Parrandata
Datathon 2025
Orderfox Challenge

Location | Company Name | DeepSupplyThink | Ask about suppliers, logistics, or procurement insights.

Analytics Club
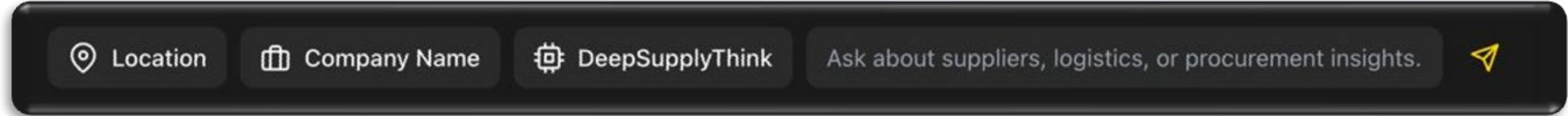
Our approach

Hybrid retrieval

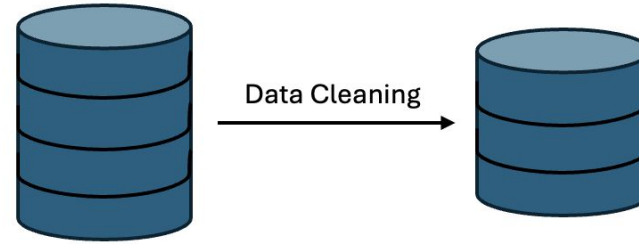| -0.005 | 0.012 | -0.008 | ... | -0.010 |
| -0.005 | 0.012 | -0.008 | ... | -0.010 |
| -0.005 | 0.012 | -0.008 | ... | -0.010 |

Vector database
with embeddings

SQL queries
Exact matching

Multi-agent LLM
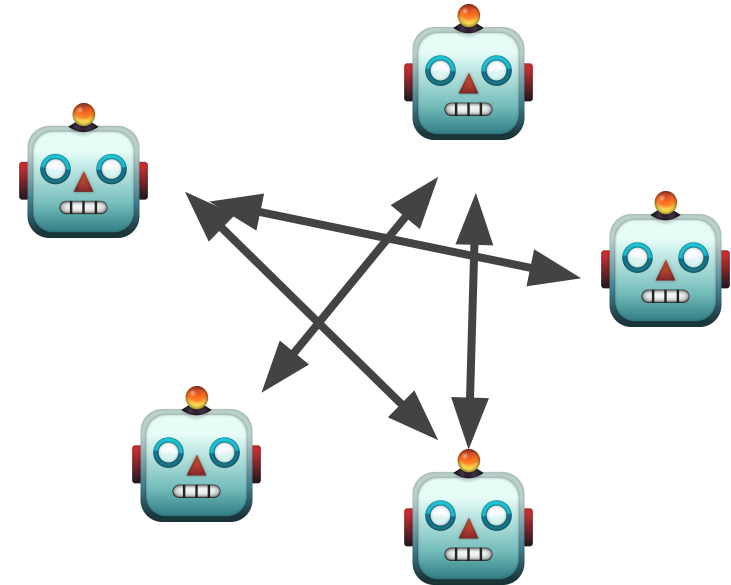
1. Data Cleaning & Preprocessing

- Remove raw files (png, jpg, pdf, mp4, ...)

- Remove language duplicates (es, fr, de) and irrelevant pages (terms-of-use, privacy-policy, cookies, ...)

- Remove huge noisy files
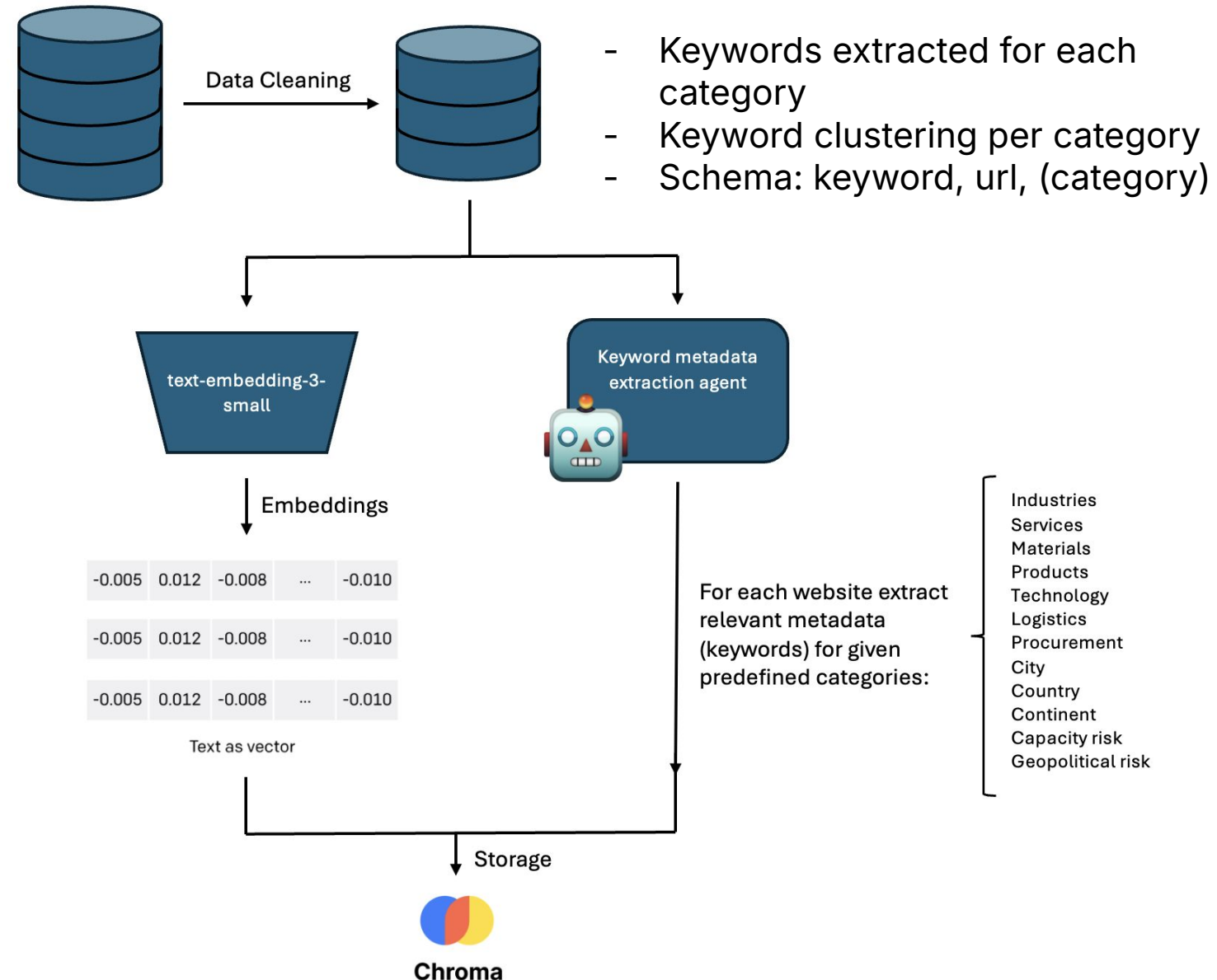

Data Cleaning

1. Data Cleaning & Preprocessing

- Remove raw files (png, jpg, pdf, mp4, …)

- Remove language duplicates (es, fr, de) and irrelevant pages (terms-of-use, privacy-policy, cookies, …)

- Remove huge noisy files

2. Knowledge Base Design

2.1 Embedding DB

- We have used OpenAI's text-embedding-3-small model to encode each of the companies' landing page.

- We store the embeddings in a ChromaDB

2.2 Keyword No-SLQ DB

- Keywords extracted for each category
- Keyword clustering per category
- Schema: keyword, url, (category)

Data Cleaning

text-embedding-3-small

Embeddings

-0.005	0.012	-0.008	...	-0.010
-0.005	0.012	-0.008	...	-0.010
-0.005	0.012	-0.008	...	-0.010

Text as vector

Keyword metadata extraction agent

For each website extract relevant metadata (keywords) for given predefined categories:

Industries
Services
Materials
Products
Technology
Logistics
Procurement
City
Country
Continent
Capacity risk
Geopolitical risk

Storage

Chroma

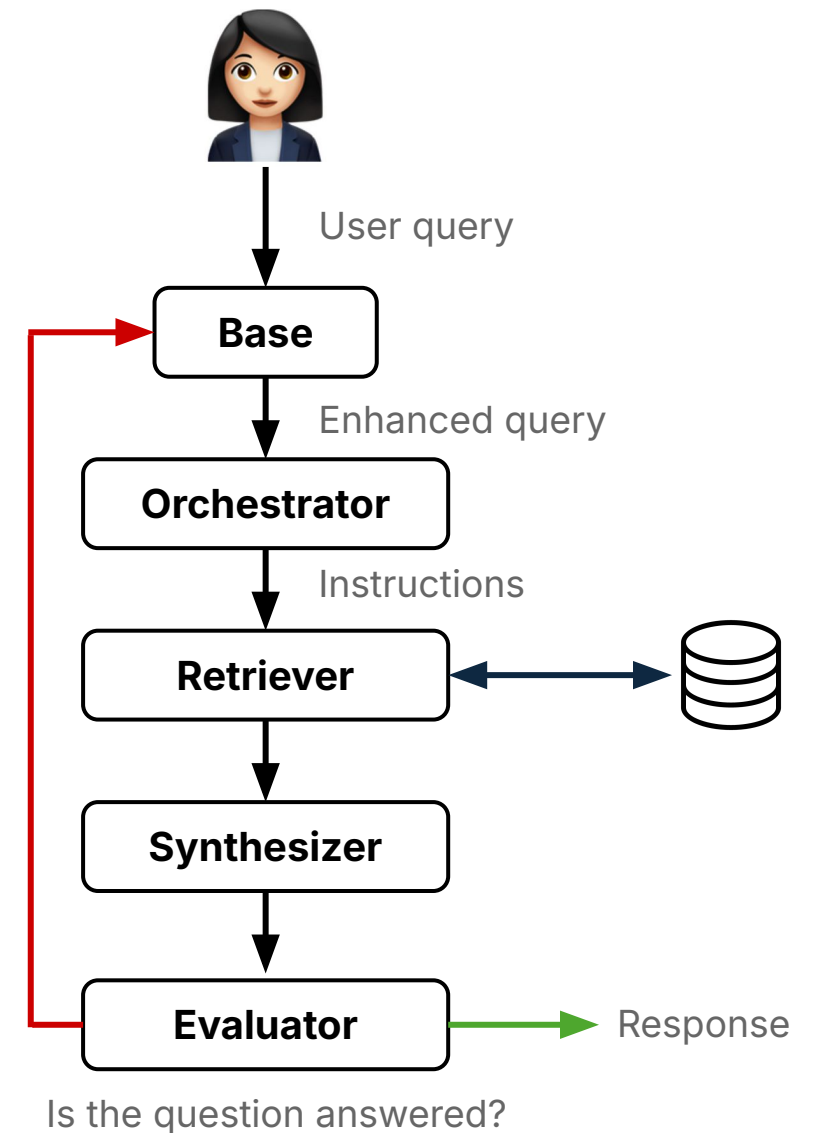# 3. Retrieval system

3.1. Fast mode: Single-shot RAG

- Semantic search with the embedding of the **expanded query** and the ChromaDB embeddings from the **document homepage**.

- Chooses top 10 documents and includes the raw text as context.

- Generation: Returns a detailed answer based on the query and the content of the documents.
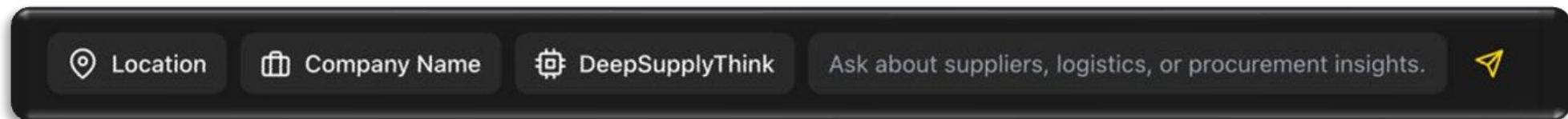
3. Retrieval system

3.2. DeepSupplyThink: Multi-agent RAG

- Steps:
    1. Query Enhancer
    2. Orchestrator
    3. Specialized Retrievers
    4. Synthesizer
    5. Evaluator
- Uses **LangGraph** & LangChain
- Relevant categories for the RAG agents: "industries", "services", "materials", "products", "technology", "logistics", "procurement", "regions", "capacity_risk", "geopolitical_risk"
- For each category, the agents dive deeper into selected keywords within each category, chosen by the enhancer.

User query → Base → Enhanced query → Orchestrator → Instructions → Retriever → Synthesizer → Evaluator → Response

Is the question answered?
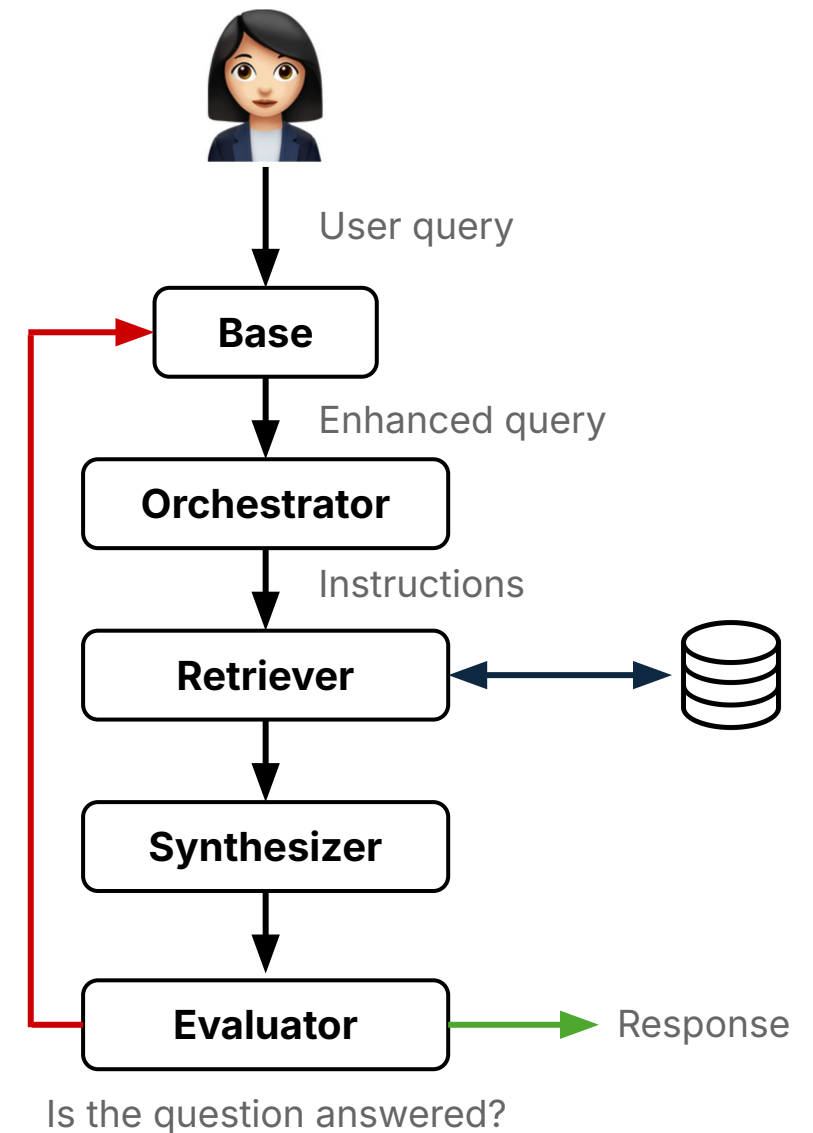
Location | Company Name | DeepSupplyThink | Ask about suppliers, logistics, or procurement insights...

3. Retrieval system

3.2. DeepSupplyThink: Multi-agent RAG

1. The **base** agent receives the query and enhances it to pass it to the other agents.
2. The **orchestrator** receives the query and decides what information is needed to answer it
3. With the **retriever**, the specific queries that need to be done to the databases are chosen, as well as which methods to use (vector embeddings, sql queries...).
4. The queries are executed and the **synthesizer** aggregates all of the answers.
5. The **evaluator** decides if the result answers the original query properly, and provides an **assessment**, **score** and **feedback**.
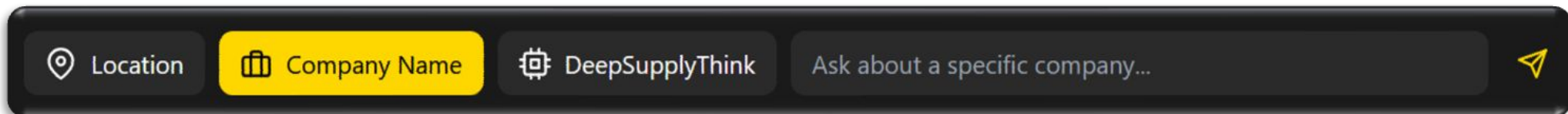
3. Retrieval system
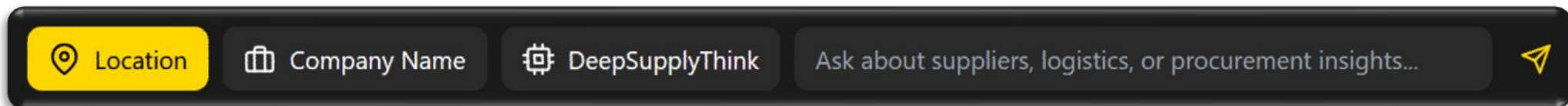
3.3. Company deep-dive: fuzzy matching

- When prompted about a particular company, we find the relevant documents through **fuzzy matching** on the company name

- The company name is **extracted from the query** using an LLM call

- With the company-related documents (.json), we craft a data-driven response to the query

- If a company is not in the database it will let the user know (next step: augment the database)
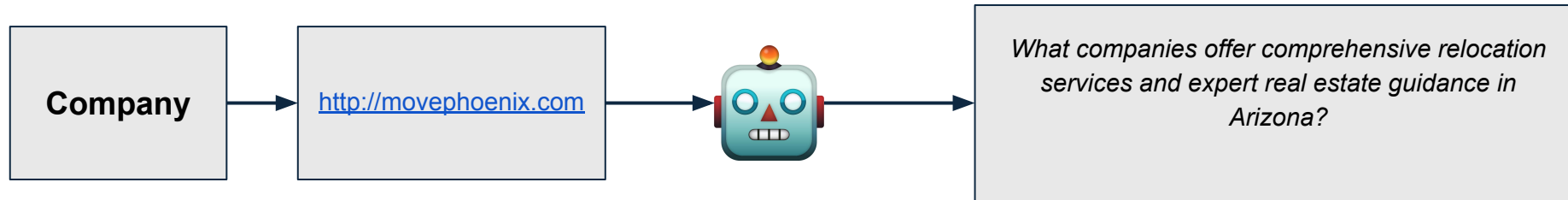
3. Retrieval system

3.4. Geospatial analysis: SQL

- We use the **keyword database** to find all the companies located in the location relevant to the query

- The retrieval system **formulates a SQL queries** to get the relevant companies from the database at different geographic scales

- Generation: The **statistics derived from the query** are presented to the user based on city, country, and continent

- Limitations: Currently we support count-based queries

Location   Company Name   DeepSupplyThink   Ask about suppliers, logistics, or procurement insights...

4. Evaluating the embedding RAG

QA datasets: The LLM provides a question about a webpage

| Company | → | http://movephoenix.com | → | 🤖 | → | *What companies offer comprehensive relocation services and expert real estate guidance in Arizona?* |

→ Our RAG system obtains the following recalls@10: Easy: 0.36 Hard: 0.26.

4. Key Features

- **Conversational capabilities**: The Multi-Agent system stores the conversation history. All agents use it. Allows for follow-up questions.

- **Location extraction**: We allow the user to specify to the system that it is looking for locations.

Evaluation metrics

- Evaluation score based on Recall@10

5. Challenges

- Dataset Size

- OpenAI API Call cap (60 per minute)

- LLM Hallucination → Requires thorough and detailed prompts & the evaluation agent.

5. Demo!

6. Groves Capital

- **Overview**: A family-owned business based in San Diego, Groves Capital offers a wide range of mortgage products, including those for new construction.

- **Timely Closings**: They emphasize a smooth closing process and maintain communication with clients throughout.

- **Unique Features**: Groves Capital provides a tech-savvy approach to mortgage lending, allowing clients to apply online and track their loan status easily.

- **Customer Satisfaction**: They have received high marks for their professionalism and efficiency, with many clients noting a seamless experience.

Summary of Key Features

- **Timely Closings**: Most lenders aim for closings within 30 days of completion, with a strong emphasis on communication and coordination with builders.

- **Unique Benefits**: Many companies offer personalized service, paperless processes, and various loan options tailored to new construction.

- **Customer Satisfaction**: High ratings across the board for responsiveness, support, and overall experience.

These lenders represent a mix of national and regional options, providing a range of services tailored to new construction home buyers. If you have specific preferences or need further details about any of these lenders, feel free to ask!

Give me more details on Groves Capital

Groves Capital is a financial services company that offers a variety of mortgage and lending options tailored to meet the needs of different borrowers. They provide services for various loan types, including FHA loans, VA loans, commercial refinance, hard money loans, and more. Their offerings also extend to specialized loans for investors, such as fix-and-flip loans, long-term rental loans, and loans for foreign nationals.

Additionally, Groves Capital features resources to assist clients in navigating the borrowing process, including mortgage calculators, guides, and success stories. They cater to a diverse clientele, including self-employed individuals and those in the gig economy, by offering tailored solutions like bank statement loans and stated income loans. Their website contains comprehensive information about their services, along with

Location | Company Name | DeepSupplyThink | Ask about a specific company...