**Meeting with Anna-Hellena:**

- Senior portfolio researcher at DWS, with 8 years of experience.

- Pursuing a part-time PhD at UCL in computer science, focusing on AI in financial market analysis.

- Explains **DWS**:

* Formerly Deutsche Asset & Wealth Management, asset management arm of Deutsche Bank.
* Multi-asset manager: from liquid investments (ETFs, equities, fixed income) to illiquid ones (private credit).
* Works with retail and institutional clients (sovereign wealth funds, pension funds, insurance companies).
* Anna’s department handles institutional client relationships and quantitative strategies.

**Feedback on the project idea:**

Suggesting **considering cost aspects** (local vs large general LLMs).

Smaller models (e.g., BERT) may be better for tasks like sentiment analysis - "Occam’s Razor" principle.

Corporations often choose LLMs based on existing cloud partnerships (e.g., Gemini on GCP).

**Our value can lie in:**

Defining **unique characteristics** for recommendations.

Using academic and industry papers to justify choices.

Providing a **library of models** (general and finance-specific).

Including RAG setups and prompt engineering examples.

A **simple explanatory tool** could help non-technical colleagues.

**Tasks where LLMs could be useful (Financial Use Cases for LLMs):**

1. **Synthetic data generation** (using human-coded rules to produce training data).
2. **Sentiment analysis** (BERT better).
3. **Table understanding** & **LLM-to-SQL** no-code queries.
4. **Reporting**:
   * Combining performance data with news to generate fund reports.
5. **Compliance & risk controls**.
6. **Language / policy checks** (e.g., political sensitivity).
7. **Text generation** for pitches/reports.
8. **RAG-based document retrieval** to reduce hallucination.
9. **Proposal responses** automation using historical Q&A databases.

**Suggested starting point**: sentiment analysis or RAG-based information extraction.

**Hardest part**:

Defining and keeping updated the **characteristics** that drive the recommendations.

**Use diverse sources:**

* Academic papers
* News articles
* Social media trends (e.g., Reddit consensus)

**Trust-building:**

* Start with small-scale feedback (20–30 users).
* Use feedback collection similar to RLHF (compare two outputs).

### **Clarifying Synthetic Data Generation:**

* Synthetic data can represent **expert analysis patterns** (e.g., ranking news events by importance).
* Useful when real annotated datasets are scarce.
* LLMs can generate data tagged with expert-like scores, which can train other models.
* Numerical data generation is harder and often experimental.

**Reliability of Reddit & Informal Sources:**

Reddit can be useful **only if**:

* There is **consensus** (many agreeing posts).
* Supplemented with higher-trust sources (papers, trusted blogs).

Provide users with **certainty levels** (e.g., “40% confidence”).

### **Internal Evaluation Practices at DWS**

- LLM usage driven by **cost** and **security**.

- Internal models: Gemini, LLaMA, and a Deutsche Bank fine-tuned LLaMA.

- External vendors allow **side-by-side LLM comparisons.**

- No fully systematic evaluation, often subjective preferences.

High priority on:

* **Hallucination reduction** (RAG-based fact linking).
* **Numeric value accuracy** (values inserted from databases, not generated).

### **AI as of now in Finance**

- Regulatory acceptance depends on **human-in-the-loop.**

- Fully automated AI not allowed in production for finance (EU regulations).

For this project:

* Recommender system’s explainability is more about **method transparency** (data, characteristics, weights).
* Industry shifting from strict determinism to pragmatic use of LLMs with safeguards.

### **Our Model Recommendations:**

* No system will be perfect - aim for **approximation with confidence scores**.
* Academic research, blogs, and social data can inform choices.
* Taxonomy of tasks is useful for grouping recommendations.

### **Module Prioritisation**

* Core priority: **recommender module** with high-quality, unique characteristics.
* Output layer (explainer) still important for non-technical users.
* Dashboard can be built after the core logic is ready.