Examining Engineers' Lived Experiences Deploying Machine Learning Production Models: A Phenomenological Study

Durga Devi Papineni
University of the Cumberlands

Mary L. Lind
Louisiana State University Shreveport

This qualitative phenomenological study investigated machine learning (ML) model deployment challenges during the ML lifecycle using the theoretical framework of the technology acceptance model (TAM). Researchers have designed several frameworks for understanding the ML lifecycle, but those frameworks remain untested, and many ML model deployments still fail. The study's central research question asked, what challenges do organizations face when deploying ML models in production environments? The phenomenological research design identified users' perceptions and lived experiences deploying ML models in production environments. Data were collected via semi structured interviews with 15 ML experts. The phenomenon from the interviews was described in textural, structural, and textural-structural descriptions of participants' lived experiences.

Keywords: machine learning, deployment, phenomenological, technology acceptance

INTRODUCTION

This qualitative phenomenological study investigated the challenges of machine learning (ML) model deployment in production environments by investigating the perceived usefulness and ease of use of five stages of the ML lifecycle. ML models play a significant role in wide-ranging business fields, including technology, health, security, and manufacturing (Chen et al., 2020). Scholars have noted that many companies rely on ML models to increase process efficiency, develop innovative products, and improve service provision (Canhoto & Clear, 2020; Liu, 2020; Vincent-Lancrin & van der Vlies, 2020). Unfortunately, organizations can face many challenges when deploying ML models in production environments (Baier et al., 2019; Cai et al., 2019; Garcia et al., 2020).

Background and Problem Statement

As technology has become increasingly important today, ML has arisen as a fundamental method of extracting meaning from enormous quantities of data (Olowononi et al., 2021). ML offers organizations many benefits, health outcome predictability (Engelhard et al., 2021), and organizational efficiency (Dankwa-Mullan et al., 2019; Wuest et al., 2016). ML models also allow organizations and researchers to use novel analysis techniques that address poor data quality or complex datasets (Cai et al., 2019; Garcia et al., 2020). Unfortunately, the many benefits of ML models cannot be obtained if the models are ineffectively deployed (Mueller & Massaron, 2021; Oakden-Rayner et al., 2020).

Because of the challenges associated with ML, most development projects underperform (Mueller & Massaron, 2021; Oakden-Rayner et al., 2020; Pohle, 2018). Approximately 80% of ML projects never reach the deployment stage, and of those that do, only 60% are productive (Pohle, 2018). Based on these numbers, most ML projects never produce effective results, and the high failure rates suggest that investment in ML solutions can present risks for organizations (Mueller & Massaron, 2021). ML deployment success and adoption rates are predicted to increase because of increased data availability and exceptionally adaptable models capable of adjusting to explicit business needs, but more research is needed to ensure ML models operate efficiently (Agrawal et al., 2020; Correia et al., 2021).

This study used a qualitative research design and a phenomenological approach to investigate ML model deployment challenges in production environments examining the perceived usefulness and ease of use of five stages of the ML lifecycle. The selected stages of the ML lifecycle were grounded in scholarly literature. These five stages included (a) requirements analysis, (b) data management, (c) benchmarking metrics, (d) user acceptance testing, and (e) privacy policy.

Purpose of the Study

The purpose of this qualitative phenomenological study was to identify the challenges organizations face when deploying ML models in production environments. This purpose was achieved by exploring the lived experiences of ML engineers and evaluating their perceptions of the five stages of the ML lifecycle. Exploring these perceptions and attitudes generates a better understanding of the challenges associated with ML model deployments. The focus on the usefulness and ease of use of the ML lifecycle can improve scholarly and practitioner understanding of why deployments result in project success or failure (Singh, 2021). The study's findings can help ML teams maximize success during the model deployment process. The target population consists of 15 ML engineers recruited from all over the United States. By highlighting the challenges associated with ML model deployment and linking those challenges to specific stages within the ML lifecycle, solutions to those challenges can also be identified (Baier et al., 2019). Thus, the specific purpose of this study was to investigate the challenges associated with ML model deployment in production environments to fill a gap in the body of knowledge and provide scholars and practitioners with a better understanding of the most critical elements of the ML lifecycle.

Significance of the Study

This qualitative phenomenological study was significant because it investigated the challenges associated with ML model deployments by examining the lived experiences of professionals responsible for deployments. Studying ML deployments is critical because increasing the use of these methods can increase efficiency and improve decision-making processes in various settings (Char et al., 2020; Gu et al., 2020; Guezzaz et al., 2021). This goal was achieved by identifying relevant model deployment challenges through the lens of the TAM. The goal was to identify what elements of the ML lifecycle might present challenges based on ML engineers' perceptions of their usefulness and ease of use (Dankwa-Mullan et al., 2019; Tripathi et al., 2020). The use of the TAM to address ML deployment challenges was supported by previous research by scholars, including Alshurideh et al. (2020), Arpaci et al. (2020), Kamble et al. (2021), and Tripathi et al. (2020). Concentrating on ML professionals in different industries made it possible to provide a substantive analysis of common deployment difficulties (John et al., 2021).

Research Question

The central research question for this qualitative phenomenological study asked, how do Perceived Ease of Use (PEOU) and Perceived Usefulness (PU) impact engineers deployment of ML models in production environments?

Theoretical Framework

The literature indicates that scholarly disagreement exists regarding how to assess ML deployments and what steps precisely constitute the ML lifecycle (A. Chen et al., 2020; John et al., 2021; Paleyes et al., 2021; Xie et al., 2021). In addition to conflicting opinions on ML deployments and the ML lifecycle, the extant literature is also somewhat limited and narrowly focused (Paleyes et al., 2021). The TAM was especially well-suited for the present study because of the scholarly disagreement over the ML lifecycle. Using the TAM to evaluate participants' perceptions and attitudes toward different stages in the ML lifecycle helped identify specific challenges associated with ML deployments.

LITERATURE REVIEW

This literature review is structured to support a phenomenological research design. The review contains sections identifying structural themes and sections supporting the TAM as a theoretical framework. Phenomenological research relies on identifying universal structures associated with a phenomenon (Neubauer et al., 2019). Osborne (1990) noted that these structures provide researchers "with descriptions of experience which are then interpreted by the researcher from a particular theoretical perspective" (p. 81). These universal elements of the phenomenon constitute structural themes. Researchers use structural themes to explain *how* a phenomenon is experienced (Moustakas, 1994). In the present study, the literature review identified five universal structures (i.e., elements of the ML lifecycle), described in greater detail later in this section.

The Role of ML in AI

The first step in understanding the role of ML is to define learning in the context of probabilistic frameworks (Ghahramani, 2015). Ghahramani (2015) characterized ML as "inferring plausible models to explain observed data" (p. 452), and Attaran and Deb (2018) asserted that the basis of ML is the belief that machines can learn and adapt based on their independent analysis of data. The computerized analysis of data and the ability to learn from it or use it to predict events using models to account for periods of uncertainty is the essence of ML (Ghahramani, 2015). ML is frequently used as an element of AI, which is explained concisely as the use of computers to simulate human cognition with the goal of problem-solving and improving human behavior. In modern society, ML has become ubiquitous, and there are many ways in which ML improves daily life (Lo Piano, 2020).

Attaran and Deb (2018) characterized ML as a subset of AI and suggested that the purpose of ML is to "turn data into useful information" (p. 285). ML uses sophisticated algorithms to analyze data that inform organizational decision-making (Attaran & Deb, 2018). There are many different ML algorithms, and examples include linear and logistic regression, decision trees, k-means, random forest, and dimensionality reduction algorithms (Attaran & Deb, 2018). Algorithms are designed to solve different learning problems, and it is critical to use the right algorithm for a given process (Mahesh, 2018).

The diverse ways ML can be used to improve aspects of life and business illustrate the power of ML algorithms (Lo Piano, 2020). However, many concerns are associated with the growing use of ML, including increasing dependence on ML in decision-making, loss of human autonomy, and inadvertent or systemic bias in ML algorithms (Lo Piano, 2020). Additionally, there are concerns about the success rates of ML deployments. The following subsections address the benefits of ML and the challenges associated with ML deployments in production environments.

Benefits of ML Deployments

Researchers have studied the benefits of ML systems and techniques in different settings, including finance, research, and healthcare (Engelhard et al., 2021; Luo et al., 2019; Medeiros et al., 2019; Wuest et al., 2016). Medeiros et al. (2019) studied the ability of ML methods to improve inflation forecasting. Medeiros et al. began their study by stating the importance of forecasting inflation when making economic decisions. While banks typically release inflation forecasts to aid in financial decision-making, these forecasts can be inaccurate, improving forecasting quality has proven difficult. The researchers noted that previous inflation forecasting had been hampered by methodological approaches that limited the predictability of existing models and suggested the use of ML methods to improve forecasting. Medeiros et al. compared traditional forecasting models to ML methods and found that ML could improve inflation forecasting by 30%.

Engelhard et al. (2021) identified several benefits of ML in healthcare settings, including the ability to analyze large-scale clinical data to identify predictive factors in health outcomes. However, Engelhard et al. noted that ML had not been a universal success in healthcare. Engelhard et al. cited conflicting results from various studies comparing the results of data analysis using regression models and ML algorithms where there was no discernable improvement when using ML methods. Despite the lack of superiority in some healthcare contexts, Engelhard et al. noted that image classification is one area where ML vastly improves performance over human analysis. The use of ML to analyze images has critical implications for healthcare specialties like dermatology, ophthalmology, and radiology.

Research in the manufacturing sector identified similar benefits to ML as the previously cited works in the finance and healthcare sectors. Like (Medeiros et al., 2019) and Engelhard et al. (2021), Wuest et al. (2016) noted that ML techniques are highly effective when used to analyze “high-dimensional, multi-variant data” when analyzing chaotic, dynamic, or complex environments (p. 28). However, Wuest et al. suggested that different ML models could produce different advantages because of their suitability to specific settings and uses. Wuest et al. explained that distributed hierarchical decision tree methods handle high-dimensional data better than other algorithms.

Wuest et al. (2016) suggested that a significant advantage of ML techniques is that they are effective ways to uncover implicit knowledge or previously unknown information. ML methods can be instrumental in rapidly changing manufacturing environments because they can easily change and adapt rapidly when necessary. Wuest et al. noted that ML could even be used in manufacturing settings to cut costs and waste by predicting future consumer behavior or outcomes. While Wuest et al. cited numerous benefits of ML, they cautioned that it was critical to ensure that the right ML algorithm was used in any given setting because each manufacturing problem is different.

Challenges Associated With ML Deployments

The literature highlights many advantages to ML in a business context, but there are also many challenges associated with ML deployments (Lee & Shin, 2020). Canhoto and Clear (2020) noted that ML could help organizations save money through improved business processes. However, the authors cautioned that AI, driven by ML, could erode organizational value if managers become overly dependent on the algorithms and lose the ability to recognize and manage risks. Other challenges noted by scholars include poor data quality (Garcia et al., 2020), difficulty working with complex datasets et al, and a widespread lack of algorithm transparency (De Laat, 2017).

Baier et al. (2019) documented consistency between participants’ experiences and the existing ML literature. However, their research did have some limitations. The small sample size meant that they did not interview multiple experts within each industry. Thus, their study might not have produced easily replicable findings. While they were able to triangulate the responses from experts in different industries, comparing responses from multiple participants in each industry would have demonstrated that the participants’ perspectives were representative of their industries. Most of the participants lived in Germany, which might have introduced some bias into the findings. Increasing the number of participants from other countries or surveying equal numbers of individuals from multiple countries might have produced different results. Even with these limitations, the findings aligning with existing literature supported the credibility of Baier et al.’s conclusions.

ML Deployment Research

Research on ML deployment is extensive because of increasing scholarly interest and the rapidly expanding trend among organizations of using business intelligence tools to improve data analysis and organizational outcomes (John et al., 2021). ML deployment research is also diverse because the process of ML development requires constant experimentation with new datasets, models, and tuning parameters

(A. Chen et al., 2020). Because of the diversity, many scholars have suggested different ways to assess ML deployment and the ML lifecycle (A. Chen et al., 2020; John et al., 2021). ML deployments are also fundamentally different from traditional software engineering and deployment projects (Paleyes et al., 2021). Paleyes et al. (2021) noted the main differences as the unique activities required in such deployments, including data discovery, the preparation of datasets, model training, and success measurements. Based on these differences, Paleyes et al. argued that ML deployments require different approaches than the approaches used in traditional software engineering projects.

Some scholars use the term MLOps to describe the development of an ML model deployment lifecycle (Sputh et al., 2021; van der Goes, 2021; Zhou et al., 2020). However, there is little scholarly agreement over the use of the term MLOps, and it is sometimes associated with proprietary frameworks or processes associated with specific types of ML deployments (Sputh et al., 2021; van der Goes, 2021; Zhou et al., 2020). As the present study focused on understanding the foundational processes critical to general ML deployment, the term ML lifecycle was used rather than MLOps.

A. Chen et al. (2020) developed an open-source platform called MLflow, designed to help streamline the ML lifecycle and address deployment challenges. A. Chen et al.'s research was associated with an ML development company called Databricks and based on the company's experience with ML customers. In developing their MLflow platform, A. Chen et al. identified four challenges users experience during ML deployments. The first of these challenges was the multitude of available tools to address each phase of the ML development process. A. Chen et al. noted that ML developers generally want to try as many available algorithms as possible when developing an ML model, adding complexity and development time.

Chen et al. (2020) were not the only scholars to propose an ML lifecycle management system. Melgar et al. (2021) also proposed a system for managing the ML lifecycle. In developing their system, Melgar et al. focused on automating the ML deployment lifecycle rather than simply improving application development steps. Like A. Chen et al. (2020), and Melgar et al. (2021) developed their automated ML lifecycle management system based on their experiences helping customers deploy ML models. Melgar et al. identified Ease.ML as an 8-step process, but the steps were neither well defined nor described.

The elements of Melgar et al.'s (2021) framework differed significantly from the elements of Chen et al.'s (2020) framework. These framework differences illustrate the complexity of ML and the deployment process. Many systems and frameworks are designed to aid in the ML deployment process, yet ML usability remains a critical issue for many users and ML practitioners (Melgar et al., 2021). The source of some of the complexity is identifying what steps should be included in the ML lifecycle to support more successful deployments. The following section identifies elements of the ML lifecycle that served as a focus in the present study. Each element of the present study's ML lifecycle framework was chosen based on extant scholarly literature. Discussions of each element are provided with sources from the literature.

ML Lifecycle

Scholarly disagreement exists regarding ML deployment assessments and what steps precisely constitute the ML lifecycle (A. Chen et al., 2020; John et al., 2021; Paleyes et al., 2021; Xie et al., 2021). In addition to conflicting opinions on ML deployments and the ML lifecycle, the extant literature is also somewhat limited and narrowly focused (Paleyes et al., 2021). For example, Paleyes et al. (2021) surveyed available case studies on challenges to deploying ML and noted that while practitioner reports were available, "the challenges of the entire machine learning deployment pipeline are not covered nearly as widely in the academic literature" (Paleyes et al., 2021, p. 2). Other researchers such as Melgar et al. (2021) and John et al. (2021) suggested the creation of frameworks, programs, or platforms to aid in the deployment process.

A review of scholarly articles on common aspects of the ML lifecycle was conducted to ensure that the present study was grounded in literature. That literature review identified five main components of the ML lifecycle: requirements analysis, data management, benchmarking metrics, user acceptance testing, and privacy policy. The following sections examine relevant literature on these aspects of ML deployments.

Requirements Analysis

The first component of the ML lifecycle identified as part of the literature review was requirements analysis. Before implementing an ML deployment, systems engineers conduct a requirements analysis to identify all necessary project elements (Cho & Lee, 2020). A requirements analysis is critical to the success or failure of system implementation (Paleyes et al., 2021). Cho and Lee (2020) noted that requirements analyses should provide actionable, measurable, and fully documented outcomes sufficiently detailed to implement systems design. Cho and Lee further stated that these outcomes should be supported by testable and traceable logic identifying organizational needs and opportunities. As summarized in a survey of case studies, many scholars and practitioners consider a requirements analysis to be a foundational activity necessary before a deployment can begin, while others consider it the first step in deployment (Paleyes et al., 2021).

At the conceptual level, a requirements analysis consists of three activities: eliciting, recording, and analyzing requirements (Vogelsang & Borg, 2019). Vogelsang and Borg (2019) conducted a qualitative study and interviewed data scientists to determine how requirements analyses are characteristic of or unique to ML deployments. Vogelsang and Borg identified a gap in the body of knowledge and sought perspectives from data scientists to expand the literature on a requirements analysis methodology for ML systems. Using thematic collaborative coding, Vogelsang and Borg documented three dimensions where requirements analyse for ML systems differed from requirements analyses for other types of systems. These three dimensions were related to identifying, documenting, and analyzing an ML system's requirements.

Data Management

Data management was the second component in the ML lifecycle that emerged from the literature review and the first step in the active deployment process (Paleyes et al., 2021). The main focus of the data management step is preparing the data required to build any proposed ML system (Paleyes et al., 2021). Data management can often be overlooked or under-prioritized despite being an early step in the ML lifecycle (Xie et al., 2021). Xie et al. (2021) conducted a systematic mapping study on the ML lifecycle and examined 405 publications between 2005 and 2020. Xie et al. noted that few studies focused on data management and model production problems. Xie et al. recommended that more researchers address system lifecycles from a holistic perspective because of the importance of data management in successful ML system deployments.

Polyzotis et al. (2017) elucidated that data management challenges in production ML are represented by four dimensions (a) understanding the data, (b) validation of the data, (c) cleaning of the data, and (d) data enrichment. Polyzotis et al. noted that a thorough understanding of the data was critical for the engineers involved with any deployment. Specific issues arising from this include identifying and comprehending explicit and implicit data dependencies to avoid introducing new dependencies and supporting a maintainable ML learning pipeline. Polyzotis et al. stated that existing techniques for data-provenance management are often applicable in ML contexts but that there are issues specific to such contexts, such as identifying the short- and long-term impacts of legacy feature removal from pipelines. Additionally, provenance tracking is an ongoing issue due to the highly heterogeneous infrastructure of ML pipelines, as Schelter et al. (2018) noted as an issue in many deployments.

The last dimension of data management addressed by Polyzotis et al. (2017) was enrichment. Enrichment is the process of adding new features or augmenting the training and serving data. Polyzotis et al. noted that the enrichment process had to be addressed carefully as enriching data in a meaningful way was a significant task. An additional issue with enrichment, identified by Polyzotis et al., was fostering a complete understanding of all the ramifications of any proposed enrichment by the deployment team. Such teams need to decide whether proposed enrichment strategies will return the resources required to implement them (see also Kumar et al., 2019).

Benchmarking Metrics

The third component of the ML lifecycle identified in the literature was benchmarking metrics. Relevance was related to measuring important features of an ML system (Dai & Berleant, 2019).

Representativeness was the quality of benchmarks representing performance metrics that both practitioners and academics broadly accept (Dai & Berleant, 2019). Equity was related to benchmarks allowing all systems to be fairly compared (Dai & Berleant, 2019). Repeatability reflected the quality of verifiability with benchmark results being reproducible (Dai & Berleant, 2019). Cost-effectiveness was related to benchmark tests providing good value for the resources expended on running the test (Dai & Berleant, 2019). The next characteristic of good benchmarks was scalability. Good benchmarks should scale from single to multiple servers (Dai & Berleant, 2019). Finally, the last characteristic was transparency, reflecting the need for benchmark tests to be easily comprehensible (Dai & Berleant, 2019). These seven characteristics represent the qualities identified as vital to benchmarks allowing the proper measurement of current-generation ML systems and the ability to evaluate the strengths and weaknesses of these systems (Dai & Berleant, 2019).

User Acceptance Testing

User acceptance testing was the fourth component of the ML lifecycle that emerged as critical during the literature review. Schelter et al. (2018) characterized user acceptance testing as crucial to the deployment of ML systems both in the domain of offline back tests and even more critically assessing models in production. User acceptance testing is performed by end-users or ML system clients (Galli, 2021). Such tests verify that the system functions correctly and as per its specifications when fed client-supplied datasets (Dalton, 2019). User acceptance testing represents the final test on a new system after all functional testing and integration and system testing (Dalton, 2019). The overall goals of user acceptance testing are to identify any remaining defects or bugs in the system, capture changes to existing use cases, and gain client approval (Dalton, 2019).

Best practices in the performance of user acceptance testing include ensuring a strong focus on requirements, building provisions that support testing in the system, and ensuring that usability testing is supported (Pandit & Tahiliani, 2015). Pandit and Tahiliani (2015) also highlighted the need to create comprehensive testing checklists to support best practices in user acceptance testing. Ganesh et al. (2014) noted the end-user expectations system specifications divide as a potential obstacle in user acceptance testing. Ganesh, in the context of that divide, provided the following caveat, “ideally, the acceptance criteria should match what the users think the system should do; however, the acceptance test should only be validated against predefined acceptance criteria, and not against what the users wish the new system would do” (Ganesh et al., 2014, p. 123).

Privacy Policy

The fifth and final component of the ML lifecycle identified as part of the literature review was establishing a privacy policy. The privacy of individuals has been under increasing threat in the last two decades with, sequentially, the rise of electronic data storage, the Internet, the cloud, and, latterly, big data analytics driven by ML (Zhao et al., 2018). ML and its subcategories, deep learning, and neural networks represent algorithms and technologies that present unique challenges to the privacy of individuals due precisely to their ability to analyze multidimensional datasets to create unique and actionable insights into that data (Zhao et al., 2018). Quasi-identifiers such as age or gender may be included where cross comparison may, in some cases, allow the re-identification of an individual (Janmey & Elkin, 2018; Xia et al., 2021; Zhao et al., 2018). With the use of both publicly available data and enhanced datasets purchased from data brokers, ML systems have consistently demonstrated the ability to re-identify individuals from datasets both intentionally and inadvertently (Janmey & Elkin, 2018; Khalfoun et al., 2021; von Thenen et al., 2018; Xia et al., 2021).

Research Method and Paradigmatic Perspective

This study used a qualitative phenomenological research design to study challenges affecting ML model deployment success. Studying ML model deployments using a qualitative methodology was appropriate because the study was primarily concerned with identifying challenges to deployment success based on insights from industry professionals experienced in the phenomenon. Qualitative studies highlight individuals' unique experiences and allow researchers to examine patterns that explain phenomena of interest (Creswell & Poth, 2016). Thus, a qualitative methodology was preferable for this study rather than a quantitative approach focused on the relationships between measurable variables. The main features of qualitative research include smaller samples, in-depth exploration of a complex phenomenon, and narrative data types (Creswell & Poth, 2016). These features are instrumental in generating accurate, comprehensive, and diverse findings (Creswell & Poth, 2016).

Sampling Procedures and Data Collection Sources

Purposive sampling requires searching for participants who have experienced the phenomena of interest (i.e., deploying ML models). The sample frame consisted of ML engineers who reside in the United States. Purposive sampling also requires qualitative researchers to know the target population's characteristics (& Garity, 2008). Purposive sampling methods are well-suited to qualitative studies, as researchers can locate participants who are experts in the study's area of interest (Fawcett & Garity, 2008). The purposive sample for this study consisted of 15 ML engineers with experience deploying at least two ML projects within the last 2 years. Only ML companies that deploy models in production environments were considered part of the purposive sample frame.

Data Collection

In qualitative phenomenological studies, researchers collect data in real-life settings where the phenomenon of interest occurs. The collection of empirical evidence is a major component of a successful phenomenological study. The data collection process should remain open-ended and flexible, so researchers can adapt to unexpected events affecting the study (Leedy & Ormrod, 2020). The study's purpose is fulfilled by collecting data from purposely selected participants who have experience with the phenomenon under study (Creswell & Creswell, 2018).

During the interviews, the researcher took field notes to highlight significant keywords and phrases for reference during the data analysis. At the end of the interview, the researcher saved the data files associated with that interview using participant identification codes. Participants then received a copy of the interview transcripts to review and validate that the transcriptions accurately reflected the content of the interview responses. Following any corrections or additions to the transcripts, the data were loaded into qualitative data analysis software for review.

Data Analysis

Following the preparation of the data, the interview transcripts were examined to determine if there was consistency or irregularity within the participants' responses, and the level of data saturation was assessed to determine if more interviews were necessary (Bryman & Burgess, 2002). Once data saturation was reached, no additional interviews were necessary. During the data reflection, the researcher created code categories to help sort the data. Coding is the process of appointing basic words or short expressions describing the essence of a response (Yin, 2017). The motivation behind coding was to filter information and identify patterns and themes in the data. Comparable responses were indexed utilizing similar codes.

The four criteria for judging the soundness of qualitative research are credibility, transferability, dependability, and confirmability (Amin et al., 2020; Stahl & King, 2020). These criteria provide an alternative to quantitatively-oriented criteria (Amin et al., 2020; Stahl & King, 2020). The four criteria associated with qualitative research better reflect the underlying assumptions involved in phenomenological research (Lincoln & Guba, 1985). These criteria were supported by this research.

Research Findings

This qualitative phenomenological study investigated the challenges associated with machine learning (ML) model deployment in production environments to fill a gap in the body of knowledge and provide scholars and practitioners with a better understanding of the most critical elements of the ML lifecycle. The study answered a central research question: What challenges do organizations face when deploying ML models in production environments? The technology acceptance model (TAM) served as the study's theoretical framework, and a phenomenological approach was used to analyze interview data collected from 15 participants. This section presents the findings of the data analysis.

This section begins with a description of the research setting and the participants. Following the description of the participants, the section presents a project analysis. In the project analysis, the findings were organized in alignment with phenomenological methods. Phenomenology focuses on textural and structural themes to describe a phenomenon. Textural themes describe what individuals experience concerning a phenomenon, and structural themes describe how they experience it. The project analysis includes textural and structural descriptions of each participant's experiences. Following the project analysis, the section includes a section that focuses on how well the study's findings answer the research question. In this section, a combined textural-structural description representing the essence of the participants' lived experiences with the phenomenon is presented.

Participants and Research Setting

The focus of this study was on ML engineers who have experience deploying ML models in production environments. A purposive sampling approach was used to identify and select participants with experience in the study's area of interest (Fawcett & Garity, 2008). As part of the purposive sampling approach, individuals were selected from companies specializing in deploying ML models in production environments. The participants selected for this study lived and worked all over the United States. An effort was made to identify and select experts in the ML field who worked on ML model deployments in production environments within the last 5 years. Participants provided information on their areas of ML expertise and their specific job titles. Participants' job areas of ML expertise included deep neural networks, developing neural network models for financial data, images, videos, text, and time-based data, Python, Java, SQL, Matlab, R, RStudio, Pandas, TensorFLow, scikit-learn, Matplotlib, cloud service, Tableau, data visualization tools, data analysis, data processing, data mining, data security, data management, natural language processing, deep learning, statistics, Azure, RESTful APIs, web scraping, computer vision, data science strategy and development, AI, predictive analytics, visual analytics, extract-transform-load, and automation. Participants' job titles included machine learning expert, data scientist, machine learning engineer, predictive analytics data scientist, team lead for deployment, data security expert, and data analyst.

Project Analysis

Phenomenological research relies on identifying textural and structural themes to describe the participants' lived experiences (Neubauer et al., 2019). Textural themes describe what happens concerning a phenomenon, and structural themes describe how the phenomenon is experienced (Neubauer et al., 2019). The textural and structural themes are then synthesized to create a textural-structural description of the phenomenon's essence. The first step in the project analysis was to identify and describe the textural themes from participants' accounts of their lived experiences. The textural themes are presented in the first subsection. The second step in the project analysis process was to identify and describe the main structural themes associated with the participants' lived experiences. The structural themes are presented in the second subsection. Once the textural and structural themes were identified, then each participant's lived experiences were described using the identified themes, and these descriptions are provided in the third subsection. Five textural themes emerged from the data analysis.

Each textural theme emerged based on the analysis of participants' responses to the interview questions and a coding process that identified units of meaning within the data (Neubauer et al., 2019). The coding process involved identifying initial codes, second-level codes, and patterns resulting from the generation of the textural themes. The initial codes were generated by reading the interview transcripts, identifying meaningful words and phrases, and highlighting them. Second-level codes were then developed by looking for patterns and commonalities within the initial codes. Finally, themes were developed based on the prominence and clustering of second-level codes. The following subsections identify the textural themes and illustrate how the initial and second-level codes were developed based on participants' responses and the coding process.

Textural Theme 1

The first textural theme that emerged from the data was that communication and model adjustment make ML models easier to deploy. This theme is most closely aligned with the TAM construct of perceived ease of use. Textural Theme 1 was developed by reviewing the participants' interview responses and coding the units of meaning within each participant's responses. Reviewing the interview transcripts using the coding process clarified that participants identified communication and the need for model adjustment as challenges during ML deployments. Theme 1 was relevant to all but one of the participants' lived experiences. P1 stated that clients "have different levels of understanding of what machine learning can do." Thus, without proper communication, misunderstandings could easily occur. P2 noted that it was critical for ML engineers to have "good communication skills" because clients and team members often came from "different background[s]."P3 explained that projects were subject to scope creep without effective communication. Scope creep occurs when client expectations and project parameters continuously change because of poor communication. P3 stated, "I think scope creep is the greatest failure."

Textural Theme 2

The second textural theme that emerged from the data was that the ML lifecycle is useful when deploying ML models in production environments. Textural Theme 2 was developed by reviewing the participants' interview responses and coding the units of meaning within each participant's responses.

The data analysis indicated that Textural Theme 2 was relevant to every participant's lived experience. P1 noted that the ML lifecycle provided ML engineers and deployment experts with a way to plan projects. P1 stated, "You need to gather data, and then clean it, and then train a model, and then deploy right." P1 continued, "It's really important [to plan] because when you're deploying a model, you face some challenges." P2 also cited the value of the ML lifecycle stating that several lifecycle stages were "very important when deploying ML models." P3 noted that the ML lifecycle helped regularize deployments because "sometimes it is very easy, some other times it is very difficult." P3 advised ML engineers to use the ML lifecycle to develop a plan. "You have to get a plan, so the planning defines the journey."

Collectively, the participant's responses related to the ML lifecycle support the importance of this tool when deploying ML models in production environments. Each participant identified slightly different stages of the ML lifecycle as most important, but overall, the responses overwhelmingly indicated that the ability to use the ML lifecycle as a tool was critical to project success. As part of the data analysis, participant's responses to the interview guide questions were analyzed to determine links between individual questions and the generation of themes.

Textural Theme 3

The third textural theme that emerged from the data was that ML models require engineering expertise and access to adequate data to deploy. Textural Theme 3 was developed by reviewing the participants' interview responses and coding the units of meaning within each participant's responses.

The essence of Textural Theme 3 was that effective ML model deployment relies on experienced ML engineers and access to quality data. When referencing Textural Theme 3, P1 stated that success "really depends on how much data you have." P1 further noted that "data can be "really expensive, and in a lot of domains, there's not enough data to train powerful networks to predict well." P1 added, "cleaning the data and finding the best way to enter data into the model is a challenge." P2 cautioned that gathering data could be a challenge and that expertise was needed to overcome that challenge. P2 noted that sometimes the data sets included outliers and that ML engineers need to know "what to do to the data set" to make sure no "problem persists in the model." P2 elaborated,

As with the second textural theme, every participant in the sample provided interview responses that contributed to Textural Theme 3. As a result, Textural Theme 3 and the notion that expertise and access to adequate data are useful and make ML models easier to deploy were one of the strongest findings from the data analysis. The challenge of accessing data was experienced differently by some participants, and those differences are examined in more detail in the individual textural and structural descriptions. As part of the data analysis, participants' responses to the interview guide questions were analyzed to determine links between individual questions and the generation of themes.

Textural Theme 4

The fourth textural theme that emerged from the data was that ML models must reconcile with client expectations before and during deployment. Textural Theme 4 was developed by reviewing the participants' interview responses and coding the units of meaning within each participant's responses.

Textural Theme 4 was not mentioned by as many participants as the first three themes, but the data analysis indicated that the need to reconcile client expectations before and during deployment was still a critical element of most participants' lived experiences. When talking about client expectations, P1 offered several comments. P1 started by saying that "it's important to understand what kind of problems can be solved with ML." P1 also stressed the importance of being "realistic." P1 continued, "You need to understand what the expectations of users are when they are using that model."

While Textural Theme 4 was not identified in every participant interview, there were strong commonalities regarding this theme throughout the sample. Several participants highlighted the need to manage expectations, the difficulties in doing so at times, and the consequences if expectations were not managed. More information on how participants managed client expectations during different phases of the ML lifecycle is provided in the participants' individual textural and structural descriptions. As part of the data analysis, participants responses to the interview guide questions were analyzed to determine links between individual questions and the generation of themes.

Textural Theme 5

The fifth and final textural theme that emerged from the data was that well-maintained ML models are easier to retrain during and after deployment. Textural Theme 5 was developed by reviewing the participants' interview responses and coding the units of meaning within each participant's responses. As with the first and fourth textural themes, Textural Theme 5 was not cited by all participants. Nonetheless, over half of the participants acknowledged the importance of maintaining and retraining models during and after deployment. P1 noted that training was a critical element that should be included right from the planning stages of deployment. P1 stated, "I asked for more dates on training, again, and because, like whenever you're training a model, you need to do a lot of things." P1 also observed that training was a time-consuming element of the deployment process. P3 observed that the process of training required judgment and compromise. 'After training, it's, you know, they said that's the attitude. I have a question, and again in ML, there's always a trade-off. Like, am I ready to sacrifice speed for accuracy to sacrifice size for accuracy? And those are the trade-offs we often discuss. "

Summary of the Textural Themes

The generation of the textural themes depended on the participant's responses, the coding process, and the determination of data saturation. Patterns in the data contributed to themes based on the prominence of specific codes and units of understanding. Each textural theme that emerged from the data was identified by at least half of the participants, and two themes were identified by all the participants. The commonality of the participants' responses supporting the emergence of textural themes was consistent across the entire sample. The fact that no new material was provided in support of alternate themes demonstrated data saturation. The determination of data saturation allowed data a collection to end after 15 participants were interviewed. The analysis did not yield any findings that would be considered supplementary or not related to the study's research question.

Structural Themes

While textural themes focus on what is experienced, structural themes are identified as universal elements of a phenomenon (Neubauer et al., 2019). Five main structural themes were identified as part of the research design based on the literature reviewed. These five structural themes highlight critical stages of the ML lifecycle recognized by other scholars.

The structural theme of requirement analysis addressed the need to identify necessary project elements when deploying an ML model in a production environment. The universal structure of requirement analysis was initially identified as critical within scholarly literature. Paleyes et al. (2021) consider a requirement analysis a foundational activity representing the first step in an ML model deployment. Paleyes et al. noted that it was critical for ML engineers to identify all the necessary project elements before starting a project by conducting a requirement analysis. The goal of requirement analysis is to specify actionable, measurable, and fully documented outcomes.

While the structural theme of requirement analysis was identified initially during the literature review, the participants also stressed the importance of this stage in the ML lifecycle. P3 described the requirements analysis stage of the ML lifecycle as the most useful stage. “It is very useful, and if I may say, is the most useful because for requirement analysis, one of the things you want to look at like you would define what you want your model to be like.” P6 provided a similar answer regarding the importance of requirement analysis. P6 explained that during the requirement analysis, developers “create a common set of understandings of what the requirements are.”

Structural Theme 2: Data Management

The structural theme of data management addressed the role of data preparation and use in the successful deployment of ML models. Along with requirement analysis, the universal structure of data management was selected based on a review of the ML deployment literature. Data management refers to preparing and acquiring data when building an ML model (Paleyes et al., 2021). Xie et al. (2021) cautioned that ML engineers and developers often overlooked and under-prioritized data management. Polyzotis et al. (2017) explained that developers often struggle to understand, validate, clean, and enrich data, and this assertion aligned with several of the participants’ responses.

P5 noted the importance of the data management phase of the ML lifecycle both because it was one of the “most critical” phases as well as being “the hardest one.” P9 also cited the importance of data management So, being able to wrangle and move the data around and manipulate it and pull it cleanly and mirrors the way it was developed during model training. It’s useful because it’s the beginning step of the entire deployment process and necessary.

Participant 13 agreed that data management was the most important phase “because what I thought I could do, and what I thought was there will change at that point.” Collectively, 12 of the 15 participants cited the importance of the data management phase of the ML lifecycle, indicating that it did represent a universal structure within the ML deployment process.

Structural Theme 3: Benchmarking Metrics

The structural theme of benchmarking metrics addressed measuring performance when deploying ML models in production environments. Benchmarking metrics was a universal element of the ML lifecycle based on extant ML deployment scholarship. During the benchmarking metrics stage of the ML lifecycle, developers use standard tests and trials to assess the model’s performance (Dai & Berleant, 2019). Often benchmarking metrics measure whether models are relevant, representative, equitable, repeatable, cost-effective, scalable, and transparent (Krizhevsky et al., 2017; Nurvitadhi et al., 2017; Ovtcharov et al., 2015).

P6 described benchmarking as critical and explained that “it allows you to set a standard and essentially be scientific in the test of your performance at each stage to see if you’re having any degradation. P3 provided a detailed explanation of the importance of benchmarking metrics, “By the time you are deploying you have like a set goal that you have defined in your requirement analysis, like, so again, if you have benchmark metrics that working, you know where your most, you know what your minimum should be.”

P12 considered benchmarking metrics a useful “way for helping monitor the real-time performance” of a model. Ten of the 15 participants commented directly on the importance of the benchmarking metrics phase of the ML lifecycle. These comments supported the identification of benchmarking metrics as a universal structure within the ML deployment process.

Structural Theme 4: User Acceptance Testing

The structural theme of user acceptance testing addressed the need to gain user approval when deploying ML models in production environments. User acceptance testing was considered a universal element of the ML lifecycle based on the deployment literature. Schelter et al. (2018) identified user acceptance testing as a crucial part of the ML deployment process, and Dalton (2019) noted that user acceptance testing allowed developers to identify remaining defects in the system, capture system changes, and obtain final approval. The participants' interview responses also supported the identification of user acceptance testing as a universal structure within ML deployment.

When asked about different stages of the ML lifecycle, P1 stated, "User acceptance testing is really important and useful. I mean, you need to know how the user is experiencing this technology. So I think it is pretty useful." P7 described user acceptance testing as "a necessity" and then stated, "I'd say that it's probably the most important thing there in terms of use, and it gets back to the use case, and it's a situation where the client only cares about results and accuracy." P14 referred to user acceptance testing as "super useful," and P15 stated it was "a very useful stage in model deployments." Ten of the 15 participants specifically acknowledged the importance of user acceptance testing as a critical part of the ML lifecycle. The participants' comments buttressed the scholarly literature on user acceptance testing and established this phase as a universal structure associated with ML deployments.

Structural Theme 5: Privacy Policy

The structural theme of privacy policy addressed the need to consider data privacy when deploying ML models in production environments. Like the other structural themes, the literature review contributed to selecting privacy policy as a universal theme relevant to the ML lifecycle. Zhao et al. (2018) noted the importance of incorporating privacy policy best practices within the ML deployment process to address rising concerns about data privacy resulting from big data analytics. Several scholars also noted the trend toward privacy legislation in the United States and abroad (Pohle, 2018). Based on the extant scholarly literature, privacy policy was initially identified as a universal structure, but participants' responses supported privacy policy as a structural theme in the data.

P2 and P3 both described the privacy policy of the ML lifecycle as "very useful," with P3 noting that "the policies define what you do." P8 noted that privacy policies were useful because they helped define the parameters of a project, stating, "If they're well-defined, like 'you can and cannot do this,' that's useful, but if it's not well defined, then it's a hindrance." P14 explained that it was often important to ensure that models were "not collecting data that you shouldn't be." A total of 10 out of 15 participants cited the importance of the privacy policy stage of the ML lifecycle. These statements supported the scholarly literature on the role of privacy policies in ML deployments and the use of this stage as a universal structural theme within the present study. More information on the intersections between the textural and structural themes is in the following descriptions of the participants' lived experiences.

Descriptions of Participants' Lived Experiences

This subsection identifies the lived experiences of each of the study's 15 participants. To create a holistic description of the experience, a textural description of each participant's experience is provided, using the relevant textural themes to identify what each participant experienced. The textural description is followed a structural description that focuses on how the phenomenon was experienced in the context of the five universal structures of the ML lifecycle: requirement analysis, data management, benchmarking metrics, user acceptance testing, and privacy policy. Following the individual participant's textural and structural descriptions, a combined textural-structural description of the phenomenon is provided as part of the analysis of the research question.

P1's Textural and Structural Descriptions

All five textural themes were relevant to P1's lived experience working with ML deployments in production environments. When talking about Textural Theme 1, P1 noted that communication was so critical because clients "have different levels of understanding of what machine learning can do." P1 also noted that communication is also essential for ML engineers. “It’s really hard to even for machine learning experts. It’s really hard to know how much data is needed for solving a particular problem.”

Regarding Textural Theme 2, P1 explained that the ML lifecycle was useful because developers can use it “to gather data and then clean it and then train a model and then deploy it right.” P1 went on to explain that “if you have a plan, the chances are, your success is high.” P1 concluded that the ML lifecycle was important because when developers implement ML models in production environments, they “face some challenges.”

P1 also cited Textural Theme 3 when talking about the necessity of expertise and access to adequate data. P1 indicated that data volume and quality impacted the success of a deployment. “It depends on how much data you have, and a lot of these problems can be solved with huge amounts of data.” P1 also talked about the costs of expertise and data. “It’s really expensive, and in a lot of domains, there’s not enough data to train powerful networks to predict well.”

Textural Theme 4, which addressed reconciling client expectations, was also relevant to P1’s lived experiences with ML model deployment. P1 explained that ML experts “need to understand what the expectations of the users are when they are using the model.” P1 also noted that clients have to “be realistic about their expectations” when considering what can be solved by ML models.

Regarding the final textural theme, P1 agreed that well-maintained ML models are easier to retrain during and after deployment. Recounting his experiences on one project, P1 stated, “I asked for more dates on training again, and because like whenever you’re training a model, you need to do a lot of things, you need to clean the data, and then you need to tune the hyperparameters of the ML model, and then do it over and over again. It’s time-consuming.”

P2’s Textural and Structural Descriptions

A review of the interview data indicated that the first four textural themes were central to P2’s lived experiences. Regarding Textural Theme 1, P2 found communication critical, especially when working to resolve misunderstandings with clients. P2 noted that clients often do not state their objectives clearly. “When you apply your ideology to the approved agenda, they come and say, “No, this is not what we asked for. This is different from what I expected.” Sometimes, it makes you look as if you don’t know what you’re doing, but sometimes they don’t even get a chance to make you continue to work because this is not what they want and don’t give you a chance.” For these reasons, P2 felt ML engineers needed to develop good communication skills.

In support of Textural Theme 2, P2 highlighted the value of the ML lifecycle as a tool when deploying ML models. P2 specifically characterized the lifecycle as “very important when deploying ML models.” P2 also noted the value of planning and documenting communication between clients and ML engineers during several different lifecycle stages.

One of the most prominent themes for P2 was Textural Theme 3, which dealt with the need for engineer expertise and access to adequate data during an ML deployment. P2 noted that gaining access and permission to use data could be a challenge when deploying ML models. P2 gave the example that some respondents “don’t want to share age, which some of them just put any age.” P2 explained that this could cause problems in training the model. P2 also noted that outliers create problems with datasets, and in these instances, refining a model “goes from very easy to a very difficult task.” In these instances, P2 explained that an engineer’s expertise could make a big difference.

The last theme represented in P2’s lived experience was Textural Theme 4, which indicated that reconciling ML models with client expectations is useful before and during deployment. P2 noted that when models were not reflective of client expectations, engineers “have to go back all over again to start doing a lot of troubleshooting.” For this reason, P2 advised that ML engineers should be cautious about communicating with clients and using the ML lifecycle structures to improve alignment between project goals and client expectations.

Like P2, four textural themes were relevant to P3’s lived experiences. However, the four themes that emerged from P3’s interview data were Textural Themes 1, 2, 3, and 5. When discussing the first theme of communication, P3 noted that projects are people-centered. P3 stated that success “depends on the people and if the people on the project can manage the stakeholders." P3 specifically noted the issue of scope creep when discussing communication and noted that communication was a requirement to avoid a project's scope from going "beyond what was defined at the user requirements phase." P3 noted that clients remain aware of problems with good communication, allowing developers to move on to production faster and more successfully.

P3 noted that the ML lifecycle was useful when deploying ML models in production environments. P3 stated, "It is very useful because again, it's that this thing comes from the idea of all your work, all the ML training, deployment, and everything." P3 was very adamant about the value of the ML lifecycle starting toward the end of the interview that "the planning phase decides project, or so I like to tell my colleagues."

Textural Theme 3, which addressed the need for expertise and data access, was part of P3's lived experience. When considering specific data expertise, P3 noted that it was critical for ML engineers to examine questions like "How do you manage the data? How do you talk about collecting the data? What are you collecting?" These questions were important because P3 noted that "some [data are] easy to get. Some are difficult, especially depending on the data sets you want to use."

The final textural theme relevant to P3's lived experience was Textural Theme 5, which addressed maintaining and retraining models during and after deployment. "After training, it's, you know, they said that's the attitude. I have a question, and again in ML, there's always a trade-off. Like, am I ready to sacrifice speed for accuracy to sacrifice size for accuracy? And those are the trade-offs we often discuss."

From a structural perspective, P3's interview data related to four universal structures: requirements analysis, data management, benchmarking metrics, and user acceptance testing. The importance of the requirements analysis phase emerged in discussions of Textural Themes 2 (communication) and 3 (access to data). During the data management phase, the structure of the ML lifecycle and access to appropriate data were the most critical textural themes. The same was true during the benchmarking stage, but maintenance and retraining of the model was also an issue experienced when using benchmarking metrics. P3 indicated that communication was most critical during the user acceptance testing phase.

P4's Textural and Structural Descriptions

All five textural themes were evident in P4's lived experiences with ML model deployment in production environments. P4 cited Textural Theme 1, which aligned with the TAM construct of perceived ease of use and addressed the need for communication and model adjustment when discussing client expectations. P4 noted, "I think people have realistic expectations, but sometimes people, they're looking for high-level ML, and we're just not there yet." P4 continued, "You know, I talked to clients, and they want like 100% accuracy for the forecast, and then question, like what you need to get to that level requires a ton of work." P4 noted that sometimes it was impossible to meet client expectations, and communication was imperative when that happened.

P4 noted that the ML lifecycle had value, but he also believed there was room for improvement. "I don't think there's been much careful planning about the ML lifecycle. I think people just don't think about it." Despite this observation, P4 was hopeful for improvement. "I think better methodology will come about in terms of how to do lifecycle planning."

Regarding Textural Theme 3, P4 observed that it was imperative to "know what data is available." P4 explained, "Most of the time, data is all over the place, and that's a problem." P4 identified labeling as a significant data access problem. "It's just a lot of labeling problems. It's hard to get that high-quality data. In some other projects, there are just holes in the data. Right? There are gaps. And then how do you properly handle the holes?" P4 believed that more ML engineer expertise was needed to address data access and labeling problems.

As part of his lived experience and in support of Textural Theme 4, P4 noted that ML engineers need to reconcile ML models with client expectations. P4 observed, "You know there's kind of what people want and then what technology can do. You know, sometimes people want this extreme accuracy, which is impossible without a bigger dataset right? Or people underestimate how much it's going to cost to get all the data labeled right. That's also a considerable cost." P4 asserted that the failure to reconcile ML models with client expectations inevitably caused problems for developers and dissatisfaction among clients.

The final textural theme, which was associated with maintaining and retraining models during and after deployment, also factored into P4's lived experiences. P4 noted that fresh data was necessary when retraining a model. P4 advised that "you want to keep [data] organized; you want to keep it clean" when training and maintaining a model. "Let's continue updating and refreshing the data and kind of managing the incoming data and how you handle that and how you retire the old data and how you clean it up." P4 concluded by saying, "Yeah, data management is just, it's not something people think about is that it's a continual process."

P5's Textural and Structural Descriptions

Like many participants, P5's lived experience with ML model deployment included every textural theme. Regarding the theme of communication and model adjustment, which aligned with the TAM construct of perceived ease of use, P5 observed that it was essential to constantly solicit feedback from clients because in doing so, developers and ML engineers can ensure that their projects run smoothly. "If I'm talking to and really soliciting feedback consistently from the people I'm building the model for, they should have a hand in designing the system. So, it should be an instance where there are no surprises in the system to them. They shouldn't come back and go, "What's all this stuff, you know why, why is it doing this?" It should be, "Yes, it's doing the thing that I had wanted it to do.""

In support of the second textural theme, P5 identified the importance of the ML lifecycle, stating, if you don't plan at the beginning, you're going to fail." P5 explained that there are "a lot of moving components" within an ML deployment, and because of that, the structure provided by the ML lifecycle was beneficial for ML engineers. "It's most useful because it sets up the success for all of the other steps."

Referencing Textural Theme 3, P5 noted that ML models relied on "high variety, high volume, and high velocity" data but that the most critical data consideration was finding the appropriate data for the model. "The biggest data management challenges are going to be making sure the data you put in is correct," P5 explained that ML engineers make assumptions about what data is needed, and if those assumptions are incorrect, the models will not work well.

The fourth textural theme addressed reconciling ML models with client expectations. For P5, this process relied primarily on trust.["Clients] either are highly skeptical of machine learning, so there's a trust factor, or they think it's going to do you know solve all their problems like they're just gonna they're not going to need to do anything they're just the machine learning going to tell them what to do the model is going to tell them what to do and so really kind of tempering their expectations and helping them, you know, identifying what pain points they have." P5 stressed the importance of helping clients develop reasonable expectations. "That's the hardest part for the more technical requirements."

Finally, P5 shared thoughts regarding the final textural theme of maintaining and retraining models during and after deployment. "It's common now to retrain your model pretty consistently." P5 continued, "Training time is a huge consideration. If you can't retrain your model in less than 30 minutes or an hour, it's prohibitive. You need to find another model." P5 further explained that reducing training time was necessary because of the large volumes of data the models had to process.

The ML lifecycle stages of requirement analysis, data management, benchmarking metrics, and user acceptance testing were most important to P5's lived experiences from a structural perspective. The textural theme of communication was most relevant during the user acceptance testing phase of the ML lifecycle. P5 identified the lifecycle's requirement analysis and data management phases as the most useful. Expertise and access to adequate data were primary concerns during the requirements analysis and data management stages. P4 indicated that managing client expectations was crucial during the requirements analysis stage. Finally, the maintenance and retraining of the ML models generally occurred during the first three stages of the ML lifecycle based on P4's experiences.

P6's Textural and Structural Descriptions

Each of the five textural themes was evident in P6's interview response. Regarding Textural Theme 1, P6 noted that it was critical to communicate about ML model deployment elements associated with requirement analysis, benchmarking metrics, and user acceptance testing. P6 noted, "There's a lot of a nuance in requirements, so we generally have to go back and forth to get more detail in the requirements." P6 explained that the goal of communication was to "create a common understanding" between clients and developers.

When discussing the second textural theme associated with the value of the ML lifecycle, P6 observed that the most critical phases included requirements analysis, data management, and benchmarking metrics. Explaining how the ML lifecycle helped with deployment, P6 noted, "Let's organize the data, put it into, you know, really plan out the data architecture, so that we could then have a habit efficient running model." Emphasizing the early stages of the model, P6 noted that "planning is important from the beginning." P6 suggested that the main benefit of the ML lifecycle was that it "saves a whole lot of headache in the future."

P6's responses included extensive support for the third textural theme that expertise and access to adequate data are useful and make ML models easier to deploy. P6 explained how "understanding the capabilities of the data or even the availability of the data" was critical to success. P6 also noted that expertise was necessary as data was often "siloed" in "legacy systems," requiring experienced engineers. P6 concluded that "just finding people that even have that domain-specific capability[y]" was both problematic and also critical to model success.

P6 cited the importance of reconciling ML models with client expectations. P6 observed that "It does take quite a bit of time to make sure we understand the, you know, the criteria because it impacts what we're looking for in the requirements." P6 highlighted the importance of taking time with clients to go over their expectations, "The biggest challenge is humans. Because what is acceptance sometimes changes, so you would like to define your acceptance testing in the requirements phase. So there are a couple of problems with that. One is the people doing the requirements analysis are often not the ones writing the acceptance test, so that's a bit of a challenge."

The final textural theme relevant to P6's lived experience was that well-maintained ML models are easier to retrain during and after deployment. P6 noted that training was both necessary and time-consuming. "It takes some rework and monitoring in the production environment, but for the most part, I wouldn't call [training] the most difficult phase." P6 also emphasized the importance of new data when training models, "What I found after running my early models was that what you have to plan for is future model degradation because over time, if you are not ingesting new data, you know, then the accuracy and the effectiveness of the model decline."

Examining the structural themes related to P6's lived experience indicated that all five universal structures associated with the ML lifecycle were relevant. The textural themes associated with communication and model adjustment (perceived ease of use), the usefulness of the ML lifecycle as a planning tool when deploying models (perceived usefulness), the need for expertise and access to adequate data (perceived usefulness and ease of use), and the need to reconcile models with client expectations (perceived usefulness) were all present during the requirement analysis phase. In the data management phase, the main textural themes that defined P6's experience included the value of the ML lifecycle, the need for expertise and data access, and the need to maintain and retrain models during and after deployment. Textural Themes 1, 2, 3, and 5 were also relevant when using benchmarking metrics. Textural Themes 1 and 4 were associated with user acceptance testing, and the privacy policy phase of the ML lifecycle was most closely associated with Textural Theme 3.

P7's Textural and Structural Descriptions

Each of the five textural themes was relevant to P7's lived experience deploying ML models in production environments. P7 specifically highlighted the importance of communicating with clients to understand their project needs. P7 indicated that he spends time "looking at what the client is looking to get out of the project." P7 continued, "Most of the time, it's just figuring out what the client wants, but first of all, what is the data available, and then as well, what is the client seeking to predict based on the level or kind of behavior [the data contains]. "

P7 also offered feedback regarding the value of the ML lifecycle as a tool for model deployment in support of Textural Theme 2. P7 felt that the value in the ML lifecycle was partly because of "the ability to make sure the project is actually on schedule. I would say that the logistic aspect of it is the most useful part for scheduling purposes." When citing specific phases that were most useful, P7 identified the data management, benchmarking metrics, and privacy policy stages.

Textural Theme 3, which addressed the need for ML expertise and access to adequate data, was highlighted in P7's interview responses. P7 noted that it was a common challenge "when the customer wants to study data that just isn't there." P7 highlighted the lack of available data and scalability as significant challenges facing ML engineers.

P7 contributed to Textural Theme 4 by noting the iterative nature of the reconciliation process when working with clients. P7 stated, "Some cases are easier than others, but there's also a lot of back and forth when you're getting into feature engineering and especially when becoming more exploratory with your project." For P7, satisfying customers was critical when finalizing a model and deploying it successfully.

The final theme mentioned by P7 addressed the maintenance and retraining of models. P7 stated, "I think the introduction of new data is always going to be a factor and especially if you're looking at retraining and model potentiality in the future." P7 suggested that the real issue was how often to retrain a model rather than whether to retrain it, "The particular challenge there was deciding at what interval we regenerate new reports to a new time series forecast to deliver to this theoretical client. Again, it was just a proof of concept, but I think the challenge there regarding data management was just deciding at what intervals to kind of retrain new forecasting, for example."

Examining the structural themes relevant to P7's lived experience demonstrated that each of the ML lifecycle stages was associated with specific textural themes. Textural Theme 1, which dealt with communication, was most closely associated with the structural themes of requirement analysis and user acceptance testing. When discussing Textural Theme 2, which identified the ML lifecycle as useful, P7 focused primarily on the universal structures of data management, benchmarking, and privacy policy. Textural Theme 3, access to adequate data, was closely linked to requirement analysis, data management, and benchmarking metrics in P7's lived experiences. The fourth textural theme was linked to requirement analysis, while Textural Theme 5 was linked to data management.

P8's Textural and Structural Descriptions

Each textural theme was relevant to P8's lived experiences. Referencing Textural Theme 1, P8 explained that communication is critical because sometimes "the customer doesn't always know what they want." P8 recounted that a lack of user understanding, specifically during the user acceptance testing phase, can cause problems without effective communication, "I would say, and trying to reason with someone who's using an emotional judgment or a psychological bias, that's, that can be heard, but for those who have been around the block and know what to reasonably expect from ML, if they have a scientific way of doing user acceptance testing, then it's easy."

Like many other participants, P8 supported Textural Theme 2, identifying the ML lifecycle as a useful planning tool. P8 said, "You have to think ahead. What are the biggest risks going into this project." P8 noted that the ML lifecycle helped developers "identify the high-risk items, taking care of those first before more heavy investments happen." P8 concluded, "When you've got a lot of parts that depend on each other, and you've got limited resources, and the deadline of a nontrivial project, you have to look ahead."

When discussing the importance of accessing adequate data (Textural Theme 3), P8 noted a different challenge than other participants. This challenge was data trustworthiness. P8 stated, "Just because you see a number doesn't mean you can trust it." P8 explained that ML engineers need to have enough experience to understand when to trust data and test it effectively. P8 concluded, "You can't know that just by looking at a number, right?"

P8's interview responses highlighted the importance of Textural Theme 4. P8 stated that reconciling client expectations depended on "how reasonable their expectations are." P8 cautioned against ML engineers working with unrealistic clients, "Hopefully, you never even start working with them once you find out their head is in the clouds, but if you are, then you're just in a bad spot, and I don't see how it's going to end well, so I think it's entirely dependent on how informed the end-user is because ML is kind of a newish thing right? So there isn't a standard. There isn't an established way of saying whether this work is good enough or not. So given that, I'd say, on average, it's harder than other fields."

Regarding Textural Theme 5 and the need for model training, P8 focused on the lack of standard solutions when structuring a training approach. P8 stated that he did not feel that keeping multiple model versions was the best practice, “There are a lot of scattered solutions for model versioning. Usually, I have not found it necessary to keep more than the previous model around because as long as the model is better than the last one, I don’t remember a time where we’ve had to revert, say four versions, to get things working properly again. Maybe that’s happened to someone. I just haven’t experienced it.”

Examining the structural themes associated with P8’s lived experiences demonstrated a link between each of the five universal structures and the study’s textural themes. Specifically, the universal structure of requirement analysis was associated with Textural Themes 1, 2, and 5. This intersection indicated that P8 believed communication was critical to the requirements analysis, that that stage in the ML lifecycle was particularly useful, and that the requirements analysis was critical to retaining and maintaining ML models. Textural Themes 2, 3, and 5 were relevant to both the data management structural theme and the benchmarking metrics structural theme. The similarities between the theme interactions in those two categories suggest that data management and benchmarking metrics may have been closely linked for P8. When examining the universal structure of user acceptance testing, the most important themes were Textural Theme 1 and Textural Theme 4. P8 indicated that Textural Theme 2 was the main factor when working on privacy policy issues.

P9’s Textural and Structural Descriptions

An analysis of P9’s interview data supported the first four textural themes. Concerning Textural Theme 1, Participant 9 advised that ML engineers needed to take a proactive approach to communication. “So I know it’s very generic, but it starts as a conversation with the stakeholders, more of a consultative approach rather than, you know, talking directly about analytics off of that.” By focusing on communication, P9 explained that it was possible to answer the question “what are they [clients] exactly trying to solve?” P9 also noted the importance of continued communication and flexibility, “It’s kind of a two-way arrow between business and data understanding because sometimes we go to the data and we have more questions about the business. Sometimes we go to the business, and we have more questions about the data, and then, obviously, some things are discovered. You know, whether it’s errors or data understanding things that the business can tell us about or business understanding things that we’ve developed that we can communicate to the business because of the data. “

P9 provided a brief comment in support of Textural Theme 2. P9 stated, “If there’s no plan and there’s no process, then I don’t think you even set yourself up for an opportunity of success.” When discussing the ML lifecycle’s value as a planning tool, P9 cited the importance of each stage in the ML lifecycle.

P9 also highlighted the role of ML engineers’ expertise in comments related to Textural Theme 2. When asked to identify the most critical aspects of the ML lifecycle, P9 replied, “Data prep. If it’s not done responsibly and clearly and accurately, then you’re not going to be able to create a model that’s reliable or useful, or accurate in production. In which case, everything you do post-data prep is useless if data prep is not done in a way that’s fair and generalizable and responsible.”

P9’s interview responses also supported Textural Theme 3. P9 noted that ML model deployment could be challenging because “sometimes you can’t get data points that you want.” Specifically, P9 cited this challenge within the healthcare industry, “So, like for images, for example, a lot of images that are taken for patients have patient health information attached to them so whether that’s their date of birth, their name and the physician’s name. The facility that they’re currently attending or pictures of themselves or tattoos and jewelry all that stuff actually has to be blurred out in the image before the image can be fed into a machine learning model which makes deployment a little bit more complex than data that’s not sensitive in that in that case.”

Referencing Textural Theme 4, P9 highlighted the differences between client and ML engineers’ perspectives and the importance of bridging that gap, “I think it’s challenging because sometimes people try to answer people that are not analytics-minded or aren’t doing the analysis, don’t know how to interpret it, or think that a model can answer more questions than just the one that was originally discussed. And sometimes, the visualization changes because when it’s shown in the product, people will change their minds. So, I wouldn't say the word difficult. I would say it's more like an ever-evolving and changing process."

When viewing P9's lived experiences from a structural perspective, all five phases of the ML lifecycle constituted universal themes relevant to the participant's experiences. Communication was most important in the requirements analysis and data management structural themes. P9 highlighted all five universal structures as useful when discussing the value of the ML lifecycle. ML models' reliance on expertise and adequate data was most relevant in the data management, benchmarking metrics, and privacy policy phases of the ML lifecycle. The final textural theme that P9 cited, reconciling client expectations, was identified as necessary during the requirement analysis, user acceptance testing, and privacy policy stages of the ML lifecycle.

P10's Textural and Structural Descriptions

P10 cited four of the textural themes as part of his lived experience. The four themes relevant to P10 were Textural Themes 1, 2, 3, and 5. Referencing Textural Theme 1, P10 highlighted the role of communication both before and during the deployment. "Being able to speak to customers before you even start." P10 referred to early communication as "customer interviewing" and noted that the goal of that process was to "really understand what it is the customer needs." Referring to customer interviewing, P10 asserted, "I think that's the most important step because the rest of the process becomes a bit simpler because you have a very clear target."

Discussing Textural Theme 2, P10 strongly supported the value of the ML lifecycle as a planning tool. P10 stated, "Planning the ML cycle is a very important step. It's something I've always done." P10 continued, "It's useful to plan ahead of time and then use that plan as like a guideline and then be open to the plan changing but always start with a plan."

P10 provided several insights related to Textural Theme 3. P10 suggested that the volume of data could also cause difficulties for ML deployments because it could be overwhelming, "when you're working with data that is coming in quickly, let's say your users are producing 10,000 data points an hour, you know or more, I think that is where it gets a little more challenging. Because then you have to start looking at tools that are built for this like a continuous stream of data and being able to organize it in real-time, and I think that's where it gets more challenging."

P10 suggested that ML engineer expertise improves data handling and access through careful system design. P10 clarified, "The best system that you can design is one that can pull the new data automatically into your data management system and then actually automatically send the new data up to the ML model."

Finally, referencing Textural Theme 5, P10 observed that monitoring and training were essential to validating an ML model and ethically analyzing data. P10 stated, "Just thinking about things like feedback loops, you know. If your model continues to train, maybe more of the ethics side. This stuff I would think about." P10 concluded by stating that "I would say, setting up these monitoring tools is pretty straightforward."

Like most of the other participants, all five structural themes were relevant to the lived experiences of P10. The requirements analysis phase of the ML lifecycle resonated with Textural Themes 2, 3, and 5, meaning that during this phase, P10 focused on the value of the ML lifecycle, the need for expertise and appropriate data, and the process of maintaining and retraining models. P10 related the data management phase of the ML lifecycle to the importance of communication, the value of the ML lifecycle, the need for appropriate data, and the need to maintain and retrain models. P10 cited the value of the ML lifecycle, the need for adequate data, and the process of maintaining and retraining models as the most relevant issues when talking about the universal structure of benchmarking metrics. Communication and the value of the ML lifecycle were the most prominent textural themes associated with the user acceptance testing phase, and the privacy policy phase was linked to the overall value of the ML lifecycle.

P11's Textural and Structural Descriptions

P11's interview materials reflecting the participants' lived experiences deploying ML systems in production environments embodied the first four textural themes. Although P11 only touched on communication with clients briefly, P11 emphasized the consequences of poor communication noting a case in which all benchmarks were met, but the client would not accept the system, having developed a different understanding of what it was. P11 further noted, “you want to catch” problems early to avoid redoing or scrapping projects. In that context, P12 noted that “people, are to date, the most complicated machine” but that, if asked, “People are quite happy to express their concerns.”

Textural Theme 2 was the most dominant in P11’s interview materials and thinking. P11 placed a heavy focus on the importance of requirements analysis. P11 noted that there was “no replacement for getting into the real world” and that getting the requirements for a project right was vitally important. However, P11 also warned that “that requirement which you know it’s made with the best of intentions” could turn out to be “horrible” because not all facts and circumstances are knowable before a project deploys. P11 characterized data management as “hugely” important while noting that 99% accuracy was not enough in projects in which 1% could represent human mortality rates and 100% was challenging. P11 characterized the benchmarking step as much easier as it was a ‘quantifiable’ activity. P11 also noted that benchmarking helped keep teams focused and in alignment. P11 noted user acceptance testing as a useful step but characterized privacy policy as “difficult, drudgery, painful, necessary!”

P11 emphasized the importance of Textural Theme 3, stating, “data is the new oil, and it’s not just a little bit of data it’s, not even a lot of data, its massive data and that you do get to that you need big teams and new technology hadoop, spark, Haskell, scala.” P11 also noted that the necessary expertise did not yet exist for managing the risks inherent in excess data. “The data security component is much more complicated than really anyone predicted because it’s not just humans anymore it’s really you know our models are learning things we don’t want them to learn that exist in the data and how to solve that is, you know we have federated learning a lot of attempts to deal with these things, but we have not yet they’re not solved, we have some fixes but it’s tough.” (P11) P11 also noted in the context of access to data that there were challenges inherent in who could get access. P11 specifically noted that organizations with privileged data access “can win, essentially, whereas those who aren’t able to get access to sensitive data have almost no recourse which a lot of societal impacts there.”

Textural Theme 4, which addressed reconciling client expectations, was also relevant to P11’s lived experiences with ML model deployment. However, P11 noted that good client communications also relied on trust. Trust was critical for clients and end-users. P11 stated,

> Trust, and safety those kinds of go hand in hand, both in you know literal you know, will somebody use this application if they know there’s AI involved? Or you know, maybe technology, but also that have, you know, for instance, robotics and the quite literal physical safety and trust, confidence, calmness, that is risk and certainly hard to get back from users and employees.

P11 highlighted that a key component in ML models is trust, especially when these models use increasingly hard-to-comprehend algorithms. Users must trust that ML developers are knowledgeable and will deliver a system that will not violate data privacy or other regulatory requirements. P11 emphasized that communication relies on trust to function and that this trust dimension continues to increase in importance.

From the perspective of a structural description, all five phases of the ML lifecycle were relevant to P11’s lived experiences. In the case of P11, all five phases, identified as universal structures in the literature, contributed to the ML lifecycle’s value. P11 also linked the universal structure of data management to the textural theme associated with access to appropriate data. In addition to supporting the value of the ML lifecycle, user acceptance testing also was linked to communication and the need to reconcile ML models with client expectations. P11 linked the universal structure of privacy policy to the ability to access appropriate data.

P12's Textural and Structural Descriptions

The first four textural themes were found in the interview data of P12's lived experiences. P12 highlighted the critical importance of client communications. P12 stated that "understanding the business and making sure me and the client are on the same page" were crucial to ensuring project success. P12 also noted that good client communications at the beginning of a project ensured that "both sides are on the same page in the beginning," which was also helpful in defining the business problem. In the dimension of business analytics, P12 noted an important part of the job was, "to make sure the client understands and accepts and metrics you come up with and so it's like in if the client also has also had the data science no knowledge, it will be easier to translate, but most of the time they don't so it's challenging for the data scientists to translate the technical or words into business words and make that make them understand."

Concerning Textural Theme 2, P12 made several remarks. P12 noted that requirements analysis was "very important" and that "planning is it's very important for the whole life cycle." P12 observed that without requirements analysis, it was quite possible for "the whole time" spent on a project "to be wasted." P12 stated that the second step in the ML lifecycle, data management was "the most important thing when it comes to machine learning" and that "data management could help with production." P12 also noted the importance of benchmarking metrics, stating that they are useful "for helping monitor the real-time performance" of a system. P12 noted that user acceptance testing was particularly helpful in combination with benchmarking. Overall, P12 opined that "the lifecycle itself is a product and then making sure like you know the businesspeople know how to use it it's another part, so I'll say 50 to 60%. It's important for the business to succeed."

P12 touched twice on Textural Theme 3, which dealt with the need for engineer expertise and access to adequate data during an ML deployment. P12 noted that "it's hard to find a way to aggregate or clean the data and well eventually will need to set our pipeline, but they will require a very big amount of data size." This comment echoed other participants who noted the increase in difficulty for data management when handling massive datasets. P12 also noted the need for expertise in handling data at user acceptance testing, stating that in cases where data was not well handled, "the data will receive in real-time does not align with what we see in the development cycle, then the acceptance testing won't be passed, then the result won't look good."

The final textural theme found in P12's interview materials was Textural Theme 4. P12 noted the importance of client communication to a project, stating, "Have like good talk to the client, and then that would be helpful." P12 returned to client communications later in the interview, noting the importance of having "a good template to talk to the client to agree with privacy policy and then for data management or how to set up the database."

The structural description of P12's lived experience included the four elements of requirement analysis, data management, benchmarking metrics, and user acceptance testing. While dealing with the requirement analysis stage of the ML lifecycle, P12 indicated that the themes associated with communication, the value of the ML lifecycle, and the need to reconcile models with client expectations were the most relevant. When discussing the data management phase, P12 cited the themes associated with the value of the ML lifecycle and the need for ML engineer expertise and access to adequate data. The benchmarking metrics phase was associated with the need for client communication and the value of the ML lifecycle. User acceptance testing, the final structural theme of importance to P12, was linked to communication, the value of the ML lifecycle, and the need for adequate data.

P13's Textural and Structural Descriptions

The first four textural themes were relevant to the lived experience of P13 working with the deployment of production-based ML models. P13, when discussing Textural Theme 1, highlighted the importance of ongoing communication with clients, or else the number of model adjustments would substantially increase. P13 relating his experiences in this regard with one client, stated, "it made it challenging because they changed what they wanted and when you're into a model getting having to go back and rework it is, is quite time-consuming." P13 also provided an example. "I had a model, complete and had done all the work, and then, when we went to deploy it in the deployment environment and also the data which they were going to use was completely different than what we had originally and what I had trained the model on."

P13 also highlighted the importance of communication at the user acceptance testing stage. P13 stated,

> "I've learned through doing different models, and working with different groups that the closer I keep them in the loop of what's happening, meaning the closer I work with them and the requirement phase, the more information, I give them and enable the show through analyzing the data and the data management and the things that had to be done in the database and showing them the different models along the way, and keeping them abreast of kind of where I am and the model and why and some changes, the more I include them in the earlier steps, the easier acceptances."

P13, in the context of Textural Theme 2, elucidated the value of the ML lifecycle as it encourages developers to find the "business reason or you know, a reason for doing a model to understand what the purpose" of that model is. P13 also noted that it encouraged the developers to "spend a lot of time getting to know the data" and to figure "out if I know where it is insufficient, where, perhaps I need to clean it up." P13 also characterized the ML lifecycle as "critical, because if you don't, if you don't know." Finally, P13 noted that in a good project, "a good 70% of the time is spent upfront before we before I even touch training modeling your development at all."

P13 echoed Textural Theme 3 when noting, "one of the biggest challenges is getting the data in a usable form and making sure that I have complete data." P13 also noted that the biggest challenge was "just the size of the data it's unruly" and "the sheer size of the data is limiting." P13 noted the need for expertise. "If you can't control and rely on your data, then you don't have a workable model and so to establish a structure and a flow of data and even look into new software systems that can handle it is vital because your machine learning is nothing without the data."

Concerning Textural Theme 4, P13 highlighted the potential consequences of not maintaining ongoing client communication. P13 provided an example of a client who started a project with "one idea in mind of what they were looking for and halfway through it changed again and by the end, they wanted something different," resulting in a "really challenging" and "quite time-consuming" project. P13 also noted that client communication was both challenging and necessary to align "reality and what we can do and what [the client] want[s] and finding a happy medium." P13 noted that in ML, "you can't meet all the demands and all the expectations all the time."

P13's structural description included interactions between four textural themes and all of the universal structures within the ML lifecycle. As with many other participants, P13 identified communication, ML lifecycle value, the need for adequate data, and reconciling models with client expectations as important issues to resolve during the requirements analysis phase of the ML lifecycle. The universal structure of data management was related to the value of the ML lifecycle and access to adequate data. P13 associated benchmarking metrics with the value of the ML lifecycle, access to adequate data, and the ability to reconcile client expectations. User acceptance testing was linked to the textural themes of communication and the value of the ML lifecycle, and the privacy policy structural theme was most closely linked to P13's discussion of the value of the ML lifecycle.

P14's Textural and Structural Descriptions

Three textural themes were relevant to P14's lived experience deploying ML models in production environments. These were themes 2, 3, and 5. Textural Theme 2 stated that the ML lifecycle is useful when deploying ML models, and this theme recurred multiple times throughout the interview. P14 noted the usefulness of determining the projects' requirements characterizing discovery of the requirements as a "very crucial step of the whole machine learning process. P14 further noted that determining such requirements allowed developers to "stay focused on what question you're trying to answer."

P14 characterized data management as "super useful." P14 noted that there were many dimensions to successful data management, stating, "There's a lot of considerations to take when it comes to being able to manage all this data" and that these also included technical challenges related to the system's reliability beyond data storage challenges. P14 also noted that all steps of the ML lifecycle had value. In terms of benchmarking metrics, P14 stated that benchmarking was "super useful for deployment purposes." However, P14 also noted that "benchmarking metrics is actually getting the benchmark metrics themselves right" and that getting the metrics right was challenging in many instances. Finally, P14 opined that "it would be difficult for a team that probably perhaps isn't as experienced in machine learning" to perform benchmarking successfully.

P14 also highlighted the importance of the last two steps in the ML lifecycle. P14 characterized user acceptance testing as "super useful" and "important to continue." In the context of user testing, P14 noted that the most significant challenge in a larger organization was "to get the user acceptance testing done." User acceptance testing "dynamically changes, and availability might not be there just because--not because that they don't want to do [it, but] because it's been pretty busy." P14 noted that other user acceptance testing challenges included staffing shortages and time constraints. The final step in the ML lifecycle, privacy policy, was also noted by P14 as being "pretty useful" as it ensured that the organization was "not collecting data that you shouldn't be."

Textural Theme 3 stated that expertise and access to adequate data are useful and make ML models easier to deploy. P14 touched on this, stating that "we're just not talking about a little bit data we're talking about a lot of data that comes in" and noted that "especially streaming data trying to get some idea about" its nature was challenging in terms of "actually being able to understand that data, the question is, what are we, what questions what our requirements right."

The final Textural Theme, 5, stated that well-maintained ML models are easier to retrain during and after deployment. P14 focused on the privacy dimension of ML model maintenance and retraining in deployment. P14 noted that ML models often process "information that shouldn't be there." P14 meant that datasets might include private data that should not be analyzed or that a model could recreate private data by cross-referencing multiple available datasets. P14 noted that privacy was critical but challenging technically when coordinating with lawyers "who might not necessarily be the most technical people."

Developing a structural description of P14's lived experiences focused primarily on the role each phase in the ML lifecycle played in creating value when viewing the lifecycle as a deployment tool. P14 noted that the requirement analysis phase was also linked with the theme of accessing adequate data, and the privacy policy phase was linked with the need to maintain and retrain models during and after deployment.

P15's Textural and Structural Descriptions

P15's lived experiences of ML deployment as reflected by the interview data were reflected by the first four textural themes. Regarding Textural Theme 1, P15 noted that the "majority of the clients that access to work on those models knew what we want," at project commencement. P15 further noted that "then the requirement changes during the project" resulted in the need to "timeline communication" with the clients to discuss adjustments to the model and realign client and project requirements.

P15, in support of Textural Theme 2, highlighted the importance of the ML lifecycle to project development, stating that "the requirement analysis stage is useful to a machine learning deployment in the production environments." P15 also noted that requirement analysis was important because it "allows the user to assess the effectiveness" of a proposed model. Furthermore, P15 emphasized that using the ML lifecycle impacted project quality, stating that "it forces the machine learning to ensure that the strong quality get along with the definite entry and exit criteria in the model are implemented."

Additionally, P15 noted that structured, cleaned, and validated data was critical to project success noting that if these data parameters were not established as feasible early in the process that it would make later ML lifecycle steps "challenging" and create problems "at the end at towards the deployments." P15 also noted the value of later steps of the ML lifecycle, specifically noting that "data management is the most important part in the deployment" of ML models and characterizing benchmarking metrics as one "of the biggest challenges" and a "really convincing stage." P15 also noted the importance of the two final steps in the ML lifecycle. P15 noted that user acceptance testing was "a very useful state in model deployments" and also noted that developing an appropriate privacy policy," the final step in the ML lifecycle, was "really, really important in model deployments."

P15's lived experiences also reflected Textural Theme 3, which stated that expertise and access to adequate data are useful and make ML models easier to deploy. P15 emphasized the importance of good data during the requirements analysis phase and the importance of data management. P15 expanded on this in the context of data management expertise, stating, "the biggest data management challenges is understanding the data like trying to determine the usefulness of the data to the model and outreach and meet the needs of the of the problem statements."

Textural Theme 4, which stated that reconciling ML models with client expectations is useful before and during deployment, was the final theme reflected in P15's interview materials. P15 only addressed the need to reconcile client expectations during deployment once in the interview materials in the context of user acceptance testing. P15 noted that user acceptance testing occurs when the user tests the model to see if it meets requirements. P15 noted that "the client understands the need for the model," but if user requirements were previously "not correctly determined" and the model, as a result, was not "able to address the real-world situation," when the client performed user acceptance testing that the model "might be returned." After return, the model would undergo adjustment to reconcile it with actual client expectations.

In addition to exploring the textural themes influencing P15's lived experiences, it was essential to develop a structural description of the intersection of themes. The four textural themes that represented P15's lived experience corresponded with specific universal structures of the ML lifecycle. The dominant textural theme in P15's lived experience was Textural Theme 2. The ML lifecycle is useful when deploying ML models was the only theme that intersected all the universal structural themes. In the requirement analysis phase of the ML lifecycle, the two textural themes that were relevant were the textural themes of communication and the value of the ML lifecycle. By using TAM constructs, Themes 1, 3, and 5 correspond to the TAM construct of perceived ease of use (PEOU) whereas Themes 2, 3, and 4 correspond to the TAM construct of perceived usefulness (PU).

The textural themes most relevant to the data management phase were the value of the ML lifecycle and Textural Theme 3, which stated that expertise and access to adequate data are useful and make ML models easier to deploy. The benchmarking metrics phase of the ML lifecycle was supported by Textural Theme 2. User acceptance testing was supported by two relevant themes, Textural Theme 2 and Textural Theme 4. For privacy policy, the sole relevant textural theme was Textural Theme 2.

Summary

This section presented the results of the data analysis and described the essence of the lived experiences of the study's participants. This qualitative phenomenological study was guided by a research question: What challenges do organizations face when deploying ML models in production environments? The data analysis process identified several textural and structural themes that described the essence of the phenomenon. Table 1 summarizes the textural and structural themes that emerged from the interview data and the literature review.

The composite textural-structural description indicated that the main challenges organizations face when deploying ML models in production environments involve the need for continuous communication and model adjustment, the need for engineer expertise and access to adequate data, the ability to reconcile client expectations with model performance, and the need to continuously maintain and retrain models during and after model deployment.

TABLE 1
SUMMARY OF THE TEXTURAL AND STRUCTURAL THEMES

Theme Type	**Theme Description**
Textural Themes	
Theme 1	Communication and model adjustment make ML models easier to deploy.
Theme 2	The ML lifecycle is useful when deploying ML models.
Theme 3	Expertise and access to adequate data are useful and make ML models easier to deploy.
Theme 4	Reconciling ML models with client expectations is useful before and during deployment.
Theme 5	Well-maintained ML models are easier to retrain during and after deployment.
Structural Themes	
Requirement Analysis	A stage in the ML lifecycle where engineers identify all necessary project elements (Cho & Lee, 2020).
Data Management	A stage in the ML lifecycle where engineers prepare the data required to build a successful model (Paleyes et al., 2021).
Benchmarking Metrics	A stage in the ML lifecycle where engineers use tools to measure the performance of the ML system being deployed (Dai & Berleant, 2019).
User Acceptance Testing	A stage in the ML lifecycle where users test a system to determine whether it is functioning according to their expectations (Schelter et al., 2018).
Privacy Policy	A stage in the ML lifecycle where policies are established to protect system data (Zhao et al., 2018).

Summary, Discussion, and Implications

Machine Learning (ML) is a field heavily dependent on data. The purpose of this qualitative phenomenological study was to investigate the challenges associated with ML model deployment in production environments to fill a gap in the body of knowledge and provide scholars and practitioners with a better understanding of the most critical elements of the ML lifecycle. A qualitative research method was appropriate because it enabled the researcher to explore the participants' subjective experiences and comprehend the ML model deployment phenomenon from their perspectives (Eshlaghy et al., 2011). The research method included (a) the collection of data via online scheduled interviews, (b) a review of transcribed and recorded data, (c) invariant constituent classification and category creation, (d) identification of concepts and themes to highlight patterns from the data using Dedoose, (e) and cross-validation of data to reduce misrepresentation and overlap (Eshlaghy et al., 2011).

The focus of phenomenology is to treat lived experiences as critical data used to understand human nature (Moustakas, 1994). The present study explored ML experts' lived experiences deploying ML models in production environments. The participants in the research study were ML experts located across the United States. The participants had lived experiences with ML model development and deployments in different domains, including healthcare, banking, and defense. Textural information from the 15 semi-structured interviews captured the participants' explained perceptions during various stages of ML deployment, while structural information drawn from the interviews provided insights into how the participants experienced the phenomenon of ML deployment (Moustakas, 1994).

Practical Assessment of Project Analysis

The central research question for this qualitative phenomenological study was: What challenges do organizations face when deploying ML models in production environments? That research question was answered by creating textural and structural descriptions of the participants' lived experiences deploying ML models. The textural descriptions addressed what the participants experienced, and the structural descriptions described foundational elements of the experience. As noted in the previous section, the data analysis for this phenomenological study resulted in five textural themes describing what participants experienced when working with the ML lifecycle to deploy ML models in production environments. Participants experienced those themes in the context of the universal structures of the stages of the ML lifecycle (i.e., requirements analysis, data management, benchmarking metrics, user acceptance testing, and privacy policy).

Theme 1: Communication, Model Adjustment, and Perceived Ease of Use

Participants indicated that communication and adjustment were essential to project success when implementing or deploying ML models. Some participants (e.g., P6 and P7) highlighted the importance of documenting client needs before and during an ML deployment project, while others (e.g., P1 and P8) suggested that communication and adjustment were most important because clients did not always understand ML capabilities. Regardless of the specific reasons communication was necessary, the participants overwhelmingly supported ongoing communication and model adjustment throughout the ML lifecycle. In answering the present study's research question, Theme 1 indicated that a lack of communication is a significant challenge to ML deployment success.

Theme 2: The Perceived Usefulness of the ML Lifecycle

Many participants cited the usefulness of the ML lifecycle as a planning tool. P3 described the ML lifecycle as a plan that defines the entire deployment process, and P9 echoed this sentiment by stating, "If there's no plan and there's no process, then I don't think you even set yourself up for an opportunity of success." Other participants (e.g., P12, 13, and 15) attributed the value of the ML lifecycle to its usefulness as a framework for documenting changes, fostering collaboration, and facilitating the deployment process. P15 stated, "The planning of the ML lifecycle will make the ML deployment process easier because the planning helps to itemize and determine what's required for each stage of the ML process."

Theme 3: Expertise and Access to Data, Perceived Usefulness, and Perceived Ease of Use

Participants overwhelmingly observed that engineer expertise and access to quality data were fundamental requirements of a successful ML model deployment. Participants' responses indicated that the lack of these factors posed significant challenges to deployment success. Several participants indicated that data access and availability were important (P1, P3, and P7). Other participants cautioned that well-designed models could not be effective without the right data (P2, P4, and P11). However, some participants noted that quality data would not produce effective ML models if engineers were inexperienced and made errors during the model design or adjustment (P10, P14, and P15).

Theme 4: The Perceived Usefulness of Reconciling Client Expectations

Participants' responses indicated that reconciling client expectations was a complex consideration when deploying ML models. Some participants noted that clients and ML engineers often have vastly different perspectives on ML (P1, P9), and other participants used the term *manage* when discussing the differences between expectations and reality (P2, P5, P13). Several participants noted that reconciling client expectations was necessary at different points throughout the ML lifecycle (P6, P7). However, the participants indicated that reconciling client expectations was necessary for ML model deployments to succeed.

Theme 5: Maintenance, Retraining, and Perceived Ease of Use

An important consideration among the participants was that model maintenance and retraining provide ML engineers with more control over the deployment process and greater opportunities for success. Some participants indicated that training was often time-consuming, but they argued that time was a worthwhile investment (P5, P6, P7). When discussing training and model maintenance, participants also noted the ability to get more value from data (P14), the ability to validate a model through training (P10), and the need for more standard training approaches (P8). However, all of the participants who mentioned training and maintenance expressed that the refinement process improved deployment success.

Implications for Future Study

The results of this phenomenological study have important implications for future research, and researchers can use the findings from this study to further the investigation and understanding of the ML lifecycle and ML model deployment in production environments. One way to identify a study's implications for future research is to examine the research design constraints and limitations. As previously noted, this study was limited to the qualitative analysis of narrative data from 15 ML engineers. Thus, the findings were not easily generalizable to the larger population of ML organizations and practitioners. Future researchers could address this limitation by conducting a quantitative study and surveying a larger population to determine statistically which elements of the ML lifecycle are most significant to ML model deployment success. Many stakeholders can benefit from generalizable ML deployment research (Wang et al., 2014).

While the present study included participants working in healthcare, banking, and defense, no efforts were made to compare or contrast the experiences of ML engineers in different domains. Future research could examine whether different aspects of the ML lifecycle were more important in specific domains. For example, the privacy policy stage of the ML lifecycle might be more critical in a field like healthcare compared to a field like manufacturing or retail sales. Cabitza et al. (2017) argued that ML use could have unintended consequences in fields like healthcare. Research should focus on how the ML lifecycle could be adjusted to make ML deployments more effective in different settings.

A final recommendation would be to use the information derived from this study to develop a more extensive ML lifecycle framework. Several researchers have proposed ML lifecycle frameworks (Melgar et al., 2021; Chen et al., 2020; John et al., 2021). However, no universally accepted conceptual model exists to guide deployment processes, despite scholars noting the benefits of conceptual models (Chen et al., 2020; Dai & Berleant, 2019; John et al., 2021). Information provided by the participants could be used to identify specific factors that are essential to each stage of the ML lifecycle. Future research could identify strategies and steps practitioners could use to facilitate each of the five stages of the ML lifecycle (i.e., requirements analysis, data management, benchmarking metrics, user acceptance testing, and privacy policy).

Summary

This phenomenological study investigated the challenges associated with ML model deployment in production environments. Semi-structured interview data were collected from 15 ML engineers working in the United States. The goal was to explore the lived experiences of ML experts responsible for deploying ML models in production environments. A single research question guided the study: What challenges do organizations face when deploying ML models in production environments?

The use of a phenomenological study resulted in identifying five textural themes and five universal structures relevant to ML model deployments. Together, the textural and structural themes defined participants' lived experiences and answered the study's research question. The five universal structures of the ML lifecycle identified as part of the research were requirements analysis, data management, benchmarking metrics, user acceptance testing, and privacy policy. The five textural themes were as follows:

1. Communication and model adjustment make ML models easier to deploy. (perceived ease of use)
2. The ML lifecycle is useful when deploying ML models. (perceived usefulness)
3. Expertise and access to adequate data are useful and make ML models easier to deploy. (perceived usefulness and ease of use)
4. Reconciling ML models with client expectations are useful before and during deployment. (perceived usefulness)
5. Well-maintained ML models are easier to retrain during and after deployment. (perceived ease of use)

Based on the participants' responses and an analysis of the literature, several challenges were identified, including poor communication, lack of a planning framework, inexperienced ML engineers, limited access to data, poor data quality, failure to manage client expectations, and failure to maintain and retrain ML models when necessary. The participants' lived experiences indicated that the ML lifecycle is an important tool for successful ML model deployment, and more research is needed to develop structures that increase deployment success rates.

REFERENCES

Agrawal, A., Gans, J., & Goldfarb, A. (2020). How to win with machine learning: And how to catch up if you're lagging behind. *Harvard Business Review*, *2020*(9/10), 126–133. Retrieved from http://eli.johogo.com/Class/WMU/How_to_win_with_machine_learning.pdf

Alshurideh, M., Al Kurdi, B., Salloum, S.A., Arpaci, I., & Al-Emran, M. (2020). Predicting the actual use of M-learning systems: A comparative approach using PLS-SEM and machine learning algorithms. *Interactive Learning Environments*. Advance online publication. https://doi.org/10.1080/10494820.2020.1826982

Amin, M.E.K., Nørgaard, L.S., Cavaco, A.M., Witry, M.J., Hillman, L., Cernasev, A., & Desselle, S.P. (2020). Establishing trustworthiness and authenticity in qualitative pharmacy research. *Research in Social and Administrative Pharmacy*, *16*(10), 1472–1482. https://doi.org/10.1016/j.sapharm.2020.02.005

Arpaci, I., Al-Emran, M., Al-Sharafi, M.A., & Shaalan, K. (2020). A novel approach for predicting the adoption of Smartwatches using machine learning algorithms. In M. Al-Emran, K. Shaalan, & A. Hassanien (Eds.), *Studies in systems, decision and control* (vol. 295, pp. 185–195). Springer. https://doi.org/10.1007/978-3-030-47411-9_10

Attaran, M., & Deb, P. (2018). Machine learning: The new 'big thing' for competitive advantage. *International Journal of Knowledge Engineering and Data Mining*, *5*(4), 277–305. https://doi.org/10.1504/ijkedm.2018.095523

Baier, L., Jöhren, F., & Seebacher, S. (2019). *Challenges in the deployment and operation of machine learning in practice*. Paper presented at the Twenty-Seventh European Conference on Information Systems (ECIS2019), Stockholm-Uppsala, Sweden. Retrieved from https://tinyurl.com/3d4smxph

Cabitza, F., Rasoini, R., & Gensini, G.F. (2017). Unintended consequences of machine learning in medicine. *Journal of the American Medical Association*, *318*(6), 517–518. https://doi.org/10.1001/jama.2017.7797

Cai, J., Zeng, H., Yong, H., Cao, Z., & Zhang, L. (2019). Toward real-world single image super-resolution: A new benchmark and a new model. In *2019 IEEE/CVF International Conference on Computer Vision (*pp. 3086–3095). IEEE. https://doi.org/10.1109/ICCV.2019.00318

Canhoto, A.I., & Clear, F. (2020). Artificial intelligence and machine learning as business tools: A framework for diagnosing value destruction potential. *Business Horizons*, *63*(2), 183–193. https://doi.org/10.1016/j.bushor.2019.11.003

Char, D.S., Abràmoff, M.D., & Fuedtner, C. (2020). Identifying ethical considerations for machine learning healthcare applications. *The American Journal of Bioethics*, *20*(11), 7–17. https://doi.org/10.1080/15265161.2020.1819469

Chen, A., Chow, A., Davidson, A., DCunha, A., Ghodsi, A., Hong, S. A., . . . Zumar, C. (2020). Developments in MLflow: A system to accelerate the machine learning lifecycle. *Proceedings of the Fourth International Workshop on Data Management for End-to-End Machine Learning*, (6), Article 5. https://doi.org/10.1145/3399579.3399867

Cho, B.-S., & Lee, S-W. (2020). A comparative study on requirements analysis techniques using natural language processing and machine learning. *Journal of the Korean Society for Computer Information and Information Science*, *25*(7), 27–37. https://doi.org/10.9708/JKSCI.2020.25.07.027

Creswell, J.W., & Creswell, J.D. (2018). *Research design: Qualitative, quantitative, and mixed methods approaches*. Sage.

Creswell, J.W., & Poth, C.N. (2016). *Qualitative inquiry and research design: Choosing among five approaches*. Sage.

Dai, W., & Berleant, D. (2019). Benchmarking contemporary deep learning hardware and frameworks: A survey of qualitative metrics. *2019 IEEE First International Conference on Cognitive Machine Intelligence*, pp. 148–155. https://doi.org/10.1109/cogmi48466.2019.00029

Dalton, J. (2019). *Great big Agile: An OS for Agile learners*. Apress. https://doi.org/10.1007/978-1-4842-4206-3

Dankwa-Mullan, I., Rivo, M., Sepulveda, M., Park, Y., Snowdon, J., & Rhee, K. (2019). Transforming diabetes care through artificial intelligence: The future is here. *Population Health Management*, *22*(3), 229–242. https://doi.org/10.1089/pop.2018.0129

De Laat, P.B. (2017). Algorithmic decision-making based on machine learning from big data: Can transparency restore accountability? *Philosophy & Technology*, *31*(4), 525–541. https://doi.org/10.1007/s13347-017-0293-z

Eshlaghy, A.T., Pourebrahimi, A., & Nobari, B.Z. (2011). Presenting a model for ranking organizations based on the level of the information security maturity. *Computer and Information Science*, *4*(1), 72–78. https://doi.org/10.5539/cis.v4n1p72

Fawcett, J., & Garity, J. (2008). *Evaluating research for evidence-based nursing practice*. F.A. Davis.

Galli, S. (2021, March 19). *How to test and monitor machine learning model deployments*. Medium. Retrieved from https://trainindata.medium.com/how-to-test-and-monitor-machine-learning-model-deployments-952382f4884a

Gamkrelidze, T., Zouinar, M., & Barcellini, F. (2021). Artificial intelligence (AI) in the workplace: A study of stakeholders' views on benefits, issues and challenges of AI systems. In *Proceedings of the 21st Congress of the International Ergonomics Association* (pp. 628–635). Springer. https://doi.org/10.1007/978-3-030-74614-8_78

Ganesh, K., Mohapatra, S., Anbuudayasankar, S.P., & Sivakumar, P. (2014). User acceptance test. In *Enterprise resource planning management for professionals* (pp. 123–127). Springer. https://doi.org/10.1007/978-3-319-05927-3_9

Garcia, P., Darroch, F., West, L., & BrooksCleator, L. (2020). Ethical applications of big data-driven AI on social systems: Literature analysis and example deployment use case. *Information*, *11*(5), Article 235. https://doi.org/10.3390/info11050235

Ghahramani, Z. (2015). Probabilistic machine learning and artificial intelligence. *Nature*, *521*(7553), 452–459. https://doi.org/10.1038/nature14541

Gu, Z., Wang, J., & Luo, S. (2020). Investigation on the quality assurance procedure and evaluation methodology of machine learning building energy model systems. *2020 International Conference on Urban Engineering and Management Science*, pp. 96–99. IEEE. https://doi.org/10.1109/icuems50872.2020.00031

Guezzaz, A., Asimi, Y., Azrour, M., & Asimi, A. (2021). Mathematical validation of proposed machine learning classifier for heterogeneous traffic and anomaly detection. *Big Data Mining and Analytics*, *4*(1), 18–24. https://doi.org/10.26599/bdma.2020.9020019
Janmey, V., & Elkin, P.L. (2018). Re-identification risk in HIPPA de-identified datasets: The MVA attack. In *AMIA Annual Symposium proceedings* (vol. 2018, pp. 1329–1337). Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6371259
John, M.M., Holmström Olsson, H., & Bosch, J. (2021). Architecting AI deployment: A systematic review of state-of-the-art and state-of-practice literature. In E. Klotins, & K. Wnuk (Eds.), *Software business: ICSOB 2020: Lecture notes in business information processing* (vol. 407, pp. 14–29). Springer. https://doi.org/10.1007/978-3-030-67292-8_2
Kamble, S.S., Gunasekaran, A., Kumar, V., Belhadi, A., & Foropon, C. (2021). A machine learning based approach for predicting blockchain adoption in supply chain. *Technological Forecasting and Social Change*, *163*(2), Article 120465. https://doi.org/10.1016/j.techfore.2020.120465
Khalfoun, B., Ben Mokhtar, S., Bouchenak, S., & Nitu, V. (2021). EDEN: Enforcing location privacy through re-identification risk assessment: A federated learning approach. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, *5*(2), Article 68. https://doi.org/10.1145/3463502
Krizhevsky, A., Sutskever, I., & Hinton, G.E. (2017). ImageNet classification with deep convolutional neural networks. *Communications of the ACM*, *60*(6), 84–90. https://doi.org/10.1145/3065386
Kumar, V.S., Cuddihy, P., & Aggour, K.S. (2019). NodeGroup: A knowledge-driven data management abstraction for industrial machine learning. *Proceedings of the 3rd International Workshop on Data Management for End-to-End Machine Learning*, Article 10. https://doi.org/10.1145/3329486.3329497
Leedy, P.D., & Ormrod, J.E. (2020). *Practical research: Planning and design* (12th Ed.). Pearson.
Lincoln, Y.S., & Guba, E.G. (1985). *Naturalistic inquiry*. Sage.
Liu, S.Y. (2020). Artificial intelligence (AI) in agriculture. *IT Professional*, *22*(3), 14–15. https://doi.org/10.1109/mitp.2020.2986121
Lo Piano, S. (2020). Ethical principles in machine learning and artificial intelligence: Cases from the field and possible ways forward. *Humanities and Social Sciences Communications*, *7*, Article 9. https://doi.org/10.1057/s41599-020-0501-9
Luo, X., Zhang, Y., He, Z., Yang, G., & Ji, Z. (2019). A two-step environment-learning-based method for optimal UAV deployment. *IEEE Access*, *7*, 149328–149340. https://doi.org/10.1109/access.2019.2947546
Mahesh, B. (2018). Machine learning algorithms: A review. *International Journal of Science and Research*, *9*(1), 381–386. https://doi.org/10.21275/ART20203995
Medeiros, M.C., Vasconcelos, G.F., Veiga, Á., & Zilberman, E. (2019). Forecasting inflation in a data-rich environment: The benefits of machine learning methods. *Journal of Business & Economic Statistics*, *39*(1), 98–119. https://doi.org/10.1080/07350015.2019.1637745
Moustakas, C. (1994). *Phenomenological research methods*. Sage.
Mueller, J.P., & Massaron, L. (2021). *Machine learning for dummies*. John Wiley & Sons.
Nurvitadhi, E., Venkatesh, G., Sim, J., Marr, D., Huang, R., Ong Gee Hock, J., . . . Boudoukh, G. (2017). Can FPGAs beat GPUs in accelerating next-generation deep neural networks? *Proceedings of the 2017 ACM/SIGDA International Symposium on Field-Programmable Gate Arrays*, (2), 5–14. https://doi.org/10.1145/3020078.3021740
Oakden-Rayner, L., Dunnmon, J., Carneiro, G., & Re, C. (2020). Hiddent stratification causes clinically meaningful failures in machine learning for medical imaging. In *CHIL '20: Proceedings of the ACM Conference on Health, Inference, and Learning* (pp. 151–159). Association for Computing Machinery. https://doi.org/10.1145/3368555.3384468
Olowononi, F.O., Rawat, D.B., & Liu, C. (2021). Resilient machine learning for networked cyber physical systems: A survey for machine learning security to securing machine learning for CPS.

IEEE Communications Surveys & Tutorials, *23*(1), 524–552. https://doi.org/10.1109/comst.2020.3036778
Osborne, J.W. (1990). Some basic existential-phenomenological research methodology for counsellors. *Canadian Journal of Counselling*, *24*(2), 79–91. Retrieved from https://cjc-rcc.ucalgary.ca/article/view/59497
Ovtcharov, K., Ruwase, O., Kim, J., Fowers, J., Strauss, K., & Chung, E.S. (2015). Toward accelerating deep learning at scale using specialized hardware in the datacenter. In *2015 IEEE Hot Chips 27 Symposium* (pp. 1–38). IEEE. https://doi.org/10.1109/HOTCHIPS.2015.7477459
Paleyes, A., Cabrera, C., & Lawrence, N.D. (2021). *Exploring the potential of flow-based programming for machine learning deployment in comparison with service-oriented architectures*. Cornell University arXiv. https://arxiv.org/abs/2108.04105
Pandit, P., & Tahiliani, S. (2015). AgileUAT: A framework for user acceptance testing based on user stories and acceptance criteria. *International Journal of Computer Applications*, *120*(10), 16–21. https://doi.org/10.5120/21262-3533
Pohle, J. (2018). Data privacy legislation in the European Union member states: A Practical overview. *Computer Law Review International*, *19*(4), 97–116. https://doi.org/10.9785/cri-2018-190402
Polyzotis, N., Roy, S., Whang, S.E., & Zinkevich, M. (2017). Data management challenges in production machine learning. *Proceedings of the 2017 ACM International Conference on Management of Data*, *(5), 1723*–1726. https://doi.org/10.1145/3035918.3054782
Singh, S., & Srivastava, P. (2019). Social media for outbound leisure travel: A framework based on technology acceptance model (TAM). *Journal of Tourism Futures*, *5*(1), 43–61. https://doi.org/10.1108/jtf-10-2018-0058
Singh, P. (2021). *Deploy machine learning models to production: With Flask, Streamlit, Docker, and Kubernetes on Google Cloud Platform*. Apress. https://doi.org/10.1007/978-1-4842-6546-8
Soares, M.S., & Do Nascimento, R.P. (2014). Evaluation of SysML diagrams to document requirements using TAM. *Proceedings of the 7th Euro American Conference on Telematics and Information Systems*, (4), Article 9. https://doi.org/10.1145/2590651.2590661
Sputh, O., Frid, J., & Hellander, A. (2021). The machine learning life cycle and the cloud: Implications for drug discovery. *Expert Opinion on Drug Discovery*, *16*(9), 1071–1079. https://doi.org/10.1080/17460441.2021.1932812
Stahl, N.A., & King, J.R. (2020). Expanding approaches for research: Understanding and using trustworthiness in qualitative research. *Journal of Developmental Education*, *44*(1), 26–29. Retrieved from https://tinyurl.com/hwettnfn
Tripathi, G., Ahad, M.A., & Paiva, S. (2020). S2HS: A blockchain based approach for smart healthcare system. *Healthcare*, *8*(1), Article 100391. https://doi.org/10.1016/j.hjdsi.2019.100391
van der Goes, M. (2021). Scaling enterprise recommender systems for decentralization. In *RecSys '21: Fifteenth ACM Conference on Recommender Systems* (pp. 592–594). Association for Computing Machinery.
Vincent-Lancrin, S., & van der Vlies, R. (2020). Trustworthy artificial intelligence (AI) in education. *OECD Education Working Papers*. https://doi.org/10.1787/a6c90fa9-en
Vogelsang, A., & Borg, M. (2019). Requirements engineering for machine learning: Perspectives from data scientists. *2019 IEEE 27th International Requirements Engineering Conference Workshops*, pp. 245–251. https://doi.org/10.1109/rew.2019.00050
von Thenen, N., Ayday, E., & Cicek, A.E. (2018). Re-identification of individuals in genomic data-sharing beacons via allele inference. *Bioinformatics*, *35*(3), 365–371. https://doi.org/10.1093/bioinformatics/bty643
Wang, L., Zhan, J., Luo, C., Zhu, Y., Yang, Q., He, Y., Qiu, B. (2014). BigDataBench: A big data benchmark suite from internet services. *2014 IEEE 20th International Symposium on High Performance Computer Architecture*, pp. 488–499. https://doi.org/10.1109/hpca.2014.6835958

Wuest, T., Weimer, D., Irgens, C., & Thoben, K. (2016). Machine learning in manufacturing: Advantages, challenges, and applications. *Production & Manufacturing Research*, *4*(1), 23–45. https://doi.org/10.1080/21693277.2016.1192517

Xia, W., Liu, Y., Wan, Z., Vorobeychik, Y., Kantacioglu, M., Nyemba, S., Clayton, E.W., & Malin, B.A. (2021). Enabling realistic health data re-identification risk assessment through adversarial modeling. *Journal of the American Medical Informatics Association*, *28*(4), 744–752. https://doi.org/10.1093/jamia/ocaa327

Xie, Y., Cruz, L., Heck, P., & Rellermeyer, J.S. (2021). Systematic mapping study on the machine learning lifecycle. *IEEE/ACM 1st Workshop on AI Engineering – Software Engineering for AI (WAIN)*, pp. 70–73. https://doi.org/10.1109/wain52551.2021.00017

Zhao, J., Mortier, R., Crowcroft, J., & Wang, L. (2018). Privacy-preserving machine learning based data analytics on edge devices. *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, (12), 341–346. https://doi.org/10.1145/3278721.3278778

Zhou, Y., Yu, Y., & Ding, B. (2020). Towards MLOps: A case study of ML Pipeline platform. In *2020 International Conference on Artificial Intelligence and Computer Engineering* (pp. 494–500). IEEE. https://doi.org/10.1109/ICAICE51518.2020.00102