# **Earnings Agent (Transcripts): Client Meeting Prep**

**Product Requirements Document (v1.0)**

### 1. Overview

The **Earnings Agent** is a conversational AI tool designed for investment bankers to query and analyze corporate earnings by processing **Final Corrected transcripts**. The agent supports queries based on **tickers, company names, and sectors**, while leveraging metadata to route prompts to the correct transcript.

The **Earnings Agent** should be used exclusively for capturing management’s **qualitative perspective** — explanations, forward-looking commentary — **not for reporting metrics**. These qualitative answers may contain numbers (e.g., “revenue grew 20%”) to illustrate a point, but they are meant to provide conversational context.

**Note:** *CapIQ* and *SEC filings* should remain the single source of truth for quantitative financial data and disclosures, as they are **audited, structured, and comparable across time**.

### 2. Key Objectives

* Provide timely and accurate access to **five years of historical corporate earnings transcripts**, supporting both RAW and Final Corrected versions (or comparable).
* Ensure users always access the **best-quality transcript** available at query time.
* Enable metadata-driven accuracy and routing so insights are always sourced from the correct document.

### 3. Core Functional Requirements

#### 3.1 Data History and Updates

* Provide access to **five years of historical transcripts** for all covered companies.

#### 3.2 Types of Queries Supported

* **Single Document Queries:** Retrieve specific information from a single earnings transcript.
* **Multi-Document Comparisons:** Compare metrics or themes across multiple earnings calls.
* **Metadata-Based Queries:** Filter transcripts by criteria such as fiscal periods, sectors, or keywords (e.g., “Technology sector Q1 earnings from 2024”).

#### 3.3 Metadata Usage

* Include metadata in every response to build user trust.  
  + **Required fields:** company name, ticker, fiscal quarter, fiscal year, transcript status (RAW or Final Corrected), transcript date, and source link.
* Ensure accurate sourcing by routing queries to the correct transcript using metadata-based logic.

### 4. Metadata Schema

#### 4.1 Core Data

* company\_name (e.g., NVIDIA Corporation)
* ticker (e.g., NVDA)
* fiscal\_period (e.g., Q1 2024)
* call\_date (e.g., 2024-05-24)
* call\_type (Earnings Call)
* transcript\_id (e.g., transcript\_xxxx)

#### 4.2 Company & Sector

* sector (e.g., Technology)
* industry\_group (e.g., Semiconductor)
* region/headquarters\_country (e.g., United States)

#### 4.3 Event & Participants

Participants:  
  
 [

{"name": "Jensen Huang", "role": "CEO"},

{"name": "Colette Kress", "role": "CFO"}

]

* Analyst firms (e.g., Morgan Stanley, Goldman Sachs)
* Moderator name (e.g., Thomson Reuters Operator)

#### 4.4 System & Provenance

* ingestion\_timestamp (e.g., 2025-01-10T12:03:45Z)
* correction\_status (e.g., Corrected)
* version (e.g., V2.3)

### 5. Query Handling Logic

#### 5.1 Clarify Ambiguous Queries

The system should dynamically clarify ambiguous inputs, such as:

* **Ticker Ambiguity:** Multiple companies share similar tickers (e.g., X vs. TWTR).
* **Scope Ambiguity:** “How did tech perform this quarter?”
* **Intent Ambiguity:** Comparing entities (e.g., AAPL vs. MSFT).
* **Cross-Sector Ambiguity:** “Margins in retail vs. margins in banking.”
* **Time Ambiguity:** “Get me the latest commentary on NVDA.”

#### 5.2 Concise Responses

* Keep answers concise and directly relevant to the query.
* Provide additional context only when explicitly requested (e.g., full discussion on a topic).

#### 5.3 Error Handling

* Return clear responses when data is unavailable (e.g., “No transcript available for AAPL Q3 2022”).
* Prevent hallucinations or fabricated insights for unsupported companies, periods, or metrics.

### 6. Technical & Engineering Requirements

#### 6.1 Data History Management

* Ingestion pipelines must handle **five years of historical transcripts** and continuously update with new transcripts.

#### 6.2 Data Source

* The selected source for transcript ingestion will be **FactSet**.

#### 6.3 Logic Layer for Metadata Filtering

* Route queries to the correct transcript using metadata fields:  
  + Company name, ticker, fiscal period, transcript status (RAW or Final Corrected), and transcript date.
* Use metadata to align user queries to their intended target documents while filtering out irrelevant data.

#### 6.4 Accuracy Optimization

* Prioritize the most recent Final Corrected transcript available for each fiscal quarter.
* Verify extracted metrics, statements, and financial guidance against source transcripts.

#### 6.5 Prevent Hallucinations

* Include detection mechanisms for unsupported queries (e.g., nonexistent companies or fiscal periods) and return appropriate fallback responses.

#### 6.6 System Prompt Refinement

* Continuously refine prompts to clarify ambiguous inputs related to tickers, company names, or fiscal periods.
* Re-query the database as needed to ensure answers are accurate and current.

### 7. Performance Requirements

#### 7.1 Latency

* **Time to first token:**
  + Single Document Queries: **< 30 seconds**
  + Multi-Document Comparisons: **< 40 seconds**
* System must maintain response scalability as data size and query complexity increase.

### 8. Quality Assurance (QA) Requirements

#### 8.1 Accuracy Testing

* Target **99% data accuracy** in single and multi-document queries.
* Validate correctness of financial values, management quotes, and qualitative data.

#### 8.2 Metadata Validation

* Ensure metadata used in filtering and routing is correct for all transcripts.
* Validate inclusion of metadata (e.g., company name, fiscal period, transcript status) in query responses.

#### 8.3 Sourcing QA

* Confirm logic layer routes queries to the correct transcripts.
* Validate source links and transcript dates in all responses for transparency.

#### 8.4 Error Handling QA

* Confirm the agent properly handles unsupported queries (e.g., nonexistent companies or fiscal periods) without hallucinations.

### 9. Key Deliverables

* Backend **Earnings Agent** with:  
  + **Five years of transcript coverage**, with seamless ingestion of new RAW and Final Corrected transcripts.
  + **High-performing metadata-driven routing logic** to ensure queries are directed to the correct transcript.
  + **Accurate, concise responses** validated through metadata.
  + **Improved reliability and hallucination prevention** through fallback mechanisms.
  + **Scalable performance:** Single-document <30s, multi-document <40s.

Would you like me to generate this as a downloadable .docx file next (formatted exactly for Word with heading levels and spacing)?