

AIKO (AI-POWERED KNOWLEDGE ORGANIZER)
A SYNOPSIS
ON
MEGA PROJECT
SUBMITTED IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE AWARD OF THE DEGREE
OF
BACHELOR OF TECHNOLOGY
IN
COMPUTER SCIENCE AND ENGINEERING

Submitted By

Kaustubh Warade
CSE/RN/23
687545

Aditya Deshmukh
CSE/RN/5
687505

Devansh Parapalli
CSE/RN/13
687506

Yashasvi Thool
CSE/RN/77
687548



DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING
GOVERNMENT COLLEGE OF ENGINEERING, NAGPUR

July 2024

1. INTRODUCTION

The exponential growth of digital information across platforms has created challenges in knowledge management and retrieval. Artificial Intelligence-powered Knowledge Organizer(AIKO) will address these issues through an innovative approach leveraging advanced AI and data storage techniques. Core principles of AIKO include:

1.1. Intelligent Knowledge Integration

AIKO will employ sophisticated AI algorithms to integrate diverse information sources into a cohesive knowledge bank. It will utilize NLP and machine learning to analyze, categorize, and organize information, facilitating novel retrieval techniques.

1.2. Multimodal Information Processing

AIKO will ingest, analyze, and synthesize information from various content types, including textual documents, audio, video, and structured/unstructured datasets. AIKO's multi-modal approach enables comprehensive knowledge capture and cross-modal information retrieval.

1.3. Modular Architecture

AIKO's modular design incorporates multiple ingress and egress points, enhancing extensibility and adaptability. Clear interfaces allow for easy modification as requirements evolve or new technologies emerge.

1.4. Transparent AI Systems

AIKO prioritizes transparency and interpretability in its ML components, incorporating explainable AI (XAI)^[17] techniques.

In summary, AIKO will offer a pragmatic solution to current information management challenges and serve as a paradigm for developing transparent, adaptive, and intelligent knowledge management systems.

2. BRIEF LITERATURE SURVEY

The following section provides an overview of content organization, generation, and retrieval platforms, analyzing capabilities and limitations in modern knowledge management systems.

2.1. mem.ai^[1]

mem.ai focuses on content organization and generation within its own ecosystem.

- Exclusive support for files/documents created within mem's ecosystem
- No support for importing external documents
- Emphasis on content organization and generation capabilities

The closed ecosystem limits versatility but potentially allows for tighter integration of core functionalities.

2.2. Brain Assistant^[2]

Brain Assistant is a browser-centric approach to generative AI, built on open-source models.

- File processing capabilities
- Browser-only accessibility
- Support limited to text-based files
- Manual file upload required

2.3. AI Brain Bank^[3]

AI Brain Bank provides a straightforward retrieval solution using open-source models for content embedding.

- Simple design and functionality
- Open-source models for content embedding
- No support for additional communication channels
- No content generation capabilities

2.4. iWeaver^[13]

iWeaver focuses on website link analysis using OpenGraph for similarity scores.

-
- Works exclusively with website links
 - Uses OpenGraph for similarity score generation
 - Paid support for video/audio processing
 - Uses public models without user-selectable options

2.5. keepi.ai^[4]

keepi.ai offers a diverse range of supported content types.

- Supports URLs, text documents, and images
- No support for video and audio content

3. PROBLEM FORMULATION

In today's digital age, individuals and organizations grapple with an overwhelming influx of information spread across various platforms such as emails, social media, cloud storage, and more. The fragmented nature of available information makes it challenging to efficiently manage, access, and utilize knowledge. This leads to decreased productivity, missed opportunities, and increased cognitive load. A non-exhaustive list of problems is provided below:

- **Information Overload:** The vast amount of data generated daily makes it difficult for users to keep track of important information. New information overshadows the older ones, leading to increased cognitive load.
- **Fragmented Data:** Information is scattered across various platforms, devices and formats, making it hard to efficiently retrieve any content.
- **Inefficiencies in Information Retrieval:** Finding specific information can be time-consuming, especially when dealing with large volumes of data.
- **Relevance and Accuracy:** Most information is stored in an unstructured manner, making it a challenge to determine the relevance with any given query.
- **Contextual Understanding:** Most existing search mechanisms are often inadequate^[16] and operate on a one-and-done methodology, lacking the

ability to understand the context within and across pieces of information.

- Lack of Personalization: Users have unique needs and preferences for organizing and accessing information, support for which is sorely lacking in current solutions.

4. OBJECTIVES

AIKO aims to develop an advanced information storage and retrieval system, addressing challenges in managing and accessing diverse, large-scale data.

4.1. Data Integration

- Implement unified data layer for heterogeneous sources
- Develop APIs for seamless data ingestion
- Create centralized hub for consolidated information

4.2. Advanced Search Functionality

- Develop high-performance search engine
- Implement advanced indexing and NLP for semantic search
- Enable efficient cross-platform information retrieval

4.3. Security Implementation

- Design multi-layered security architecture
- Implement encryption at rest and in transit
- Integrate OAuth 2.0^[9] and SAML for authentication/authorization

4.4. Automated Synchronization

- Design publish-subscribe system for real-time updates
- Implement conflict resolution algorithms for multi-source updates
- Ensure automatic data consistency across platforms

4.5. Intuitive User Experience

- Create cross-platform accessible frontend
- Implement collaborative filtering and user preference customization

-
- Enhance user engagement through personalized interfaces and information feeds

5. METHODOLOGY

AIKO implementation follows the Waterfall^[15] SDLC model, divided into backend and frontend segments.

5.1. Backend

- Components: Data layer, NLP-enabled content processor, high-performance search engine
- Design focus: Large-scale data handling, cross-platform integration, information retrieval
- Security: Advanced encryption, robust authentication mechanisms
- Technologies:
 - NLP & Deep Learning: NLTK^[14], TensorFlow^[12, 11]
 - LLMs^[18], Hugging Face Transformers^[19], API integration
 - Authentication: OAuth 2.0, Supabase Auth (server-side sync)
 - Database: PostgreSQL^[10] with Supabase Realtime^[5]

5.2. Frontend

- Design principles: Cross-platform accessibility, user-centric interface, seamless information interaction
- Key features: Customizable information flow, adaptive user preferences
- Technologies:
 - Framework: SvelteKit^[6]
 - Styling: Tailwind CSS^[7]
 - Deployment: Vercel^[8]
 - Enhanced functionality: Progressive Web Application (PWA) for offline access

6. DELIVERABLES

The deliverables for AIKO are summarized in the following points:

- Frontend Interface: A user-friendly interface accessible across all platforms, providing a seamless user experience.
- Backend System: A robust backend system capable of being scaled horizontally and deployed on Open Container Interface-compliant Cloud Platforms.
- AI, ML, DL, and GenAI Models: A suite of AI, ML, DL, and Generative AI models for content processing, semantic search, and information retrieval.
- API & Platform Documentation: Comprehensive API and Platform Documentation for developers and users to understand the system and its capabilities.
- Self-hostable Instance of AIKO: A self-hostable instance of AIKO for organizations and individuals to deploy on their own infrastructure.

7. PROPOSED TIMELINE

AIKO will be developed using Waterfall Model of SDLC. The proposed timeline is given below:

- Feasibility Study: 2 weeks
- Design Phase: 2 weeks
- Development Phase: 8 weeks
- Testing Phase: 2 weeks
- Maintenance Phase: 15th week onwards

8. BUDGET

AIKO will be built using open-source technologies and frameworks. The budget for the project is estimated as follows:

- Machine Learning Cloud: ₹5,000/month
- Domain and Hosting: ₹2,000/year
- API Access for GenAI models: ₹2,000/month

Total estimated budget for the project is ₹7,500/month.

9. FACILITIES REQUIRED

The following facilities are required for the successful development and deployment of AIKO:

- High-performance computing resources, possibly with GPU acceleration, for training complex models and handling large datasets.
- A secure and scalable data storage system for managing and storing research data.
- Access to cloud computing platforms for deployment and hosting of AIKO's backend and frontend components.

10. PROPOSED OUTCOMES

The successful development and deployment of AIKO are expected to yield the following outcomes:

10.1. Unified Personalized Knowledge Repository

A centralized, organized database of personalized information accessible across platforms, enabling users to efficiently manage and retrieve data.

10.2. Enhanced Information Retrieval

Advanced search capabilities leveraging NLP and semantic search techniques, enabling users to quickly locate relevant information across AIKO.

10.3. Better Knowledge Management

Improved knowledge management through automated synchronization, conflict resolution, and real-time updates, reducing manual data management efforts.

10.4. User Empowerment

Users will be able to change models, algorithms, and data sources as per their requirements, empowering them to customize their information flow.

REFERENCES

- [1] Mem - the AI notes app that keeps you organized. Retrieved July 14, 2024 from <https://get.mem.ai/>
- [2] Brain assistant. Retrieved July 14, 2024 from <https://mybrain.zone/dashboard>
- [3] AI Brain Bank. Retrieved July 14, 2024 from <https://aibrainbank.com/>
- [4] Personal knowledge AI | Keepi. Retrieved July 14, 2024 from <https://www.keepi.ai/>
- [5] Supabase | the open source Firebase alternative. Retrieved July 14, 2024 from <https://supabase.com/>
- [6] SvelteKit. Retrieved July 14, 2024 from <https://kit.svelte.dev/>
- [7] Tailwind CSS - Rapidly build modern websites without ever leaving your HTML. Retrieved July 14, 2024 from <https://tailwindcss.com/>
- [8] Vercel: Build and deploy the best web experiences with the Frontend Cloud. Retrieved July 14, 2024 from <https://vercel.com/>
- [9] 2012. *The OAuth 2.0 Authorization Framework*. <https://doi.org/10.17487/rfc6749>
- [10] 2024. PostgreSQL. Retrieved July 14, 2024 from <https://www.postgresql.org/>
- [11] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Rafal Jozefowicz, Yangqing Jia, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dan Mané, Mike Schuster, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. 2015. TensorFlow, Large-scale machine learning on heterogeneous systems. <https://doi.org/10.5281/zenodo.4724125>
- [12] Martín Abadi, Paul Barham, Jianmin Chen, Zhifeng Chen, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat, Geoffrey

-
- Irving, Michael Isard, Manjunath Kudlur, Josh Levenberg, Rajat Monga, Sherry Moore, Derek G. Murray, Benoit Steiner, Paul Tucker, Vijay Vasudevan, Pete Warden, Martin Wicke, Yuan Yu, and Xiaoqiang Zheng. 2016. TensorFlow: a system for large-scale machine learning. *Operating Systems Design and Implementation* (2016), 265–283. <https://doi.org/10.5555/3026877.3026899>
- [13] iWeaver AI. 2024. Remember Recall Reuse your knowledge | iWeaver AI Memory Tool. Retrieved July 14, 2024 from <https://www.iweaver.ai/>
- [14] Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural Language Processing with Python*.
- [15] Kai Petersen, Claes Wohlin, and Dejan Baca. 2009. *The waterfall model in Large-Scale development*. https://doi.org/10.1007/978-3-642-02152-7_29
- [16] Nicholas J. Radcliffe and Patrick D. Surry. 1995. *Fundamental limitations on search algorithms: Evolutionary computing in perspective*. <https://doi.org/10.1007/bfb0015249>
- [17] Wojciech Samek, Grégoire Montavon, Andrea Vedaldi, Lars Kai Hansen, and Klaus-Robert Müller. 2019. *Explainable AI: Interpreting, explaining and visualizing deep learning*. <https://doi.org/10.1007/978-3-030-28954-6>
- [18] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. 2017. Attention is All you Need. *arXiv (Cornell University)* 30, (2017), 5998–6008. Retrieved July 14, 2024 from <https://arxiv.org/pdf/1706.03762v5>
- [19] Thomas Wolf, Lysandre Debut, Victor Sanh, Julien Chaumond, Clement Delangue, Anthony Moi, Pierric Cistac, Tim Rault, Rémi Louf, Morgan Funtowicz, Joe Davison, Sam Shleifer, Patrick von Platen, Clara Ma, Yacine Jernite, Julien Plu, Canwen Xu, Teven Le Scao, Sylvain Gugger, Mariama Drame, Quentin Lhoest, and Alexander M. Rush. 2020. HuggingFace's Transformers: State-of-the-art Natural Language Processing. Retrieved July 14, 2024 from <https://arxiv.org/abs/1910.03771>