

Convera AI

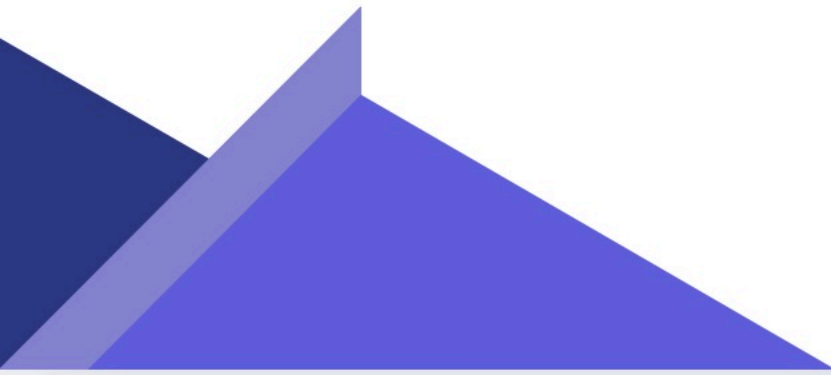
Made by -
Nehal Singh-E23CSEU0852
Group - 8, Batch - 29

Submitted to -
Dr. Nitin Arvind Shelke

CSET - 301
Artificial Intelligence and Machine Learning


Problem Statement

Even though low-code AI platforms are becoming more popular, non-technical users still struggle to prepare the structured and personalised data needed for the AI to perform effectively. Most current tools rely on manual input without offering proper guidance, which makes the process confusing and inefficient. As a result, the AI agents often don't align well with the users' actual goals or workflows—especially during the critical early phase of exploring and understanding the data, known as Exploratory Data Analysis (EDA).



Complex UI OF Existing Platforms

Metadata

 This feature is released to all users for a limited time during beta.

Define metadata that your agent can extract autonomously so you can filter, search and analyse tasks

Try some examples: ☒ Prospect replied ☒ Meeting booked ☒ # Lead score ☒ Prospect Industry


Metadata	Type	Method
----------	------	--------

No data...

Add new metadata

Override the system prompt for running the metadata tool

RELEVANCE AI



Data Pool


Add Data Source

×

Label your data source * (required)

Type data source name

Add File (.xml, .docx, .pdf)


Drag and drop your file here or Click to upload.
XML, DOCX or PDF (MAX. 800×400px)

OR

Enter URL * (sitemap or website URL)

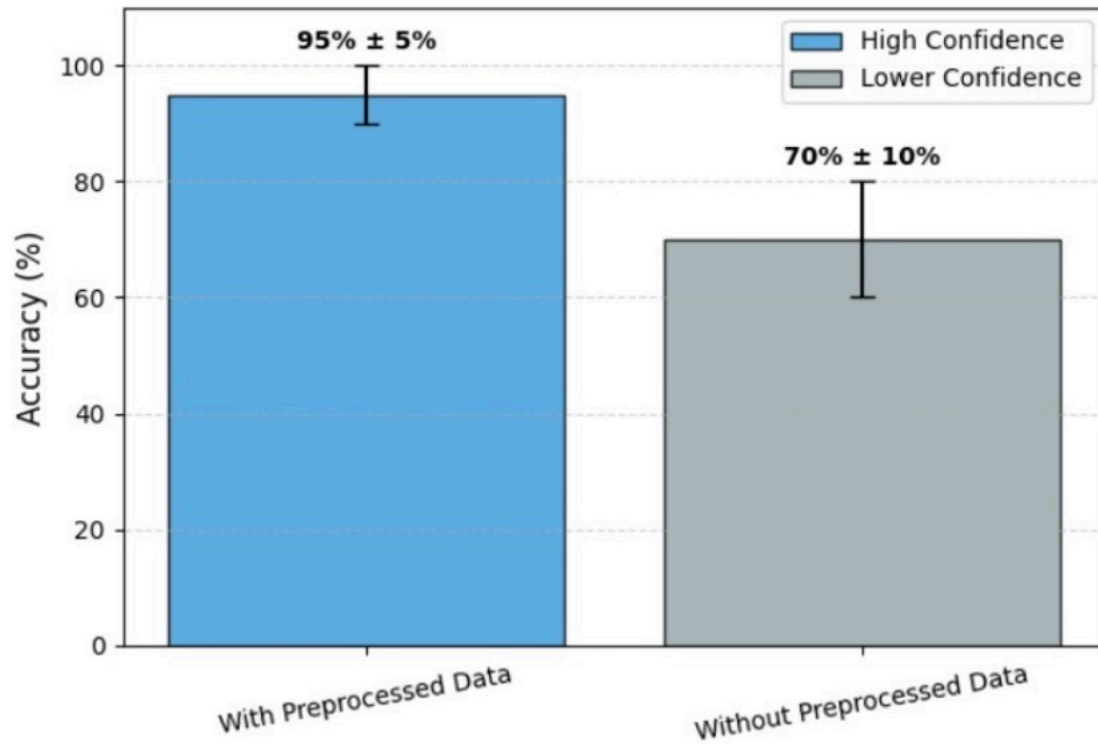
Type data source URL

Add Data Source

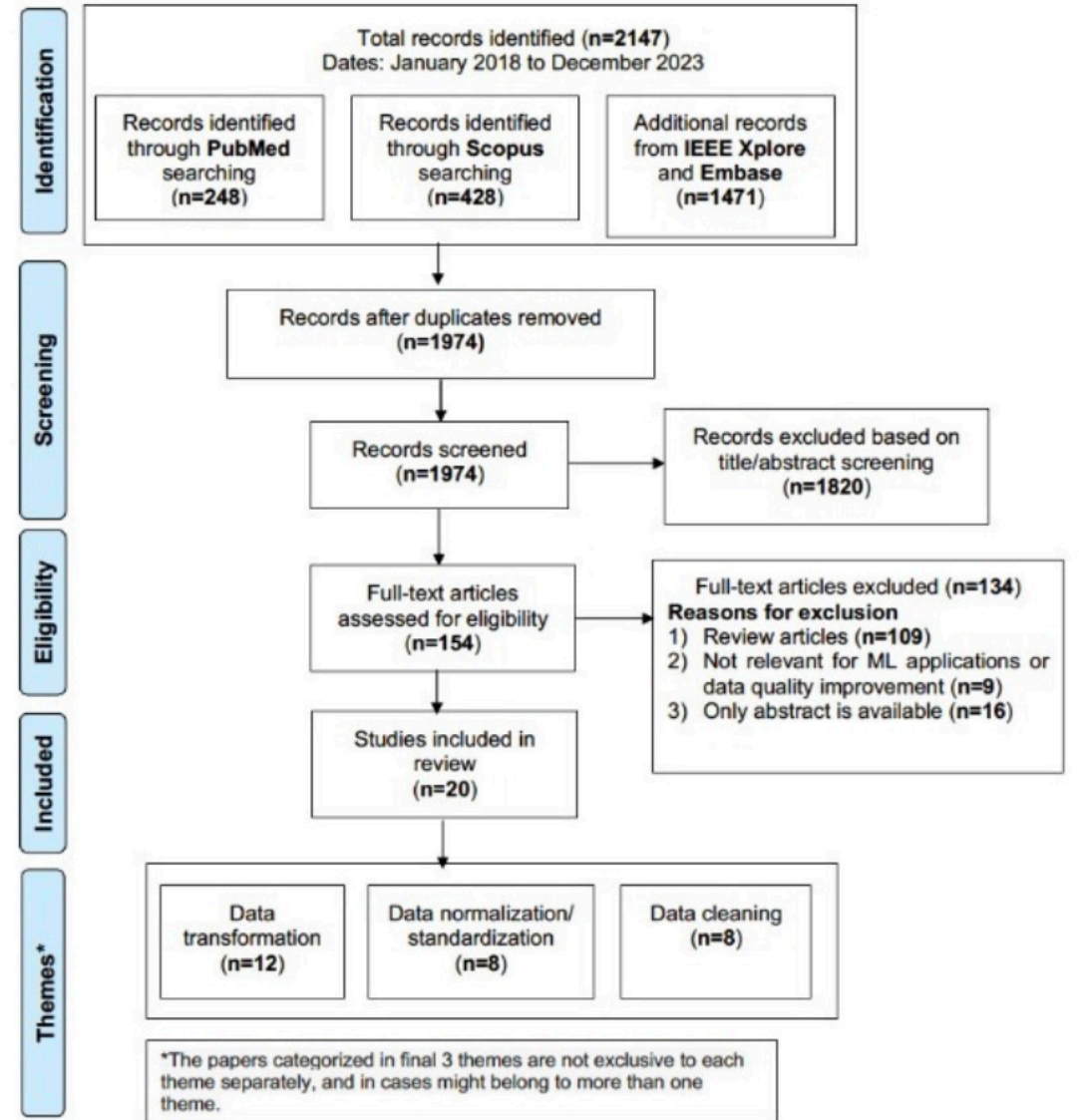
SMYTHOS

IMPACT OF DATA PRE-PROCESSING

Accuracy of Assistant AI: Preprocessed vs. Unprocessed Data



HOW ITS DONE



Key Challenges

Companies that succeed with modern generative AI tools—such as [Microsoft's MSFT -3.02% ▼](#) Copilot, [Salesforce's CRM -2.82% ▼](#) Agentforce, or offerings from a [raft of new startups](#)—are discovering that in order to get real value from AI, they have to organize their data in ways they might not have before. And this isn't a one-and-done effort. To keep their shiny new AIs up-to-date, the information they feed them must be kept constantly updated—creating more work for humans.

Every company needs a 'knowledge base'

[link](#)

Every company I talked with mentioned that to get real value out of their shiny new generative AI systems—no matter the application—they needed to overhaul or double down on their strategy for feeding it the kind of data that today's AI excels at processing—"unstructured" data.

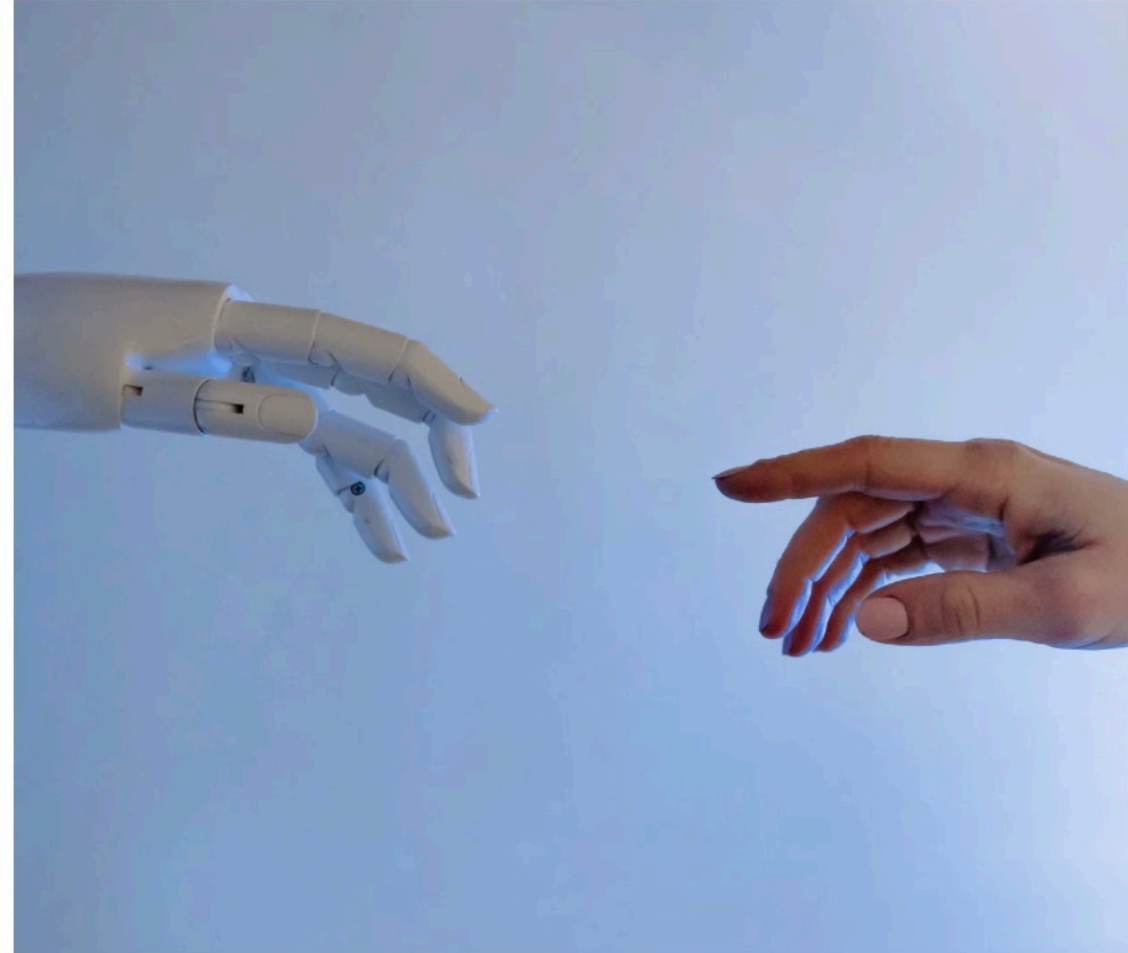
After prompting your LLM and receiving a response back in plain text. Quite straightforward for simple exchanges. It is not very practical for more complex applications and tasks. What we will need is some form of structured output. A common practice within the community and among various tools is to instruct the LLM to output in a JSON format, with specified keys where values can be parsed. These instructions are specified in your prompt. Here is an example from LangChain [5]. It doesn't have to be in JSON format, it can be even a simpler structure that you may want to engineer that can be parsed with Regex.

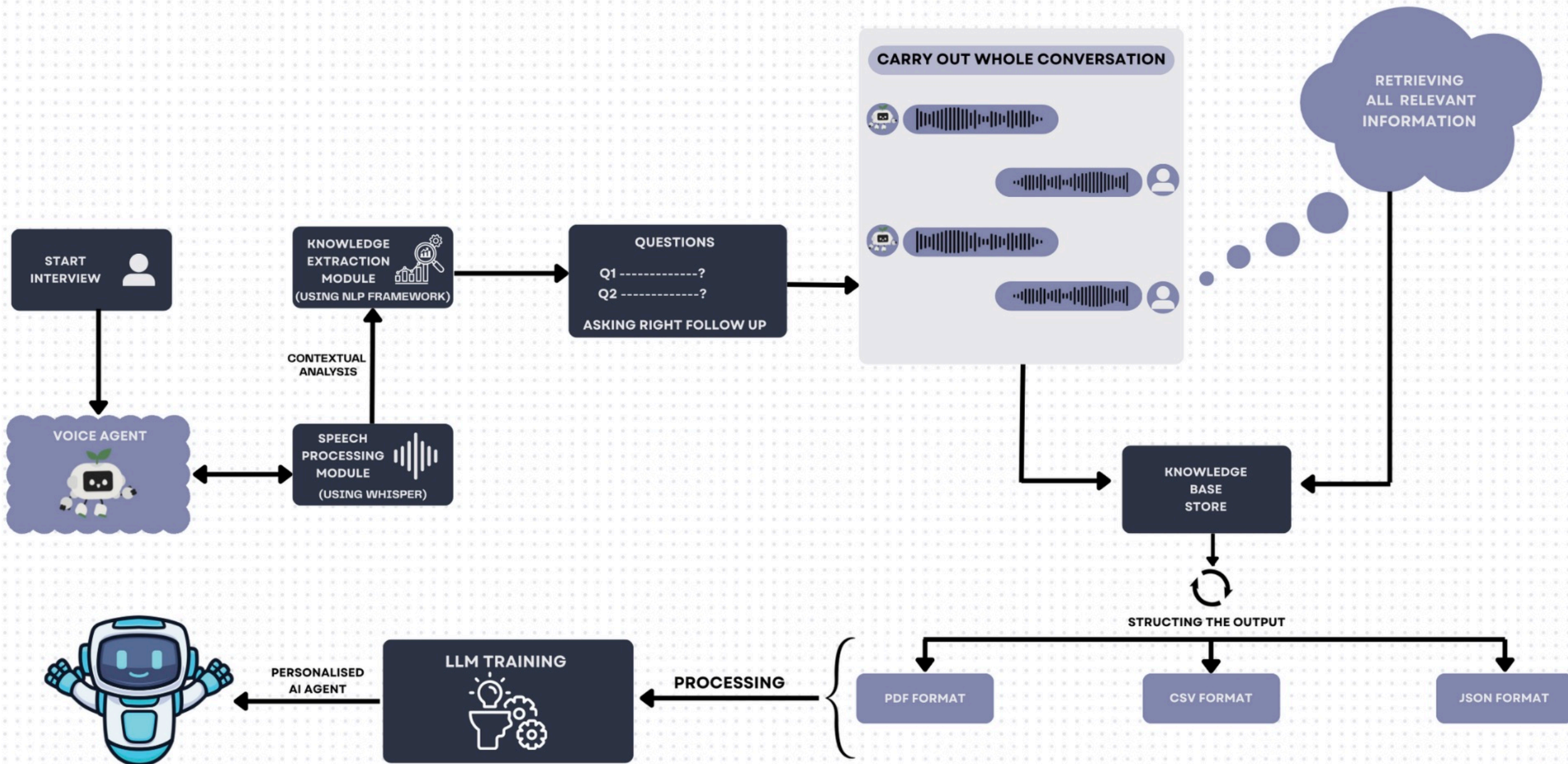
[link](#)

- Data Collection Complexity
- Lack of Guided Interaction
- Data Structuring & Contextualisation
- Misalignment Between User Intent and AI Interpretation
- Limited Accessibility of Existing Tools

Solution

- We propose a voice-first conversational engine that guides users through intelligent, adaptive interviews.
- The system asks context-aware questions to extract relevant goals, workflows, and domain knowledge.
- It dynamically adjusts its dialogue based on real-time responses for a personalized experience.
- Collected information is automatically structured into clean, usable knowledge bases.
- These outputs are optimized for direct use in low-code AI platforms, simplifying agent creation.





TECH STACK

FRONT-END



HTML

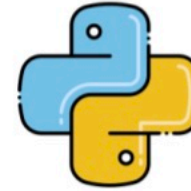


CSS



JAVASCRIPT

BACK-END



Python



LLM



Fast API

Impact and benefits

- **Empowers Non-Technical Users**

Makes AI agent creation accessible to domain experts without requiring coding or data science skills.
Enables intuitive voice-based interaction to gather and organise knowledge effortlessly.

- **Bridges the Data-to-Agent Gap**

Automates the Exploratory Data Analysis (EDA) phase through guided conversational input.
Reduces friction in preparing domain-specific data for AI agent deployment.

- **Personalised, High-Quality Agents**

Ensures agents are context-aware, aligned with real-world workflows, goals, and constraints.
Minimises misalignment between agent behaviour and user expectations.

- **Accelerates AI Adoption**

Drastically reduces time from ideation to deployment of intelligent agents.
Facilitates smoother onboarding and faster prototyping for AI solutions.

- **Scalable Across Domains**

Applicable to a wide range of industries including healthcare, education, customer support, and more.
Adapts to diverse terminologies and workflows through intelligent, domain-sensitive dialogue.

Feasibility

- **Proven Tech:** Uses reliable voice, NLP, and low-code integration tools.
- **Scalable Design:** Easily deployable with structured outputs and adaptive dialogue.

Viability

- **Strong Demand:** Solves key pain points for non-technical users.
- **Efficient & Unique:** Saves time, reduces cost, and offers a voice-first edge.

REFERENCES

- Breddels, Maarten A., and Jovan Veljanoski. "Vaex: big data exploration in the era of gaia." *Astronomy & Astrophysics* 618 (2018): A13.
- Vere, Steven, and Timothy Bickmore. "A basic agent." *Computational intelligence* 6.1 (1990): 41-60.
- Maynard, Diana, Yaoyong Li, and Wim Peters. "NLP Techniques for Term Extraction and Ontology Population." (2008): 107-127.
- An, Shengnan, et al. "Make your llm fully utilize the context." *Advances in Neural Information Processing Systems* 37 (2024): 62160-62188.
- Luo, Yajing, et al. "Characteristics and challenges of low-code development: the practitioners' perspective." *Proceedings of the 15th ACM/IEEE international symposium on empirical software engineering and measurement (ESEM)*. 2021.