Chapter 2

Generative AI and the Strategic Redefinition of Enterprise Information Systems

Cevher Özden[1]

Abstract

The rapid progress of artificial intelligence has pushed the field of Management Information Systems (MIS) into a period of major transformation. Organizations must reconsider how data is structured, how daily operations are executed, and even how strategic decisions are made. For many years, AI systems in business primarily handled tasks with clear rules and predictable outcomes. With the rise of generative AI (GenAI), this narrow focus has widened dramatically, bringing creativity, synthesis, and exploratory analysis to the forefront of technological strategy. This chapter examines how GenAI is reshaping MIS and clarifies its departure from traditional discriminative AI approaches. Discriminative models learn decision boundaries to classify or predict predefined categories, excelling at tasks like fraud detection or credit scoring but incapable of producing new content. GenAI, by contrast, models the joint probability structures of data, enabling the generation of novel outputs (text, code, images, etc.) beyond the training examples. This paradigm shift compels organizations to adopt hybrid MIS architectures: traditional discriminative tools remain essential for high-speed, precision tasks, while GenAI introduces a new layer dedicated to experimentation, content creation, and innovation. Technologically, this transition is underpinned by sequence-to-sequence Transformer architectures and large language models (LLMs). To sustainably integrate GenAI into MIS, firms must actively manage new risks by establishing robust ethical and governance frameworks. Three strategic priorities emerge for MIS leaders: incorporating retrieval-augmented generation (RAG) for more reliable, fact-grounded outputs, expanding the use of low-code/no-code platforms to democratize analytics, and investing in reinforcement learning from human feedback (RLHF) to align AI behavior with human values. By balancing innovation with responsible governance, enterprises can leverage GenAI to radically enhance decision-making and operational performance without compromising security or ethics.

[1] Assist. Prof. Dr., Cukurova University, Faculty of Arts and Sciences, Department of Computer Sciences, cozden@cu.edu.tr, ORCID ID: 0000-0002-8445-4629

https://doi.org/10.58830/ozgur.pub1143.c4718

1. Defining the Paradigm Shift: Generative vs. Discriminative

Understanding how generative AI diverges from earlier discriminative AI approaches is essential for interpreting the sweeping transformation underway in enterprise IT. Although both fall under the AI umbrella, the two methodologies serve fundamentally different business purposes and rely on separate architectural logics. Discriminative models (e.g. logistic regression, support vector machines) are engineered to make decisions about predefined classes by learning the boundaries that distinguish one category from another (Cao et al., 2023). These models excel in operational tasks where accuracy and consistency are crucial such as fraud detection, credit risk scoring, spam filtering, or biometric verification. However, because discriminative models are limited to interpreting existing data and cannot generate new information, they are ill-suited for domains that demand original content or creative synthesis.

Generative AI, by contrast, learns the underlying statistical patterns of its training corpus, effectively modeling the joint probability distribution of inputs and outputs (Feuerriegel et al., 2024). This deeper understanding allows GenAI not only to assess information but also to produce artifacts that never existed in the original dataset ranging from natural-language text and working software code to synthetic images and video. In other words, where a discriminative model focuses on selecting the most likely label for a given input, a generative model's role is to generate *new* possibilities consistent with what it has learned. This capability makes GenAI a powerful tool for automating creative tasks and building personalized content in areas like marketing copy generation, customer service chats, or R&D prototyping.

The architectural gap between these two model classes carries practical implications for MIS deployment, particularly regarding computational overhead and latency. Discriminative systems are usually lightweight, train relatively quickly on labeled data, and deliver low-latency inferences, making them cost-effective for real-time or high-volume use cases on standard IT infrastructure (Gozalo-Brizuela and Garrido-Merchan, 2023). For example, a logistic regression model or small decision tree can score a credit application or detect a fraudulent transaction in milliseconds on commodity servers. Generative systems especially modern LLMs differ dramatically: they often consist of hundreds of millions to billions of parameters and require powerful GPU clusters both for training and for inference at scale. Serving an LLM-based application (e.g. an AI assistant) can incur substantial computational cost and response time compared to a simple classifier. These constraints mean organizations cannot simply replace all discriminative models with GenAI. Instead, a two-tiered strategy is needed: high-speed, routine workflows (like transaction processing or sensor analytics) continue to rely on efficient discriminative models, while generative models are deployed in environments designed to support their heavier resource needs and interactive, exploratory nature.

In essence, attempting to use a giant GenAI model for ultra-fast processes (such as high-frequency trading or on-device IoT sensor interpretation) would usually be too slow or costly to be practical. The immense computational demands of GenAI mean that, today, the vast majority of its processing (including LLM training and inference) happens in cloud data centers with scalable hardware. Edge deployments of GenAI are emerging only for specialized cases and often rely on compressed "small" models due to device limitations (Davenport and Mitta, 2023). This reality reinforces the need for a blended MIS architecture that aligns model complexity with latency and cost requirements. High-throughput, real-time tasks stay on discriminative rails, whereas generative models are invoked where their creative capacity adds unique value worth the expense (e.g. generating a custom report or simulating scenarios). At a deeper level, this technical bifurcation reflects a conceptual evolution in MIS: a shift from using AI purely to *predict* or classify known outcomes toward using AI to *explore* open-ended questions and imagine new possibilities. Traditional MIS frameworks focused on minimizing uncertainty by extrapolating from past data to future outcomes; with GenAI, the emphasis shifts to synthesizing alternatives and identifying patterns not present in historical records. This broadened analytic scope requires new governance models to manage the creative and ethical risks of AI-generated content. Table 1 summarizes the core differences between discriminative and generative AI approaches and illustrates how they reshape MIS decision-making processes.

Table 1. Generative vs. Discriminative AI – A Comparative Analysis

Dimension	Generative AI (e.g., LLMs, GANs)	Discriminative AI (e.g., SVMs, Logistic Regression)
Primary Function	Creates novel content or data (synthetic output)	Classifies or predicts outcomes (decision boundary identification)
Underlying Principle	Models the joint probability distribution of inputs and outputs	Learns the conditional decision boundary between classes
Training Data Requirement	Can leverage large volumes of unlabeled data (self-supervised learning)	Requires extensive labeled datasets (supervised learning)
Computational Footprint	Large model architectures (hundreds of millions to billions of parameters); high training and inference latency	Smaller models; faster training and low-latency inference
Typical Use Cases	Content generation, creative design, scenario simulation, synthetic data augmentation	Classification, anomaly detection, risk scoring, rule-based decisions
Example Applications	Chatbots, code generation, marketing content, product design prototypes	Fraud detection, spam filtering, credit scoring, biometric ID verification

The comparison in Table 1 can also be interpreted through the lens of classical Management Information Systems typologies. Discriminative AI aligns closely with Transaction Processing Systems (TPS) and traditional MIS, where speed, accuracy, and rule-based decision-making dominate. These systems rely on structured data and predefined logic, making discriminative models well suited for operational control and routine managerial reporting.

Generative AI, in contrast, exhibits stronger alignment with Decision Support Systems (DSS), Executive Support Systems (ESS), and Knowledge Management Systems. These systems operate under conditions of uncertainty, ambiguity, and strategic exploration, where the ability to synthesize information, simulate scenarios, and generate alternative narratives adds value. From this perspective, generative AI does not replace existing MIS layers but extends the upper tiers of the MIS hierarchy by enhancing analytical creativity and strategic sense-making.

2. The Foundational Role of GenAI in Management Information Systems (MIS)

Generative AI should be regarded not as a mere add-on to existing systems, but as a strategic force capable of reshaping the core of Management Information Systems. Its impact spans three critical domains: strengthening decision-making, improving operational performance, and redefining how knowledge is created and shared across the enterprise. Recent industry research demonstrates that the most immediate economic gains from GenAI are emerging in areas where creativity, customization, and direct customer engagement are central. For instance, a McKinsey analysis estimates that about 75% of GenAI's near-term business value will concentrate in customer operations, marketing and sales, software engineering, and R&D use cases (Chui et al., 2023). This finding indicates where organizations should prioritize early GenAI investments to capture rapid returns. If a company aims for "quick win" applications of GenAI, it would do well to start with customer support virtual agents, generative marketing content, AI-assisted programming tools, or research & development accelerators – domains where GenAI can immediately drive innovation and efficiency.

At the same time, GenAI acts as a catalyst for accelerating core management processes. By streamlining the flow of information and enabling faster production of actionable insights, it enhances an organization's capacity to adjust operations swiftly. For example, generative AI systems can rapidly summarize market trends or customer feedback, allowing management to respond in near-real-time. This heightened responsiveness improves organizational adaptability in the face of competitive pressures and volatile economic conditions. In today's unpredictable markets, the ability to pivot quickly is a critical determinant of sustained competitive advantage. McKinsey's research suggests that current generative AI technologies (along with other automation) could automate 60–70% of employees' time spent on routine activities, freeing humans to focus on high-impact strategic work (Chui et al., 2023). By assuming responsibility for repetitive, low-value tasks, GenAI allows staff to concentrate on innovation, problem-solving, and strategic thinking. The resulting boost in organizational agility and creativity can strengthen competitive positioning, as companies that learn and adapt faster are better poised to seize new opportunities or mitigate emerging threats.

However, it must be emphasized that these advantages will only be fully realized if enterprises proactively manage the ethical and operational challenges accompanying GenAI adoption. Issues such as "hallucinated" outputs (AI-generated misinformation), vulnerability to data leaks, and biases embedded in model training data pose non-trivial risks to business integrity and trust. It is clear that incorporating GenAI into MIS is not a plug-and-play endeavor – it requires strategic foresight and governance. MIS leaders should therefore treat GenAI as both a source of innovation and a driver of organizational change, warranting careful alignment with corporate objectives and risk management practices.

Generative AI stands to become a foundational element of next-generation MIS because it amplifies what organizations can do with their information: create new knowledge and content, not just process what already exists. By doing so, GenAI redefines the scope of MIS from systems of record and analysis to systems of imagination and innovation. The following sections examine the technological underpinnings of this shift – namely, the rise of large language models and advanced Transformer architectures – and how these technologies can be harnessed within robust enterprise architectures.

3. Structural Foundation: Large Language Models and Sequence-to-Sequence Architectures

The extent to which generative AI can be effectively leveraged in MIS depends largely on the technological architecture supporting it. At the center of modern GenAI implementations are foundation models – deep learning models trained on massive, broad datasets that can be adapted to a wide range of tasks (Rombach et al., 2022). The most prominent examples today are *large language models (LLMs)*, which represent the most widely deployed class of foundation models. Trained on enormous corpora of unstructured text (and sometimes code or images), LLMs exhibit a remarkable ability to both generate and interpret complex natural language. Yet the notion of a "foundation model" extends beyond language alone, encompassing multimodal systems capable of producing code, structured tables, images, audio, and other sophisticated content types. Early foundation models included text-only LLMs like OpenAI's GPT series and Google's BERT. Today, similar approaches are being applied to other data modalities: e.g., models like DALL-E and Stable Diffusion for image generation, MusicGen for music, and large code models for software development. This breadth means GenAI can touch virtually every information asset in an enterprise – documents, databases, logs, designs – blurring the lines between data creation and consumption.

Underpinning many of these generative systems is the sequence-to-sequence (Seq2Seq) architecture, often implemented via an *encoder-decoder* Transformer model. Sequence-to-sequence frameworks are specifically designed to transform one sequence into another, even when the input and output lengths differ. This flexibility makes Seq2Seq invaluable for tasks like language translation (input sentence to output sentence), abstractive summarization (long text to short summary), or conversational Q&A

(user query to answer). In a classical Seq2Seq model, an encoder network processes the input sequence and distills its information into a latent representation (often a fixed-length vector). A decoder network then generates the output sequence from this representation, step by step (Min et al., 2023). Importantly, modern Transformer-based implementations extend this design with a multi-head self-attention mechanism, allowing both the encoder and decoder to dynamically focus on the most relevant parts of the sequence (or on each other's outputs via cross-attention). This attention-driven architecture significantly improves the contextual accuracy and richness of the generated outputs, compared to older recurrent neural network approaches.

The strategic value of the encoder-decoder model reaches beyond linguistics. Many high-impact organizational functions – ranging from manufacturing and logistics to finance and IT operations – are fundamentally driven by sequential data such as event logs, time-series sensor readings, transaction sequences, or process workflows (Min et al., 2023). The Seq2Seq paradigm provides a unifying computational foundation for applying generative techniques to these non-linguistic sequences as well. For example, researchers have developed Seq2Seq-based models for detecting anomalies in industrial time-series data (treating an equipment's sensor readings as a sequence to reconstruct and flag deviations). Others have encoded numerical time-series into token sequences so they can be processed by transformer models originally built for text. By using a shared sequence framework, MIS teams can standardize tooling across text-based and numeric data workflows, leveraging similar model architectures, libraries, and talent skillsets for diverse tasks. This cross-domain alignment simplifies MLOps (machine learning operations) and improves infrastructure efficiency, since the same GenAI platform might support both an NLP application and, say, a network intrusion detection system that analyzes event sequences. In short, encoder-decoder architectures serve as a *lingua franca* of generative modeling, allowing enterprises to apply common techniques to many forms of data.

3.1. Low-Code / No-Code Platforms as Enablers of Generative MIS

Alongside advances in large language models and retrieval-based architectures, the practical integration of generative AI into Management Information Systems increasingly depends on Low-Code / No-Code (LCNC) platforms. These platforms function as organizational interfaces that lower the technical barrier for interacting with complex AI systems, enabling non-technical users—such as managers, analysts, or domain experts—to build workflows, query data, and experiment with generative applications without writing software code. In this sense, LCNC environments act as a socio-technical bridge between advanced AI capabilities and everyday managerial decision-making.

When combined with generative AI, LCNC platforms accelerate experimentation cycles within MIS. Business users can rapidly prototype AI-assisted reports, scenario simulations, or customer insights dashboards, while IT departments retain oversight over data governance and model deployment. This shift redistributes analytical agency across the organization, transforming MIS from a centralized technical function into a distributed innovation infrastructure. Rather than replacing traditional development practices, LCNC platforms complement them by enabling faster iteration, improving cross-functional collaboration, and enhancing organizational learning.

From a strategic perspective, LCNC adoption amplifies the value of generative AI by ensuring that advanced models do not remain confined to specialized data science teams. Instead, GenAI-enabled LCNC tools democratize access to enterprise knowledge, thereby reinforcing MIS's evolving role as an adaptive system that supports exploration, creativity, and informed decision-making under uncertainty.

We next delve into the evolution of these architectures – in particular, how the Transformer design supplanted earlier sequence models – and then examine the variants of Transformer models (encoder-only, decoder-only, and combined) that now drive different enterprise use cases.

4. The Evolution of LLMs: From Sequence-to-Sequence to Transformers

The Transformer architecture, now the foundation of modern LLMs, was originally developed to overcome limitations in earlier natural language processing approaches. Classical sequence models like recurrent neural networks (RNNs) and long short-term memory (LSTM) networks had difficulty capturing long-range dependencies in text and were constrained by their sequential processing nature. Because RNNs process tokens one-by-one in order, they struggled to retain context from far-back in a long sequence, and they could not be easily parallelized during training (each step depended on the previous one). These issues made training on very long sequences or very large datasets slow and less effective (Chui et al., 2023).

The landmark 2017 paper *"Attention Is All You Need"* introduced the Transformer as a solution to these problems (Vaswani et al., 2017). The Transformer architecture eliminated recurrence entirely, relying instead on self-attention mechanisms to process the entire input sequence in parallel. In self-attention, each token in a sequence can attend to (i.e., consider) every other token directly, weighted by learned attention scores, regardless of their positions. This enabled the model to capture long-range relationships more effectively (since distance in the sequence matters less when any token can directly attend to any other) and allowed for much greater parallelization during training on GPUs. The result was a dramatic leap in scalability: Transformers could be trained on orders of magnitude more data than RNNs, leading to the massive model sizes that characterize today's LLMs (with tens or hundreds of billions of parameters).

Initially, Transformers were conceived with an encoder-decoder structure (e.g., for translation tasks). As the technology matured and organizations applied Transformers to a wider range of needs, three specialized architectural formats emerged: encoder-only models, decoder-only models, and the combined encoder-decoder configuration. These correspond to different subsets of the full Transformer and are suited to different categories of tasks. Below, we unpack each of these model types and their relevance to enterprise MIS.

4.1 Architectural Classification: Encoder, Decoder, and Hybrid Designs

Encoder-Only Models: These models use only the Transformer's encoder stack. The encoder processes an input sequence (e.g., a sentence or document) by iteratively applying self-attention and feed-forward layers to build a rich, context-aware representation of the entire sequence. Because the encoder's self-attention is bi-directional (each token attends to all other tokens, left and right), encoder-only systems are highly effective for tasks that require deep understanding of input text (Devlin et al., 20119). Examples of encoder-focused models include BERT (Bidirectional Encoder Representations from Transformers) and its variants like RoBERTa. During pretraining, these models often use objectives like masked language modeling, where some words are hidden and the model must infer them from context, thereby learning nuanced language representations. Encoder-only models do *not* have a generative decoder component, so they are typically not used to produce free-form text output; instead, they shine in analytical tasks on text (Rogers et al., 2020). This includes classification (e.g., classifying a document's topic or sentiment), entity extraction (identifying names, dates, etc. in text), and similarity or semantic search tasks. Within enterprises, encoder models are valuable for workflows like document classification (categorizing emails or tickets), sentiment analysis of customer feedback, or information extraction from contracts and forms. They can be fine-tuned on domain-specific data to achieve very high accuracy, even with relatively modest model sizes (BERT-base has ~110 million parameters, BERT-large ~340 million). This balance of strong language understanding at moderate scale translates to efficient inference for real-time applications – a key reason these models are often deployed for tasks requiring quick, on-the-fly analysis of text.

Decoder-Only Models: These models use only the Transformer's decoder stack, which is designed for text *generation*. A decoder produces output sequences autoregressively, meaning it generates one token at a time and, at each step, can only attend to previously generated tokens (via masked self-attention) (Bengio et al., 2003). This setup ensures that the model cannot "see" future output tokens during generation, which forces it to construct the output in a left-to-right manner consistent with natural language production. Decoder-only models are thus optimized for fluent and coherent text generation. The most notable examples are the GPT (Generative Pre-trained Transformer) family (e.g., GPT-2, GPT-3, GPT-4) and similar large language models like Meta's LLaMA (Touvron et al., 2023). These models are typically pretrained with a simple objective: predict the next word given all prior words in the sequence (also known as causal language modeling). Despite the simplicity of this training task, when scaled up with huge datasets and parameters, decoder-only LLMs acquire an impressive range of capabilities. They can continue a prompt in a contextually relevant way, compose answers to questions, summarize texts, write code, and much more. In business workflows, decoder-only models are the engines behind generative applications: producing marketing content drafts, generating responses in chatbots, writing software boilerplate, or summarizing long reports. They have also been adapted to extract information by framing extraction tasks as "fill in the blank" or Q&A generation problems. One trade-off is that these models tend to be extremely large (often many billions of parameters) to achieve high performance, which brings higher inference costs and latency. Studies have found that for understanding-focused tasks, using a giant decoder model is often overkill: an encoder-based model can achieve equal or better accuracy with a fraction of the computational demand. Thus, many enterprises use decoder LLMs specifically when text generation is needed, and switch to smaller architectures for pure analysis tasks.

Combined Encoder-Decoder Models: These models (also called full Transformer or Seq2Seq models) integrate both an encoder and a decoder (Sutskever et al., 2014). The encoder first reads and encodes the source sequence; then the decoder generates the target sequence, with each decoder step attending *both* to earlier outputs and to the encoder's outputs via an *encoder-decoder attention* mechanism (Lewis et al., 2020) (Figure 1). This architecture is ideally suited for any task where the system must transform an input into a distinct output – classic examples being machine translation (input sentence in French, output in English) or document summarization (input a report, output a summary). Modern encoder-decoder LLMs like T5 (Text-to-Text Transfer Transformer) and BART exemplify this design. They are often pretrained on mixed objectives that blend understanding and generation – for instance, T5 treats every task (translation, Q&A, etc.) as a "text-to-text" problem where it conditions on some input text and generates output text. The presence of the encoder means these models grasp input context deeply, while the decoder allows free-form output, striking a balance between the other two model types (Ji et al., 2023).

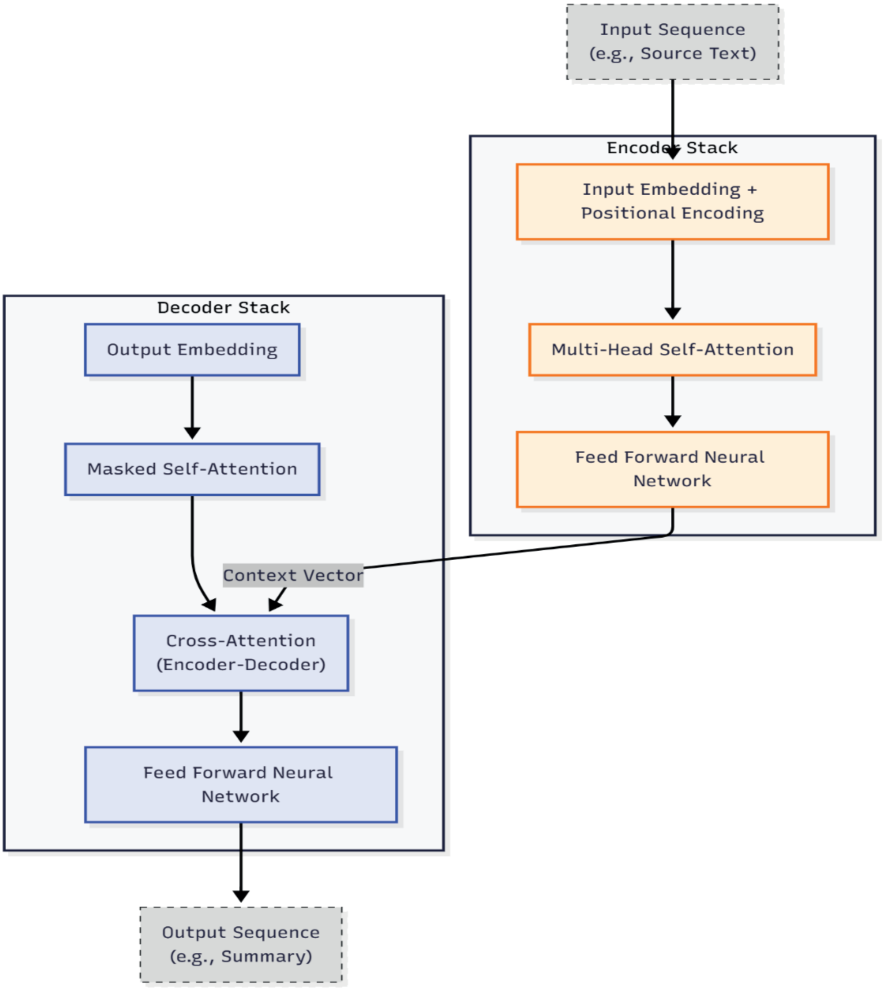

Figure 1. The Transformer Encoder-Decoder Architecture for Sequence-to-Sequence Tasks. *Note:* As depicted, the architecture consists of two main blocks: the Encoder (right), which processes the input into a context vector, and the Decoder (left), which generates the output. The critical Cross-Attention mechanism connects them, allowing the generative process to attend to specific parts of the input sequence, a fundamental design for tasks like translation and summarization.

In enterprise settings, encoder-decoder models are extremely useful for multi-step knowledge tasks. For example, they can translate documents for multilingual operations, convert unstructured text into structured outputs (parsing an invoice into a database entry), or generate executive summaries of long reports. A notable advantage of encoder-decoder systems is that the encoder's full awareness of the input can help ground the decoder's generation, reducing the chance of deviating off-topic or hallucinating irrelevant content. Because the decoder is constantly "paying attention" to the encoder's representation of the source, the output tends to stay faithful to the input's facts and context. This makes such models especially reliable for high-stakes applications like summarizing legal documents or translating compliance materials, where accuracy is paramount.

4.2 The Financial and Operational Impact of Architectural Choice

From a business perspective, choosing the right AI model architecture is not just a technical decision but also a financial and operational one. Developing an effective enterprise AI roadmap requires aligning the organization's goals with the model type best suited for the intended task – and doing so in a cost-efficient manner. In large-scale LLM deployments, inference (model runtime) often represents a significant portion of total cost of ownership. Therefore, understanding the nature of the workload (comprehension-intensive vs. generation-intensive vs. transformation) should guide the architectural decision.

For tasks that emphasize **text understanding** rather than generation, encoder-only models can offer strong results at lower cost. For example, models like BERT or RoBERTa achieve state-of-the-art accuracy on many language understanding benchmarks with on the order of only a few hundred million parameters. RoBERTa-base (~125 million parameters) or RoBERTa-large (~355 million) can be fine-tuned to perform sentiment analysis, document classification, or named entity recognition with high accuracy and relatively fast inference times (Minaee et al., 2021). Deploying such a model for an internal analytics dashboard or real-time alerting system is financially attractive because it can run on CPUs or modest GPU instances, processing many requests per second. In contrast, assigning a multi-billion-parameter decoder-only model (like GPT-3) to the same classification task would incur far greater computational overhead without improving results. In essence, if a task does not require the*generation* of new content, using a massive text generator is inefficient. Organizations have found that they are better served by using smaller, well-tuned encoder architectures for workloads that primarily involve extracting insight from text rather than producing it. This approach improves both performance and cost-efficiency, as shown in research comparing model types: for certain NLP tasks, compact encoder models not only ran faster but even outperformed much larger decoder models in accuracy.

On the other hand, if the task *does* require substantial text generation or complex interaction (such as a customer-facing chatbot that must produce answers, or an AI writing assistant), then the decoder or encoder-decoder models are justified despite their higher compute needs. Even here, a nuanced approach can optimize costs (Li et al., 2023). One strategy is to employ a two-stage pipeline: an encoder model first filters or analyzes inputs to identify when generative output is needed, then a decoder model is invoked only for those cases. Another strategy is model distillation or parameter reduction – for instance, using a smaller distilled version of a large model for most queries and falling back to the large model only for particularly difficult or high-value queries.

The key principle is *right-sizing* the model to the task. This not only ensures technical fit but also aligns with FinOps (financial operations) goals of controlling cloud expenditure and maximizing the ROI of AI initiatives. Organizations that successfully balance these factors treat model selection as a component of business architecture, not just IT architecture. They develop guidelines for when to use which type of model, considering factors like response latency requirements, privacy (smaller models can often be deployed on-premises or at the edge, avoiding cloud costs), and the criticality of accuracy vs. creativity in the task at hand.

5. Challenges, Ethics, and Future Perspectives

The widespread integration of generative AI into MIS brings a set of complex challenges related to data governance, ethical responsibility, and organizational alignment. Enterprises must tackle these concerns in parallel with technical innovation to ensure long-term, sustainable adoption of GenAI. Ignoring these factors could lead to financial, legal, or reputational damage that undermines the gains from AI.

5.1 Data Management, Privacy, and Security Challenges

Deploying GenAI in corporate environments often entails handling large volumes of sensitive information – customer records, proprietary financial data, internal strategy documents, intellectual property, and more. Feeding such material into AI models, especially those hosted on public cloud services, heightens the risk of privacy violations and security breaches. A key issue is that many advanced GenAI models (like GPT-4) run on third-party infrastructure (OpenAI, Azure, etc.), meaning any data input to them leaves the organization's direct control. This raises concerns about unauthorized access or retention of data by the service provider. Indeed, if confidential business data or personally identifiable information (PII) is inadvertently leaked through an AI service, companies could face severe consequences under privacy laws (e.g., GDPR fines) and breach of contract or secrecy claims.

To illustrate, consider an employee who uses a cloud-based GenAI tool to help draft an internal report, and in doing so, inputs excerpts of a confidential strategy memo. If the GenAI provider retains that input and it later becomes part of the model's training data or is otherwise exposed, the strategic secret could be revealed to others – a nightmare scenario for data governance. In one publicized case, it was reported that proprietary code input to a GenAI service was later found in responses given to other users, indicating a data leak through model retraining. These scenarios highlight that *data governance policies must evolve* when GenAI is in use.

To mitigate such vulnerabilities, organizations are instituting rigorous security controls and usage policies around GenAI. Strong encryption should protect data in transit to and from AI services and at rest. Data masking and anonymization techniques can be applied so that, whenever possible, sensitive identifiers (names, SSNs, etc.) are removed or tokenized before sending data to an AI model. Strict access controls and authentication ensure only authorized personnel and systems can invoke the GenAI with certain data. Moreover, comprehensive governance policies are being crafted to define what data is permissible to use with GenAI and under what conditions. Many companies now maintain an internal "allowed vs. disallowed" list for GenAI usage – for instance, public cloud GenAI might be allowed with non-sensitive data, but any customer PII or secret project info may only be used with GenAI models deployed in a private cloud or on-premises environment.

A technical approach gaining traction to enhance data protection is Retrieval-Augmented Generation (RAG). Under a RAG framework, sensitive enterprise knowledge (documents, databases) is stored in a secure vector database under the company's control (Gao et al., 2023). When the GenAI needs information, relevant chunks are retrieved and provided to the model as context, rather than the model being trained on the raw data itself. This way, the generative model's output remains grounded in enterprise data, but the data itself isn't absorbed into the model's weights where it could be regurgitated arbitrarily. The model effectively acts as a *reader* of corporate data, not a *repository* (Figure 2). By keeping proprietary data in an isolated retrieval system separate from the model, RAG significantly limits the possibility that raw sensitive data will be reproduced in outputs or learned by the model in a way that could leak it (Lewis et al., 2020). Furthermore, companies are choosing GenAI solutions that operate entirely within their private cloud or on-premises servers for highly sensitive applications. For example, some vendors offer on-prem LLM deployments or appliances so that all data processing stays behind the corporate firewall (Yao et al., 2024).

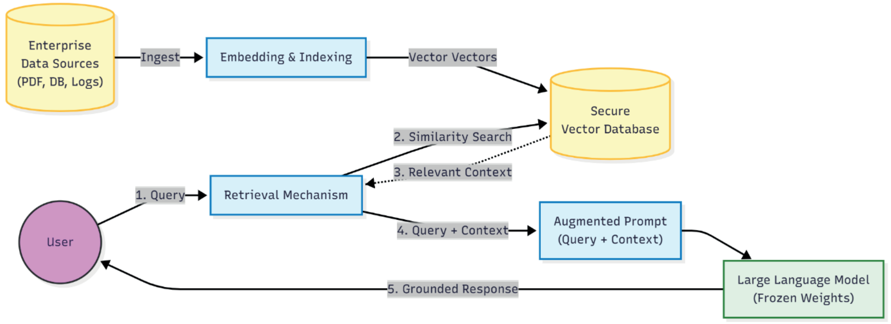

Figure 2. Architectural Data Flow of Retrieval-Augmented Generation (RAG) in Enterprise Systems. Note: This schematic illustrates the decoupling of proprietary enterprise data from the Large Language Model (LLM). By dynamically retrieving relevant context from a secure vector database only during query execution, the system ensures that responses are grounded in factual records while maintaining data privacy, as the model weights remain frozen

In addition to confidentiality, GenAI introduces new wrinkles to security policies. Traditional IT security measures (firewalls, DLP systems, etc.) must be updated to monitor GenAI usage. For instance, some organizations have implemented filters on corporate networks to detect and prevent users from pasting large dumps of data into external AI web services. Others require that any prompt to an AI that contains business data be logged (if possible) for audit purposes, or they disable GenAI access entirely on networks with classified data.

Another concern is that GenAI can unintentionally output sensitive info it *saw in training data* (a phenomenon where a model reveals parts of its training set). Thus, even if your company doesn't give an AI any internal data, if the model was trained on something like leaked passwords or personal data from the internet, it might regurgitate that. Addressing this requires thorough vetting of model providers – companies need assurances (and ideally technical proof via red-teaming or audits) that the model isn't going to spout confidential data from other sources. The legal agreements with cloud AI providers are also crucial; many have updated terms to promise they won't use customer inputs to retrain models without permission, precisely to alleviate these privacy concerns.

Introduction of GenAI into MIS necessitates a reassessment of data governance and security frameworks. Key steps include: establishing clear policies on allowed data usage for GenAI, leveraging architectural solutions like RAG that keep sensitive data segregated, insisting on privacy-protective terms from AI vendors, and layering multiple security controls (encryption, monitoring, user training) to safeguard information flows. Organizations that approach GenAI deployment with a "security by design" mindset – integrating these protections from the outset – will be far better positioned to exploit GenAI's benefits without suffering unintended data leaks or privacy infractions.

5.2 Model Bias, Accuracy (Hallucinations), and Ethical Challenges

Two major factors threaten the reliability and acceptance of GenAI outputs: bias within the models and the phenomenon of hallucination. Bias can lead AI systems to produce systematically unfair or prejudiced results, while hallucinations refer to the model generating information that is false or not grounded in reality. Both issues directly erode trust in MIS-generated insights and can carry legal or reputational repercussions.

Bias in AI models arises because the models learn from historical data that often contains societal or institutional biases. GenAI can inadvertently amplify stereotypes or discriminatory patterns present in its training set. For instance, an AI recruiting assistant might, if naively trained on past hiring data, favor male candidates for engineering roles due to historical imbalance, thereby perpetuating gender bias. Or a generative image model might produce mostly images of men when asked for "CEO" and mostly women for "assistant," reflecting biased associations (Köchling and Wehner, 2020). In a text context, biases can manifest in subtler ways in content or recommendations. These outputs can be not only embarrassing for a company but could also violate anti-discrimination laws. For example, if an AI customer service agent consistently responds less helpfully to queries in certain dialects or from certain locations due to biased training, that could be a compliance issue (fair lending laws, etc., in finance context).

Generative AI's biases have real-world impact. A 2023 study on a generative image model (Stable Diffusion) found it amplified both gender and racial stereotypes in the images it created. If an enterprise were to use such a model to, say, generate marketing visuals or profile illustrations, it might unintentionally produce content that marginalizes certain groups – e.g., depicting professionals overwhelmingly as one race/gender. The "veneer of objectivity" around AI can also make people less likely to question these outputs, which is dangerous. Thus, managing bias is not just a moral imperative but necessary for business inclusivity and compliance.

Hallucinations refer to the tendency of LLMs to sometimes fabricate facts, figures, or citations that sound plausible but are incorrect. This occurs because the model's goal is to produce fluent, contextually relevant text, not to guarantee truthfulness. It will fill gaps with its best guess, which can be entirely wrong (Luccioni et al., 2023). In enterprise settings, hallucinations are a serious issue if GenAI is used for any decision support or informational purpose. Consider an AI assistant that summarizes legal cases for lawyers: there was a notable incident (Mata v. Avianca case) where an attorney submitted a brief containing case citations that ChatGPT had *invented*, thinking they were real. The result was professional embarrassment and a stern warning from the judge. If a financial analyst used GenAI to answer "What were our Q3 profits in 2019?" and the AI confidently gives a wrong number, decisions made on that could be harmful. The risk is magnified in customer-facing scenarios. Imagine a chatbot giving a customer incorrect instructions that cause harm or a medical advisory bot hallucinating a nonexistent treatment recommendation – the liability for the enterprise could be substantial.

Tackling bias and hallucination requires a multi-pronged approach. One strategy is rigorous evaluation and testing of models before deployment. Organizations are adopting "bias bounties" and internal red-team exercises to probe their GenAI models with diverse inputs and see where problematic outputs occur. Bias testing might involve inputting prompts that describe individuals of different demographics in various roles and checking for skew in responses. For hallucinations, factual QA tests are run – e.g., ask the model a set of questions with known answers (often drawn from the company's actual data) and measure accuracy.

On the procedural side, implementing recurring ethical audit cycles is crucial. Just as financial processes are audited, AI systems should be periodically audited for ethical and performance issues. This could be done by an internal AI governance committee or external experts. They can review a random sample of GenAI outputs for appropriateness and correctness, and examine logs to catch patterns of errors or bias.

Technical mitigations are also emerging. One effective method to improve factual accuracy is integrating retrieval mechanisms (RAG) as discussed – by grounding the model in up-to-date, authoritative data sources when answering questions, the incidence of hallucination is greatly reduced. In fact, research shows that retrieval-augmented models not only are more accurate but users trust them more, especially if the sources are cited. Some GenAI systems now output references or highlight which parts of the answer come from which document, to provide transparency.

To combat bias, fine-tuning models on carefully curated datasets and applying debiasing algorithms can help adjust model weights. For instance, if an enterprise finds its model tends to produce gendered assumptions, they can fine-tune on data that counteracts this or explicitly instruct the model via prompt or system message to avoid certain stereotypes. Tools exist to post-process model outputs and scrub them of biased language (though they are not foolproof). Another angle is diversifying the human feedback in RLHF (reinforcement learning from human feedback) – ensuring that the people rating AI outputs come from diverse backgrounds so that their feedback teaches the model a more balanced perspective.

There is also a role for user education and user interface design. For example, whenever a GenAI system presents an answer, especially internally, it can come with a disclaimer like “This is AI-generated and may contain inaccuracies.” Encouraging users to verify critical information and not blindly trust AI is part of building a healthy AI-aware organizational culture. Some companies implement features where the AI will only provide answers with a certain confidence threshold or will explicitly state when it’s unsure or when multiple interpretations are possible, prompting a human to double-check.

Finally, regulation is likely on the horizon requiring companies to address AI bias and transparency. New York City already has a law mandating bias audits for AI hiring tools. The EU’s proposed AI Act might classify certain enterprise AI uses as high-risk, requiring strict oversight. Forward-thinking enterprises are preparing by documenting their AI development processes, decisions made to mitigate bias, etc., creating an audit trail that could be shown to regulators or external stakeholders.

In essence, bias and hallucinations are the Achilles’ heel of GenAI in MIS. If not addressed, they can undermine all the potential value by leading to flawed analyses, offended customers or employees, or even legal sanctions. Addressing them is an ongoing process: as AI models are updated or encounter new inputs, new forms of bias or error could emerge, so vigilance is required. By implementing rigorous evaluation, combining AI with knowledge bases, involving human feedback, and setting up strong governance, organizations can significantly safeguard against these failure modes of GenAI.

5.3 Future Focus: Reinforcement Learning and Autonomous Decision-Making

Looking ahead, a major frontier for generative AI in MIS is enabling AI systems to make *autonomous decisions* that remain aligned with human intentions, ethical norms, and organizational goals. A central methodology driving progress in this area is Reinforcement Learning from Human Feedback (RLHF). RLHF blends reinforcement learning techniques with curated human input to fine-tune AI model behavior in complex or subjective tasks (Ouyang et al., 2020).

Traditional AI optimization uses predefined reward functions – mathematical proxies for the task objective. However, for many high-level goals (like being helpful, truthful, or avoiding offense), it's extremely difficult to hand-craft an adequate reward function. RLHF tackles this by learning a reward model from human preferences: humans evaluate the AI's outputs (e.g., rank multiple responses to a prompt from best to worst), and the AI learns from these judgments to internalize what humans consider good behavior. In essence, RLHF injects a human value system into the training loop, allowing the model to optimize not just for likelihood of data, but for human approval according to specific criteria.

OpenAI's ChatGPT is a prime example of RLHF in action: after initial pre-training, the model was refined through RLHF by showing it prompt-response pairs and having human raters score which responses were more helpful or correct. The result was a model that, compared to its pre-RLHF version, is far more aligned with user expectations (e.g., it politely refuses requests for disallowed content, it follows instructions more rigorously, etc.). Many state-of-the-art LLMs from Anthropic, DeepMind, and others also leverage RLHF for alignment.

In the context of MIS, RLHF can be pivotal for ensuring AI-driven decision support systems act in accordance with company values and policies. For example, a future AI-powered decision support system (DSS) might autonomously suggest business strategy changes, adjust pricing in real-time, or negotiate with suppliers' AI agents. We would want such a system to optimize for profit and efficiency *while* respecting legal, ethical, and reputational boundaries. A purely algorithmic reward (like profit maximization) might lead to strategies that, say, exploit customers or violate regulations if unchecked. With RLHF, the model can be trained to incorporate human-defined constraints and soft goals – like fairness, transparency, or customer satisfaction – into its decision criteria. Essentially, RLHF becomes a mechanism to embed a conscience or policy adherence into AI agents.

As an example, consider an AI system in finance that manages a portfolio. Beyond just maximizing return, the firm might have ethical investment guidelines (no investing in certain industries, or considering ESG scores). Through RLHF, the system can learn a reward model that penalizes strategies conflicting with those values, because human feedback would rate such strategies poorly even if they yield profit. The AI then seeks strategies that find a balance – good return but within ethical boundaries – mirroring how a human portfolio manager would operate under guidelines.

Another emergent use of RLHF is in tuning models to local legal requirements and cultural norms. AI that interacts with the public in different regions may need to adjust its responses to align with local sensibilities or regulations (for instance, privacy laws differ by country, or what is considered offensive varies culturally). By collecting human feedback from different demographics and locales, an AI could learn to modulate its behavior appropriately depending on the user's context.

Importantly, RLHF is not just about avoiding negatives; it's also about *enhancing* positive, desired behaviors that are hard to encode otherwise. For instance, "Write code that is easy to read and well-commented" – a classical reward function can't capture code readability, but human programmers can judge it. Using their feedback, an AI code assistant can improve not just on functional correctness, but style and clarity.

In the long run, RLHF may evolve into broader reinforcement learning from human *interaction*. Rather than static feedback datasets, models could continuously learn from how humans actually use and react to them in deployment. We see glimpses of this in personalization algorithms (like recommendation systems tweaking based on user clicks), but applying it to large generative models at scale is an active area of research. Some envision AI "agents" that observe human colleagues or managers and learn from their reactions or corrections in real-time to refine their policy, analogous to a junior employee learning on the job.

There are challenges: human feedback can be inconsistent, biased, or costly to obtain. It also raises questions of whose values are being taught to the AI (hence the need for diverse feedback providers to avoid injecting bias). However, the alternative – AI learning values purely from data or static rules – seems insufficient for complex social and ethical alignment.

For enterprises, investing in RLHF infrastructure means putting in place the workflows to gather quality feedback. This could involve employing human reviewers (or leveraging crowd-sourcing platforms) to rate AI outputs in the context of the company's use cases. Some firms might create "AI ethics boards" that define the guidelines and oversee the RLHF training processes, essentially serving as the teachers of the AI's value system. Over time, as models become more self-directed (like autonomous business agents negotiating deals or adapting supply chain parameters on the fly), RLHF will be essential to trust these agents with more control. It establishes guardrails in the form of learned human-approved policies within the AI.

6. Conclusion and Strategic Recommendations

Generative AI has moved far beyond its early role as a niche automation tool within MIS. It is now becoming a central driver of economic and strategic value across industries. Recent estimates suggest that GenAI could contribute an additional $2.6–4.4 trillion annually to the global economy, spanning a vast range of use cases and sectors (Chui et al., 2023). This transformative potential is unfolding along three strategic pillars for enterprise MIS: (1) ensuring dependable information access through retrieval-augmented generation (to ground AI in facts), (2) expanding analytical capacity with low-code/no-code platforms (to democratize AI use), and (3) securing long-term ethical alignment through reinforcement learning from human feedback.

By taking over tasks that once demanded extensive human labor – from drafting routine communications to analyzing data – GenAI is freeing employees to focus on higher-value activities. Staff can spend more time on strategic thinking, creativity, and complex problem-solving, while AI handles repetitive or highly data-intensive chores. The resulting boost in organizational productivity and adaptability strengthens competitive positioning. Employees augmented with GenAI tools can iterate faster, explore more ideas, and respond to changes with greater agility. In a very real sense, GenAI, when properly deployed, acts as a force multiplier for human talent.

However, this evolution will only be sustainable if companies carefully manage the attendant ethical and operational challenges. Issues such as hallucinated misinformation, data security vulnerabilities, and embedded model biases need ongoing attention, as discussed in Section 6. For GenAI to remain an asset rather than a liability, MIS leaders must instill robust governance – from validating outputs to protecting data and aligning AI goals with human values. Technical fixes like RAG and RLHF are part of the solution, but an organizational culture that treats AI outputs with healthy scrutiny and emphasizes continuous improvement is equally important.

In this regard, we highlighted RAG as a technical necessity for mitigating immediate accuracy risks (grounding AI answers in real data reduces errors and prevents leaks), and RLHF as providing the structural basis for long-term ethical robustness (teaching AI systems the "rules of the road" for acceptable behavior). Both are complementary: RAG addresses *what* the AI knows and cites (ensuring it knows the right facts), while RLHF shapes *how* the AI uses that knowledge in alignment with human expectations.

To fully harness GenAI's strategic advantages and minimize its risks in an enterprise setting, MIS leadership should orient efforts around the following top priorities:

1. **Building a RAG-Centered Foundation for Reliability:** The reliability of GenAI-powered systems hinges on eliminating hallucinations and ensuring factual correctness. Therefore, investment in Retrieval-Augmented Generation should be a top strategic requirement. By integrating verified enterprise data sources and specialized vector databases into the AI pipeline, RAG ensures that GenAI outputs remain grounded in trusted information. Beyond improving accuracy, RAG also aids privacy: since models retrieve sensitive info as needed instead of storing it, there's less chance of unintended disclosure. In practice, MIS teams should develop a solid data indexing and retrieval layer before widescale deployment of GenAI applications. This might include curating high-quality knowledge bases (wikis, document repositories, FAQs) and using tools to embed and index this content. The GenAI can then consult this "single source of truth" for the organization when generating outputs, making it a *foundational infrastructure layer* for any durable MIS-GenAI implementation.
2. **Democratizing Analytics and Accelerating Innovation with Low-Code/No-Code:** To enhance agility and truly foster a data-driven culture, organizations should expand the use of low-code/no-code (LCNC) platforms across their MIS ecosystem. These tools empower non-technical employees to perform advanced analytics or even build AI-driven processes without deep programming skills. For instance, a business analyst could use a no-code interface to train a custom GenAI model on recent customer feedback and ask natural language questions about emerging trends, all without writing a line of code. By shifting day-to-day analytical tasks from

centralized IT or data science teams to domain experts on the front lines, LCNC dramatically speeds up internal innovation. Marketing teams can prototype personalized GenAI-driven campaigns, HR can develop AI-enabled hiring pipelines, etc., with minimal developer involvement. Meanwhile, technical teams are freed to focus on more complex, long-term projects. This broadening of analytical autonomy means every department can meaningfully participate in the digital transformation and AI innovation of the enterprise. MIS leaders should invest in training and governance for LCNC usage, to ensure quality and security as citizen development expands, but the payoff is a more nimble organization where AI and analytics are pervasive.

3. **Embedding Ethical Alignment through RLHF and Governance:** As AI systems (especially autonomous agents and DSS) become more deeply embedded in business operations, ensuring they remain aligned with human values and clear corporate ethics is *mandatory*. Thus, investment in Reinforcement Learning from Human Feedback and related governance processes is crucial to manage emerging legal, ethical, and reputational risks. Concretely, organizations should incorporate structured ethical reviews into their AI development lifecycle. Techniques such as bias audits, fairness testing, and adversarial robustness checks (somewhat analogous to security penetration tests, but for ethics) need to be standardized. For instance, an AI output audit might be conducted quarterly, similar to financial audits. Moreover, developing internal guidelines or adopting frameworks (like fairness principles or model cards documentation) will help consistently evaluate models. RLHF can be the mechanism to refine model behavior when automated metrics fall short – by explicitly training models on human preferences for correct vs. incorrect or appropriate vs. inappropriate outputs, the AI's decision-making gets aligned with complex human values that are hard to encode otherwise. MIS governance should also involve cross-functional committees (including legal, compliance, HR, and technical leaders) to oversee AI ethics. By embedding these practices, RLHF and ongoing human oversight become the foundational mechanisms for building *trustworthy, ethically consistent* autonomous decision-making into enterprise MIS platforms.

In closing, generative AI presents a once-in-a-generation opportunity to redefine how enterprises leverage information. It can elevate MIS from a primarily reactive, report-generating function to a proactive, innovation-driving one – where AI not only informs decisions but also generates creative solutions and strategies. The organizations that succeed with GenAI will be those that approach it strategically: marrying technical excellence with governance, and bold innovation with responsibility. By focusing on dependable information access (RAG), democratized innovation (low-code GenAI), and aligned, ethical AI behavior (RLHF + governance), enterprises can confidently integrate generative AI into their core and secure its benefits for the long haul. The journey involves challenges and learning, but the reward is an MIS function – and by extension, an organization – that is smarter, faster, and more creative than ever before.

References

Bengio, Y., Ducharme, R., Vincent, P., & Jauvin, C. (2003). A neural probabilistic language model. *Journal of Machine Learning Research*, *3*(Feb), 1137-1155.

Cao, Y., Li, S., Liu, Y., Yan, Z., Dai, Y., Yu, P. S., & Sun, L. (2023). A comprehensive survey of AI-generated content (AIGC): A history of generative AI from GAN to ChatGPT. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *46*(7), 1-1. doi:10.1109/TPAMI.2023.3295471

Chui, M., Hazan, E., Roberts, R., Singla, A., Smaje, K., Sukharevsky, A., and Zemmel, R. (2023). *The economic potential of generative AI: The next productivity frontier*. Retrieved from https://www.mckinsey.com/capabilities/tech-and-ai/our-insights/the-economic-potential-of-generative-ai-the-next-productivity-frontier

Davenport, T. H., & Mittal, N. (2023). How generative AI changes productivity. *Harvard Business Review*, *2023*(11), 1-10. Retrieved from https://hbr.org/podcast/2023/05/how-generative-ai-changes-productivity

Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies* (pp. 4171-4186). Minneapolis, MN: Association for Computational Linguistics.

Feuerriegel, S., Hartmann, J., Janiesch, C., & Zschech, P. (2024). Generative AI. *Business & Information Systems Engineering*, *66*(1), 111-126. doi:10.1007/s12599-023-00834-7

Gao, Y., Xiong, Y., Gao, X., Jia, K., Pan, J., Bi, Y., ... & Wang, H. (2023). Retrieval-augmented generation for large language models: A survey. *arXiv preprint arXiv:2312.10997*.

Gozalo-Brizuela, R., & Garrido-Merchan, E. C. (2023). ChatGPT is not all you need. A state of the art review of large generative AI models. *arXiv preprint arXiv:2301.04655*.

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, *55*(12), 1-38. doi:10.1145/3571730

Köchling, A., & Wehner, M. C. (2020). Discriminated by an algorithm: A systematic review of discrimination and fairness in algorithmic decision-making in the context of HR recruitment and HR development. *Business Research*, *13*(3), 795-848. doi:10.1007/s40685-020-00134-y

Lewis, M., Liu, Y., Goyal, N., Ghazvininejad, M., Mohamed, A., Levy, O., & Zettlemoyer, L. (2020). BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension.

In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics* (pp. 7871-7880). Online: Association for Computational Linguistics.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Goyal, N., ... & Kiela, D. (2020). Retrieval-augmented generation for knowledge-intensive NLP tasks. *Advances in Neural Information Processing Systems*, *33*, 9459-9474.

Li, G., Wu, L., Yan, C., Li, H., & Deng, L. (2023). Efficient large language models on edge devices: A survey. *arXiv preprint arXiv:2309.12395*.

Luccioni, A. S., Akiki, C., Cimino, A., & Mitchell, M. (2023). Stable bias: Analyzing societal representations in diffusion models. *arXiv preprint arXiv:2303.11408*.

Min, B., Ross, H., Sulem, E., Veyseh, A. P. B., Nguyen, T. H., Sainz, O., & Roth, D. (2023). Recent advances in natural language processing via large pre-trained language models: A survey. *ACM Computing Surveys*, *56*(2), 1-40. doi:10.1145/3605943

Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M., & Gao, J. (2021). Deep learning-based text classification: A comprehensive review. *ACM Computing Surveys*, *54*(3), 1-40. doi:10.1145/3439726

Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C. L., Mishkin, P., ... & Lowe, R. (2022). Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, *35*, 27730-27744.

Rogers, A., Kovaleva, O., & Rumshisky, A. (2020). A primer in BERTology: What we know about how BERT works. *Transactions of the Association for Computational Linguistics*, *8*, 842-866. doi:10.1162/tacl_a_00349

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., & Ommer, B. (2022). High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (pp. 10684-10695). New Orleans, LA: IEEE.

Sutskever, I., Vinyals, O., & Le, Q. V. (2014). Sequence to sequence learning with neural networks. In *Advances in Neural Information Processing Systems* (pp. 3104-3112). Montreal, QC.

Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M. A., Lacroix, T., & Lample, G. (2023). LLaMA: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971*.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., & Polosukhin, I. (2017). Attention is all you need. In *Advances in Neural Information Processing Systems* (pp. 5998-6008). Long Beach, CA.

Yao, Y., Duan, J., Xu, K., Cai, Y., Sun, Z., & Zhang, Y. (2024). A survey on large language model (LLM) security and privacy: The good, the bad, and the ugly. *High-Confidence Computing*, *4*(2), 100211. doi:10.1016/j.hcc.2024.100211