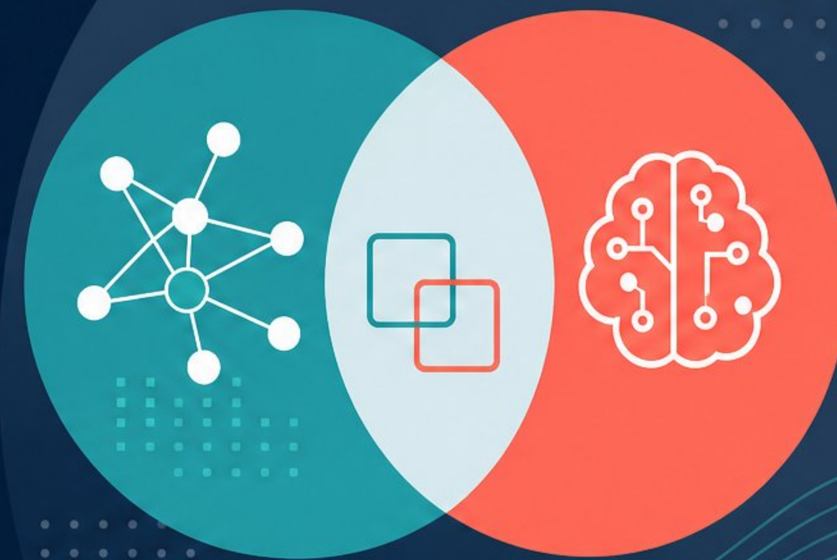
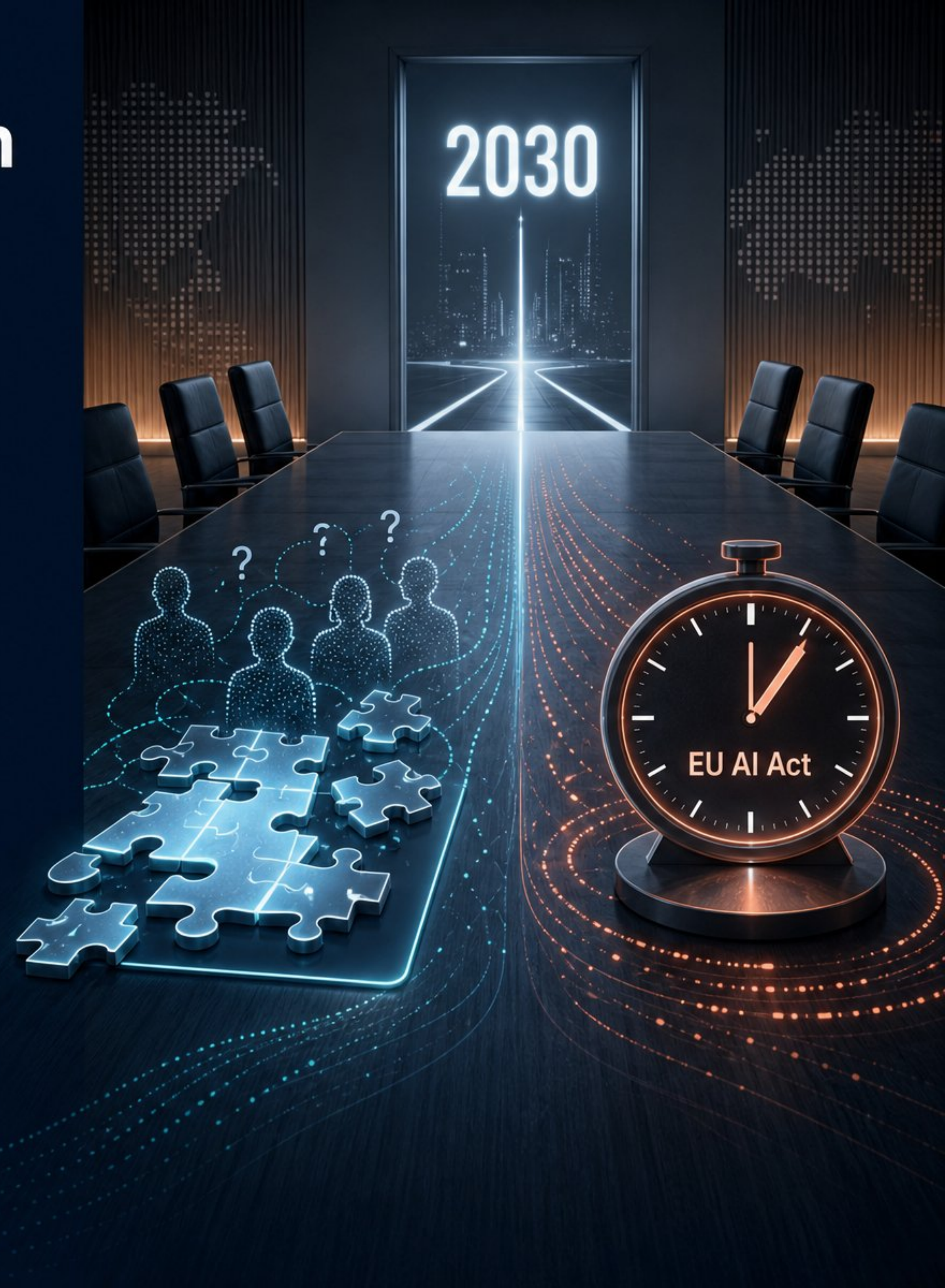
Data Science vs. Machine Learning: Roles, Methods, and Use Cases



- Persistent confusion exists in job postings and business strategy talks.
- This session clarifies definitions, boundaries, and overlaps without technical jargon.
- Designed for PMs, business leaders, analysts, students, and professionals.
- Focus on practical decisions when evaluating data and AI capabilities.

Why This Distinction Matters Now

- Heavy AI investment with misaligned roles leads to failed projects and burnout
- EU AI Act now demands clear accountability for automated decisions
- Compliance deadlines phased: high-risk Annex III from December 2027
- McKinsey: AI/ML specialists see over 80% net job growth by 2030

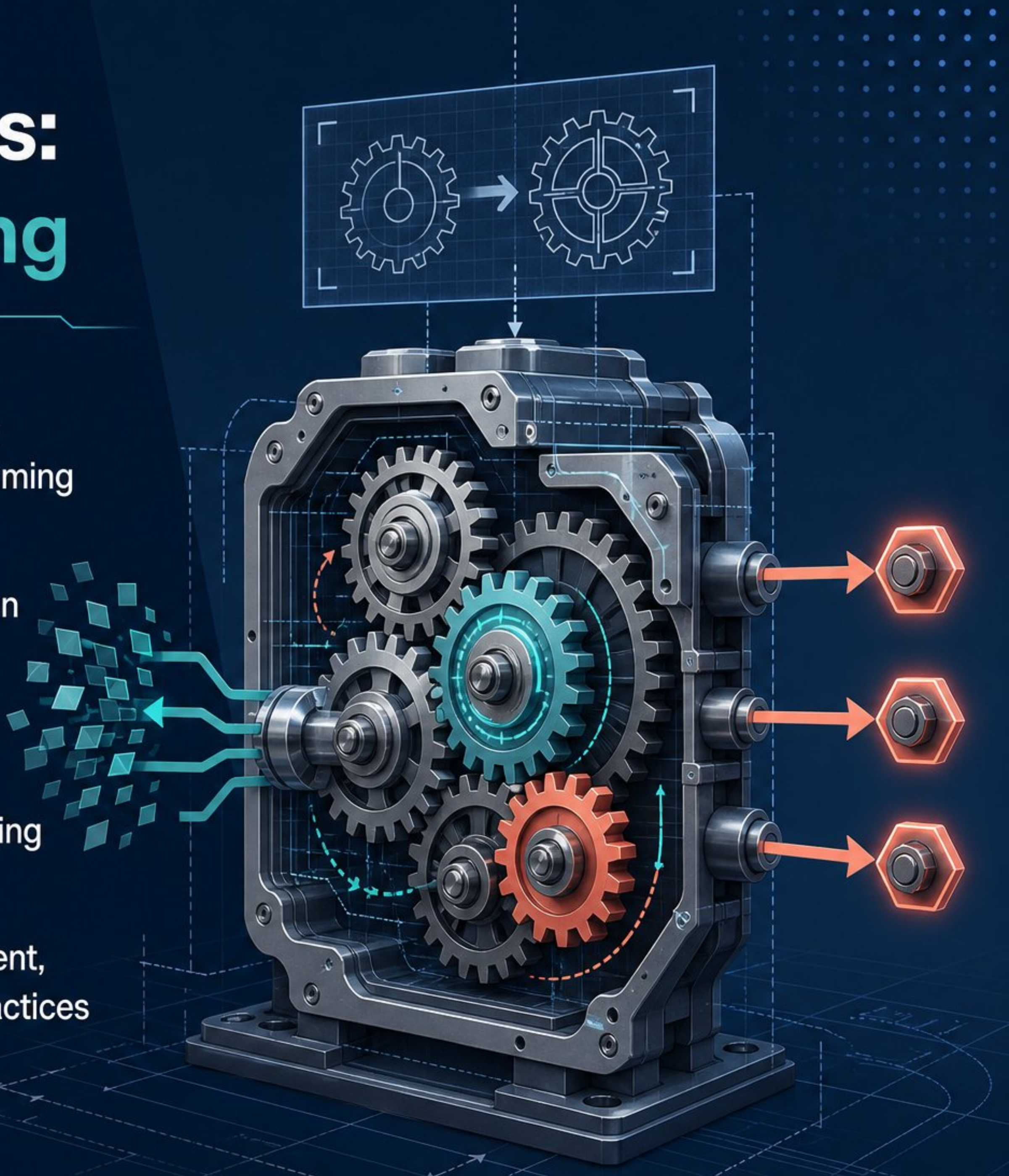


Core Definitions:
Data Science

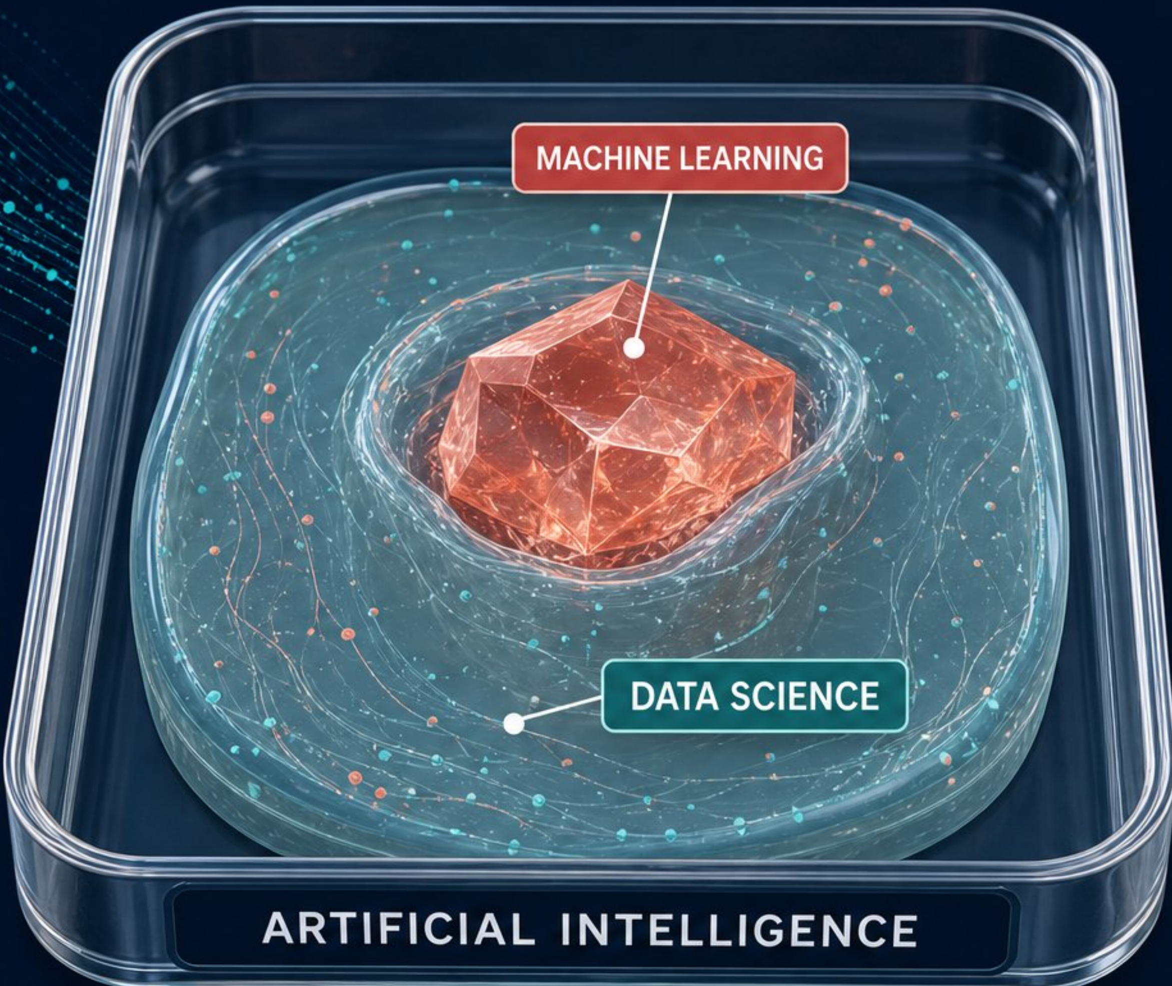
- Interdisciplinary field extracting insights for human decisions
- Answers: What happened, why, and what should we do?
- Full lifecycle: framing, collection, analysis, storytelling
- Output is typically a report, dashboard, or presentation

Core Definitions: Machine Learning

- Subset of AI that learns from data without explicit programming
- Core question: How do we run this model reliably at scale?
- Automates predictions and decisions via robust engineering
- Requires dedicated deployment, monitoring, and retraining practices



The Relationship: Venn Diagram of AI, ML, and DS

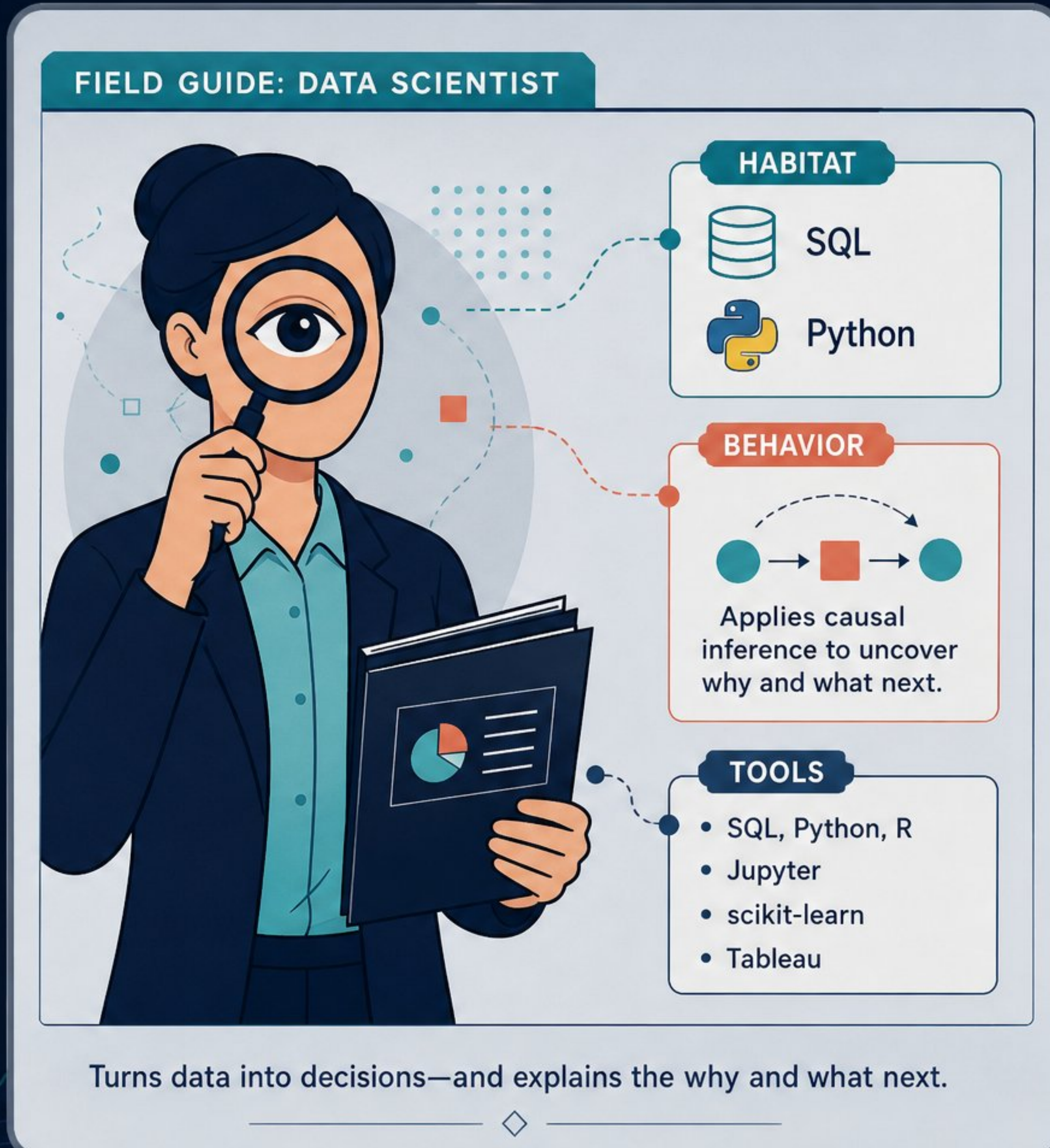


- ML is a tool within the Data Science lifecycle, but goals differ.
- Data Science targets insight; ML targets automated predictions.
- Output a slide or report? That's Data Science.
- Output a running system or API? That's Machine Learning.

Distinct Roles: Data Scientist

- Turns messy data into clear business decisions and recommendations.
- Runs A/B tests, defines metrics, and applies causal inference.
- Core tools: SQL, Python, R, Jupyter, scikit-learn, Tableau.
- Backgrounds in statistics, economics, or quantitative fields.
- Delivers the "why" and "what next" to skeptical stakeholders.

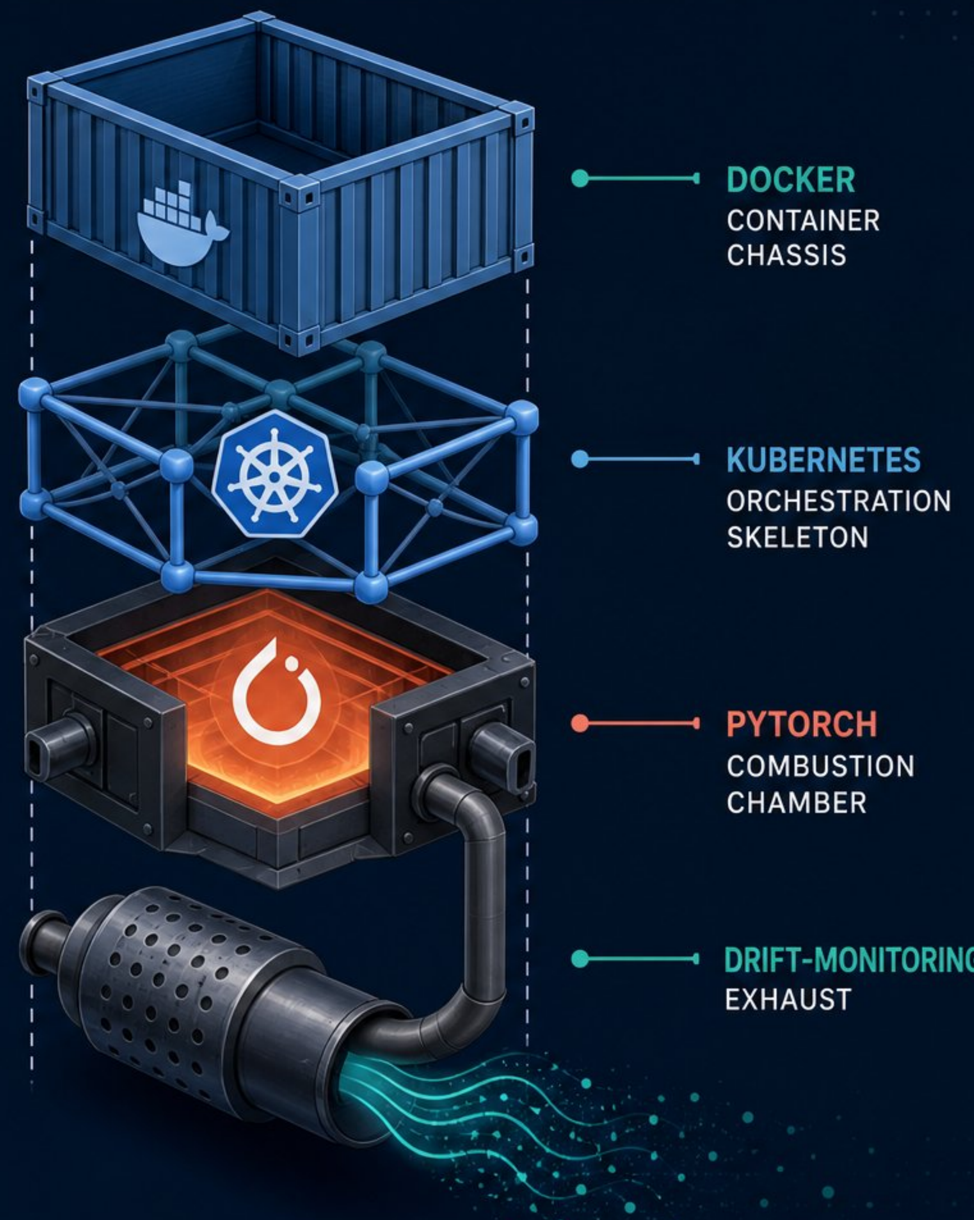
Human-in-the-Loop Silhouette



Distinct Roles:
Machine Learning Engineer

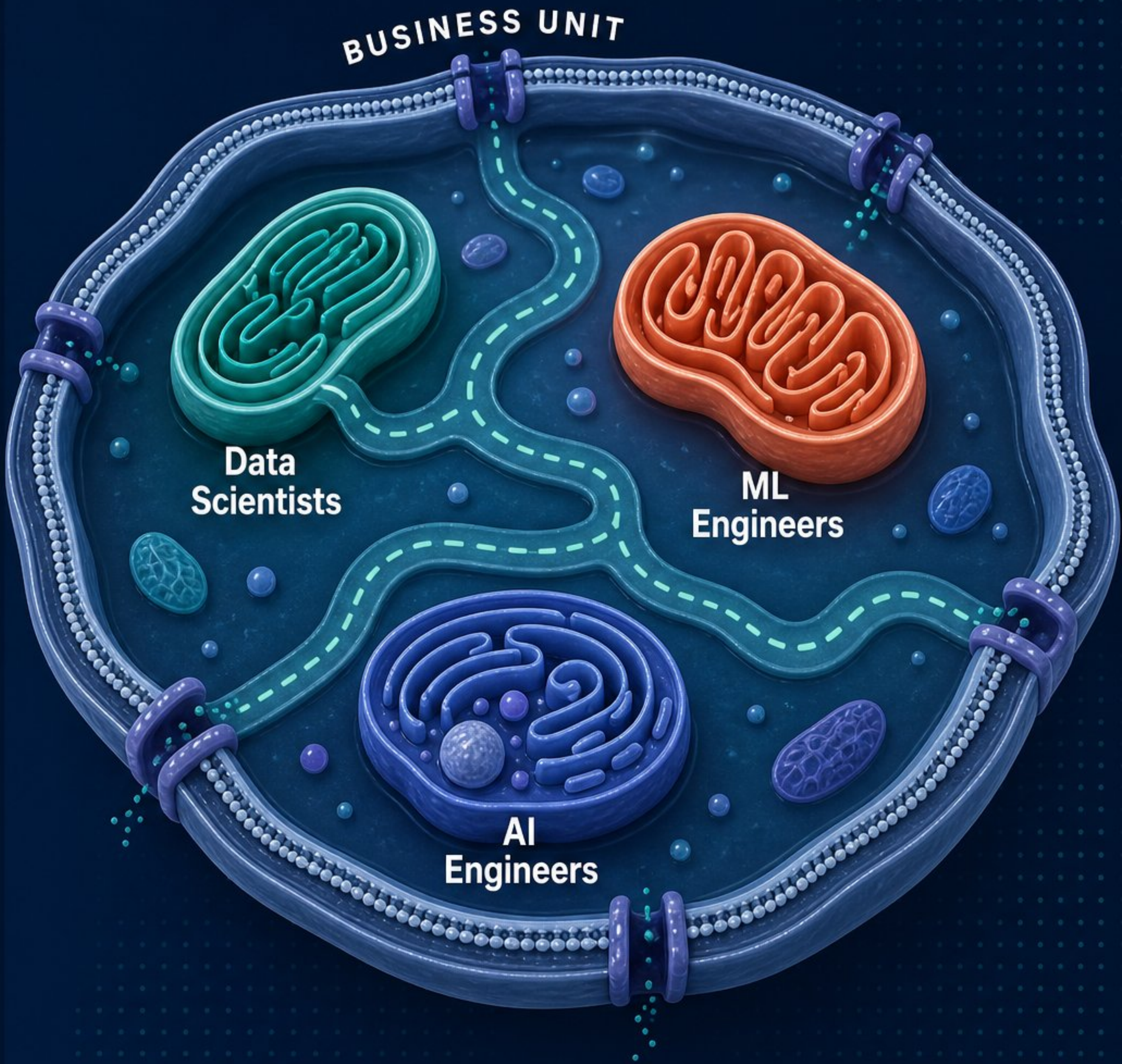
- Owns model productionization: serving, latency, throughput, monitoring, and drift detection
- Toolset: PyTorch, TensorFlow, Docker, Kubernetes, CI/CD, cloud platforms, feature stores
- Daily work: building infrastructure to serve predictions reliably at scale, forever
- Background: computer science or software engineering with production shipping experience
- Senior US salary $180K–$240K; commands a ~$43K premium over Data Scientists

PREDICTION ENGINE
EXPLODED VIEW



Emerging Roles and Team Structures

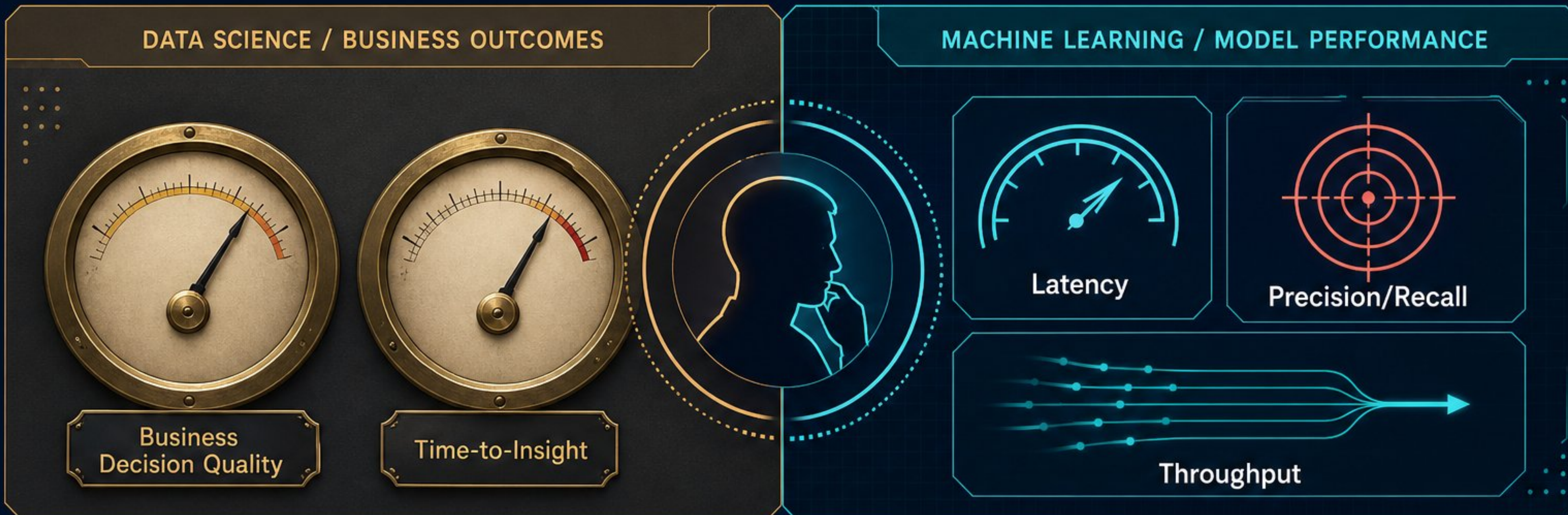
- **AI Engineer:** builds apps on top of foundation models, not training from scratch
- **~200** data scientists and ML engineers embedded across business units
- When ML is load-bearing, put both roles on one team
- Shared roadmap eliminates the "game of telephone" at handoff



Methodology: How Each Discipline Works

- **Data Science:** hypothesis testing, causal inference, time-series decomposition, clustering, and visualization.
- **Machine Learning:** supervised, unsupervised, reinforcement, and deep learning.
- **Data Scientists** validate the right question; **ML Engineers** ensure the system survives production.

Success Metrics and Evaluation

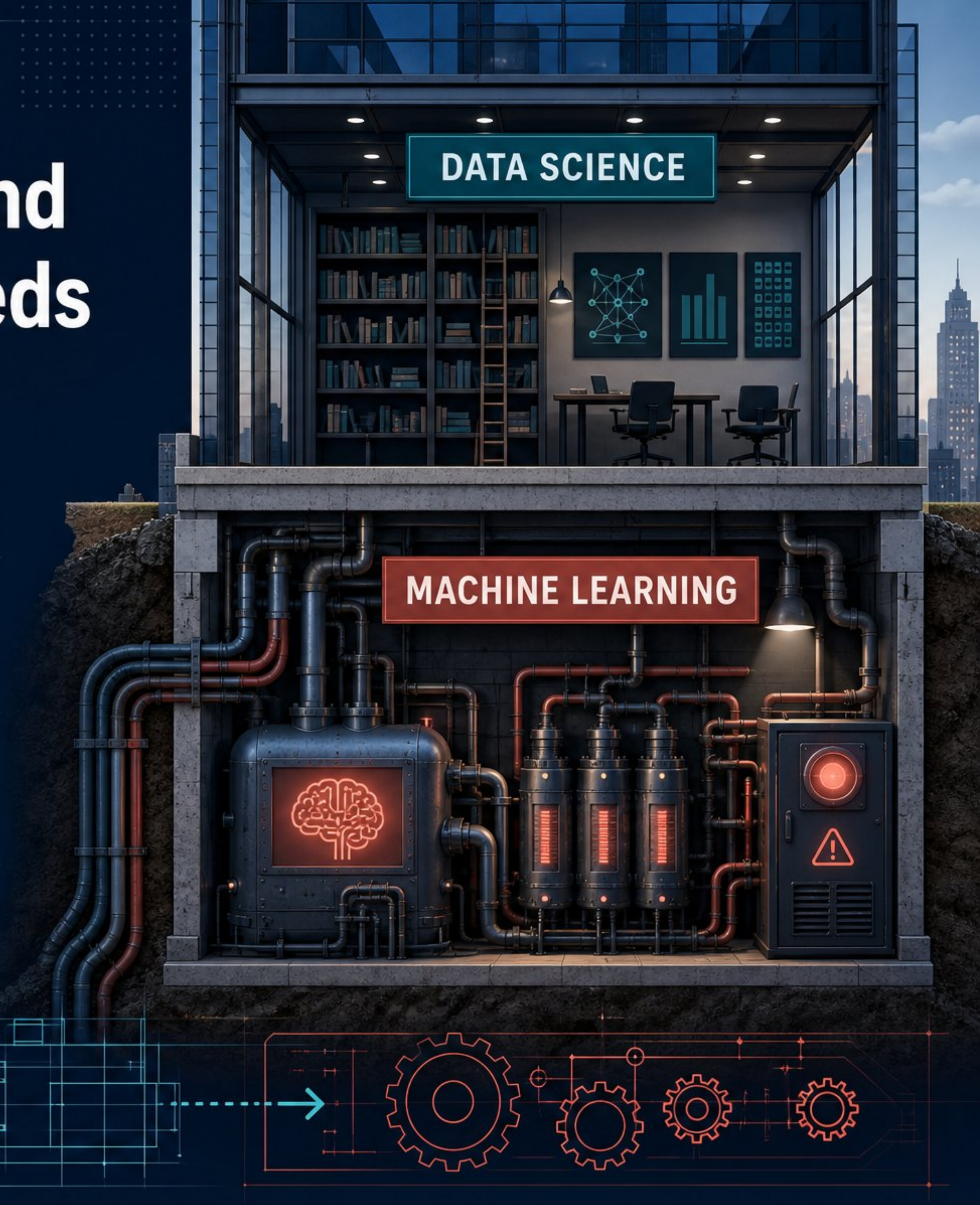


- **Data Science:** decision quality, time-to-insight, experiment velocity, data literacy
- **Machine Learning:** accuracy, precision/recall, latency, throughput, business impact
- Offline evaluation (ML) vs. observational and experimental evaluation (DS)
- A model that passes a test set can still fail in production

BLUEPRINT-TO-GEARS

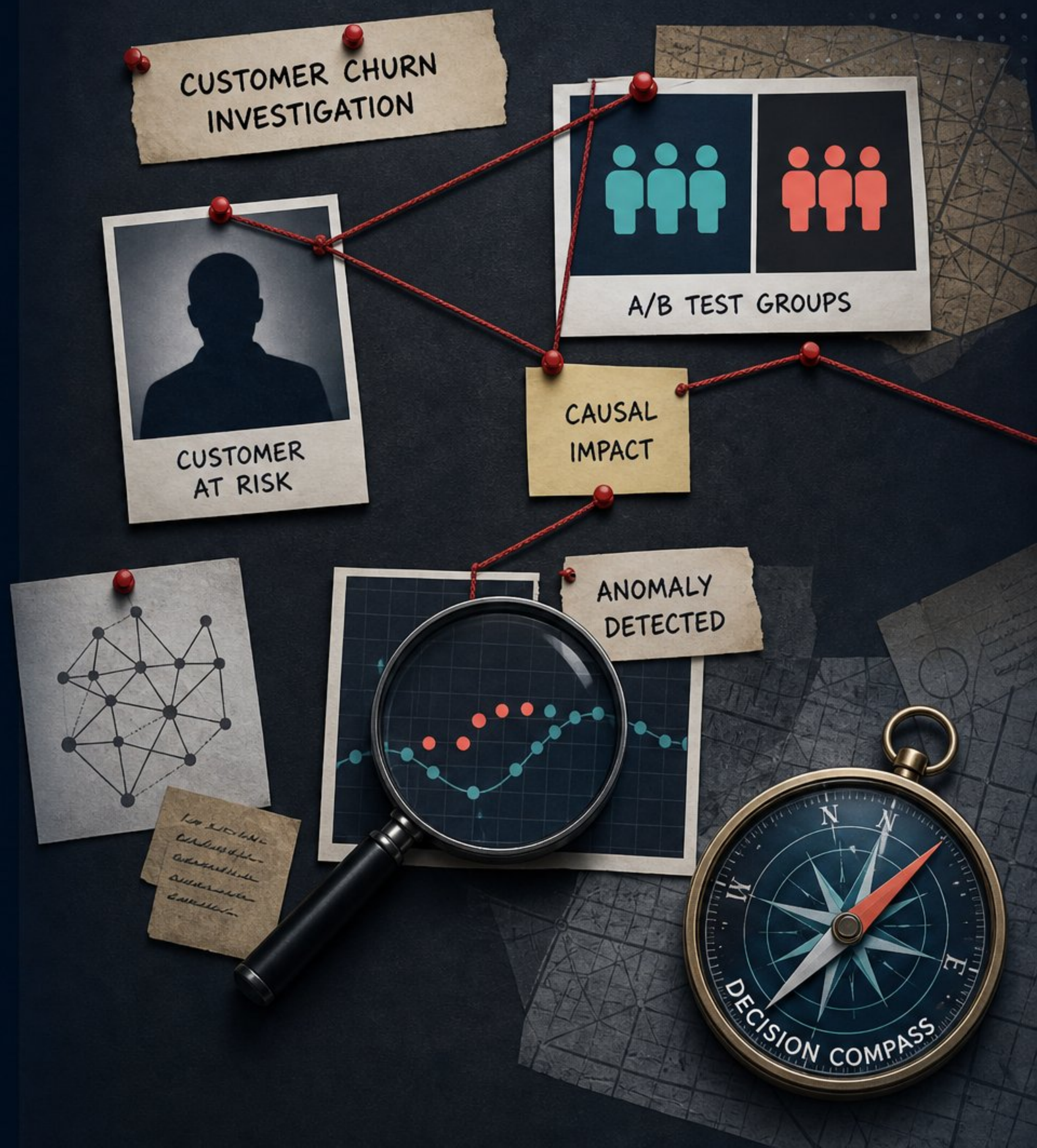
Infrastructure and Operational Needs

- **DS:** data warehouses, BI tools, analytical databases, experiment tracking
- **ML:** training pipelines, serving endpoints, model registries, drift monitoring
- ML needs its own operational discipline (MLOps) to avoid reliabiity failures
- Treating ML as just another analytical step leads to cost overruns



Use Cases: Data Science in Action

- Churn analysis and prediction to understand why customers leave and what interventions work
- A/B testing platforms to measure the causal impact of product changes and marketing campaigns
- Fraud investigation and anomaly detection using statistical patterns and exploratory analysis
- Example: BMG's StreamSight uses forecasting models to predict music royalty revenue and detect anomalies

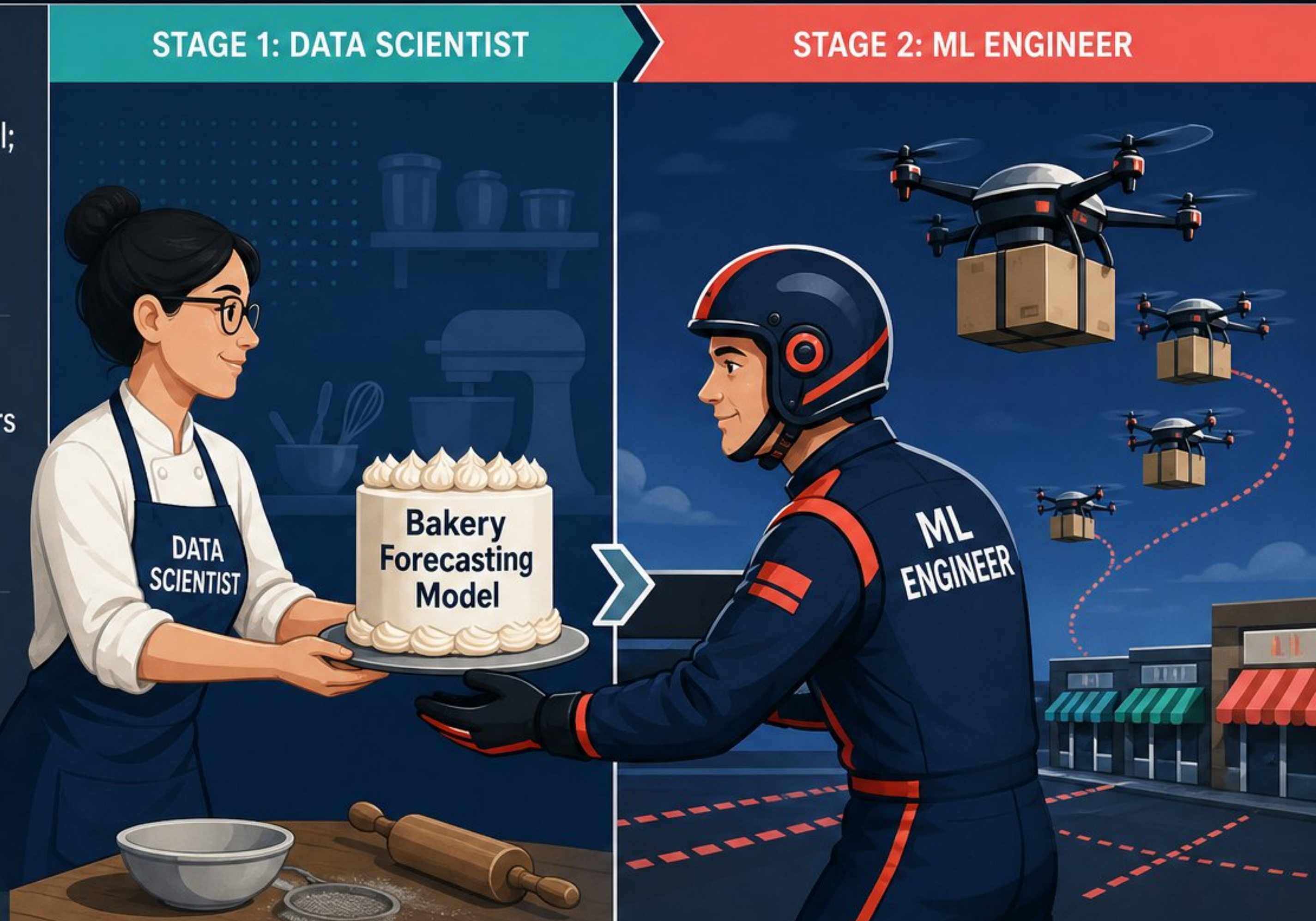


Use Cases: Machine Learning in Action

- Real-time recommender engines for personalized content and product suggestions
- Automated credit scoring and underwriting models at scale
- Predictive maintenance in manufacturing and autonomous driving perception
- Hapag-Lloyd: ML vessel schedule predictions improved schedule reliability (key industry KPI)

Collaboration: When Both Disciplines Must Work Together

- **JLP:** Data Scientists built a bakery forecasting model; ML Engineers scaled it across stores, reducing waste and boosting sales
- **BBVA** unifies data scientists and ML engineers on a shared AWS platform for model development and deployment
- **Hapag-Lloyd:** Data Scientists designed ML-driven vessel predictions; ML Engineers built scalable multi-model inference infrastructure
- **The pattern:** Data Scientists prove what's worth building; ML Engineers make it durable, monitored, and scalable



Strategic Decisions for Leaders

- Hire a Data Scientist first if the business question is still open
- Hire an ML Engineer first to scale a proven model to production
- EU AI Act: Annex III high-risk rules apply Dec 2027, not optional
- Build vs. buy must factor MLOps maturity and a 28% MLE salary premium
- Misaligning roles costs time and talent—don't bet on one becoming the other