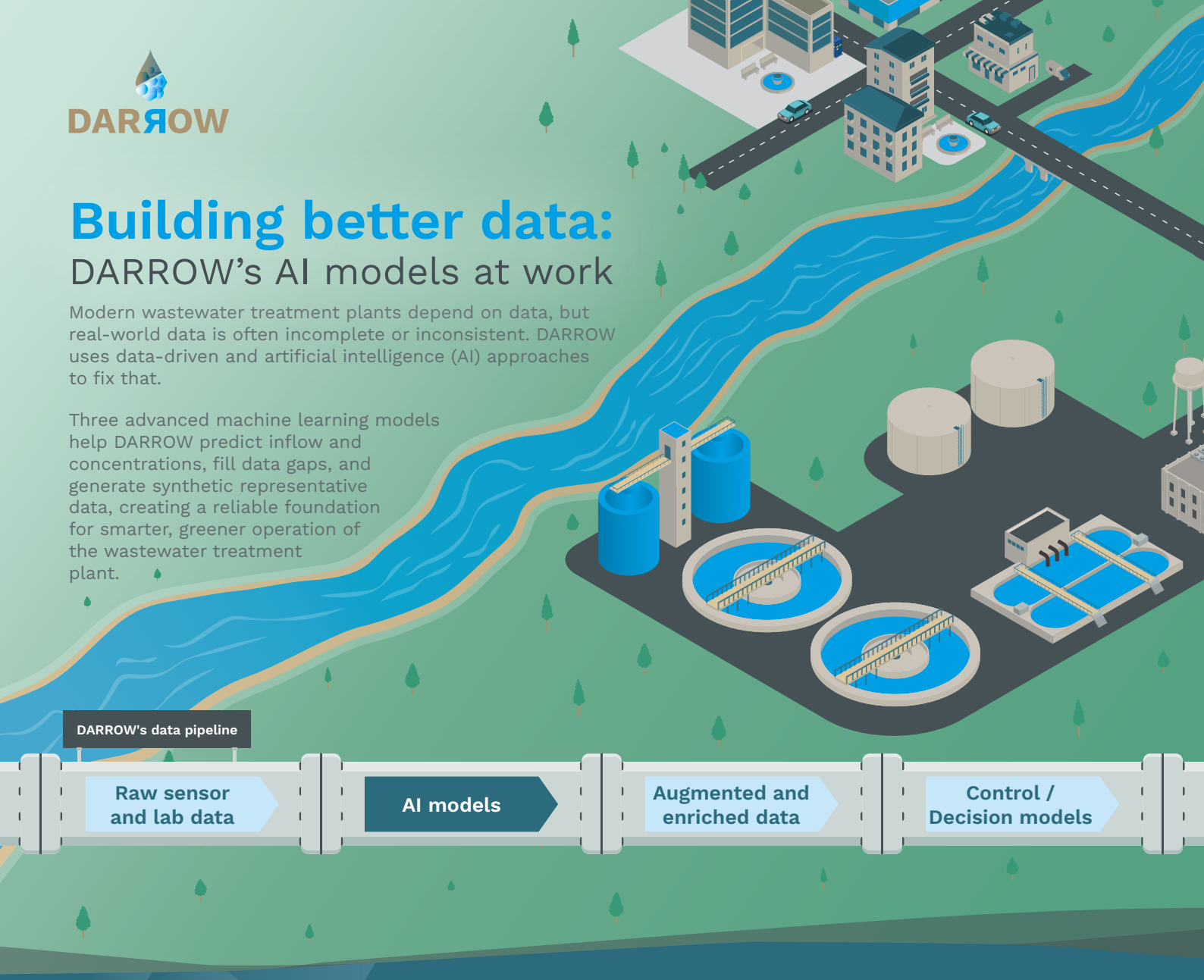
DARROW

# Building better data:

## DARROW's AI models at work

Modern wastewater treatment plants depend on data, but real-world data is often incomplete or inconsistent. DARROW uses data-driven and artificial intelligence (AI) approaches to fix that.

Three advanced machine learning models help DARROW predict inflow and concentrations, fill data gaps, and generate synthetic representative data, creating a reliable foundation for smarter, greener operation of the wastewater treatment plant.



**DARROW's data pipeline**

Raw sensor and lab data → **AI models** → Augmented and enriched data → Control / Decision models

## Inflow prediction model

### What it does

Predict the amount of water arriving at the plant in the next 24 hours.

### Why it matters

Helps the plant prepare for changing volumes due to rain, weather, or daily routines.

### How it works

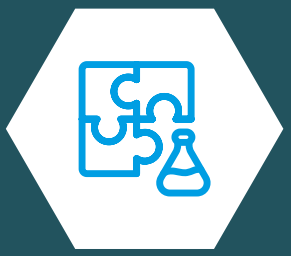
When water level, time, and weather data are available, the model draws on historical patterns to forecast how much water will enter the plant. It uses a Long Short-Term Memory (LSTM) and an Artificial Neural Network (ANN) to predict water levels and inflow volumes, similar to how experienced plant operators anticipate daily surges.

**AI models used:** LSTM (Long Short-Term Memory): A neural network that recognises patterns over time, ideal for time series like inflow data.

ANN (Artificial Neural Network): A flexible model that translates predicted water levels into actual flow volumes based on plant-specific behaviour.

**Output:** Hourly flow forecasts for the next 24 hours.

# Concentration prediction model



## What it does
Predict the quality of the water arriving at the plant in the next 24 hours.

## Why it matters
Helps the plant prepare for changing water quality due to rain, weather, pollution discharges or daily routines.

## How it works
When time, weather data and sensor data on water quality are available, the model draws on historical patterns to forecast the quality of the water that will enter the plant. It uses Artificial Neural Networks (ANN) to predict concentration of several parameters.

**AI model used:** ANN (Artificial Neural Network) in different configurations.

**Output:** Hourly concentration forecasts for the next 24 hours for parameters such as:

- COD (Chemical Oxygen Demand)
- BOD (Biochemical Oxygen Demand)
- $P_{tot}$ (Total Phosphorus)
- $N_{tot}$ (Total nitrogen), $NH_{4+}$ (Ammonium), $NO_{3-}$ (Nitrate)
- pH
- TSS (Total Suspended Solids)

# Lab data gap filler

## What it does
Estimate lab results on days when no samples were taken (e.g. semi-weekly sampling, weekends).

## Why it matters
Ensures continuity for key parameters (e.g. COD, BOD, TOC, Ptot) to support real-time decision-making.

## How it works
When lab results are unavailable, the model uses patterns in the plant's operational data to estimate missing values for key water quality indicators. The model combines three machine learning methods into a robust ensemble model.

**AI models used:** Ensemble of three different models stacked using an elastic net model. This leverages the unique strengths of each individual model, resulting in a more robust and accurate ensemble.

**Output:** Daily estimates for key lab parameters such as:

- COD (Chemical Oxygen Demand
- BOD (Biochemical Oxygen Demand)
- TOC (Total Organic Carbon)
- $P_{tot}$ (Total Phosphorus)
- $PO_4$ (Phosphate)
- $N_{tot}$ (Total nitrogen)
- $N_{Kj}$ (Kjeldahl nitrogen)

## What's next?
These AI models are ready to be deployed at the Tilburg wastewater resource recovery facility, where the DARROW platform will be tested.

Co-funded by the European Union