ĐẠI HỌC QUỐC GIA TP. HỒ CHÍ MINH
TRƯỜNG ĐẠI HỌC CÔNG NGHỆ THÔNG TIN
KHOA CÔNG NGHỆ PHẦN MỀM

PROJECT CHARTER

UIT AI ASSISTANT

NHÓM THỰC HIỆN:
Quách Gia Kiệt - 23520819
Nguyễn Tuấn Kiệt - 23520815

GV HƯỚNG DẪN:
Th.S Nguyễn Công Hoan

TP. HỒ CHÍ MINH, 2025

Document Control

Document Information

- **Document Id:** UIT.SE.AIAGENT-2025
- **Document Owner:** Quách Gia Kiệt
- **Issue Date:** 28/12/2024
- **Last Saved Date:** 28/12/2024
- **File Name:** UIT.SE.AIAGENT - Project Charter.pdf

Document History

Version	Issue Date	Changes
1.0	28/12/2024	Initial version

Document Approvals

Role	Name	Signature	Date
Project Sponsor	Th.S Nguyễn Công Hoan	✓	
Project Manager	Quách Gia Kiệt	✓	
Product Manager	Nguyễn Tuấn Kiệt	✓	

Acronyms List

Acronym	Full Form	Description
PM	Project Manager	Người chịu trách nhiệm quản lý dự án
RAG	Retrieval-Augmented Generation	Kỹ thuật kết hợp retrieval và generation
MCP	Model Context Protocol	Giao thức chuẩn hóa tool integration
LLM	Large Language Model	Mô hình ngôn ngữ lớn
DAA	Điểm - Attendance - Assignment	Portal học vụ của UIT
MVP	Minimum Viable Product	Sản phẩm tối thiểu khả thi

Mục lục

Document Control **1**

1 Project Summary, Purpose, Goals, and Success Criteria **4**
1.1 Summary 4
1.2 Purpose and Vision 4
1.2.1 Purpose 4
1.2.2 Vision 4
1.3 Project Goals and Objectives 5
1.3.1 Goals (SMART) 5
1.3.2 Objectives 5
1.4 Success Criteria 7

2 Project Scope, Quality Management, and Timeline **8**
2.1 Project Scope 8
2.1.1 Inclusions 8
2.1.2 Exclusions 9
2.2 Quality Management 10
2.2.1 Standards 10
2.2.2 Control Procedures 10
2.2.3 Cam kết bảo mật (NDA) 11
2.3 Timeline and Milestones 11

3 Resources, Costs, and Budget **12**
3.1 Resource Management 12
3.2 Estimated Costs and Budget 12

4 Stakeholders **13**

5 Risk Management **14**

6 Additional Information **16**
6.1 Technical Stack 16
6.1.1 Backend Technologies 16
6.1.2 AI & ML Technologies 16
6.1.3 Data Layer 16
6.1.4 Frontend Technologies 16
6.1.5 Infrastructure & Communication 17
6.2 Current Progress 17
6.3 Next Steps 17

Chương 1 Project Summary, Purpose, Goals, and Success Criteria

1.1 Summary

UIT AI Assistant là một dự án phát triển trợ lý ảo AI toàn diện, được thiết kế riêng để hỗ trợ sinh viên trường Đại học Công nghệ Thông tin (UIT). Hệ thống sử dụng các công nghệ AI tiên tiến như RAG (Retrieval-Augmented Generation), LangGraph agent orchestration, và Model Context Protocol để cung cấp trải nghiệm hỗ trợ thông minh, liền mạch cho sinh viên.

Agent này sẽ là một người bạn đồng hành, giúp giải quyết các vấn đề từ học vụ (tra cứu quy định, chương trình đào tạo, điểm số, lịch thi), quản lý thời gian cá nhân, đến hỗ trợ học tập, nhằm đơn giản hóa và nâng cao trải nghiệm học tập tại trường.

1.2 Purpose and Vision

1.2.1 Purpose

Mục đích chính của dự án là xây dựng một công cụ tập trung, thông minh để giải đáp thắc mắc, tự động hóa các tác vụ hành chính, và cung cấp hỗ trợ học tập cá nhân hóa cho sinh viên UIT. Dự án mong muốn giảm bớt gánh nặng cho sinh viên trong việc:

- Tìm kiếm thông tin về quy định, chính sách học vụ
- Tra cứu chương trình đào tạo, điều kiện tốt nghiệp
- Theo dõi điểm số, lịch thi, thời khóa biểu
- Quản lý công việc và thời gian học tập

Từ đó giúp sinh viên tập trung hơn vào việc học và phát triển bản thân.

1.2.2 Vision

Tầm nhìn của UIT AI Assistant là trở thành một trợ lý không thể thiếu đối với mỗi sinh viên UIT, đồng hành cùng họ từ lúc nhập học cho đến khi tốt nghiệp. Sản phẩm hướng tới việc xây dựng một hệ sinh thái hỗ trợ thông minh, liền mạch, giúp sinh viên khai thác tối đa tiềm năng của mình trong suốt quãng đời đại học.

1.3 Project Goals and Objectives

1.3.1 Goals (SMART)

- **Specific:** Hoàn thành phiên bản MVP (Minimum Viable Product) của AI Assistant có khả năng:
  - Trả lời câu hỏi về quy định và chương trình đào tạo thông qua RAG
  - Tra cứu điểm số, lịch thi, thời khóa biểu từ DAA portal
  - Giao tiếp với người dùng qua web interface
- **Measurable:** Hoàn thành 100% các tính năng đã đề ra cho MVP:
  - Knowledge base với 27 file PDF quy định + 100+ file Markdown chương trình đào tạo
  - 4 MCP tools hoạt động: retrieve_regulation, retrieve_curriculum, get_grades, get_schedule
  - Agent có thể demo thành công các luồng nghiệp vụ chính
  - Hệ thống deploy production với domain public
- **Achievable:** Các công nghệ (LangGraph, LlamaIndex, FastMCP, ChromaDB) đã được lựa chọn và triển khai thành công, có kế hoạch chi tiết theo 4 giai đoạn, đảm bảo tính khả thi về mặt kỹ thuật.
- **Relevant:** Dự án đáp ứng trực tiếp nhu cầu thực tế của sinh viên UIT, giúp giải quyết các vấn đề thường gặp trong quá trình học như tìm kiếm thông tin quy định, tra cứu điểm số, quản lý lịch học.
- **Time-bound:** Hoàn thành và báo cáo thành công phiên bản MVP trong vòng 4 tháng của Đồ án 1 (từ tháng 9/2024 đến tháng 1/1/2025).

1.3.2 Objectives

Dự án được chia thành 4 giai đoạn với mục tiêu cụ thể:

Giai đoạn 1: Setup project và build RAG knowledge base

- Setup ChromaDB + LlamaIndex
- Thu thập tài liệu từ các trang web của trường (27 quy định + 100+ chương trình đào tạo)

- Build data pipeline xử lý dữ liệu (processing → chunking → indexing)
- Build query engine (blended retrieval + reranking) để truy vấn

Giai đoạn 2: Build MCP Server với các tools

- Setup MCP Server với FastMCP
- Implement retrieval tools sử dụng query engine
- Implement scraping tools để lấy thông tin từ DAA (điểm số, thời khóa biểu, lịch thi)
- Test tool calling với MCP Inspector

Giai đoạn 3: Viết logic agent và kết nối với MCP tools

- Setup LangGraph agent với ReAct pattern
- Prompt engineering cho agent
- Kết nối MCP client với MCP Server
- Implement tool executor (parallel execution + timeout handling)
- State management với PostgreSQL checkpointer
- gRPC server để expose agent API

Giai đoạn 4: Build GUI, API server và deploy

- Build API Gateway (Go + Gin) expose REST API endpoints, kết nối với agent qua gRPC
- Build web frontend (React + Vite + Tailwind)
- Build browser extension (Svelte) để sync cookies từ DAA
- Deploy production lên VPS với domain public
- Setup CI/CD tự động deploy

1.4 Success Criteria

Dự án được coi là thành công khi đạt được các tiêu chí sau:

- **Giảng viên hướng dẫn hài lòng:** Đánh giá cao về mặt kỹ thuật, tính ứng dụng, và độ hoàn thiện của hệ thống.
- **Sản phẩm MVP hoạt động ổn định:**
  - Hệ thống chạy end-to-end: user → web → API Gateway → agent → MCP → RAG/DAA
  - Agent có thể trả lời đúng các câu hỏi về quy định và chương trình đào tạo
  - Agent có thể tra cứu thành công điểm số, lịch thi từ DAA
  - Web interface hoạt động mượt mà, responsive
- **Deployment thành công:**
  - Live production trên VPS với domain public
  - CI/CD pipeline tự động deploy khi có thay đổi
  - Hệ thống ổn định, không có lỗi critical
- **Documentation đầy đủ:**
  - Báo cáo đồ án chi tiết về kiến trúc, công nghệ, kết quả
  - Tài liệu kỹ thuật (tech stack, data preparation)
  - README và hướng dẫn cài đặt
- **Phản hồi tích cực:** Nhận được feedback tích cực từ nhóm sinh viên dùng thử (nếu có).

Chương 2 Project Scope, Quality Management, and Timeline

2.1 Project Scope

2.1.1 Inclusions

Các tính năng và nội dung sẽ làm trong dự án:

RAG System:

- Knowledge base chứa 27 file PDF quy định và 100+ file Markdown chương trình đào tạo
- Data pipeline: parsing (LlamaParse), cleaning, metadata generation (Gemini Flash), chunking, indexing
- Query engine: semantic search (ChromaDB) + reranking (ViRanker)
- Embeddings: OpenAI text-embedding-3-small

MCP Server:

- Retrieval tools: retrieve_regulation, retrieve_curriculum
- DAA integration tools: get_grades, get_schedule
- Expose tools qua HTTP endpoint /mcp

AI Agent:

- LangGraph agent với ReAct pattern
- Tool calling: gọi MCP tools và native tools
- State persistence với PostgreSQL checkpointer
- Conversation memory qua nhiều lượt chat
- gRPC server expose agent API

Backend Services:

- API Gateway: Go + Gin, REST endpoints, gRPC client

- MongoDB: lưu chat sessions và messages
- PostgreSQL: lưu agent state
- Redis: cache credentials và session data

Frontend:

- Web app: React + Vite + Tailwind, chat interface
- Browser extension: Svelte, sync cookies từ DAA

Infrastructure:

- Docker + Docker Compose: containerization
- VPS deployment với domain public
- CI/CD pipeline tự động deploy

2.1.2 Exclusions

Các tính năng và nội dung sẽ không làm trong Đồ án 1:

- Google Calendar integration
- Notion integration
- Study tracker (theo dõi tín chỉ chi tiết)
- Material search (tìm tài liệu học)
- Web search tool
- Notification scraping tool
- FAQ database (sẽ làm ở Đồ án 2)
- Dashboard hiển thị trực quan điểm số, thời khóa biểu
- Mobile app (chỉ có web + browser extension)
- Advanced analytics và reporting
- Multi-language support (chỉ tiếng Việt)

2.2 Quality Management

2.2.1 Standards

- **Accuracy:**
  - RAG trả lời đúng >70% câu hỏi về quy định và chương trình đào tạo trên bộ test
  - DAA scraping tools lấy đúng 100% dữ liệu điểm số, lịch thi
  - Agent gọi đúng tools theo prompt ≥80% trường hợp
- **Performance:**
  - Response time <10s cho query thông thường
  - Response time <5s cho DAA scraping (khi có cookie)
  - Hệ thống xử lý được ≥10 concurrent users
- **Reliability:**
  - Uptime ≥95% trong thời gian demo
  - Không có lỗi critical khi golive
  - Tối đa 3 lỗi minor (UI glitches, slow responses)
- **Security:**
  - Cookies được encrypt trước khi lưu vào Redis
  - API endpoints có rate limiting
  - Không log sensitive information (passwords, cookies)

2.2.2 Control Procedures

- **Code review:** Peer review sau mỗi feature hoàn thành
- **Testing:**
  - Unit tests cho data pipeline và query engine
  - Integration tests cho MCP tools
  - End-to-end tests cho agent workflows
- **Documentation:** Cập nhật README và technical docs sau mỗi milestone

- **Version control:** Git workflow với feature branches và pull requests
- **Monitoring:** Logs và error tracking trong production

2.2.3 Cam kết bảo mật (NDA)

Cam kết không tiết lộ hoặc chia sẻ các thông tin liên quan đến:

- Dữ liệu cá nhân của sinh viên (điểm số, thông tin DAA)
- Credentials và cookies sync từ browser extension
- Source code và implementation details (trừ khi được phép)
- Infrastructure và deployment configuration

2.3 Timeline and Milestones

Date	Milestone Description
01/09/2024	Khởi động dự án, setup repository, xác định roadmap
30/09/2024	**Giai đoạn 1 hoàn thành:** Knowledge base sẵn sàng, query engine hoạt động cơ bản
31/10/2024	**Giai đoạn 2 hoàn thành:** MCP Server expose 4 tools qua HTTP, test thành công với MCP Inspector
30/11/2024	**Giai đoạn 3 hoàn thành:** Agent service chạy qua gRPC, gọi được MCP tools, hoạt động với ReAct pattern
31/12/2024	**Giai đoạn 4 hoàn thành:** MVP end-to-end hoàn chỉnh, deploy production, CI/CD setup
05/01/2025	Hoàn thiện documentation, báo cáo đồ án, chuẩn bị demo

Bảng 2.1: Timeline và milestones chính của dự án

Chương 3 Resources, Costs, and Budget

3.1 Resource Management

Role	Name	Assignment Status	Start Date	End Date	% Effort per day
PM	Quách Gia Kiệt	Đã Giao	01/09/2024	01/01/2025	80%
Tech Lead	Quách Gia Kiệt	Đã Giao	01/09/2024	01/01/2025	100%
Backend Dev	Nguyễn Tuấn Kiệt	Đã Giao	01/09/2024	01/01/2025	90%
AI/ML Engineer	Quách Gia Kiệt	Đã Giao	01/09/2024	30/11/2024	100%
Frontend Dev	Nguyễn Tuấn Kiệt	Đã Giao	01/12/2024	31/12/2024	100%
DevOps	Nguyễn Tuấn Kiệt	Đã Giao	15/12/2024	01/01/2025	50%
QA/Tester	Cả nhóm	Đã Giao	20/12/2024	01/01/2025	40%
Documentation	Cả nhóm	Đã Giao	01/09/2024	01/01/2025	20%

Bảng 3.1: Phân bổ nguồn lực dự án

3.2 Estimated Costs and Budget

Task	Service Costs	Infrastructure	Budget
OpenAI API	150,000		150,000
LlamaParse API	0 (free tier)		0
Modal GPU (ViRanker)	0 (free tier)		0
VPS Hosting (4 tháng)		0 (free tier)	0
Domain Name		0 (free)	0
Development Tools		0 (free/open source)	0
Total			**150,000**

Bảng 3.2: Ước tính chi phí dự án (VNĐ)

Lưu ý: Đây là chi phí ước tính cho giai đoạn Đồ án 1 (4 tháng). Chi phí thực tế có thể thay đổi tùy theo mức độ sử dụng API và infrastructure.

Chương 4 Stakeholders

Name	Role	Power	Interest	Contact
Th.S Nguyễn Công Hoan	Project Sponsor/ Giảng viên hướng dẫn	Cao	Cao	hoannc@uit.edu.vn
Quách Gia Kiệt	Project Manager/ Tech Lead	Cao	Cao	23520819@gm.uit.edu.vn
Nguyễn Tuấn Kiệt	Product Manager/ Backend Dev	Cao	Cao	23520815@gm.uit.edu.vn
Sinh viên UIT	End Users	Trung Bình	Cao	Via web app
Khoa CNPM	Academic Department	Cao	Trung Bình	Via official channels

Bảng 4.1: Danh sách stakeholders

Chương 5 Risk Management

Risk	Likelihood	Impact	Mitigation
RAG accuracy thấp, agent trả lời sai	Cao	Cao	Tối ưu chunking strategy, sử dụng reranker (ViRanker), test thường xuyên trên bộ câu hỏi chuẩn
Agent hallucination, không follow system prompt	Trung Bình	Cao	Prompt engineering cẩn thận, thêm examples, constrain output format, test nhiều scenarios
API rate limits (OpenAI, LlamaParse, Modal)	Cao	Trung Bình	Implement caching, retry logic với exponential backoff, monitor usage, có backup plan
DAA website thay đổi cấu trúc HTML, scraping tools fail	Trung Bình	Cao	Modular scraping code, dễ update selectors, error handling tốt, fallback mechanisms
Performance issues khi scale (nhiều users)	Thấp	Trung Bình	Load testing trước khi golive, optimize database queries, implement caching, horizontal scaling nếu cần
Security vulnerabilities (credentials leak, injection attacks)	Thấp	Cao	Encrypt cookies, sanitize inputs, rate limiting, security audit, không log sensitive data
Infrastructure downtime (VPS, ChromaDB, databases)	Thấp	Cao	Monitoring và alerting, backup strategy, documentation để recover nhanh
Scope creep (thêm features ngoài MVP)	Cao	Trung Bình	Scope management nghiêm ngặt, ưu tiên MVP features, features khác defer sang Đồ án 2
Team member unavailable (ốm, bận)	Trung Bình	Trung Bình	Knowledge sharing, documentation tốt, cross-training, flexible task assignment

Risk	Likelihood	Impact	Mitigation
Integration issues giữa các components (agent, MCP, API Gateway)	Trung Bình	Cao	Integration testing sớm, clear API contracts, mock services cho testing

Bảng 5.1: Risk analysis và mitigation strategies

Chương 6 Additional Information

6.1 Technical Stack

6.1.1 Backend Technologies

- **API Gateway:** Go + Gin framework
- **AI Agent:** Python + LangGraph
- **MCP Server:** Python + FastMCP
- **Knowledge Builder:** Python CLI tool

6.1.2 AI & ML Technologies

- **LLMs:** OpenAI GPT-4, Google Gemini Flash
- **RAG Framework:** LlamaIndex
- **Agent Orchestration:** LangGraph
- **Embeddings:** OpenAI text-embedding-3-small
- **Reranking:** ViRanker (deployed on Modal GPU)
- **Document Parsing:** LlamaParse

6.1.3 Data Layer

- **MongoDB:** Chat sessions và messages
- **PostgreSQL:** Agent state (LangGraph checkpointer)
- **Redis:** Cache credentials và session data
- **ChromaDB:** Vector database cho embeddings

6.1.4 Frontend Technologies

- **Web App:** React + Vite + Tailwind CSS
- **Browser Extension:** Svelte (Manifest V3)

6.1.5 Infrastructure & Communication

- **Containerization:** Docker + Docker Compose
- **Protocols:** gRPC (API Gateway ↔ Agent), HTTP/REST (Frontend ↔ API, Agent ↔ MCP), MCP over HTTP
- **Deployment:** VPS (Ubuntu), CI/CD pipeline

Chi tiết đầy đủ về tech stack có thể tham khảo tại tài liệu *Tech Stack - UIT AI Assistant*.

6.2 Current Progress

Tính đến ngày 28/12/2024:

Giai đoạn	% hoàn thành	Chi tiết
1 - Knowledge Base	100%	Knowledge base và query engine hoạt động, độ chính xác chấp nhận được
2 - MCP Server	100%	MCP Server expose 4 tools qua HTTP
3 - AI Agent	80%	Agent chạy được qua gRPC, tuy nhiên đôi khi không follow system prompt, còn hallucination, trả lời chưa đúng câu hỏi phức tạp, chưa có test
4 - GUI & Deploy	100%	API Gateway, web app, browser extension hoàn thành. Deploy production lên VPS với domain public, CI/CD tự động

Bảng 6.1: Tình hình hiện tại của dự án

6.3 Next Steps

Công việc ưu tiên cho giai đoạn tiếp theo (Đồ án 2):

1. **Xây dựng bộ test** (QUAN TRỌNG NHẤT): Unit tests, integration tests, end-to-end tests
2. **Tối ưu Agent:** Cải thiện system prompt, giảm hallucination, tăng accuracy
3. **Cache optimization:** Cache tool output để tránh gọi dư thừa
4. **Chunking optimization:** Tối ưu chunking strategy cho chương trình đào tạo
5. **PostgreSQL Checkpointer:** Migrate từ InMemory sang PostgreSQL cho persistent chat history

6. **Dashboard enhancements:** Hiển thị trực quan điểm số, thời khóa biểu
7. **Additional tools:** Web search, notification scraping, FAQ database
8. **Third-party integrations:** Notion, Gmail, Google Calendar

---

Tài liệu này cung cấp một cái nhìn tổng quan ở cấp cao về dự án **UIT AI Assistant**, đảm bảo sự phù hợp với các mục tiêu học thuật và kỳ vọng của các bên liên quan.