DTETI, maka penulisan referensi akan mengikuti standar yang ditetapkan oleh IEEE (*International Electronics and Electrical Engineers*). Aturan penulisan ini bisa diunduh di <http://www.ieee.org/documents/ieeecitationref.pdf>. Gunakan Mendeley sebagai *reference manager* dan *export* data ke format Bibtex untuk digunakan di Latex.

Berikut ini adalah sampel penulisan dalam format IEEE:

#### L.3.1 Book

##### Basic Format:

- [1] J. K. Author, "Title of chapter in the book," in Title of His Published Book, xth ed. City of Publisher, Country: Abbrev. of Publisher, year, ch. x, sec. x, pp. xxx-xxx.

##### Examples:

- [1] B. Klaus and P. Horn, Robot Vision. Cambridge, MA: MIT Press, 1986.
- [2] L. Stein, "Random patterns," in Computers and You, J. S. Brake, Ed. New York: Wiley, 1994, pp. 55-70.
- [3] R. L. Myer, "Parametric oscillators and nonlinear materials," in Nonlinear Optics, vol. 4, P. G. Harper and B. S. Wherret, Eds. San Francisco, CA: Academic, 1977, pp. 47-160.
- [4] M. Abramowitz and I. A. Stegun, Eds., Handbook of Mathematical Functions (Applied Mathematics Series 55). Washington, DC: NBS, 1964, pp. 32-33.
- [5] E. F. Moore, "Gedanken-experiments on sequential machines," in Automata Studies (Ann. of Mathematical Studies, no. 1), C. E. Shannon and J. McCarthy, Eds. Princeton, NJ: Princeton Univ. Press, 1965, pp. 129-153.
- [6] Westinghouse Electric Corporation (Staff of Technology and Science, Aerospace Div.), Integrated Electronic Systems. Englewood Cliffs, NJ: Prentice-Hall, 1970.
- [7] M. Gorkii, "Optimal design," Dokl. Akad. Nauk SSSR, vol. 12, pp. 111-122, 1961 (Transl.: in L. Pontryagin, Ed., The Mathematical Theory of Optimal Processes. New York: Interscience, 1962, ch. 2, sec. 3, pp. 127-135).
- [8] G. O. Young, "Synthetic structure of industrial plastics," in Plastics, vol. 3,

Polymers of Hexadromicon, J. Peters, Ed., 2nd ed. New York: McGraw-Hill, 1964, pp. 15-64.

### **L.3.2 Handbook**

#### **Basic Format:**

- [1] Name of Manual/Handbook, x ed., Abbrev. Name of Co., City of Co., Abbrev. State, year, pp. xx-xx.

#### **Examples:**

- [1] Transmission Systems for Communications, 3rd ed., Western Electric Co., Winston Salem, NC, 1985, pp. 44-60.
- [2] Motorola Semiconductor Data Manual, Motorola Semiconductor Products Inc., Phoenix, AZ, 1989.
- [3] RCA Receiving Tube Manual, Radio Corp. of America, Electronic Components and Devices, Harrison, NJ, Tech. Ser. RC-23, 1992.

### **Conference/Prosiding**

#### **Basic Format:**

- [1] J. K. Author, "Title of paper," in Unabbreviated Name of Conf., City of Conf., Abbrev. State (if given), year, pp.xxx-xxx.

#### **Examples:**

- [1] J. K. Author [two authors: J. K. Author and A. N. Writer ] [three or more authors: J. K. Author et al.], "Title of Article," in [Title of Conf. Record as ], [copyright year] © [IEEE or applicable copyright holder of the Conference Record]. doi: [DOI number]

### **Sumber Online/Internet**

#### **Basic Format:**

- [1] J. K. Author. (year, month day). Title (edition) [Type of medium]. Available: [http://www.\(URL\)](http://www.(URL))

#### **Examples:**

- [1] J. Jones. (1991, May 10). Networks (2nd ed.) [Online]. Available: <http://www.atm.com>

### **Skripsi, Tesis dan Disertasi**

#### **Basic Format:**

- [1] J. K. Author, "Title of thesis," M.S. thesis, Abbrev. Dept., Abbrev. Univ., City of Univ., Abbrev. State, year.

[2] J. K. Author, "Title of dissertation," Ph.D. dissertation, Abbrev. Dept., Abbrev. Univ., City of Univ., Abbrev. State, year.

**Examples:**

[1] J. O. Williams, "Narrow-band analyzer," Ph.D. dissertation, Dept. Elect. Eng., Harvard Univ., Cambridge, MA, 1993. [2] N. Kawasaki, "Parametric study of thermal and chemical nonequilibrium nozzle flow," M.S. thesis, Dept. Electron. Eng., Osaka Univ., Osaka, Japan, 1993

## LAMPIRAN

### L.4 Contoh Source Code

#### L.4.1 Sample algorithm

---

**Algorithm 2** Kruskal's Algorithm

---

```
1: procedure MAKESET( $v$ )
2:   Create new set containing  $v$ 
3: end procedure
4:
5: function FINDSET( $v$ )
6:   return a set containing  $v$ 
7: end function
8:
9: procedure UNION( $u, v$ )
10:  Unites the set that contain  $u$  and  $v$  into a new set
11: end procedure
12:
13: function KRUSKAL( $V, E, w$ )
14:   $A \leftarrow \{\}$ 
15:  for each vertex  $v$  in  $V$  do
16:    MakeSet( $v$ )
17:  end for
18:  Arrange  $E$  in increasing costs, ordered by  $w$ 
19:  for each  $(u, v)$  taken from the sorted list do
20:    if FindSet( $u$ )  $\neq$  FindSet( $v$ ) then
21:       $A \leftarrow A \cup \{(u, v)\}$ 
22:      Union( $u, v$ )
23:    end if
24:  end for
25:  return  $A$ 
26: end function
```

---

## L.4.2 Sample Python code

```
1 import numpy as np
2
3 def incmatrix (genl1 , genl2):
4     m = len (genl1)
5     n = len (genl2)
6     M = None #to become the incidence matrix
7     VT = np.zeros ((n*m,1) , int) #dummy variable
8
9     #compute the bitwise xor matrix
10    M1 = bitxormatrix (genl1)
11    M2 = np.triu (bitxormatrix (genl2) ,1)
12
13    for i in range (m-1):
14        for j in range (i+1, m):
15            [r,c] = np.where (M2 == M1[i , j])
16            for k in range (len (r)):
17                VT[(i)*n + r[k]] = 1;
18                VT[(i)*n + c[k]] = 1;
19                VT[(j)*n + r[k]] = 1;
20                VT[(j)*n + c[k]] = 1;
21
22    if M is None:
23        M = np.copy (VT)
24    else:
25        M = np.concatenate ((M, VT) , 1)
26
27    VT = np.zeros ((n*m,1) , int)
28
29    return M
```

### L.4.3 Sample Matlab code

```
1 function X = BitXorMatrix(A,B)
2 %function to compute the sum without charge of two vectors
3
4 %convert elements into unsigned integers
5 A = uint8(A);
6 B = uint8(B);
7
8 m1 = length(A);
9 m2 = length(B);
10 X = uint8(zeros(m1, m2));
11 for n1=1:m1
12     for n2=1:m2
13         X(n1, n2) = bitxor(A(n1), B(n2));
14     end
15 end
```