AI fundamentals for cement glossary

Artificial Intelligence (AI)

Software that can reason, make decisions or predictions and perform complex tasks.

Machine Learning (ML)

A type of AI where models learn patterns and relationships directly from data, rather than being explicitly programmed with rules. Instead of telling the system *how* to solve the problem, we show it examples - and it figures out what drives performance.

Model

The mathematical engine that turns inputs into predictions. It's the learned representation of how your process behaves — built from historical data. In a cement context, a model is also shaped by physics and constraints, and tuned to a specific objective.

Algorithm

An algorithm is a sequence of steps for how to do something, for example, how to train a model or make a prediction

The term has come to be confused with 'model' in everyday speech, largely due to how people talk about social media feeds. For example, you might hear someone say the Facebook/Twitter/Instagram/TikTok 'algorithm', is showing them certain content, but what they are actually describing is what the platform's model has predicted they'll find interesting. The algorithm is the process; the model is what is doing the predicting.

Deep learning

A subfield of machine learning that uses neural networks with many layers to learn complex relationships from data. The 'deep' bit refers to the number of layers - so the more layers, the deeper the network, and the more sophisticated the patterns it can learn from.

Supervised learning

A type of machine learning where the model is trained using labelled examples, where we know the correct answer. For example training a model to predict free lime based on fuel in the kiln.

Label

In supervised learning, a label is the 'answer' attached to an example in the historical data. If your inputs are kiln temperature, feed rate and airflow, the label might be fuel consumption, clinker quality, or NOx emissions. A model will use this labelled data to learn relationships, to help it make correct predictions when it is exposed to unlabelled data in real life.

Unsupervised learning

A type of machine learning where the model is tuned on unlabelled data. Typically, this is used to group similar bits of data together, like the operating modes of the kiln e.g. is the kiln warming up, cooling down etc.

Reinforcement Learning

A type of machine learning where a model learns how to control something by trial-and-error interactions and improves its behaviour in response to the feedback it receives. For example, it might learn to maintain kiln stability while maximising alternative fuel.

Feature

Also called input feature.

An input variable to a model e.g kiln temperature, feed rate, oxygen levels, or fan speed.

Usually, some kind of engineering has been applied to the raw variables in the data. For example, a raw variable would be precalciner temperature. A feature which is used by a model might be the rolling mean of precalciner temperature over a 5-minute window.

Output Feature

The variable that the model is trained to predict - for example, fuel consumption, clinker quality, or NOx levels. Also commonly referred to as the target variable.

Time series data

A sequence of data points recorded over time, where the order and timing of the observations are critical to understanding the data. Modelling for time series differs from traditional modelling because consecutive data points are correlated. Examples of time series data include financial data (stock markets), meteorological data and IoT and sensor data.

Training a model

Training

Teaching a model by adjusting parameters to improve performance based on data.

Historical data

Past recorded data used to train and evaluate models.

Training set

The portion of historical data used to teach a model.

Test set

A separate portion of data that the model hasn't seen before, used to evaluate performance. If a model performs well here, it's learning real relationships, not just memorising history.

Underfitting

When a model is too simple to capture underlying patterns in the data, resulting in high bias and poor performance on both training and test sets.

Overfitting

When a model learns noise or random fluctuations in the training data rather than the underlying pattern, leading to poor generalisation on unseen data.

Generalisation

A model's ability to perform well on new unseen data, not just the historical data it was trained on.

Parameters

A value the model calibrates during training that represents relationships discovered in the data. Think of it as the tuning dial inside the model. In a linear regression, parameters are the coefficients that define the slope of the line.

Hyperparameter

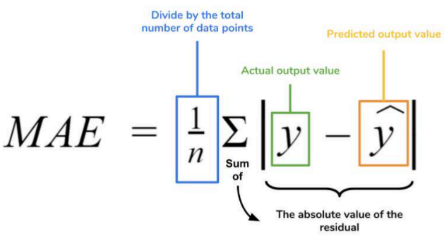
A setting chosen before training begins that defines how the model will learn, e.g., how many trees sit in a random forest, the number of layers in a deep model or even how long the model will train for

Hyperparameter tuning

The process of selecting the best hyperparameters for a model using methods like grid search, random search, or Bayesian optimisation.

Evaluating a model

Mean absolute error (MAE)



The average absolute difference between predictions and actual values. MAE is robust to outliers. The lower the number the better. In plain terms: on average, how close are we to the true value, regardless of direction?

Mean squared error (MSE)

The average of squared differences between predictions and targets. The lower, the better. Because errors are squared, large mistakes are penalised more heavily. Useful when big misses are especially costly.

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2$$

Directionality

Whether the model correctly predicts the *direction* of change, not just the exact value. Useful for situations like: knowing whether fuel consumption will go up or down can matter as much as the precise number.

Naive baseline

Using the previous known value as the benchmark for the model to beat. Also called persistence or the naive forecast. If we didn't have a model, the best guess at the next value is the previous value. If a model can do better than the last value, we know we're getting value from having a model to forecast. *Note: this definition is specific to time series modelling*

Shapley values

A method for explaining model predictions by quantifying how much each feature contributed to the outcome.

Types of models

Linear regression

A model that assumes a straight-line relationship between inputs and outputs.

Non-linear models

Models that can capture curved, interacting, or more complex relationships between variables.

Random forest

An ensemble of decision trees that work together to make predictions.

Ensemble model

A model built by combining multiple models to make better predictions than any single model alone.

Neural network

A model inspired by the structure of the brain, made up of interconnected “nodes” (artificial neurons). Each node processes signals and passes them forwards.

Transformer

A neural network architecture widely used in language models and computer vision.

LLMs (Large language models)

AI models that are trained on vast amounts of text to understand and generate language. They’re great at conversation and summarisation. But optimising a cement kiln? That requires models trained on physics, constraints and plant data — not internet text.

AI Applications in cement

Soft sensor / software sensor

A virtual sensor that estimates the value of some physical variable.

Recommender/optimiser

A type of model which recommends values for one or more setpoints at the plant.

Objective function

In control, this is the goal of the recommender/optimiser, i.e. what it is trying to minimise or maximise when choosing which action to take. It encodes a combination of the overall goal - e.g. maximise thermal substitution rate (TSR) - along with any constraints.