PART I

AI Technology Stack: Evolution of AI Systems

ML vs. LLMs, Transforming Healthcare Through AI

By:
Assaf Mischari, Managing Partner, Team8
Eyal Eliakim, Head of AI, Team8

TEAM8™

In the unfolding narrative of technological advancement, artificial intelligence represents more than a mere shift in computational paradigms—it embodies our evolving understanding of how to scale human intelligence while preserving the nuanced contexts of human interaction. This white paper explores two fundamental approaches to AI solution development: the traditional, domain-specific ML pipeline, commonly utilized up to the materialization of LLMs, and the emerging paradigm built around large, pre-trained foundation models. As we speed forward on this technological fast track, our challenge extends beyond implementation to fundamentally reimagining how AI can amplify and enhance human capability.

Key Insights You'll Gain

1. Why the FDA draws a clear line between traditional machine learning and generative AI in healthcare—and what that means for future approvals.

2. How classic ML pipelines and modern LLM-based systems differ across the entire AI lifecycle, from data gathering to deployment and monitoring.

3. What makes LLMs uniquely powerful yet difficult to regulate in clinical settings, especially given their non-deterministic nature.

4. The technical and operational trade-offs between narrow, purpose-built models and general-purpose foundation models in healthcare.

5. Stay tuned - A strategic framework for evaluating, adapting, and safely integrating LLMs into high-stakes healthcare environments.

What are the implications on healthcare?

The FDA's position leads us with a clue. In late 2024, the FDA convened its Digital Health Advisory Committee to specifically discuss generative AI in healthcare, highlighting the agency's recognition that LLM-based medical applications may require new oversight considerations An FDA executive summary prepared for that meeting noted that the rapid rise of generative AI "may present challenges to FDA's laws and regulations" and that new regulatory approaches could be needed to ensure safety and effectiveness.

To date, the FDA has approved a few, if any, medical devices that use generative AI or LLM technology.
All AI/ML-enabled devices authorized so far have used more traditional forms of machine learning (e.g. computer vision or rule-based algorithms) rather than large language models. In practice, most FDA-cleared AI software has "locked" algorithms that do not change once approved, whereas LLMs are more adaptive and can generate new free-text outputs, a capability not yet seen in any cleared device.

As of early 2025, the FDA has authorized over

1,000 ML-powered medical devices

vs.

0 LLM-powered devices

A vast majority of the approved AI devices are in diagnostic imaging fields – for example, about 77% of AI medical devices are in radiology as of late 2023 – and they typically perform pattern recognition or classification tasks on medical images. These "traditional" AI devices have demonstrated clinical performance in narrow domains (such as flagging abnormalities on X-rays or ECGs). LLM-based medical AI, on the other hand, would involve interpreting or generating human language (e.g. summarizing patient records or engaging in a diagnostic dialogue). Such applications are still in experimental or pilot stages and have not yet reached FDA authorization.

AI Pipeline Evolution in healthcare
A Technical Overview

1 Classic ML Pipelines: The Foundation of Systematic Intelligence

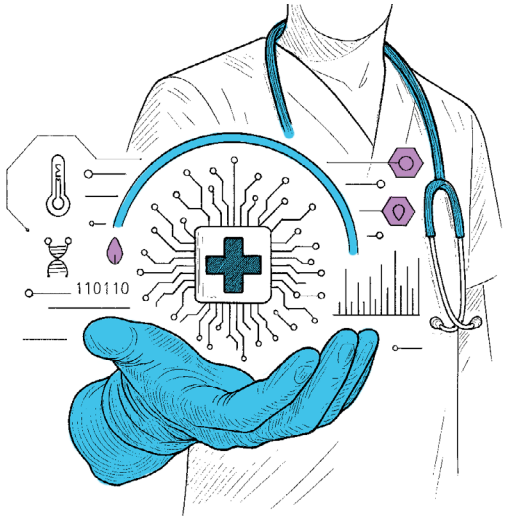
In this article, the term "classic ML pipelines" refers to AI systems as they were commonly built before the emergence of large foundation models. Such pipelines utilized extensive data gathering and feature engineering, as well as small proprietary models designed to tackle specific tasks and provide deterministic outputs.

Data Gathering

At the foundation of classic ML pipelines lies a comprehensive data governance framework. **This approach emphasizes strategic data collection that aligns precisely with specific use cases and intended outcomes.** Organizations implementing these pipelines must establish enterprise-grade data infrastructure, typically leveraging modern cloud architectures, while maintaining robust security and compliance frameworks throughout their data ecosystem. **Importantly, sufficient data collection is considered an entry barrier when looking to develop classic ML systems.**

Data Engineering Excellence

The transformation of raw data into meaningful insights requires sophisticated data engineering processes. Advanced ETL workflows ensure data integrity across the pipeline, while feature engineering processes are carefully guided by domain expertise to extract maximum value from available data. Such feature engineering is normally a key facilitator for models to perform their designated tasks in classic ML pipelines. **These processes are orchestrated through enterprise-grade platforms that maintain completeness, consistency and reliability across all data transformations.**



Model Development

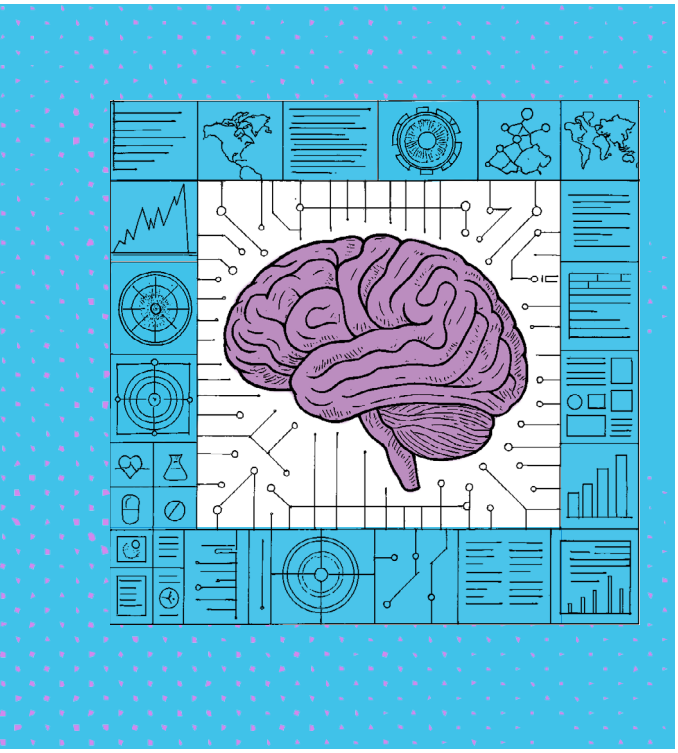
When developing classic ML systems, teams create purpose-built models designed for specific scenarios, ensuring tight integration with established decision support frameworks. **Each model is developed from scratch** and undergoes rigorous validation against defined metrics, creating a reliable foundation for automated decision-making processes. **Such models are typically small** (compared to foundation models, which will be discussed later in this article) and **provide a deterministic output** (i.e., an image classifier that categorizes x-rays as pneumonia cases or non-pneumonia cases).

Deployment Architecture

System integration focuses on seamlessly embedding ML capabilities within existing enterprise systems. **The architecture prioritizes real-time decision support delivery while minimizing workflow disruption.** This careful balance ensures that AI capabilities enhance rather than complicate existing business processes.

Operations & Monitoring

Quality assurance in classic ML pipelines **requires continuous monitoring of accuracy and outcomes. The framework includes comprehensive audit trails that track system inputs and outputs, as well as system security and data privacy.** In some cases, organizations implement solutions to increase the interpretability of AI-driven recommendations, ensuring stakeholders can understand and trust the system's decisions.



2 Modern AI Pipelines: Amplifying Human Intelligence

Foundation Models

Modern AI pipelines center around foundation models (e.g., LLMs, image/video generators, etc.). **These models are usually trained by frontier labs on vast, non-domain-specific data and can generate non-deterministic outputs. The models demonstrate comprehensive understanding of contextual nuances and facilitate enhanced human-machine interaction.** Their flexibility allows for broad application across various use cases while maintaining high performance standards.

Model Selection Framework

Foundation models represent a new class of AI systems trained on vast, diverse datasets that enable broad applicability across multiple domains. These models - including language processors like OpenAI's GPT and Anthropic's Claude, image generators like DALL-E, and multimodal systems like CLIP - demonstrate remarkable versatility with minimal task-specific adaptation.

Developing modern AI systems normally **requires selecting a foundation model that is aligned with organizational workflows and objectives.** This process includes a rigorous evaluation of model performance, bias, and fairness to ensure the system performs its designated task at a sufficient level and that it's ethically deployed. In domains like healthcare, **teams must verify compliance with regulatory requirements, creating a framework that balances innovation with responsibility.**

Model Adaptation for Specific Use Cases

Foundation models are usually designed to perform general tasks well. Adaptation is required for these models to excel at domain-specific tasks normally carried out by highly skilled professionals. **Organizations implement fine-tuning processes using high-quality domain data to specialize these models for specific applications. Additionally, Teams develop sophisticated prompting strategies while integrating domain expertise to enhance model performance in targeted scenarios.**

Operations & Monitoring

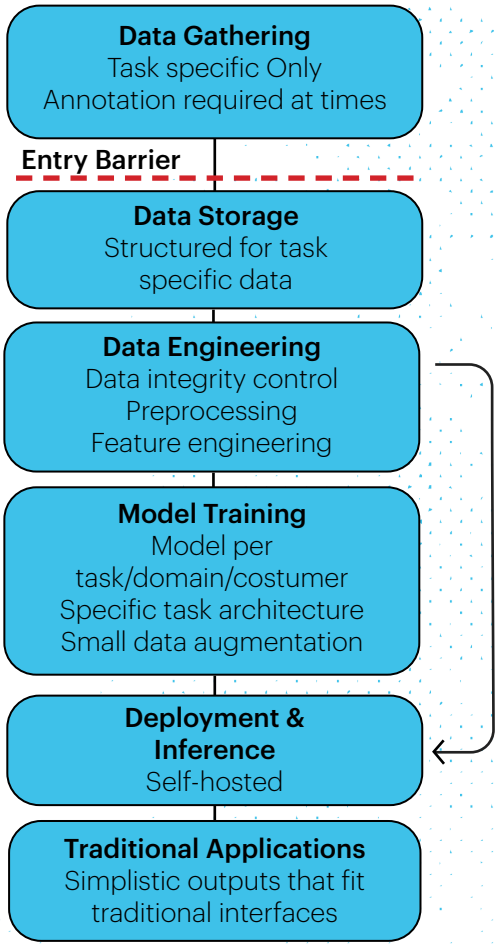
Operational excellence in modern AI pipelines demands continuous monitoring of model appropriateness and performance. **Organizations implement privacy-preserving architectures to protect sensitive data while enabling broad model capabilities.** The framework emphasizes cost-effective scaling of capabilities, ensuring sustainable deployment of AI systems. Notably, the non-deterministic output of modern AI systems introduces new challenges around monitoring model output reliability.

Application Layer

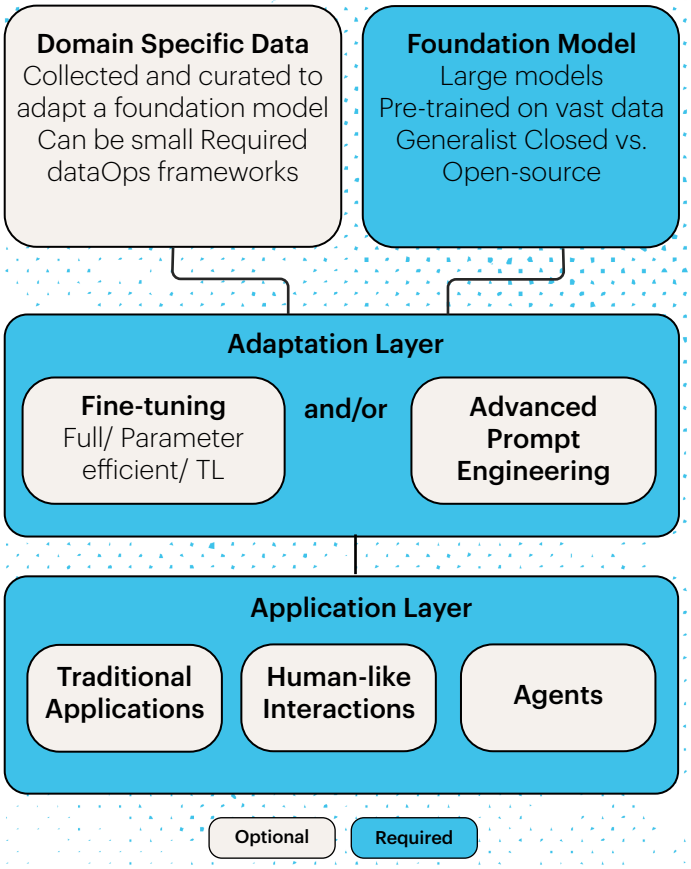
The application layer focuses on enhanced interaction between users and AI systems. **Modern pipelines implement multi-modal natural language interfaces that facilitate intuitive communication with AI systems.** These interfaces transform applications as we knew them and support AI-augmented decision support. They enable personalized engagement solutions that adapt to individual user needs and preferences.

3 Comparative Analysis: Implementation Considerations

Classic ML Pipeline



Modern AI Pipeline



Classic ML considerations

- Models evaluated against task organization specific test sets.
- Security & privacy maintained as long as the organization's network and apps are kept secure.
- Data & MLOps solutions required to achieve observability, maintain data integrity and monitor model performance.
- Explainability required on a per use-cases basis and achieved with algorithm/data based methods.

Modern AI considerations

- Models evaluated against benchmark datasets + task/organization specific test sets.
- Complex security & privacy challenges due to reliance on third-party foundation models Data integrity maintainable only on parts of the training data.
- Explainability achievable with algorithm based methods alongside prompt engineering techniques.
- Observability becomes key due to foundation model related challenges around latency.
- Cost and resource utilization become key to monitor and manage due to extensive hardware requirements.

Aspect	Classic Pipeline	Modern Pipeline
Data gathering	Entry barrier. Domain-specific. Requires extensive resources.	Foundation models rely on vast training data already available to them. Small domain-specific data gathering is required.
Data engineering	Heavy data integrity, preprocessing and feature engineering.	Handle raw input efficiently. Minimal controls are required on small domain-specific datasets.
Model training	Task-specific. Trained from scratch. Small models per domain/ customer.	Foundation models are pre-trained. Fine-tuning to a specific task might be required.
Deployment	Normally self-hosted with low latency.	Reliant, at times, on third-party APIs. Requires optimized infrastructure for serving resource-intensive models.
Ops & monitoring	Evaluated against organization-specific test sets. Security & privacy maintained as long as the org's infra is kept secure. Ops solutions required for observability, data integrity, and to monitor performance. Explainability is required on a per-use-case basis.	Evaluated against benchmark datasets + organization-specific test sets. Complex security & privacy challenges. Data integrity maintainable in parts of the training data. Explainability continues to be a challenge. Observability becomes key due to latency and cost challenges.
Application layer	Simplified model outputs that fit traditional interfaces.	Open avenues for improved user experiences through human-like interactions.

4 Navigating the Next Chapter: A Strategic Approach to LLM Adoption

As we navigate this transition, the conversation shifts from merely adopting AI to strategically integrating it into enterprise ecosystems in a way that enhances decision-making, preserves human expertise, and mitigates risks. **The rapid advancement of large language models (LLMs) introduces unprecedented opportunities—but also requires a deliberate framework for evaluation, implementation, and oversight.**

In the next chapter, we will dive into our perspective on LLMs, exploring:

- **The Team8 Health Compass for LLMs** – A structured approach to assessing AI's role in healthcare and other high-stakes domains.
- **Critical Questions to Ask** – A practical guide for decision-makers to uncover hidden risks and maximize AI's potential.
- **Avoiding Pitfalls** – Common mistakes organizations make when vetting LLMs—and how to proactively address them.

We stand on the cusp of an AI-driven future, where human intelligence and machine intelligence work in tandem to create new possibilities, new efficiencies, and new breakthroughs. But this future will only be realized if we take a deliberate and informed approach. **The next chapter will outline how to navigate this journey with confidence—leveraging AI's strengths while understanding where it can provide the highest value while limiting risks.**