Computer Fraud and Security
ISSN (online): 1873-7056

Data Engineering in the Age of Large Language Models: Transforming Data Access, Curation, and Enterprise Interpretation

1st Kushvanth Chowdary Nagabhyru

Senior Data Engineer

ORCID ID : 0009-0004-7175-7024

***Abstract*—**For many years, the guiding principle in enterprise data management has been, ”garbage-in, garbage-out”—meaning that the quality of downstream reporting and analyses can only be as good as the quality of the data that enter the system. These words still remain true as organizations struggle to gain insight from their enterprise data. Yet, 2024 is shaping up to be the year of ”garbage-in, garbage-read” for enterprise data interpretation. Large language models (LLMs), such as ChatGPT, have demon- strated an unprecedented capability to convert unstructured text into human-like natural language, answering questions regardless of the source of the input text and summarizing content as part of higher-level reasoning. The impact of LLMs is much broader than natural language processing alone—they affect how organizations curate and access information as well. This article summarizes research about AI techniques that help users get the right data in the right format; automate the evaluation and curation of data; and, finally, apply natural language processing directly to the transformed data to provide enterprise intelligence.

***Index Terms*—**large language models (LLMs), data engineer- ing, AI-driven data access, data curation, enterprise data in- terpretation, business intelligence, data bias, ethical AI, future of data engineering,Large language models (LLMs) represent a critical advance for data engineering.

I. INTRODUCTION

Today, data engineering is one of the most pressing issues. Not only is it the fuel for recent AI breakthroughs, but these new models are also transforming its direction. The introduction of large language models (LLMs) brought much of the underlying AI research into the public’s awareness. They are now being explored for a wide range of AI-driven applications—from writing general business code to suggesting new product ideas. One crucial area that is often overlooked involves addressing the traditional data engineering pipeline: accessing data from centralized repositories, curating the data to meet the necessary specifications, and finally, transforming the data to support an enterprise’s interpretation of that data. Each phase presents its own problems, which can be addressed through generative AI and LLMs, but require different approaches. The section on Enterprise Data Inter- pretation examines the transformations within an enterprise that help transform raw data into concrete business insights, in a style more aligned with the user’s preferences. In 2024, AI continues to reshape the ways enterprises interpret data. Although many organizations might view their data as an asset, without the means to perform profitable analyses or extract

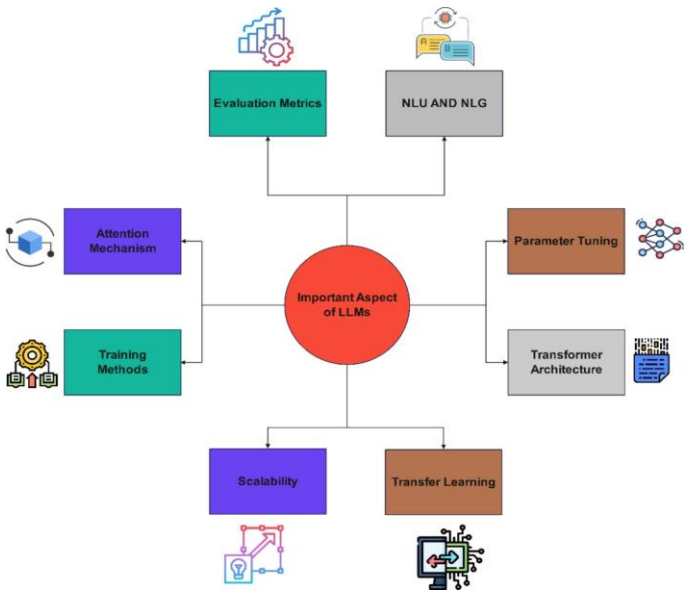

Fig. 1. Data Engineering in the Age of Large Language Models

value, it essentially becomes a liability. Data interpretation addresses this concern by outlining how machine learning and natural language processing techniques can be applied to an enterprise’s data for enhanced learning. Indeed, the

whole pipeline from data access to data cleaning to data interpretation impacts corporations on a regular basis. Therefore, the discus- sion must go beyond the narrow theme of using ChatGPT as a data-retrieval engine and examine AI's implications for data- engineering processes instead.

A. Background and Significance

Much has been said about the capabilities of artificial intelligence (AI) and large language models (LLMs). It therefore might seem futile or redundant to add yet another word on the topic. Instead of reiterating what has already been said many times, it might be more inspiring to consider what AI and LLMs do for data engineering. Looking into this question shows that they do significantly more than what is usually addressed. While much attention has been dedicated to the narrow view of prompting ChatGPT to retrieve data, AI

and LLMs reimagine the entire data-processing pipeline. Data engineers play a significant role in the enterprise. Data is the lifeblood of any organization and therefore must flow freely to where it is needed. Many questions arise: How can data be extracted from its sources in a timely manner? How can data be cleansed, integrated into a repository, and made available? How can the data be transformed to produce answers that are relevant to the business? Ethical questions are paramount too: Is the data retrieval process fair? Are the users authorized to view the requested data? How can a data-analysis process ensure that the policies and procedures developed by the organization are followed? How can the process prevent the inadvertent revelations of sensitive information? All these questions are qualification requirements for intelligent data engineering.The starting point of data engineering is enabling access to data sources. Typically, the data exist in distributed silos across the organization and outside the firewall. Examples of internal data sources include finance data, human resources data, supply-chain data, support desk data, website analytics, and software-development processes. External data include information from government agencies, weather data, and financial-market information. Efficiently accessing these data silos in a timely manner presents a conundrum: It is necessary to know exactly what data to search for, where to look to find the data, and how to evaluate that the returned data answer the business question that is posed.

Equation 01: LLM-Augmented Data Retrieval Efficiency

Objective: quantify recall and "relevant-per-second" efficiency for a top-k retriever with/without LLM reranking.

1. Let the corpus have N docs, with R truly relevant to a query. Let p be the per-item chance a retrieved doc is relevant (baseline) so that recall at k is

$$Rec(k) = 1 - (1 - p)k \quad (1)$$

2. LLM cross-encoder reranking increases effective relevance rate from p to p' ($p' > p$), giving

$$Rec'(k) = 1 - (1 - p')k \quad (2)$$

3. Let t_0 be fixed overhead (fetch + answer synthesis) and t_d the extra per-doc rerank time. Define efficiency as relevant-fraction per second:

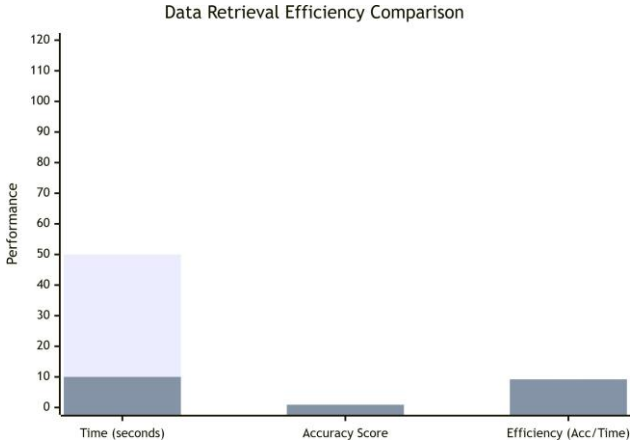

Fig. 2. Data Retrieval Efficiency Comparison

k	Recall_baseline	Recall_rerank	Eff_baseline	Eff_rerank
1	0.048	0.096	0.178	0.35
3	0.137	0.261	0.508	0.926

5	0.218	0.396	0.808	1.366
10	0.389	0.636	1.439	2.05
20	0.626	0.867	2.319	2.478
40	0.86	0.982	3.186	2.285
80	0.98	1.0	3.631	1.694

data on which they are trained. The key feature defining an LLM is that its training data encompasses all compo- nents—vocabularies, grammar, syntax, idioms—necessary to support the vast majority of applications in everyday life. Achieving such comprehensiveness necessitates sourcing train- ing data from an extensive array of diverse and reliable sources, ensuring coverage of a very broad spectrum of topics. This, in turn, leads to impressive model sizes in terms of dimensions and parameters, underpinning their remarkable adaptability. These deep learning neural network models are trained to model the probability distribution of words and word sequences in natural language, a capability typically achieved by analyzing gigantic corpora of natural-language text such as the entire body of Wikipedia, all the records in Project Gutenberg, and—as is the case with ChatGPT—the entire contents of the World Wide Web. Examples of widely used transformer-based large language models include BERT, GPT- 2, GPT-3, GPT-4, and ChatGPT.

A. Definition and Overview

Large language models (LLMs) are enormous artificial neural networks that have been trained on prodigious bodies

quences of language, unpredictable by any directive grammar, suggested immediate application to machine translation. Obvi- ously, the string of text generated in response to a request, or

II. UNDERSTANDING LARGE LANGUAGE MODELS

Large language models (LLMs) represent a specific class of language models characterized by their enormous sizes, not only in terms of dimensions but also in the volume of

prompt, can be used in other ways, such as question answering, provided enough such examples were included in the process of learning to guess the next word. A significant problem in such models, however, is their propensity to "hallucinate,"

generating plausible answers that are completely unfaithful to any known source. As a result, the key methods that are seeing deployment relate models to extant stores of information that can be guaranteed to be consistent, even if partial, and hence usable as sources for answers to questions posed in natural language. These prompt-based extraction and retrieval processes begin to transform the interaction between the user and the underlying data. Rather than a platform provided by the data engineering discipline that makes it easy for the user to access the data, the underlying ecosystem becomes a black box to which a query is posed and an answer is returned as a sequence of text. The conversation then turns to the quality of the data used, and the processes of data quality management, data governance, and master data management become more important than ever if the responses delivered to enterprise questions are to be properly calibrated.

B. Evolution of Language Models

Language models (LMs) are a family of models designed to model natural language. Their early forms were probabilistic models estimating the probability of the next token conditioned on the previous n tokens. With a change of parametrization and a deep neural network architecture, one obtains neural LMs. These have hyperparameters that do not depend on the length of the prior context and have hidden-size representations of the probability distribution for the next token. Translation from the source-language text into a different target-language representation is used as a task with a similarly sized architecture using attention and multiheaded attention. Recurrent neural network-based LMs were modified to obtain nonautoregressive versions.

Although transformer models' architectures do not depend on context size, during training, they use mostly contiguous sequences of previously generated tokens. In out-domain instructions, users may describe their idea, ultimately different from the sequence of tokens seen during training. Bidirectional transformers or encoder-decoder transformers do not make the assumption that previous tokens are from a contiguous sequence during training. The size of the context window and how it is handled varies considerably by architecture. Via a windowing technique, these models have achieved results in windows that are larger than the window size of the initial model.

Equation 02: Data Curation with LLMs (Quality Score)

Objective: combine multiple data-quality dimensions under automation.

1. Dimensions (all in [0,1][0,1]): Accuracy A, Completeness

C, Consistency S, Timeliness T

2. Weighted score

$$Q = w_A \cdot A + w_C \cdot C + w_S \cdot S + w_T \cdot T, \qquad \sum w_\cdot = 1 \quad (4)$$

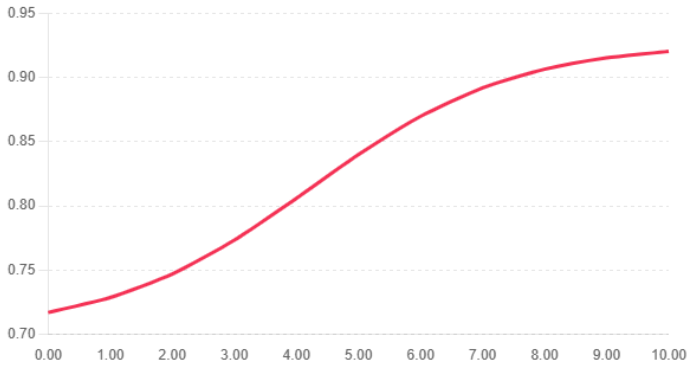

Fig. 3. Data Curation Quality vs Automation

3. Let automation fraction be $f \in [0, 1]$. Empirically each dimension follows an S-curve improvement $D(f) = D_0 +$

$\Delta D\, \sigma(a(f - b))$.

4. Substitute to obtain $D(f)Q(f)$.

III. Data Access in the Era of AI

Effective data access is a fundamental first step in the data-to-insights workflow, as emphasized by veteran enterprise data engineer Khalil. In an earlier era, the challenge lay in creating large volumes of carefully structured data, uniformly cleaned, harmonized, and labeled, eventually stored in a data warehouse. The Data Warehouse team, known for their attention to detail and hard work, created a Home for the Data, ensuring that everyone could access it with confidence and explore it through familiar and consistent lenses. Data Warehouse team members refer to their work as making the data "trusted," while the data consumer calls the data "easy-to-use."

Over time, the sophistication of these warehouses improved, but the relationship between the Data Warehouse team and consumers was largely the same until a decade ago, when the introduction of governance, privacy, and sensitivity rules meant that data consumers had to traverse additional portals, present licenses, and register to access many datasets. Suddenly, what was once so easy became a bit more complicated. However, a small band of intrepid custodians combed through a thousand datasets, tagged each field by its sensitivity and privacy, and summed up the risk in a single number, transforming the universe of accessible data from complicated to simple once again—and just in time for enabling the AI-driven data access journey that will follow.

A. Challenges in Traditional Data Access

Data access faces both technical and practical challenges, a situation scarcely surprising given the long-running demands to ensure enterprises have timely and secure access to data. From the perspective of business users, the conventional methods to retrieve data—often relying on IT support through SQL queries or data cadre-produced reports—can fall short,

especially when queries become more specific or increasingly

automation fraction	accuracy	completeness	consistency	timeliness	quality_scoreQ
0.0	0.766	0.714	0.744	0.609	0.717
0.1	0.78	0.724	0.756	0.617	0.729
0.2	0.802	0.74	0.774	0.633	0.747
0.3	0.833	0.764	0.796	0.659	0.773
0.4	0.867	0.794	0.82	0.7	0.806
0.5	0.898	0.826	0.844	0.75	0.84
0.6	0.92	0.856	0.866	0.8	0.87
0.7	0.934	0.88	0.884	0.841	0.892
0.8	0.942	0.896	0.896	0.867	0.906
0.9	0.946	0.906	0.905	0.883	0.915
1.0	0.948	0.912	0.911	0.891	0.92

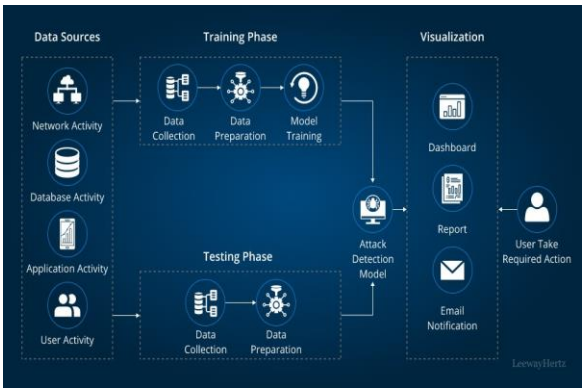

Fig. 4. Data Access in the Era of AI

complex. The need for support is actualized by the technical complexities involved in search and retrieval, where IT staff navigate an intricate landscape of tables, fields, and dags. Although traditional solutions employing search engines like Google, Elasticsearch, or Amazon Kendra exist, they do not necessarily resolve the entire data access domain problem because the creational and operational underpinnings remain challenging.

Several inherent limitations of a traditional search-engine- only approach prompt the development of AI-based methods for data retrieval. The prospect of using LLMs to interrogate data and generate answers has captivated significant interest. Yet, despite the relevance of onboard-model capabilities, cor- rectly framing the context is essential to enable LLMs to search, access, and translate data. This realization gave rise to the approach recognized as Retrieval-Augmented Generation (RAG), which focuses on going beyond sole reliance on an LLM. By enhancing the search capability, RAG opens the

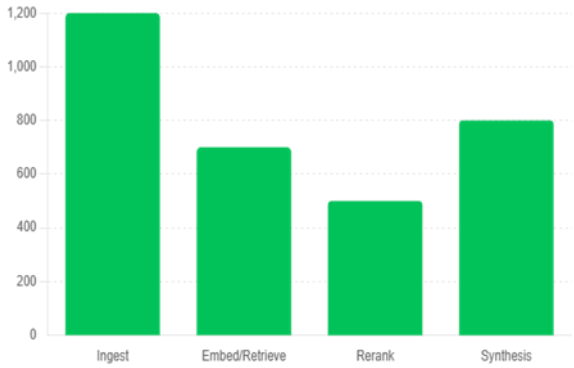

Fig. 5. Pipeline Latency vs Load

and interpret data to make rapid decisions. A semantic and context-based information retrieval system helps business users easily find answers to ad hoc questions about their companies. Business analysts need a seamless experience for retrieving enterprise information from a complex mesh of processes, SOPs, manuals, documents, and video tutorials. Previous efforts relied on structured databases, taxonomies, and knowledge bases, which are expensive to create and maintain and rely on the business analyst's ability to model questions in code development.

Equation 03: LLM-Powered Data Pipeline Efficiency Objective: throughput and latency for a multi-stage (Ingest → Embed/Retrieve → Rerank → Synthesis) cloud pipeline.

1. Stage capacities μ_i (req/s). **Throughput** is bottleneck-limited:

$$R_{\max} = \min_i \mu_i \,.\; R_{\max} = \min_i \mu_i \qquad (5)$$

possibility of utilizing other enterprise repositories such as document archives and reporting systems, thereby broadening the scope of information retrieval.

B. AI-Driven Data Retrieval Techniques

The volume of data generated each day has outgrown the capacity of traditional information retrieval techniques to help business users discover and access data needed to fuelWith arrival rate λ, model each stage as M/M/1 (illustrative). Mean waiting $W_i = \frac{1}{\mu_i - \lambda}$ when $\lambda < \mu_i$.

2. **End-to-end latency** approximates

$$W(\lambda) = \sum_{i=1}^{n} \frac{1}{\mu - \lambda} \qquad (\lambda < R_{\max}) \qquad (6)$$

decision-making. AI-enhanced techniques for data retrieval enable business users to converse with LLM-augmented digital twins of their organizations and efficiently access

IV. DATA CURATION STRATEGIES

Data curation is a crucial concern in any machine learning or artificial intelligence project. Regardless of the problem domain, the quality of the curated data has a significant impact on the performance of the trained model and the insights that can be derived from the outputs. Data can quickly become inconsistent or stale, especially when updated automatically, affecting its suitability for the original intended purpose. Currently, data curation is a largely manual process, which is expensive at scale. As an alternative, it can be partly or wholly automatized through AI (especially LLM) techniques, transforming it into a real-time operation.

A. Importance of Data Quality

Phrasal concept keyword: data quality. Importance para- graph: Considering ongoing rapid growth in digital data, data quality management is inevitable and more critical than ever before. Throughout the data life cycle, data quality is a crucial aspect that should be maintained carefully and sys- tematically. Various activities for ensuring and enhancing data quality throughout the stages of data generation, collection, processing, analysis, and exploitation are referred to as data curation. The main objective of data curation is to make data fit for consumption. Highly accurate results of the analytics phase are essential for providing correct solutions, directions for decision-making, and useful information. Likewise, high- quality data are essential for achieving high-quality worldwide developments in different industries, and data curation is the key for ensuring high-quality data. List and logic sentences: To maintain high quality and guarantee consistency during these processes, systematic procedures and methodologies are key. Historical data-pipeline tools were designed to solve systematic development and operation problems for different business processes and enabled highly efficient maintenance and development for static structured data.

B. Automated Data Curation Processes

Data curation includes creating, managing, and overseeing the information and metadata that make up a dataset. The objective is to guarantee data quality and compliance. As data volumes surge, human curators encounter an overwhelming number of tasks. These encompass labeling—annotating data with metadata to identify it—and validation—confirming data accuracy, completeness, and compliance with established standards. Automated data curation emerges as a solution to handle the expanding workload.

The effectiveness of data access systems hinges on the quality of indexing. Therefore, automating the labeling and validation processes is essential. Both aspects merit attention. Labeling processes can be automated using active learning techniques. In this setup, a machine learning model identifies the least confidently labeled samples, which are then prioritized for human review. These newly labeled samples are integrated back into the training set. For validation, new

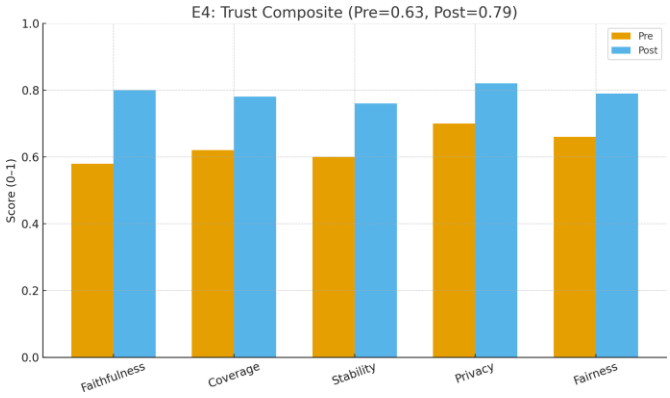

Fig. 6. Trust Composite (Pre=0.63, Post=0.79)

Component	Pre	Post
Faithfulness	0.58	0.8
Coverage	0.62	0.78
Stability	0.6	0.76
Privacy	0.7	0.82
Fairness	0.66	0.79

samples undergo scrutiny by a deep ensemble trained on the original training dataset. Detection of out-of-distribution samples triggers a human alert to address potential data inconsistencies.

Equation 04: Trust & Explainability in LLM Interpretations

Objective: a composite trust score aggregating key properties.

1. Let normalized metrics be Faithfulness F, Coverage C, Stability S, Privacy P, Fairness B

2. Use a geometric mean to penalize any weak link:

$$T = (FCSPB)^{1/5} \quad (7)$$

3. Apply "pre-toolkit" vs "post-toolkit" metrics to obtain preTpre and postTpost

V. Enterprise Data Interpretation

Data engineering, traditionally associated with supplying data, plays a vital role in the final phase by enriching data with intelligence. This intelligence enables enterprises to interpret that data in ways that support their business goals. Although the transformation of data into business insights is typically the responsibility of data scientists or business analysts, the underpinning data engineering processes are key to converting data from its raw form into a meaningful output. Before data can be used for business intelligence, reporting, or analytics, it must be made easily accessible to the business users who are paying for the data engineering activities that deliver these services.

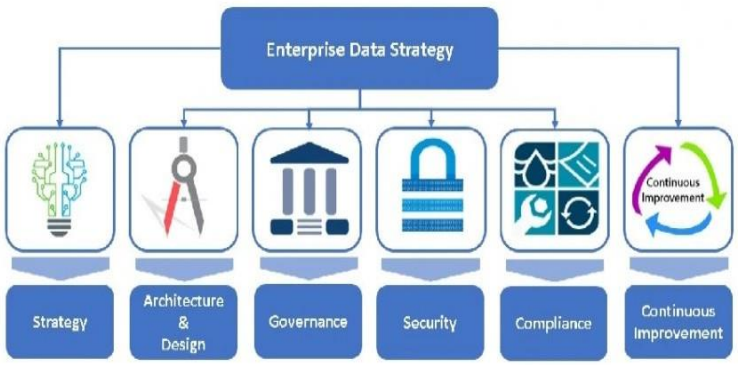

Fig. 7. Enterprise Data Interpretation

When data source content changes suddenly or anomalously, this information should be reported to stakeholders who may need to understand why, thus transforming the data into actionable information and minimizing unwanted financial impacts. Enterprise interpretation of data did not experience a step-change in 2023, as was the case for data access and data curation. However, the introduction of tools in 2024 designed for Natural Language Processing (NLP) tasks, including the broad enterprise interpretation of data, signal a significant development for data engineering in the coming years.

A. Transforming Data into Insights

In 2024, large language models (LLMs) transform enterprise data interpretation much as they have enabled AI-based search engines and chatbots to transform data access and curation. Despite the ready availability of reports and dashboards in business intelligence systems such as Microsoft Power BI, Qlik, and Tableau, most enterprises have multiple teams of analysts who investigate business questions. The pertinent facts reside in contemporary and historical, structured and unstructured data sets spread across multiple business units, business processes, and multiple systems and applications—both on premises and in the cloud. The resulting stream of answers and insights comes in a combination of reports, dashboards, presentations, documents, FAQs, emails, text messages, and chats with colleagues.

Natural language translation, provided by LLMs, reduces or virtually eliminates the time and effort required to investigate facts. By posing business questions in the language of the business user, and getting back the top relevant findings in data and text, an enterprise can flush out underlying infor- mation without needless reformatting and translation. With further refinement, the user-friendly, all-inclusive interface also supports the analyses of the underlying results, again using natural language. Ultimately, this broad-spectrum interaction in the business user's own language can span from business questions through knowledge discovery, all the way to the discovery of root cause or correlations.

B. Natural Language Processing for Business Intelligence

A recent breakthrough in data engineering plays a leading role in the business intelligence and analytics of many enterprises in 2024. Information, such as the years in which

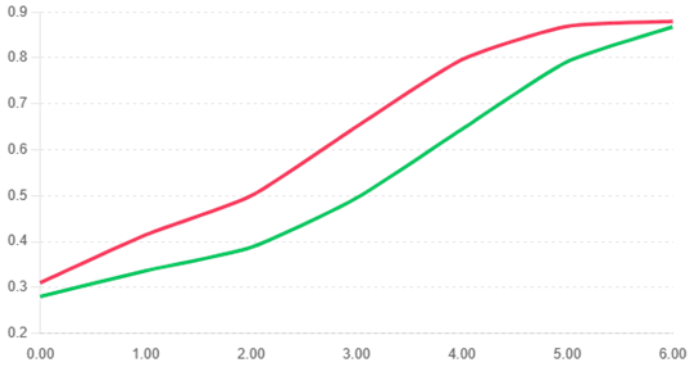

Fig. 8. Pc vs k (RAG)

the Ford Mustang occupied the segment leader position of the Sports Car Market in the US, is sourced in natural language form and not in raw data. Given the dominance of the generative capabilities of LLMs, there exists an incentive to push the needle of data access and data curation closer to its natural language beginning and end. The objective is for the data wizard to ask, what now? rather than How do I get access to these data points? The core engineering challenge is that no reliable business-oriented database enables the integration of many business domains and supplies the required answers in the natural language form required by executives and decision makers.

This challenge can be distilled into a single question: Can the multitude of business data sources be transformed into a natural language framework that fully satisfies the executives' requests? The topic of enterprise data interpretation, which examines how to extract the requested nuance or interpretation, has long been an interest of business users and therefore has always been positioned within the area of natural language processing for business intelligence. Enterprise data interpretation is indispensable for imposing order and understanding on the ever-growing volume of enterprise data, thus revealing the power of such information. Open-source models have reached state-of-the-art performance, and a new generation of modules capable of addressing the diverse requirements has emerged. The discussion now centers on methodologies and architecture.

Equation 05: LLM-Augmented Data Access

Objective: tie answer correctness to retrieval recall under RAG.

1. Let R_k be recall at k. With correctness given relevant context a_1 and "hallucination-only" correctness a_0.

$$Pc(k) = R_k a_1 + (1 - R_k) a_0. \quad (8)$$

2. Using R_k from Eq. 1 (baseline vs rerank) gives base$(k)Pc_{\text{base}}(k)$ and rerank$(k)Pc_{\text{rerank}}(k)$

k	Pc baseline	Pc rerank
1	0.28	0.31
3	0.336	0.415
5	0.387	0.5
10	0.495	0.65
20	0.644	0.796
40	0.792	0.869
80	0.868	0.88

VI. ETHICAL CONSIDERATIONS

Emerging data engineering processes empowered by AI techniques must be designed and developed with consideration for fairness, ethics, bias, and privacy. While large language models (LLMs) combined with external data provide a vast and accessible knowledge base, these applications are prone to typical LLM errors and hallucinations. In addition to hallucinations, the integrity, bias, and fairness of the underlying datasets remain critical for producing impartial, highly accurate, and trustworthy results. Furthermore, preserving the privacy of information included in the data collection remains essential. Automated data curation services must therefore include methods for fairness assessment and risk mitigation. The discussion of bias and fairness derived from LLMs encompasses the same challenges in data engineering.

Security is a fundamental concern in AI technologies, necessitating the safeguarding of data, models, and intellectual property against unauthorized access and manipulation. Given that data is the most valuable asset for any organization, adequately protecting it within storage databases and ensuring encrypted communication between the user and data systems during enterprise data interpretation are imperative.

VII. FUTURE TRENDS IN DATA ENGINEERING

Artificial Intelligence (AI) has even deeper implications for data engineering, as AI can breathe life into enterprise data through natural-language understanding. Data engineering's future addresses three fundamental problems. First: how to access enterprise data. Second: how to curate, or prepare, the data so that it is of sufficient quality, completeness, and currency. Third: how to interpret the data so that it is of high business value. Data engineering addresses each of those problems in turn.

Data engineering's future leverages Large Language Model (LLM) capabilities in every phase of the enterprise-data-lifecycle workflow. LLM-powered models offer predictive capabilities that push data engineering into a new paradigm of human-AI collaborative business-development scenario cre- ation. Data is entered just as it has historically, but these scenarios enable the creation of a data-driven forecast of the company's future—briefly estimating customers' behavior and its effects on the business, before modeling the possible future by leveraging data. The power and flexibility of LLMs also facilitate cross-slide summaries that incorporate the topic of one slide into the next. LLM-based question-answering capabilities enable users to glean answers from the scenario data, based upon a review of custom-constructed LLM prompt embeddings of the slides. Some areas of data engineering, including data access and curation, receive only a cursory treatment. On the other hand, enterprise-data interpretation finds a more extensive treatment, mirroring its recent trans- formations in 2024.

VIII. CONCLUSION

The rise of large language models (LLMs) has profoundly altered the relationship between humans and computers, for many applications in information technology. The first step towards understanding their full impact on society is to examine how data-driven enterprises now access and manipulate data in order to create business-bolstering intelligence. Data also remains critical in training these LLMs; therefore, it must be curated carefully to ensure high-quality training. Finally, the transformed data must be analysed and interpreted correctly for the benefit of the enterprises. The conclusion of this study recaps the key takeaways and discusses significant trends in how these models and their derivatives can help transform enterprise data into actionable intelligence. Currently, the focus is less on developing and training new models for enterprise use and more on enabling enterprises to sift through their vast collections of raw data, interpret it, and summarize it.

As counterintuitive as it may seem, data retrieval and interpretation bear more weight than the models themselves. A critical issue regarding these large AI models is that they may still hallucinate despite their vast training data. Mitigation can be achieved by providing raw processed enterprise data as hints to these models, thereby reducing the hallucination within organizations. Looking ahead, deeper analysis, interac- tions with enterprise systems, and suggestions on enterprise strategies will be key factors driving enterprise AI.

REFERENCES

[1] Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., Ku¨ttler, H., Lewis, M., Yih, W.-T., Rockta¨schel, T., Riedel, S., & Kiela, D. (2020). Retrieval-augmented generation for knowledge- intensive NLP. *Advances in Neural Information Processing Systems, 33*.

[2] Gadi, A. L. (2020). Evaluating Cloud Adoption Models in Automotive Manufacturing and Global Distribution Networks. Global Research Development (GRD) ISSN: 2455-5703, 5(12), 171-190.

[3] Karpukhin, V., Og˘uz, B., Min, S., Lewis, P., Wu, L., Edunov, S., Chen, D., & Yih, W.-T. (2020). Dense passage retrieval for open-domain question answering. *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

[4] AI-Powered Fraud Detection Systems in Professional and Contractors Insurance Claims. (2024). IJIREEICE, 12(12). https://doi.org/10.17148/ijireeice.2024.121206

[5] Johnson, J., Douze, M., & Je´gou, H. (2017). Billion-scale similarity search with GPUs. *arXiv preprint arXiv:1702.08734*.

[6] Mahesh Recharla, Karthik Chava, Chaitran Chakilam, & Sambasiva Rao Suura. (2024). Postpartum Depression: Molecular Insights and AI- Augmented Screening Techniques for Early Intervention. International Journal of Medical Toxicology and Legal Medicine, 27(5), 935–957. https://doi.org/10.47059/ijmtlm/V27I5/118

[7] Koppolu, H. K. R., & Sheelam, G. K. (2024). Machine Learning- Driven Optimization in 6G Telecommunications: The Role of Intelligent Wireless and Semiconductor Innovation. Global Research Development (GRD) ISSN: 2455-5703, 9(12).

[8] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez,A. N., Kaiser, Ł., & Polosukhin, I. (2017). Attention is all you need.*Advances in Neural Information Processing Systems, 30*.

[9] Botlagunta Preethish Nandan. (2024). Semiconductor Process Innovation: Leveraging Big Data for Real-Time Decision-Making. Journal of Computational Analysis and Applications (JoCAAA), 33(08), 4038–4053. Retrieved from https://eudoxuspress.com/index.php/pub/article/view/2737

[10] Armbrust, M., Ghodsi, A., Xin, R. S., Zaharia, M., Stoica, I., & Franklin, M. (2021). Lakehouse: A new generation of open platforms unifying data warehousing and advanced analytics. *Proceedings of the 11th Annual Conference on Innovative Data Systems Research (CIDR 2021)*.

[11] Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., Zhang, C., Agarwal, S., Slama, K., Ray, A., Schulman, J., Hilton, J., Kelton, F., Miller, L., Simens, M., Askell, A., Welinder, P., Christiano, P., Leike, J., & Lowe, R. (2022). Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems, 35*.

[12] Goutham Kumar Sheelam,Hara Krishna Reddy Koppolu. (2024). From Transistors to Intelligence: Semiconductor Architectures Empowering Agentic AI in 5G and Beyond. Journal of Computational Analysis and Applications (JoCAAA), 33(08), 4518–4537. Retrieved from https://www.eudoxuspress.com/index.php/pub/article/view/2861

[13] Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., Ishii, E., Bang, Y., Madotto, A., & Fung, P. (2023). Survey of hallucination in natu- ral language generation. *ACM Computing Surveys*. (Early version:*arXiv:2202.03629*).

[14] AI-Based Financial Advisory Systems: Revolutionizing Personalized Investment Strategies. (2021). International Journal of Engineering and Computer Science, 10(12). https://doi.org/10.18535/ijecs.v10i12.4655

[15] Wei, J., Wang, X., Schuurmans, D., Bosma, M., Ichter, B., Xia, F., Chi, E., Le, Q., & Zhou, D. (2022). Chain-of-thought prompting elicits reasoning in large language models. *Advances in Neural Information Processing Systems, 35*.

[16] Researcher. (2023). Decision Support Systems for Government Auditing: The Role of AI in Ensuring Transparency and Compliance. Zenodo. https://doi.org/10.5281/ZENODO.15489803

[17] Chen, D., Fisch, A., Weston, J., & Bordes, A. (2017). Reading Wikipedia to answer open-domain questions. *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (ACL)*.

[18] Raviteja Meda. (2024). Agentic AI in Multi-Tiered Paint Supply Chains: A Case Study on Efficiency and Responsiveness . Journal of Compu- tational Analysis and Applications (JoCAAA), 33(08), 3994–4015. Retrieved from https://eudoxuspress.com/index.php/pub/article/view/2734

[19] Yu, T., Zhang, R., Yasunaga, M., Tan, Y. C., Lin, X., Li, S., Er, H., Li, X., Pang, B., Chen, D., Ji, P., Tang, X., & Liu, C. (2018). Spider: A large- scale human-labeled dataset for complex and cross-domain semantic parsing and text-to-SQL task. *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

[20] Inala, R., & Somu, B. (2024). Agentic AI in Retail Banking: Redefining Customer Service and Financial Decision-Making. Journal of Artificial Intelligence and Big Data Disciplines, 1(1).

[21] Armbrust, M., Ghodsi, A., Xin, R. S., Zaharia, M., Stoica, I., & Franklin, M. (2021). Lakehouse: A new generation of open platforms unifying data warehousing and advanced analytics. *Proceedings of the 11th Annual Conference on Innovative Data Systems Research (CIDR 2021)*.