Deep Science Publishing, 2026
https://doi.org/10.70593/978-93-7185-592-1_6

Chapter 6: Generative AI in Data Workflows and Decision Systems

6.1. Introduction

While the conceptual underpinnings of generative AI have been established for more than half a century, the applications that have captured the public's imagination emerged only in the last decade. These applications include the generation of natural language, 3D realistic computer graphics, and convincingly photorealistic images of people and animals, the development of conversational agents, and text-to-image synthesis.

These generative processes create new data products using data as input; however, such data are rarely the users' ultimate goal. Instead, organisations strive to streamline and optimise the mission applications shaped by pipeline-like data workflows interlinked with decision systems that give meaning to the provided data. These data workflows and decision systems reside at the core of most data-and-analytics transformations. Their architecture incorporates, integrates, and ultimately completes all end-to-end data-and-analytics processes, including data gathering, transformation and synthesis, decision making, and insight sharing. Data workflows supply data assets for a wide range of analytical models and, increasingly, in-house decision systems.

6.1.1. Background and Significance

This section provides a brief overview of the current state of generative artificial intelligence (GAI) in relation to data work and decision systems. The broad definition, key capabilities, and core innovations central to GAI are first established. These lay the foundations for a closer examination of two main applications in the context of data flows: data workflows and decision systems—how GAI can help organize and align data to the needs, preferences, and priorities of an organization, and how GAI can assist in generating insights and supporting automation.

The rising interest in GAI is, in many respects, commensurate with a convergence of forces working to detract from the quality of the data that form the foundation of data-driven decision making. Whether the issue is a shortage of high-fidelity data for new businesses, including those in established sectors beginning to operate in the virtual world, an increase in the use of easily detectable signals such as credit card information in the credit risk space that are often poorly predictive for real-world outcomes, regulatory scrutiny that restricts access to sensitive variables, or the relevance of historically located business problems, organizations are searching for new ways for addressing data challenges. In this context, GAI-enabled data workflows and decision systems represent important opportunities for providing, securing, augmenting, and synthesizing data in order to better support data-driven decision making.

Fig 6.1: Generative AI in Data Workflows and Decision Systems

6.2. Foundations of Generative AI

Generative AI allows for the creation of original content through Machine Learning (ML) models that have been trained on vast corpuses of textual, audio, visual, and multimodal data. Natural Language Processing has particularly captured the public's attention due to advancements in large language models (LLMs) such as ChatGPT and BERT. Transformer-based architectures, which use a self-attention mechanism, have been proven to be effective in generating coherent and well-structured text sequences by ground-truth comparisons and human evaluations. GPT-3, for instance, has 175 billion parameters. The training objective, which is formulation-dependent, allows the model to generate human-like responses to queries.

Generative AI can be deployed to support a variety of tasks, including business operations or creative endeavors. Alternatively, it can enable Voice Cloning, Text-to-Image conversion, and Text-to-3D model creation. Vision-language models are particularly multifaceted because they can be used for either Visual Question Answering or Text-to-Image generation. Examples of foundational and specialized vision-language models include CLIP, DALL-E, BLIP, and Flamingo. Most of the current models are built on a pretrained, frozen vision model in combination with a language model.

6.3. Data Workflows: Architecture and Integration

Generative AI applications within data workflows are considered here. Data workflows perform specific data-related functions considered foundational for decision systems. Applied generative AI creates synthetic output-data or injects domain expertise into ingestion and transformation. These operations prepare-pipelines for pipeline and further-automation. Potential user-concerns with generated-results—accuracy, integrity, safety—are relevant for all data-work producers. An initial risk factor framework is outlined and linked to fairness, accountability, and bias mitigation in pipelines and systems.

Fig 6.2: Data Workflows Architecture and Integration

A data workflow is an aggregation of pools-swimming in the data ocean but acting to fulfill distinct purposes by performing or supporting specified data-related operations. Data ingestion and data transformation are two data functions critical to all decision systems and are listed in Figure 6.1 as foundation-layer pipelines. Data ingestion is the process of collecting and preparing data from existing sources for analysis. Data transformation is the process of changing the format, structure, or values of data to make it more appropriate for analysis. Data sources cannot be controlled or mandated. Like all aspects of data access and quality—temporal granularity, frequency, range, currency, reliability—source quality is unknown until valuable analysis is accidently produced or known value ceases. Directive transmutation of data sources by overall fairness care is unrealistic; injected expertise is needed. Generative AI is one candidate, its application opening pipelines up for further automation.

6.3.1. Data Ingestion and Transformation

The wide adoption of generative AI makes it a valuable assistant in many data workstreams. Within decision pipelines, generative AI can perform complex data-augmentation operations, expose underlying structure in data, provide explanations for data transformation processes, and offer translation services to facilitate integration of data across sources. Generative AI can also generate synthetic data for model evaluation and testing. However, deploying generative AI in these scenarios requires careful considerations. Decisions must be made regarding the operationalisation of generative AIs, their governance, and the means to ensure ethical and responsible use.

The initial phases of data workflows, namely ingestion and transformation, are prerequisites for downstream decision stages. These phases focus on data sourcing, structuring, and cleaning, and support generation of actionable information. Existing data transformation processes are often repetitive and time-consuming due to the complex nature and high dimensionality of modern data. Generative AI can assist or automate such tasks, freeing skilled data specialists to focus on more impactful work and encouraging broader community contributions via easier tooling.

6.4. Decision Systems: From Data to Insight

Generative AI offers new possibilities for decision-making systems designed to extract insights and knowledge from large volumes of data. These systems are characterized by decision models, learned from data and expressed as algorithms or rules, that directly convert the data into new information. Such data-driven approach relies primarily upon data discovery and automated learning for decision model development, removing the need for human-data and human-formulation-expert interaction. Generative AI can play an important role in the full decision process both within individual decision pipelines and across full decision workflows.

Decision systems, including pipelines, can benefit substantially from the use of generative AI techniques at every stage. In particular, generative approaches can shape workflow design and integrate capability, automate the synthesis of decision pipelines from a set of candidate models, and capture uncertainty in system inputs, decisions and underlying models. These aspects together allow for a concrete formulation of decision automation and the progression from data – often described as data-to-decision – to automated decision paths. Work in these areas is already yielding practical results in at least three applications of particular interest: data storytelling, the generation of narrative data summaries and the automated development of synthetic data.

6.4.1. Decision Pipelines and Automation

Decision systems include the decision pipelines and appropriate automation of machine learning model training and scoring. The decision pipeline organizes the models needed for a decision into a well-defined sequence. Measurement and evaluation, like data workflows, is a decision pipeline, using conventionally organized model scoring to update traditional dashboards. Powerful population-based optimization techniques like hyperparameter tuning and automated machine learning simplify model selection.

A different aspect of automation addresses the operational lifecycle of supervised machine learning. Training set preparation, feature engineering, class distribution monitoring, evaluation, and retraining can be automated by a combination of generative AI and conventionally defined business rules. A less conventional component is the training set preparation, where generative AI provides rules for populating missing features.

6.5. Applications of Generative AI in Data Workflows

Generative AI will transform data-centric workflows through applications that enable rapid data synthesis and automation of routine tasks. Within large language models, a number of software engineering tasks have already been successfully automated. For instance, GitHub Copilot leverages OpenAI Codex to assist in code completion, synthesis, documentation, and testing—among numerous other tasks. A similar capability for enhancing the development of training datasets would address a major bottleneck in data-intensive artificial intelligence deployments. The capacity to generate synthetic datasets based on specified requirements and characterizations would dramatically reduce the workload of data engineering teams and, when combined with

LLMs or diffusion models that support high-fidelity image or text generation, would enable the straightforward creation of training datasets with support for data augmentation. As such, Generative AI tools are likely to empower data engineers and largely automate the generation of routine training and evaluation datasets.

Generative data workflows support data-centric jobs through three applications: synthetic dataset generation, data augmentation to enrich an existing dataset, and generation of metadata (e.g., mappings, labels, constraints, and descriptions) or transform definitions for artisanal datasets. The demand for such services is driven by the inability to engage fully crowdsourcing to overcome the data-sparsity risks associated with supervised models. The motivation for Generative AI is to relieve data engineering by rapidly synthesizing data need or availability.

6.5.1. Synthetic Data Generation and Augmentation

Generative artificial intelligence is broadly applicable across data workflows, encompassing data ingestion, transformation, preparation, and cleansing. At the same time, it also has functions related to the generation of synthetic data for training data-hungry systems and augmentation of existing datasets to overcome limitations in sample size or diversity. For instance, the statistical characteristics of training datasets can be learned by trained neural networks, which can subsequently sample entirely new synthetic data instances with the same statistical properties. Such synthetic data can be generated in various domains, particularly image, audio, and natural language processing.

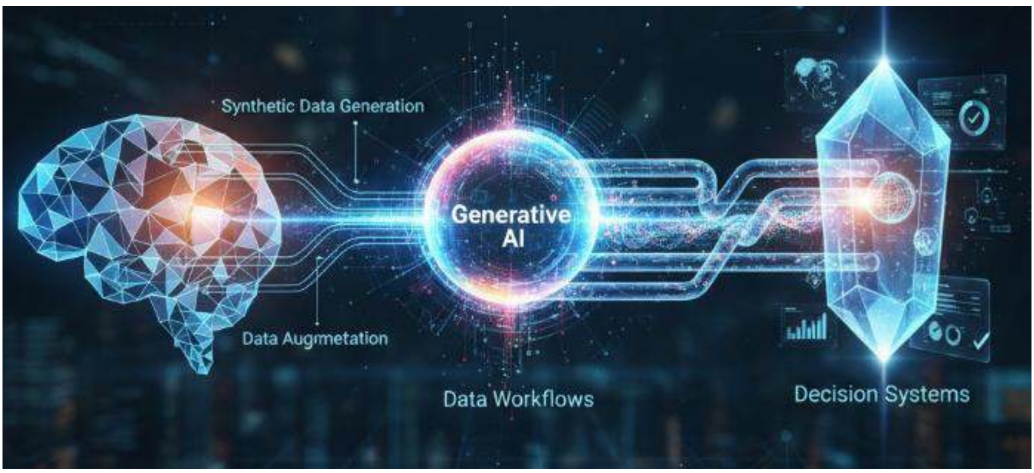

Fig 6.3: Synthetic Data Generation and Augmentation

The data distribution learned by generative models can also be exploited for augmenting other datasets. Whenever the available training dataset is small, specialized, or imbalanced (having a significant class imbalance), the classifier's generalization capability is limited. Augmenting such datasets through data transformations or by sampling from the learned conditional distribution at the decision-region margins can alleviate the low generalization problem. Generative models have been shown to successfully augment data through condition-based sampling, either by representing the entire dataset as a latent distribution with variational autoencoders or by utilizing GANs. The effect of augmenting datasets with data sampled through conditional generation is found to be better than random transformation-based augmentations.

6.6. Governance, Ethics, and Compliance

Governance of generative AI-based systems in general, and their application in data workflows particularly, is an essential priority, if not the priority during the recent years of broad and rapid adoption. Searching any term related to "Generative AI" using Twitter or Google reveals the growing chorus of observers calling for immediate and stringent implementation of responsibly governance frameworks and mechanisms focussed specifically on addressing the perceived risks. While the focus of governance conversations appears to be on declaring a pause in AI product launches for a few months while an appropriate regulatory framework is implemented, suitably-centred independence of effect and efficiency to be produced by such regulation remains contestable.

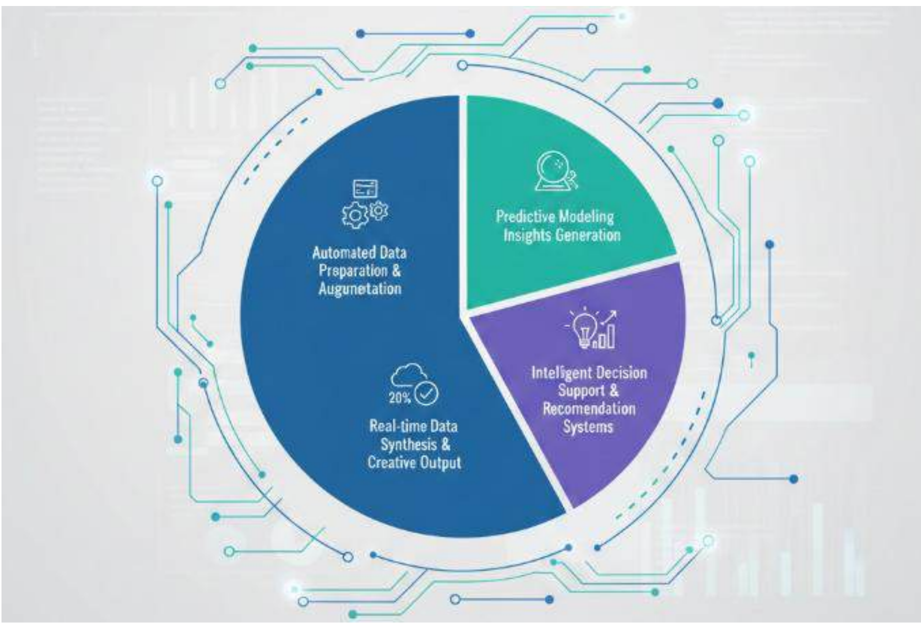

Fig 6.4: Data Workflows and Decision Systems

Many users and producers of text generation systems herald Generative AI as the democratization of producing naturally coherent text based on short prompts that effectively describes what the user wants and how the output may be structured. The users and producers see a world-changing societal shift where large corporate text websites' profits are on the verge of collapse, yet the companies and products behind the tools are monetizing something people have been doing at little or no cost for years, with little consequence. The increasing availability of cheap computation enables anyone without either cost or competency in coding to produce books, legal documents, reports, stories and more, in multiple languages or styles – even fake product reviews and PhD theses. The prevalent view is that these Generative AI tools have the same benefits and risks as any other teaching tool. They achieve work and school objectives, with increasing risk of detection, where cheating has been detected and found unacceptable and learners subsequently been penalized.

6.6.1. Fairness, Accountability, and Bias Mitigation

The decision systems outlined in this section enable automatic feedback loops, where generated data are fed back into the data pipeline to confront model bias or ensure representation for specific groups. When generative AI solutions are sufficiently automatable, automated systems evaluate candidate model versions and deploy decisions in production. Underlying these applications are governance frameworks addressing risks posed by generative AI. Ethical, political, and social movements implore companies to propose systems that bolster rather than undermine public safety and clamour for greater accountability and fairness in AI systems.

Fairness is a difficult problem, with its definition debated and often tied to specific political goals. The behaviour of a system may also become fairer after auditing and adaptation, including using modern fairness-aware classifiers. Adversarial training using important safeguards, testing using distinguishable high-fidelity data and noise robustness are appreciated for enhancing the fairness of machine-learning models. Bias may even be contained in data flows generated by natural systems, and work at the intersection of machine learning and causality has begun to quantitatively address the connection between causality and fairness. Yet despite the surge in literature on fairness and machine learning, the fairness of developers' definitions or value choices is seldom questioned, nor the implications of deploying apparently fair classifiers scrutinised. The ICML 2019 paper 'Big Data's Disparate Impact' contends that algorithmic fairness efforts should not distract attention from the emerging interplay between machine learning and the creation, maintenance, and acceleration of social inequality in the United States and elsewhere.

6.7. Conclusion

Generative AI has made a transformative impact on many areas and applications across society. Its ability to maintain and enhance the functionality of data pipelines and decision systems–-even without any training specific to those requirements–-presents compelling utility for data scientists and organizations as a whole. In data ingestion and preparation, generative AI enables automation as well as intelligent system augmentation. In decision systems, it allows decision pipelines to process information more efficiently through enabling component parallelization. Further, automatic training, tuning, and testing of decision models enhances the speed and reliability of decision pipelines. Finally, generative AI provides solutions for more advanced segments of data science including the augmentation and creation of new datasets for model training.

Generative-AI technologies are thoroughly integrated into everyday tools such as cloud-based office applications and programming environments. Though these technologies are not yet fully-adopted tools for data workflows and decision systems, their accelerating integration into cloud development platforms and new-generation data tools suggests that generative AI will become another mainstream technology in the field of data science. Nevertheless, there are broader considerations than just technology. Knowledgeable data scientists and data analysts are imperative to monitor model behaviour, assess results and reports, and comply with ethical business practices.

6.7.1. Emerging Trends

Over the past decade, the technologies and tools available to organizations for deriving business insights from data have improved rapidly, and the cost of helpful technology has been decreasing. Industries

have begun to advocate for a more advanced type of data practitioner—the data engineer and data analyst. More recently, the advent of generative AI shows promise for data workflows and decision-making systems, with the prospect of users being able to produce sophisticated results simply by describing what they want with natural language. Generative AI has the potential to play an assistance or user-empowerment role for many tasks in the data-processing workflow, involving data ingestion, data transformation, and handling of all aspects of decision-making systems. It specifically refers to those models that can generate new data points similar to their training data.

Generative models have been applied extensively for generating images, video, or text, with more recent interest in generating or supplementing data intended to train other models or using generative AI for synthetic data establishment and augmentation in data processing pipelines. Synthetic data generation and augmentation can aid data-limited use cases, improve classification performance, and reduce undesirable bias for risk-sensitive applications. Still, there are multiple important risks associated with the usage of synthetic data, and key topics on data governance must rigorously address these areas in order to ensure proper alignment of model decisions with social values relevant to fairness, accountability, transparency, security, and privacy..

References

Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). Generative AI at work. arXiv. https://doi.org/10.48550/arxiv.2304.11771

Polamarasetti, S., Kakarala, M. R. K., Goyal, M. K., Butani, J. B., Rongali, S. K., & kumar Prajapati, S. (2025, May). Designing Industry-Specific Modular Solutions Using Salesforce OmniStudio for Accelerated Digital Transformation. In 2025 International Conference on Advancements in Smart, Secure and Intelligent Computing (ASSIC) (pp. 1-13). IEEE.

Chen, Y., Yan, Z., & Zhu, Y. (2023). A unified framework for generative data augmentation: A comprehensive survey. arXiv. https://doi.org/10.48550/arXiv.2310.00277

Varri, D. B. S. (2021). Cloud-Native Security Architecture for Hybrid Healthcare Infrastructure. Available at SSRN 5785982.

Cheng, X., Li, J., & Bing, L. (2023). AI agents for complex workflows: A new paradigm in software engineering. arXiv.

Sambath Narayanan, D. B. G. (2024). Data engineering for responsible AI: Architecting ethical and transparent analytical pipelines. International Journal of Emerging Research in Engineering and Technology, 5(3), 97–105. https://doi.org/10.63282/3050-922X.IJERET-V5I3P110

Yandamuri, U. S. (2021). A Comparative Study of Traditional Reporting Systems versus Real-Time Analytics Dashboards in Enterprise Operations. Universal Journal of Business and Management.

Sambath Narayanan, D. B. G. (2025). AI-driven data engineering workflows for dynamic ETL optimization in cloud-native data analytics ecosystems. American International Journal of Computer Science and Technology, 7(3), 99–109. https://doi.org/10.63282/3117-5481/AIJCST-V7I3P108

Kolla, S. H. (2021). Rule-Based Automation for IT Service Management Workflows. Online Journal of Engineering Sciences, 1(1), 1–14. Retrieved from https://www.scipublications.com/journal/index.php/ojes/article/view/1360

Sambath Narayanan, D. B. G. (2025). Generative AI–enabled intelligent query optimization for large-scale data analytics platforms. International Journal of Artificial Intelligence, Data Science, and Machine Learning, 6(2), 153–160. https://doi.org/10.63282/3050-9262.IJAIDSML-V6I2P117

Segireddy, A. R. (2020). Cloud Migration Strategies for High-Volume Financial Messaging Systems.

Sindhu, A., & Arumugam, S. (2025). Agent-based generative AI model for cost-aware automation in machine learning pipelines. IEEE Access, 13, 192190. https://doi.org/10.1109/ACCESS.2025.3574660

Boone, T., et al. (2025). Applications of GenAI across key areas of supply chain management: Security and governance. Journal of Operations Management.
Gottimukkala, V. R. R. (2022). Licensing Innovation in the Financial Messaging Ecosystem: Business Models and Global Compliance Impact. International Journal of Scientific Research and Modern Technology, 1(12), 177-186.
Doreswamy, N., & Horstmanshof, L. (2025). Generative AI decision-making attributes in complex health services: A rapid review. Cureus, 17(1), e78257. https://doi.org/10.7759/cureus.78257
Nagubandi, A. R. (2024). Breakthrough Real-Time AI-Driven Regulatory Intelligence for Multi-Counterparty Derivatives and Collateral Platforms: Autonomous Compliance for IFRS, EMIR, NAIC, SOX & Emerging Regulations. Journal of Information Systems Engineering and Management, 9.
Dubey, R., et al. (2024). Theoretical toolbox for generative AI in supply chain management. International Journal of Production Research.
A Scalable Web Platform for AI-AugmentedSoftware Deployment in Automotive Edge Devices via Cloud Services. (2024). American Advanced Journal for Emerging Disciplinaries (AAJED) ISSN: 3067-4190, 2(1). https://aajed.com/index.php/aajed/article/view/12
Ibrahim, M., et al. (2025). Generative AI for synthetic data across multiple medical modalities: A systematic review. Computers in Biology and Medicine, 189, 109834. https://doi.org/10.1016/j.compbiomed.2025.109834
Vijay, M., Kumar, B. P., Vardhan, S. V., & Reddy, V. S. G. (2025, June). AI-Driven Multimodal Authentication: Strengthening Healthcare Biometric Security with GRNs. In 2025 International Conference on Computing Technologies (ICOCT) (pp. 1-7). IEEE.
López-Solís, O., et al. (2025). Effect of generative artificial intelligence on strategic decision-making in entrepreneurial business initiatives: A systematic literature review. Administrative Sciences, 15(2), 66. https://doi.org/10.3390/admsci15020066
Agentic AI in Data Pipelines: Self OptimizingSystems for Continuous Data Quality, Performance,and Governance. (2024). American Data Science Journal for Advanced Computations (ADSJAC) ISSN: 3067-4166, 2(1). https://adsjac.com/index.php/adsjac/article/view/23
Pahune, S., Akhtar, Z., Mandapati, V., & Siddique, K. (2025). The importance of AI data governance in large language models. Big Data and Cognitive Computing, 9(6), 147. https://doi.org/10.3390/bdcc9060147
Guntupalli, R. (2025). Intelligent cloud networking: Applying ai and reinforcement learning for dynamic traffic engineering, QoS optimization and threat detection in software-defined cloud architectures. Available at SSRN 5267809.
Sun, M., Han, R., Jiang, B., Qi, H., Sun, D., & Yuan, Y. (2025). A survey on large language model-based agents for statistics and data science. The American Statistician. https://doi.org/10.1080/00031305.2025.2561140
Aitha, A. R. (2022). Cloud Native ETL Pipelines for Real Time Claims Processing in Large Scale Insurers. Available at SSRN 5532601.
Vaddepalli, R. K. (2025). Smart governance for AI: Can metadata automation keep up with real-time ML pipelines? International Journal of Artificial Intelligence, Data Science, and Machine Learning, 6(2), 119–124.

Amistapuram, K. (2024). Generative AI in Insurance: Automating Claims Documentation and Customer Communication. Turkish Journal of Computer and Mathematics Education (TURCOMAT), 15(3), 461–475. https://doi.org/10.61841/turcomat.v15i3.15474

Wamba, S. F., et al. (2024). GenAI use across different stages of supply chain projects: From proof-of-concept to full-scale. Journal of Business Research.

Zhang, K., et al. (2025). Revolutionizing health care: The transformative impact of large language models in medicine. Journal of Medical Internet Research, 27, e59069. https://doi.org/10.2196/59069

EconStor (2025). Generative AI for decision-making: A multidisciplinary perspective. https://www.econstor.eu/bitstream/10419/327646/1/S2444569X25000964.pdf

IIT Bombay - IRCC (2026). White Paper: GenAI for Business: Insights from India. https://rnd.iitb.ac.in/sites/default/files/2026-01/Sonne%20Vyas%20Arur%20GenAI%20for%20Business%20WP%20Jan%202026.pdf

Rani, P. R. S., Kummari, D. N., Yellanki, S. K., Meda, R., Reddy Koppolu, H. K., & Inala, R. (2025). Blockchain and AI for Securing Electrical Infrastructure. In 2025 2nd International Conference on Computing and Data Science (ICCDS) (pp. 1–6). IEEE. 2025 2nd International Conference on Computing and Data Science (ICCDS). https://doi.org/10.1109/iccds64403.2025.11209487

Preprints.org (2025). A comprehensive review of generative AI adoption in hedge funds: Trends, use cases, and challenges. https://doi.org/10.20944/preprints202506.0931.v1