



**Hakutaku - TCC**

# SUMMARY

<b>1. Project Members.....</b>	<b>4</b>
<b>2. Definition of the problem to be solved.....</b>	<b>4</b>
2.1 Introduction:.....	4
2.2 Problem Development:.....	5
2.3 Data/Comprovação:.....	7
2.4 Conclusion:.....	9
<b>3. Customer Segments.....</b>	<b>10</b>
3.1 Introduction:.....	10
3.2 Market Research (Glean):.....	10
3.3 Ideal Customer:.....	11
3.4 Minimum Requirements:.....	11
3.5 How the Solution Generates Value for These Customers:.....	13
3.6 Competitive Advantage:.....	13
<b>4. TAM, SAM, SOM:.....</b>	<b>14</b>
4.1 Introduction:.....	14
4.2 TAM Calculation:.....	15
4.3 SAM and SOM Segmentation:.....	17
<b>5. Analysis of Possible Technologies and Tools.....</b>	<b>19</b>
5.1. Large Language Models (LLM).....	20
5.2. Retrieval-Augmented Generation (RAG).....	21
5.3. Document Integration and Vector Databases.....	22
5.4. Tools and Infrastructure.....	22
5.4.1. Cloud Platforms (AWS, Azure, Google Cloud).....	22
5.4.2. Docker and Kubernetes.....	23
5.4.3. React and TypeScript.....	23
5.4.4. CI/CD Pipelines (GitHub Actions).....	24
5.5. Security and Privacy.....	25
5.5.1. Data Protection and Infrastructure.....	25
5.5.2. Security in AI Usage.....	26
5.5.2.1. Protection Against Data Leakage and Misuse.....	26
5.5.2.2. Defense Against AI Model Attacks.....	26
5.5.3. Compliance, Governance, and Continuous Monitoring.....	27
<b>6. Value Proposition.....</b>	<b>28</b>
6.1. Introduction.....	28

---

6.2. Value Proposition.....	29
6.3. Pains.....	31
6.4. Customer Jobs.....	31
6.5. Gains.....	32
6.6. Offered Service.....	32
6.7. Pain Relievers.....	33
6.8. Gain Creators.....	34
6.9. Conclusion.....	34
<b>7. Requirements Gathering.....</b>	<b>35</b>
7.1. Introduction.....	35
7.2. Chatbot.....	35
7.3. Data Ingestion.....	36
7.4. Administrator.....	38
7.5. Feature Roadmap.....	39
<b>8. Solution Interface.....</b>	<b>40</b>
8.1. Creation Process.....	40
8.2. Solution Interfaces.....	43
8.2.1. Homepage.....	43
8.2.1.1. Screen Objective.....	43
8.2.1.2. Use Cases.....	44
8.2.2. File Management and Integrations.....	45
8.2.2.1. Screen Objective.....	45
8.2.2.2. Use Cases.....	46
8.2.3. Chat.....	49
8.2.3.1. Screen Objective.....	49
8.2.3.2. Use Cases.....	50
8.2.4. Administrator Screen.....	53
8.2.4.1. Screen Objective.....	53
8.2.4.2. Use Cases.....	53
<b>9. Landing Page of the Solution.....</b>	<b>56</b>
9.1. Goal.....	56
9.2. Technologies Used.....	56
9.3. Content.....	57
9.3.1 Introduction.....	57
9.3.2 Features.....	58
9.3.3 How It Works.....	60
9.3.4 Benefits.....	61
9.3.5 Team.....	62
9.3.6 Form.....	63
<b>10. Initial development of the solution (Chatbot).....</b>	<b>64</b>

---

10.1. Objective.....	64
10.2. Technologies used.....	65
10.3. Screens.....	66
10.3.1 Home Screen.....	66
10.3.2 Conversation Screen.....	67
10.4. Functionalities.....	68
<b>documentationSources:</b> .....	<b>70</b>

## 1. Project Members

Names
Patrick Victorino Miranda
Raduan Oliveira Galli Muarrek
Vinicio Venancio Lugli
Rodrigo Campos Rodrigues
Pedro de Carvalho Rezende

## 2. Definition of the problem to be solved

### 2.1 Introduction:

Knowledge management is characterized as the process of capturing, organizing, sharing, and utilizing knowledge within an organization. There are two types of knowledge: tacit and explicit. According to Nonaka et al. (2000), "Explicit knowledge can be expressed in a formal and systematic language; it can be shared in data, scientific formulas, specifications, and manuals; it can be processed, transmitted, and stored easily. Tacit knowledge is highly personal and difficult to formalize."

An effective knowledge management approach encompasses both types, aiming to make them easily accessible within a company and intuitive to be consumed. However, there is often no efficient management in this regard. Frequently, for example, HR departments develop internal management systems or opt to acquire specialized knowledge management software for their specific needs. However, due to the closed nature and outdated design of most of these software solutions, knowledge is unable to be shared between the different departments of the organization. Artificial intelligence, on the other hand, has the ability to extend across multiple areas, gradually enabling the creation of knowledge management software that meets the needs of various departments within a company. This facilitates the use of knowledge generated not only within a single department but

---

across the entire organization. Despite these advantages, senior leadership still faces challenges in implementing this technology effectively.

## 2.2 Problem Development:

The problem we are addressing is knowledge management, which is characterized by the strategy and practice of managing all the content and data generated by the company in recent years.

In practice, this issue requires considerable discipline and organization. It involves everything from creating useful information for the company to storing, managing, and controlling access to it. Consequently, most companies face difficulties, especially in:

- I. Organizing their documents and knowledge;
- II. Creating new documents that effectively record data and knowledge;
- III. Sharing information between departments;
- IV. Unifying knowledge into a single platform;
- V. Establishing a robust document search system.

These challenges lead to problems such as:

- I. **Operational inefficiency**, as employees may spend hours searching for the same information;
- II. **Succession issues**, where employees who leave take tacit knowledge with them;

- 
- III. **Lengthy onboarding processes** that require the time of multiple workers to clarify questions that could be answered with proper documentation and organization;
  - IV. **Data fragmentation issues**, leading to extensive searches across multiple platforms to find desired information, which is not only inconvenient but also generates additional costs with platform maintenance.

From a financial perspective, these issues create significant problems for companies. Reduced efficiency means more time is needed to complete a task. Succession and onboarding problems ultimately relate to the same efficiency issue, as new employees need to interrupt the experienced team, occupying the time of both parties.

In addition to these financial concerns, effective knowledge management generates less tangible but equally valuable benefits: leveraging the knowledge acquired over the years to enhance daily activities and problem-solving. For example, if a consultancy has already conducted a go-to-market project, the ideal is to review both the macro and micro aspects of the previous project to improve future implementations. This allows the company to continuously learn from the information generated by past experiences.

## **2.3 Data/Comprovação:**

To validate the existence of this issue, market research, field data collection, and participation in competitions were conducted, which provided significant insights and validations regarding the relevance of the problem.

## Market Research:

According to a survey conducted by Notion for the 6th episode of their webinar (2024), **97% of leaders recognize the importance of knowledge management, but only 44% believe their organizations execute it effectively.** This data highlights the gap between the perception of the topic's relevance and its practical application within companies.

## Competitions and Hackathons:

- I. **Stark Bank Hackathon** → We participated in the Stark Bank hackathon, which focused on using AI in the financial market. During the event, the concept of Hakutaku was proposed, receiving praise from the internal team. The main issue identified was "disorganization within the company," which is common in startups experiencing rapid growth. Although it did not win the top prize, the feedback provided clear validation of the relevance of the problem.
  
- II. **Bemobi Hackathon** → In this hackathon, Hakutaku was presented with characteristics and functionalities similar to those it currently has. The project was awarded **first place**, reinforcing the acceptance and consistency of the proposed solution.

## Validation with Companies:

- III. **Farmoquímica SA** → In conversations with a trade marketing analyst, it was reported that finding information within the company is challenging.

---

Employees often have to ask on Microsoft Teams to find someone who can help. This situation illustrates how Hakutaku could address knowledge access issues.

- IV. **Vault** → In discussions with Vault's CTO, onboarding issues were highlighted, where new employees spend time on unproductive tasks, delaying their activities and those of other team members. This case emphasizes the direct impact of operational inefficiency.
  
- V. **Mavericks** → Mavericks' CFO mentioned succession problems, both in the current company and at Citi (where they previously worked). This underscores the issue of tacit knowledge loss when employees leave the organization.

These practical examples and quantitative data confirm that knowledge management challenges are widely recognized and experienced across different organizational contexts.

## 2.4 Conclusion:

Knowledge management, despite its recognized importance, remains a significant challenge for various organizations. Issues related to information organization, communication between departments, and retention of tacit knowledge are widely documented, as shown in the market research and validations conducted.

---

The data presented in section 2.3 corroborates the relevance and urgency of this issue. The gap between leaders' perceptions (97% recognize the importance) and effective execution (only 44% implement it correctly) highlights the existing disparity. Moreover, the validations obtained through hackathons and real-world companies strengthened the understanding that solutions like Hakutaku can address practical problems such as operational inefficiency, onboarding of new employees, and succession.

Therefore, it is evident that knowledge management is not just a theoretical problem but a practical and pressing need in the market. The implementation of technologies using artificial intelligence, combined with intuitive and accessible design, is a promising approach to addressing these challenges in information management, transforming the way organizations capture, store, and share internal knowledge.

Aqui está a tradução para o inglês:

### **3. Customer Segments**

#### **3.1 Introduction:**

First and foremost, before discussing how we deliver value, it is essential to understand who Hakutaku is capable of serving that is, what the prerequisites are for a company to become our client.

Additionally, it is important to highlight the profile and characteristics of the ideal customer, which will be our primary focus after brand consolidation. This group consists of companies that, beyond meeting the minimum requirements, have a

---

structure capable of maximizing the benefits of our solution, making them strategic partners for Hakutaku's growth.

### **3.2 Market Research (Glean):**

To understand the necessary requirements for adopting this type of solution, we gathered information from Glean, a global reference platform in AI for knowledge management. To achieve this, we presented ourselves as potential customers to understand what would be required for a company to become one of their clients.

After multiple email exchanges and a video conference, we obtained the following insights:

- The company must have a well established documentation base before implementing the solution.
- The minimum package offered was for 80 users.
- The annual cost of the basic package was \$80,000.

Based on the strategies and requirements of our main global competitor, we identified key metrics and factors that need to be considered when identifying potential customers, particularly because Glean's solution is the closest to the one we plan to develop.

### **3.3 Ideal Customer:**

From our research, we concluded that in an ideal scenario, especially after brand consolidation, our ideal customer would be large companies with hundreds of

---

employees, a well organized and consolidated knowledge base, and the financial capacity to afford an average monthly ticket price of at least R\$400 per employee.

However, since we are offering an entry level solution, we understand that while these are the ideal customers for our base, we must also adapt our product to serve companies that have not yet gained access to such solutions. To achieve this, we analyzed the research metrics to assess their criticality and determine how we can adjust them to make our product more accessible.

### **3.4 Minimum Requirements:**

Based on the insights gathered from Glean's research, although our goal is to create a more accessible solution, we identified some essential requirements for a company to be considered a potential Hakutaku customer. These requirements primarily relate to the volume of stored documents (internal company knowledge) and the number of users.

First, the contracting company must have a structured data and information base before implementing Hakutaku. While our plan is to ensure that the system can handle unstructured data through its search tool, there must be a significant amount of information available. This is crucial because the data provided needs to be sufficient for our chatbot to answer most employee inquiries accurately.

Additionally, it is important to consider the minimum number of users required to ensure the platform's viability. This requirement is linked to the maintenance and updating of the knowledge base. The higher the number of users, the more quality data will be available, and the more frequently missing or incorrect data will be

corrected through the system. While Glean requires a minimum of 80 users, we believe that with an efficient architecture, we can reduce this number, making the solution viable for companies with fewer users.

It is also worth noting that, beyond the dependency on user numbers for data maintenance, there is a financial barrier associated with the technologies we plan to use, such as LLM and RAG. These technologies require not only a significant number of platform users but also a considerable average ticket price to ensure financial feasibility.

While Glean's basic package costs \$80,000 per year, our goal is to lower this average ticket price, making the solution accessible to companies with fewer users without compromising quality or scalability.

### **3.5 How the Solution Generates Value for These Customers:**

Even though we understand the need to create a simpler, more accessible, and scalable solution to reach a larger portion of the market that still lacks access to such technologies, all the challenges presented in section 2.2 regarding problem development must be addressed. In other words, while simplifications may be made to create a more accessible platform, the developed features must ensure that the identified problems are effectively resolved.

Thus, to guarantee that value is delivered to customers, even with a more accessible platform, the solution must offer the following essential functionalities:

- Document and knowledge organization.
- Creation of new documents based on company knowledge.

- Information sharing across departments.
- Unification of company documents.
- Establishment of a robust document search system.

Ensuring that all the challenges outlined in section 2.2 are fully addressed by the solution will enable client companies to achieve greater operational efficiency, reduce succession issues, streamline onboarding processes, and minimize data fragmentation within the organization.

### **3.6 Competitive Advantage:**

Therefore, to ensure the planned value delivery for our knowledge management project while creating a competitive advantage over the competitor analyzed in our research, we decided to focus on making our product more accessible to medium sized companies and more suited to the Brazilian market.

To effectively deliver this competitive advantage, we have set the following goals for product development:

- 75% reduction in the number of required users, decreasing from 80 to 20.
- 75% reduction in the average ticket price per user, designing an architecture that allows for a cost below R\$100.00 per user.
- Development of a robust system for knowledge maintenance and creation, enabling work with knowledge bases that are not fully organized.
- Localization of the platform into Portuguese, ensuring greater accessibility for the Brazilian audience.

---

By designing a more accessible and optimized platform from the outset, we aim to provide a solution that reduces the need for large teams, minimizes the required pre-existing documentation, and offers a more competitive cost. This way, we seek to serve medium and small businesses that have not yet had access to such technology, making it more viable, especially for the Brazilian market.

## **4. TAM, SAM, SOM:**

### **4.1 Introduction:**

To determine the addressable market size for our product, we adopted a three-step segmentation process: TAM (Total Addressable Market), SAM (Serviceable Available Market), and SOM (Serviceable Obtainable Market). Each of these indicators represents a distinct portion of the market, considering different criteria of accessibility and penetration capacity. TAM represents the total available market, meaning the maximum potential value that could be reached if all companies and consumers in the target market used the product or service. SAM, in turn, reflects the accessible market based on more specific characteristics of companies or consumers who can effectively use the solution. Finally, SOM represents the portion of the market that can be effectively captured by the company, considering factors such as competition and execution capacity.

The following sections present the calculation logic and the sources used to determine the available market size for Hakutaku in each of these segments.

#### 4.2 TAM Calculation:

Before presenting the methodology used for the calculations and definition of our TAM, it is important to clarify which data were considered. We aimed to adopt a precise and reliable bottom-up strategy. To achieve this, we used metrics based on the number of companies with more than 50 employees, multiplied by our average ticket per employee and the average number of employees per company to determine the market size.

First, it is essential to explain how we arrived at the threshold of 50 employees. This value was defined based on our strategic objectives, particularly regarding the competitive advantage, which set a target of 20 employees per company, as described in section 3.6. To ensure, in a conservative manner, that the company has at least 20 employees with the potential to use the platform (i.e., those who work directly with computers and data handling), we opted to adopt the criterion of 50 employees. To obtain this information, we used data from Econodata, which, in March 2025, indicated the existence of 401,720 companies with more than 50 employees in Brazil.

To determine the average number of employees per company, we used data from the 2014 IBGE Trade Survey. Although this survey is outdated in terms of the total number of companies, it provides a general overview of the distribution of employees per company, allowing us to calculate the average number of employees per company. Based on the median of each company size category, we obtained the following results:

- 8,464 companies with 75 employees

- 
- 4,292 companies with 175 employees
  - 1,126 companies with 375 employees
  - 851 companies with 1,000 employees

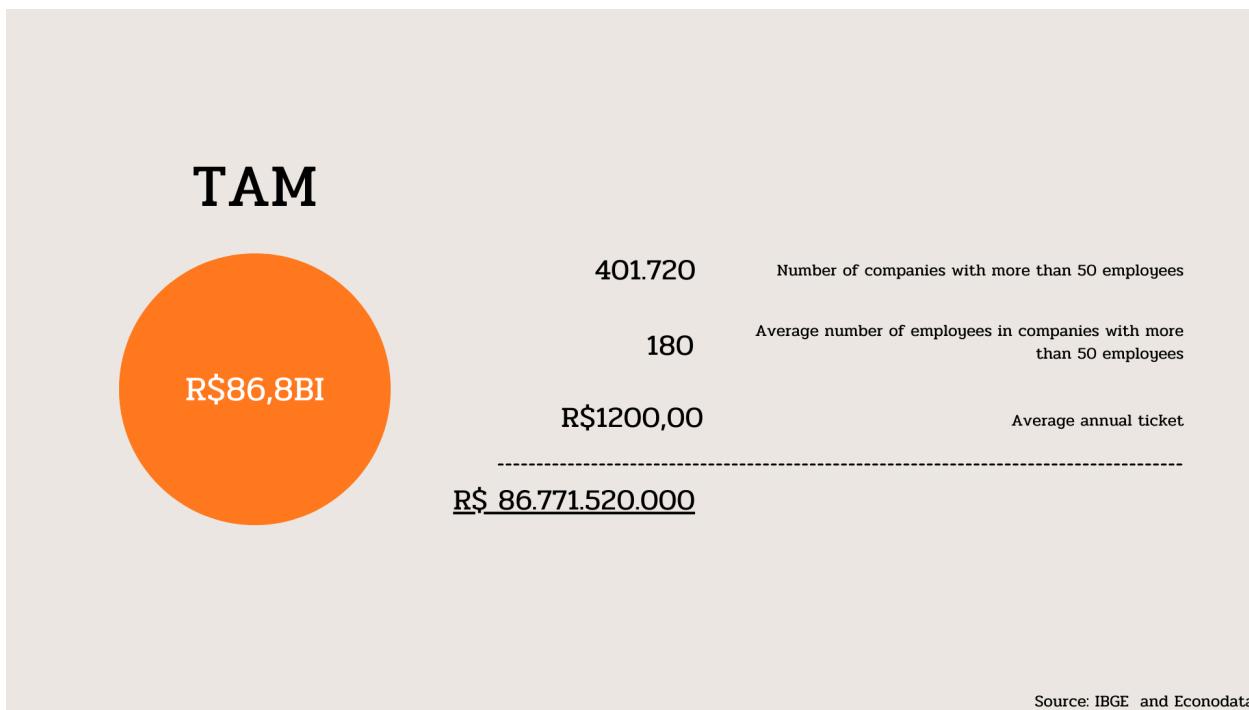
From this data, we calculated a weighted average of 180 employees per company.

By multiplying the total number of companies with more than 50 employees (401,720) by the average number of employees per company (180) and the projected average ticket mentioned in section 3.6 (R\$ 100 per month, totaling R\$ 1,200 per year), we obtain the following result:

$$401,720 \times 180 \times 1,200 = 86,771,520,000$$

That is, a TAM (Total Addressable Market) of approximately R\$ 86.8 billion, calculated using the bottom-up approach, as illustrated below:

[Figure 1: TAM Hakutaku]



*Source: Prepared by the authors, 2025.*

#### **4.3 SAM and SOM Segmentation:**

After calculating the market that represents our TAM, we carried out market segmentations to define the SAM (Serviceable Available Market) and the SOM (Serviceable Obtainable Market).

Initially, we defined our SAM as companies that use computers with internet access. According to the 2010 IBGE Census, 82.2% of companies with more than 10 employees use computers with internet for their daily activities. Based on this, we calculated the following segmentation:

$$86,771,520,000 \times 0.82 = 71,152,646,400$$

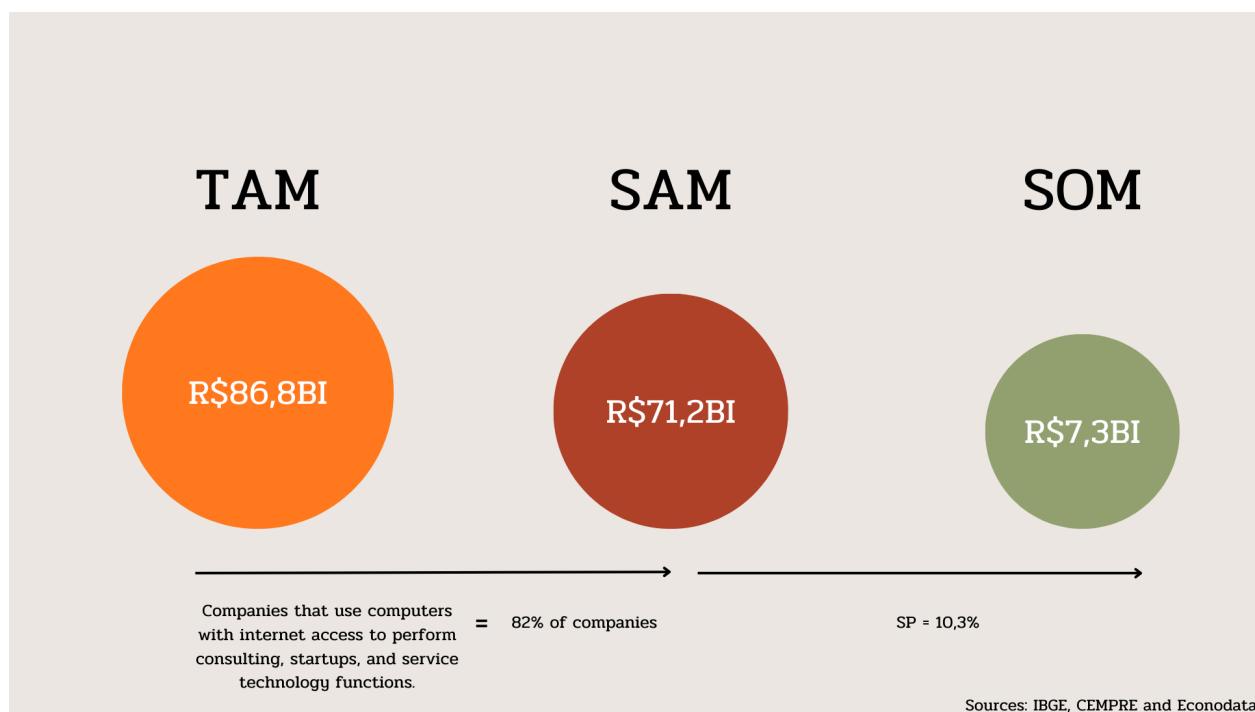
Thus, our SAM is R\$ 71.2 billion.

Regarding SOM, we applied a segmentation based on the city of São Paulo, which, according to IBGE data from 2019, accounted for 10.3% of the country's GDP. This allowed us to perform the following calculation:

$$71,152,646,400 \times 0.103 = 7,328,722,579$$

Thus, the SOM was estimated at **R\$ 7.3 billion**, as illustrated below, along with the SAM and SOM.

[Figure 2: TAM, SAM, SOM Hakutaku]



*Source: Prepared by the authors, 2025.*

After segmentation, we obtained the following results:

- TAM = R\$ 86.8 billion

- SAM = R\$ 71.2 billion
- SOM = R\$ 7.3 billion

## 5. Analysis of Possible Technologies and Tools

To efficiently address the knowledge management challenges described in the project's introduction, the Hakutaku solution requires advanced technologies capable of unifying, processing, and intuitively interacting with company data. The primary technical approach selected for this project involves integrating **Large Language Models (LLM)** with **Retrieval-Augmented Generation (RAG)**, providing an intelligent chatbot interface for internal knowledge queries.

**[Figure 3: Technology Possibilities and Structure - Hakutaku]**



---

*Source: Prepared by the authors, 2025.*

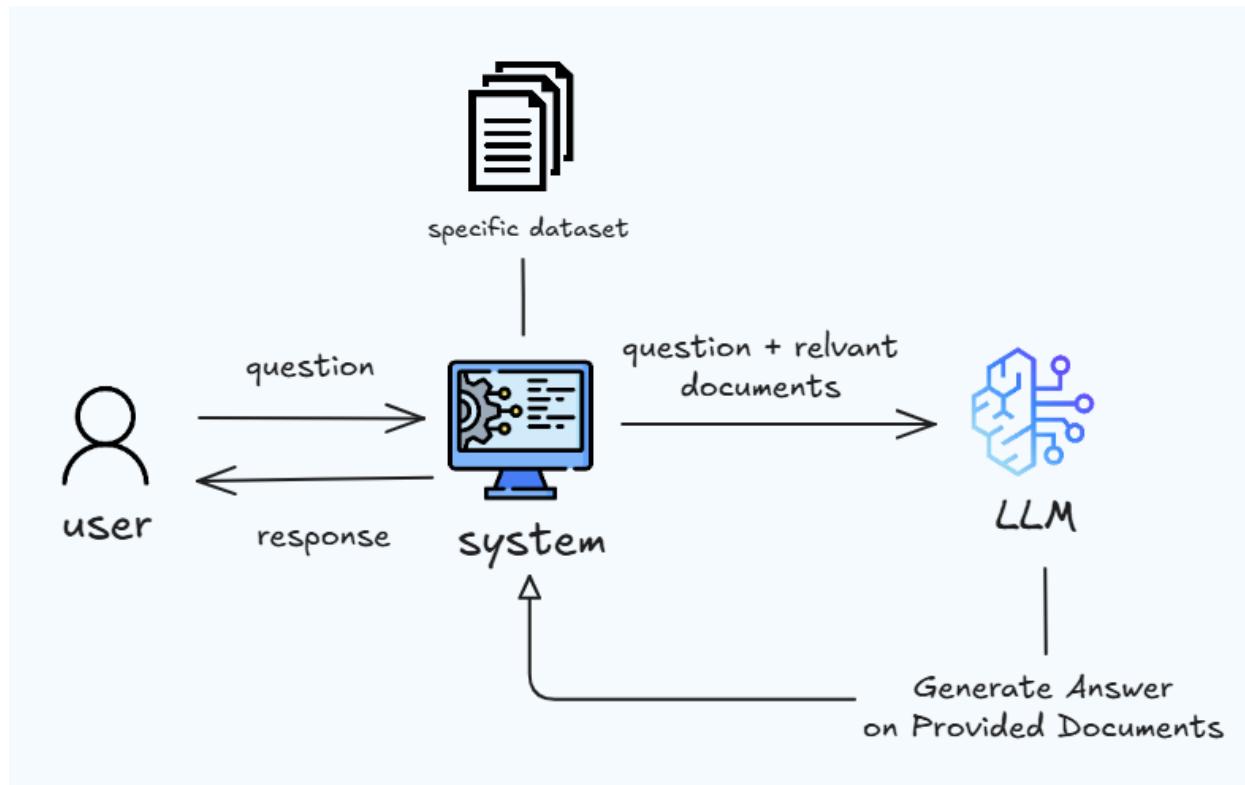
## 5.1. Large Language Models (LLM)

Large language models are advanced artificial intelligence systems trained on vast amounts of text, enabling them to understand context, semantics, and generate human-like text. Models such as GPT-4 by OpenAI, Claude by Anthropic, and LLaMA by Meta are viable options due to their advanced comprehension capabilities and adaptability to specific domains. GPT-4 stands out particularly for its widespread adoption in enterprise environments, robust developer ecosystem, and ease of integration via APIs, making it a powerful tool for enabling the conversational chatbot we plan to deliver to users.

## 5.2. Retrieval-Augmented Generation (RAG)

Retrieval-Augmented Generation combines the power of retrieval-based models with generative models. The RAG architecture addresses inherent limitations of generative models, such as hallucinations and inaccurate responses, by retrieving relevant information directly from the company's knowledge base before generating a response. This ensures that answers are accurate, relevant, and based on the most up-to-date data available from our users.

**[Figure 4: Common RAG Workflow - Hakutaku]**



*Source: Prepared by the authors, 2025.*

### 5.3. Document Integration and Vector Databases

To effectively implement the RAG architecture, Hakutaku needs efficient methods for document ingestion and retrieval. Technologies such as LangChain or LlamaIndex are recommended to create streamlined integration flows, as they facilitate data preprocessing, embedding generation, and queries. Additionally, vector databases like Pinecone, Weaviate, or Chroma can efficiently store and index documents, enabling fast similarity searches, which are essential for retrieving the files managed by the platform.

### 5.4. Tools and Infrastructure

---

For a practical and effective implementation of the Hakutaku solution, specific tools are required to ensure not only a robust technical framework but also a smooth and intuitive user experience. Below are the key tools and infrastructure components, highlighting their functions, strategic importance, and direct impact on users:

#### 5.4.1. Cloud Platforms (AWS, Azure, Google Cloud)

These platforms provide scalable and secure environments to host essential solution components, such as LLMs, vector databases, and front-end applications. Choosing cloud platforms enables:

- **Automatic Scalability:** Quickly adjusting resources to maintain high performance even during peak usage.
- **High Availability:** Minimizing downtime to ensure continuous chatbot access.
- **Advanced Security:** Offering robust protection against cyber threats, preserving corporate data integrity.

#### 5.4.2. Docker and Kubernetes

The use of containerization and orchestration tools is essential for:

- **Easy Deployment and Updates:** Containers enable frequent and seamless updates, ensuring users always have access to the latest and most secure system versions.

- **Resource Management:** Kubernetes automates resource allocation, ensuring the application responds efficiently to variable user demands.
- **Isolation and Security:** Isolated containers enhance security by preventing failures or issues in one part of the system from affecting the entire service.

#### 5.4.3. React and TypeScript

These front-end technologies are crucial for the user experience by allowing rapid and efficient development of intuitive and responsive interfaces:

- **Interactive Experience:** React provides a dynamic interface, enabling natural chatbot interactions.
- **Code Maintenance and Scalability:** TypeScript adds an extra layer of security, helping prevent errors and facilitating system maintenance and evolution.
- **Performance Optimization:** React enhances interface performance, ensuring quick user interactions.

#### 5.4.4. CI/CD Pipelines (GitHub Actions)

The use of Continuous Integration and Continuous Deployment (CI/CD) pipelines ensures **agility and reliability** in software delivery:

- **Process Automation:** Automates testing and deployments, reducing human errors and ensuring secure updates.
- **Rapid Feature Deployment:** Enables quick implementation of improvements and fixes, directly benefiting users with an always-updated system.
- **Continuous Monitoring and Feedback:** Facilitates early issue detection and resolution, continuously improving the user experience.

These combined tools ensure not only a technically robust solution but also a practical, secure, and efficient application, resulting in a highly positive user experience, covering frontend, backend, infrastructure, and monitoring/observability aspects.

## 5.5. Security and Privacy

The **sensitive nature of corporate internal data** requires rigorous privacy and security measures to ensure the confidentiality, integrity, and availability of information. The Hakutaku solution adopts multiple layers of protection, combining encryption, access control, auditing, and AI-specific strategies to mitigate risks such as **data leakage, misuse, and attacks on language models**.

### 5.5.1. Data Protection and Infrastructure

- **Data Encryption:** Implementation of **end-to-end encryption** for data in transit and at rest, using standards such as **AES-256** for storage and **TLS 1.3** for

---

secure communication.

- **Role-Based Access Control (RBAC)**: Ensures that users access only the data necessary for their roles, minimizing exposure to sensitive information.
- **Identity and Access Management (IAM)**: Implements **multi-factor authentication (MFA)** and strict permission controls for APIs and services.
- **Audit and Log Monitoring**: Detailed logging of system interactions to detect suspicious activities, ensure regulatory compliance, and maintain governance.
- **Redundancy and Secure Backup**: **Automated backup mechanisms** and **secure storage** for recovery in case of incidents.

### 5.5.2. Security in AI Usage

#### 5.5.2.1. Protection Against Data Leakage and Misuse

- **Input/Output Filtering**: Implementing techniques to sanitize queries and model responses, preventing exposure of sensitive information.
- **Data Anonymization and Obfuscation**: Reducing the risk of storing confidential information within the model using techniques like **differential privacy**.

- **AI Response Validation:** Using **Retrieval-Augmented Generation (RAG)** to ensure that answers are based solely on authorized sources.
- **Confidence Thresholds and Human Review:** Defining thresholds to **block low-confidence responses** and allow manual supervision when necessary.

#### 5.5.2.2. Defense Against AI Model Attacks

- **Prompt Injection Protection:** Implementing filters to prevent command manipulations that could lead to unauthorized responses.
- **Model Extraction Protection:** Limiting repetitive queries and applying **obfuscation techniques** to hinder reverse engineering of the model.
- **AI Response Monitoring:** Keeping detailed logs of all interactions for forensic analysis and continuous security improvements.

#### 5.5.3. Compliance, Governance, and Continuous Monitoring

- **Regulatory Compliance (LGPD, GDPR):** Implementing policies to **ensure compliance with data protection laws**, including the right to be forgotten and explicit consent.
- **Explainability and Transparency:** Maintaining **detailed documentation** of model decisions for audits and to enhance system trust.

- **AI Security Testing (Red Teaming):** Simulating attacks to identify vulnerabilities and strengthen defenses.
- **AI Incident Response:** Establishing clear procedures to handle failures or accidental data exposures.

By integrating **advanced AI technologies with robust security and privacy frameworks**, Hakutaku ensures a **reliable and effective knowledge management solution**, enhancing both **corporate data integrity** and **user experience**.

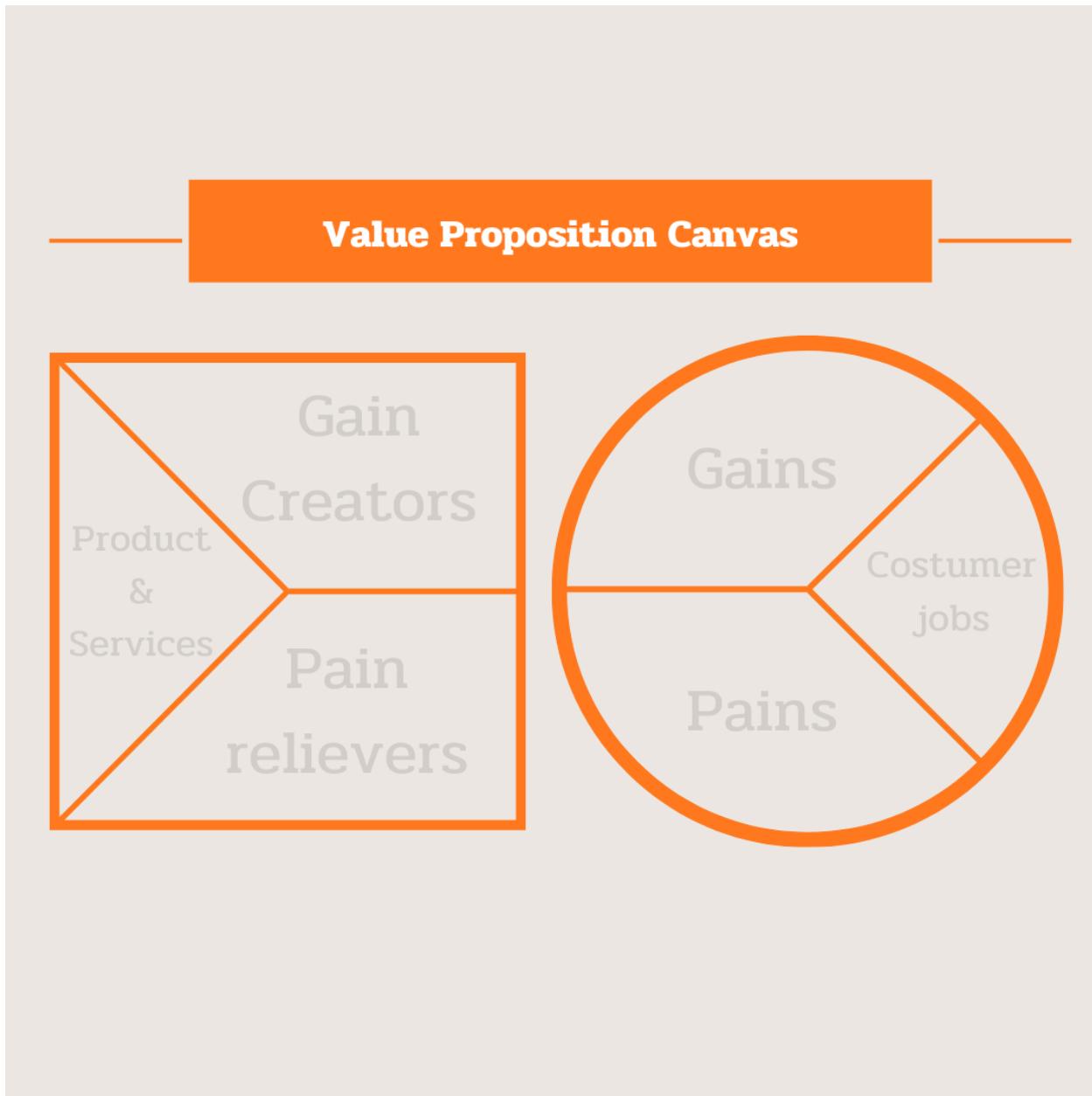
## 6. Value Proposition

### 6.1. Introduction

To ensure that Hakutaku's solution is truly aligned with the needs of our users, we used the Value Proposition Canvas, a strategic tool that allows us to clearly map customer pains, gains, and jobs, directly connecting this information to the benefits offered by the product. This helps us guide the platform's development to solve the right problems and deliver real value to our clients.

Below, you can see the layout used in the construction of Hakutaku's value proposition.

**[Figure 5: Value Proposition Canvas Template]**



Source: Prepared by the authors, 2025.

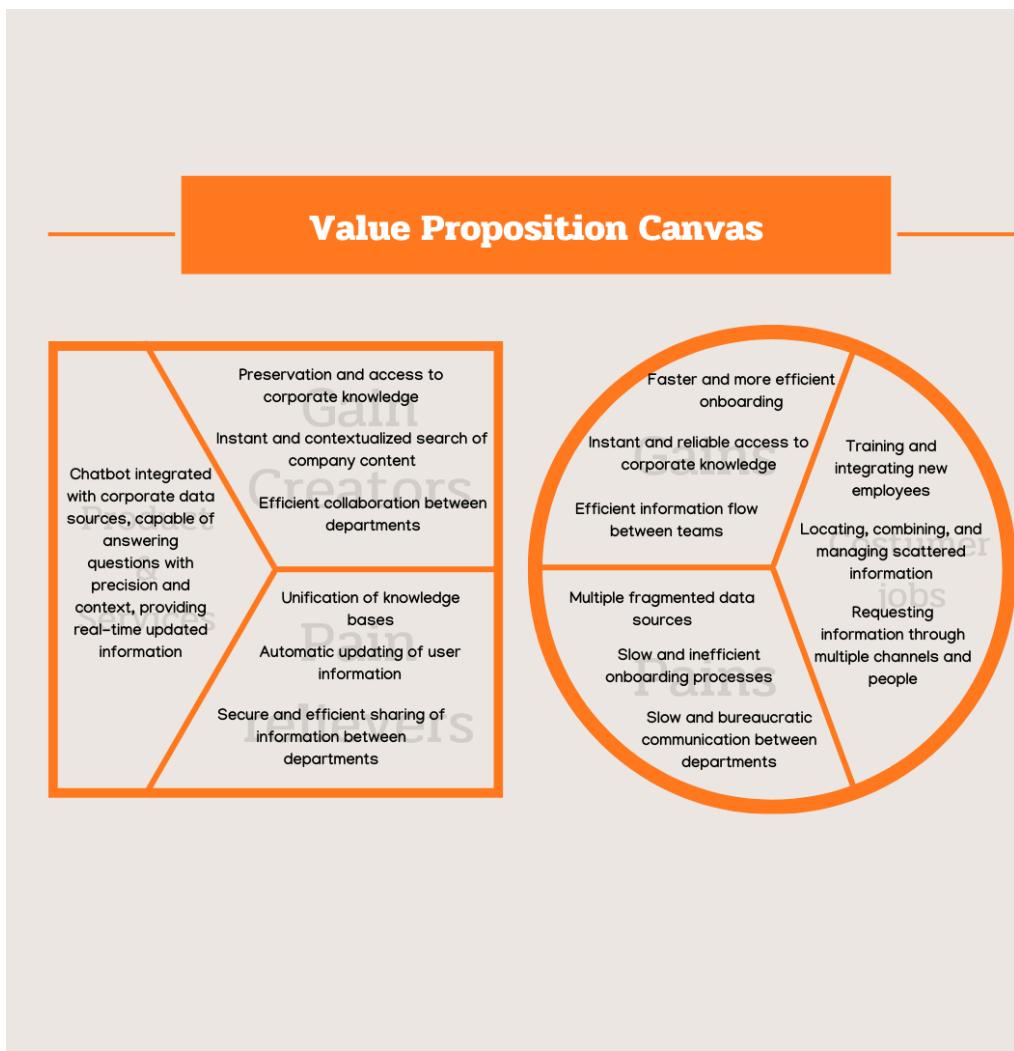
## 6.2. Value Proposition

Based on the conversations held throughout the ideation process focused on the topic of knowledge management we sought to understand the main **pains, expected**

**gains**, and **tasks** carried out by our potential clients. From these conversations, we mapped the most recurring points and built the **customer profile**, an essential step to properly guide our solution.

From this profile, we identified real opportunities for action and developed proposals aimed at solving the pains and generating concrete value. As a result, we created Hakutaku's value proposition, which can be seen below, along with a description of each mapped item.

[Figure 6: Hakutaku Value Proposition]



---

Source: Prepared by the authors, 2025.

### 6.3. Pains

Before designing any solution, we first tried to understand the pain points clients experienced regarding knowledge management. Below are the main issues identified during our conversations:

- **Multiple fragmented data sources**(Information is spread across different tools, making it difficult to find what's needed inside the company.)
- **Slow and inefficient onboarding processes**(New employees take too long to adapt due to difficulties finding and understanding company processes, which delays integration and consumes the time of other employees.)
- **Slow and bureaucratic communication between departments**(Teams waste time requesting and waiting for information from other areas.)

### 6.4. Customer Jobs

Next, we identified the main tasks employees perform on a daily basis that could be optimized or eliminated through a knowledge management solution:

- **Training and integrating new employees**(Employees need to spend time teaching processes, locating documents, and explaining how the company works to each new hire.)

- **Locating, combining, and managing scattered information**(Employees must search for data in different platforms and unstructured documents, making it harder to analyze and consolidate information.)
- **Requesting information through multiple channels and people**(It's common to send messages in Teams groups, emails, or directly contact colleagues from other departments to find documents or clarify doubts.)

## 6.5. Gains

During our interviews, after analyzing how clients currently manage their internal knowledge, we identified the main gains they expect from improving these processes:

- **Faster and more efficient onboarding**(Clients want new employees to integrate more quickly into the company.)
- **Instant and reliable access to corporate knowledge**(They want employees to find what they need quickly, with confidence in the accuracy of the information.)
- **Efficient information flow between teams**(They wish for information to circulate smoothly between departments, eliminating unnecessary back-and-forth.)

## 6.6. Offered Service

---

Based on user pains, needs, and expectations, we developed a solution to reduce the manual work clients currently face in managing internal knowledge, consolidating it into a single tool:

- **Chatbot integrated with corporate data sources, capable of answering questions with precision and context, providing real-time updated information**(This allows employees to quickly access internal knowledge based on reliable and updated data, reducing search time and optimizing processes such as onboarding, doubt resolution, and cross-team collaboration.)

## 6.7. Pain Relievers

Based on the previously identified pains, we designed functionalities to solve or reduce the problems our users face. Below are the main pain relievers offered by the platform:

- **Unification of knowledge bases**(Eliminates information fragmentation by bringing together dispersed data in a single platform, making it easier to access and reducing time spent searching.)
- **Automatic updating of user information**(Ensures that the data being accessed is always current based on the connected sources, avoiding rework.)
- **Secure and efficient sharing of information between departments**(Facilitates interdepartmental communication through a system

---

with access control, removing the need for manual message exchanges between teams.)

## 6.8. Gain Creators

Beyond solving existing problems, with Hakutaku we aim to generate significant value for companies by delivering the benefits they expressed as essential to improve knowledge management. The following are the main gains we plan to deliver through our solution:

- **Preservation and access to corporate knowledge**(Ensures that the knowledge accumulated by both current and former employees is recorded and accessible to the entire company, preventing information loss over time or due to turnover.)
- **Instant and contextualized search of company content**(Allows users to quickly and accurately find information, taking the context of their question into account and drastically reducing search time.)
- **Efficient collaboration between departments**(Improves information flow between teams, promotes synergy, eliminates communication bottlenecks, and enables more integrated operations.)

## 6.9. Conclusion

---

Therefore, the value proposition helps us clearly understand what the client expects from the solution, serving as an excellent guide for defining functionalities and requirements during product development. It also supports strategic decisions and ensures that we are building something truly aligned with the users' real pains and goals.

## 7. Requirements Gathering

### 7.1. Introduction

In order to develop a solution that truly solves the problems we have identified in companies, an essential step in the process is to fully understand what the system needs to do and how it should behave. That's why, before starting development, we chose to carry out a requirements gathering process. In this process, we documented both the functional requirements ("what" the system must do) and the non-functional requirements ("how" the system must do it). It's worth mentioning that this gathering was organized by system modules, which are: chatbot, data ingestion and file management, and the administrator area.

The main goal of this step is to avoid rework by clearly defining the expected functionalities for the MVP we are building.

### 7.2. Chatbot

This section covers the requirements related to the chatbot module. This module of the system encompasses all functionalities related to the chat screen of the

solution, aiming to ensure it connects properly to the LLM and is capable of consuming files through RAG. These files should be automatically provided by the functionalities that will be defined in the data ingestion module.

Below you will find the system requirements for this module:

Identifier	Module	Description	Dependencies	Classification
RF001	ChatBot	The system must allow the user to initiate a conversation with the chatbot through the platform's main menu.	-	Important
RF002	ChatBot	If no response is being generated for the user, the system must allow the user to send a new message in the chat.	-	Essential
RF003	ChatBot	The system must send the user's messages to the language model (LLM) for processing.	RF002	Essential
RF004	ChatBot	The system must save the user's conversation history during the session to ensure coherence in subsequent responses.	RF002	Essential
RF005	ChatBot	The system must allow the user to access previous conversations from the platform's home screen.	RF004	Important
RF006	ChatBot	The system must display a visual indicator while the model is generating a response.	RF002	Nice to have
RNF001	ChatBot	The system must respond to 95% of user requests within 30 seconds after the message is sent.	RF002	Essential
RNF002	ChatBot	The system must ensure that user information is not used to train third-party LLMs.	-	Important
RNF003	ChatBot	The system must provide feedback to the user in case of errors or connection failures.	-	Important

### 7.3. Data Ingestion

This section covers the requirements related to the system's data ingestion module.

This module includes all functionalities related to the consumption and processing of data that will be used by RAG. Its purpose is to ensure that user data is automatically consumed, stored, and filtered so that RAG can provide business context to the LLM as effectively as possible.

Below you will find the system requirements for this module:

Identifier	Module	Description	Dependencies	Classification
RF007	Data Ingestion	The system must receive and consume data from a vector database.	-	Essential
RF008	Data Ingestion	The system must allow the user to upload and process files in .docx format.	-	Essential
RF009	Data Ingestion	The system must allow the user to upload and process files in .xlsx format.	RF008	Important
RF010	Data Ingestion	The system must retrieve company information through the RAG mechanism and provide it as context to the LLM.	RF003	Essential
RF011	Data Ingestion	The system must display, along with the chatbot's response, the information sources used by the RAG mechanism.	RF008	Essential
RF012	Data Ingestion	The system must allow the user to connect to a Google Drive account and consume the files stored there.	RF008	Essential
RF013	Data Ingestion	The system must allow the user to view and manage the documents that Hakutaku has access to.	RF008	Important
RNF004	Data Ingestion	The system must update changed files in the drive within 10 minutes for at least 95% of changes.	RF013	Important
RNF005	Data Ingestion	The system must log success and error events for each data ingestion operation.	RF008	Important
RNF006	Data Ingestion	The system must complete the ingestion process of files up to 20 MB in no more than 45 seconds.	RF008	Important

## 7.4. Administrator

This section covers the requirements related to the system's administration module.

This module is responsible for centralizing access management features, controlling data access permissions, and monitoring platform usage, allowing administrators to keep the database secure and up to date.

Below are the key system requirements for this module:

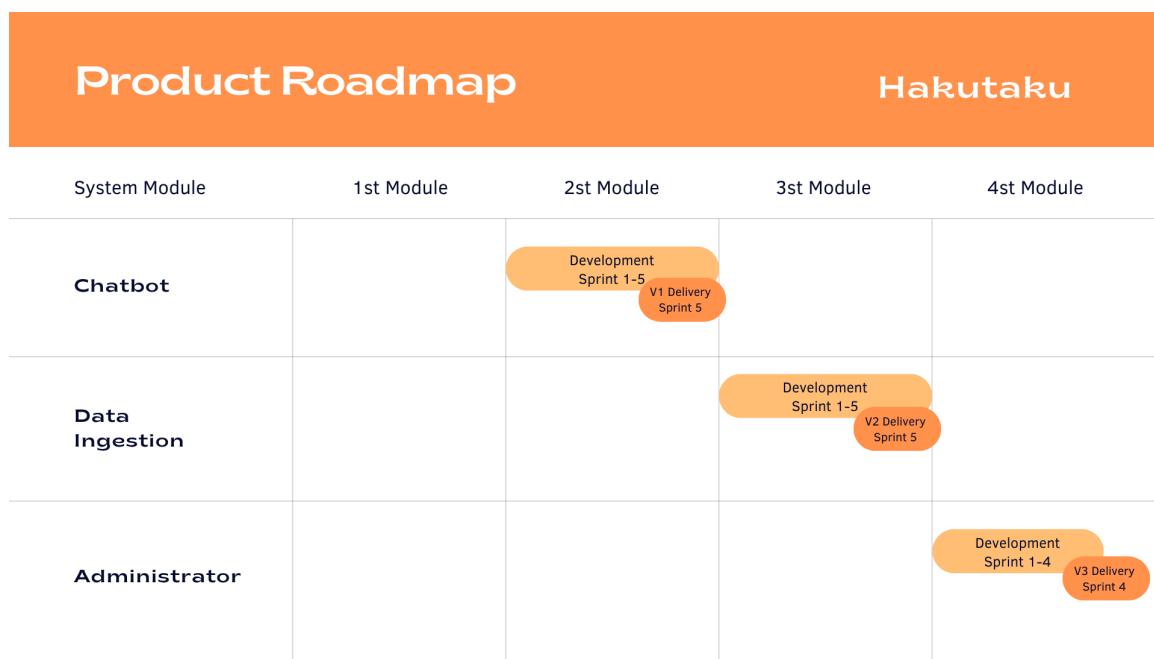
Identifier	Module	Description	Dependencies	Classification
RF014	Administrator	The system must allow the administrator to add new users to the platform.	-	Essential
RF015	Administrator	The system must allow the administrator to remove users from the platform.	RF014	Essential
RF016	Administrator	The system must allow the administrator to view the list of all users registered on the platform and their information.	RF014	Essential
RF017	Administrator	The system must allow the administrator to create, edit, and delete roles with different access levels.	-	Essential
RF018	Administrator	The system must allow the administrator to assign one or more roles to each user.	RF017	Essential
RF019	Administrator	The system must allow the administrator to define which roles have access to which documents and data connectors.	RF018	Essential
RF021	Administrator	The system must provide the administrator with a dashboard containing platform usage metrics.	RF003	Important
RNF007	Administrator	The system must ensure that only users with an administrator role have access to administration features.	-	Essential
RNF008	Administrator	The platform metrics dashboard must be updated with a maximum delay of 24 hours.	RF021	Important

## 7.5. Feature Roadmap

To ensure our solution is developed efficiently and aligned with the defined project requirements, it's crucial to have a clear understanding of when each feature will be delivered.

This roadmap aims to guide the development of our MVP by strategically prioritizing functionalities to support the system's evolution. It will help us stay focused, align deliverables with the team, and reduce the risk of rework. Below is the planned implementation order for the gathered requirements:

[Figure 7: Feature Roadmap]



Source: Prepared by the authors, 2025.

---

Thus, the delivery of one module per academic term is scheduled, with the Chatbot module being the first to be completed, expected by the end of Sprint 5 of Module 2. Next, the Data Ingestion module will be delivered by the end of Sprint 5 of Module 3. Finally, the Administrator module will be delivered in Module 4, at the end of Sprint 4.

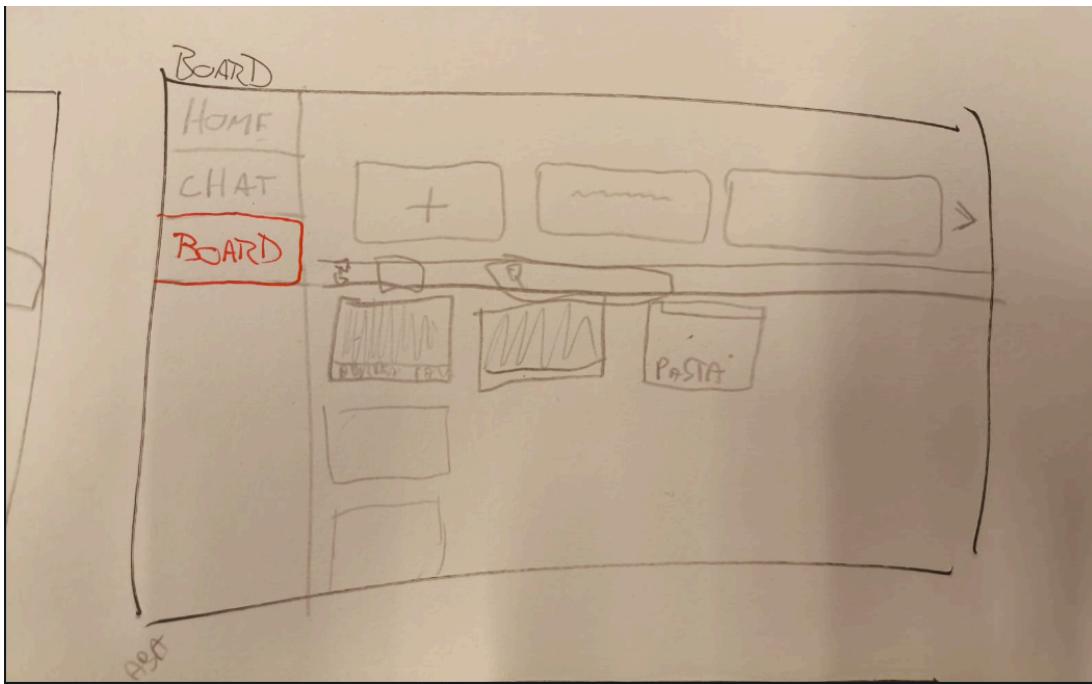
## 8. Solution Interface

### 8.1. Creation Process

The process of creating Hakutaku's screens was carried out in well-defined stages, with the aim of structuring the application flow and validating functionalities before the final implementation.

Initially, we made drawings on cardboard, where we manually sketched the main functionalities and the visual organization of the screens. This stage was important for quickly discussing ideas and defining the basic structure of the system.

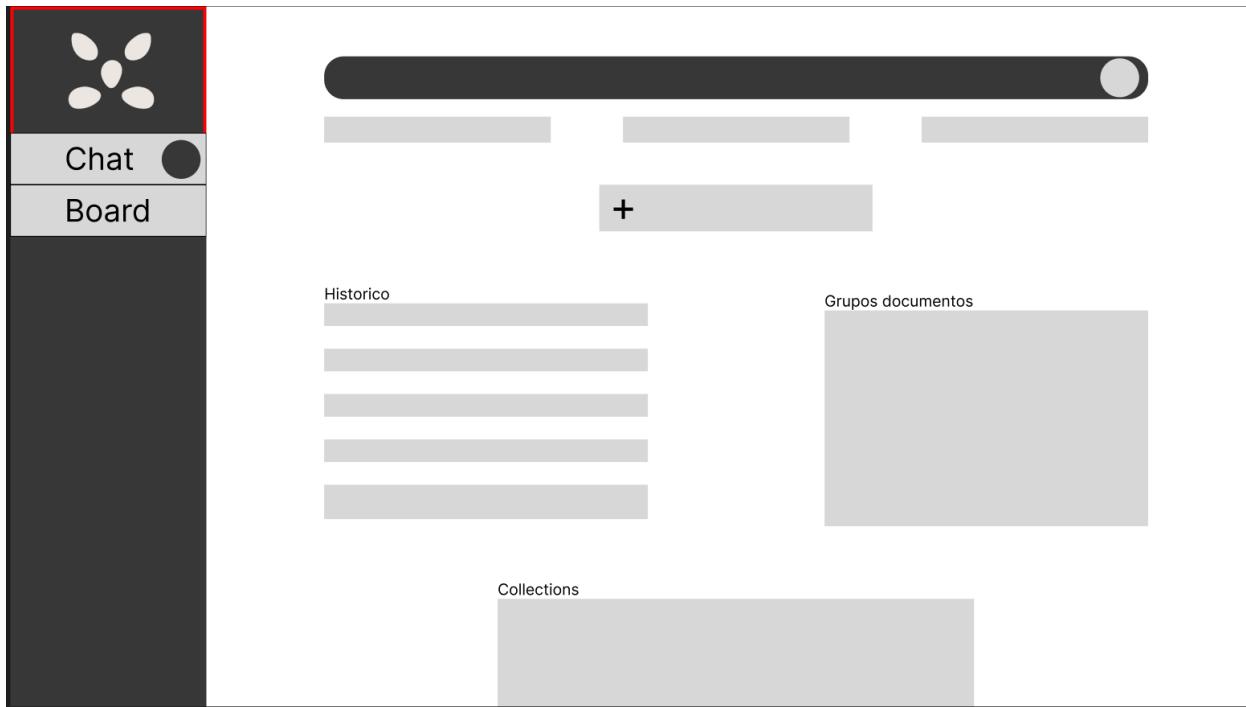
[Figure 8: Initial cardboard sketches of Hakutaku's screens]



Source: Prepared by the authors, 2025.

Next, based on the approved sketches, we developed a wireframe of the solution. In this version, we defined the hierarchy of information, the arrangement of components, and the navigation sequence between screens. The focus of the wireframe was to ensure that all necessary elements were represented and that the flows were functional, without worrying about the final visual design. This stage allowed the system logic to be validated before proceeding to graphic refinement.

[Figure 9: Digital wireframe of Hakutaku's screens]



Source: Prepared by the authors, 2025.

Subsequently, we developed a high-fidelity mockup, adding colors, typography, icons, and other graphic components. The mockup served to visualize what the complete interface would look like and enabled visual adjustments before developing the final screens.

[Figure 10: High-fidelity mockup of Hakutaku's screens]



Source: Prepared by the authors, 2025.

Finally, we created the definitive screens, consolidating the final design that was implemented on the platform. This version incorporated all functionality and visual adjustments made in the previous stages.

## 8.2. Solution Interfaces

### 8.2.1. Homepage

#### 8.2.1.1. Screen Objective

---

The main objective of Hakutaku's Homepage screen is to guide users in quickly accessing the platform's main functionalities, acting as a central navigation point. Through this screen, users can easily search for documents in the knowledge base and start conversations with the chatbot.

Additionally, the homepage displays important metrics related to the platform's use, such as the number of queries resolved, time saved, number of active users, and the total number of knowledge items stored. This provides an overview of the platform's performance and the value delivered by Hakutaku to the company.

#### **8.2.1.2. Use Cases**

The main use cases related to Hakutaku's Homepage screen are listed below, illustrated with the corresponding image:

##### **Navigation to main functionalities:**

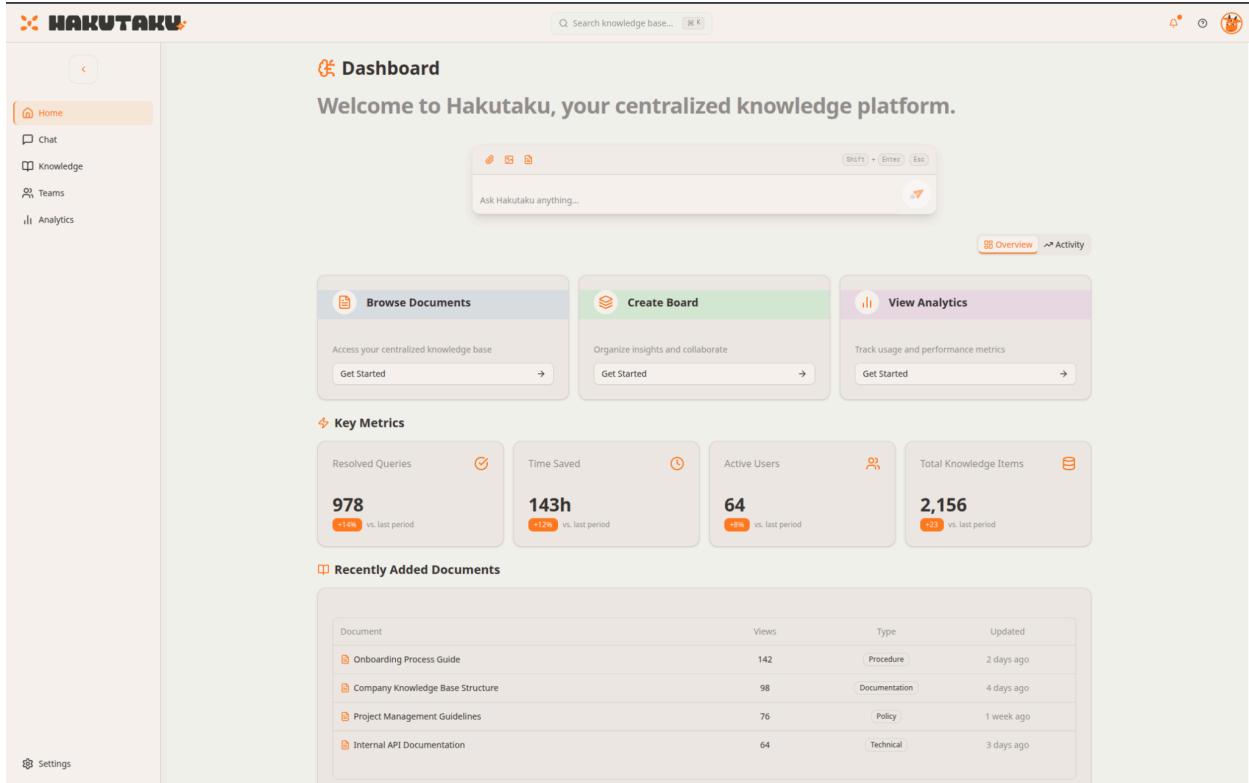
The homepage allows users to quickly access essential functionalities, such as searching for documents in the knowledge base or initiating new interactions with the chatbot, making it easier to obtain information and resolve doubts in a practical and fast way.

##### **Viewing metrics:**

The screen displays key performance indicators of the platform, such as the number of queries resolved, time saved, and active users, offering a quick and efficient overview of Hakutaku's use and impact within the organization.

These functionalities can be observed in the image below, which shows the platform's home screen with navigation shortcuts and main performance metrics.

[Figure 11: Hakutaku homepage screen with navigation shortcuts and metrics]



The screenshot displays the Hakutaku homepage with the following features:

- Navigation Bar:** Includes links for Home, Chat, Knowledge, Teams, and Analytics.
- Search Bar:** A search input field with placeholder "Ask Hakutaku anything..." and keyboard shortcuts (Shift + Enter, Esc).
- Dashboard Header:** "Welcome to Hakutaku, your centralized knowledge platform."
- Key Metrics:**
  - Browse Documents:** Access your centralized knowledge base. Get Started.
  - Create Board:** Organize insights and collaborate. Get Started.
  - View Analytics:** Track usage and performance metrics. Get Started.
- Key Metrics Summary:**

	Value	Variance
Resolved Queries	978	+14% vs. last period
Time Saved	143h	+12% vs. last period
Active Users	64	+8% vs. last period
Total Knowledge Items	2,156	+20% vs. last period
- Recently Added Documents:**

Document	Views	Type	Updated
Onboarding Process Guide	142	Procedure	2 days ago
Company Knowledge Base Structure	98	Documentation	4 days ago
Project Management Guidelines	76	Policy	1 week ago
Internal API Documentation	64	Technical	3 days ago
- Settings:** A link to the settings menu.

Source: Prepared by the authors, 2025.

### 8.2.2. File Management and Integrations

#### 8.2.2.1. Screen Objective

The main objective of Hakutaku's file and integration management screen is to centralize and facilitate the organization, visualization, and continuous updating of the company's data. This interface allows users to view the documents stored in the knowledge base along with additional information such as creators, edit history, associated tags, and access permissions.

---

Additionally, this screen has a fundamental pop-up for uploading new files, such as Word documents, PowerPoint presentations, and PDFs, as well as for integration with external data sources such as Google Drive, Slack, and OneDrive, enabling the knowledge base to be automatically and continuously updated with new documents.

### **8.2.2.2. Use Cases**

The main use cases related to Hakutaku's file and integrations management screen are listed below, along with images of the screens where these use cases can be found:

#### **File viewing:**

Users can view all documents centralized on the platform, accessing detailed information such as the area related to the content, who created the file, the edit history, and the tags used for access permissions.

#### **Search and Filters:**

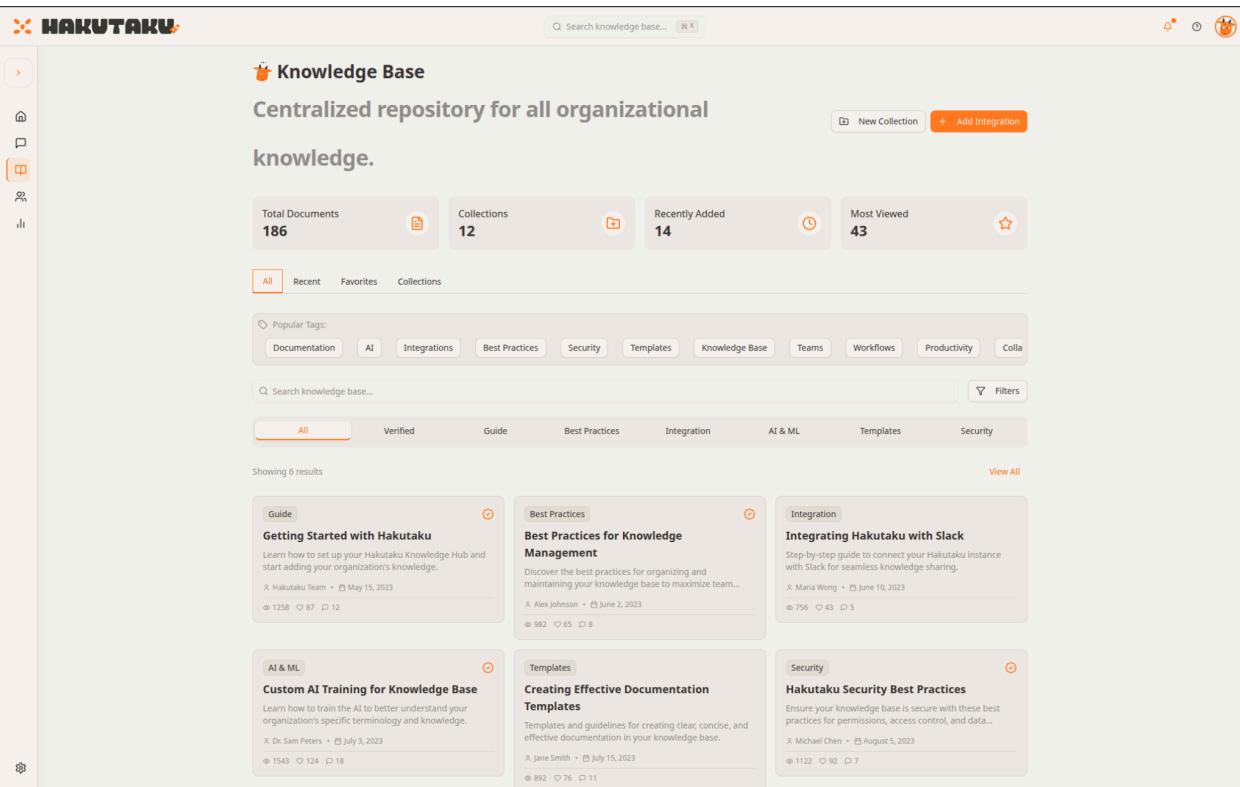
Users can quickly locate files based on multiple filters such as tags, areas, or document types. This functionality significantly optimizes the search time within the platform.

#### **Creation of Collections:**

For better organization, users can group files into thematic collections. This facilitates the organization of related content, such as training materials or technical documentation, promoting a more structured knowledge management.

These functionalities can be observed in the image below, which shows Hakutaku's main knowledge base screen with organized and categorized documents.

[Figure 12: Knowledge base screen with centralized and categorized documents]



The screenshot displays the Hakutaku Knowledge Base interface. At the top, there is a header with the Hakutaku logo, a search bar, and several navigation icons. Below the header, the title "Knowledge Base" and the subtitle "Centralized repository for all organizational knowledge." are displayed. Key statistics are shown in boxes: "Total Documents 186", "Collections 12", "Recently Added 14", and "Most Viewed 43". Below these, a navigation bar includes tabs for "All", "Recent", "Favorites", and "Collections". A "Popular Tags" section lists categories like Documentation, AI, Integrations, Best Practices, Security, Templates, Knowledge Base, Teams, Workflows, Productivity, and Collaboration. The main content area shows a grid of six document cards, each with a category label (Guide, Best Practices, Integration, AI & ML, Templates, Security) and a title. Each card also includes a brief description, author information, and view counts.

Category	Title	Description	Author	Views
Guide	Getting Started with Hakutaku	Learn how to set up your Hakutaku Knowledge Hub and start adding your organization's knowledge.	A. Hakutaku Team	1258
Best Practices	Best Practices for Knowledge Management	Discover the best practices for organizing and maintaining your knowledge base to maximize team...	A. Alex Johnson	982
Integration	Integrating Hakutaku with Slack	Step-by-step guide to connect your Hakutaku instance with Slack for seamless knowledge sharing.	A. Maria Wong	756
AI & ML	Custom AI Training for Knowledge Base	Learn how to train the AI to better understand your organization's specific terminology and knowledge.	A. Dr. Sam Peters	1543
Templates	Creating Effective Documentation Templates	Templates and guidelines for creating clear, concise, and effective documentation in your knowledge base.	A. Jane Smith	892
Security	Hakutaku Security Best Practices	Ensure your knowledge base is secure with these best practices for permissions, access control, and data...	A. Michael Chen	1122

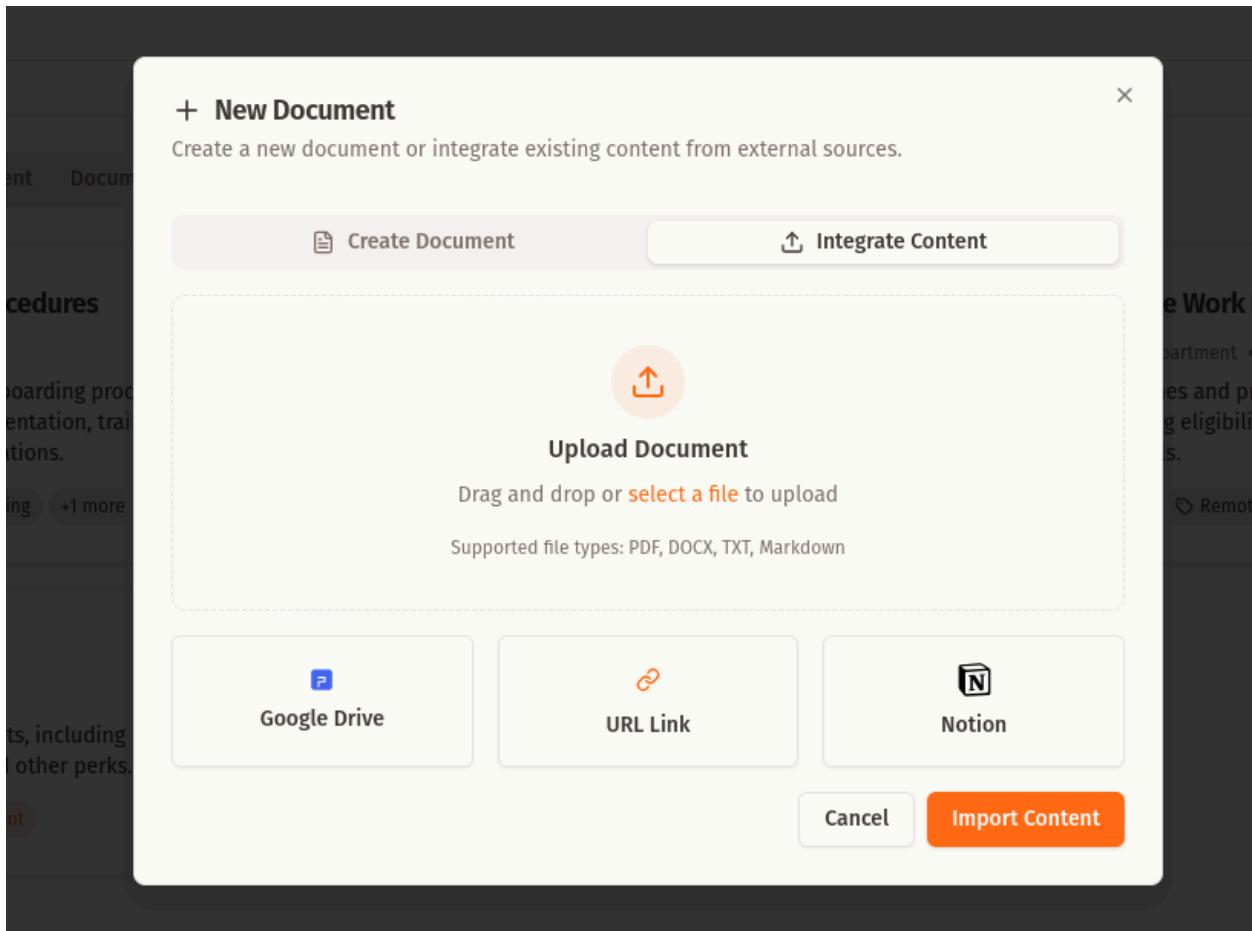
Source: Prepared by the authors, 2025.

### Adding New Files:

The pop-up screen, accessible via the "add Integration" button in the upper right corner of the page, allows for the manual import of new documents in various formats such as DOC, PDF, TXT files, and PowerPoint presentations. This functionality ensures that any relevant content can be easily incorporated into the organization's knowledge base.

In the image below, it is possible to visualize the content integration interface, where the user can directly upload documents or integrate files from external services.

[Figure 13: Content integration screen (document upload)]



Source: Prepared by the authors, 2025.

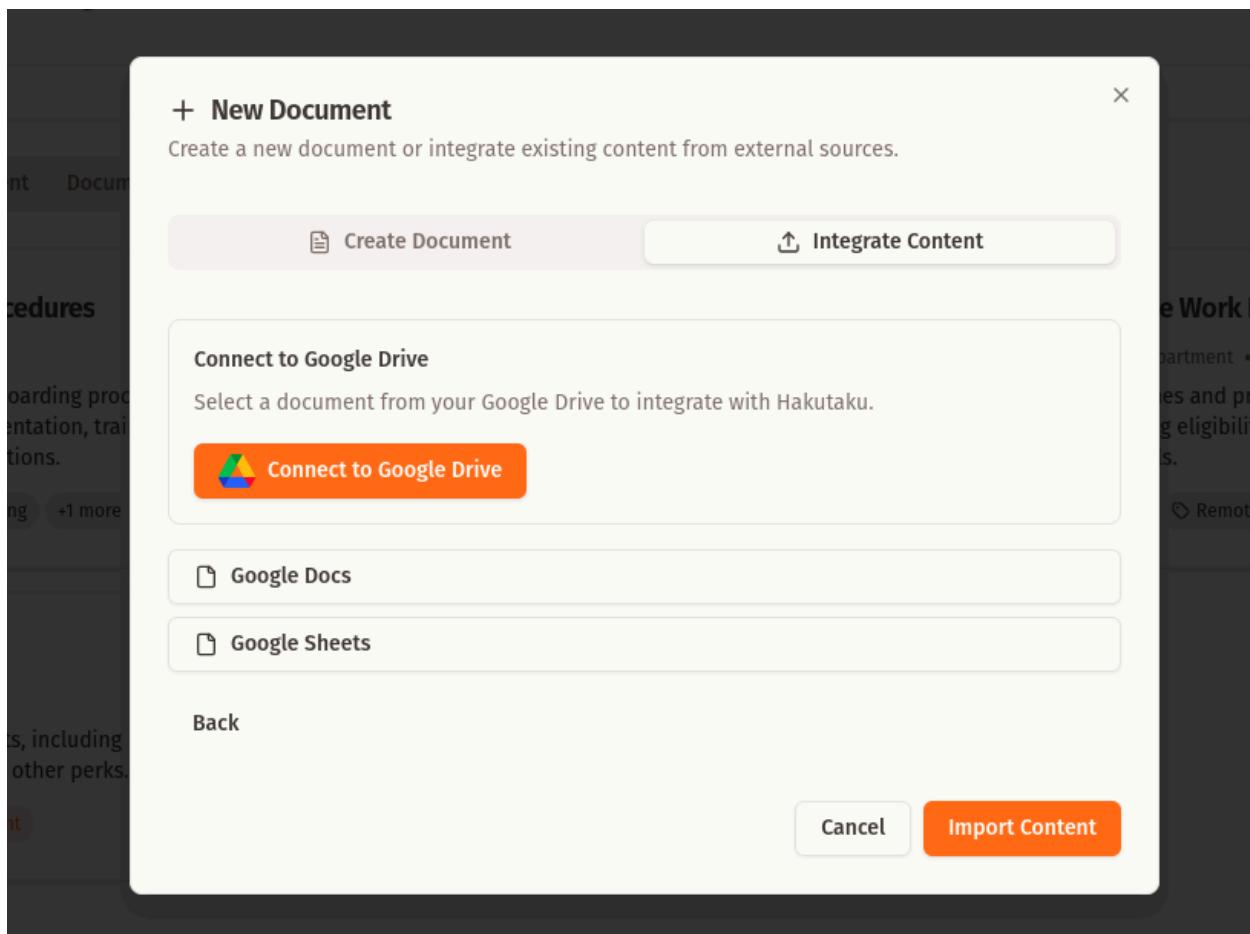
### Integration with External Sources:

The pop-up also offers the possibility of creating connections with external sources, such as Google Drive. This enables the automatic updating of the knowledge base whenever new documents are added to these sources, eliminating the need for

frequent manual imports and ensuring that the base is always up to date and synchronized with the company's operations.

In the following screen, it is possible to visualize the interface that enables these integrations:

[Figure 14: Integration screen with Google Drive]



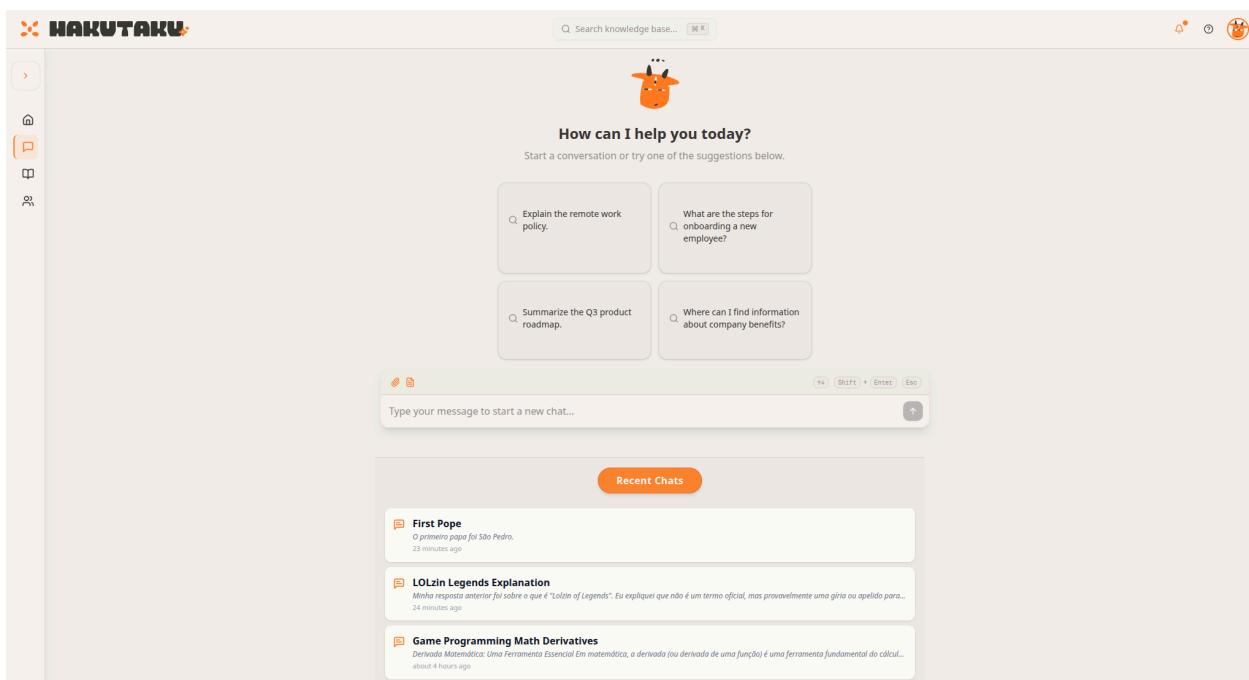
Source: Prepared by the authors, 2025.

### 8.2.3. Chat

#### 8.2.3.1. Screen Objective

Hakutaku's Chat screen is the heart of the interaction between the user and the knowledge management system. The main objective is to provide an intuitive and efficient conversational interface where users can obtain precise and contextual responses to their questions and information needs. This screen acts as a central access point to corporate knowledge, allowing company employees to ask questions in natural language and receive relevant answers based on the data and documents stored on the platform through the File Management screen. This reduces dependence on manual information searches, optimizes response time, and ensures that the company's knowledge is intuitively accessible to all members of the organization.

[Figure 15: Main Chat screen]



Source: Prepared by the authors, 2025.

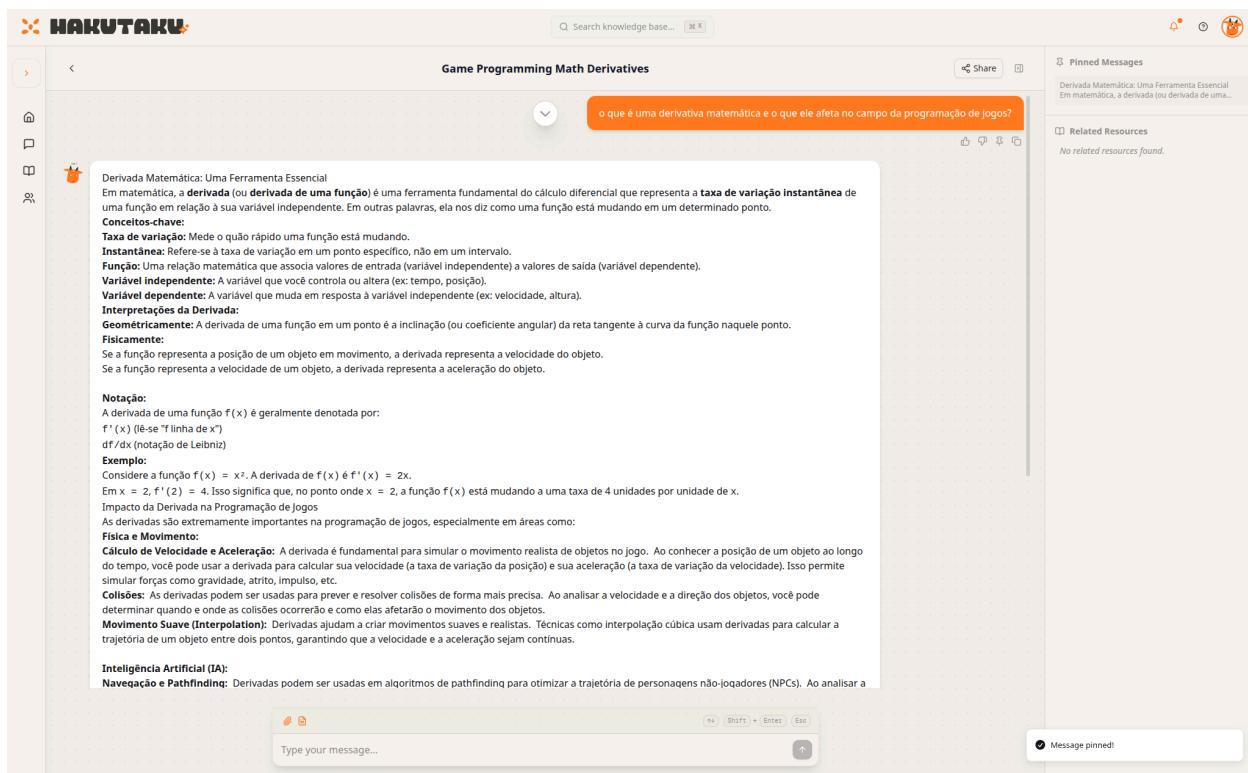
#### 8.2.3.2. Use Cases

On Hakutaku's Chat screen, users can perform various actions and interactions, obtaining information and support quickly and efficiently. Below are some of the main use cases:

### Specific information queries:

The user can ask about a specific topic, such as "What is the process for requesting reimbursement of expenses?" or "Where can I find the confidentiality agreement template?". The chatbot, using the RAG mechanism, searches the company's knowledge base and provides an accurate and relevant response, along with the information sources used.

[Figure 16: Main Chat screen after sending a question]



The screenshot shows the Hakutaku chat interface. The search bar at the top contains the query "o que é uma derivada matemática e o que ela afeta no campo da programação de jogos?". The main pane displays a search result titled "Game Programming Math Derivatives". The result is a summary of mathematical derivatives, mentioning their definition as the rate of change of a function, and their application in game programming for calculating slopes, velocities, and accelerations. It also discusses geometric interpretations, physics applications, and specific terms like Leibniz notation. A pinned message on the right side of the interface reads: "Derivada Matemática: Uma Ferramenta Essencial Em matemática, a derivada (ou derivada de uma função) é uma ferramenta fundamental do cálculo diferencial que representa a taxa de variação instantânea de uma função em relação à sua variável independente. Em outras palavras, ela nos diz como uma função está mudando em um determinado ponto." Below this message, there is a note stating "No related resources found."

Source: Prepared by the authors, 2025.

### **Resolution of specific doubts:**

The user can ask quick and direct questions, such as "Who is responsible for the marketing department?" or "What is the deadline for submitting the monthly report?".

The chatbot responds concisely, streamlining the retrieval of basic and routine information.

### **Conversation history:**

The user can access the history of their previous conversations to review information already obtained or continue an interrupted consultation. This ensures continuity and coherence in the dialogue with the system.

### **Requesting documents and files:**

The user can request specific documents, such as "Can you send me the best practices manual?" or "I need the file with the 2024 strategic plan." The chatbot, if it has access to these documents, can provide them directly or indicate where they can be found on the platform.

### **Obtaining explanations and context:**

The user can ask the chatbot to explain a complex concept or process in detail, such as "How does the performance evaluation system work?" or "What methodology is used for data analysis?". The chatbot can provide clear and contextualized explanations, assisting in understanding and learning.

These possible use cases demonstrate how Hakutaku's Chat screen can be used to facilitate access to knowledge, contributing to optimizing communication between areas and the information flow within the company, creating a more efficient work environment.

## 8.2.4. Administrator Screen

### 8.2.4.1. Screen Objective

Hakutaku's Administration screen is the central control panel for managing and maintaining the system's integrity. Its main objective is to provide administrators with the necessary tools to configure, monitor, and maintain the platform. This includes managing users and their accesses, defining permissions for file access, and monitoring activity within the platform.

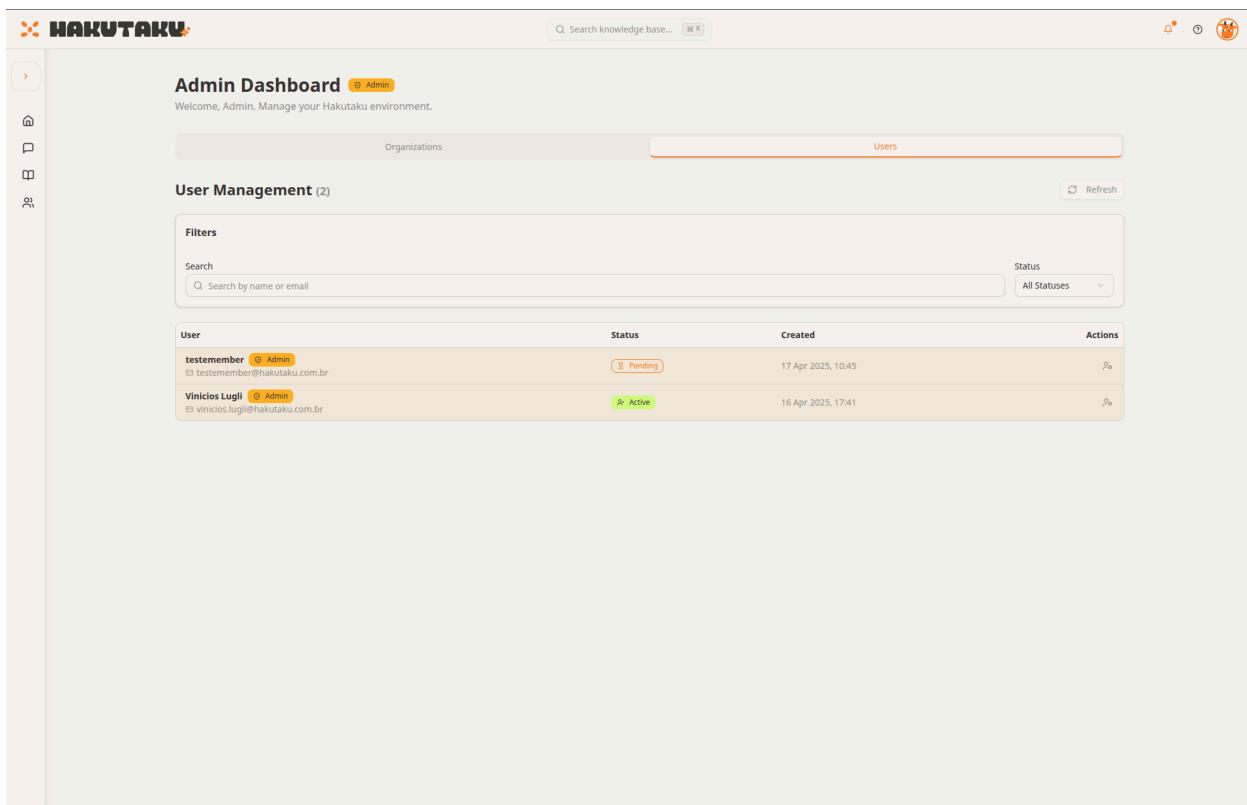
### 8.2.4.2. Use Cases

On Hakutaku's Administration screen, administrators can perform several crucial actions for the system's operation and security. Below are some of the main use cases:

#### **User management:**

The administrator can add new users to the platform, remove existing users, view the list of all registered users, and their contact information. This ensures that only authorized individuals have access to the system and that contact information is always up to date.

[Figure 17: User administration screen]



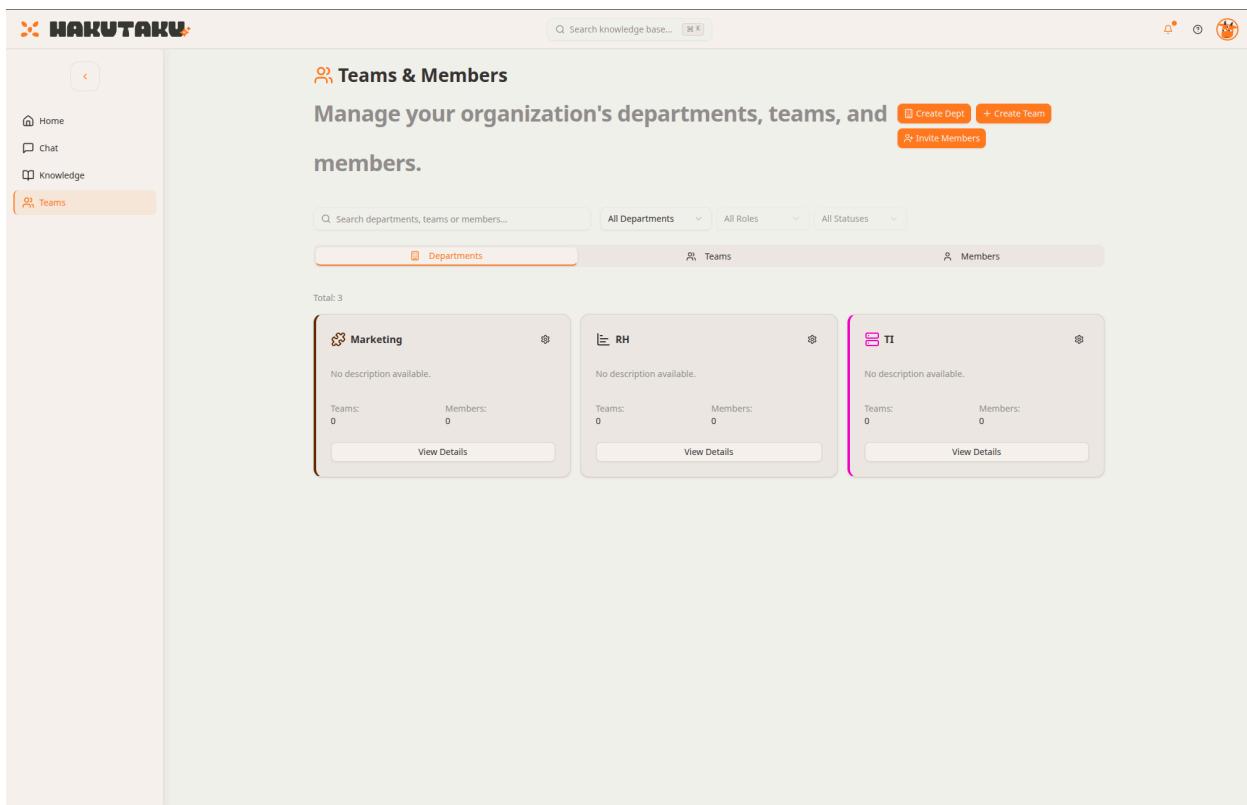
User	Status	Created	Actions
testemember (Admin) testemember@hakutaku.com.br	Pending	17 Apr 2025, 10:45	
Vinicio Lugli (Admin) vinicios.lugli@hakutaku.com.br	Active	16 Apr 2025, 17:41	

Source: Prepared by the authors, 2025.

### Management of roles and permissions:

The administrator can create, edit, and delete roles with different access levels. It is possible to assign one or more roles to each user, defining which data and functionalities each user can access. This ensures that data access is controlled and that each user has only the necessary permissions to perform their tasks.

[Figure 18: Team administration screen]



Source: Prepared by the authors, 2025.

### **Defining access to documents and data connectors:**

The administrator can specify which roles have permission to access certain documents and data connectors. This allows for control over access to sensitive information and ensures that data is accessed only by those who should view it.

### **Platform monitoring:**

The administrator has access to a dashboard with platform usage metrics, such as the number of active users, queries made, system response time, and error logs. This allows for monitoring the system's health, identifying problems, and having greater control over platform usage.

---

These use cases demonstrate how Hakutaku's Administration screen is essential to ensure the solution's efficient and secure operation.

## 9. Landing Page of the Solution

### 9.1. Goal

The **Hakutaku landing page** was created with the main objective of introducing the solution to the public, serving as a strategic step to validate the project's value proposition. By showcasing the main benefits and features of the platform, such as knowledge organization and the use of artificial intelligence to optimize internal processes and facilitate the search for documents and information, the page aims to generate interest from companies experiencing these challenges. Thus, the page plays a key role not only in generating visibility but also in validating whether the identified problem truly exists in the market.

It is worth noting that in addition to communicating what Hakutaku is, how it works, and how it can be used, the landing page was also designed to capture leads through a form. This form plays a fundamental role, allowing interested people and companies to submit their contact details, thereby making it easier to establish an initial relationship with potential clients. Therefore, the landing page is an essential tool both for validating the proposal's adherence and for kickstarting the lead-capture process for the company.

### 9.2. Technologies Used

---

Hakutaku's landing page was predominantly developed with the **Next.js framework**, chosen for being a modern technology that delivers high performance and supports both frontend and backend functionality in a unified environment. This choice aims to simplify development, scaling, and maintenance of the solution over time.

The application is hosted on **Amazon Web Services (AWS)**, a widely adopted and recognized infrastructure platform for its reliability and scalability. Additionally, Cloudflare has been integrated as the network provider to enhance stability and security.

The entire structure of the solution, as well as the source code and other technical components, is documented in our private repository: **2025-1A-T02-G70-INTERNO**.

The developed landing page can be accessed at the following address:

<https://hakutaku.com.br/pt-BR>.

## 9.3. Content

The content of the Hakutaku landing page has been carefully structured to communicate the solution's proposal in a clear, efficient, and attractive way. Each section was designed to guide the visitor from understanding the solution and its benefits to signing up as a potential customer. The main sections of the page and their roles within this flow are described below.

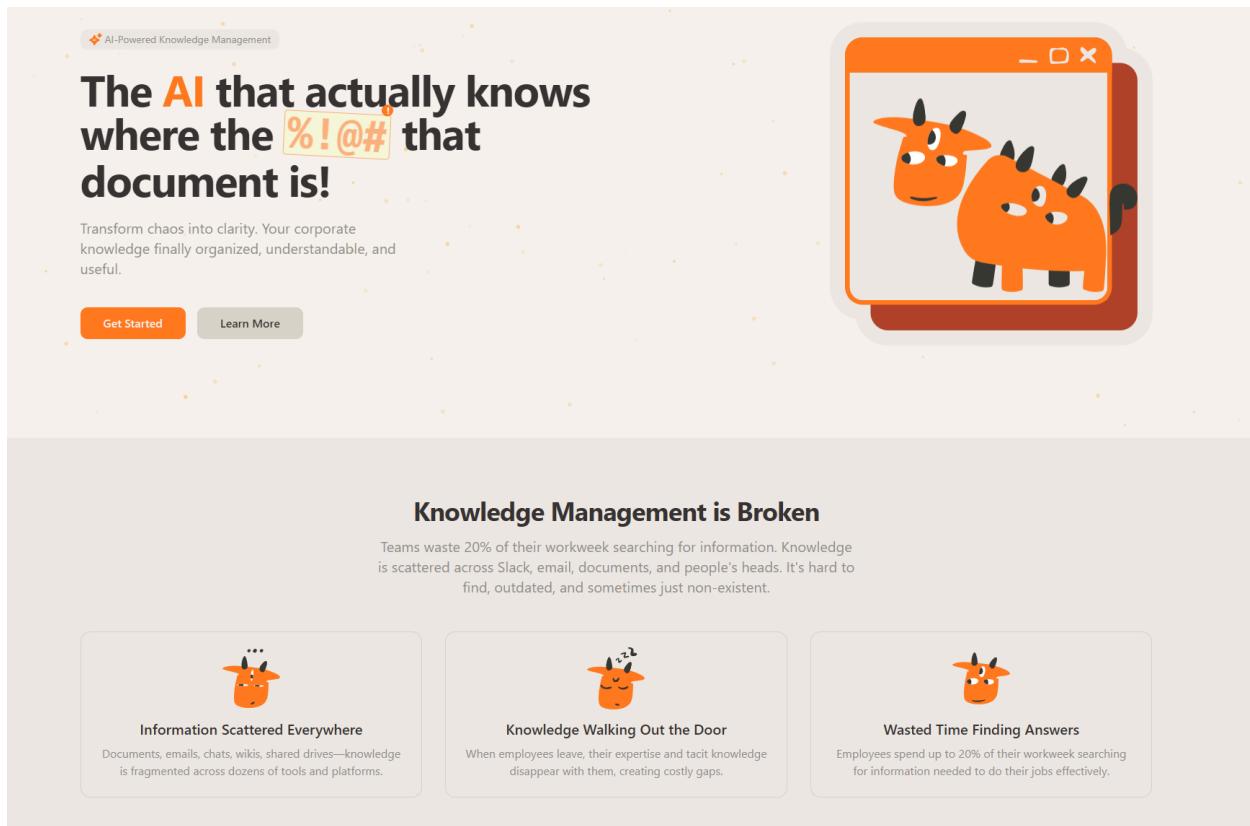
### 9.3.1 Introduction

The Introduction section aims to provide context to the visitor, explaining the problem that Hakutaku seeks to solve and addressing numerous difficulties

companies commonly face, such as disorganization of documents, wasted time on manual tasks, and difficulty accessing information. The objective is to help the target audience identify their own struggles and direct their attention toward the solution presented.

The main goal is to validate the actual existence of the problem and prepare the ground for Hakutaku to be recognized as the solution.

[Figura 19: Initial section of the landing page]



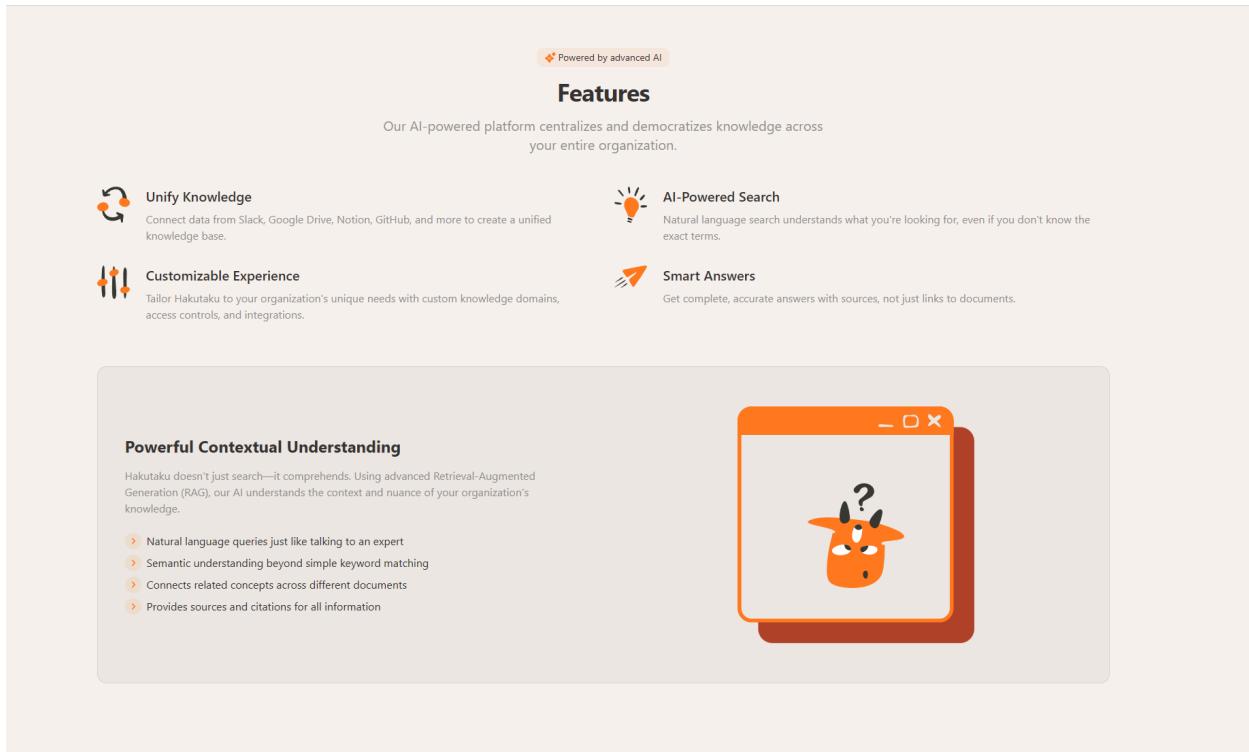
Source: Prepared by the authors, 2025.

### 9.3.2 Features

The Features section of the page highlights the main features offered by Hakutaku, such as document organization, smart file search through a chatbot, integration with other tools, and more. These features are described clearly, specifying both the technologies used and their respective use cases.

The main objective here is to demonstrate the technical value of the solution by showing how it can be applied in practice to solve previously presented problems.

[Figura 20: Functionalities and technologies Part 1]



The screenshot displays the 'Features' section of the Hakutaku platform. At the top, there is a header with the text 'Powered by advanced AI' and a small orange icon. Below the header, the word 'Features' is centered in bold. A descriptive paragraph follows, stating: 'Our AI-powered platform centralizes and democratizes knowledge across your entire organization.' Underneath this paragraph, four feature cards are listed:

- Unify Knowledge**: Represented by a circular icon with three dots. Description: 'Connect data from Slack, Google Drive, Notion, GitHub, and more to create a unified knowledge base.'
- AI-Powered Search**: Represented by a lightbulb icon. Description: 'Natural language search understands what you're looking for, even if you don't know the exact terms.'
- Customizable Experience**: Represented by a gear and wrench icon. Description: 'Tailor Hakutaku to your organization's unique needs with custom knowledge domains, access controls, and integrations.'
- Smart Answers**: Represented by a rocket ship icon. Description: 'Get complete, accurate answers with sources, not just links to documents.'

Below these cards, there is a section titled 'Powerful Contextual Understanding'. It includes a sub-section heading 'Hakutaku doesn't just search—it comprehends.' followed by the text: 'Using advanced Retrieval-Augmented Generation (RAG), our AI understands the context and nuance of your organization's knowledge.' To the right of this text is a small illustration of a cartoon character with a question mark on its head, enclosed in a window-like frame.

Source: Prepared by the authors, 2025.

[Figura 21: Functionalities and technologies Part 2]

## Academic Foundation

Hakutaku is built on solid academic research and innovative approaches to knowledge management.

 Academic Research

### Retrieval-Augmented Generation

Our platform is built on cutting-edge RAG techniques that combine the power of large language models with precise information retrieval from your organization's knowledge base.

- ✓ Semantic understanding beyond keyword matching
- ✓ Context-aware retrieval that understands user intent
- ✓ Transparent citations and evidence for all responses
- ✓ Tuned specifically for enterprise knowledge bases

Based on research methodologies from top academic institutions and our team's expertise in natural language processing.

 Platform Features

#### AI-Powered Knowledge Search

Intelligent search that understands context and intent, providing accurate answers from your organization's knowledge base with transparent citations.

#### Knowledge Base Management

Centralized repository for all organizational knowledge with collections, tags, and seamless document integration.

#### Conversational Interface

Natural chat interface with suggested prompts, message history, and feedback mechanisms to continuously improve responses.

#### Team & Department Segmentation

Organize knowledge by teams and departments with customized access controls, ensuring information is accessible to the right people at the right time.

[Explore all platform features >](#)



### Ongoing Research & Development

Hakutaku is continuously evolving through our R&D efforts. Our team of researchers and engineers are exploring new frontiers in:

<ul style="list-style-type: none"> <li>✓ Multi-modal knowledge representation</li> <li>✓ Collaborative AI knowledge systems</li> </ul>	<ul style="list-style-type: none"> <li>✓ Domain-specific language fine-tuning</li> <li>✓ Knowledge graph integration with LLMs</li> </ul>
--	---

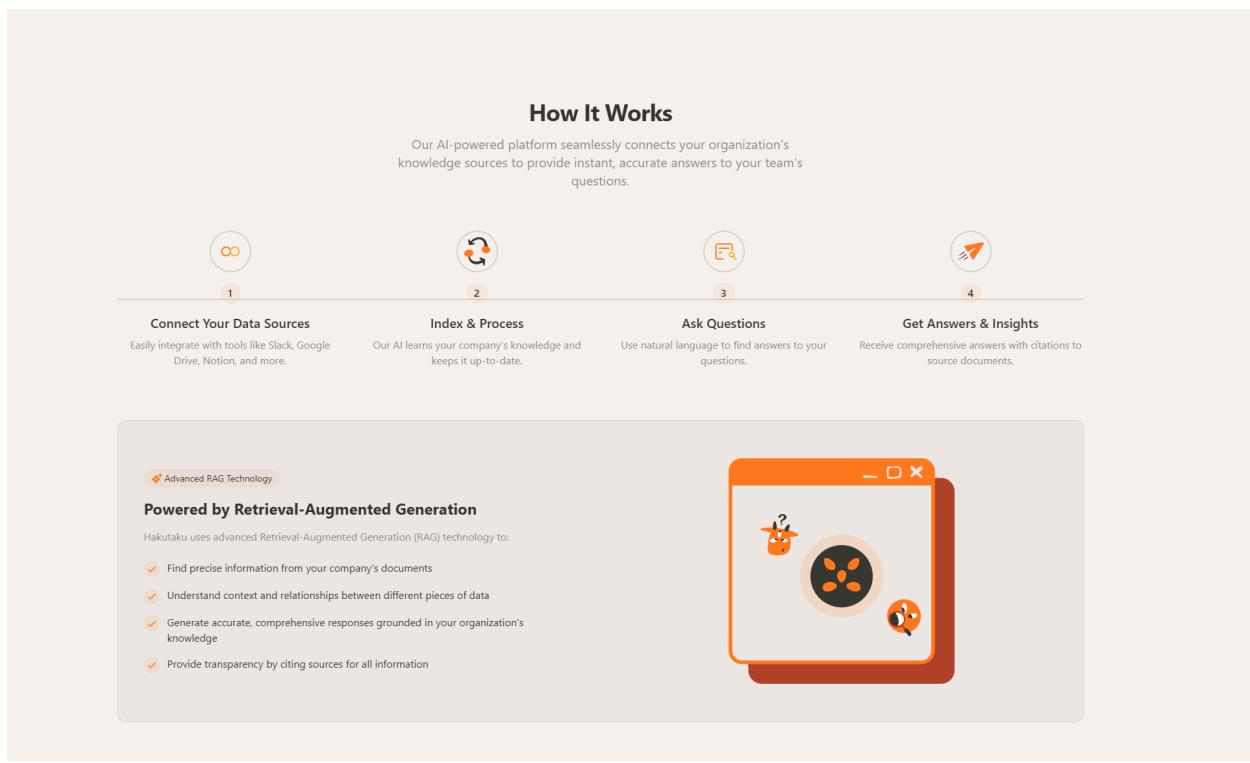
Source: Prepared by the authors, 2025.

### 9.3.3 How It Works

The How It Works section shows a simple, step-by-step view of how the platform is used. From creating connections to using its functionalities, the content is presented visually to aid understanding and help the visitor become more comfortable with the solution's workflow.

The main objective here is to reinforce the ease of use of Hakutaku, reducing entry barriers and doubts about the platform.

[Figura 22: Section explaining how the platform works]



Source: Prepared by the authors, 2025.

### 9.3.4 Benefits

The Benefits section aims to highlight the advantages that Hakutaku can bring to a company upon implementation. In addition to describing the platform's resources, this section underscores the concrete improvements it can generate in the daily operations of businesses, such as reducing time spent on manual tasks, increasing team efficiency, and improving agility in accessing information.

This is demonstrated through practical cases of how the solution can help improve a company's daily operations.

[Figura 23: Section with main benefits of the solution]

## Elevate Team Performance

Hakutaku leverages proven knowledge management principles to deliver tangible benefits, improving efficiency, collaboration, and knowledge retention across your enterprise.



### Save Time

Reduce time spent searching for information by up to 35%.<sup>[1]</sup>



### Faster Onboarding

Help new employees get up to speed 50% faster.<sup>[2]</sup>



### Mitigate Knowledge Loss

Preserve critical expertise when employees leave. Reducing knowledge loss significantly cuts hidden costs associated with turnover.<sup>[3]</sup>



### Better Decisions

Make more informed decisions with complete information.<sup>[4]</sup>

**Impact Backed by Research**

Studies and reports highlight the potential gains from implementing effective knowledge management:

- ✓ Significant productivity gains by reducing time spent searching for information (nearly 20% of workweek).<sup>[1]</sup>
- ✓ Accelerated onboarding through structured access to essential information, procedures, and best practices.<sup>[2]</sup>
- ✓ Mitigation of knowledge loss, a significant factor in employee replacement costs and lost productivity.<sup>[3]</sup>
- ✓ Improved quality and speed of decision-making, empirically linked to better organizational performance.<sup>[4]</sup>
- ✓ Enhanced innovation by connecting experts across departments and making collective intelligence readily accessible.<sup>[5]</sup>



Hover over the reference numbers **[x]** to view source details.

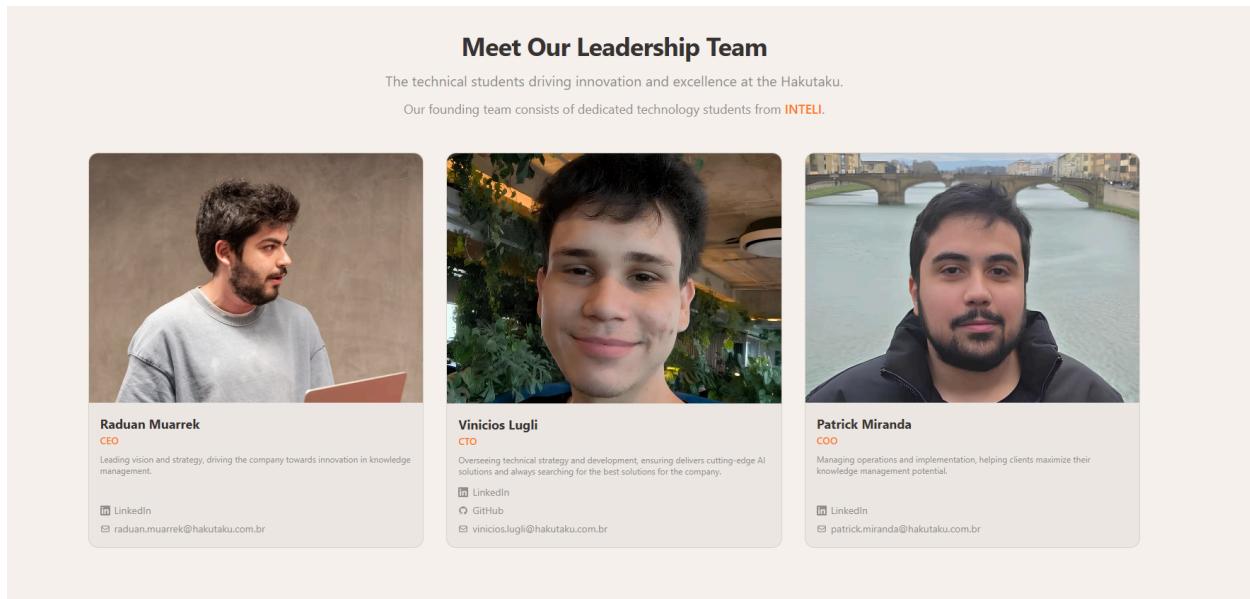
Source: Prepared by the authors, 2025.

### 9.3.5 Team

The Team section introduces the team responsible for developing the platform. It displays their information, roles, and professional experience.

The main objective is to show who is behind the project and how their skills help make it a good solution. This helps people trust the platform more.

[Figura 24: Section introducing the team]



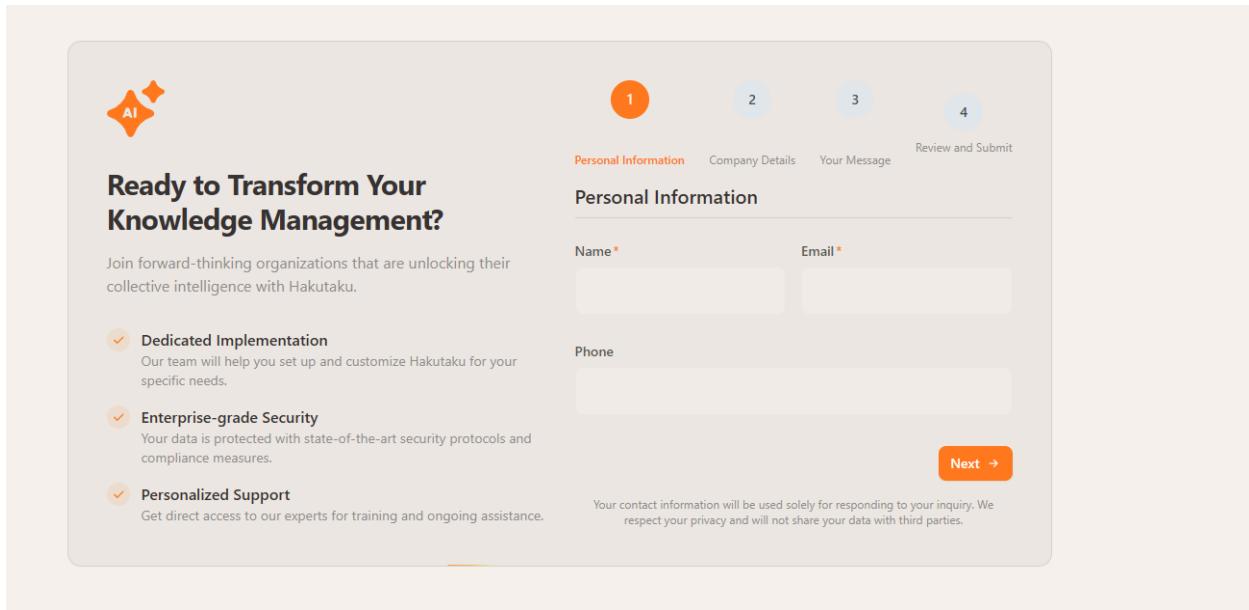
Source: Prepared by the authors, 2025.

### 9.3.6 Form

The final section of the page focuses on lead capture through a simple form where the visitor can fill in their name, email, and the number of employees in their company.

The main objective is to start building a base of interested leads, allowing the lead-capture process to start.

[Figura 25: Section with lead-capture form]



The screenshot shows a user interface for transforming knowledge management. At the top left is an orange diamond icon with 'AI' inside. To its right are four numbered circles (1, 2, 3, 4) corresponding to tabs: 'Personal Information' (orange), 'Company Details' (light blue), 'Your Message' (light blue), and 'Review and Submit' (light blue). Below the tabs is a section titled 'Ready to Transform Your Knowledge Management?'. It includes a sub-section 'Join forward-thinking organizations that are unlocking their collective intelligence with Hakutaku.' followed by three bullet points: 'Dedicated Implementation' (Our team will help you set up and customize Hakutaku for your specific needs.), 'Enterprise-grade Security' (Your data is protected with state-of-the-art security protocols and compliance measures.), and 'Personalized Support' (Get direct access to our experts for training and ongoing assistance.). A note at the bottom states: 'Your contact information will be used solely for responding to your inquiry. We respect your privacy and will not share your data with third parties.' An orange 'Next →' button is located on the right side of the form.

Source: Prepared by the authors, 2025.

## 10. Initial development of the solution (Chatbot)

### 10.1. Objective

During Module 2, the development of the chat interface was initiated, aiming to meet the previously defined requirements (as described in Topic 7.2).

This first delivery includes the essential functionalities to enable user interaction with the chatbot integrated into the Hakutaku platform.

This interaction is carried out through natural language, using a connection with a language model (LLM), currently utilizing Gemini for this functionality.

Context and conversation history retrieval features were also implemented, ensuring greater coherence in responses throughout the session.

The creation of this interface aims primarily to establish the foundation for the next stages of the platform, ensuring a functional chat that interacts with the user in an

---

intuitive and efficient way, thereby setting the stage for future integrations with external data sources through RAG (Retrieval-Augmented Generation), planned for the next module of the project.

## 10.2. Technologies used

Hakutaku's solution was developed using modern and widely adopted market technologies, with a focus on performance and easy maintenance, aiming to facilitate the addition of new functionalities and future support.

The main user communication interface is carried out through a chatbot that uses a language model (LLM) to interpret and respond to messages in natural language. Currently, the model in use is Gemini, integrated into the application via the LangChain library, responsible for orchestrating LLM calls and helping build more fluid conversational flows.

For data persistence, Prisma ORM was adopted, which facilitates communication with the relational PostgreSQL database, hosted on the Supabase provider.

This combination enables the storage of all necessary information for chat functionality, as well as user data and history.

Both the frontend and backend of the application are hosted on Amazon Web Services (AWS), ensuring reliability, scalability, and security for the solution's infrastructure.

Additionally, it is a platform widely used by developers, making future maintenance easier.

The complete solution code, including the entire application structure, backend, language model integration, database, and other technical components, is documented and versioned in the private repository: 2025-1A-T02-G70-INTERNO.

---

## 10.3. Screens

The Hakutaku platform interface was designed with a focus on simplicity and familiarity, adopting already established chatbot market standards. Navigation is organized into two main areas: the home screen and the conversation screen.

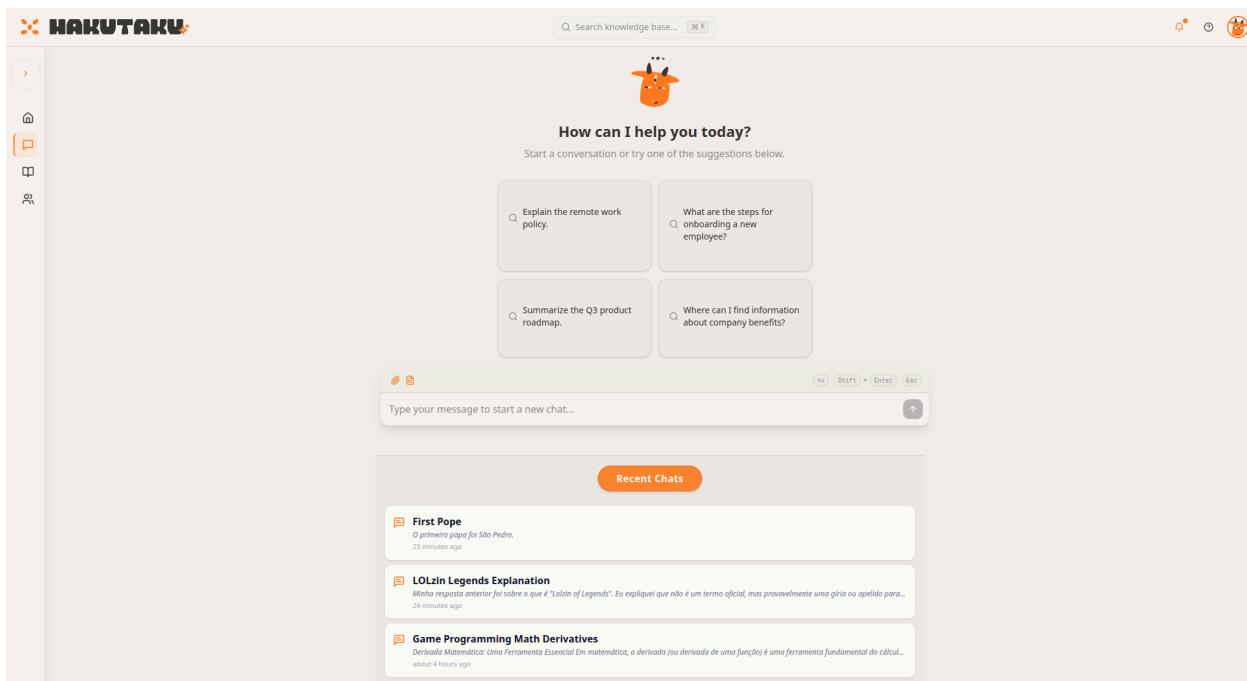
### 10.3.1 Home Screen

The main purpose of the home screen is to guide the user and facilitate the start of a new conversation with the chatbot. On this screen, the user finds:

- A text input field to start a new chat.
- Suggested questions (prompt tips and use cases) to guide tool usage, such as: “What is the remote work policy?” or “How does onboarding for a new employee work?”
- Recent interaction history, allowing the user to easily resume previous conversations.

These are the main elements of the home screen, whose objective is to direct the user to start a new chat or continue an existing conversation. The interface was designed to be simple and intuitive, reducing usage barriers and optimizing the experience from the first access.

[Figure 26: Hakutaku chat interface home screen]



Source: Prepared by the authors, 2025.

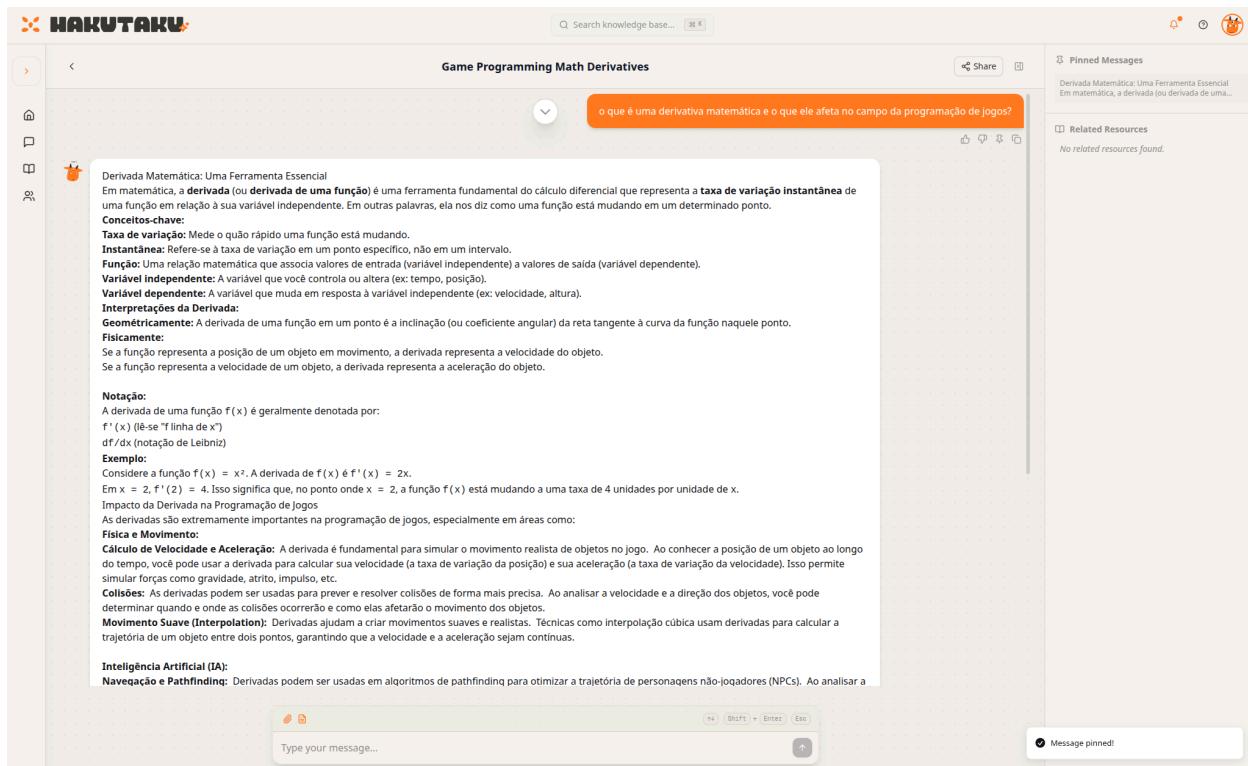
### 10.3.2 Conversation Screen

The main conversation screen was inspired by modern LLM and chatbot interfaces, such as ChatGPT, with the goal of providing ease of use and generating familiarity with tools already established in the market. With this in mind, the main elements in this section of the platform are:

- A centralized chat panel where user messages are displayed in orange and system responses in white, facilitating distinction between questions and answers.
- A fixed text input field at the bottom of the screen, with keyboard shortcut support.
- A system of pinned messages located on the right-hand side, allowing the user to mark and access important responses for consultation at any time.

The design was created to reinforce the sense of a natural and productive dialogue environment, where company knowledge can be accessed quickly and contextually.

[Figure 27: Main screen for interacting with the Hakutaku chatbot]



Source: Prepared by the authors, 2025.

## 10.4. Functionalities

The Hakutaku platform offers a set of essential functionalities to enable and optimize access to internal company knowledge. All features were designed with a focus on usability and employee productivity, addressing key challenges of organization, access, and reuse of corporate information.

Below are the main functionalities currently implemented in the platform:

### **Chat History Creation and Retrieval:**

Each chat session is recorded and can be resumed at any time.

The system maintains a complete history of conversations stored in our PostgreSQL database, allowing users to revisit previous responses, resume ongoing discussions, and preserve interaction context.

### **Suggested Questions (Prompts):**

When accessing the chat screen, the user is greeted with predefined question suggestions, based on recurring topics in the corporate environment. These tips serve as a starting point, facilitating initial use of the tool and highlighting relevant use cases.

### **Natural Language Conversation with LLM:**

The core of the interaction is carried out through a natural language conversation, using a language model (LLM) to interpret and respond to user messages intelligently and contextually. Currently, the model used is Gemini, enabling fluid and accurate conversational flows.

### **Pinned Messages:**

The interface allows important messages to be pinned, facilitating organization and quick access to critical information. These messages are available in a fixed side section, reinforcing the idea of a platform that not only responds but also helps organize knowledge.

These functionalities aim to ensure Hakutaku delivers value, turning an organization's internal knowledge into an accessible, structured, and high-value resource for employees.

## documentationSources:

- Notions - Episode 6: Supercharging company knowledge with AI. *Available at:* <https://info.notion.so/webinar/webinars-amer/notions--episode-6-the-future-of-organisational-knowledge>. Accessed on: Mar. 1, 2025.
- POPADIUK, S.; SANTOS, A. E. M. DOS. Conhecimentos Tácito, Explícito e Cultural no Planejamento da Demanda. JISTEM Journal of Information Systems and Technology Management, v. 7, n. 1, p. 207–226, Apr. 30, 2010.
- CHATGPT. Assistente de Escrita e Revisão de Texto. OpenAI, 2025. Ferramenta utilizada para correções ortográficas e coesão textual no desenvolvimento da documentação do projeto. *Available at:* <https://chatgpt.com/>.
- Glean | Work assistant & knowledge management platform. *Available at:* <https://www.glean.com/>.
- Lista de Empresas com mais de 50 funcionários no Brasil - Econodata. *Available at:* <https://www.econodata.com.br/empresas/todo-brasil/mais-de-50-funcionario-s>. Accessed on: Mar. 26, 2025.

- 
- IBGE | Brasil em síntese | comércio | número de empresas, por faixas de pessoal ocupado. *Available at:* <https://brasilemsintese.ibge.gov.br/comercio/numero-de-empresas-por-faixas-de-pessoal-ocupado.html>.
  - Cidade de São Paulo concentra 10,3% do PIB do país em 2019 | Agência de Notícias. *Available at:* <https://agenciadenoticias.ibge.gov.br/agencia-sala-de-imprensa/2013-agencia-de-noticias/releases/32596-cidade-de-sao-paulo-concentra-10-3-do-pib-do-pais-em-2019>.