Megan Martin
Modern Data Applications
Fall 2023

Generative AI in the Clinical Genetics Laboratory: Review and Innovation Strategy

In the past year, there has been a revolution in generative AI (GenAI) applications and while this may appear to be at a peak of the hype cycle, companies that aren't currently staying abreast of these advancements and looking for ways to harness this technology are at risk of falling behind. This document is intended to be a comprehensive review of GenAI and is organized into six sections. The first sections, Learn and Align, will provide essential background information about the state of GenAI and how this can contribute to company objectives. The following four sections (Assess, Design, and Evaluation) provide an example framework for integrating these technologies into a laboratory business. Note that citations are made with links within text and an additional resource list has been provided.

Learn

Artificial intelligence (AI) in its most basic description is any system that can simulate human intelligence or thought processes. In the last year, GenAI has received a lot of hype being touted as a rapid accelerator for changing how people and businesses operate and interact. Utilizing the "3WHERR" framework, we provide a general background on GenAI below:

- What: GenAI is a technique that learns from massive amounts of source content (text, audio, images, graphics, video, code, etc). After training on these massive datasets, GenAI models can produce various types of content given a prompt. A particular GenAI that has become highly discussed are large language models (LLMs). These are models trained on text which allows it to interpret and generate human-like text. Examples of well-known GenAI models include ChatGPT, stable diffusion, Bard, Gemini, and others.
- Why: GenAI has the potential to impact nearly every field from healthcare to manufacturing and even creative industries. It can greatly reduce the time and cost required to create new content, increasing productivity of workers. One of the powerful aspects of Generative AI and LLMs is that it interacts with humans in a manner which we are comfortable with- written or verbal requests. It doesn't require complex code or search terms in order to get a useful response, although prompt refinement may be required. The image below represents some of the quick progression of use cases in generative AI:

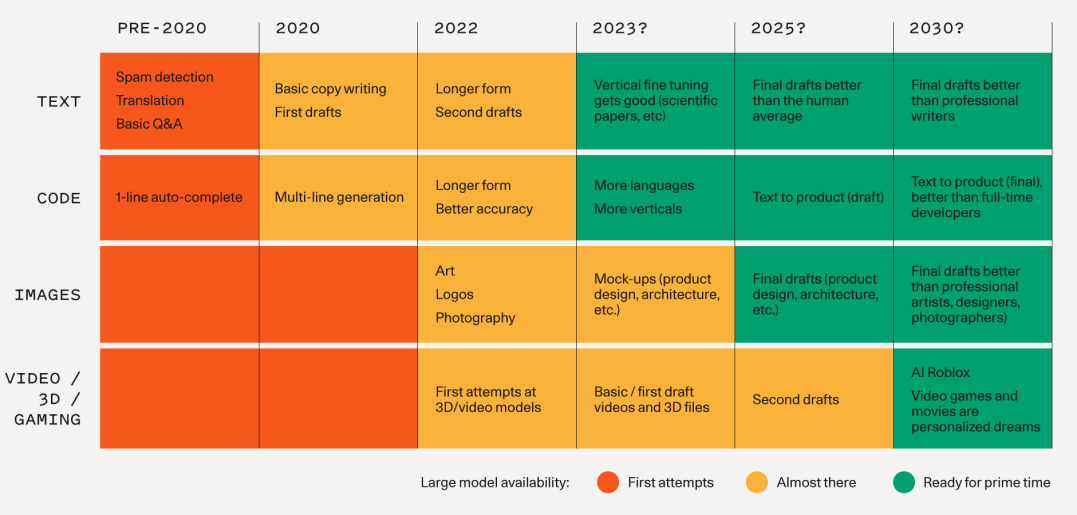

Image 1: Progression and application of Generative AI as of 2022 by Sequoia, obtained from: https://www.sequoiacap.com/article/generative-ai-a-creative-new-world/

- When: The rise of GenAI has occurred over a number of years, in fact the first GenAI were introduced in the 1960s via chatbots. In 2014, new technologies such as generative adversarial networks (GANs) and transformers enabled models to be trained without having to label all the data in advance. Without this limitation, models could be trained on ever-increasing datasets, with models including billions or even

trillions of parameters now being more common. Combined with advances in computing and storage capacity, GenAI is now available to the masses. Gartner's hype cycle for Generative AI suggests that we are still in a period of accelerated rise of expectations or at the peak across many applications.

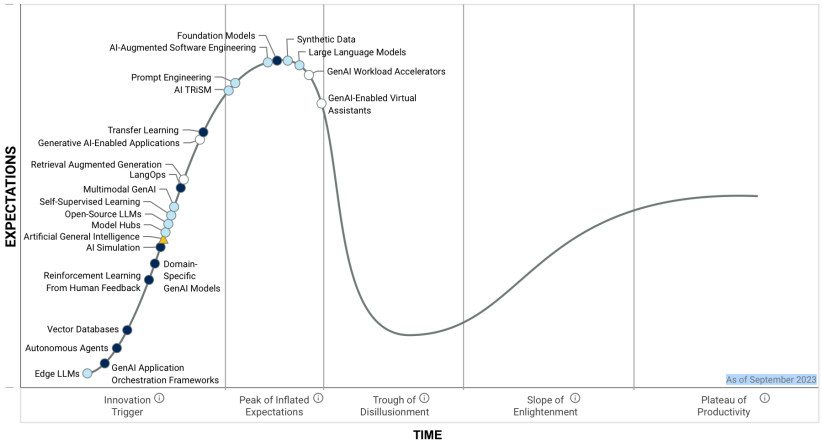

Image 2: Hype Cycle for Generative AI by Gartner, Sep 2023. Obtained from:

How Much: The costs, expertise, and time required to build and train a GenAI model are extensive. ChatGPT is estimated to cost approximately $700,000 per day for OpenAI to run. Luckily, these GenAI models are offered to customers as a monthly subscription for access and use for a certain number of prompts. These costs can start at $20 for an individual and can increase for enterprise-level access. Costs for implementation can increase, especially in the healthcare sector where hosting the model and data storage must be secure and compliant. Additionally, fine tuning a foundational GenAI model on domain-specific datasets will likely be required for custom applications, adding to additional costs. Ultimately, incorporation of GenAI into a business must have a sound financial reasoning unless brought on for a strictly competitive advantage.

- Ethical Considerations: One of the most commonly discussed ethical concerns around GenAI is the impact on the working force and resulting loss of jobs due to efficiency gains. Additionally, GenAI poses the risk of plagiarism if depending upon similarity to original content, generation of fake news and photographic evidence, and GenAI itself may provide convincing yet incorrect evidence through a phenomenon known as hallucination. Many ML models have documented bias that could result in discrimination, thus knowing the pros/cons and gaps associated with the training dataset is important. The vast amount of data that available generative models are trained on are not known to the public, making evaluation of bias extremely challenging.
- Regulatory considerations: HIPAA compliance and PHI/PII privacy and security must be ensured in any GenAI pipeline deployed in the healthcare industry. Beyond this, GDPR and CCPA/CRPA may have limitations on how companies can process and share data that may be used in GenAI training or application. Additionally, there are ongoing legal challenges related to copyright and use of material that is created from GenAI.
- Risks: One of the risks is associated with the rapid pace of improvement in available foundational models. We are in the midst of a GenAI race and there is a risk that a company invests in a model and pipeline that is quickly surpassed in capabilities. Thus, it is important to understand this evolving space and build in flexibility accordingly. As stated above in the regulatory considerations, there is currently evolving stances regarding copyrightability for GenAI content. Thus, such content generated by a company may not be protected by copyright law.

Porter's Five Forces can be applied to identify the key driving forces in this quickly evolving market:

1. Competitive Rivalry- high: There are a number of rivals in the current market. All the major companies in tech (Google, Microsoft, Meta, Amazon, OpenAi) have a GenAI model and many smaller competitors are also in this space. The competition for performance and capabilities is fierce and given the deep finances of some of these companies, it is difficult for others to compete.
2. Supplier Power- low/moderate: As this is not a manufactured product, the suppliers in this context include the AI workers themselves, the data sources, and the extensive hardware required to support the

compute and storage needs for these models. The biggest risk here is adequate access to high-demand GPU.

3. Threat of New Entrants- moderate/high: Given the extensive costs to build a GenAI model, the entry barriers are high. Established companies who have been researching and investing in AI for years have an advantage over new entrants.
4. Buyer Power- moderate: As of late 2023, there are more GenAI models and services available to customers. However, given that GenAI is for the most part still in the acceleration phase of the hype cycle, there may be more customers than services providers. To this point, GenAI services may impose limitations on use so as to manage costs and maintain performance expectations.
5. Threat of Substitutions- low: Non-GenAI substitutions are rare given the barriers of entry described above. Part of the hype for GenAI is because it provides such a different experience in natural interaction with AI than has existed before. Given the impressive capabilities, GenAI technologies are unlikely to be replaced by alternate methodologies any time soon.

Finally, a summary of the GenAI landscape diagram is provided. The landscape is presented in the following layers:

- Infrastructure: the hardware and supporting infrastructure required for GenAI applications.
- GenAI models: split into the massively trained foundation models, domain-specific models for certain use cases, and model hubs which operate as a "marketplace" for GenAI Models.
- GenAI engineering tools: describes the ecosystem of tools that are required to deploy a model into production within the business. Much of this can now be orchestrated through available MLOps tools packages.
- Applications of GenAI: Divided into horizontal apps which are business function focused, and Vertical apps which span a multitude of functions.
- Additionally two vertical layers represent the importance of legal/compliance and data privacy/security across all layers of a GenAI technology.

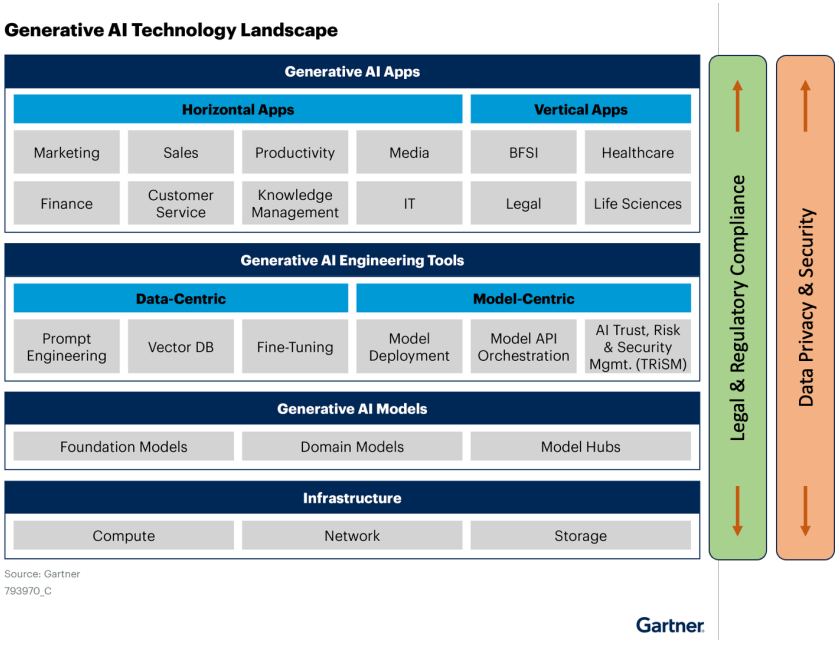

Image 3: Modified Generative AI Technology Landscape, Sep 2023. Obtained from: https://www.gartner.com/document/4750131?ref=hp-discovery&reqid=31ec12f4-82d0-4345-abe2-0d0fbdc50790

As described above, the GenAI landscape is fast evolving. While this may pose concerns if a company is not interested in being an early adopter, by not considering use of GenAI applications now, companies may be behind competitors as GenAI becomes cheaper and more broadly available. As described in the next sections, the efficiencies that GenAI can provide in a company may be critical for success.

Align

Existing and future clinical diagnostic laboratories must stand out in a competitive genomic technologies market. In order for laboratories to reach profitability, we must be able to create efficient workflows throughout the genetic testing process. The laboratory relies on highly trained and highly compensated genetics professionals which generally have a modest capacity for workload due to the hands-on nature and high quality of laboratory

tests. Workload capacity must increase the for each headcount in a non-linear manner. Gains in efficiency will also help to demonstrate a financially viable workflow in a market where genomic test reimbursement and margins are decreasing. This may also help with product differentiation if GenAI applications are incorporated as part product offerings, further contributing to the company's revenue goals.

GenAI can help improve efficiencies within a clinical genetics laboratory by providing the right information at the right time for employees. Being in a knowledge industry, much of the time spent by employees is searching and obtaining information that then determines how they respond to customers and how they complete the genetic testing workflow. One of the most impressive aspects of GenAI and LLMs is that it responds to a user in a manner that we are all comfortable with, natural language. Instead of selecting and navigating to a database, crafting a complex search term, and sifting through the results, GenAI is able to provide a generally easy to understand response based on a simple written or spoken query or an insightful analysis based on submitted data. Additionally, GenAI models like ChatGPT are able to have an ongoing back and forth conversation with the user, allowing the user to build their knowledge or tailor the response in a manner which best answers their question. GenAI can also provide suggestions for improvements or other related topics, or craft write-ups that can be used as a first draft for review by the employee. Meeting the user at a level where they are naturally comfortable, GenAI at a basic level can help as an assistant to an employee to help them obtain the information they need to act on quickly.

Assess

In order to understand how GenAI can be applied in the clinical laboratory business, we should first understand where we currently stand in relation to data management and technical capabilities. The Gartner Data Management Maturity framework is a well-regarded framework that can be used by businesses to create a master data roadmap. The current state of affairs for most laboratories is that data exists across departmental and functional silos. Data about customers, patients, and genomic data is often distributed across multiple business applications. Due to these separated datasets, managers develop inconsistent methods for insights and are often not sharing their tools or learnings with each other. Managers are also largely addressing problems as they are identified and utilizing their skills and tools to the best of their ability. A cohesive data management strategy will be essential for the embedding of any fine-tuned GenAI models.

Given the vast capabilities of GenAI, there are a multitude of applications for a clinical diagnostic laboratory. This is not an exhaustive list, but provides some high-level examples of how GenAI can be embedded to enable scaling and efficiency gains. Because we are in a regulated industry, all solutions must be evaluated to comply with HIPAA, PHI security, GDPR, and CPRA regulations. The priority and order of implementation for each idea below should be evaluated based on business value vs. costs and feasibility. Maintenance and frequency of re-training of the foundational model should also be considered.

- **LLM for an automated customer chat:** A chat functionality built on well-established LLMs can be fine-tuned to answer common questions customers (patients, families, physicians) have about testing services. Given that bottlenecks are often experienced in answering phone and emails and real-time engagement with customers is becoming increasingly challenging, this can scale communication methods. Based on a rough estimate, an automated chat functionality is expected to 2-3x a customer service team's capacity in responding to customers.
- **CAPA report creation tool:** an interactive tool utilizing GenAI to create corrective and preventive action reports for non-conforming events. The tool can walk through a series of questions with the user to collect the relevant information and generate a draft CAPA report for review. A more advanced application could reference a lab information system database to pull relevant logged notes and data to accompany the report.
- **Paper Test Requisition reader:** GenAI can be fine-tuned on an extensive database of stored paper test requisitions. Instead of requiring an employee to physically read and enter relevant information into the

LIS, scanned documents can be input into this tool and pre-fill relevant data fields. Based on missing information, flags can be visible through the user to help identify missing information or further verification required during the sample accessioning process. This functionality would likely 2-3x the capacity of the sample accessioning team.

- **Medical record summary tool:** Similar to the paper test requisition reader, a GenAI model can be fine-tuned on an extensive database of medical records. This is a more complex task as medical records are not similarly structured and may include hand-writing or print. If document reading could be successful, automated ICD-10 code assignment, medical summaries, and HPO term assignment could be pre-generated and reviewed by an employee.
- **Social media content generation:** Many marketing departments are already using ChatGPT or other similar tools to help generate pithy and effective social media content. This is an easy first application within a laboratory business.
- **GenAI for genetic test interpretation and write-up:** An interactive fine-tuned GenAI application that variant analysts can utilize to search a database of curated genetic variants and generate a summary about a genetic variant to be utilized in the final test report. This could improve efficiency of the curators and report writers up to 2x by reducing the time it takes to search for previously curated variants and crafting the interpretation for the final report.

The ability to effectively embed these GenAI applications in a laboratory business relies heavily on the right team and resources. At the center of an effective team is the team lead or architect. This person should be able to collect requirements and needs from business owners/subject matter experts and translate it to the technical team members, essentially operating as the hub in this hub and spoke team model. This lead should also be able to articulate the cost and business impact for each GenAI application. Additionally, at least one machine learning engineer who is proficient in MLOps is essential as any of the above applications will require allocation of compute and storage resources, design, deployment, and maintenance of the ML pipeline, and KPI validation of the GenAI application. Depending upon the complexity of the domain-specific data, a data scientist or data engineer may be helpful for wrangling, cleaning, prepping, and labeling data for the purposes of fine-tuning a foundational model for a targeted business application. Beyond the technical/engineering roles, a project manager role would be required to ensure timelines and milestones are met. An adjacent role to this team would be an executive/leadership sponsor who can advocate for and manage resource requirements as well as ensure the key business needs are being met by each GenAI project. For a more comprehensive review of the types of talents that would be required to form an effective team, please see "Data Science & The Art of Persuasion".

Design

It is advised that the easiest GenAI use case that does not require sharing confidential information and the output accuracy doesn't pose significant risk to the business should be built first. Instead, I would like to provide an example design for the "GenAI for genetic test interpretation and write-up" example above as it is an area I am highly familiar with. This application would embed a GenAI foundational model and fine-tune based on a robust and available genetic test result dataset for this specific use case. Building this pilot application could include the following steps:

1. **Team creation:** Assemble the team and get input from key stakeholders. Meetings to discuss the pain points, areas that could be streamlined, risks, benefits, and KPIs would help identify key areas to be aware of during the build process.
2. **Data management:** An accounting of the datasets must be established. What data is available that would be beneficial for fine tuning? How is this data stored? How messy and unstructured is the data? Here, we

will need to make an infrastructure decision. Data fabric design is a data management concept that serves an integrated layer (fabric) of data and connecting processes. Given that genetic testing data and relevant information may be found across currently disconnected applications, this metadata-driven approach can help build a connected data environment using uniform identifiers. Rather than build a complex data fabric solution, we can utilize pre-built suites of tools. In this case, Microsoft Fabric combines data engineering, Azure data storage, PowerBI integration, and data science tools in a single integrated platform. Once the strategy for data management has been established, a plan can be made on how to access and store genetic testing data from PDF reports, patient data from LIS, and genetic variant data. Cleaning, standardizing, and labeling the data is an extensive task and will require a data engineer, data scientist and a domain expert to advise.

3. **Model and platform selection:** As mentioned above in the introduction to GenAI, there is a significant barrier to entry for creating a new foundational model. This is likely not worth the company's time or finances and we should instead harness the multitude of foundational models that have been pre-trained. Here, we can utilize the Microsoft ecosystem again by utilizing their Azure OpenAI Service. This provides access to one of the most powerful and well-known foundational models: GPT from OpenAI. GPT-3.5 and GPT-4 language models have demonstrated a wide range of capabilities and impressive human interactions (see ChatGPT for examples). Here, we can harness the model used for this popular chat assistant and create an application specifically to aid in the research and writing of genetic test reports. While highly capable, it is notable that OpenAI is not open source and thus may be a "black box" if investigation into performance or inaccuracies is needed. However, given that it is a highly utilized and supported model, by going with an OpenAI model, we should be able to find talent that is familiar with it and it comes with a highly supported and documented system through Microsoft.
4. **Model customization:** Once the foundational model is selected, the team must make a decision on the level of fine-tuning required. This would involve feeding a number of labeled example datasets for the GPT model to "learn". Ultimately, the goal would be that the fine-tuned GenAI model would be able to produce a report write-up for a genetic variant that would be a well-written and well-referenced report in the similar style of the genetic test reports produced by a clinical interpretation and variant analyst team. As part of the Azure OpenAI service, a custom model wizard tool can be utilized to fine tune GPT-3.5 for a custom application. Training and validation datasets will need to be pre-identified. Training data will be utilized to fine-tune the model and validation data will be used during an iterative model evaluation process. In order to understand how the model is performing for iterative tuning, a set of established evaluation criteria will need to be identified by the project team, including review by people with domain expertise in genetics. It is imperative to have all stakeholders involved in the evaluation of the model performance prior to deployment. Of note, a fine-tuned Azure OpenAI model can access selected datasets for data retrieval. This ability to actively ingest data from an extensive genetic test data set or even

external datasets like ClinGen can be utilized to improve the usefulness of the GenAI output and could decrease the amount of data required for fine tuning (which could result in cost savings).

5. **Building and deploying the model pipeline:** After testing and validation, the model is ready to be deployed. In order to create a maintainable and sustainable GenAI application, it must be operationalized through a well-designed pipeline. Here, the frameworks within MLOps can be utilized to support the full model lifecycle from model governance, model release, monitoring, logging, and maintenance. Luckily, we can utilize yet another platform offered by Microsoft: Azure Machine Learning. As demonstrated by the Gartner MLOPs framework shown, this is a complex cycle in order to ensure that all the effort put into developing a model continues to provide business value throughout the deployment lifecycle.

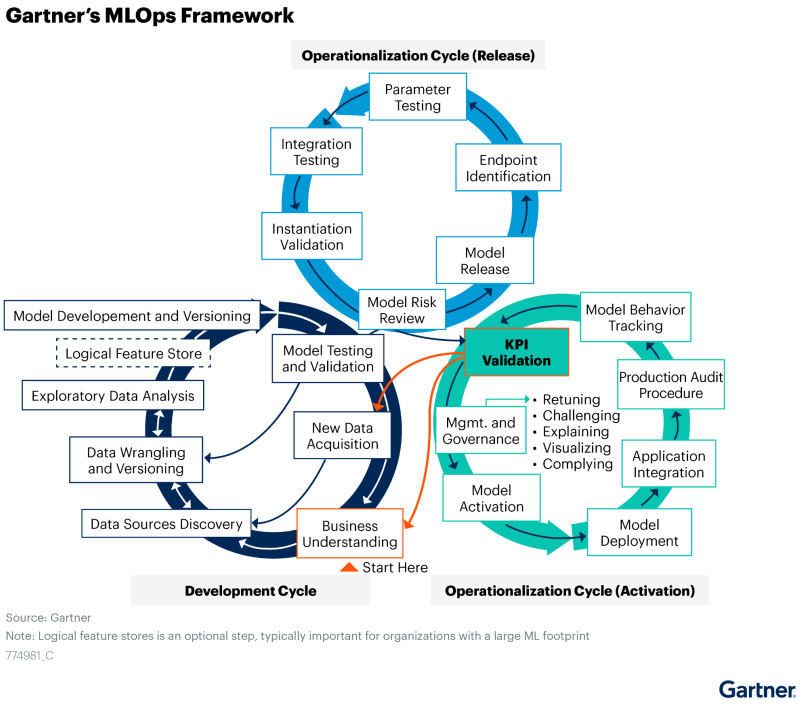

Image 4: USe Gartner's MLOps Framework to Operationalize Machine Learning Projects, August 2022. Obtained from: https://www.gartner.com/document/4018020?ref=solrAll&refval=350235091

Once the model is selected, fine-tuned, and evaluated, it can be deployed to a sandbox environment for end-user testing. Given the regulatory environment a CLIA/CAP laboratory operates under, user acceptability testing and documentation should be completed by a core team of end-users. A rough estimate for a timeline would consist of 6-8 weeks for hiring the project team, 4 weeks of input and ideation from the team and domain experts, 6 weeks for data wrangling and labeling, 4 weeks for model fine-tuning and iterations, and 2 weeks for deployment into the MLOPs pipeline. A working prototype could be in the hands of end-users for testing by 5-6 months from the approval of this initiative.

Evaluate

As mentioned, the goal of all of these proposed GenAI applications is to increase employee productivity and decrease headcount costs while a laboratory scales. As part of the evaluation of the success of the program, it would be ideal if the headcount savings from employee efficiency could more than offset the cost of the newly assembled GenAI project team. It is important to note however, beyond standard maintenance, once an application is built and deployed, the GenAI project team could move onto the next project, thus deploying an increasing number of efficiency tools over time. In order to evaluate the cost/benefit of assembling and deploying a GenAI applications strategy in a business, cost-accounting baselines would need to be established each role and task. After deploying a tool, subsequent cost-accounting can be done to determine time savings. It is important to note that should any of these tools be beneficial outside of the laboratory business unit, management could explore offering such tools as either a competitive component to the products that could be sold to other laboratories, or as a paid service, potentially creating new revenue streams for the company.

In summary, the field of GenAI is fast moving and presents many opportunities to transform business operations. While there are up-front costs to assembling a team that could design and deploy such tools, it will create a competitive advantage and a compelling story to investors on how a company is using modern technologies to tackle operational challenges. Ultimately, laboratories will have to utilize some form of GenAI if they are to remain competitive and stave off profitability challenges.

Additional Resources:

- "What is generative AI? Everything you need to know"; Oct 2023: https://www.techtarget.com/searchenterpriseai/definition/generative-AI
- "Generative AI: A Creative New World"; Sep 2022: https://www.sequoiacap.com/article/generative-ai-a-creative-new-world/
- "What CEOs Need to Know about the Costs of Adopting GenAI"; Nov 2023: https://hbr.org/2023/11/what-ceos-need-to-know-about-the-costs-of-adopting-genai
- "The Five Competitive Forces that Shape Strategy". Michael Porter. Harvard Business Review.
- "The Inference Cost Of Search Disruption – Large Language Model Cost Analysis; Feb 2023 https://www.semianalysis.com/p/the-inference-cost-of-search-disruption
- Hype Cycle for Generative AI by Gartner; Sep 2023. https://www.gartner.com/interactive/hc/4726631?ref=hp-discovery&reqid=31ec12f4-82d0-4345-abe2-0d0fbdc50790
- Generative AI Technology Landscape; Sep 2023. https://www.gartner.com/document/4750131?ref=hp-discovery&reqid=31ec12f4-82d0-4345-abe2-0d0fbdc50790
- "Data Science and the Art of Persuasion"; Feb 2019 https://hbr.org/2019/01/data-science-and-the-art-of-persuasion
- Gartner's MLOps framework; Aug 2022 https://www.gartner.com/document/4018020?ref=solrAll&refval=350235091
- "What is Data Fabric Design?"; April 2021 https://www.gartner.com/document/4000561?ref=lib
- Microsoft Fabric https://learn.microsoft.com/en-us/fabric/
- Azure Machine Learning: https://azure.microsoft.com/en-us/products/machine-learning/mlops/#features
- Azure OpenAI service: https://azure.microsoft.com/en-us/products/ai-services/openai-service-b#Features
- ChatGPT: https://chat.openai.com/