

Translating Manufacturing problems into AI models

Document not to be used for teaching. All rights reserved by the author, Dr. Mohamed ASSELLAOU.

Mohamed ASSELLAOU



Manufacturing

Document not to be used for teaching. All rights reserved by the author, Dr. Mohamed ASSELLAOU.

Data-rich environment

Real-time opportunities

AI-driven insights

Questions to ask:

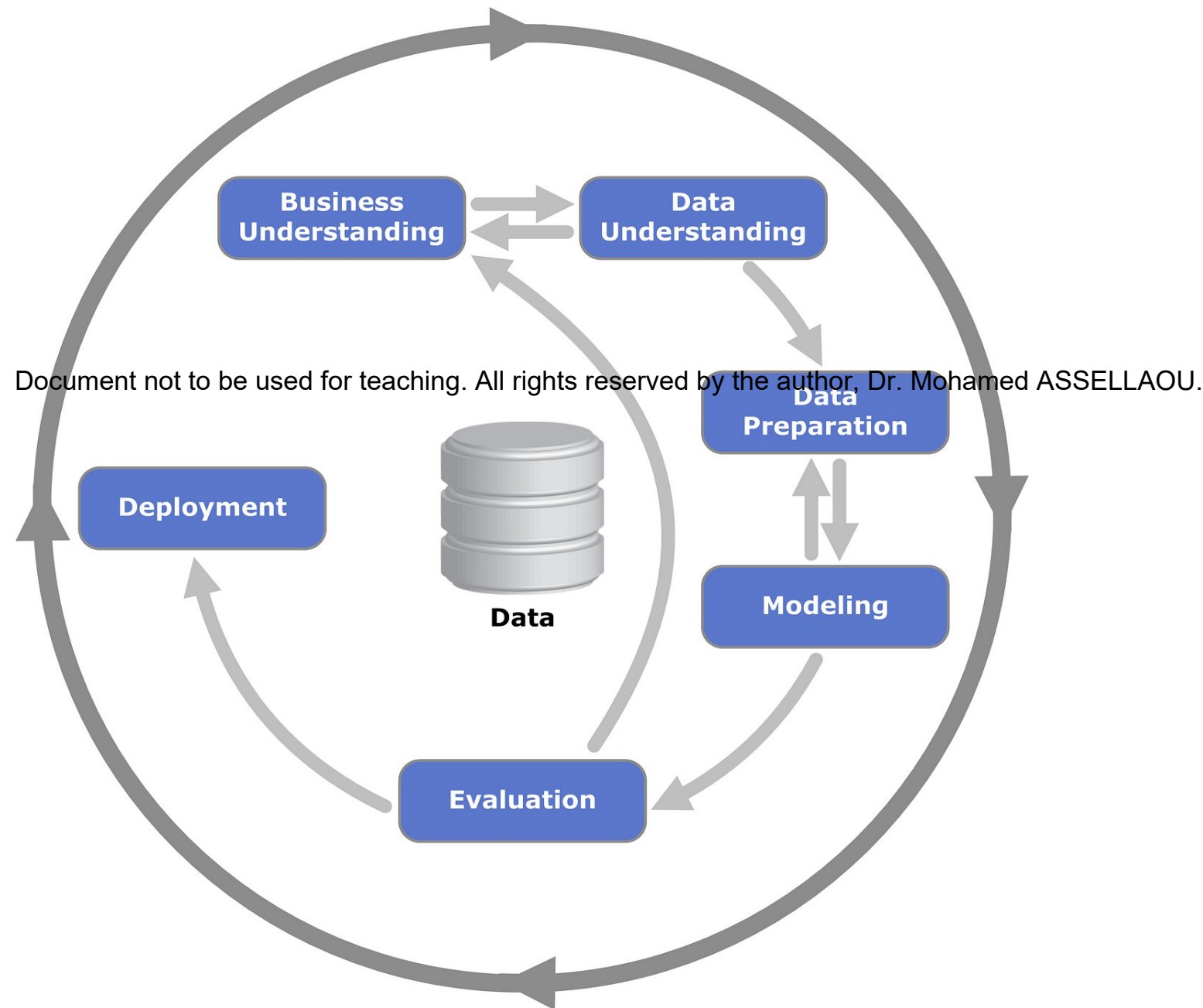
- What are the business needs and requirements?
- Is machine learning the best solution for this problem?
- What are the inputs and outputs of the problem?

Document not to be used for teaching. All rights reserved by the author, Dr. Mohamed ASSELLAOU.

- Where is my data located?

- How do I collect the data?
- Is the data ready?
- How do I transform and structure my data?
- How do I select features for the model?
- How do I measure the success of the solution?

Cross-Industry Standard Process for Data Mining (CRISP-DM) :



ML project needs and requirements

Business objectives

- Production optimization
- Operational cost reduction
- Security improvement
- Maintenance prediction

Technical constraints

- Available IT infrastructure
- Required computing capacity
- Integration with existing systems
- Acceptable latency

Human resources

- Available data science expertise
- Field team support
- End-user training
- Data governance

Budget and timelines

- Initial investment
- Expected ROI and horizon
- Deployment phases

Document not to be used for teaching. All rights reserved by the author, Dr. Mohamed ASSELLAOU.

Is ML the best solution?

Criteria for favoring ML

- The problem involves **complex patterns** difficult to model manually
- Presence of **large quantities of historical data** that can be exploited
- Need for **continuous adaptation** to changing conditions
- Business rules are **too numerous or vague** to be coded
- The solution must **evolve and improve** over time

Alternatives to consider

- Simple and deterministic business rules
- Classic statistical analyses
- Physical or engineering models
- Expert systems based on heuristics

Document not to be used for teaching. All rights reserved by the author, Dr. Mohamed ASSELLAOU.



MI problem inputs and outputs

Document not to be used for teaching. All rights reserved by the author, Dr. Mohamed ASSELLAOU.

Typical inputs

- lot sensors
- Production history
- Equipment parameters
- Weather conditions
- Maintenance data

Transformation

The ML model analyzes patterns, identifies **correlations**, and learns **complex relationships** between input variables to **generate accurate and actionable predictions**.

Expected outputs

- Equipment failure predictions
- Trajectory optimization
- Safety alerts
- Operational recommendations

Data location and accessibility

Operational databases

ERP, maintenance management systems, SQL/NOSQL databases containing operation history and real-time telemetry.

Cloud storage and data lakes

Historical archives, unstructured data, application logs, distributed backups on aws, azure or private infrastructure.

Field equipment and sensors

Locally collected IoT data, edge devices, embedded systems on vehicles or surveillance cameras.

Documents and files

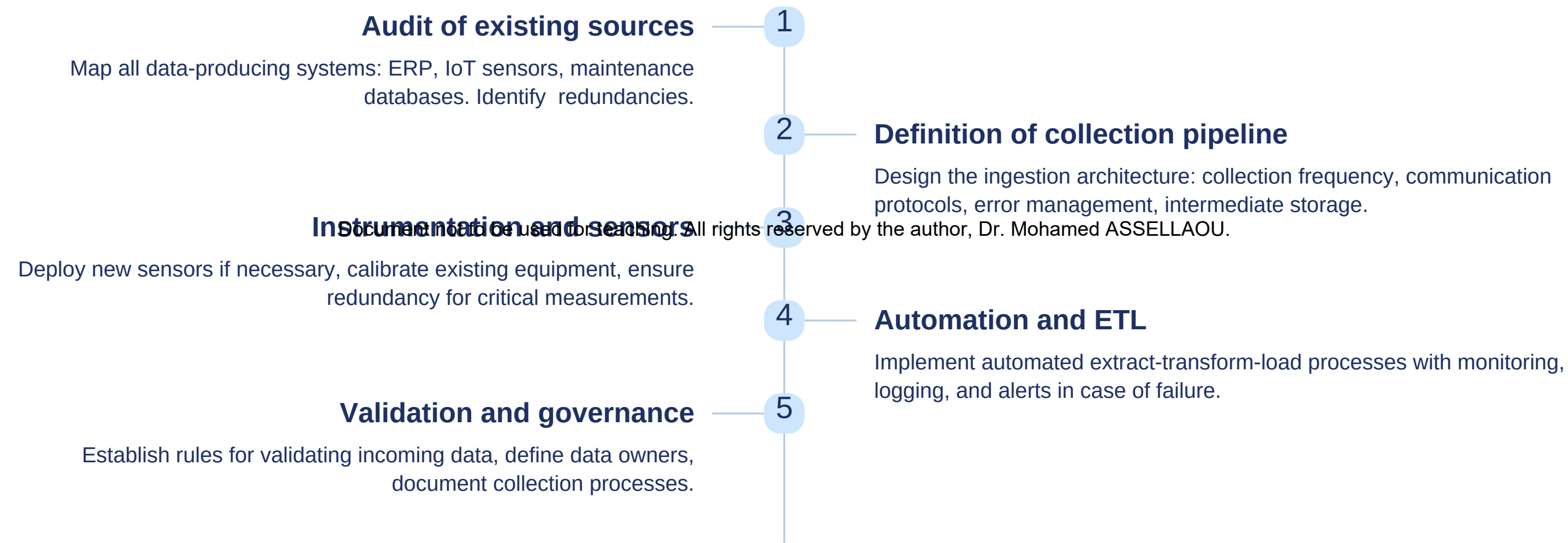
Excel reports, pdf documents, text logs, manual logbooks, surveys that require extraction and structuring.

External sources

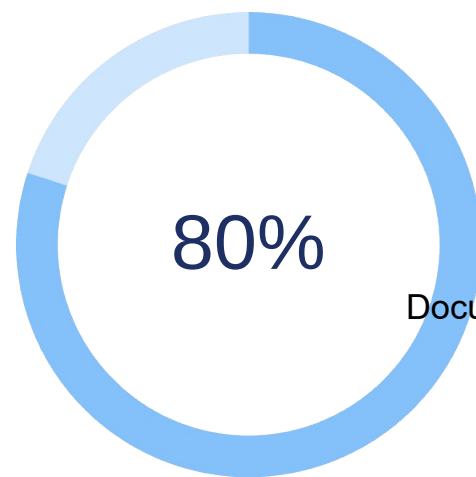
Third-party data, regulations, industry benchmarks, public data.

Document not to be used for teaching. All rights reserved by the author, Dr. Mohamed ASSELLAOU.

Data collection strategies

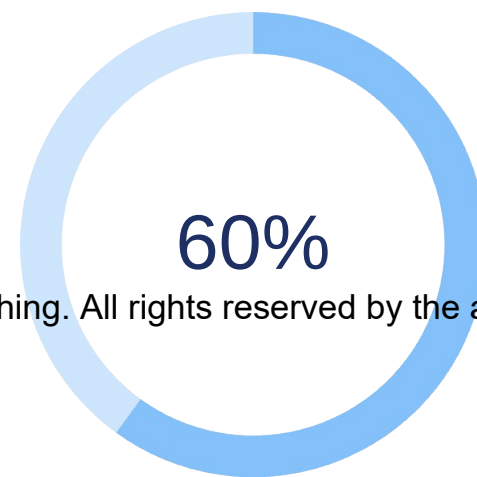


Data readiness status



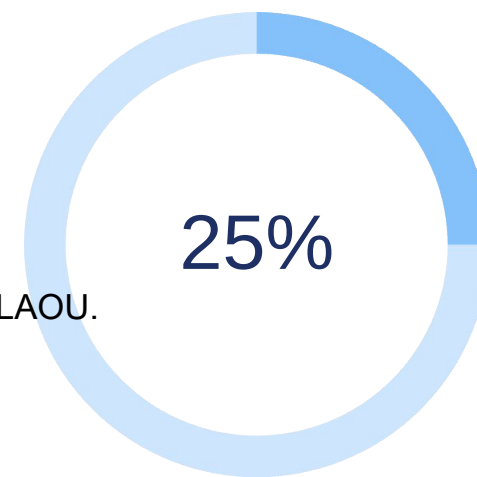
Preparation time

On average, 80% of a ML project's time is dedicated to data collection, cleaning, and preparation



Delayed projects

More than 60% of ML projects experience delays due to insufficient quality of initial data



Failure rate

Approximately 25% of ML initiatives fail due to incomplete or unusable data

Document not to be used for teaching. All rights reserved by the author, Dr. Mohamed ASSELLAOU.

Data readiness

Data completeness

Verify the absence of critical missing values, sufficient temporal coverage, and representativeness of all operational scenarios.

Consistency

Identify outliers, anomalies, sensor errors, duplicates, and inconsistencies between different data sources.

Format and accessibility

Evaluate format homogeneity, schema documentation, access rights.

Data preparation and transformation

01	02	03
Cleaning and filtering	Normalization and standardization	Temporal aggregation
Remove outliers, correct sensor errors, interpolate missing values, eliminate noise and corrupted data.	Scale numerical variables, encode categorical variables, harmonize units of measurement, manage range differences.	Create relevant temporal windows, calculate moving averages, identify trends.
04	05	
Enrichment and derivation	Structuring and labeling	
Generate calculated features (ratios, deltas), integrate contextual data.	Create labels for supervised learning, document transformation traceability, and organize train/validation/test datasets.	

Data preparation and transformation

Recommended tools

- Apache spark
- Python (pandas, polars, numpy)

Best practices

- Reproducible pipeline
- Data versioning
- Automated quality tests

Pitfalls to avoid

- Data leakage
- Information loss
- Introduced bias

Document not to be used for teaching. All rights reserved by the author, Dr. Mohamed ABDELMOU

Feature selection for the model

Statistical methods

Correlation: eliminate redundant variables

Variance: remove constant features

Statistical tests: ANOVA, chi-square

Algorithmic approaches

Feature importance: random forest, XGBoost

L1 regularization: automatic selection

PCA: dimensionality reduction

Domain expertise

Field knowledge: critical variables identified

Physical causality: mechanical relationships

Operational constraints: real-time availability

Document not to be used for teaching. All rights reserved by the author, Dr. Mohamed ASSELLAOUI

Machine learning landscape

Supervised learning

Uses historical labeled data to learn models and make predictions on new data.

- Ideal for problems where outcomes are known.

Unsupervised learning

Discovers hidden patterns in data without labeled examples.

- Perfect for exploratory analysis and anomaly detection.

Reinforcement learning

Learns optimal actions through trial and error in dynamic environments, maximizing rewards over time.

- Ideal for complex decision-making problems.

Document not to be used for teaching. All rights reserved by the author, Dr. Mohamed ASSELLAOU.



Supervised learning

Document not to be used for teaching. All rights reserved by the author, Dr. Mohamed ASSELLAOU.

Predictive maintenance: historical failure data predicts future equipment problems

Production planning: past performance data predicts future phosphate production

Site monitoring (safety and security): Detection of incidents

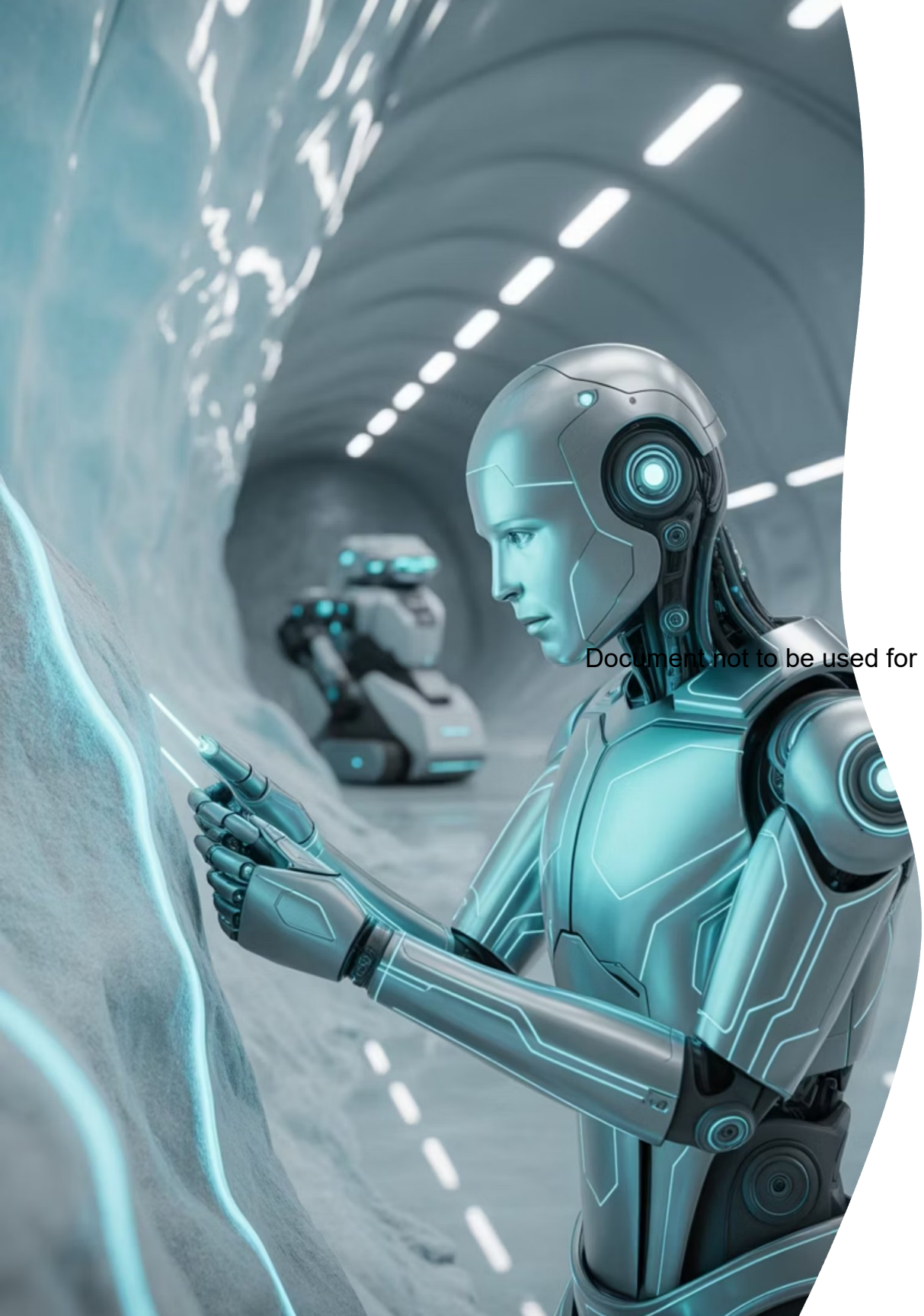
Unsupervised learning

Document not to be used for teaching. All rights reserved by the author, Dr. Mohamed ASSELLAOU.

Clustering

Anomaly detection

Dimensionality reduction

A futuristic robot with a metallic, blue-tinted body and glowing blue eyes is standing in a long, curved tunnel. The robot is holding a small, glowing blue object in its right hand. In the background, another similar robot is visible, and the tunnel is illuminated by a series of bright, rectangular lights on the ceiling.

Reinforcement learning

Agent and environment

An agent performs actions that modify the state of the environment and generate rewards.

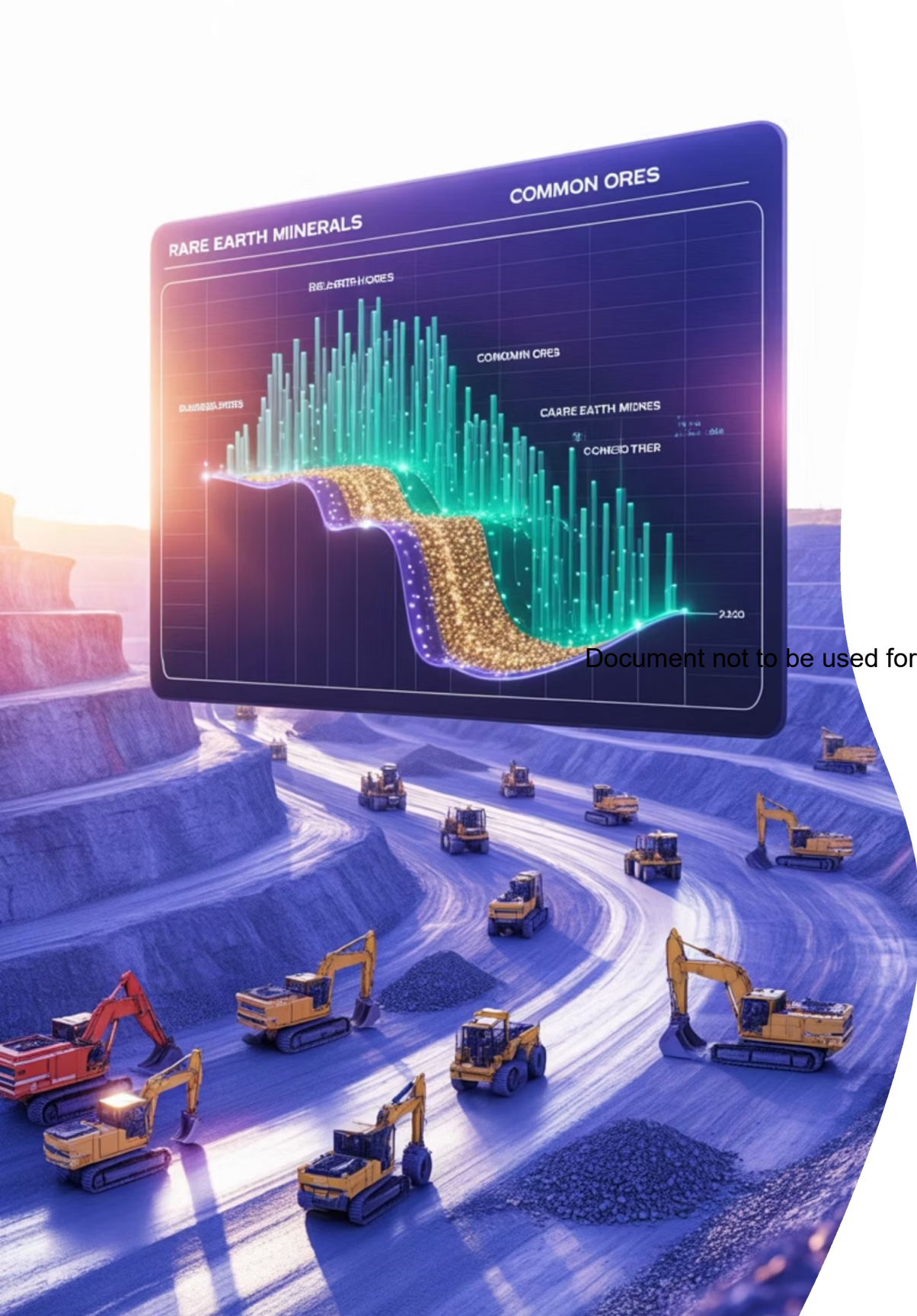
Learning policy

The agent aims to develop a policy (actions for each state) that maximizes the sum of long-term rewards.

Document not to be used for teaching. All rights reserved by the author, Dr. Mohamed ASSELLAOU.

Applications

- Autonomous machine trajectories: optimization of paths to reduce fuel consumption, wear, and cycle times.
- Dynamic inventory management
- Operator training in simulated environments.



Classification

01

Binary classification

Two possible classes or categories, such as equipment failure prediction, security risk detection, or quality assessment (pass/fail).

02

Multi-class classification

Multiple possible categories, such as equipment states (normal, warning, critical).

Regression

01

Linear regression

Simple relationships

- Predicting tonnage based on operating hours
- Energy consumption vs. production rate
- Forecasting basic maintenance costs

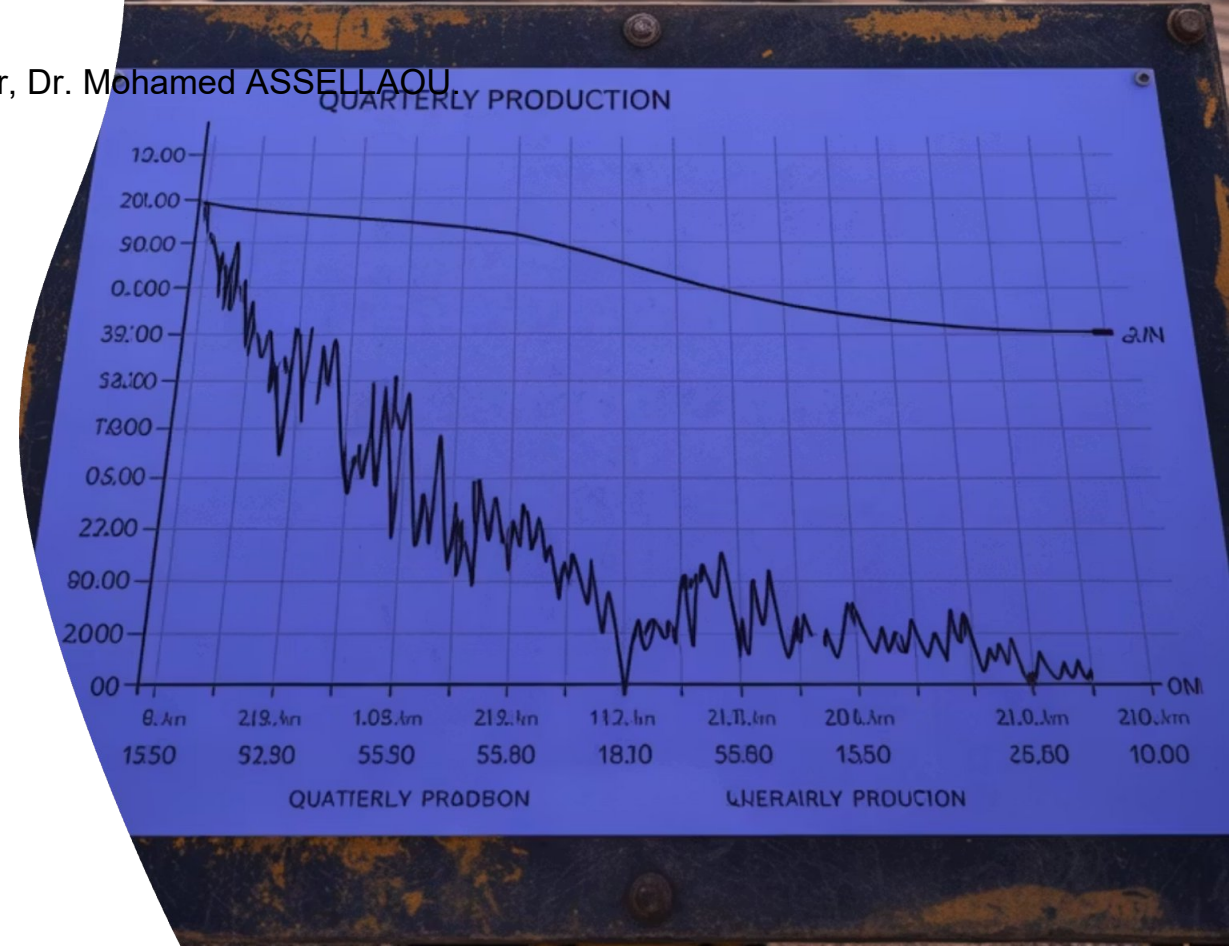
Document not to be used for teaching. All rights reserved by the author, Dr. Mohamed ASSELLAOU

02

Non-linear regression

Complex relationships

- Equipment degradation curves
- Multi-variable production optimization



Translating to ML problems

Classification problems

Problem: equipment failure prediction

ML type: binary/multi-class classification

Document not to be used for teaching. All rights reserved by the author, Dr. Mohamed ASSELLAOU.

Output: failure/no failure or specific failure type

Example: will this conveyor fail in the next 48 hours?

Regression problems

Problem: phosphate production forecasting

ML type: linear/non-linear regression

Output: continuous numeric values

Example: how many tons of phosphate will be processed tomorrow?



Classification performance metrics

Accuracy

Overall correctness of predictions across all classes. Simple but can be misleading with imbalanced datasets.

Precision

Among all positive predictions, how many were actually correct. Essential when false

Document not to be used for teaching. All rights are reserved by the author, Dr. Mohamed ASSELLAOU.

Recall

Among all actual positive cases, how many did we correctly identify. Crucial when missing cases is dangerous.

F1-score

Harmonic mean of precision and recall. Balances both metrics for a comprehensive evaluation.

Regression model performance



Mean absolute error (MAE): average absolute difference between predictions and actual values

Root mean squared error (RMSE): square root of the average squared errors, penalizing larger errors

Mean absolute percentage error (MAPE): percentage-based error measure for better interpretability

R-squared (R^2): proportion of variance explained by the model

Measuring business impact

2.5M

Annual cost savings

15%

Efficiency improvement

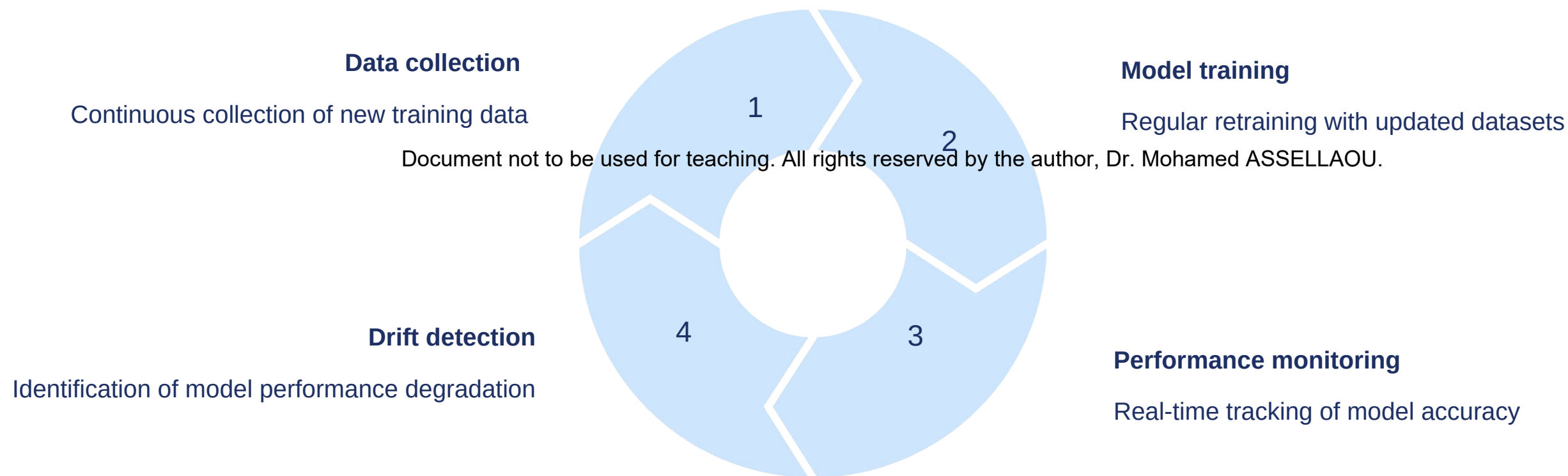
350%

Return on investment

Document not to be used for teaching. All rights reserved by the author, Dr. Mohamed ASSELLAOU.



Model validation & continuous monitoring



Overview of use cases

Document not to be used for teaching. All rights reserved by the author, Dr. Mohamed ASSELLAOU.



Predictive maintenance



Production prediction



Site surveillance

Translation of business use cases

Business problem	Expected outcome	AI approach
1- Equipment failure and unplanned downtime	25 to 30% reduction in downtime	Predictive maintenance using sensor data
2- Inaccurate production forecasts	Improved planning accuracy by 20%	Regression models with historical data
3- Safety incidents	Significant reduction in incidents	Real-time detection models

AI workflow

01

Data collection

02

Data processing

03

Model development

04

Deployment

05

Monitoring

Document not to be used for teaching. All rights reserved by the author, Dr. Mohamed ASSELLAOU.

Question 1: What are the business requirements?

Equipment availability

Reduce unplanned downtime by 40%.

Prediction window

Predict failures 7-14 days in advance.

System performance

85%+ detection, <15% false alerts.

Operational coverage

Monitor 50+ conveyors across site.

Document not to be used for teaching. All rights reserved by the author, Dr. Mohamed ASSELLAOU.

Operational needs

- SCADA integration
- Real-time SMS/email alerts
- Intuitive UI for technicians
- 20% maintenance cost reduction

Environmental constraints

- Dusty, corrosive environment
- Extreme temperatures (-5°C to 45°C)
- Intermittent network connectivity

Question 2: Is ML the best solution?

Approach	Pros	Cons	Verdict
Reactive	Low initial cost, simple	Costly, unpredictable breakdowns	✗
Preventive	Easy planning, fixed schedule	Over-maintenance, high costs	✗
Expert Rules	Simple to implement, explainable	Rigid, lacks real complexity	✗
Machine Learning	Adaptive, detects hidden patterns	Needs data & tech expertise	✓

Why ML is optimal

Complex interactions

Highly non-linear relationships are impossible to model manually with precision.

Continuous learning

System improves automatically with new operational data over time.

Inputs & outputs

INPUTS

Real-time sensors

- Vibrations
- Temperature
- Rotation Speed

Contextual data

- Load transported
- Operating hours
- Maintenance history
- Weather conditions

OUTPUTS

01

Binary classification

Failure within 7 days: Yes/No

02

Multi-class classification

Document not to be used for teaching. All rights reserved by the author, Dr. Mohamed ASSELLAOU.

Anticipated failure type: Motor, bearing, belt

03

Time regression

Remaining Useful Life (RUL) in hours

04

Risk score

0-100 evaluation for prioritizing maintenance

Data location and architecture

SCADA system

Real-time database: PostgreSQL/TimescaleDB.
Generates 500k+ data points/day. 3 years of history available.

CMMS (SAP PM)

Computerized maintenance management system. 5 years of structured intervention tickets and text notes.

Additional IoT sensors

Local edge storage with periodic cloud synchronization.

Data lake

Azure Data Lake / AWS S3 for all raw, unstructured data.

Data warehouse

Snowflake for structured data and complex analytical queries.

Document not to be used for teaching. All rights reserved by the author, Dr. Mohamed ASSELLAOU.

Data collection strategy

Pipeline acquisition & labeling

Document not to be used for teaching. All rights reserved by the author, Dr. Mohamed ASSELLAOU.

Acquisition: IoT sensors SCADA, CMMS collect raw data.

Key data: machine ID, timestamp, temperature, vibration, load, max capacity, pressure, operating hours.

Validation: verification of data quality and schemas.

Labeling: each data point is labeled (0 = normal, 1 = fault) if critical thresholds are exceeded.

Key considerations

Minimal latency: local data buffering to manage connection losses.

Validation & labeling: filtering of erroneous data and rapid identification of faults in real-time.

Enhanced security: encryption to protect data integrity and confidentiality.

Rich metadata: integration of machine identifiers and GPS locations for analysis.

Data preparation and quality

Identified problems

Missing values: 15-20% (sensor failures, network outages)

Outliers and noise: 20%

Class imbalance: 98% normal, 2% failure

Applied solutions

Imputation: linear interpolation or forward fill for short gaps

Anomaly detection: isolation forest to identify outliers

Resampling: SMOTE to balance classes (normal/failure)

Document not to be used for teaching. All rights reserved by the author, Dr. Mohamed ASSELLAOU.

Data transformation and structuring

1 Temporal aggregation

10 min windows: mean, standard deviation, min/max, median.

2 Derived features

Rate of change, ratios (temp/load), deviation from normal, trends (1h/24h).

3 Contextual encoding

Variables: machine type, shift, weather, operator.

4 Target labeling

Labeled failures: "critical" (<72h), "attention" (72-168h), "normal" (>168h).

Document not to be used for teaching. All rights reserved by the author, Dr. Mohamed ASSELLAOU.



Feature selection

Selection methodology

Reduction from 180 initial features to 20 to optimize models and improve interpretability.

Document not to be used for teaching. All rights reserved by the author, Dr. Mohamed ASSELLAOU.

Applied techniques:

Correlation analysis: eliminate highly correlated features (>0.9).

Random forest importance: identify the most contributing features.

Domain expertise: validation by maintenance engineers.

Maintenance model architecture

01

Data pre-processing

Clean sensor data, manage missing values, and normalize measurements for different equipment types and operating conditions.

02

Feature engineering

Document not to be used for teaching. All rights reserved by the author, Dr. Mohamed ASSELLAOU.
Extract moving averages, trend indicators, frequency domain features, and statistical measures from raw sensor streams.

03

Anomaly detection

Identify unusual equipment behaviors using unsupervised learning techniques.

04

Classification model

Train supervised models to classify equipment status as healthy, or requiring immediate attention, based on historical labeled data.

Measuring ML solution success

ML metrics

- recall $\geq$ 85%
- precision $\geq$ 70%
- AUC-ROC > 0.90

Business KPIs

Availability rate : 95% target

Maintenance cost : -30% goal

MTBF : +40%

ROI : 18 months

Continuous monitoring

- Drift detection
- A/B testing
- Feedback loop

Maintenance: from theory to practice

Approach summary

Business understanding

Align needs with objectives (ROI).

Data quality

Robust pipelines, relevant features.

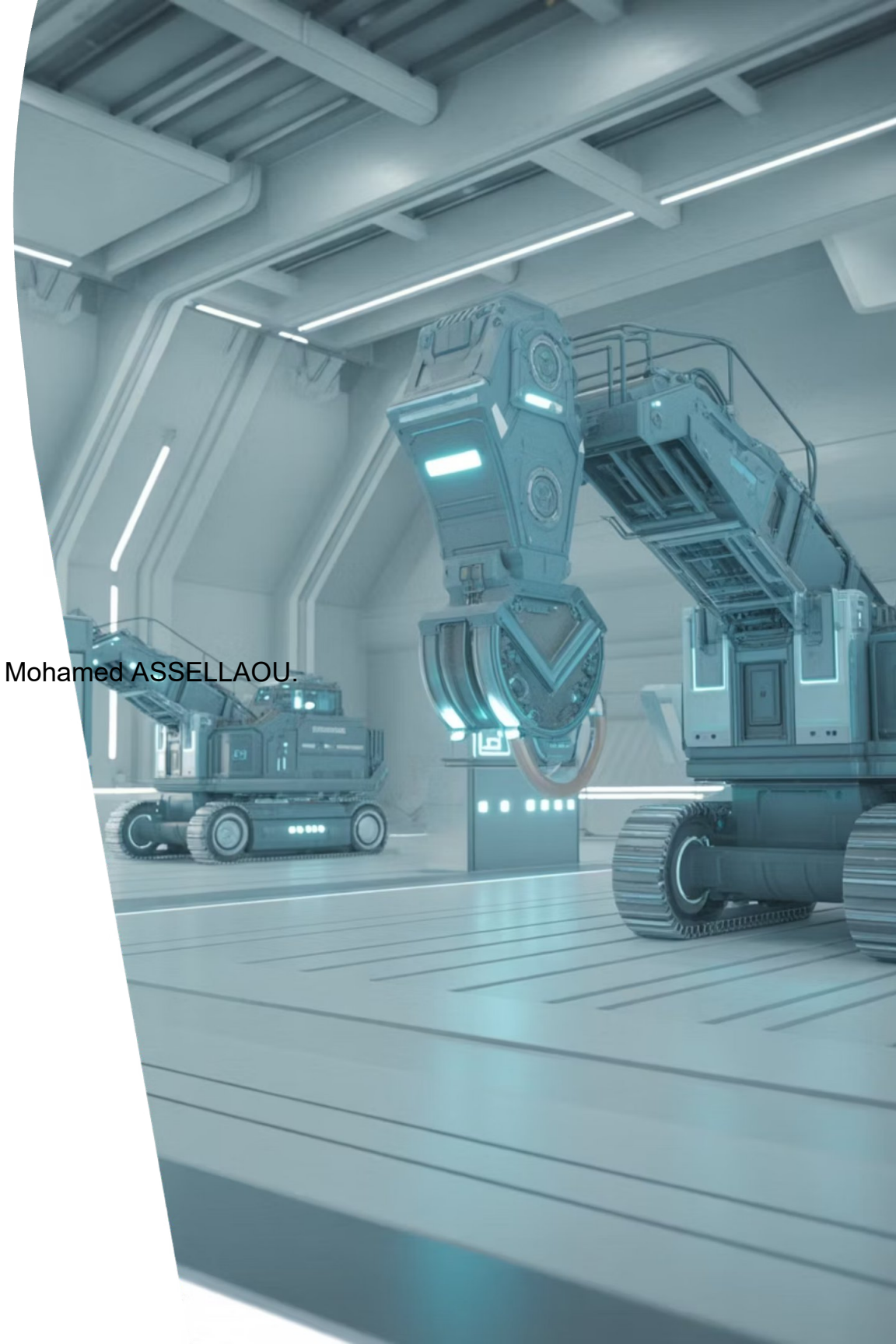
Impact measurement

Continuous KPI monitoring (tech & business).

Next steps

1. PoC on 5 pilot machines
2. Validation (3 months of operation)
3. Phased deployment on 50 pieces of equipment
4. CMMS and ERP integration
5. Expansion to other sites

Document not to be used for teaching. All rights reserved by the author, Dr. Mohamed ASSELLAOU.



Key principles of responsible AI in Manufacturing

01 - Ethical foundation

Establish clear ethical guidelines and accountability frameworks for the development and deployment of AI across all mining operations.

02 - Data security

Implement robust data governance, protection protocols, and privacy safeguards to maintain stakeholder trust and regulatory compliance.

03 - Bias prevention

Actively detect, correct, and prevent algorithmic biases through diversified datasets, regular audits, and inclusive development practices.

04 - Transparency

Ensure AI systems are explainable, interpretable, and provide clear justification for recommendations and decisions.

05 - Human-centered design

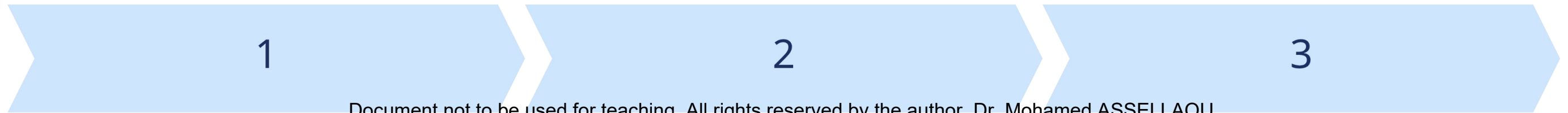
Develop AI solutions that augment human capabilities while preserving meaningful human oversight and decision-making authority.

06 - Continuous improvement

Regularly monitor AI performance and implement feedback mechanisms for ongoing optimization and ethical compliance.

Document not to be used for teaching. All rights reserved by the author, Dr. Mohamed ASSELLAOU.

The power of explainable AI



Data input

Raw data from mining equipment sensors, geological surveys, and operational metrics

AI processing

Machine learning algorithms analyze patterns and generate predictions or recommendations

Explainable output

Clear reasoning and confidence levels help engineers understand and trust AI recommendations

Défis courants

Unclear objectives

Projects often start without specific and measurable objectives.
'Improving efficiency' is too vague!

– 'Reduce unplanned downtime in phosphate processing by 25%' provides clear direction and success criteria.

Document not to be used for teaching. All rights reserved by the author, Dr. Mohamed ASSELLAOU.

Insufficient data quality

Inconsistent data collection, missing values, and unreliable sensors from phosphate extraction or processing equipment create models that fail in production!

-Data quality assessment must precede model development.

Unrealistic expectations

Expecting 100% accuracy or immediate ROI in predicting phosphate equipment failures leads to disappointment!

-Setting realistic expectations and iterative improvement goals ensures sustainable progress."

Insufficient domain knowledge

Data scientists and ML engineers without expertise in phosphate extraction may miss critical operational nuances!

-Close collaboration between technical and operational teams is essential."

Translating Manufacturing problems into AI models

Document not to be used for teaching. All rights reserved by the author, Dr. Mohamed ASSELLAOU.

Mohamed ASSELLAOU

