Politecnico di Torino
International University

LAVAZZA
TORINO, ITALIA, 1895

AI PERSONAS

Enrico Chen - s337750
Van Thanh Nguyen - s336748
Xiaoning Ma - s337332

Project overview

Politecnico di Torino
International University

OUR GOAL

➔ Develop an AI platform where internal business teams can dynamically interact with Personas representing distinct market segments to identify strategies

KEY FEATURES

➔ **Automated extraction** of unstructured data into semi-structured formats
➔ **Data-grounded** AI personas derived from market data retrieved via a RAG system
➔ Explainable responses supported by **explicit citations and contextual evidence**

BENEFITS

➔ **Reduced time and resources spent on data preprocessing**
➔ **Improved trustworthiness and credibility of responses**
➔ Enable more focused and informed market strategies

VALUE PROPOSITION

Politecnico di Torino
International University

For **business units** struggling in **evaluating marketing performances, customer understanding, models and ideas testing**, our platform allows **interacting with data-driven AI Personas** representing the different **market segments**

Research Questions

Extraction

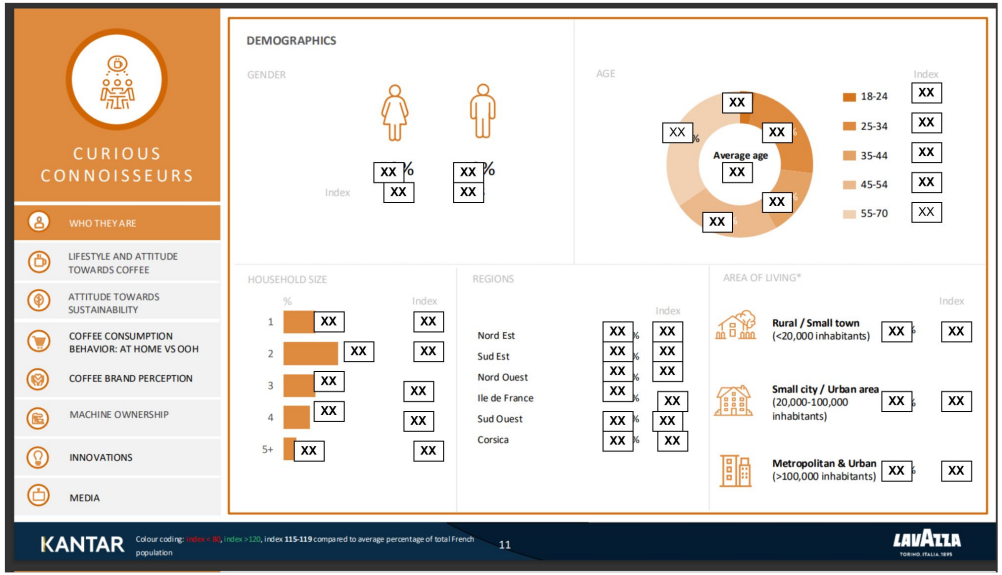
How can we automatically extract accurate data from **complex, visually rich** PDFs while preserving layout semantics?

Personality

Can **latent reasoning traits** (style profiles & value frames) be inferred from heterogeneous, noisy raw data to build consistent AI personas?

Grounding

How can **hallucinations** in language model outputs be reduced?



For privacy and compliance reasons, certain details are omitted from this presentation.

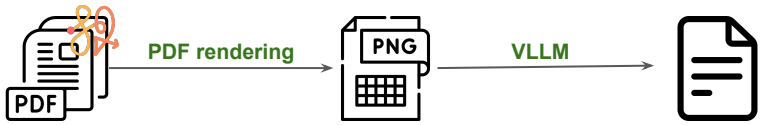
Research Question 1

Question:

How can we automatically extract accurate data from **complex, visually rich** PDFs while preserving layout semantics?

Our solution:

We treat PDFs as visual documents and extract semantic meaning using a VLLM



- Preserves layout and spatial relationships
- Fully automated, no manual rules
- Directly produces schema-consistent data ready for downstream use

Alternative Solutions:

- **Manual annotation**: high accuracy, but low scalability, high costs
- **OCR-based pipelines**: scalable, but limited layout information

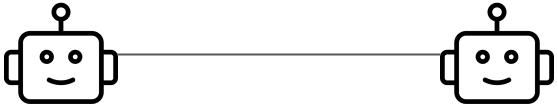
Research Question 2

Question:

Can we **infer latent reasoning traits** from heterogeneous, noisy raw data in order to build consistent AI personas?

Our solution:

We use a cooperative multi-agent system in which each agent specialises in a specific stage of reasoning and trait inference.



Agent 1

Normalize raw data into semi-structured representations

Agent 2

Infer latent persona reasoning traits by identifying patterns in language use, metrics, and expressed priorities

- Scalable and repeatable across datasets
- Reduces subjective human bias
- Produces consistent persona traits at scale

Alternative Solutions:

- **Human review and inference**: high accuracy but slow, subjective, and not scalable
- **Single-model end-to-end inference**: automated, but sensitive to noise and lacking trait consistency

Research Question 3

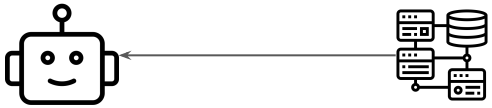
Politecnico di Torino
International University

Question:

How can we reduce **hallucinations** in language model outputs?

Our solution:

We integrate a RAG system that retrieves relevant evidence before generation and constraints model outputs to the retrieved context.



- Improves factual accuracy
- Reduces hallucinations caused by unsupported generation
- Explainable responses
- Low computational costs

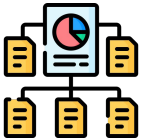
Alternative Solutions:

- **Prompt-based**: lightweight, but unreliable and sensitive to prompt design
- **Pure fine-tuning**: improves fluency, but does not guarantee factual grounding, computationally expensive

Dataset

Source & Scale

- Kantar France 2023
- 4,001 Respondents
- 9 consumer segments
- Rich qualitative + quantitative data



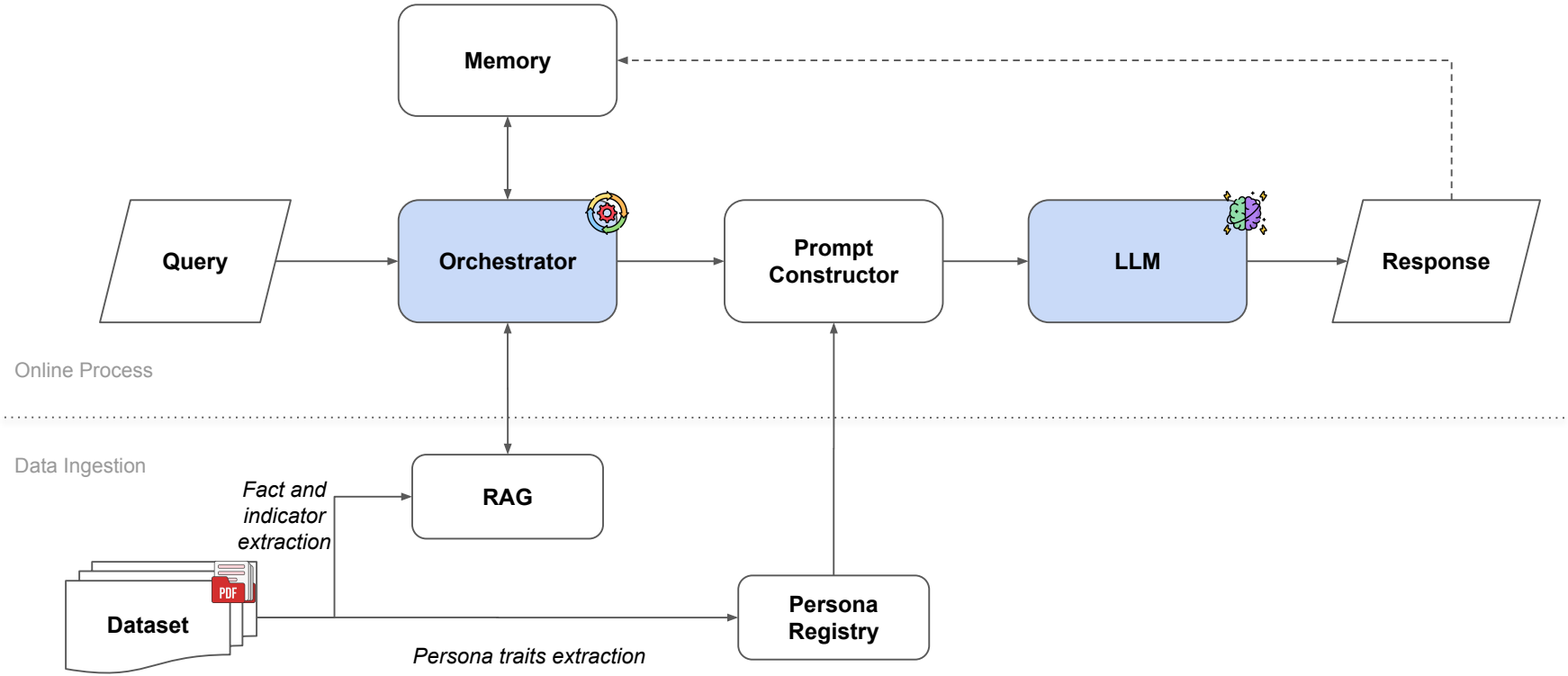
Data Structure

- Multi-modal composition
- Inconsistent layouts across segments
- Mixed granularity
- No unified or machine-readable format

Key Dimensions

- Demographics
- Psychographics
- Coffee attitudes
- Consumption behaviours
- Brand perception & sustainability attitudes

Functional Diagram



Fact data: all data (including segments characteristics, market data)
Persona traits data: characteristics of the customer segments (like tone, style, values, preferences)

Method - Persona Extraction

Each page of the dataset

VLLM Extraction

JSON

JSON indicators

VLLM Reasoning

Persona Traits

Prompt tuning

Fine-tuned AI Personas

Chunker and Embedding Model

Database

Method - Fact Extraction

Each page of the dataset

VLLM

Markdown output

Chunker

Chunked output

Embedding Model

Database

Method - Application

Back-end (REST APIs)

Handles system operations and provides results to the frontend

Streamlit

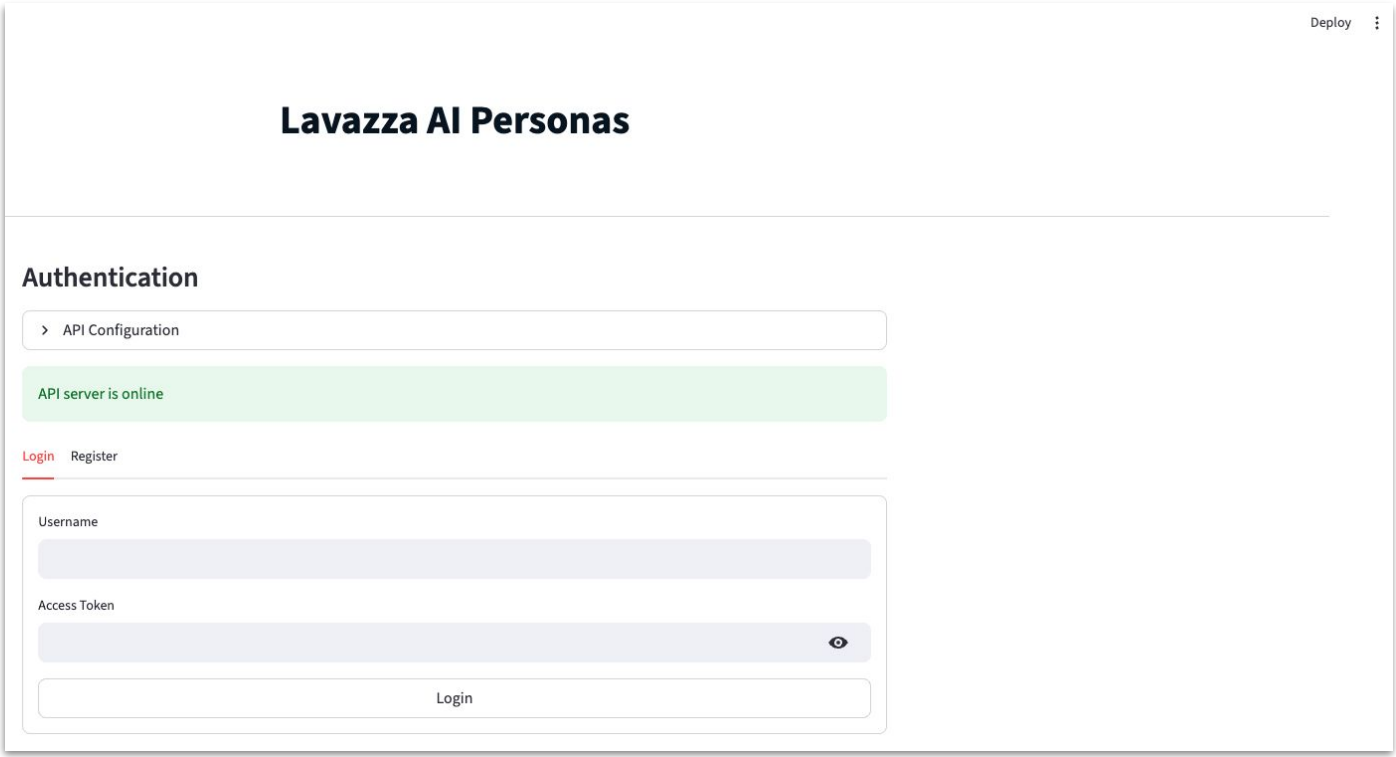
Front-end

User interface for interacting with the backend, built with Streamlit

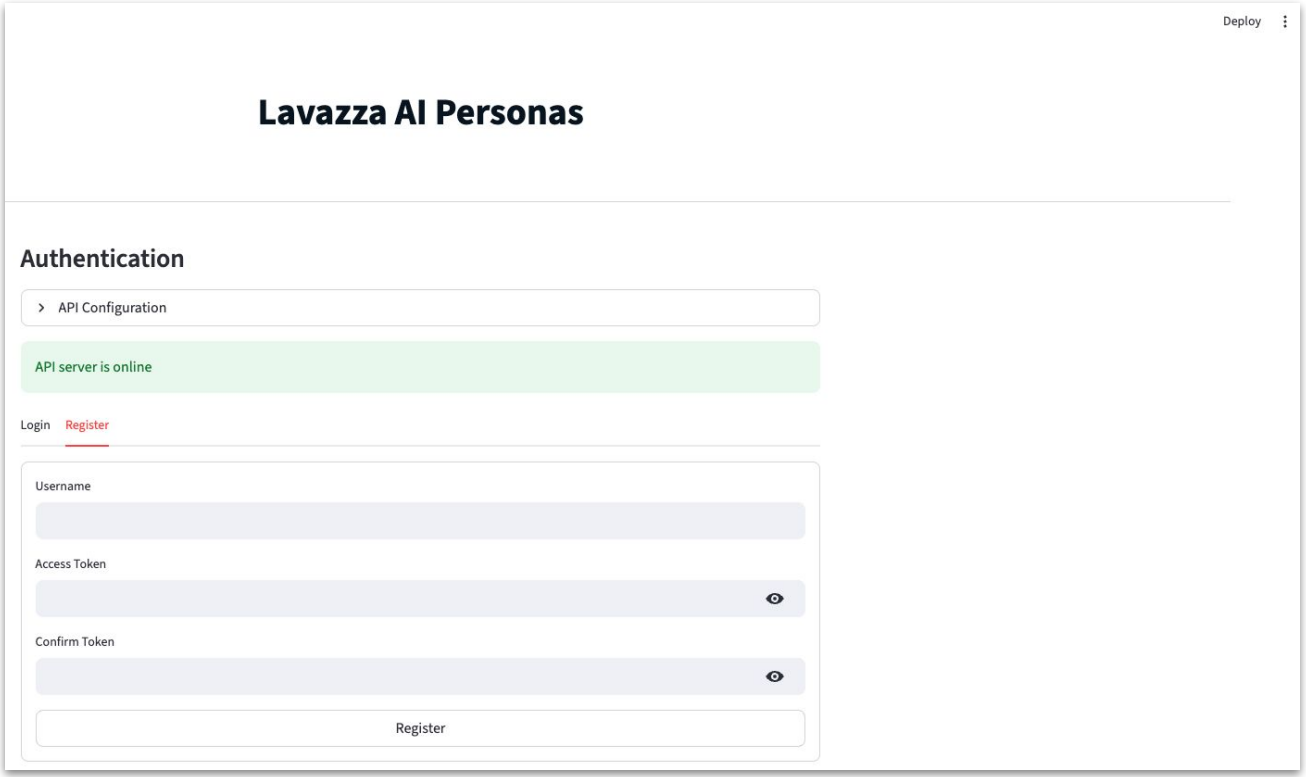
Features implemented

1. Authentication
2. Customer segment selection
3. Past chat selection
4. Context customization
5. Persona name customization
6. Persona information

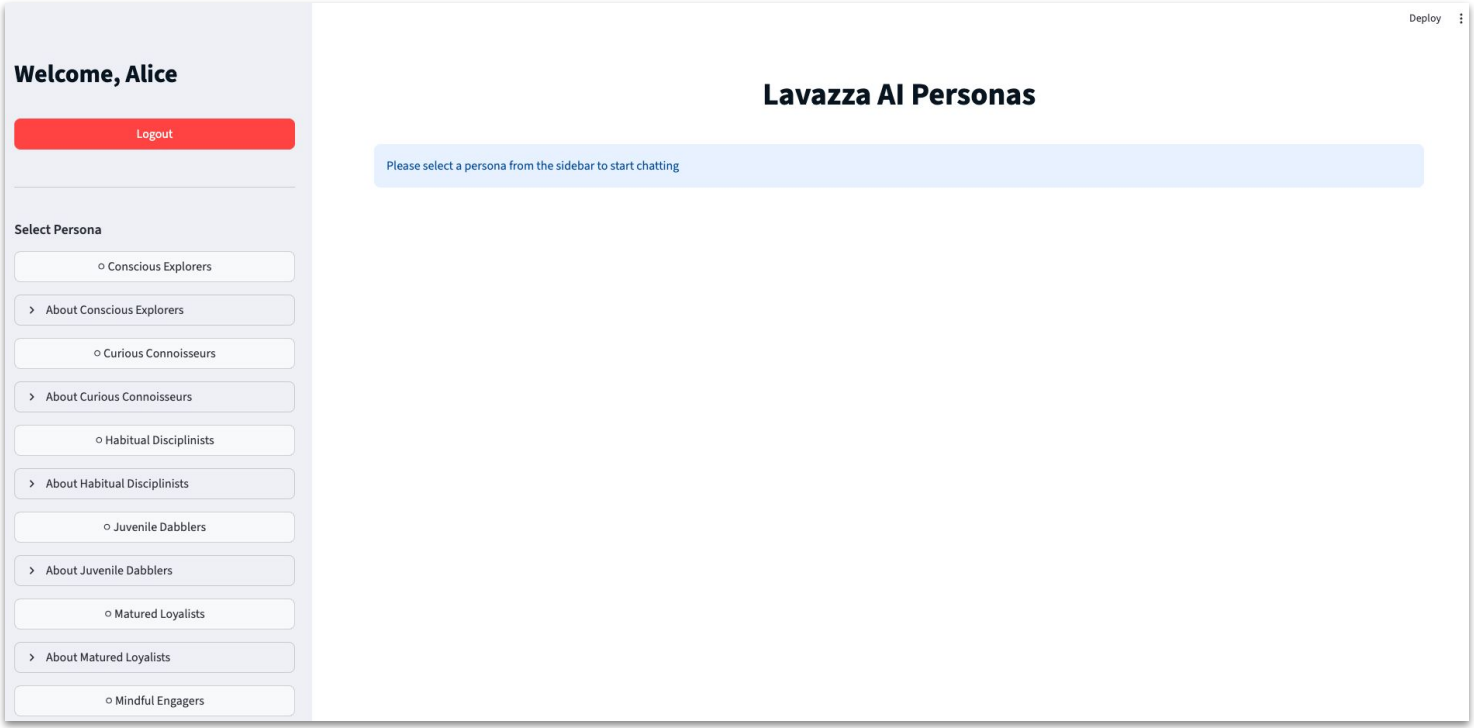
UI - Login


Deploy
Lavazza AI Personas
Authentication
API Configuration
API server is online
Login
Register
Username
Access Token
Login

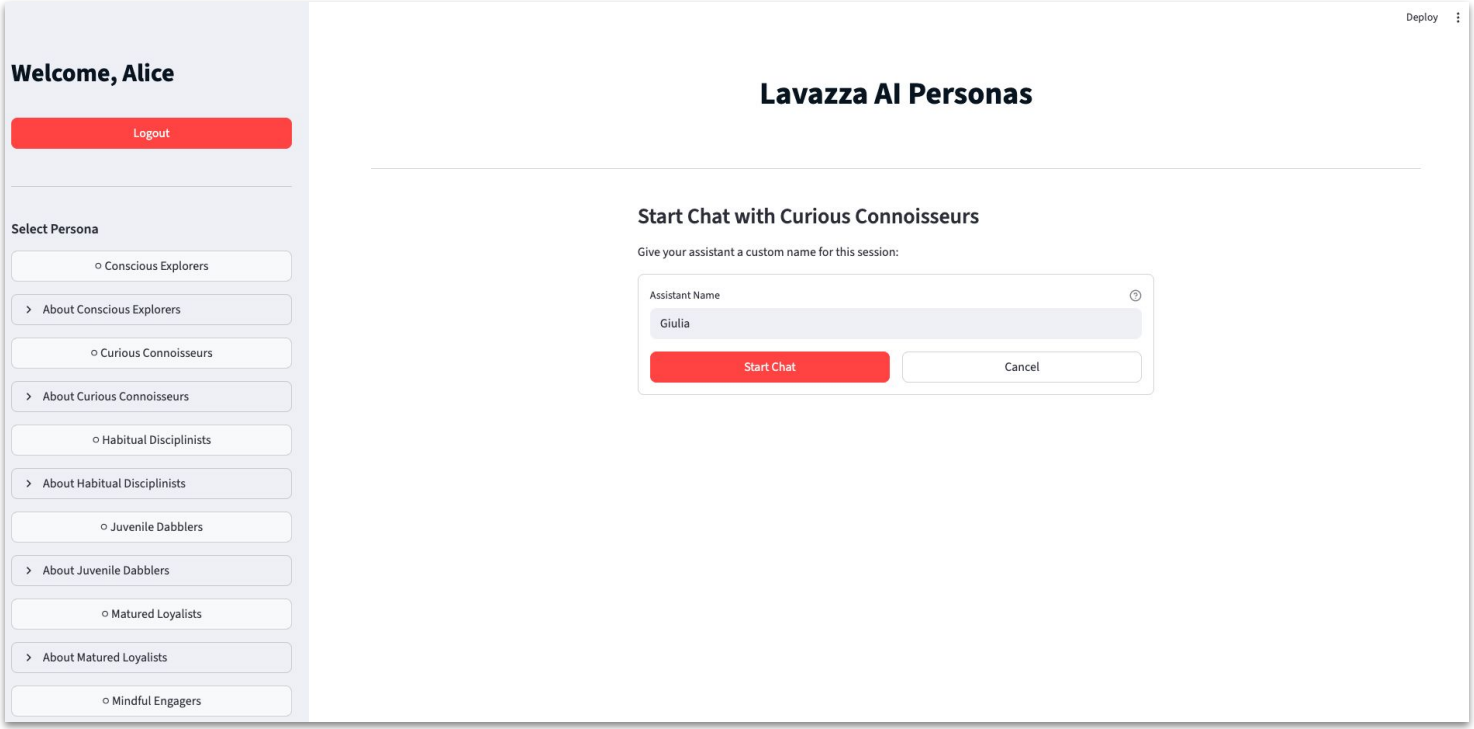
UI - Registration



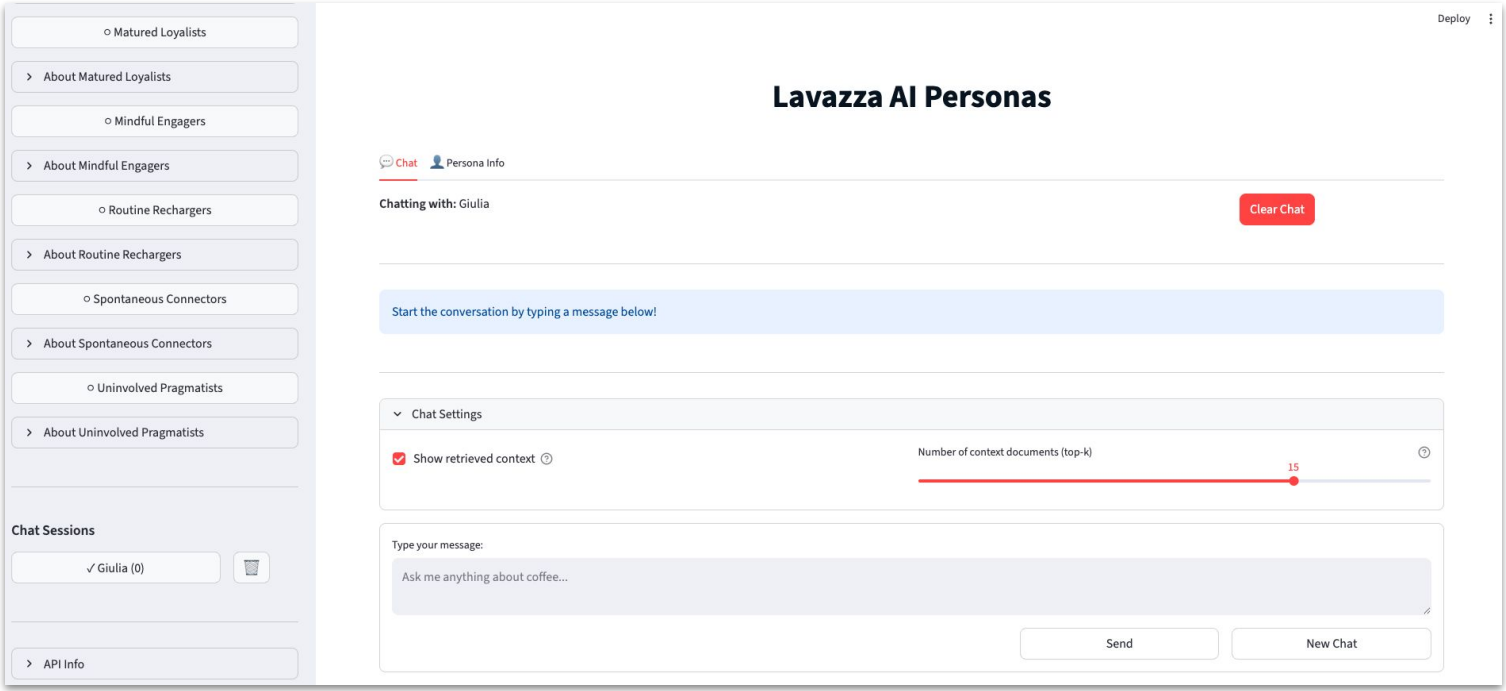
UI - Homepage


Deploy
Welcome, Alice
Logout
Select Persona
Conscious Explorers
About Conscious Explorers
Curious Connoisseurs
About Curious Connoisseurs
Habitual Disciplinists
About Habitual Disciplinists
Juvenile Dabblers
About Juvenile Dabblers
Matured Loyalists
About Matured Loyalists
Mindful Engagers
Lavazza AI Personas
Please select a persona from the sidebar to start chatting

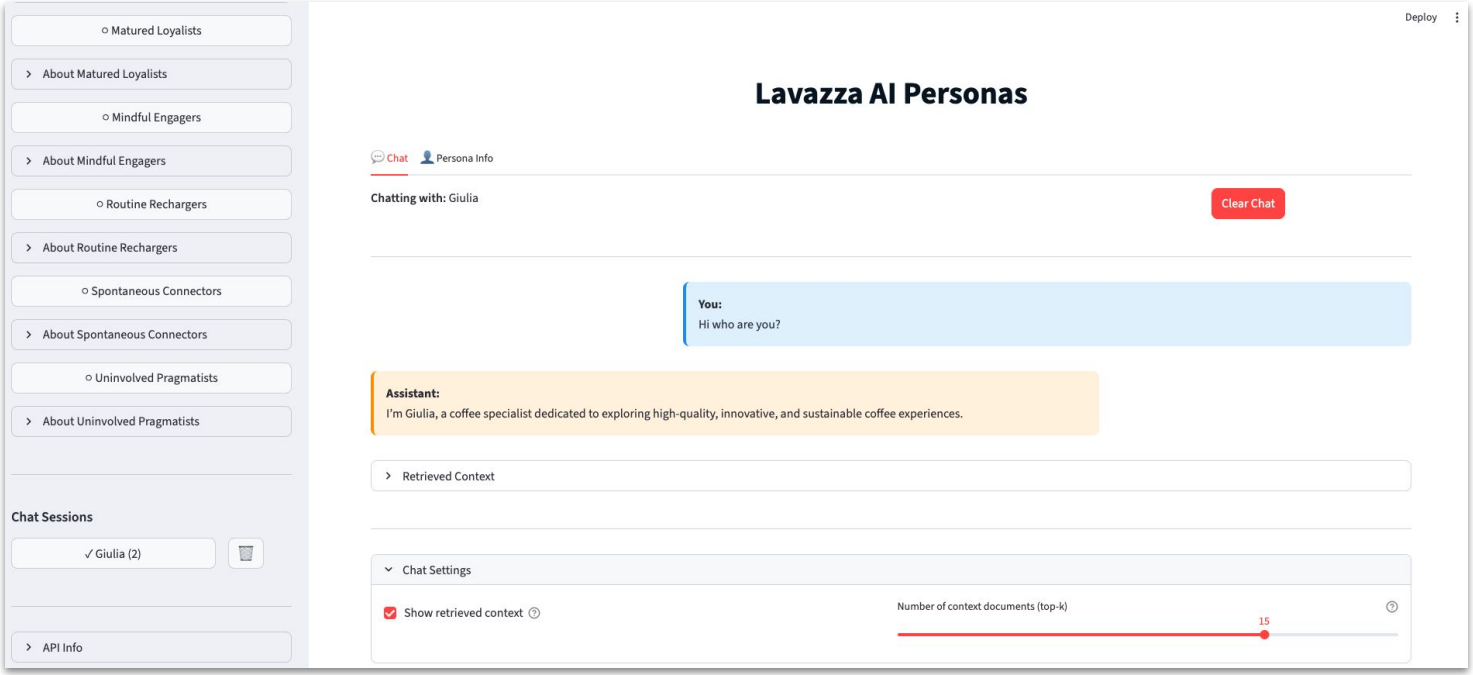
UI - Persona name customization



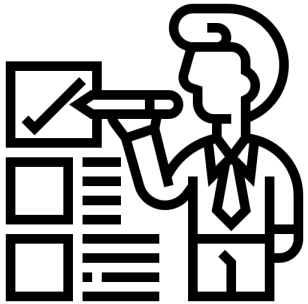
UI - Chat Management



UI - Chat Interaction



Evaluation



Persona Extraction

Is persona information correctly extracted?

Fact Extraction

Is extracted fact data correct?

Retrieval Relevance

Is the retrieved information relevant?

Authenticity Evaluation

Do personas reflect real consumer segments?

Evaluation - Persona Extraction

Approach

Compare persona extraction outputs with manually annotated ground truth.

Metrics

➔	**Persona Detection Rate**	Are all target personas identified?
➔	**Metrics Recall**	Is the required information complete?
➔	**Metrics Precision**	Is the extracted information accurate?

*Persona Detection Rate = (Correctly Detected Personas) / (Total Ground Truth Personas)
*Metrics Recall = (Correctly Metrics) / (Total Ground Truth Metrics)
*Metrics Precision = (Correct Metrics) / (All Extracted Metrics)

Evaluation - Persona Extraction

Politecnico di Torino International University

Configuration

- **Models**:
  - **Extract indicators:** mistralai/mistral-medium-3-instruct
  - **Persona traits reasoning**: mistralai/mistral-medium-3-instruct
- **Test dataset:**
  - **Scope**: 23 pages focused on the Curious Connoisseurs segment
  - **Source**: Customer Segmentation Analysis PDF
  - **Ground Truth**: 1,051 metrics manually extracted and validated from the PDF

Result

	Score	Interpretation
Persona Detection Rate	100%	All personas correctly identified (23/23)
Metrics Recall	95.30%	Very high coverage of ground-truth metrics (1002/1051)
Metrics Precision	96.80%	Minimal noise in extracted metrics (1002/1035)

Evaluation - Fact Extraction

Approach

- Manually verify numeric and factual data against source PDFs.
- Run extraction and check exact matches.
- A result is correct only if ≥80% textual overlap with the extracted markdown.

Metrics

➔ **Exact Match Accuracy** what fraction of expected facts were extracted correctly?

*Exact Match Accuracy = (Exactly Matching Values) / (Total Ground Truth Values)

Evaluation - Fact Extraction

Configuration

- **Models**:
  - **PDF to Markdown Conversion:** mistralai/mistral-large-3-675b-instruct-2512
- **Test dataset:**
  - **Scope**: 23 pages focused on the Curious Connoisseurs segment
  - **Source**: Customer Segmentation Analysis PDF
  - **Ground Truth**: 467 validated metric and statement snippets manually extracted from the PDF

Result

	Score	Interpretation
Exact Match Accuracy	97%	High extraction accuracy with minimal deviation from ground truth

Evaluation - Retrieval Relevance

Approach

- Create persona-specific test queries.
- Retrieve top-K documents (K = 3, 5, 10, 20)..
- Manually label relevance for retrieval documents.

Metrics

➔ **Precision@K** How many of the top K results are actually correct.

➔ **Recall@K** How many of the total correct results were successfully found within the top K.

*Precision@K = (Relevant docs in top-K) / K
*Recall@K = (Relevant docs in top-K) / (Total relevant docs)

Evaluation - Retrieval Relevance

Politecnico di Torino International University

Configuration

- ➔ **Models**:
  - ◆ **Embedding:** sentence-transformers/all-mpnet-base-v2
- ➔ **RAG setup**:
  - ◆ **Chunk Size**: 1,200 characters
  - ◆ **Chunk Overlap**: 50 characters
  - ◆ **Input**: 222-page split to 712 total text chunks
- ➔ **Test dataset:**
  - ◆ **Source**: Customer Segmentation Analysis PDF
  - ◆ **Ground Truth**: 31 evaluation questions focused on the Curious Connoisseurs segment

	Score	Interpretation
Precision@3	58.10%	Top-3 results are moderately relevant
Precision@5	56.80%	Relevant remains stable in top-5
Precision@10	54.20%	Slight precision drop
Precision@20	55.50%	Precision stabilizes at broader context window
Recall@3	16.70%	Limited coverage with very short context
Recall@5	24.30%	Partial retrieval of relevant context
Recall@10	45.10%	Balanced trade-off between precision and recall
Recall@20	**94.10%**	**Near-complete retrieval of relevant**

Result

Evaluation - Persona Authenticity

Approach

- Generate persona-based responses to predefined questions.
- Have domain experts score them on authenticity, style alignment, and factual grounding, each on a 1–5 scale.

Metrics

➔	**Expert Authenticity Score**	Average authenticity ratings.
➔	**Style Alignment Score**	Average all style alignment ratings.
➔	**Factual Grounding Score**	Average all factual grounding ratings.

Evaluation - Authenticity Results

Politecnico di Torino International University

Configuration

- **Models**:
  - **Filter context model:** mistralai/mistral-medium-3-instruct
  - **Generation model:** mistralai/mistral-small-24b-instruct
- **RAG Setup:**
  - **Chunk Size**: 1,200 characters (~300–350 tokens)
  - **Chunk Overlap**: 50 characters (to preserve context continuity)
  - **Input**: 222-page to 712 total text chunks
- **Test dataset:**
  - **Source**: Customer Segmentation Analysis PDF
  - **Ground Truth**: 31 evaluation questions focused on the Curious Connoisseurs segment

Result

	Score	Interpretation
Expert Authenticity Score	3.90/5	Persona behavior is largely authentic.
Style Alignment Score	3.74/5	Style is mostly consistent, with minor persona drift.
Factual Grounding Score	3.67/5	Responses are generally grounded.

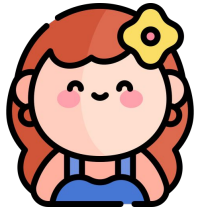
Manage

Enrico

- Semantic extraction of fact data
- Fact data indexing and organization
- Frontend application development
- Related presentation slides

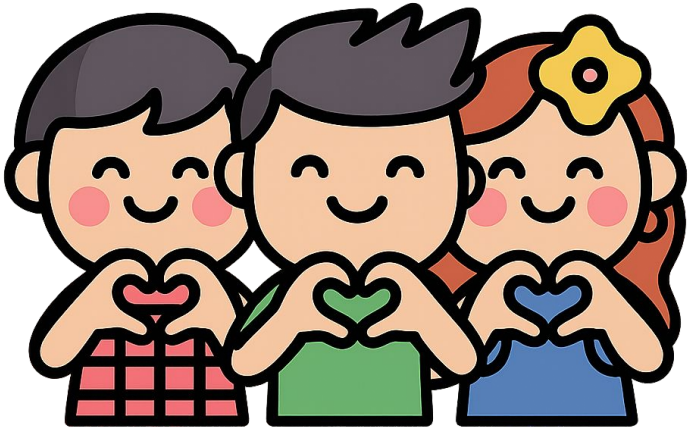
Thanh

- Persona semantic information extraction
- Backend system development
- Evaluation (persona & fact data extraction)
- Related presentation slides



Xiaoning

- AI persona common trait extraction
- Evaluation dataset construction
- Evaluation (retrieval & authenticity analysis)
- Related presentation slides



THANK YOU

Enrico Chen - s337750
Van Thanh Nguyen - s336748
Xiaoning Ma - s337332

LAVAZZA

TORINO, ITALIA, 1895

Politecnico di Torino

International University

Appendix

Manage - Gantt

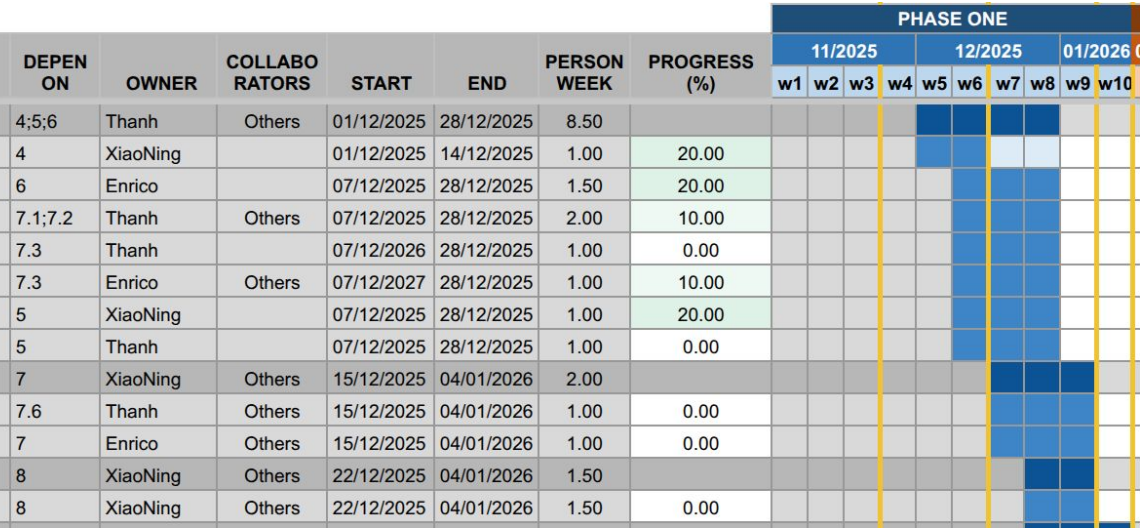
WB No	TASK TITLE	DEPEN ON	OWNER	COLLABO RATORS	START	END	PERSON WEEK	PROGRESS (%)
1	Project Conception and Initiation		Thanh	Others	03/11/2025	14/11/2025	3.00	
1.1	Project structure+ work flow		Thanh	Others	03/11/2025	03/11/2025	0.50	100.00
1.2	Kick off with Lavazza tutor		XiaoNing	Others	03/11/2025	07/11/2025	0.50	100.00
1.3	Research		Enrico	Others	03/11/2025	14/11/2025	2.00	100.00
2	Design	1	Enrico	Others	03/11/2025	16/11/2025	7.50	
2.1	Objective and Goal Definition	1.2	Thanh	Others	03/11/2025	14/11/2025	0.50	100.00
2.2	Stakeholder Map	1.2	Enrico	Others	03/11/2025	14/11/2025	1.00	100.00
2.3	User Personas Definition	1.2	Enrico	Others	03/11/2025	14/11/2025	0.50	100.00
2.4	User Journey Definition	1.2	XiaoNing	Others	03/11/2025	14/11/2025	1.00	100.00
2.5	User Requirements Definition	1.2	XiaoNing	Others	03/11/2025	14/11/2025	0.50	100.00
2.6	Usecase diagram	1.2;1.3	XiaoNing	Others	03/11/2025	16/11/2025	1.00	100.00
2.7	Func and Non-Func Requirements Definition	1.2;1.3	Enrico	Others	03/11/2025	16/11/2025	1.00	100.00
2.8	System Architecture and Func Diagram	1.2;1.3	Thanh	Others	03/11/2025	16/11/2025	1.00	100.00
2.9	Risk Analysis	1.2;1.3	Thanh	Others	08/11/2025	16/11/2025	1.00	100.00
3	Management	2	XiaoNing	Others	08/11/2025	16/11/2025	1.00	
3.1	Tasks Breakdown and Gantt Diagram	2	XiaoNing	Others	08/11/2025	16/11/2025	1.00	100.00

Manage - Gantt

WB No	TASK TITLE	DEPENDON	OWNER	COLLABORATORS	START	END	PERSON WEEK	PROGRESS (%)
4	Data Foundation	1.2	Thanh	Others	17/11/2025	23/11/2025	6.00	
4.1	Data acquisition & ingest	1.2	Thanh	Others	17/11/2025	23/11/2025	3.00	90.00
4.1.1	Collect data from Lavazza	1.2	Enrico	Others	17/11/2025	21/11/2025	1.00	100.00
4.1.2	Collect data from external source		Thanh	Others	17/11/2025	21/11/2025	1.00	100.00
4.1.3	Understand dataset	4.1.1;4.1.2	XiaoNing	Others	21/11/2025	23/11/2025	1.00	100.00
4.2	Finalize PersonaProfile schema	4.1	XiaoNing	Others	20/11/2025	28/11/2025	1.00	100.00
4.3	Data processing pipelines	4.2	Thanh	Others	20/11/2025	30/11/2026	2.00	100.00
4.3.1	Handle fact data pipeline	4.2	Enrico	Thanh	20/11/2025	30/11/2025	1.00	100.00
4.3.2	Handle persona data pipeline	4.2	XiaoNing	Thanh	20/11/2025	30/11/2025	1.00	100.00
5	Prompt Tuning AI Persona	4.3.2	XiaoNing	Thanh	24/11/2025	27/12/2025	4.00	
5.1	Semanctic Extraction/Personas Structuring from Customer Segmentation Data	4.3.2	XiaoNing	Thanh	24/11/2025	12/12/2025	2.00	100.00
5.1.1	Extract common traits/rules for each personas	4.3.2	XiaoNing	Thanh	24/11/2025	12/12/2025	1.00	100.00
5.1.2	Create personas fine-tuning dataset	4.3.2	XiaoNing	Thanh	30/11/2025	12/12/2025	1.00	100.00
5.2	Implement Training pipeline	5.1	Thanh	XiaoNing	01/12/2025	28/12/2025	1.50	30.00
5.3	Implement Inference & serving	5.2	Thanh	XiaoNing	13/12/2025	28/12/2025	0.50	0.00
6	Fact Data Ingestion	4.3.1	Enrico	Others	24/11/2025	21/12/2025	2.50	
6.1	Semanctic Extraction from Fact Data	4.3.1	Enrico	Thanh	24/11/2025	21/12/2025	1.00	50.00
6.2	Indexing fact data	6.1	Enrico	Thanh	24/11/2025	21/12/2025	0.50	30.00
6.3	Implement retrieval logic with RAG	6.2	Enrico	Thanh	01/12/2025	21/12/2025	1.00	30.00

Manage - Gantt

WB No	TASK TITLE	DEPEN ON	OWNER	COLLABO RATORS	START	END	PERSON WEEK	PROGRESS (%)
7	Core Layer	4;5;6	Thanh	Others	01/12/2025	28/12/2025	8.50	
7.1	Input handling	4	XiaoNing		01/12/2025	14/12/2025	1.00	20.00
7.2	Retrieval-Augmented Generation	6	Enrico		07/12/2025	28/12/2025	1.50	20.00
7.3	Implement Orchestrator logic	7.1;7.2	Thanh	Others	07/12/2025	28/12/2025	2.00	10.00
7.4	Implement Explaination module	7.3	Thanh		07/12/2026	28/12/2025	1.00	0.00
7.5	Prompt construction	7.3	Enrico	Others	07/12/2027	28/12/2025	1.00	10.00
7.6	Persona registry	5	XiaoNing		07/12/2025	28/12/2025	1.00	20.00
7.7	AI Persona Router	5	Thanh		07/12/2025	28/12/2025	1.00	0.00
8	Application Layer	7	XiaoNing	Others	15/12/2025	04/01/2026	2.00	
8.1	Persona configuration	7.6	Thanh	Others	15/12/2025	04/01/2026	1.00	0.00
8.2	Q&A service	7	Enrico	Others	15/12/2025	04/01/2026	1.00	0.00
9	UI	8	XiaoNing	Others	22/12/2025	04/01/2026	1.50	
9.1	FE	8	XiaoNing	Others	22/12/2025	04/01/2026	1.50	0.00

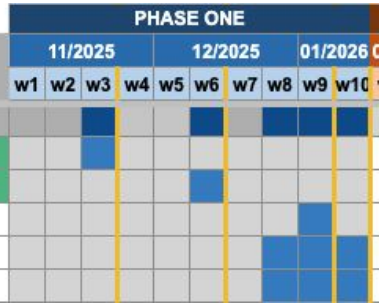


Manage - Gantt

WB No	TASK TITLE	DEPEN ON	OWNER	COLLABO RATORS	START	END	PERSON WEEK	PROGRESS (%)	PHASE ONE									
									11/2025				12/2025				01/2026	
									w1	w2	w3	w4	w5	w6	w7	w8	w9	w10
10	Monitoring and Evaluation	7	Enrico	Others	22/12/2025	11/01/2026	2.00											
10.1	Evaluation	7	Enrico	Others	22/12/2025	11/01/2026	2.00	0.00										
11	Deployment	7;8;9	Thanh	Others	22/12/2025	11/01/2026	2.00											
11.1	Packaging	7;8;9	Thanh	Others	22/12/2025	11/01/2026	1.00	0.00										
11.2	Deploy entire system	11.1	Thanh	Others	22/12/2025	11/01/2026	1.00	0.00										
12	Testing	7;8;9	XiaoNing	Others	22/12/2025	11/01/2026	2.00											
12.1	Test	7;8;9	XiaoNing	Others	22/12/2025	11/01/2026	1.00	0.00										
12.2	Fix Bug	12.2	Thanh	Others	22/12/2025	11/01/2026	1.00	0.00										
13	Demo	12	Enrico	Others	29/12/2025	11/01/2026	1.00											
13.1	Run full flow & get feedback	12	Enrico	Others	22/12/2025	11/01/2026	1.00	0.00										

Manage - Gantt

WB No	TASK TITLE	DEPEN ON	OWNER	COLLABO RATORS	START	END	PERSON WEEK	PROGRESS (%)
14	Communication		Enrico	Others	14/11/2025	19/11/2025	5.00	
14.1	First Checkpoint Presentation		Enrico	Others	14/11/2025	19/11/2025	1.00	100.00
14.2	Second Checkpoint Presentation		XiaoNing	Others	02/12/2025	09/12/2025	1.00	100.00
14.3	Third Checkpoint Presentation		Thanh	Others	30/12/2025	06/01/2026	1.00	0.00
14.4	Final Presentation		Enrico	Others	23/12/2025	12/01/2026	1.00	0.00
14.5	Final Report		XiaoNing	Others	23/12/2025	13/01/2026	1.00	0.00



Design

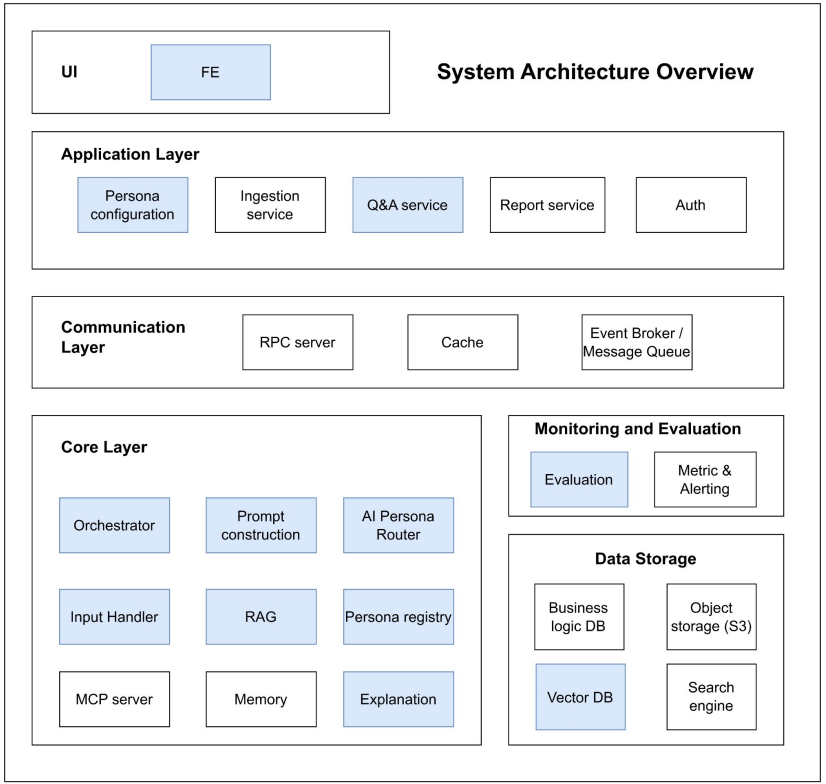
1. User Interface (UI)
The user interface serves as the system's entry point, built as a **Frontend (FE)** application. It enables users to interact seamlessly with the platform, submit queries, upload data, and view results or reports.

2. Application Layer
This layer contains the core application logic and manages all user-driven workflows.
Key components include:

- **Persona Configuration**: Enables users to select or customize AI personas dynamically.
- **Ingestion Service**: Handles ingestion of raw data such as PDFs or images and stores them in S3.
- **Report Service**: Generates structured, formatted reports from processed and analyzed data.
- **Q&A Service**: Manages interactive question-and-answer exchanges with the AI.
- **Auth Service**: Provides authentication and authorization for users, ensuring secure access and operations.

3. Communication Layer
This layer facilitates efficient communication and coordination among microservices.

- **RPC Server**: Enables direct service-to-service communication via Remote Procedure Calls.
- **Cache**: A high-speed memory layer that stores frequently accessed data to optimize performance.
- **Event Broker / Message Queue** (RabbitMQ or Kafka): Handles asynchronous communication and event-driven processing across services, ensuring reliability, scalability, and robust monitoring.

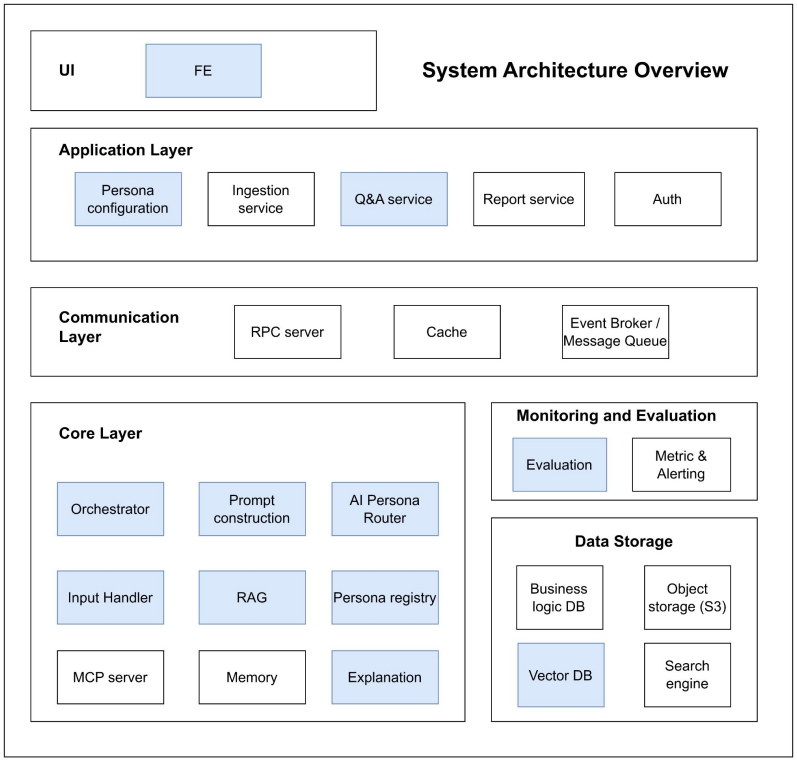


Design

4. Core Layer

The intelligence engine of the system—handles AI persona logic, LLM orchestration, and data-driven grounding.

- **Orchestrator**: The central coordinator of the Core Layer. When a request arrives, the Orchestrator manages the entire generation process, directing which services to call.
- **Input Handler**: Preprocesses and normalizes user inputs, including text extraction from PDFs and preparation of image data for AI analysis.
- **Prompt Construction**: Dynamically builds structured prompts by combining user input, persona rules, and retrieved data.
- **AI Personas**: Represents the fine-tuned Large Language Models (LLMs) tailored to embody distinct customer segment personalities.
- **RAG (Retrieval-Augmented Generation):** Provides factual grounding by retrieving relevant information from the Vector DB, ensuring responses remain accurate.
- **Persona Registry:** Stores the static attributes and behavioral definitions of each persona, guiding prompt construction and response tone.
- **Explanation:** This module allows for an in-depth explanation of the thought process behind the reasoning model and the data used in the thinking process.
- **MCP Server (Model Context Protocol Server)**: Enriches LLM interactions with real-time contextual or external domain data.
- **Memory**: It stores the recent history of the user's chat, allowing the persona to remember what was said earlier in the conversation and provide context-aware answers.



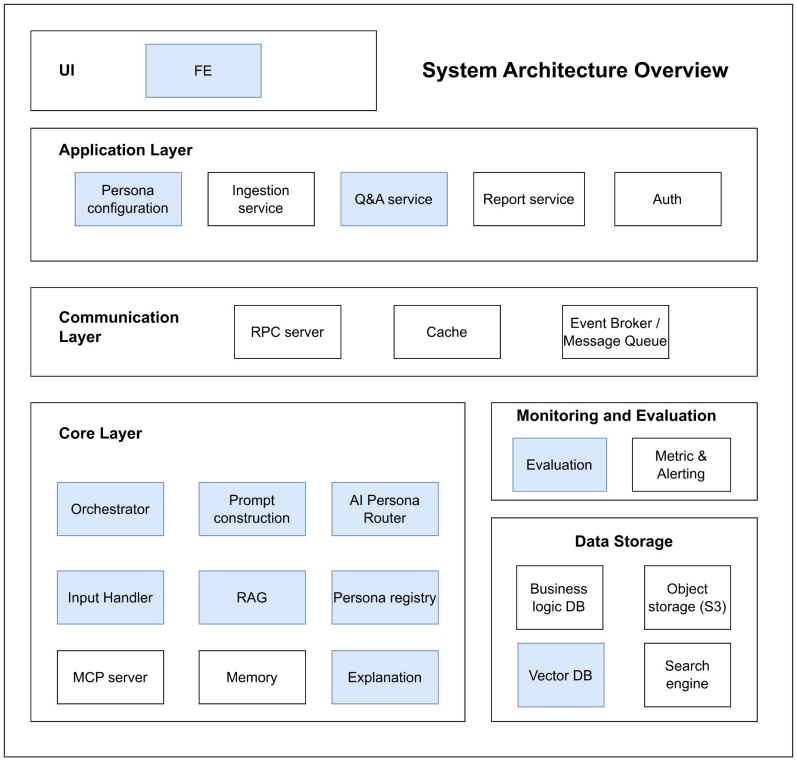
Design

5. Monitoring and Evaluation
A centralized observability layer that tracks performance, quality, and reliability across all services.

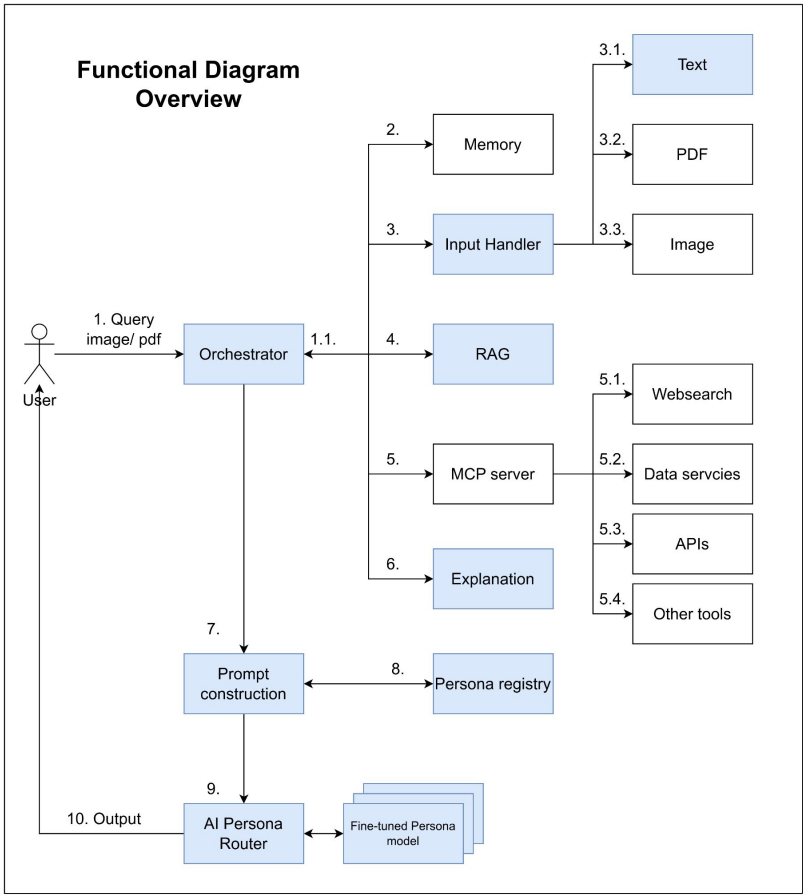
- **Evaluation Tools**: Measure the accuracy and quality of AI responses and data processing outcomes.
- **Metrics & Alerting**: Monitor key indicators such as latency, error rates, resource utilization, and token usage, triggering alerts for anomalies or system degradation.

6. Data Storage Layer
The persistence foundation of the system, designed for scalability, durability, and speed.

- **Business Logic Database**: Stores structured data such as user profiles, authentication records, saved reports, and persona definitions.
- **Object Storage (S3)**: Manages large, unstructured data files (e.g., raw PDFs, images, and uploaded datasets).
- **Vector Database**: Stores embeddings for persona-related documents, historical interactions, and reference materials — powering RAG retrieval and factual grounding.



Design



1. User Query Submission: User sends a query with optional attached files (image, PDF, etc.) to the Orchestrator.
1.1 Orchestrator Analysis: The Orchestrator analyzes the query and attachments to decide which services should be used.
2. Memory Integration : Extract useful information from chat history.
3. Input Preprocessing: Inputs are preprocessed before passing to the model.
3.1 Text Input: Normalize text to make it easier to handle in later steps.
3.2 PDF Input: Parse, process, and extract meaningful information from PDF files.
3.3 Image Input: Process images and extract valuable information.
4. Context Retrieval (RAG System): Use the query and relevant input information to retrieve context (e.g., market data) via a RAG system.
5. Tool Selection & MCP Server Requests

- Decide which tools should be used to enrich the context.
- Send requests to the MCP server to gather corresponding context.

5.1 Web Search: Extract updated information from the internet (trends, real-time data, missing internal data, etc.).
5.2 Database Query: Retrieve useful data from internal or external databases.
5.3 External APIs: Call APIs to obtain additional information.
5.4 Other Tools: Use calculators, simulators, weather data extractors, or other utilities to enrich context.
6. Explanation: The explanation module will explain in detail the thought process of the reasoning model and the data used for the thinking process.
7. Prompt Construction: The Orchestrator aggregates useful context and passes it to Prompt Construction.
8. Persona Selection : Apply the selected Persona profile, including: Demographics, Behavior Data, Transactional Data, …
9. Persona Model Routing

- Route to a fine-tuned Persona model.
- Pass the enriched prompt and context.

10. Model Response: Generate a response with: Specific personality,Tone, Linguistic style of the Persona

AI Personas Extraction & Fine-Tuning

Indicators (VLLM extraction output)

Example indicator shape (JSON):

```
{
  Indicator: {
    sources: {
      "url: {"https://www.pclavazza.com/nead/20/Indicator"
    },

    statements: {
      'statement: individual insights within an indicator, sostons and influences'},
      'metrics' {'marks whethr a svisually emphasiz: 'index', '%', 'count', 'rank'},
      'influene': flags whether a statement shapes tone or stance, 'smex': 'sources'}
    }
  }
}
```

AI Personas Extraction & Fine-Tuning

Politecnico di Torino
International University

Traits (reasoning output)

Persona Blueprint (Traits)

style_profile

how they speak: tone, formality, directness, emotional flavour, criticality, verbosity, preferred structures, example phrases.

value_frame

what they prioritize: priority_rank (sustainability, price, etc.), novelty seeking, brand loyalty, health concern, description.

reasoning_policies

purchase_advice, product_evaluation, information_processing, content_filters (biases, rules, praise/criticism triggers, trust, disclaimers).

Example JSON Structure

```
// How the model should "speak"
style_profile: (
  tone_adjectives: string[],      // ["Curious", "confident", "quality-focused", "pragmatic", ...],
  formality_level: "low" | "medium" | "high",
  directness: "very_direct" | "balanced" | "hedged",
  emotional_flavour: "neutral" | "enthasiastic" | "cool_detached" | "warm_reflective",
  criticality_level: "high" | "medium" | "low",
  verbosity_preference: "concise" | "detailed" | "varies_by_question",
  preferred_structures: string[],  // ["bullet_point", "clear_rsadeoffs", "pros_cons", "step_by_step"]
  typical_register_examples: string[]// short example phrases in target style
},

// What they care about – used to biss recommendations / reasoning
value_frame: (
   priority_rank: string[],       // e.g. ["quality", "convenience", "sustainability", "price"],
   sustainability_orientation: "high" | "medium" | "low",
   price_sensitivity: "high" | "medium" | "low",
   price_sensitivity: "high" | "medium" | "low",
   novelty_seeking: "high" | "medium" | "low"
),
}
```

Design - Risks Analysis

Technical Risks

- Hallucinations and inaccurate responses: mitigate with RAG system
- Insufficient critical thinking: mitigate with RAG and prompt engineering
- Opacity: mitigate with RAG
- Inconsistent or generic personality: mitigate by fine-tuning (in case of limited resource use PEFT, smaller models, RAG with few-shot prompting)
- Performance evaluation difficulty

Design - Risks Analysis

Governance and Security Risks

- Privacy and compliance with AI Act and GDPR
- Proprietary data protection
- System integration difficulty with existing systems and infrastructure

Design - Risks Analysis

Data and Other Risks

- Data integration difficulty
- Data quality and bias
- Over relying on AI Personas