DIGITAL & TECHNOLOGY SERVICES

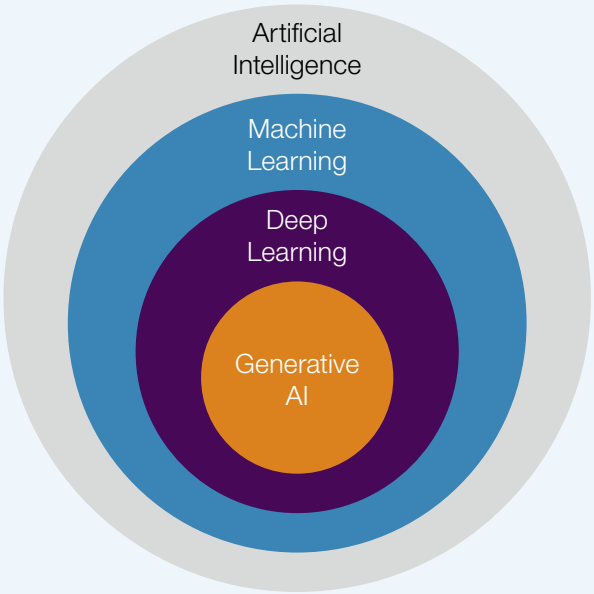
Navigating the Hype: The Strategic Use of Generative AI in Business

GenAI Is Top of Mind

There is significant buzz about Generative AI (GenAI) and how it will transform most industries. Given the tremendous potential value of GenAI and the incredible speed with which it has emerged in public use, much of this buzz is warranted. However, understanding where GenAI can be effectively deployed, as opposed to where traditional AI techniques or other technologies may be deployed, is crucial for maximizing its potential while mitigating inherent risks.

While GenAI offers compelling capabilities in content creation, interaction and exploratory analysis, its performance and cost-effectiveness diminish in certain scenarios, especially where precision, predictability and computational efficiency are paramount. Understanding the limitations of GenAI in these contexts is crucial for businesses aiming to leverage new technologies to drive performance and maximize return on critical technology investments.

GenAI's Place in the Ecosystem



Where Does GenAI Excel?

Areas of High Applicability:

1. **Content Generation:** GenAI is highly adept at producing diverse forms of content, including text, code, images and videos. This capability is particularly beneficial for marketing, creative industries and media where the need for scalable content is high. Unlike rule-based automation, GenAI can generate original content that adapts to the context or audience, often at a fraction of the time and cost required for human-only efforts.

2. **Conversational User Interfaces:** GenAI powers sophisticated conversational agents capable of understanding and generating human-like responses in both text and voice modalities. This makes it a great fit for customer service chatbots, virtual assistants and interactive applications where human engagement is crucial.

3. **Knowledge Discovery:** GenAI can sift through vast datasets to identify patterns, summarize content and extract actionable insights, which are invaluable for speeding up discovery and decision-making processes.

4. **Innovation and Brainstorming:** Similar to content generation, GenAI can produce a diverse area of ideas and simulate potential outcomes at tremendous scale and speed relative to human workers. This can facilitate ideation in creative industries but also generate novel test or configuration scenarios in diverse fields like manufacturing, engineering and life sciences.

Areas of Limited Applicability:

While GenAI can be effective in the following areas, it may not always be the best choice due to limitations in precision, decision-making capabilities or risk-related considerations:

1. **Segmentation/Classification:** For tasks that require highly accurate and explainable segmentation or classification — such as medical diagnostics — traditional machine learning models may be preferred due to their transparency and reliability.

2. **Recommendation Systems:** GenAI can personalize content and product recommendations effectively. However, systems that require deep personalization might benefit from hybrid models that combine GenAI's scalability with traditional algorithms' precision.

3. **Perception and Intelligent Automation:** In fields like robotics and surveillance, perception tasks can be enhanced by GenAI for initial processing, but the final decision often relies on more deterministic algorithms that ensure reliability and safety.

4. **Anomaly Detection/Monitoring:** GenAI can detect patterns indicating anomalies; however, critical systems such as financial monitoring or network security often require the robustness and predictability of specialized statistical and machine learning techniques.

Areas of Minimal Applicability:

In scenarios demanding high-stakes decision-making, forecasting or fully autonomous control, GenAI's current capabilities will fall short:

1. **Prediction/Forecasting:** For financial or operational forecasting, traditional statistical models or specialized predictive analytics offer greater accuracy and less speculative risk.

2. **Planning and Decision Intelligence:** Complex decision-making processes in business strategy or logistics require an intricate understanding of numerous variables where simpler, more interpretable models might outperform GenAI.

3. **Optimization:** In contexts requiring stringent accuracy, reliability and transparency, such as logistical optimization or financial portfolio management, traditional optimization algorithms are typically more effective due to their ability to consistently deliver precise, optimal solutions and their greater capacity for clear, traceable decision-making processes.

4. **Autonomous Systems:** In critical applications like autonomous vehicles or medical procedure robots, the stakes are too high to rely solely on GenAI due to concerns about reliability and ethical implications.

Note: For the above applications of more traditional machine learning techniques, one area where GenAI can help is by generating data manipulation code such as python, SQL and excel to speed using these traditional approaches. Some data scientists using GenAI to aid in coding traditional data science models report 5–20 percent increase in coding productivity.

Evaluating GenAI for Common Business Problems

While GenAI offers unparalleled capabilities in some areas, its resource intensity and operational costs can be prohibitive in contexts where simpler, more traditional approaches provide sufficient accuracy and efficiency. Whether or not any specific analytics or AI technique, including GenAI, is appropriate for a given business problem is going to depend on the individual situation and should be determined with a rigorous, collaborative process. The following examples explore considerations for why or why not to leverage GenAI, including use cases where GenAI has at least some applicability.

1. Forecasting and Prediction

Systemic forecasting of sales, revenues or other operational metrics, which requires precise and predictable outputs to be used in reporting or downstream systems, is typically incompatible with the flexible nature of GenAI outputs. In addition, data and computational requirements may render GenAI technically and financially unfeasible, even if it were accurate and precise enough.

- **Model Complexity and Resource Intensity:** GenAI models require significant computational power due to their complexity and the size of the datasets they process. This translates into higher operational costs in terms of both energy consumption and the need for advanced hardware.

- **Data Efficiency:** Traditional statistical models and some non-generative ML models typically achieve superior accuracy using much less data than GenAI models. Statistical models for time series forecasting or even simpler heuristic models demand less computational resources, which makes them more cost-effective for tightly scoped prediction tasks.

- **Explainability:** Current LLMs are entirely "black boxes" with virtually no way to reverse engineer how a model makes any particular decision. Not only is explainability useful in providing confidence about how a model reached a conclusion, in highly regulated areas such as credit scoring, insurance risk and hiring it is critical to be able to provide that a decision was not made via a biased process.

- **Speed and Scalability:** For real-time forecasting needs, the processing speed of simpler models generally exceeds that of GenAI systems. This is particularly relevant in operational contexts where decisions need to be made rapidly and at scale.

2. Recommender Systems

Recommender systems are pivotal in enhancing user experience and driving conversion in industries such as e-commerce, media and others. GenAI might be appropriate in some instances but is frequently not the most efficient choice.

- **Complexity vs. Benefit:** GenAI can generate personalized content and recommendations; however, the benefits may not justify the additional computational costs when simpler models could suffice. For example, matrix factorization techniques or even collaborative filtering processes are less resource-intensive and can be nearly as effective if the recommendation domain is not exceedingly diverse.
- **Maintenance and Update Cycles:** GenAI models require continuous updates with new data to remain effective, which can be resource-intensive. In contrast, simpler models might be updated less frequently and with fewer computational resources, reducing ongoing operational costs.
- **Precision and Predictability:** While GenAI models excel at capturing complex user preferences and behaviors, they can sometimes produce less predictable results compared to more traditional methods. This unpredictability can be a drawback in environments where consistent performance is necessary for user satisfaction and business outcomes. Additionally, current GenAI models lack sufficient version control — GenAI publishers constantly refine their models, sometimes causing systems that rely on them to behave consistently to "break" when the model behavior changes.

3. Customer-Facing Chatbots and Virtual Agents

Chatbots and virtual agents are increasingly common in customer service environments driven by the need to handle high volumes of customer interactions efficiently. However, the costs associated with GenAI-powered solutions can accumulate, particularly in high-volume environments:

- **Scalability Costs:** GenAI models, especially those sophisticated enough to handle a range of customer intents with nuance, require substantial computational resources. For enterprises handling millions of interactions, the infrastructure and operational costs can be considerable.
- **Complexity vs. Necessity:** The complexity of customer queries varies widely; not every interaction requires the advanced capabilities of GenAI. For simpler inquiries that make up a substantial portion of customer interactions, less complex and more cost-effective techniques such as rule-based automated responses or traditional ML-driven chatbots may be sufficient. These alternatives require less computational power and can be scaled more economically.
- **Predictability:** There are two sources of unpredictability. First, the models themselves can sometimes produce surprising or undesirable output. This can be somewhat mitigated by implementing multiple layers of logic; for example, having one LLM process generate content and another process check it for suitability. This, however, adds cost. The second driver of unpredictability is users. When users realize they are conversing with an LLM, they can sometime trick the LLM into answers that might be embarrassing or offensive by bypassing its internal controls.

A Summary of GenAI Risk Parameters

It is essential for organizations leveraging GenAI to develop a set of policies and best practices for secure use of GenAI. Those best practices are outside the scope of this article, but in considering where GenAI might be appropriate, leaders should be aware of several risk factors that are particularly relevant with GenAI, even compared to other analytics and AI techniques:

- **Unreliable Outputs:** The stochastic nature of GenAI might generate novel outputs, but these can also be unpredictable and occasionally nonsensical.
- **Data Privacy and Intellectual Property:** The training processes involved in GenAI may inadvertently expose sensitive data or infringe on copyrights.
- **Liability and Regulatory Compliance:** As laws struggle to keep pace with AI advancements, using GenAI could lead to liability issues, especially if outputs result in harm, financial loss or the appearance of biased decision-making.
- **Cybersecurity:** GenAI systems are susceptible to novel attacks, including data poisoning or model inversion, which can compromise user data. The "attack surface" of GenAI is currently not sufficiently understood, even by its creators.