

THARVA

CONSULTANCY

 Data

 AI

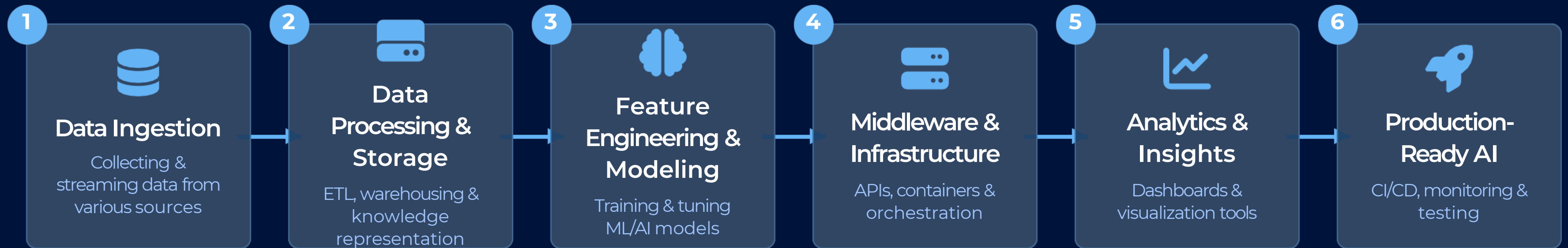
 MLOps

 Analytics

From Data Ingestion to Production AI Systems



The AI Pipeline at Tharwa



Our end-to-end AI pipeline transforms raw data into production-grade AI systems with enterprise-ready infrastructure and tooling



Our Capabilities & Technologies

Data Ingestion & Processing

- Automated metadata ingestion
- ETL & real-time streaming: Kafka, Airflow, dbt
- OCR + multimodal ingestion: Tesseract, OpenCV

Storage & Knowledge Representation

- Vector databases: FAISS, Pinecone, Weaviate
- Knowledge graphs: Neo4j, RDF4J
- Data lakes/warehouses: Snowflake, BigQuery, Delta Lake

Modeling & Training

- Tuning LLMs: LoRA, PEFT, RLHF
- Traditional ML: XGBoost, Scikit-learn
- Deep learning: PyTorch, TensorFlow, HuggingFace Transformers

Middleware & Infrastructure

- Microservices & APIs: FastAPI, gRPC
- Containerization & orchestration: Docker, Kubernetes
- MLOps pipelines: MLflow, Kubeflow, SageMaker

Analytics & Visualization

- Dashboards: Power BI, Tableau, Superset
- Real-time monitoring: Prometheus, Grafana

Production AI

- CI/CD for models: ArgoCD, GitHub Actions
- Model monitoring: Evidently AI, WhyLabs
- A/B testing frameworks

✓ Our full-stack AI capabilities enable seamless implementation of end-to-end intelligent systems





Selected Projects

The Art of AI speaks louder than words —————

E-commerce Recommendations & Healthcare Diagnosis



Personalized E-Commerce Recommendations

Pipeline:

Clickstream Ingestion → Embeddings Creation → Recommendation API

Technologies:

Vector DB (FAISS) Matrix Factorization
</> BERT Embeddings FastAPI Redis Cache MLflow

Our e-commerce recommendation system processes user behavior data in real-time, creating semantic embeddings of products and user preferences. The system delivers personalized recommendations via a high-performance API with automated retraining cycles to continuously improve accuracy.



AI-Driven Healthcare Diagnosis Assistant

Pipeline:

Medical Images + EHR → Diagnosis Suggestions → Explainable Results

Technologies:

OCR (Tesseract) FHIR Ingestion CNNs (PyTorch)
HuggingFace Models Captum XAI HIPAA Kubernetes

This healthcare system combines medical imaging analysis with electronic health record data to provide diagnostic assistance to healthcare professionals. Using explainable AI techniques, the system provides transparent reasoning behind its suggestions while maintaining strict HIPAA/GDPR compliance.

Document Analysis & Energy Forecasting

Automated Legal Document Analysis

Pipeline

OCR → LLM Summarization → Clause Extraction

Technologies

Tesseract OCR

spaCy

GPT-4

LoRA adapters

Neo4j

Streamlit

Key Results

- ♦ 90% reduction in document processing time
- ♦ Automated extraction of legal clauses and entities
- ♦ Knowledge graph for relationship visualization

Implemented for a Fortune 500 law firm, our legal document analysis system processes over 10,000 pages daily with 97% extraction accuracy. The knowledge graph has identified previously undetected contractual relationships, reducing litigation risk by 35% and saving \$24M annually in manual review costs.

Dynamic Energy Demand Forecasting

Pipeline

Weather + IoT → Demand Prediction → Grid Optimization

Technologies

Apache Kafka

Delta Lake

Prophet

LSTM (PyTorch)

Prometheus

Key Results

- ♦ High-accuracy 24-hour forecasting (95% accuracy)
- ♦ Automated alerts for demand spikes
- ♦ Reduced grid balancing costs by 15%

Deployed across 3 major utility providers serving 4M+ customers, our energy forecasting system integrates real-time weather and usage data from over 50,000 IoT devices. The system has prevented 7 potential brownouts during peak demand periods and optimized renewable energy integration, resulting in 12% reduction in carbon emissions.

Document Analysis & Energy Forecasting



Retail Inventory Optmization

Pipeline

POS → Demand Forecasting → Replenishment

Technologies

Airflow ETL

Kafka

ARIMA

Github Actions

Neo4j

FastAPI

Key Results

Implementation reduced inventory costs by 28% while improving product availability by 15%. The hybrid forecasting model achieved 94% accuracy across seasonal product lines, outperforming previous systems by 22%.



Financial Fraud Detection

Pipeline

Transactions → Anomaly Detector → Alert Dashboard

Technologies

Apache Kafka

PowerBI

Prophet

Isolation Forest

Graph Sage

Key Results

System detected 97% of fraudulent transactions with a false positive rate of only 0.3%. Real-time graph analysis reduced fraud detection time from hours to seconds, saving an estimated \$4.2M annually in prevented losses.

Why Tharwa Consultancy Stands Out

Full AI Lifecycle Coverage

- We cover the **complete AI journey** from data ingestion to production deployment
- Integrated approach eliminates **technology silos** and handoff friction
- End-to-end **accountability** with single-team ownership

Cutting-Edge Technology Stack

- Advanced LLM fine-tuning with LoRA and RLHF techniques
- Sophisticated **knowledge representation** with vector DBs and knowledge graphs
- **Enterprise-grade** infrastructure with proven scalability

Cross-Domain Expertise

- Proven impact across **diverse industries**: smart cities, healthcare, energy
- Deep **domain knowledge** in retail, finance, and manufacturing
- Transferable **AI patterns** that accelerate implementation

Future-Proof Infrastructure

- **Scalable** architecture that grows with your business needs
- **Explainable AI** with transparency in decision-making processes
- **Regulatory compliance** built into every solution (GDPR, HIPAA, CCPA)

★ Our unique combination of technical expertise and domain knowledge enables us to deliver AI solutions that create measurable business value.



Empowering Smarter Futures with Scalable AI

Riyadh, Kuala Lumpur, Paris, London, Amsterdam

Tharwa Consultancy

info@tharwaa.co

www.tharwaa.co

