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's trajectory mirrors the early internet's transformative growth.
- Success relies not only on model sophistication, but on trust and ethical integration.
- Fairness, accountability, and transparency enable sustainable and reputable AI adoption.

Lessons from the Past Applied to GenAI

- Move beyond limited internal prototypes—users need reliable, trusted ML solutions.
- Non-tech firms with established credibility can surpass tech giants by centering ethics.
- Ethical frameworks must guide scaling and adoption of AI tools.

Ethical AI Development is Not Optional

- Early-stage AI errors can scale harm exponentially.
- Developers bear ethical and legal responsibility for their code.
- **Case: United States v. Liang**: Engineer held liable despite following orders.

(Ref: DOJ Press Release)

Mobley v. Workday: AI Accountability in Hiring

- **Mobley v. Workday (2024)**: Demonstrates that AI vendors can be liable for bias in hiring.
- Automated decisions face the same legal scrutiny as human decisions.
- We must integrate fairness checks from data preparation to model output.

Personal Responsibility

- “Just following orders” is no defense—developers must question unethical directives.
- Fairness and transparency protect users, businesses, and professionals.
- Ethical vigilance is key to long-term stability and trust in AI solutions.

From Notebooks to Production

- Transition from prototype notebooks to production-grade pipelines is complex.
- Requires stable preprocessing, model versioning, compliance, and monitoring.
- Proper “infrastructure” ensures models reach real users effectively.

Addressing Complexity and Regulation

- Regulated fields (finance, healthcare) demand full auditability.
- Without robust infrastructure, even the best models remain idle.

Improving infrastructure ensures compliance with privacy laws and industry standards.

Building a Comprehensive, Ethical Pipeline

- Automation fosters consistent preprocessing, reproducible training, and smooth deployment.
- Integrate bias checks and fairness metrics throughout, not as afterthoughts.
- Ethical design upfront prevents costly retrofits and reputational damage later.

Key Technologies

- **GitHub Actions**: CI/CD ensuring continuous testing and integration.
- **Docker**: Consistent environments, reducing “works on my machine” issues.
- **Terraform**: Infrastructure as Code (IaC) for scalable AWS deployments.
- **Hugging Face Spaces**: Intuitive model hosting and demos.

FTC & IntelliVision: Compliance Matters

- IntelliVision claimed high accuracy and zero bias without evidence.
- FTC's action shows that bold claims require rigorous validation and documentation.
- Regulatory scrutiny is intensifying—substantiation is non-negotiable.

Operation AI Comply

- Launched by FTC (Sept 2024), targets deceptive AI claims.
- Enforcement actions stress that AI hype is no shield from liability.
- Misleading consumers about AI capabilities leads to penalties.

Key Enforcement Cases

DoNotPay: Claimed an “AI Lawyer” lacking real legal qualifications. Settlement included consumer warnings.

Rytr: Generated deceptive AI-based reviews, eroding trust and prompting enforcement.

Ascend Ecom, Ecommerce Empire Builders, FBA Machine: Promised easy earnings via AI but failed to deliver genuine value.

Lesson: Claims must be truthful, backed by evidence, and free from deceptive hype.

Implications for Developers and Businesses

- Transparency and honesty in AI performance claims are essential.
- Document training data, known limitations, and testing procedures.
- Compliance integrated into pipeline design reduces risk and builds long-term trust.

Our Series Roadmap

Presentation 1: Automated ML Pipeline Foundations & Ethics

Presentation 2: Data Preparation & Ethical Handling

Presentation 3: Automating Training & Fairness Checks

Presentation 4: Deployment Automation & Compliance

Presentation 5: Code Quality, Security & Ethical Governance

Next Steps

- Next: Ethical data preparation—ensuring representative, high-quality datasets.
- Building pipelines that are robust, fair, and compliant from the ground up.

Additional Resources and References

- FTC Guidelines: FTC.gov for ongoing regulatory insights.
- Legal Case Studies: Learn from *Mobley v. Workday* and *IntelliVision* scenarios.
- Operation AI Comply: Lessons in accountability and evidence-based claims.

Closing Thoughts

- Ethical AI is essential, not optional.
- By integrating fairness, transparency, and compliance, we safeguard users, developers, and organizations.
- We ensure our ML pipelines serve humanity responsibly and sustainably.