

friendly

# **Building an Automated ML Pipeline — Introduction and Ethical Foundations**

AI Masters Capstone Project

---

Jonathan Agustin

November 2024

## A New Paradigm: GenAI as the Next Internet

- GenAI echoes the early internet's journey from “bubble” to massive value
- Emphasize trust, distribution, and integration over mere technical sophistication
- Acknowledge ethics and fairness as central pillars of sustainable AI success

*We stand at a turning point where human-centric values matter more than ever.*

## Lessons from the Past Applied to GenAI

- Move beyond internal experiments; user-facing solutions drive lasting impact
- Non-tech firms can outshine tech giants by leveraging trust & distribution
- We'll combine ethics, fairness, and scale into a sustainable AI pipeline approach

*A new era demands human-centric strategies as much as technical prowess.*

## Series Overview

- Presentation 1: Automated ML Pipeline, Ethical Foundations
- Presentation 2: Data Preparation & Ethical Data Handling
- Presentation 3: Automating Model Training & Fairness
- Presentation 4: Deployment Automation & Ethical Deployment
- Presentation 5: Code Quality, Security & Ethics

*A complete journey from raw data to ethical, user-focused deployments.*

# Who is better positioned to lead in AI?

## IBM:

- Decades of AI leadership
- Vast R&D resources
- Global infrastructure
- Thousands of top-tier researchers

## JPMorgan Chase:

- Traditional, legacy banking systems
- Limited internal AI expertise
- Highly regulated environment
- No significant AI products yet

*The answer may surprise you—let's find out why.*

## Paradigm Shift in Leadership

- JPMorgan's massive, trusting user base outweighs raw R&D might
- AI models are accessible; trust and user reach are the real differentiators
- Real impact: Getting tech into users' hands, leveraging existing relationships

*We must design pipelines that prioritize trust, integration, and real-world adoption.*

## The ML Production Challenge (1/2)

- Complex preprocessing steps need standardization, not ad-hoc scripts
- Model versioning & tracking ensure accountability, reproducibility, trust

*Bridging research and production demands careful planning, not heroics.*

## The ML Production Challenge (2/2)

- Security & compliance: must-haves, not afterthoughts
- Regulated environments magnify complexity & oversight

*We need robust pipelines—“roads”—to let our “race car” models truly shine.*



## Toward a Comprehensive, Ethical Pipeline (1/2)

- Embrace automation: reduce manual effort across preprocessing, training, deployment
- Aim for stable, reproducible processes with minimal human intervention

*Speed is good, but doing it ethically and responsibly is transformative.*

## Toward a Comprehensive, Ethical Pipeline (2/2)

- Integrate bias checks early to prevent injustice downstream
- Include fairness metrics, not just accuracy or speed
- Secure, ethical deployment and continuous fairness monitoring

*We're creating ML pipelines that are not only efficient, but also principled and human-focused.*

## Key Technologies

- **GitHub Actions:** CI/CD to streamline development pipelines
- **Docker:** Consistent environments from dev to production
- **Terraform:** Infrastructure as code on AWS for reliable scaling
- **AWS:** Stable, secure cloud backbone
- **Hugging Face Spaces:** Effortless model deployment & user interfaces

*These tools represent fundamental pillars for robust, user-centric, and ethical ML deployments.*

## Ethical Foundations (1/2)

- Automation can propagate hidden biases into every stage
- Ethical AI requires deliberate fairness metrics & transparent processes
- Real-world scandals show the cost of neglecting these principles

*Ethical considerations must be embedded in every decision, every line of code.*

## Ethical Foundations (2/2)

- Introduce fairness metrics & bias checks at every stage
- Maintain transparency and accountability: audit trails, documentation, ethical reviews
- Strive for outcomes that respect all users equally

*Responsible AI isn't a luxury—it's a fundamental expectation in the modern world.*

## Pipeline Preview (1/2)

- Preprocessing modules: Validate data quality, detect bias, ensure consistency
- Training pipeline: Monitor performance metrics and fairness side-by-side

*A seamless, principled approach—from data ingestion to model refinement.*

## Pipeline Preview (2/2)

- Deployment: Track versions, maintain audit trails, ensure compliant releases
- Continuous monitoring: Keep models fair, accurate, and aligned with ethical standards

*A holistic blueprint merging efficiency, ethics, and user trust end-to-end.*

## Next Steps

- Next Presentation: Data Preparation & Ethical Handling
- Laying the groundwork for trustworthy, equitable ML solutions

*Onward to constructing fair and impactful AI, one principle at a time.*