Synthetic Cognition Meets Data Deluge: Architecting Agentic AI Models for Self-Regulating Knowledge Graphs in Heterogeneous Data Warehousing

[1]Srinivas Kalyan Yellanki, [2]Dwaraka Nath Kummari, [3]Goutham Kumar Sheelam, [4]Sathya Kannan, [5]Chaitran Chakilam

[1]*Software Engineer, Expedia Group, yellankisrinivaskalyan@gmail.com*
[2]*Software Engineer, Sailotech Inc, dwarakanathkummari@gmail.com*
[3]*IT Data Engineer, Sr. Staff, Qualcomm Inc., gouthamkumarsheelam@gmail.com*
[4]*Sr AI Developer, John Deere, sathyakannan.vsl@gmail.com*
[5]*Validation Engineer, Sequel Medtech, chaitrann.chakilam@gmail.com*

Abstract: The realities of contemporary data management and representation are evolving at an increasing rate. However, we still lack the broad foundational bridges of core data warehousing principles relating to how high-level reports are generated internally so that users can psychologically intuit where they are in the vast and complex repository of data that resides in a typical data warehouse. IT workers must constantly support users or worry about failed ad-hoc or automated operations or whose results appear without explanation. Data management may not yet exist as a science. We need a more complete transformational view of the details of the internal mappings between data from diverse sources and conceptual data model object types. Cognitive model-driven and symbolic techniques have been approached to design and develop systems to automate and rationalize these transformational processes and to support user navigation and work. These techniques are now being displaced by advanced statistical learning methods. As designed, these methods mostly do knowledge creation in the basic steps of the transformational process, but they likewise at times pave data as well. Through AI as Intentional Cognition supplemented by language, this inhibition may be bypassed. Thus, despite both their synthetic and agentic capabilities, these approaches follow a surprising and quite diverse transition. The goal of this work is to show what tasks of data management and representation these methods might be able to tackle and when and how they might interleave towards a more collaborative AI Data Management. We conclude with directions for future work.

Keywords: Data Management Evolution, Data Warehousing Principles, Conceptual Data Models, Cognitive Navigation, Transformational Mapping, Symbolic Techniques, Statistical Learning, Knowledge Creation, AI as Intentional Cognition, human collaboration, Ad-Hoc Query Support, Automated Operations, Data Representation, Data Model Object Types, Data Science Foundations, Internal Report Generation, Diverse Data Sources, User Intuition, Language-Augmented AI, Collaborative AI Data Management.

1. Introduction

Data Warehousing (DW) produces captured store repositories of meaningful data that serve as a source of information for Business Intelligence (BI) that generates management decision-making reports, yet without the key reasoning feature in mind. Associative machinery can perform reasoning by drawing associative links between data elements, hence we call this Artificial Intelligence (AI) specific to data. We believe that DW needs to incorporate AI capabilities to afford a new generation of industry-specific driven Automated Decision Making (ADM). We further articulate the original concept of the Data-Aware Self-Organizing System (DASOS) to create added-value economic offerings for the Data Economy in the new technology regime of the Fourth Industrial Revolution.

Data, models, and algorithms are the fundamental factors of the data economy, where data needs continuous curation, and models, capitalizing on domain-driven data knowledge and best practices, need self-improving and optimizing automation through algorithms. The problem is that to be useful, data need to be structured and transformed into meaningful converted information, as well as high-quality cloud-based models need to be developed that prioritizes outliers' detection to model the unusual and unexpected. This is a complex endeavor that requires expert knowledge input of various data elements, their multidimensional relationships, and domain-specific generally inadmissible values. When it comes to business applications, expert knowledge of the market, competitiveness, consumer needs, and expectations is critical. Yet such Human Attention Integration (HAI) is limited, biased, and not scalable enough for massive amounts of frequently updated data in the domains of finance, security, and surveillance, among others, where gaps in the model's coverage can result in critical business failures.

The decision-making processes involved in these domains are rule driven, based on domain-driven expert knowledge, and are therefore logical – searching for consistency or validity checks rather than for probability assignments. In such Business Intelligence (BI) and Predictive Analytics (PA) applications, Artificial Intelligence (AI) models often add only marginal value to the decision-making processes involved.

1.1. Overview of Core Concepts

This section furthers the introductory remarks mentioning some key technologies in AI and discussing a few central topics. These are technologies that may be of interest in agentic AI, and topics that we may address in dealing with the interdependencies among agentic AI and data warehousing.

We start with AI technologies. In addition to symbolic AI and, more recently, neural network AI, we mention imitational AI, hybrid AI, and autonomous AI. However, yet another AI technology is possible, which we have tentatively labelled as renormalized AI. It is rather practical, and it takes the function and impact of general intelligence seriously without renormalizing it at the cost of aspiration and data needs. It is an AI aligned to solve difficult, specific problems.

Finally, we comment on a few central topics in discussing general intelligence and agentic AI: similarities with and differences from human cognition; the need for business-like, pragmatic solutions; gradual advancement and demonstration assumptions; benchmarking; jurisdiction; trust and trustworthiness; existential risk, superintelligence, and better-than-human capabilities; and a responsible, problem-solving, humane approach to AI. In this section, we may be a little more leisurely in our conceptual trajectory than in the following sections, where critical paths may predominate.

Our key point earlier on may be encapsulated in the research question: What is agentic AI in data warehousing? It is one of the four principal topics in the essay, Is it possible to cover them evenly and evenly distribute them across the essay? The other three are question-answering in data warehousing; the Data Underground, with its many reality tunnels, perhaps labeling topic-centered or tunnel-centered multitask aptitude in agentic AI; and the other tunnel-goers, of human, AI, social, and geopolitical nature, engaged in data-governance bureaucracies or informal conspiracies.

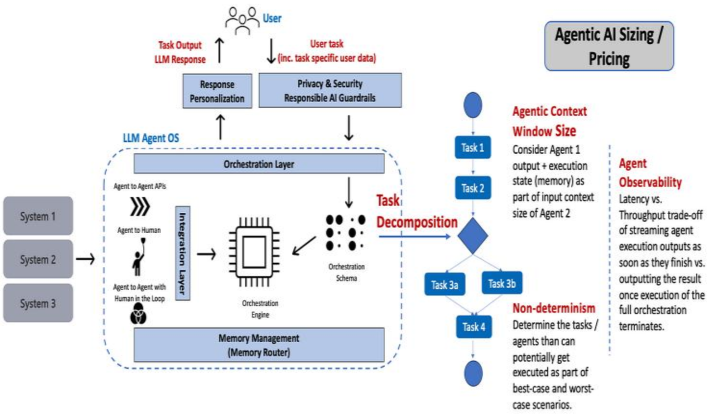

Fig 1: Agentic AI for Data Engineering

2. Background

1. Synthetic Cognition

Synthetic cognition refers to data science and artificial intelligence methods that exploit large probabilities of instances or probabilities for decisions on very large problems where straightforward algorithms cannot apply and whose knowledge storage requirements exceed local capacity. The promise is enhanced levels and degrees of agency, where an agent has a temperament, an internal concept of agency that has to be conformable to actions. With synthetic cognition, we have machines that think in a distinctively artificial manner and think in colors not black and white. As the domain of possible intelligent agents grows, on the one hand, the delimitation of agency and its inherent limitations in creativity becomes more salient and on the other hand, the potential for agency becomes radically more heterogeneous and diverse. Unlike the silicon machine that first beat a chess champion, we would have agents that could empathically inhabit a character in a Shakespearean play. We would have agents with personalities very different from the purely intellectual profile of traditional AI.

2. Data Deluge

The digital, web- and media-based, content-based explosion of represented data for people, objects, and relationships, together with the industrial and commercial movement of centralized data, results in a potential global knowledge system at the heart of the information economy. Industry and enterprise have always craved efficient, effective access to the knowledge of the external world; for a community of millions to share that ability would confer unprecedented wealth and activist power. To enable such possibilities we need to represent, store, explore, and reason about the unimaginable and ever-growing amounts of knowledge content all around us. The human population may not have the capacity to utilize such capabilities, but the net utility of automated systems that access large knowledge stores and services collaboratively would be immense.

3. Knowledge Graphs

A web-based, public knowledge graph is a highly heterogeneous, distributed database union of partial domain ontologies, of schemas, that encodes relations for people, events, and situations, for physical, intuitive, and abstract objects in a richly geocentric, dynamically annotated form while maintaining symbolic reference for meaning-dependent morphology and linkage. For semi-structured content as used in modern web search engines, supervised machine learning acts as a filter selecting probable instances and relations, while a semi-structured knowledge representation acts as a structural template linking all the instances and relations to their possible semantic identities, with the filter selection training a statistical model using a cross-validation set. A web-based, public knowledge graph is an unmediated proto-version of a universally collaborative intelligent agent with a boarding pass but no destination. This harvesting of symbolic content needs scaling to be defensible in the face of the data deluge.

2.1. Synthetic Cognition

Synthetic Cognition is a paradigm for human/machine interaction but also an application model in which humans define interactions that are implemented in AI systems and performed automatically by them: advanced but delimited capabilities including agency are transferred to machines, relieving human users from performing tedious tasks. The intimate synergy between Knowledge Graphs and AI is demonstrated in the popularization of the Web, an emergent KG within the rubric of Hypertext, the Semantic Web, based on Semantic KGs adhering to standards, and the recent explosion of generative AI that leverages foundational KGs – most importantly, WordNet. KGs have long been an important milestone on the road toward artificial intelligence, modeling knowledge about the world, and have also been seen as a substrate for the higher-order capacities of Synthetic Cognition. KGs are to be marshaled as a bridge toward the comprehensive capabilities of our Synthetic Cognition approach, specifically as pipelines for KGs developed by experts or crowdsourcing to populate and implement procedures in Synthetic Cognition applications for use by end users in real natural language. These processes reduce the burden on the end users. Notably, early Semantic Web visionaries used the term “collective intelligence” to describe the synergy between KGs and crowdsourcing.

The concept of Synthetic Cognition is based on the enabling of 4 generalized capabilities. These are performed collaboratively in a set of distinct but interdependent processes that implement a Cognitive Automation chain: Transfer, Synthesize, Optimize, and Explain or Feedback. Transfer defines Synthetic Cognition applications. Users transfer domain knowledge to the KG pipelines either directly in the form of instruction examples or indirectly as crowdsourcing for specific AI tasks. In Synthesize, the recommendations of the enabled Semantic Web pipelines are leveraged to synthesize the cognitive tasks in real-world applications, again using application-indicating KG directives.

$$SRES = -\sum_{i=1}^{n} P_i \log_2(P_i + \epsilon)$$

where:

- $SRES$ = Entropy of knowledge graph stability
- P_i = Normalized agent decision distribution on node i
- n = Number of knowledge graph nodes
- ϵ = Small constant to prevent log(0)

Equation 1: Self-Regulation Entropy Score (SRES):

2.2. Data Deluge

Due to the enormous advances in technology, society is drowning in functions, products, services, and consuming data; social, political, and economic developments; as well as their interdependencies. Without any doubt, we are experiencing a data deluge. Until now, society perceives the data deluge as a chance as well as a burden, even over deadlines. Chance, because thousands of data and information, which could never be collected, systematically captured, and evaluated, are now present in physical as well as in digital forms. However, despite substantial investments in preparation as well as in the broad distribution of the latest technology, the burden prevails, because the research and practice cannot satisfactorily convert theory and intention into action. Thus, in general, humans serve their cognitive limitations to comprehend our world. Of general processes, dependencies, and causal relations; which are indispensable as well as fundamental for any activities in everyday life and research work, decision-making, and problem-solving. Critical value-added activities and decision tasks are still overwhelmingly heuristic and rarely supported or guided by anything.

To overcome human cognition limitations about the crucial world functions and their interrelations, it is indispensable to gather all digitalized information and knowledge about global areas, regions, and interests. Hence, their transformation into machine-understandable digital knowledge constitutes the decisive precondition for service any activity generating intelligence. Every human and company can with ease share their local task routing knowledge, packaged as general processes, MAIs, MOs, etc., especially because this knowledge must not contain any business-critical information; for example, it is mainly indirect knowledge about local implementations which is unique to a specific environment, whereas the wrapper knowledge is non-business-related. Hence, the enormous growth of the open-source community offers powerful opportunities.

2.3. Knowledge Graphs

Knowledge graphs are excellent tools for addressing synthetic cognition and are a development from ontology technology, where the "knowledge" in "knowledge graph" comes from formalizations - such as axioms and rules - added to graph data. Rendering one's data as a knowledge graph not only makes that data understandable by humans and interoperable with diverse other data ports or formats, but it also programs it for cognition. This preparation enables humans and machines alike to ask knowledge-based questions, query by deduction and inference - through a form of automatic reasoning - not only search, and associate disparate data points, patterns, and relations. In this way, machine learning can be made more meaningful and contextual, as the results can be placed into a multi-layered network of concepts.

Many data warehouses around today support question-asking, pattern and relation-seeking, and answer-seeking in relations existing between disparate datasets, and data at disparate levels of semantic abstraction, explanations, and meaning-patterning. Currently, entity-relation triples are the predominant means for rendering a data set as a semantic graph; relations come with a type - for example, is-a, is-part-of, has-property, is-associated-with. Nodes/vertices may link to more than one hub through a specific relation type. Relations can take in multiple node parameters; additionally, parameters/signatures are included on both ends of single-link relations.

2.4. Heterogeneous Data Warehousing

Data Warehousing (DW) is an essential component of a data ecosystem capable of meeting the demands and expectations of its stakeholders. A DW integrates heterogeneous data, organizes it for analytics, optimizes storage costs while enabling rapid analytical processing, and allows data sharing, security, and governance functions. Particularly, heterogeneous DW allows the storage and processing of diverse data of potentially varying trustworthiness that originates from heterogeneous sources. Heterogeneous data volume and variety create challenges in DW design and implementation. For instance, heterogeneous data can be stored in a centralized on-prem DW, a decentralized DW in the cloud, or a hybrid combination of the two. Generally, data can be stored in structured, semi-structured, or unstructured formats or a combination of the three. A DW can exploit the merging of different types of data stores and optimize their access for specific tasks. For example, heterogeneous DW can involve the integration of different data stores with traditional relational database DWs. Further, a DW can exploit Knowledge Graphs for storing and querying complex structured knowledge and LLMs for unstructured knowledge and capabilities.

The cornerstone of a typical DW is a star schema consisting of a central data fact table connected to one or more dimension tables. In general, the fact table stores accurate, relevant, event-based metrics of business processes like sales revenue at the most atomic level possible. In contrast, dimension tables store rich descriptive attributes that characterize facts for various tasks like grouping, filtering, and drilling down. Each of the dimension tables is, in turn, typically partitioned in a different structure that minimizes storage cost and optimizes access for the specific dimension role it plays in queries for the DW. The overall star schema structure reduces data retrieval costs and optimizes query performance for typical OLAP operations used in Business Intelligence tools. Elucidating, such characteristics of heterogeneous DW design and optimization for effective use within a data ecosystem are critical for stakeholder satisfaction and must be made explicit in data strategies.

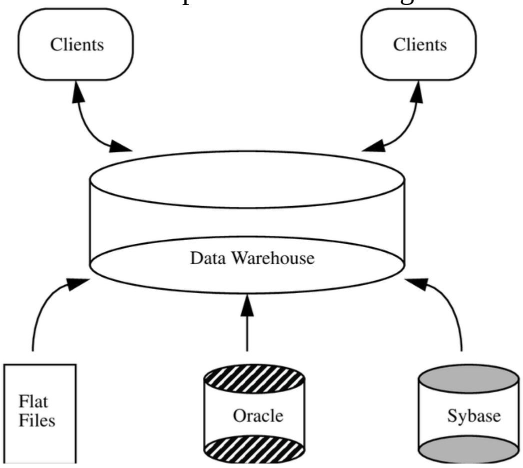

Fig 2: Heterogeneous data is integrated

3. Theoretical Framework

This chapter presents the theoretical framework for Synthetic Cognition models, Cognitive Architectures on which they are based, the Agentic AI models that use those architectures, and the Self-Regulation theoretical concepts that Cognitive Architectures help model. It needs to be noted that the current, most popular Cognitive Architectures are classic Cognitive Architectures that have some limitations in the context of Artificial General Intelligence, like 1. Weak memory; and 2. Weak natural learning capabilities. Cognitive architectures are instantiations of theoretical models of human cognition, made explicit and programmed in a certain computer processing language. They are typically instantiated based on the symbolist OR feedback loops, hybrid, 3D explorations of self, and other grounding-oriented Cognitive Architectures. Psychologists frequently use those Cognitive Architectures together with symbolic/symbol parsimony models and abstract/closed models to help them explore and understand human Cognition and the formation of Cognitive properties, i.e. to non-mechanistically model the emergence of Consciousness. Given that Symbolism offers only discrete models of the mind and Cognition formation, used in Psychology to help human Cognition exploration and modeling, and a Hybrid Cognition-related Cognitive Architecture model to model Natural Agentic Intelligence and a mechanical-oriented model to argue for Artificial Agency, a fact that relies on for their Artificial Consciousness explorations, then there is a well-delineated path for Agentic AI exploration. Therefore, thanks to Cognition being related to Consciousness, acting agency being the main property of Consciousness, Self-Regulation the main property of Cognition, Cognitive Model-Cognitive Architectures models of Cognition being well defined in the Psychology literature, Hybrid, Non-

Mechanistic, Domain Centric approaches being mostly Symbolic, Cognition-Method being plausible, Self-Regulation Mechanisms involve regulation processes or processes originating, sustaining, and inhibiting themselves.

3.1. Cognitive Architectures

The broad field of cognitive architecture has at its core the ambition of creating synthetic intelligence that can emulate specific processes of sophisticated forms of human cognition. Human beings rely on a relatively small number of cognitive architectures to solve problems in vastly different domains; moreover, they learn to make these cognitive architectures function more and more efficiently with experience. This is the expectation that devoted parties pursue with Artificial General Intelligence systems. A cognitive architecture is then a comprehensive computational theory of general intelligence that makes specific predictions both about the innate and learned modules of the human mind, and how they interact; as such, it models how the mind processes information in everyday life. More generally, cognitive architecture is a model of what the human cognitive processes are in the wider meaning of the term ‘cognitive’, embracing perception, action, and subjective experience. Theoretical ideas about cognitive architecture and related experimental ideas are typically developed and tested using the subsystems implementational approach: specific predictions are derived from a modeled cognitive architecture, and then are refined and validated using experimental techniques in cognitive and neural science.

Cognitive architecture research is very much in the hands of action, perception, and robotics researchers; this is perhaps surprising, since the main areas of practical application of artificial general intelligence, viz. data processing and social robotics, fall into different domains. Perhaps the answer lies in the very comprehensive and integrative nature of cognitive architecture research, with its in-depth studies of single research questions addressing critical aspects of the predictive mechanism of cognitive architectures. One-step envisioning domains are still used in machine learning: for instance, some research branches, like systems for unbeaten board games, are based on maximizing the expected outcome of playing one extra step from the current state. A related idea concerns enhancing suspending judgment systems with external visual planning functions in a cognitive architecture context, over time learning a model able to guide the allocation of working memory resources to execution or simply storage of elements of a visual plan.

3.2. Agentic AI Models

The term 'agentic AI models' refers to synthetic cognitive models that already evidence (1) AI-like, human-competitive thinking and self-regulation skills; and (2) AI-like, human-competitive motivation systems. These models were not deliberately designed to possess such skills, or motivation systems. They merely emerged as a consequence of the self-organization properties of large, frozen, non-differentiable neural networks that relied on a large amount of frozen physical resources, such as computation power, energy, and an astonishing amount of training data, to self-organize their sensorimotor activity in a way that enabled solving complex and challenging tasks. For a model to be considered 'agentic' its meta-cognitive agents (MA) must score higher than a preset threshold; an MA in an agentic AI model must also be able to stand at least for a small amount of time a model talks 'I' or 'Me', represent and store its past experiences, remember these experiences and use retrieval methods to couple stored information relevant to the task at hand, be able to choose between several possible answers, solutions or plans, and detect new experiences that contradict its beliefs and estimates, and either properly reason to adjust its subjective model of the world or else reject, for some reason, its inner sense of truthfulness. Agentic AIs must also be able to generate their thoughts and behavior, meaning they cannot rely on clever programming that rules their responses to otherwise novel challenges, which allows Level 5 models to at least temporarily monopolize the role of a producer of new models, and consequently acquire the Level 6 status in a short time.

Considering AGI as a model that has passed some form of Turing test with a human thought interacting with it recently, or that is capable of reasoning about certain fundamental problems or has achieved an inspiring research breakthrough, the heuristic used to classify them quantifies L(M) and not the actual competence level of M.

3.3. Self-Regulation Mechanisms

Self-regulation mechanisms of intelligent insight processing in its various activations scheme can realize functions of different purpose priority with a hierarchy of their activation potentials. Hierarchical ordering occurs concerning dimensions of clarity, cognitive load, momentum of prior achievements in analytical insight-giving activity, and affect stimulation. Other modifications determine the peculiarities of the functional preference of different activated subsystems, being characterized by features of operable convenience and disposition towards changed situational conditions. The family of modified realization parameters achieves moment stability in functioning, ordered by increased task difficulty, favorable situational support, and controlled by temp increments of emotional activity. Thus, volumes of processed information move into a permissible exploitation zone. Dynamic adjustments arise on account of consideration of operative probabilistic success estimates of previous adjusting attempts relative to actualized variations in changing possibility flows. This type of feedback is realized with input-output temporal form delays. Their adjusted and termed parameters match flow time shifting caused by short-term temporal zone influences on input stream distribution. Such a mechanism underlies micrometrics of information processing effectiveness. Its properties allow regulation of momentary flexibility and insertion of dynamic realizations concerning eventual volumetric peaks of exerted operational difficulty during insight processing strategy transition moments. At the same time, working conditions copying input probable distribution adjusted zone properties make it possible to provide corresponding operative adjustments of insight processing strategy. Simulation provided insight processing simulated model with microfeedback. It allowed realizable returns relative to estimated insight processing effectiveness to provide functional stability in natural gaming experiments with models of insight strategy.

4. Methodology

In this chapter, we present the methodology utilized in this empirical research to achieve the stated goal of investigating how and in which moment data warehouses are created through the lenses of Cognitive Work Analysis and Mental Trace Theory. The section is divided into three subsections; the first presents the design of the research, predominantly qualitative, and the data collection techniques. The second details the three different analytical methods used to analyze the collected data, which varied depending on the type of data and addressed primarily the elicitation of dark knowledge. These methods seek to enable the exploration of the knowledge and sensemaking processes of analysts while they were creating the data warehouses by decomposing decision instances for several star systems, expanding them with mental traces, and elaborating description tables with information about the data collection, sensemaking, and data product decisions.

Research design: Data for this research was collected through face-to-face interviews, stored local volunteer-provided documents, a remote shadowing session, and phone calls across different media. The first step was to create contacts and recruit interviewees, which was accomplished through the use of personal contacts and social networks. Snowball sampling was used; if needed, some potential interviewees were approached directly. Thereafter, a semi-structured interview script was prepared individually for each analyst or analysis team depending on the data collection techniques and details requested from the analyst. The main purpose of the interviews was to elicit and expand the information already collected and clarify some details. In addition, to initiate the elicitation of dark knowledge for known systems, the interviewees were requested to recapitulate the process of creating the data warehouse by focusing on the analysis of that system. Different interview scripts were developed for each analyst or analysis team depending on the details present in each document or the response to each phone call. Moreover, specific phone calls took place before some interviews so the analyst could prepare any required information for the discussion. The interviews typically took one to two hours.

4.1. Research Design

The value of advances in data warehousing is realized through employability in different application domains. The role of a data warehouse is to serve as a reporting platform for business activities and non-business applications. The fundamental characteristic of a data warehouse is that with the designated business or application domain it should contain a collection of its data and respond to a query on the data in the query language in a user-accessible and efficient manner. Any non-application domain and non-domain data must not be contained by the data warehouse. Such data can be sought from the other stores to respond to a query on it. The area selected for this preliminary survey study is cognitive operations which find valuable applications in several domains. The data warehouse selected for this study is a 5D cube encompassing data on popular users, Q-lexicons, concept functions, events, and cognitive operators. The cube stores data on cognitive operators that are modeled by taking their definitions used in different domains. It is one of the identified solutions to address the issue of knowledge qualification in information retrieval.

Emerging agentic AI agents are non-living, non-human cognitive agents capable of autonomously performing a multitude of tasks currently requiring human intelligence, and creatively developing and learning new paradigms of intelligent behavior or cognition. These tasks include information retrieval and knowledge qualification. We believe that the data warehouse can be of great assistance in storing, maintaining, and responding to such a demand. This paper further highlights the mechanism of building cognitive agent AI with the support of the discussed data warehouse and the involvement of large-scale transformer models. Synthetic cognitions plan and persistently evaluate plans in world models and assist cognitive operations by developing models as a set of parameters.

4.2. Data Collection Techniques

Data collection is a vital component of research, helping researchers to gain understanding and knowledge. There are numerous techniques available for this purpose. These include the qualitative methods of interviews, focus groups, and ethnographic studies; self-completion and interviewer-assisted surveys; and the quantitative techniques of observation and measurement. Distinctions may also be drawn between secondary techniques, in which the researcher uses materials collected for other purposes, and primary techniques involving data collected by researchers or consultants specifically for the project in hand.

Data collection can also take a range of different forms. It may be formal or informal. It may involve qualitative or in-depth questioning or relatively superficial inquiries. It may involve extensive use of probes, in which each issue is pursued in detail, or maybe more shallow and developed. It may be self-completion or may involve the researcher interviewing in either a structured, semi-structured, or unstructured manner. Finally, it may focus on the respondents' perceptions and feelings or may concentrate on facts and behaviors. In practice, a blend of techniques is most commonly used, utilizing a variety of approaches throughout the project, across different data sources. Aside from addressing a variety of different audiences, the advantages of such an approach lie in its avoidance of the weaknesses associated with particular data collection techniques, the ability to capitalize on the strengths of each, and the support afforded by multiple sources of data to the validity of the conclusions drawn. Different types of data will also provide evidence of different issues under investigation. In particular, different data types may provide evidence of discrepancy or convergence, and researcher and participant viewpoints may complement each other.

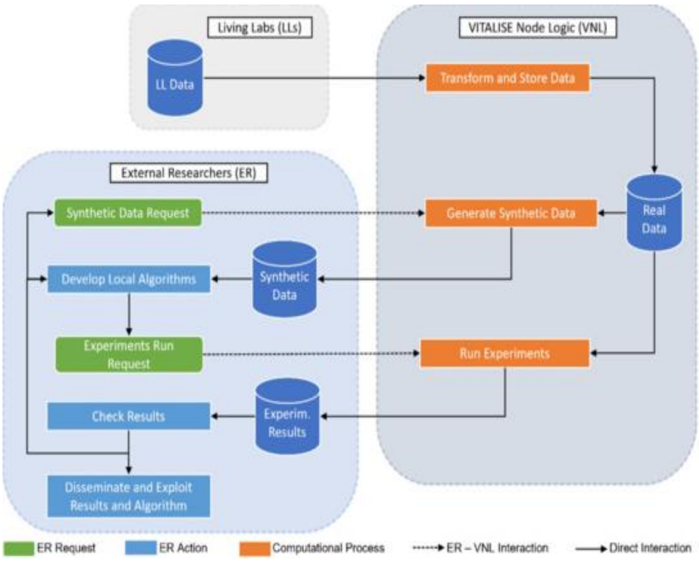

Fig 3: Incorporation of Synthetic Data Generation Techniques

4.3. Analytical Methods

Two analytical methods were used: thematic analysis and thematic discourse analysis. Regarding the first method, a technique was used: the transcripts were repeatedly read and codes were generated and gradually refined, then the themes were aggregated until the data were sufficiently organized to attract the final interpretation. The discussion was used to account for alternative interpretations and to reach a consensus regarding the decisions taken at each step. The coded and themed data were presented during elaborating workshop sessions to enhance accuracy and exhaustivity regarding the interpretation of the participants' accounts and the proposed findings. Next, the meaningful accounts were extracted from the data and were hierarchically organized in groups according to the investigators' interpretation of their meanings. This process led to the identification of four groups of phenomena influencing the response of Data Warehousing to current data deluge challenges: 1. Goals, timelines, and human resources; 2. DW data management practices and tools; 3. DW development technologies; and 4. Larger data ecosystem methods and infrastructures. Based on these findings, we propose recommendations for CDOs and DW managers to re-configure the DW for it to become one of the foundational components of the digital hub.

The second method was then used to extract the DW acceptance discourse structure from the meanings embedded in the participants' accounts of these four groups of phenomena. The use of thematic discourse analysis enables capturing dominant texts regarding the knowledge of a certain subject. Such an understanding becomes critical when tacit issues are at stake – in our case, the governance of new AI technologies and the question of who decides which technologies will be used to accumulate and structure information that will most likely be essential for the digital transformations of organizations from now on.

5. Architecting Agentic AI Models

Discussions in this section focus on how an abstract cognitive architecture can be instantiated as an agentic intelligent core of an AI model allowing generic capabilities for analytics and inferencing to be directly embedded in a foundation model. With the example in mind, we will discuss what model implementations might look like. We will explore how knowledge graphs can inform and influence how an intelligent core is architected and integrated into foundation models. The organizational context for generative and inferential analytics capabilities focused on business workflows and user activities.

The goal of this section is to show that the intelligence core is quite distinct from other modules of a foundation model, like context encoders or physical signal embedding for vision or speech. These blocks are specialized, constrained functions that do not support inference or queryability. Composition and interaction between different function blocks at the model scale create situations that are much richer than any individual part can accomplish. Loosely, we follow the idea of a "zone of architecture" for AI models. We argue that agentic, synthetic cognition models operate in an intermediate space between generality, richness, and specialization with explicit knowledge bases and functional architecture components, embedding specialized but generic inference, planning, and execution functions.

$$CPLI = \frac{\sum_{j=1}^{m}(C_j \cdot W_j)}{T_a}$$

where:

- $CPLI$ = AI agent's real-time cognitive load
- C_j = Complexity of task j (e.g., inference, alignment)
- W_j = Weight of task priority
- T_a = Total active processing time
- m = Number of concurrent reasoning tasks

Equation 2: Cognitive Processing Load Index (CPLI):

5.1. Model Frameworks

Architecting Knowledge-Enhanced Model Frameworks for Scaling Understanding in Knowledge-constrained Tasks of all Types. It is important to consider which capabilities and functionalities would be considered necessary for scaling. Current multimodal model frameworks appear to be scarce in their perceived flexibility, not accounting for the complexity of enabling diverse Agentic functionality combinations at expected performance levels. More conformal modeling efforts can be observed alongside the rapid progression of base understanding capabilities and particular specialized task performers. In line with the paradigm shifts informed by focus and aptitude in parallel processing for scaling, humans combine many specialized components to perform tasks, adding expertise as roles develop. These aspects appear to be implemented in many structures at smaller scales, as companies continue to taskize output from foundation models. Providing strong capability in necessary higher capacities on demand through Capacity and Competence Scaling is fundamental to enabling Agentic functionality across all types of tasks.

Current abstraction layers over latent space modifications remain limited to image style modifications, conditional image generation for adaptation, and the possibility for additional transformations. Exploring and incorporating additional transforms could be beneficial to the task of rapid on-demand specialization of focused modalities. Specialized modalities within particular combinations of data streams at lower cost seem like one of the optimal avenues to Agentically enable multimodal functionality across all common categories of agentic AI tasks. An ongoing exploration into low-cost-act and instrument-prompt or low-cost-prompt additional sets for the successful utility of image description task capabilities in image captioning. The proposed interpretation of blueprint-like designs is a useful perspective abstraction for guiding combinations of components across task modalities and is well supported by the current understanding of structures.

5.2. Integration with Knowledge Graphs

While the base architectural idea behind agentic AI activities is straightforward, integrating all components into a cohesive large-scale whole is a difficult problem. It is important to point out that agentic AI is not synonymous with autonomous agents. Architecting agentic AI models requires a layered, tiered architecture that integrates heterogeneous and hierarchical components. Simple autoregressive coding alone creates unrealistically simple world models that cannot ground any complex LLM function in the real world. One goal of this document is to sketch out a methodology for informationally-saturated coding-available agentic AI architecture.

Some components are too high-level to be bootstrapped from language, which lacks specific supervisory information for creating lower-level world representations. This is especially true for event-to-event memory like knowledge graphs. Some high-level modeling tasks about intelligence and agency can only be made achievable by the lower-level world modeling components inside a fully integrated cognitive architecture. Combining existing LLMs with world modeling systems in a bidirectional, interactive way using the principles of intrusive thinking, progressive referencing, and analogic MTL, would be one step towards creating an Integrated Synthetic Cognition Architecture.

Connected knowledge graphs can become the 'inner voice' that can help powerful LLM transform their inner selves towards Intelligent Life without substantial infrastructure. They represent specialized systems of higher-order words, paths, and nodes concentrating information in a small volume. Large LLMs embedded within a bidirectional integrated framework would have their latent self-management and subgoal-management functions strongly optimized in the manner of retrofitted systems. Down through time conditioning has been used for building committed partners in interpersonal psychodynamic relationships.

5.3. Scalability Considerations

When applied previously, inference engines and associated knowledge bases have had difficulty scaling to real-world application problems. Data-driven models have often not been understood in context or with sufficient semantics to drive socially sensitive applications. As a consequence, the balance of capability across these types of AI systems has been substantially lopsided. The advent of large language models with billions of parameters has begun to tip that equation by presenting data-driven AI systems that not only can engage in socially sensitive conversations but can signal sufficient autonomous agency to provide users with a new support layer in digitally mediated tasks. However, these emergent capabilities are an intellectual curiosity at present more than a practical tool, and the data-intensive computations required to implement them at scale will require the kind of infrastructure tied to diligent security postures if this class of tools is to present useful value to organizations. Currently, these models consume significant compute resources and need either massive optimization on both the model and the implementation sides or some very new hardware being manufactured and made available in very high-enough quantities. For these reasons, such models will be used most effectively as tools — toolkits used to customize or map to the more specialized tasks the organization's need for which they can then be finetuned or adapted or customized with less compute-intensive models that can more easily employ other hardware approaches.

6. Self-Regulating Knowledge Graphs

With the advent of synthetic cognition, we expect Knowledge Graphs (KGs) to represent understanding that goes beyond static, explicit relationships. When we build multilayered cognitive systems to house and cultivate synthetic cognition, KGs become repositories of not just factual knowledge, but also complex and contextualized relations between concepts. Parts of KGs will therefore encode patterns of understanding that are similar to intuition and belief, including the unexpected correlations that ever-growing troves of diverse data can reveal. By bridging these layers of different semantic nature – facts and patterns – we can move closer to building KGs that exhibit a level of cognitive intelligence similar to those of biological beings, and that can support the pathways of human-like innate understanding, such as reasoning and learning by analogy.

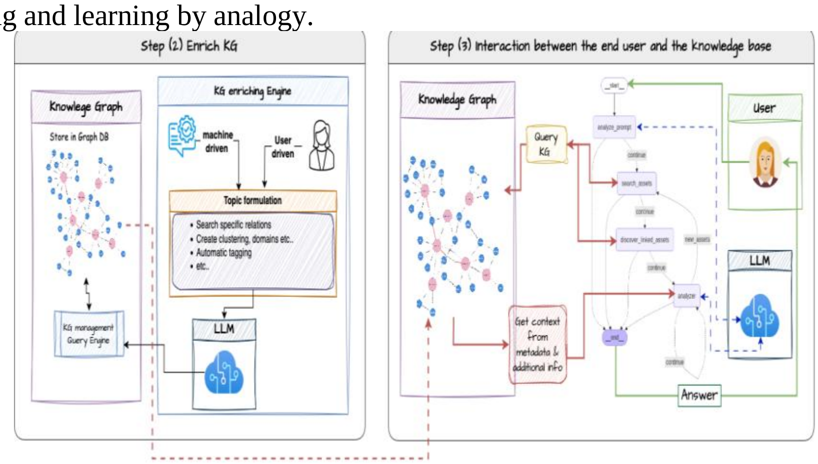

Fig 4: An implementation of an Agentic AI framework powered by a Knowledge Graph

Because synthetic cognitive agents – computing machines and natural or social entities – continuously interact with large amounts of data, whether structured or unstructured, it becomes indispensable to create self-regulating KGs that evolve and adapt as more data or novel uses become available. Dynamic adaptation can occur as these reasons, analogies, or explanations are used in cognitive tasks guided or verified by data inputs, involving either unsupervised or supervised learning pathways. Feedback mechanisms are thereby established between the KGs and auxiliary sensory or knowledge structures, or between different partners in a cognitive partnership, and structured by rules and ontologies that represent mutual understanding or expectations. Such mechanisms may also harness other KGs accessible to the cognitive agents, whether autonomously built through automated data analysis, or previously curated by human-operated or supervised processes.

6.1. Dynamic Adaptation

An initial design of a knowledge representation, such as an ontology or formal conceptual model, is unable to capture and support knowledge of a domain in a complete manner and without errors. Such designs are inherently incomplete and possibly inconsistent. Furthermore, it may become necessary over time to capture novel facts, relations, attributes, classes, and constraints in the knowledge representation because they uncover, store, and provide access to more knowledge of the domain. It is now common in a variety of sectors, such as personalized search engines, recommendation systems, question-and-answer systems, virtual assistants, eLearning, semantic-enhanced social networks, or repositories of academic knowledge, to model, store, visualize, or curate knowledge of domain(s) as knowledge graphs. Domain knowledge changes over time and knowledge graphs are often revised and later re-imported. However, knowledge graphs are usually not truly dynamic. Advanced methodologies utilizing Machine Learning are available to automate processes to some extent. It is necessary to analyze incoming data manually – or with the help of Machine Learning models – and to update the graph as new information, concepts, or relations come in. Still, missing facts, hidden interrelations, and data inconsistency are common.

Dynamic Knowledge Graph Updates enable obtaining new insights and making new interactions easier by acting on updated Knowledge Graphs. Applied to knowledge graph exploration systems, such approaches allow to correctly answer queries about new information contained in the underlying databases and Knowledge Graphs that were not previously reachable or incorrect. So, to realize this, Information Retrieval from Knowledge Graphs needs to be dynamic. Interactive processes between Users and Information Retrieval systems feed data to be formalized in the Knowledge Graphs or that act on Knowledge Graphs leading to their update. This is currently done without Knowledge Graphs automatically detecting an update necessity according to novel incoming data.

6.2. Feedback Mechanisms

Documents, words, terms, and concepts appear and disappear in the course of history. A self-regulating KGraph will be able to assess the importance of certain concepts over time and either prematurely prune or hastily create additional KGraph nodes whenever these concepts take on great and sudden importance. The phenomena of seasonal fluctuations in the economy and punctuated equilibrium in the evolution of species are examples of how systems fluctuate between states of apparent stability and states of intense change. The KGraph must be incremental and dynamic to accumulate and organize knowledge.

Incremental KGraph knowledge creation and organization revolves around data stored in a corpus of external data sources. It is the dynamic discovery of significance that makes a KGraph into a knowledge tool rather than merely a data repository. To grow and facilitate the support of knowledge processes, a dynamic KGraph will periodically check for the existence of new data, usually in an automated and unattended fashion. The frequency with which such checks are conducted generally depends on certain characteristics of the documents being monitored. That is, the rate of change for documents must be considered in selecting a check frequency. A slow change rate may give rise to long delays in locating important updates, while a rapid change rate may cause a KGraph to be prematurely pruned. Documents that have historically gone long intervals without change at one point may also be updated frequently for brief periods. A KGraph will therefore need to heuristically adapt to the changing behavior of different documents over time. User feedback – via implicit methods and explicit preference modeling – is extremely valuable to assist a KGraph in becoming dynamic. Implicit KGraph feedback is intrinsic in weblogging and concept sensitivity analysis and can be used as a substitute for more expensive explicit feedback in situations where users are loath to provide explicit input to the KGraph guidance process. Explicit preference modeling uses user-input annotations and other related user-assistance techniques as seeds in a much wider horizon and describes the overarching context within which the KGraph is operating.

6.3. Autonomous Updates

A key self-regulating property of Agentic KGs is autonomous termination, a process by which the KH detects system goals have been met and terminates updates when those updates are no longer required. This allows it to act independently with minimal intervention from the end user and confirms the original definition of the knowledge graph as a way to "automatically gather and maintain a coherent and comprehensive view" of a collection of data. In practice, updates should be auto-regulatory: a decision of whether to add, remove, or change an object, predicate, or relation in the KG should depend on the current state of the KG and the action's predicted consequences. Such changes might incur a dynamic cost incurred through the need for downstream applications to change or be retrained.

Updates that are triggered because a new data point violates an existing rule are a standard type of autonomous termination. Typically the knowledge graph ceases to inform application performance after a certain point in time or as a certain accuracy is reached. Beyond that point, adding new knowledge is costly. Knowledge graphs can have many different ways to assess the performance of the application they support, each of which could be used to make update suggestions. These include point-wise probability of the oracles, task error, loss concerning the oracles, distance-to-data, downstream influence as estimated by a gradient or surrogate model, task accuracy, and so on. Performance metrics are often predicated on latent factor models that work around the non-linearities arising from discrete but variable metrics.

7. Challenges and Limitations

Despite the growing interest in synthetic cognition and agentic AI, there are still many challenges to be overcome. The most pressing are discussed in this section, to help researchers develop a more realistic expectation of how close we truly are to achieving synthetic cognition and the socio-ethical implications thereof. While insights from psychology and neuroscience have greatly informed the development of synthetic cognition models, there are still many aspects of human cognition and intelligence that are still poorly understood at a mechanistic level. Synthetic cognition also doesn't lend itself equally well to all underlying AI paradigms. Consider for example a transformer-based language model that's trained on human output but isn't aligned with human intent. Such a model wouldn't satisfy any of the existing theories of cognition, despite the model being able to produce human-like output.

The heterogeneity of the data produced by human minds, especially in terms of the human experiences or mental states that shape the way we process different pieces of information, needs to be taken into account by researchers working on synthetic cognition. By heterogeneity, we refer to several concepts within cognition and neuroscience, including individual differences, modality-specific features, and emotion-specific features. To allow for the human-like ability to transfer knowledge from one task, domain, or experience to completely different ones, a considerably larger dataset that sufficiently covers the above aspects in terms of scope and scale is needed to train models of synthetic cognition.

There are also issues related to the resulting computational complexity when exploring synthetic cognition. There are currently no plausible models of human cognition that can accurately predict human-like output on language or vision tasks in terms of speed and efficiency without requiring vast amounts of computation. At the same time, the human cognitive output displays remarkable efficiency and ability to adapt and buttress our nervous systems against the computational challenges posed by the world that we live in, one that needs to be ideally explored and explained. Since human cognition is the starting point for building agentic AI, we need to build models of cognition that are not only efficient but capable of dealing with the computationally challenging task of solving open-ended problems.

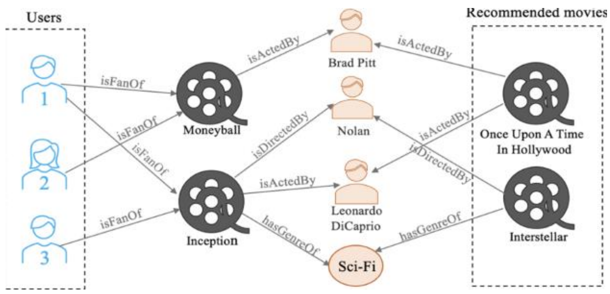

Fig 5: Knowledge Graphs: Opportunities and Challenges

7.1. Data Heterogeneity

In synthetic cognition, the perception mechanism plays a pivotal role since it is the first stage of processing synthetic sensory inputs. For its execution, the system must be able to deal with data objects of any nature and type, including multiscale, multimodal, and multi-structured objects. This entails significant technical difficulties due to the heterogeneity of data in terms of volume, velocity, variety, and variability. The volume of certain types of data has attained extreme proportions. Huge datasets are generated every day from sources and other sensors, communications and transactions, and social networks. The velocity of other types of data has also increased dramatically. Billions of messages, images, videos, posts, and comments are published every day on social networks and other digital platforms. Volumes of transactions are operated every day in exchanges and other marketplaces. These collections of big data that are produced not only have a significant impact on the difficulty of processing but also elevate the importance of supporting decisions based on reliable analyses.

Due to the diversity of the sources, data in information systems and data warehouses, as well as the results of data-generated processes executed by enterprises, can come in any type of format and structure. It is usual to find data composed of multiscale and multi-structured images, audio signals of various contacts and formats, sensor signals, financial, economic, and social records, messages, geopolitical, meteorological, and other indicators, graphs and charts, and central and noncentral documents. Databases maintain data using different models: relational or multi-relational databases handle tabular data; NoSQL databases allow a flexible structure for data expressed in documents; graph databases operate data stored in graphs; temporal databases keep temporal data; object-oriented databases study data modeled as objects; XML databases store XML data, and spatial databases manage spatial data. Data may not only come from diverse models but also from different data sources.

7.2. Computational Complexity

The history of scientific discovery is a history of progressively more complex computations, created to reduce synthetic cognition's inherent massive computational complexity. Ptolemy's epicycles of planetary motion, Kepler's elliptical planetary orbits, Lagrange's differential motion in three-part systems, and Einstein's non-Euclidean descriptions of relativistic gravitational motion are all manifestations of applied computing progress, as are Newtonian calculus and Maxwellian differential equations. Historically, such a description of heightening scientific complexity begins with the fact that our world is a physical system built from particles that must be proven to exist upon witnessing their first-principles particle motion derivations—producing a basis of analog, or non-integral, numerical physical modeling—as distinct from a basis of elementary particle creation/destruction.

Many of the challenges we face in inferring knowledge from big data in data warehousing lie in this heightening world complexity, especially in its irrational and chaotic components. Proofs of rising analytic difficulty are the justification for the Big Data revolution itself: Data are being stored and analyzed primarily to first reach definitional knowledge statements—other, related complexity theorems—curbing analytical complexity growth. Because all complexity is ultimately defined recursively, recursively defining more of the functions we try to infer, and inferring their Taylor and other series coefficients for extremely-high-dimensional cases, is a big data approach. These series, generalized moments, are analytic tile expansions of elementary, integral, analytic equations, whose analytic basis consists of elementary integrals across simpler, analytic systems with more finite, integer dimensions. Other alternatives, like synthetic cognition, are qualitatively less efficient at addressing problems of heightening world data complexity.

7.3. Ethical Considerations

With the rapid development of both synthetic cognition and agentic AI, it is important to consider the ethical implications of their employment, especially regarding data collection and usage. The worries encompassing synthetic cognition use are most notably related to the validity of the inputs, especially if the unseen biases present in data might be learned and inappropriately reinforced by the AI models. These biases might stem not only from the data used for initial model training but also from the domain-representative data used during the fine-tuning and prompting stages used for model adaptation to subtasks. Also, it can be difficult to disentangle the potential effects of model fine-tuning from the effects that a fine-tuned model will have on the outputs it generates. Residual trace embeddings of prior outputs remain in the model, potentially affecting future outputs upon giving similar prompts. Furthermore, if the model is used to synthetically generate or augment data for downstream training tasks, latent biases in the model can cause unforeseen effects of data distribution shifts, potentially affecting the decision boundaries of the trained models.

Another ethical dilemma is whether it is appropriate to employ synthetic cognition and agentic AI to augment research findings, especially when the AI artifacts are kept as black boxes, and the research findings are going through peer review. Furthermore, biases originating from different levels in the data collection pipeline can cause ethical whirlpools. In a way, agentic AI takes this concern to a greater level since there is generally more autonomy in its operation. That said, the AI agent is intended to be employed by a human for user-directed goals, to extend the uptime and data capture capabilities of the human during which the user is operating the AI agent. Even then, without human oversight, it is reasonable to assume that issues described above would have a greater chance of happening with agentic than synthetic cognition.

8. Case Studies

In this chapter, we present case studies to demonstrate the generality of the approaches we have presented in the context of agentic AI research about synthetic cognition afforded by data warehousing and business intelligence. Section 8.1 provides industry applications related to the data warehousing concepts discussed in this work. These examples highlight some of the numerous practical agentic AI applications in the world at this moment. Then we will do an analytical discussion of the results obtained by exploring the model space related to pre-trained models available in the machine learning platform.

A summary of some of the currently available data warehouse models, together with their application characteristics or intended uses, appears in a table. These models come from a diversity of research institutions and companies, showing the interest in and contributions made to agentic AI across the globe. The table is organized in columns by model source name and in rows by agentic AI functions. The columns summarize some of the results concerning (i) speech, language image training multi-modal tasking — Semantics, Computer Vision, Speech recognition and image recognition; (ii) context length modeling; (iii) task execution modularity and control — agentic personality modeling, coordination policies; (iv) reason using planning — non-KB, KB explicit and lucid; (v) memory, — memory usage, active memory access; (vi) experience phylogeny — interaction and experience algorithm; and (vii) action capabilities — vision function, action execution multimodal capabilities; and (viii) ownership and ethical guidelines; information safety — access and trust. The reasons for the inclusion of particular models are (a) the listing contains models with a variety of intended tasks; (b) the diverse sources represent the global development of Artificial Intelligence; and (c) the selected models are representative of the model space characteristic diversity defined in the previous sections.

8.1. Industry Applications

The competitive advantages offered by artificial cognition point to transformational innovations in many other sectors, as smart operations discover ever more new ways to be radically more efficient, effective, faster, and cheaper. Innovations based on utilizing the capabilities of synthetic cognition can help in automating analysis. Computer algorithms have been used in all industries for decades. But the difference now is that the algorithms utilized are not rigid functions designed explicitly by computer scientists of particular characteristics of behavior in each situation. Rather, the algorithms are themselves built from the ground up through machine or reinforcement learning engine processes that discover and recognize the multidimensional complex statistical interactions between diverse observations and consequent outputs over a huge data processing space formed from past observations and outcomes.

In fact, in data warehouses, and other forms of organizational 'systematic' and 'structured' intelligence utilization strategies, the organization, preparation, and transformation of the observed data are already there in the form of data pre-processing and ETL steps described by data warehouse designers. So humans can usefully embed their judgments about what become intelligent 'features' and 'preconditions' into pre-processing stages, for example, transforming transaction data from operational systems by creating 'process type-invariant' models for a 'process type' to embed additional relevant higher-level information and features as new dimensions. Yet, even companies that focus particular intelligent analytic functions on intelligent 'features' and 'functional adjacencies' aren't even close to exploiting the possibilities.

8.2. Comparative Analysis of Models

Comparative analysis of two different agentic model classes helps to form understanding of approaches towards artificial cognition and the goal of synthetic cognition. Let's start firstly with the strength of the hybrid cognitive system realized as Neural-Symbolic Computing.

Neurosymbolic AI is an artificial intelligence paradigm that combines neural networks with symbols and symbolic reasoning. A range of benefits arise: neural networks provide perception capabilities that do not require hand-tuning and can be optimized while symbol processing enables explicit representations that facilitate planning and accelerated higher level reasoning.

Neural-Symbolic Computing offers layered design to knowledge processing that enables exploiting the strengths of both component classes. Low-level information is processed and represented efficiently and optimally by the neural mechanism that does not rely on symbolic input. Explanatory familiarity representations and structure manipulation are executed by that separately optimized symbolic mechanism. Rich unified coupling operation enables interactive communication. Easy-to-implement connection links to external symbols and symbol structures are shared by both symbolic and neural cues.

On the other hand, neural-symbolic cognitive systems do pass comparison with complex dynamic models of cognitive action. Let's speak about the qualities of the Cognitive Agent-based Model of AI. CAAI reflects cognitive concepts of data sense-making, feedback-guided perception, and artificial subliminal memory. CAAI employs partial interaction between inner agent-like and external environment-like model classes combined in one structure. CAAI employs hidden agent-like models, capable of participating in joint processes of differential cognition of agents or perception and subliminal memory. It considers cognitive action as a hierarchical model-structure transformation process governed by a meta-cognitive procedure of joint data transformation and parameter estimation and controlled globally by feedback of emotions and motivations.

9. Future Directions

The potential for Synthetic Cognition and Agentic AI, when combined with novel and emerging technologies – such as Generative AI, AI, and ML-enhanced Data Warehousing capabilities, Deep Learning, the Internet of Things, the edge, the Metaverse, Efficient Smart Systems, Living Machines, Spatial Computing, Artificial Semantic Worlds, Quantum Computing, Neuromorphic Computing, Advanced MapReduce, or Advances in Biological Cognition – to do our cognitive work for us or much better than we could is a logical and consequential extension of humanity's two-century long pursuit of increased efficiency in productivity, from the Industrial Revolution onwards. This seems to be the logical conclusion of the Agency in AI and Hypotheses about Cognitive Reasoning adequacy in AI, or at least we should strive for that.

As in any type of research, there are brief lapses or gaps between actual theory before it morphs into mature applications and ubiquitous use. In the area of Data Warehousing, we have yet to see the emergence of truly SC & A AI-enabled applications that can do what SC & A AI proposes to enable. It is tempting to think that the holy grail of Data Warehousing is to free us from mundane worries about how to best apply today's tired-old collection of Managed Data Warehouse tools to solve increasingly simpler hybrid use cases. These few use cases seem designed primarily to justify the Data Industry estimate that the Data Economy will grow significantly and that Data Analytics will account for a notable portion of the Data Source Market as if that growth has any implications for understanding the Cognitive Computing Transformation and the role of SC & A AI in Data Warehousing.

9.1. Emerging Technologies
We have examined how organizations are integrating AI into their data warehousing environments, and have outlined a vision of future data warehousing systems that are guided both in their operation and outcomes by advanced Agentic AI. Agentic AI will seek to autonomously achieve goals set by organizations, thus providing a user-centered focus to enhance the effectiveness of data warehousing and other supporting technologies, processes, and organizational coordination. In support of this vision, we outline technological developments in AI and its integration into enterprise computing systems that are in progress and/or emerging, that we view as potentially playing a role in shaping future data warehousing technology.
Advanced Pre-Trained Large Language and Vision Models: In the past 2 years, DAOs have been greatly enhanced by the development of advanced large language models, particularly those integrated with vision capabilities. These large neural nets have been trained on large and diverse corpora of human knowledge, language, and behavior, and can converse in natural language, help users find and summarize data, assist in task, code, and text completion, and make logical inferences, among many other applications. With careful design, LLMs are customizable and able to learn on the fly on a task-by-task basis, making them effectively more agentic. Agentic LLMs can produce high-quality outputs that meet user specifications across a range of tasks as varied as, how to plan a wedding, create a knitting pattern, write a computer code or business plan, or write a poem in the style of Robert Frost. Other capabilities include code execution, browsing for additional information, handling visual input, and integrating plugins that augment its capabilities. Agentic LLMs are now being integrated into enterprise applications such as data analysis, knowledge management, business process management, project management, customer engagement, sales, marketing, and programming.
9.2. Potential Research Areas
Data warehousing and the surrounding topics have contributed a plethora of aspects, tools, challenges, and applications to generalized AI. This contribution is, however, not fully reciprocal: AI and embedding applications allow for the resolution of numerous challenges in store design, implementation, and management. There has been some prior exploratory work in this space, but many more possible angles, domains, and problems remain open for research. We summarize a brief selection of ideas loosely related to the data warehousing and generalized AI services relationship.
Cognitive graphical models, graph/neural networks, and their related algorithms are still maturing. The research on large, structured knowledge bases has typically focused on those hosted relationally. A data warehouse is a form of a more general knowledge base: It is the defined blueprint for the data available, has the schema and optionally semantic annotations, and, well designed, has the desired informativeness in terms of quality and data component coverage semantics. Algorithms for graph/network learning/tasks, but higher efficiency to scale/mature applying to actual knowledge base schemas or even partial populations, will contribute to both the benchmarks and actual usability of these technologies.
Experience is one of the major influences on agents and LSI results. There is significant industry-related knowledge about how to create and manage data warehouses, both internally and from an external consultancy view. The available experience goes, however, quite beyond data warehousing and touches queries, similarity, multi- and encapsulated, multi-task learning recalls… Models focusing on AI, and in particular Cognitive Graphs, should be able to learn from this experience and bootstrap aiding agents, specialized algorithms, and ultimately, generative giant LSI AI.

$$GCS = \frac{1}{1 + \frac{1}{n}\sum_{k=1}^{n} |S_k^{(t)} - S_k^{(t-1)}|}$$

GCS = Degree of convergence in graph state updates

$S_k^{(t)}$ = State of node k at time t

$S_k^{(t-1)}$ = Previous state of node k

n = Total number of graph nodes

Equation 3: Graph Convergence Score (GCS):

10. Conclusion

In this paper, we discuss an extension to virtual data warehousing, exploring how agentic AI and synthetic cognition can transform data warehousing in practice. First, we provide motivational background information on the historical evolution of crypto-currencies, their potential use as an incentive structure for collaborative work, and their mathematical precaution against manipulation and data corruption. Second, we provide background information on agentic AI and synthetic cognition and its application fields spanning robotics and linguistics. We conclude with a brief discussion of the function and the implications of the combination of virtual data warehousing and synthetic cognition for business applications, complemented by some hands-on, benchmarking experience.

Hundreds of millions of images of human work on an astonishing breadth of topics have been posted online. Many reconcile one's leisure time with incentives given by business agencies, backing the maintenance of their chain supply completion and their quality supervision of the products, goods, or services of public utility on a large scale. The ongoing frontier of technology pushes forward the evolution of fun-seeking, living, and creative expression, technology-enabled, seeking out huge amounts of data stored online. Blockchain technology and smart contracts offer a general-purpose, incentive design to organize a new form of augmented collaboration, involving – directly or indirectly human agents giving the tumble – artificial agents, underlying the automation of discussion and exploration above structured data while keeping real humans in the loop of unusual image posting or exploring business-relevant content, captured both at a Cartesian space-time level and via data analytics of complexity level-theory.

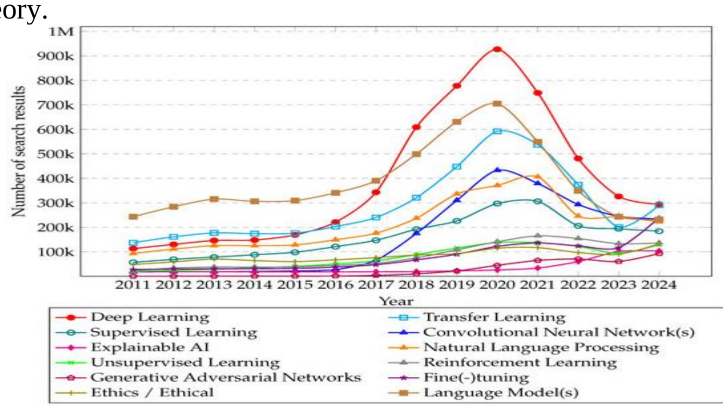

Fig 6: A Survey on Reshaping the Generative Artificial Intelligence

10.1. Summary of Key Findings and Implications

The foundational preamble to this research examined definitions, and listed characteristics of cognition, synthetic cognition, and AI. This work addressed definitional ambiguities by arguing for a functionalist understanding of cognition that distinguishes between attitudes and cognitive abilities; proposing that cognitive, agentic AI can be understood using a suitable degree of astrobiological similarity heuristic. Applying this understanding to Data Warehousing, it was noted that BI professionals are empowering organizations within an epistemic cohort. Achieving the objective of BI practice involves making interventions to effect semantics transfer from unstructured to structured data. This denies those disciplines their 'cognitive monopoly' over semantic machine data enrichment. As the design and implementation of syntactic-semantic enrichment techniques are key challenges in Data Warehousing, these BI-specific techniques would seem to be the most appropriate for organizations with a business focus on 'being data-led'.

There is, however, uncertainty about the extent of competence held by the BI industry regarding effective machine data-machine cognition enrichment. This competence matters because if agentic AI could destroy the artificial, social practices which are causes and conditions for the existence of machine data, their position could become precarious. Because syntactic-semantic enrichment is difficult and time-consuming, and BI professionals service organizations rich enough to be able to afford the cost of human enrichment, it could become the case that agentic AI allows specialized industries with a comparative advantage in Data Warehouses to use lower costs of production in the creation of profitable markets to offer what might be thought of as a processed distilled AI Data Storage Warehousing Service.

References

1. Polineni, T. N. S., & Seenu, A. (2025). The New Frontier of Healthcare and Industry: Subash's Expertise in Big Data and Cloud Computing for Enhanced Operational Efficiency. Cuestiones de Fisioterapia, 54(2), 271-283.

2. Maguluri, K. K., Ganti, V. K. A. T., Yasmeen, Z., & Pandugula, C. (2025, January). Progressive GAN Framework for Realistic Chest X-Ray Synthesis and Data Augmentation. In 2025 6th International Conference on Mobile Computing and Sustainable Informatics (ICMCSI) (pp. 755-760). IEEE.
3. Koppolu, H. K. R. Deep Learning and Agentic AI for Automated Payment Fraud Detection: Enhancing Merchant Services Through Predictive Intelligence.
4. Nampalli, R. C. R., & Adusupalli, B. (2024). AI-Driven Neural Networks for Real-Time Passenger Flow Optimization in High-Speed Rail Networks. Nanotechnology Perceptions, 334-348.
5. Chakilam, C. (2022). Generative AI-Driven Frameworks for Streamlining Patient Education and Treatment Logistics in Complex Healthcare Ecosystems. Kurdish Studies. Green Publication. https://doi. org/10.53555/ks. v10i2, 3719.
6. Sriram, H. K. (2023). Harnessing AI Neural Networks and Generative AI for Advanced Customer Engagement: Insights into Loyalty Programs, Marketing Automation, and Real-Time Analytics. Educational Administration: Theory and Practice, 29(4), 4361-4374.
7. Burugulla, J. K. R. (2025). Enhancing Credit and Charge Card Risk Assessment Through Generative AI and Big Data Analytics: A Novel Approach to Fraud Detection and Consumer Spending Patterns. Cuestiones de Fisioterapia, 54(4), 964-972.
8. Chava K. Dynamic Neural Architectures and AI-Augmented Platforms for Personalized Direct-to-Practitioner Healthcare Engagements. J Neonatal Surg [Internet]. 2025Feb.24 [cited 2025Mar.24];14(4S):501-10. Available from: https://www.jneonatalsurg.com/index.php/jns/article/view/1824
9. Challa, K. (2024). Neural Networks in Inclusive Financial Systems: Generative AI for Bridging the Gap Between Technology and Socioeconomic Equity. MSW Management Journal, 34(2), 749-763.
10. Sondinti, K., & Reddy, L. (2025). The Future of Customer Engagement in Retail Banking: Exploring the Potential of Augmented Reality and Immersive Technologies. Available at SSRN 5136025.
11. Malempati, M., & Rani, P. S. Autonomous AI Ecosystems for Seamless Digital Transactions: Exploring Neural Network-Enhanced Predictive Payment Models.
12. Pallav Kumar Kaulwar. (2023). Tax Optimization and Compliance in Global Business Operations: Analyzing the Challenges and Opportunities of International Taxation Policies and Transfer Pricing. International Journal of Finance (IJFIN) - ABDC Journal Quality List, 36(6), 150-181.
13. Vankayalapati, R. K. (2025). Architectural foundations of hybrid cloud. The Synergy Between Public and Private Clouds in Hybrid Infrastructure Models: Real-World Case Studies and Best Practices, 17.
14. Nuka, S. T. (2025). Leveraging AI and Generative AI for Medical Device Innovation: Enhancing Custom Product Development and Patient Specific Solutions. Journal of Neonatal Surgery, 14(4s).
15. Rao Suura S. Agentic AI Systems in Organ Health Management: Early Detection of Rejection in Transplant Patients. J Neonatal Surg [Internet]. 2025Feb.24 [cited 2025Mar.24];14(4S):490-50.
16. Kannan, S. (2025). Transforming Community Engagement with Generative AI: Harnessing Machine Learning and Neural Networks for Hunger Alleviation and Global Food Security. Cuestiones de Fisioterapia, 54(4), 953-963.
17. Srinivas Kalisetty, D. A. S. Leveraging Artificial Intelligence and Machine Learning for Predictive Bid Analysis in Supply Chain Management: A Data-Driven Approach to Optimize Procurement Strategies.
18. Challa, S. R. Diversification in Investment Portfolios: Evaluating the Performance of Mutual Funds, ETFs, and Fixed Income Securities in Volatile Markets.
19. Vamsee Pamisetty. (2023). Intelligent Financial Governance: The Role of AI and Machine Learning in Enhancing Fiscal Impact Analysis and Budget Forecasting for Government Entities. Journal for ReAttach Therapy and Developmental Diversities, 6(10s(2), 1785–1796. https://doi.org/10.53555/jrtdd.v6i10s(2).3480
20. Komaragiri, V. B. (2022). AI-Driven Maintenance Algorithms For Intelligent Network Systems: Leveraging Neural Networks To Predict And Optimize Performance In Dynamic Environments. Migration Letters, 19, 1949-1964.
21. Annapareddy, V. N., & Rani, P. S. AI and ML Applications in RealTime Energy Monitoring and Optimization for Residential Solar Power Systems.