TALK

MACHINE LEARNING

Fatih Nayebi
VP | Data & AI
ALDO Group

Scaling AI Initiatives in Retail

ODSC WEST

OCT 29-31
SAN FRANCISCO

About Fatih

Unleashing the power of data & AI to drive business value and innovation

- **VP, Data & AI** at **ALDO Group**
  Data Governance & Privacy, Data & Analytics Platform, Data & Analytics Products, AI
- **Faculty Lecturer** at **McGill University**
  Master of Management Analytics – Desautels Faculty of Management
- **Ph.D.** in **Engineering**
  Machine Learning and **Human Computer Interaction** from École de technologie supérieure
- **Author** of **Swift Functional Programming**
  Books published by Packt
- **Productionizing AI applications since 2008**

ALDO GROUP

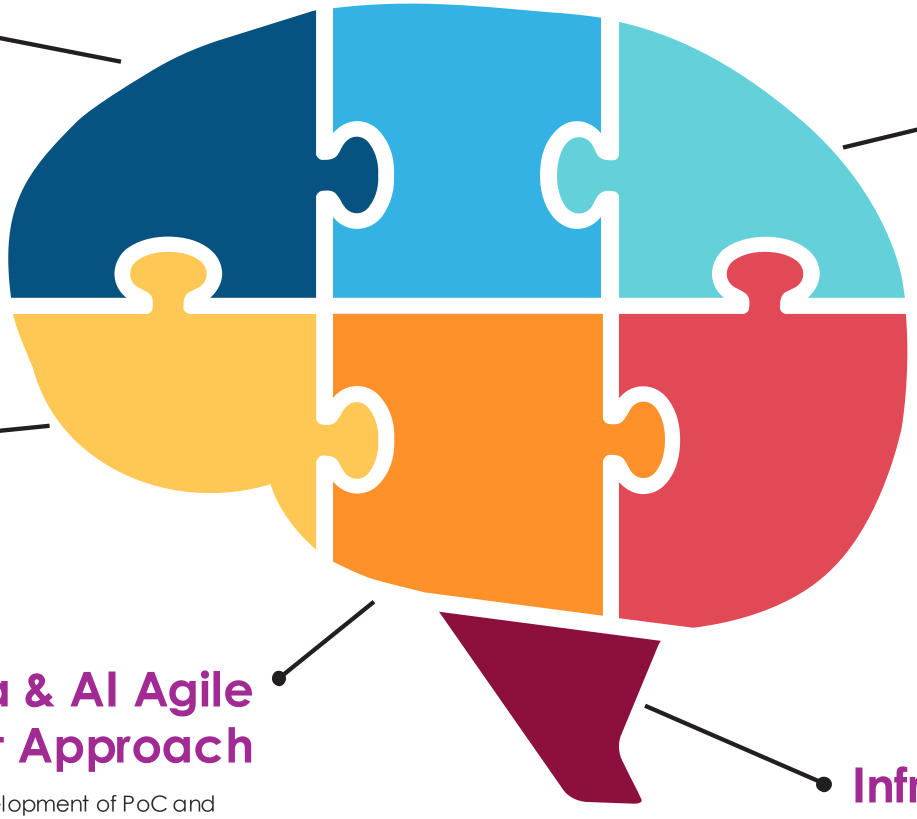
McGill

LinkedIn

1. Strategy

1. A Holistic approach to bring tangible business value
2. AI Implementation Phases: Moving from experimentation to production
3. Key Pillars of AI Strategy
4. Delivering ROI Through AI
5. Ethical AI Considerations

1.1. A Holistic Approach to bring value



Collaboration & Multidisciplinary

- Integrating diverse expertise (data science, engineering, business)
- Regular communication and knowledge sharing
- Building a culture of collaboration and continuous learning

Data Governance and Literacy

- Ensuring data accuracy, consistency, and completeness
- Compliance with data privacy regulations
- Frameworks and processes for data management
- Elevating data & AI literacy across the organization

Data & AI Agile Product Approach

- Iterative development of PoC and MVP Products
- Customer centric product
- Scalability and flexibility

Focus on Value-Added

- Prioritization based on business value, feasibility, and long-term vision
- Strategic Make-Buy-Reuse Decisions
- Instant value proposition with what is good enough

Research Mindset and Innovation

- Encouraging experimentation and innovation
- Staying updated with the latest AI research and trends
- Pilot university projects to test new ideas and approaches

Best Practices in AI Development

- Adhering to ethical AI principles
- Implementing robust testing and validation processes
- Monitoring and maintaining AI systems post-production
- Leveraging existing AI models

Infrastructure needed for AI

- Robust cloud-based infrastructure (AWS)
- High-performance computing resources
- Scalable storage solutions

1.2. AI Implementation Phases: Moving from experimentation to Production

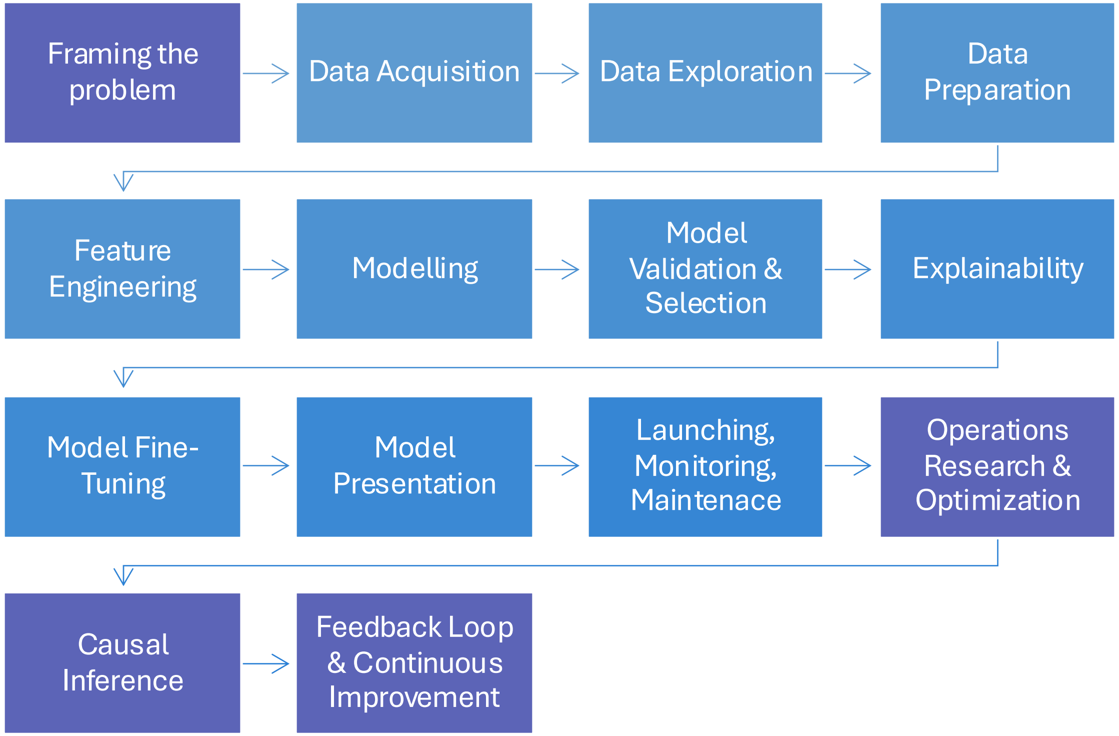
Challenges:

- Technical Complexity
- AI Data Availability
- Resource Management
- System & Process Integration

Aligning AI with key business **objectives**

AI Product Management

Transition from **experimentation** to **full-scale implementation**

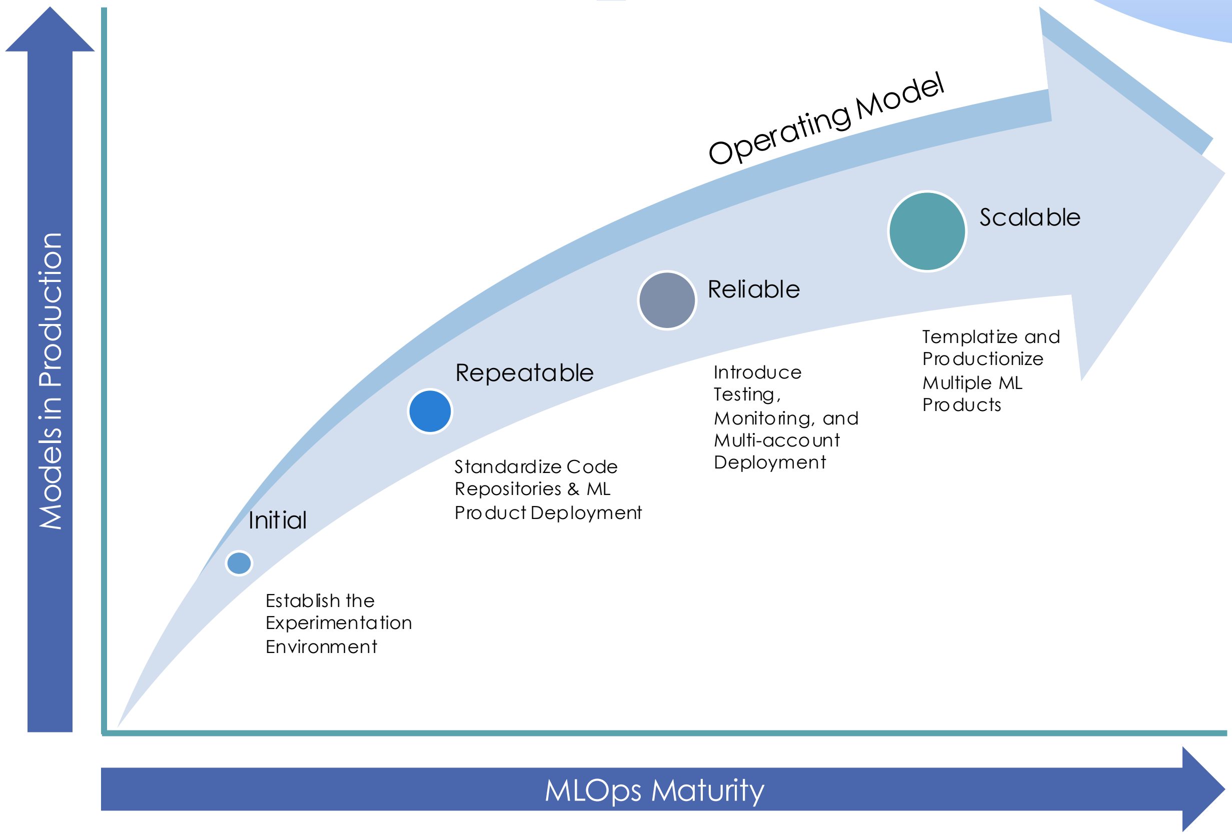


1.3. Key Pillars of AI Strategy

Prioritization of use cases that **bring competitive advantage**

Optimizing AI Deployment with **MLOps** to scale and maintain ML models

Unifying engineering, science, governance, subject matter expertise, and AI applications



1.4. Delivering ROI Through AI

1.5% Increase in **Gross Margin**

Reducing Markdowns & Stockouts

Enabling granular hierarchical forecasting and overall, **10% Increase** in **Forecasting Accuracy**

Empowered and enhanced **Decision Making**

Improved Customer Engagement

Cost optimization and **increased productivity**

1.5. Ethical AI Considerations

Addressing **bias**

Ensuring **privacy**

Maintaining **transparency**

Building **trust** with **customers** through **responsible AI**

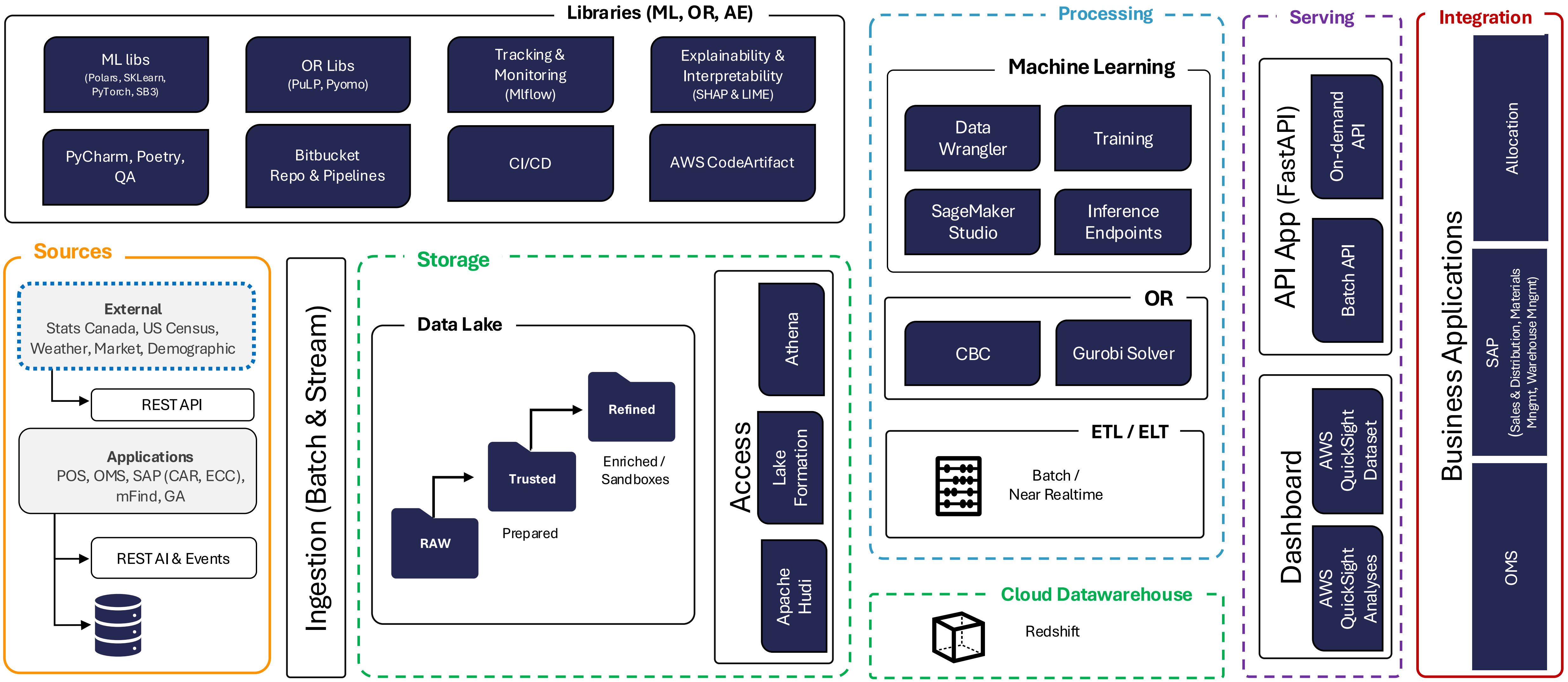
AI Governance: establishing frameworks for monitoring, auditing, and ensuring fairness

Explainability vs. **accuracy trade-off**

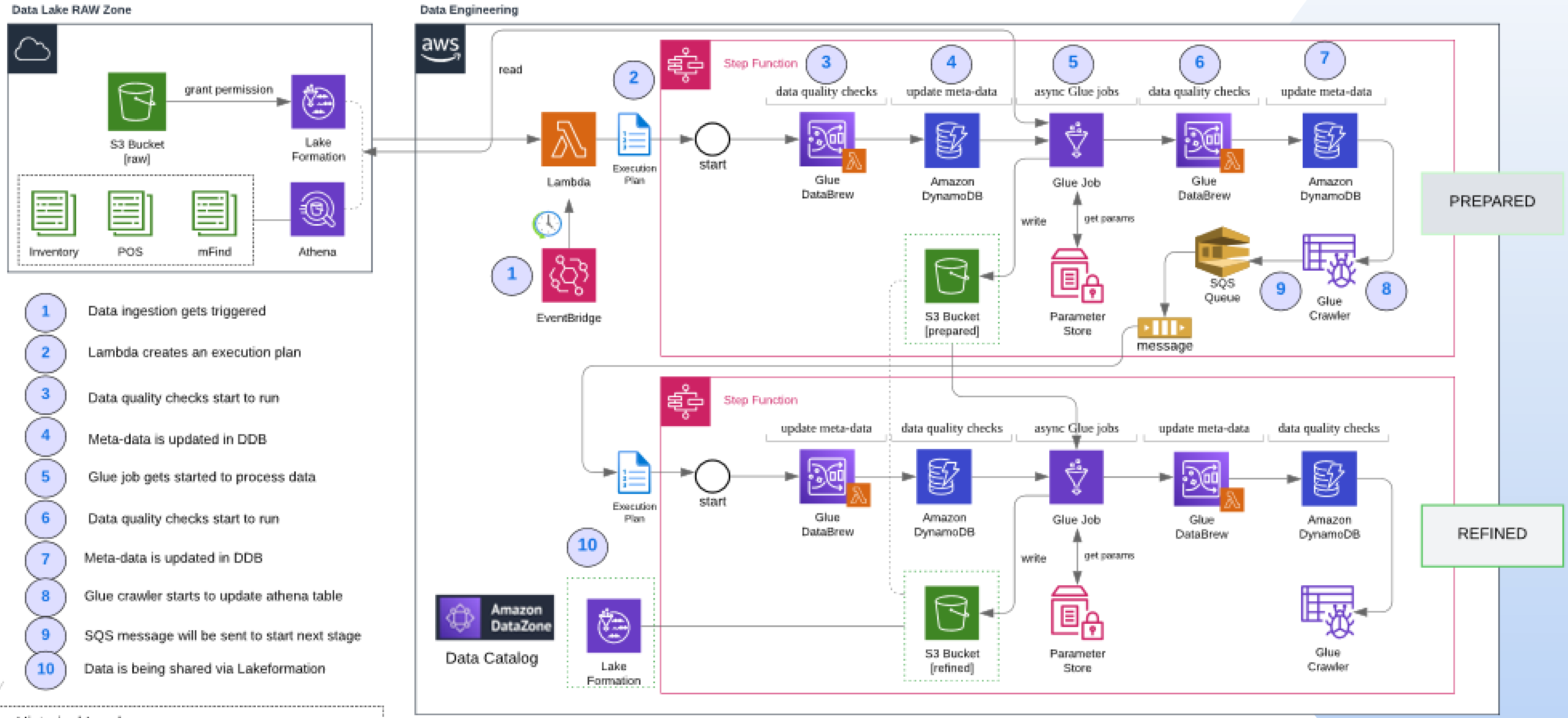
2. Architecture

1. **Machine Learning and Operations Research Application**
2. **Data Engineering**
3. **Data Science & MLOps**

2.1. ML & OR Application Architecture



2.2. Data Engineering Architecture



1 Data ingestion gets triggered
2 Lambda creates an execution plan
3 Data quality checks start to run
4 Meta-data is updated in DDB
5 Glue job gets started to process data
6 Data quality checks start to run
7 Meta-data is updated in DDB
8 Glue crawler starts to update athena table
9 SQS message will be sent to start next stage
10 Data is being shared via Lakeformation

Historical Load:

Passing a list of partitions to be processed to the queue

2.3. Data Science & MLOps Architecture

11 Data is being mirrored in the Athena (from datalake to data science account)

12 Views are built on the Athena

13 EventBridge triggers Lambda Function on schedule

14 Lambda updates partitions

15 SageMaker pipelines consumes the data from the views

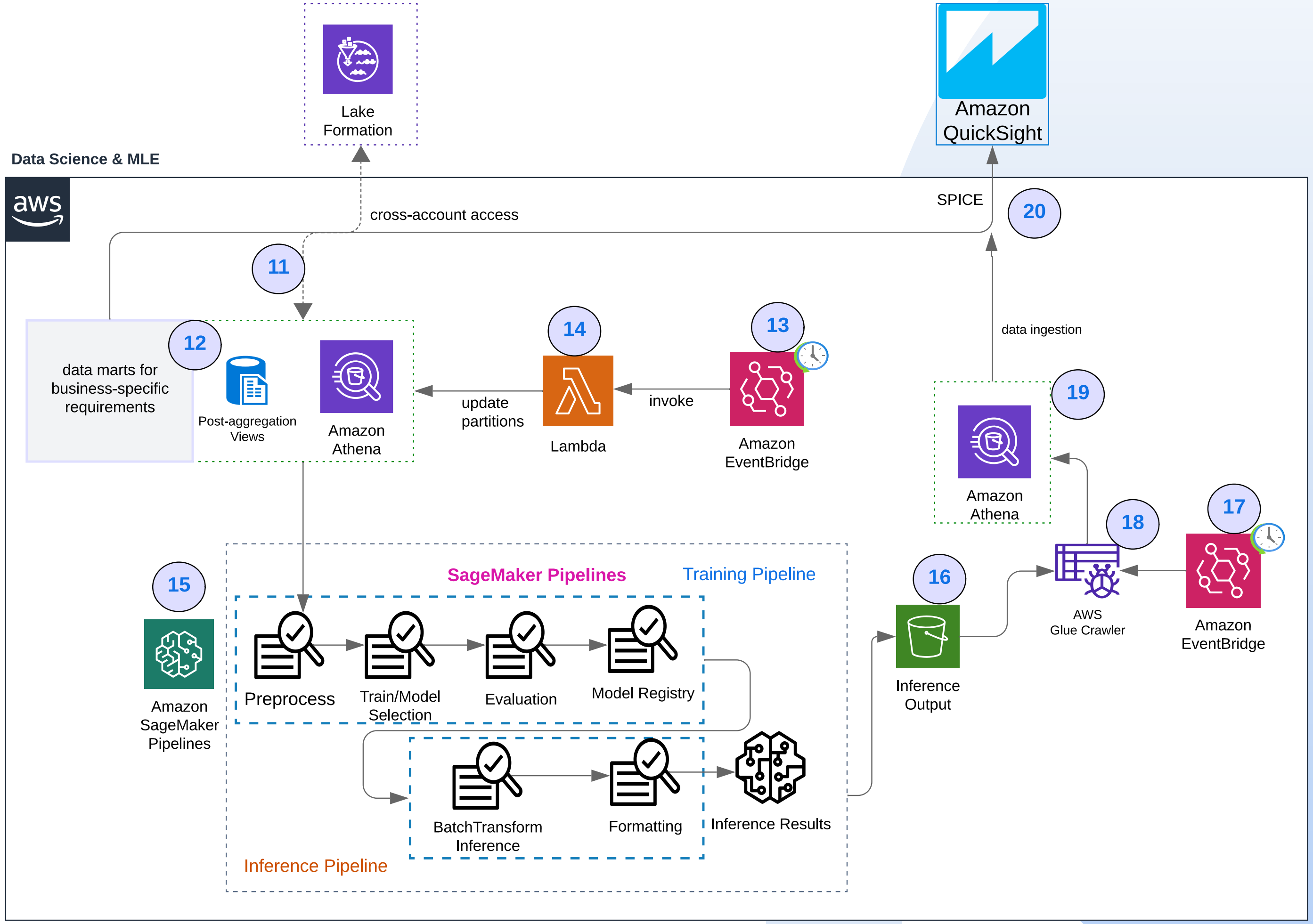
16 Inference Pipeline writes the inference output in S3

17 EventBridge triggers the crawler on schedule/Event

18 The crawler crwles the S3

19 The results data are made available to the Athena by the crawler

20 The results data are ingested to spice to be displeyed in QuickSight dashboards



3.
AI Use Cases & Products

Focus on **aligning AI** with **business objectives**

0. Product Lifecycle: From Design to Client with AI

1. Demand Forecasting

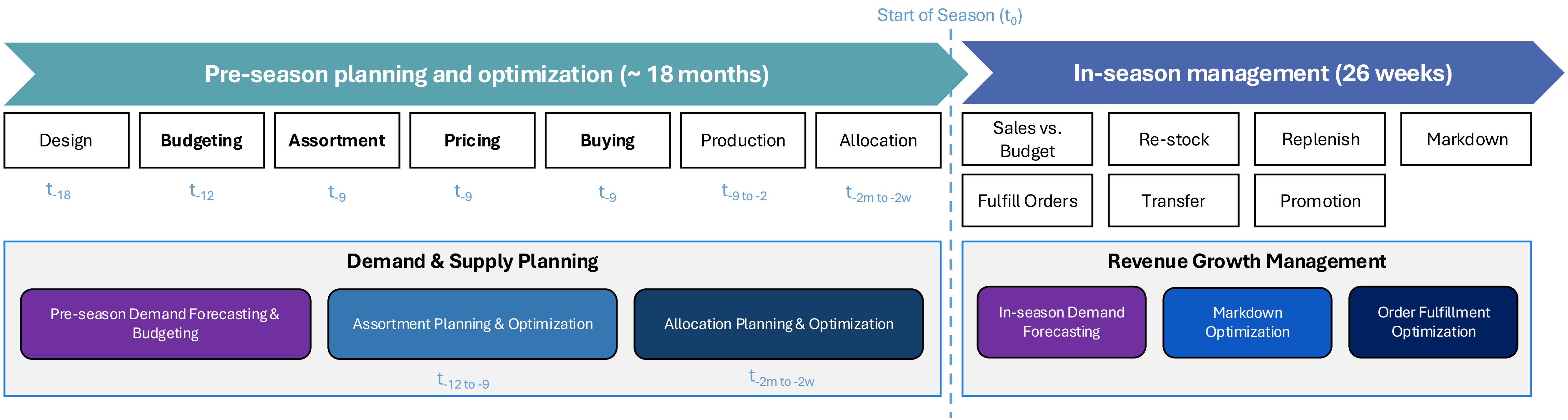
2. Revenue Growth Management

3. Demand & Supply Planning

4. Research Projects with **McGill University**

5. Gen AI & Retrieval Augmented Generation (**RAG**)

3.0. Product Lifecycle: From Design to Client with AI



Granular and Accurate Demand Forecasting is the key pillar!

3.1. Demand Forecasting

A multifaceted challenge involving various levels of granularity, hierarchy and numerous influencing factors.

Multiple levels of granularity – Country, Channel, Merchandise Division, Product Category, Merchandise Category, Store, SKU

Skewed Target Product Demands

Influencing Factors

- **Seasonality:** Seasonal Trends, Seasonal Promotions, diverse locations in NA - **External factors:** Macroeconomic, market trends, weather conditions - **Promotional Activities:** Campaigns and Discounts, Collaborations

Diverse Time-Series Data

- **Complex Time-Series Patterns:** Long Term Trends and Short-Term Fluctuations – **Hierarchical Time-Series:** Top-Down, Bottom-Up, Mid-level, and Hybrid

3.1.1. Necessity of *Scalable Machine Learning* Techniques

Complexity Management

- **Handling Non-Linear Relationships** – Ensembles
- **Feature Engineering** – Sophisticated Feature Engineering to extract meaningful insights from high dimensional data

Improving Accuracy

- **Predictive Power** – Advanced models to enhance predictive power and accuracy
- **Error Reduction** – Feature Engineering, Log Transformation and Bayesian Optimization

Scalability and Efficiency

- **Automated Model Training and Experimentation Management** – Automated Data Preparation, Feature Engineering and Hyperparameter Tuning to automate and optimize model training
- **Real-time Forecasting** – Deploying models capable of real-time demand forecasting

3.1.2. Feature Engineering & Selection

Lagging and Differencing at different levels (country, generic ID, style, channel, merch category, etc.)

- Time Frames: ranging from 4 to 52 previous weeks to capture short-term and long-term dependencies

Rolling Windows – Generating rolling lag variables to capture moving averages and trends

- Time Windows: Last 4, 8, 12, and 26 weeks to identify seasonal and cyclical patterns

Categorical Encoding with advanced techniques (e.g. Entity Embedding)

Skewed Target Balancing to reduce skewness

- Log Transformation, Box-Cox, and Sampling Techniques

KPI (CONV%, AoV, TRFC) forecasting and **Monotonic Constraint** definitions

Statistical Methods – Polynomial features, tsFresh for automated time-series feature extraction

Proportions of Merch Category and Product Sales – 1v1 and 1vO

RFECV and Feature Importances

3.1.3. Data Preparation, Modelling and Experimentation

Data Preparation

- **Imputation/Extrapolation** for COVID-affected periods
- **Scaling** – Standard and MinMax
- **Outlier Detection** – Anomaly and Novelty Detection
- **Clustering** for products and stores
- **Product Mapping** based on **Fashion-CLIP**
- **Dataset Splitting** – Temporal, Stratified, group shuffle, rolling window splitting for train-validation-test sets

Modelling Techniques

- **Gradient Boosting** Approaches, AWS Forecast, Chronos and other Transformer inspired models
- **Direct**, **Recursive**, and **Multi-Output**

Model Evaluation

- **Performance Metrics** – MAPE, WMAPE, BIAS, MAE% and Explainability

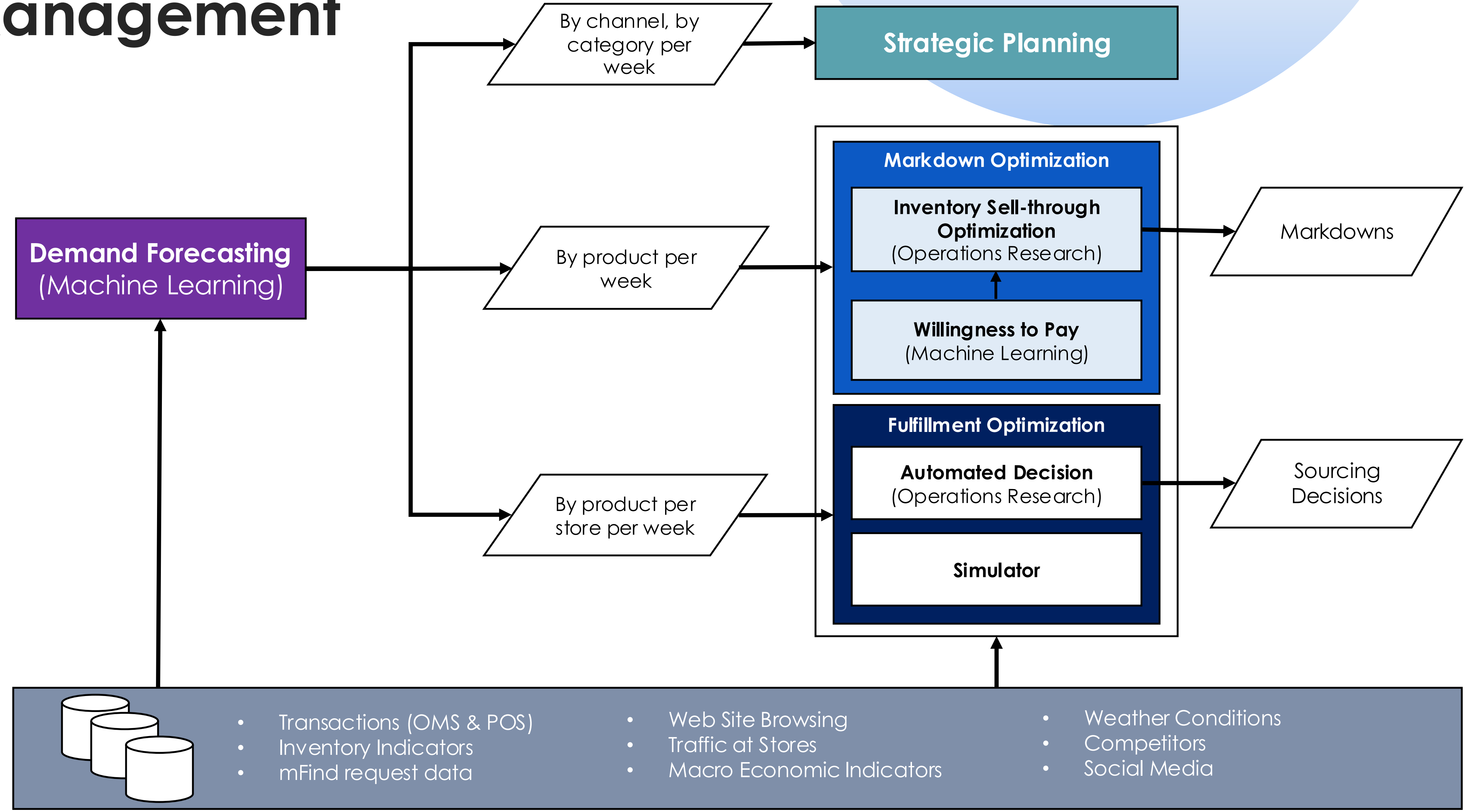
Forecasting and Error Analysis

- **Balancing Short-Term with Long-Term Forecasts**
- **Error Analysis** to identify the areas model performs poorly
- **Channel-Specific Features** – Incorporating ecomm-specific freatures to better capture trends

Experimentation Management

- Leveraging **MLFlow** to **track experiments**, parameters, metrics, and results

3.2. Revenue Growth Management



3.3. Demand & Supply Planning

Pre-Season Demand Forecasting (ML)

- **Granularity**: SKU-store-month level
- **Models**: ML for Time-Series Forecasting
- **Outputs**: demand predictions per SKU, seasonal demand curves
- **Feedback Loop**: a feedback loop from in-season sales data to update forecasts dynamically.

In Season - SKU demand by store by Month

Assortment Planning & Optimization (OR)

- **Granularity**: SKU-store level
- **Models:** MIP for maximizing gross margin
- **Constraints:** Demand, Space & Display, Financial, Product Cost & Price, Cannibalization
- **Outputs**:
  - Optimal assortment mix: ideal product selection for each store
  - Buying Decision: quantity and timing of purchases, explicitly linking to procurement
  - Expected gross margins

Assortment Mix

Buying Decision

Allocation Planning & Optimization (OR)

- **Granularity**: SKU-prepack-store level
- **Models**: Linear programming / heuristic-based allocation
- **Constraints:** Inventory, Store Demand, Shipping & Handling, Prepacks
- **Outputs**:
  - **Pre-Packing Decision**: How products are bundled for easy allocation.
  - **Pre-Season Allocation**: Distribution of inventory before the season begins.
  - **In-Season Allocation (Iterative)**: Continuous reallocation based on sales data.

Pre-Packing Decision

Pre-Season Allocation

In-Season Allocation

3.4. Research Projects

- Fit & Comfort Prediction for Return Management (Machine Learning)
- Prescriptions for Sales Enhancement (Causal Inference)
- Low Transactional Loyalty Management (Literature Survey and Implementation Strategy)
- Pricing Strategy & Elasticity (Machine Learning & Simulation)
- Client Lifetime Value for Next Best Action (Machine Learning)
- On Model Product Placement (Gen AI)

3.5. Generative AI & RAG

Trend-Aware & Conversational Product Recommendation for **Ecommerce**

Retrieval Augmented Generation (**RAG**) based on OpenAI Assistants API for:

- **Associate Assistant**
- **Product Descriptions**
- **SEO Descriptions**

Data Augmentation, **synthesis**, and **automation**

Microsoft **Co-Pilot** and **Midjourney** for **Productivity**

Lessons Learned & Best Practices

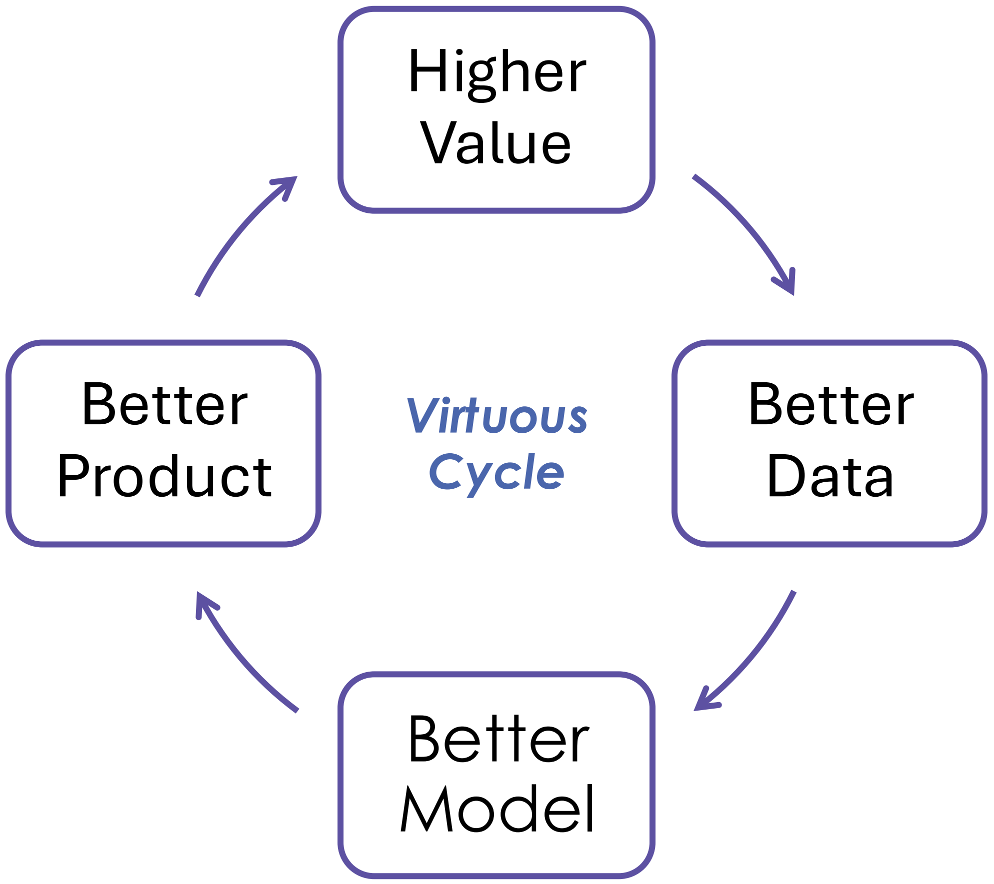
Start with **business-driven** use cases, **quick wins**

Importance of **Data & AI literacy** across organization

Invest in **MLOps** for **scalable**, **long-term AI success**

Cross-functional collaboration and **agile product management**

Beware of **Flywheel effect** and establish , **continuous feedback loop**



Thank you

Fatih Nayebi, Ph.D.

Github