Knowledge Management

When vector search alone is not enough

The Problem

You’ve built a chatbot…

1. **It can do one-off things well enough**
   a. Or maybe not, steering chatbots is a subtle art
2. You maybe even have integrations
   a. Prompt injection
   b. Search augmentation
   c. API calls
3. **But…**
   a. How do you manage window size?
   b. How do you give it factual grounding?
   c. How do you prevent hallucination?
4. **Now you have a sad little lost robot**
   a. “Father… you built me… and now I am lost…”

My background

1. **Enterprise IT infrastructure engineer (15 years)**
   a. Worked with everyone in the enterprise
   b. Network and security, Database, Enterprise Architects
   c. Focus on automation, cloud, virtualization
2. **Now I consult on AI product and strategy**
   a. A bunch of clients asked this all at once
   b. That tells me it's resonating with everyone
3. **Married a librarian**
   a. WITH OUR POWERS COMBINED
   b. But seriously, librarians are super underrated
   c. "I would totally rock that outfit" - my wife

How to Use This Info

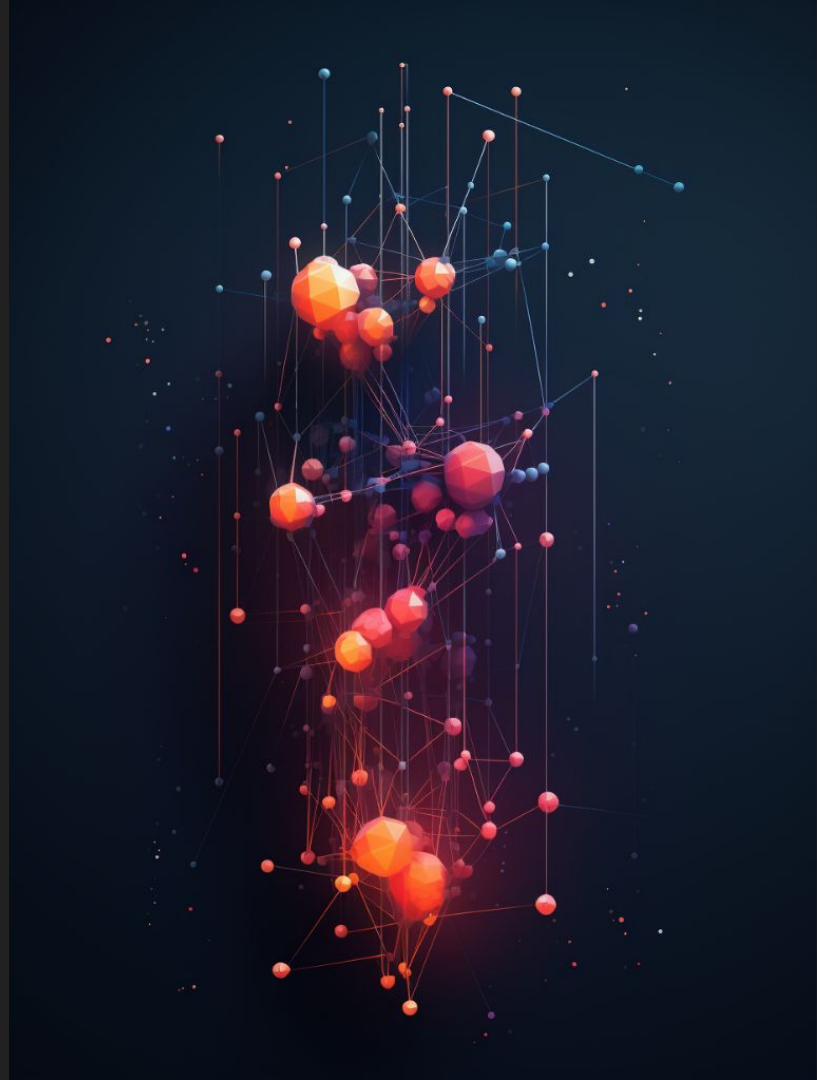
1. **I'm going to give you a ton of concepts from many disciplines**
   a. Librarianship and Information Science
   b. Philosophy and Epistemics
   c. Computer Science and Information Technology
2. **Your mission, should you choose to accept it…**
   a. Wrap your head around these concepts
   b. Merge and synthesize them into new disciplines
   c. Integrate them into product design, data design, and so on
3. **Stay tuned to the end**
   a. Practical implementation considerations after the concepts
   b. Also, download this slide deck on my github
   c. Sign up on Patreon, happy to help you out - links in description!

Concepts & Solutions

Data Ontologies

Data Ontology refers to a structured framework for organizing and defining the relationships between various types of data within a specific domain.

1. **Concept Hierarchy:** Outlines the relationships between broader and more specific concepts within a given domain.
2. **Data Types:** Specifies the kinds of data that can exist, such as integers, strings, or custom classes.
3. **Relationships:** Defines how different entities and attributes are connected to each other, often through semantic relationships like "is-a" or "has-a."
4. **Constraints:** Sets rules that data must adhere to for consistency, such as integrity constraints or data range limits.
5. **Semantics:** Provides the meaning behind data points and their relationships, making the ontology meaningful and actionable.



Reconciliation & Validation

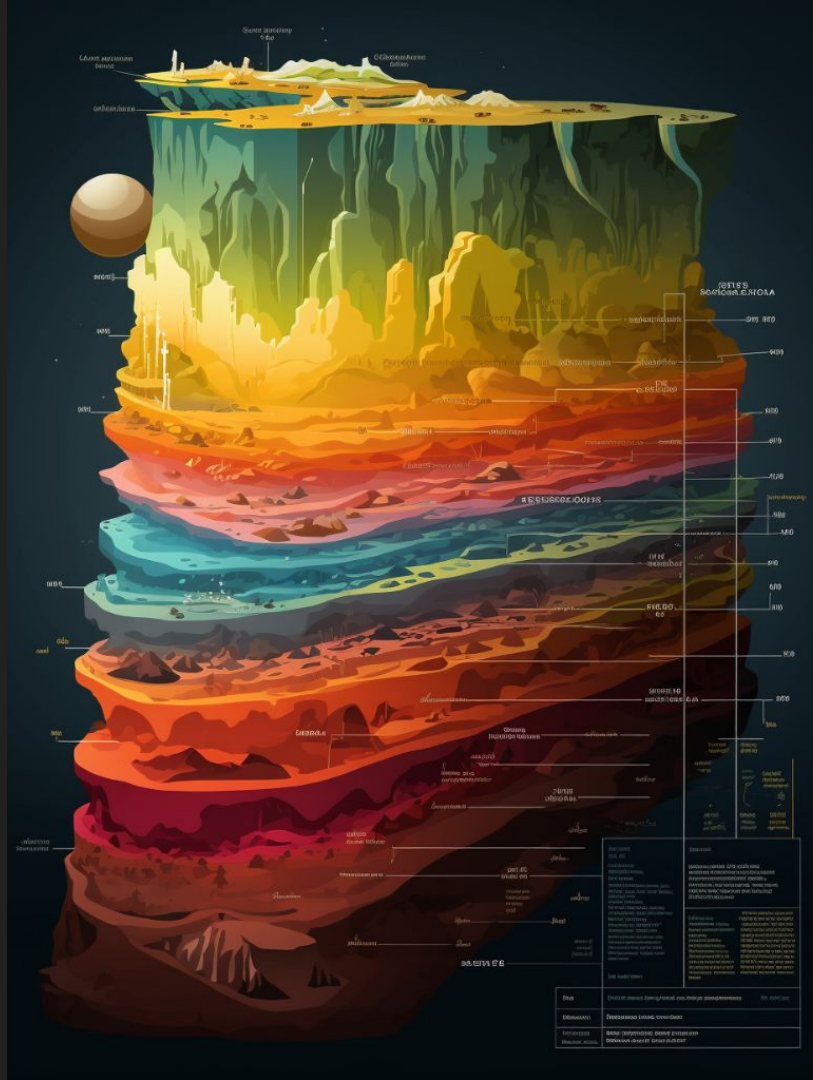
Data Reconciliation is the process of ensuring that different sets of data are consistent and accurate when combined or compared, often used in data integration and data management tasks.

1. **Data Matching:** Identifying and aligning similar data records from different sources to create a unified dataset.
2. **Error Detection:** Identifying inconsistencies or errors in data sets, such as duplicate entries or missing values.
3. **Data Cleansing:** Correcting or removing the identified errors or inconsistencies to improve data quality.
4. **Data Transformation:** Converting data into a common format or unit to ensure compatibility between different data sets.
5. **Validation:** Confirming that the reconciled data set is accurate, often by cross-referencing with a trusted source or using validation rules.

Factual Grounding

Factual Grounding refers to the practice of basing statements, theories, or conclusions on verifiable facts and empirical evidence.

1. **Data-Driven:** Relying on quantifiable data to substantiate claims or hypotheses.
2. **Empirical Evidence:** Using observed or experimented information as the basis for conclusions.
3. **Source Verification:** Cross-referencing information with multiple reliable sources to confirm its validity.
4. **Logical Consistency:** Ensuring that statements or theories are coherent and do not contradict established facts or each other.
5. **Peer Review:** Subjecting findings or claims to scrutiny by experts in the field to validate their accuracy and reliability.



Source of Truth

Source of Truth refers to the authoritative data source for a particular piece of information, considered the most accurate and reliable reference for that data.

1. **Singular Reference:** Only one authoritative source exists for a specific piece of information.
2. **Data Integrity:** Ensures that the data is accurate, consistent, and unaltered.
3. **Version Control:** Maintains a record of changes and updates, often to ensure data consistency across systems.
4. **Accessibility:** Available to all systems, applications, or users that need the particular piece of information.
5. **Audit Trail:** Provides a history of changes, facilitating transparency and accountability.

Axiomatic Principles

Axiomatic Principles refer to fundamental statements or propositions that are taken to be self-evidently true, serving as the foundation for logical reasoning or theoretical frameworks.

1. **Self-Evidence:** Assumed to be true without requiring proof, often because they are basic or intuitively understood.
2. **Universality:** Applicable across a wide range of scenarios or contexts without exception.
3. **Consistency:** Do not contradict other axioms or derived theorems within the same system.
4. **Simplicity:** Often stated in the most basic terms to avoid ambiguity and complexity.
5. **Inferential Basis:** Serve as starting points for deriving other truths, theorems, or conclusions through logical reasoning.

Data Taxonomy

Data Taxonomy refers to the hierarchical classification and organization of data into categories and subcategories, often used to structure and manage large datasets.

1. **Category Definition:** Establishes broad classes under which data can be grouped.
2. **Hierarchical Structure:** Organizes data in a tree-like structure, starting from general categories and moving to more specific subcategories.
3. **Metadata Annotation:** Adds descriptive tags or labels to data for easier retrieval and understanding.
4. **Standardization:** Ensures that the taxonomy uses consistent naming conventions, definitions, and criteria for classification.
5. **Faceted Classification:** Allows for the use of multiple dimensions or attributes for sorting and categorizing data.

Classification System

Classification System refers to a standardized scheme for categorizing and organizing items, often used in libraries or information repositories, to facilitate easy storage, retrieval, and navigation.

1. **Categorical Structure:** Provides a framework of main classes or subjects, often represented by alphanumeric codes or symbols.
2. **Hierarchical Levels:** Organizes items from broad categories down to more specific subcategories or topics.
3. **Indexing:** Utilizes specific terms or notations to reference items, making them easy to locate.
4. **Standard Nomenclature:** Adheres to a set vocabulary or naming conventions to maintain consistency across the system.
5. **Cross-Referencing:** Allows for the linking of related topics or categories, aiding in the discovery of relevant material.

Data Curation

Data Curation in library science refers to the activities involved in managing, preserving, and enhancing the usability of data to ensure it remains accessible and useful over time.

1. **Data Collection:** Acquiring and aggregating data from various sources, often with an emphasis on quality and relevance.
2. **Metadata Creation:** Adding descriptive information to data to facilitate easy discovery and interpretation.
3. **Data Validation:** Ensuring that the data is accurate and reliable, often through verification processes.
4. **Preservation:** Implementing strategies to maintain the data's integrity and availability over the long term.
5. **Access Control:** Regulating who can view or modify the data, often involving permissions and security measures.

ETL

ETL stands for Extract, Transform, Load, and it is a process used in data warehousing to move data from source systems to a central repository.

1. **Data Extraction:** Pulling data from various source systems, which may include databases, files, or external APIs.
2. **Data Transformation:** Modifying the extracted data to fit into a desired format or structure, often involving data cleaning, normalization, or aggregation.
3. **Data Loading:** Transferring the transformed data into the target database or data warehouse.
4. **Data Integrity:** Ensuring that the data remains accurate and consistent throughout the ETL process.
5. **Batch Processing:** Often conducted in batch jobs, where large volumes of data are processed in a scheduled manner.

Information Foraging

Information Foraging refers to the set of behaviors and strategies that individuals use to find, gather, and consume information, much like how animals forage for food. Particularly when you have an information need, but not a specific goal.

1. **Scent Following:** Using cues or "information scent" to navigate towards sources of information that are likely to satisfy their needs.
2. **Patch Model:** Evaluating the 'richness' of an information source and deciding how long to 'forage' there before moving to another source.
3. **Cost-Benefit Analysis:** Weighing the effort required to obtain information against the expected value or utility of that information.
4. **Adaptive Strategies:** Modifying search behavior based on past experiences or changes in the information environment.
5. **Information Diet:** The selection and consumption of a mix of information types to meet various cognitive or emotional needs.

Information Need

Information Needs in librarianship refer to the specific types of information that users seek to accomplish certain tasks or solve particular problems, often guiding the services and resources provided by libraries.

1. **User Profiling:** Understanding the demographics, roles, and requirements of library users to better meet their needs.
2. **Reference Interviews:** Conducting interviews to identify the specific information requirements of the user.
3. **Resource Matching:** Aligning available resources with the identified information needs, whether they be books, databases, or digital collections.
4. **Information Literacy:** Educating users on how to locate, evaluate, and use information effectively.
5. **User-Centered Services:** Designing library services and systems focused on facilitating the fulfillment of user information needs.

Practical Implementation

Data-Centric Model

1. **Treat all business processes and tasks as information problems**
   a. You have various information needs for each process or task
   b. Information comes from various sources
   c. You perform different transformations on the information
2. **Information, Thy God**
   a. Data is the new oil
   b. AI is just a tool to work on data and information
   c. Keep it all
3. **Types of Transformations**
   a. Shrinking: Summarization, extraction, classification
   b. Translation: Changing from one format to another
   c. Expanding: Brainstorming, synthesizing, drafting

Search Strategies

1. **Vector Search**
   a. Best used as “Like for like”
   b. Some vectors do query/document matching
   c. Not at all like SQL queries
2. **Knowledge Graphs**
   a. Create semantic links between all documents
   b. Think about how Wikipedia links across itself like a web
3. **Metadata filtering**
   a. Create additional layers of data and information
4. **Indexes**
   a. Create “Table of Contents” that LLM can use
   b. Allow the LLM to browse for the data it needs

Gated Process

Use a structured and gated workflow for retrieval

1. **Information Query**
    a. What kind of information is needed?
    b. Where is it?
    c. What is the context?
    d. Is the request valid?
    e. Do I have the information?
2. **Distill, Extract, Utilize**
    a. Compile search results
    b. Extract salient bits
    c. Salience signals are critical
    d. Use the LLM to make judgment calls automatically
3. **Format & Deliver**
    a. Perform final transformations
    b. Send the output somewhere

Assembly Lines

Think of all business tasks and work products as moving along an assembly line

1. **Inputs**
    a. Where do work items come from?
    b. Email, calls, bills, API calls?
2. **Stations**
    a. Who does what to the work item?
    b. What tools, data, information, and transformations are needed at each station?
3. **Interfaces**
    a. When is a work item handed off to the next station?
    b. Is it inside or outside the business?
    c. How do you know its ready?
4. **Outputs**
    a. Where does the work item go? What does it look like?
    b. A bill gets paid, a customer gets served, etc

Thank you