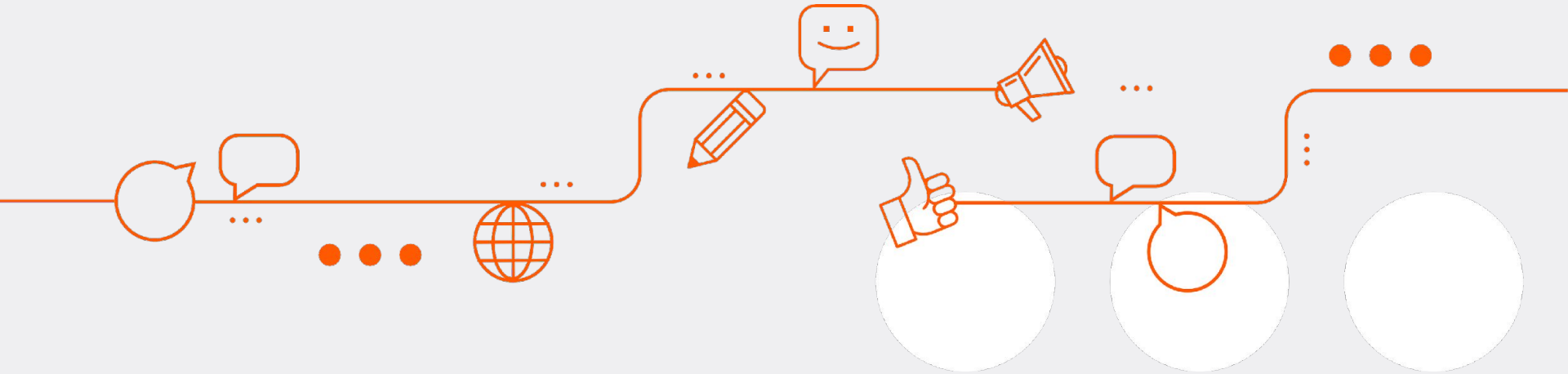
The Future is Now: Embracing Allegro Transactional Data for Risk Modeling

Maciej Wysocki

allegro

ML in PL Conference, 27th October 2023



Impact and scale @ allegro pay.

A brief history of Allegro Pay

Allegro Pay is a safe and convenient payment method for purchases on Allegro

allegro

35m
accounts

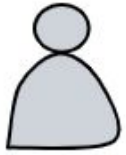
allegro

14.2m
active buyers

09.2021
official launch

allegro pay.

1.6m
onboarded users



PLN 5.5B
financed with Allegro Pay in 2022

Tech stack used at Allegro Pay

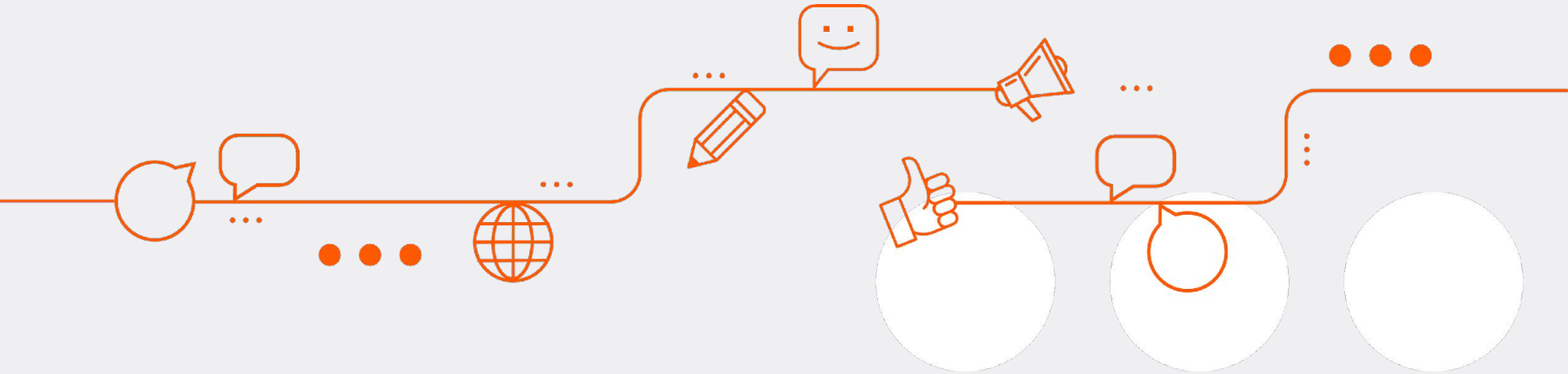
We leverage industry-standard technologies best fitted to our needs

Backend & Cloud

Analytics & Monitoring

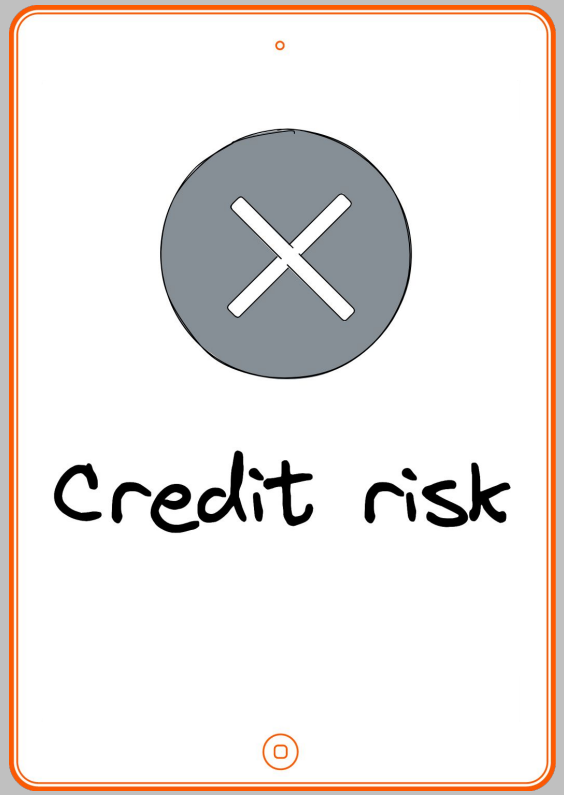
Data Engineering

Data Science



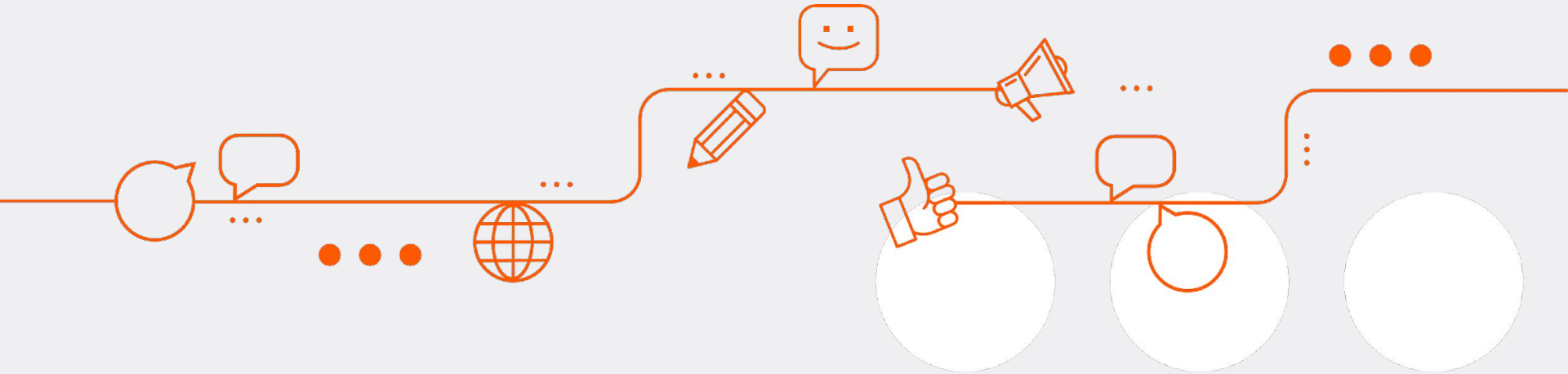
Machine Learning @ allegro pay.

We use Machine Learning models at every step of customer journey!



Credit risk

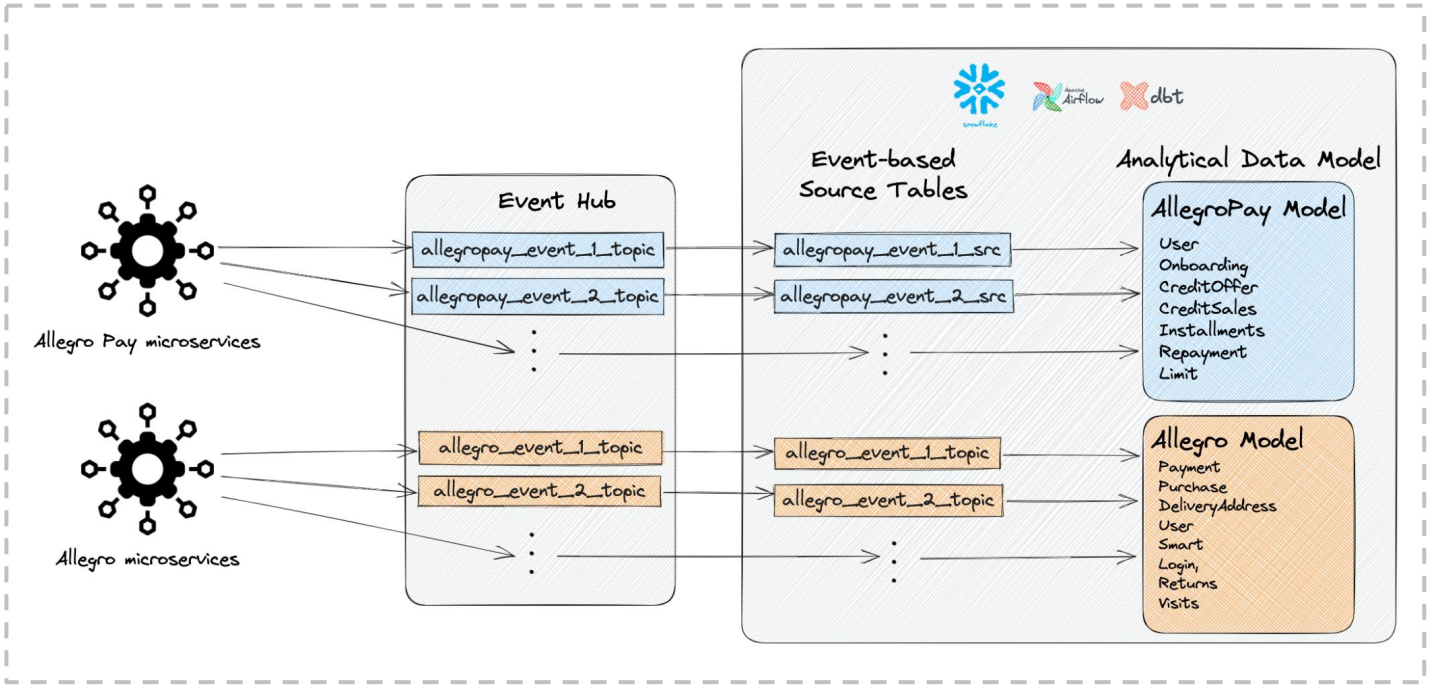
Possibility of incurring a loss due to customers' failure to repay a loan.



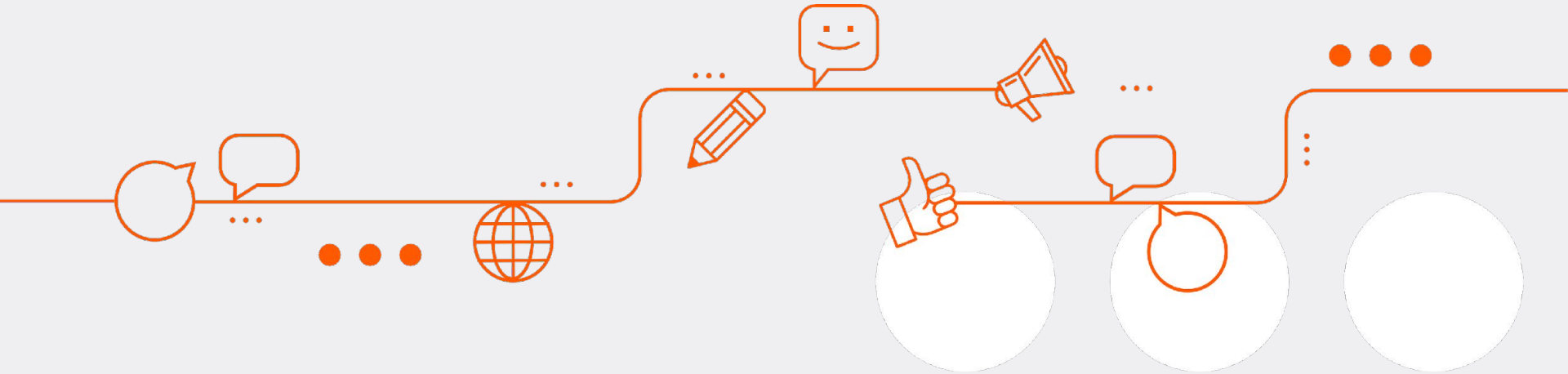
Data and Events @ allegro pay.

Event-driven architecture

A journey from petabytes of raw events to convenient analytical tables



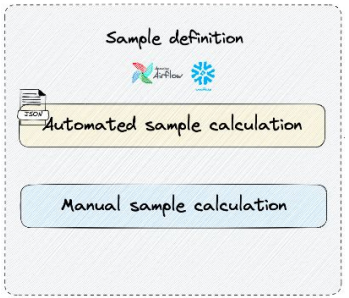
Data Engineers set up pipelines **transforming raw events** emitted by microservices into **meaningful analytical entities**



Model Training @ allegro pay.

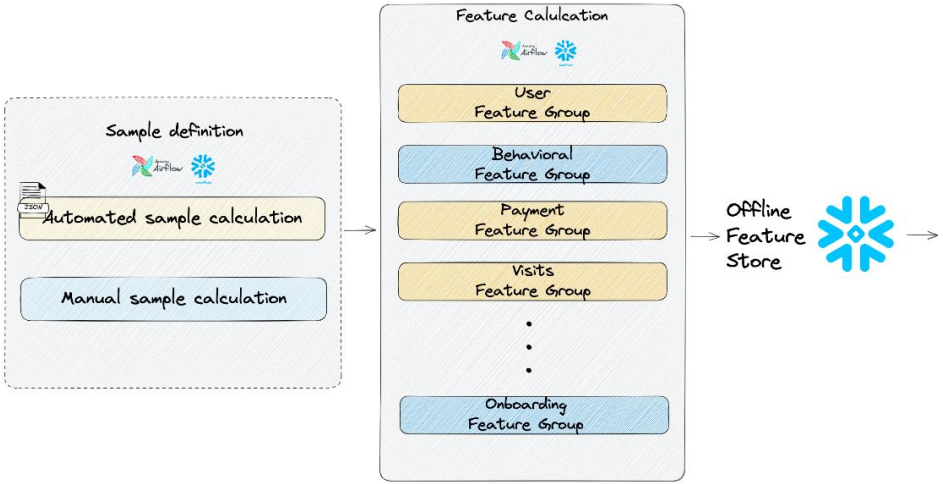
The Modeling Framework

From business hypotheses to a machine learning model in no time



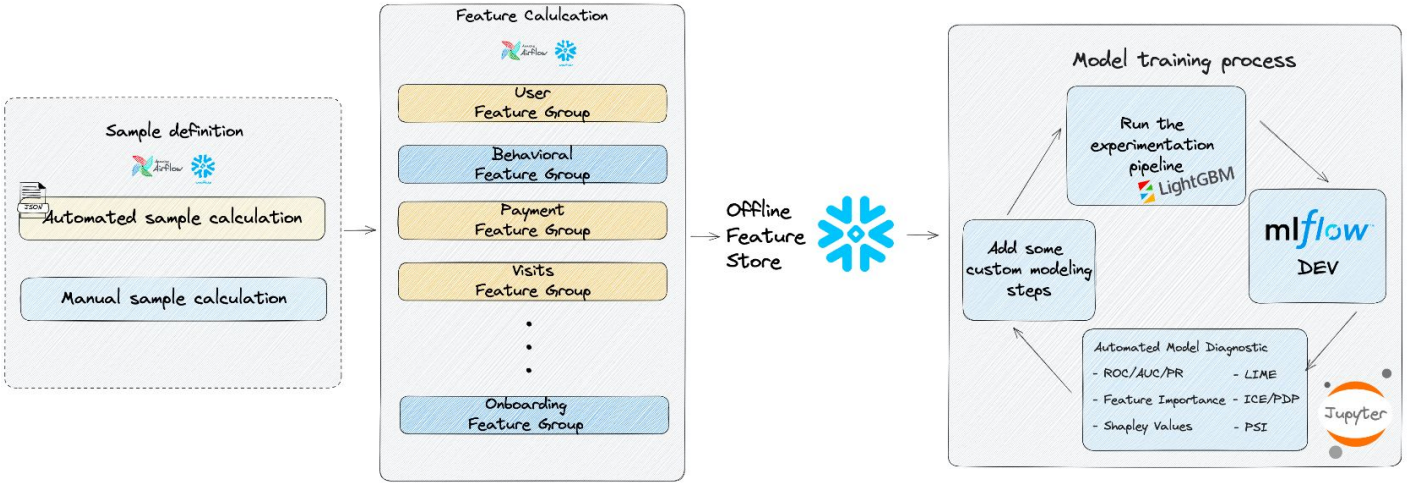
The Modeling Framework

From business hypotheses to a machine learning model in no time



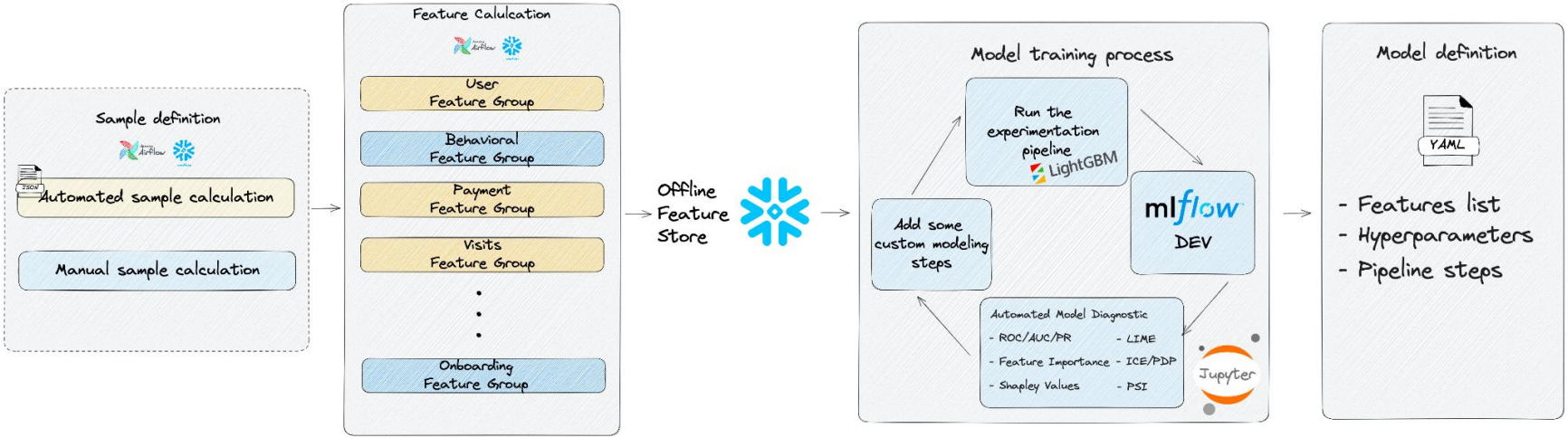
The Modeling Framework

From business hypotheses to a machine learning model in no time



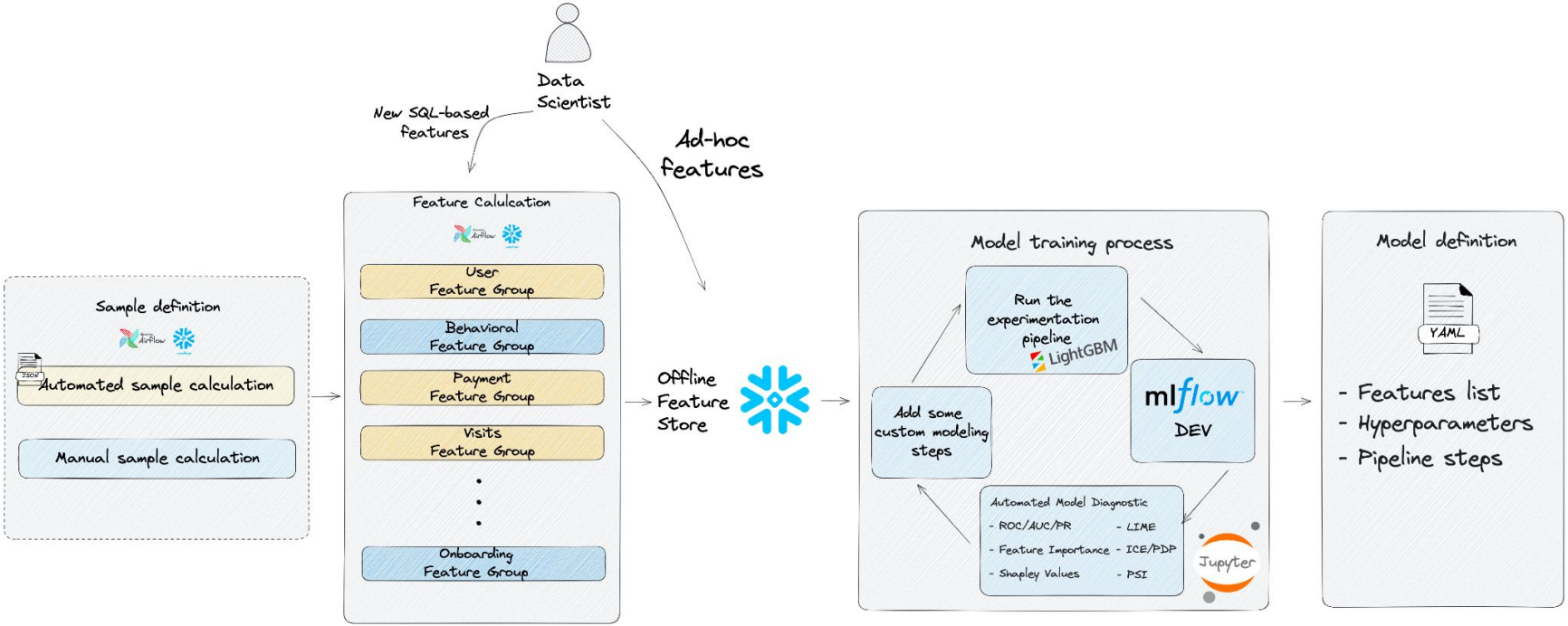
The Modeling Framework

From business hypotheses to a machine learning model in no time



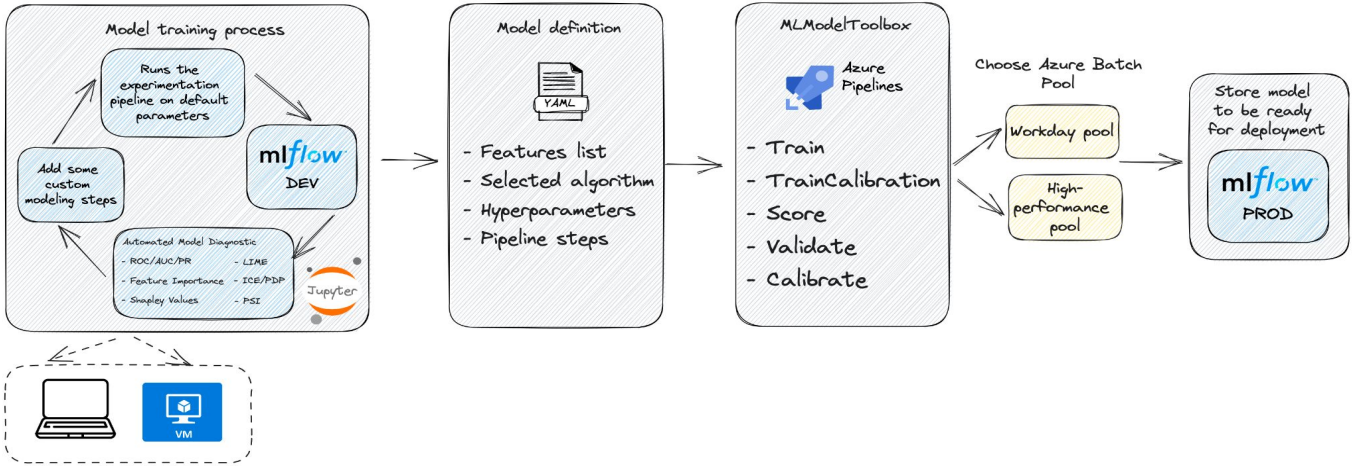
The Modeling Framework

From business hypotheses to a machine learning model in no time

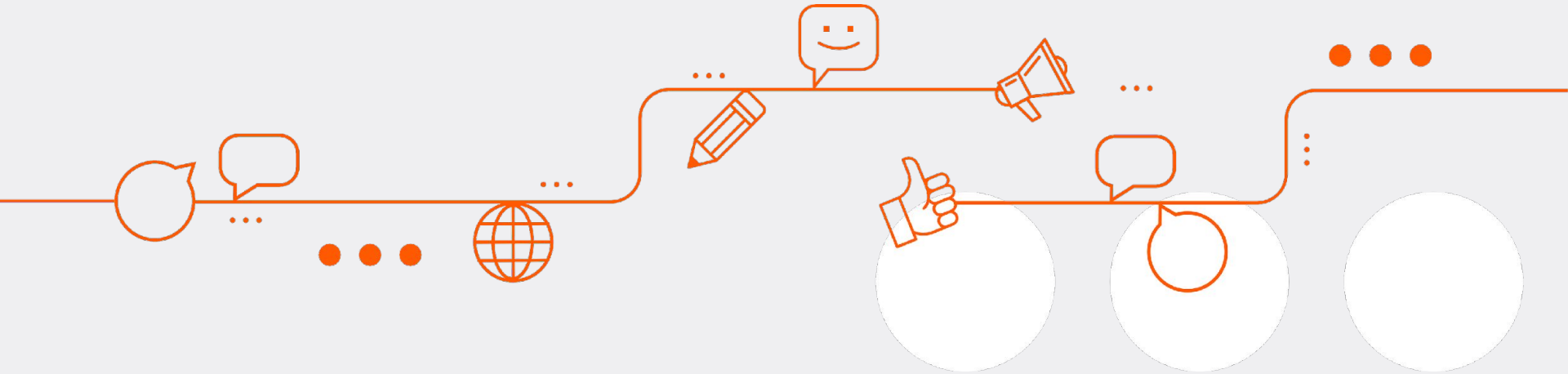


Modeling Framework with Very Little Supervision

Moving from experiments to models ready for deployment



Configurable building blocks make it possible to train a benchmark model within one day with **little to no involvement** from a Data Scientist



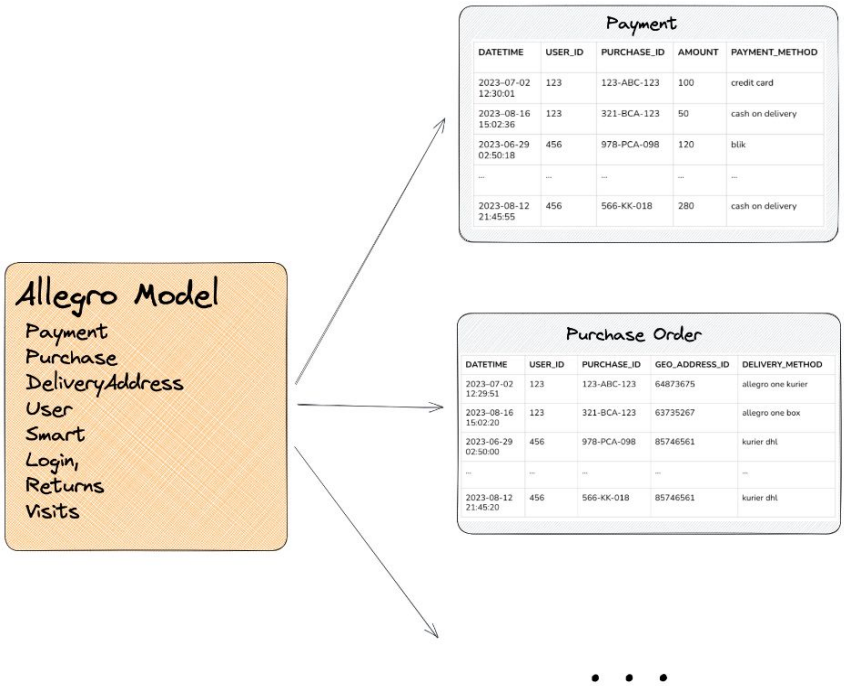
Introducing Time Series

Time Series Data

Introducing sequences of purchases and payments

Time Series Data

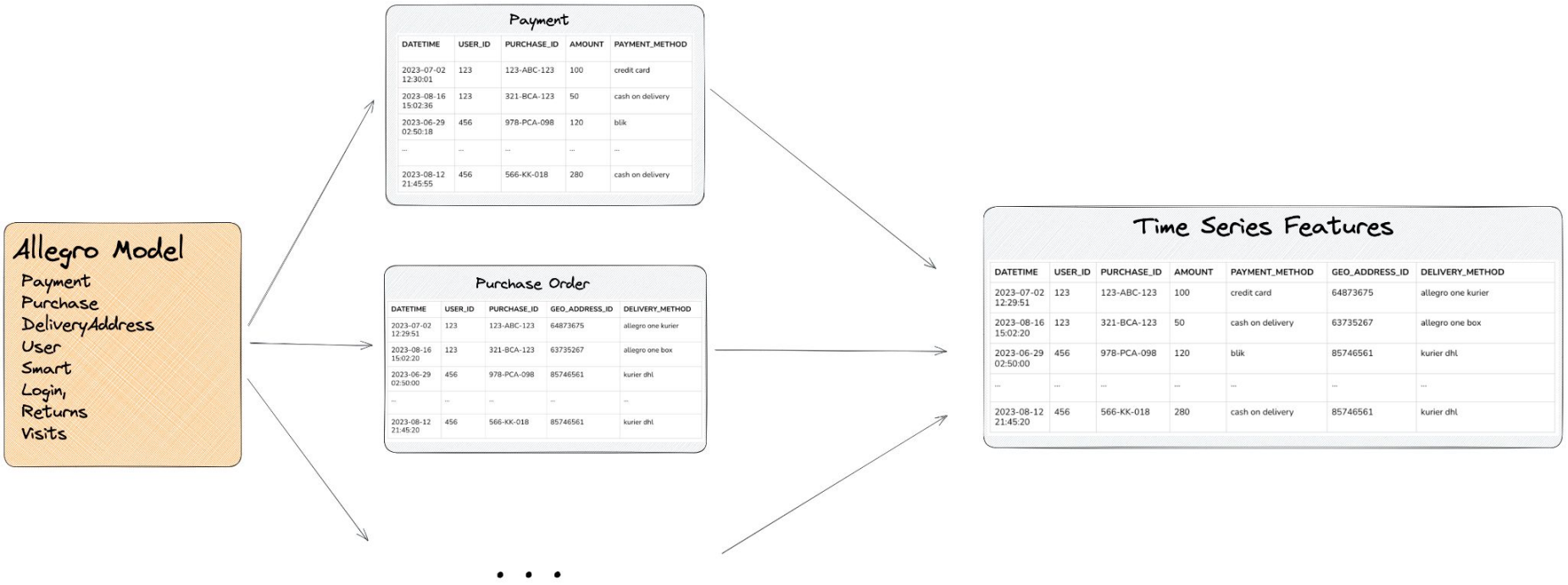
Introducing sequences of purchases and payments

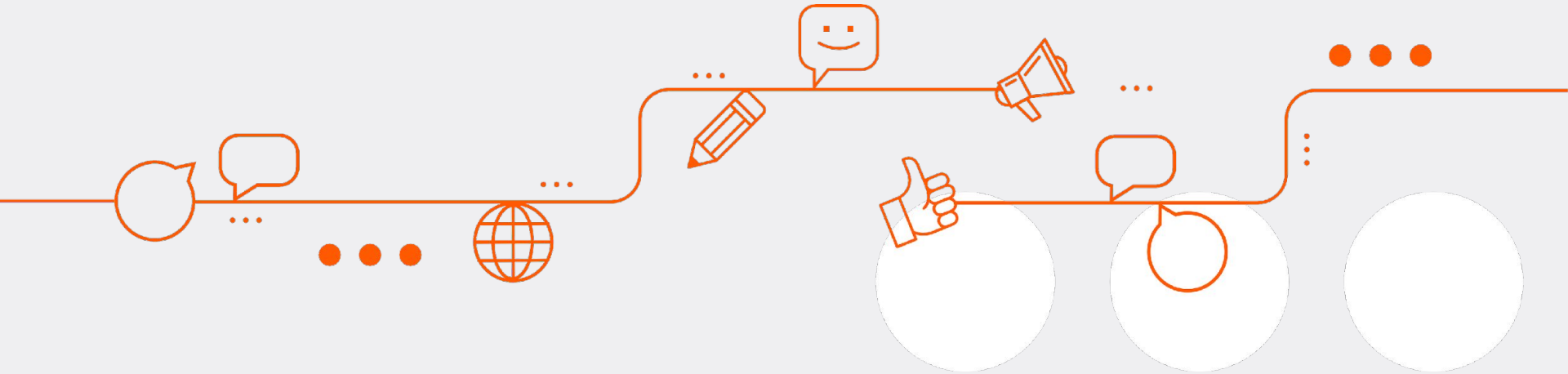


. . .

Time Series Data

Introducing sequences of purchases and payments

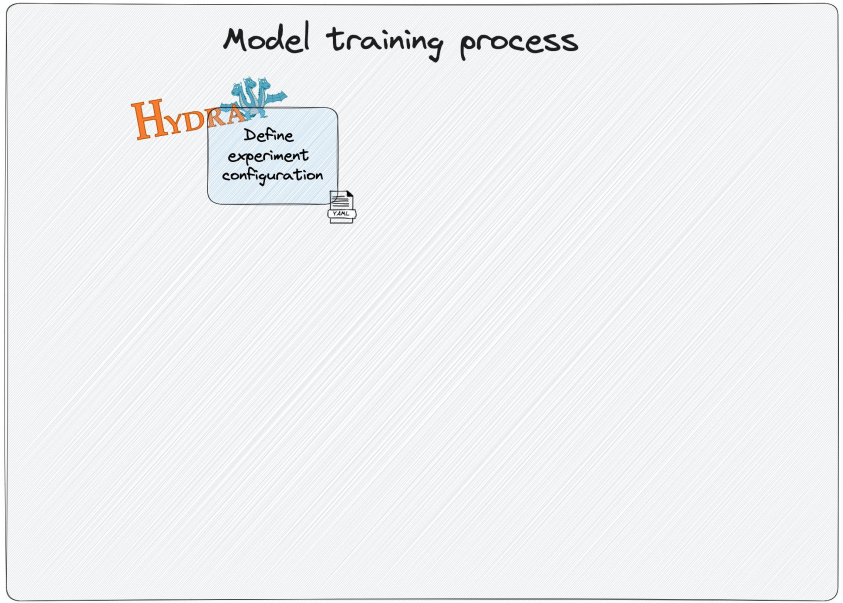




Experimentation Framework

Experimentation Framework

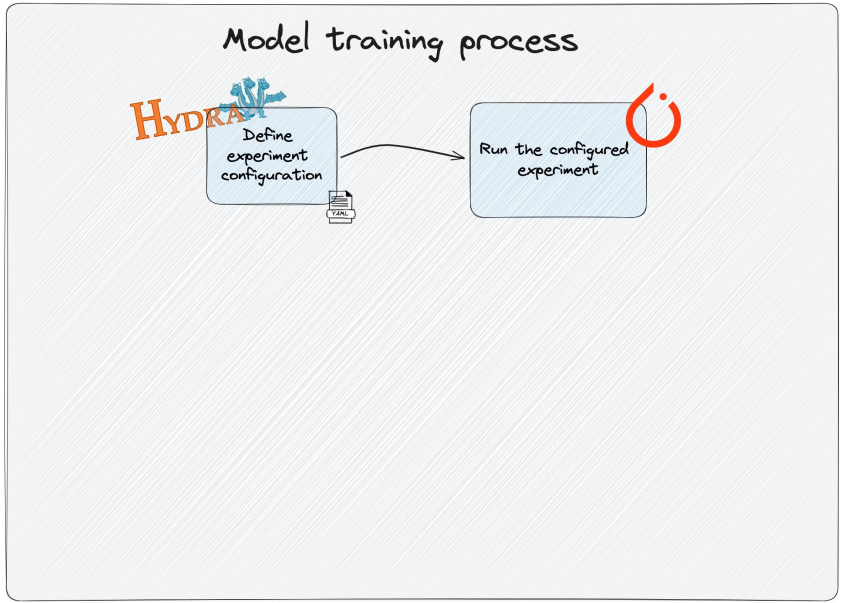
Scalable building blocks for experiments with deep learning models



Configurable pipeline that combines **pre-existing solutions** with a few **new components** suitable for experiments with neural networks

Experimentation Framework

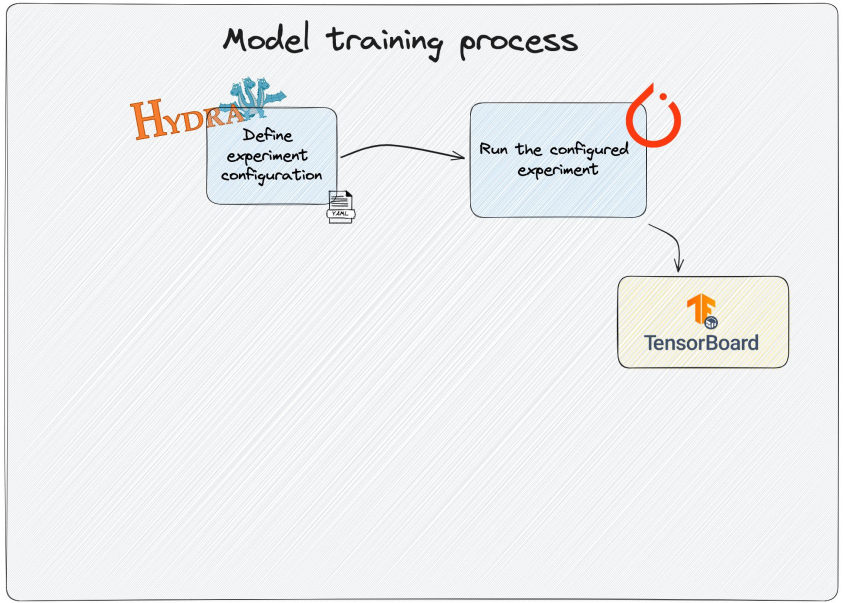
Scalable building blocks for experiments with deep learning models



Configurable pipeline that combines **pre-existing solutions** with a few **new components** suitable for experiments with neural networks

Experimentation Framework

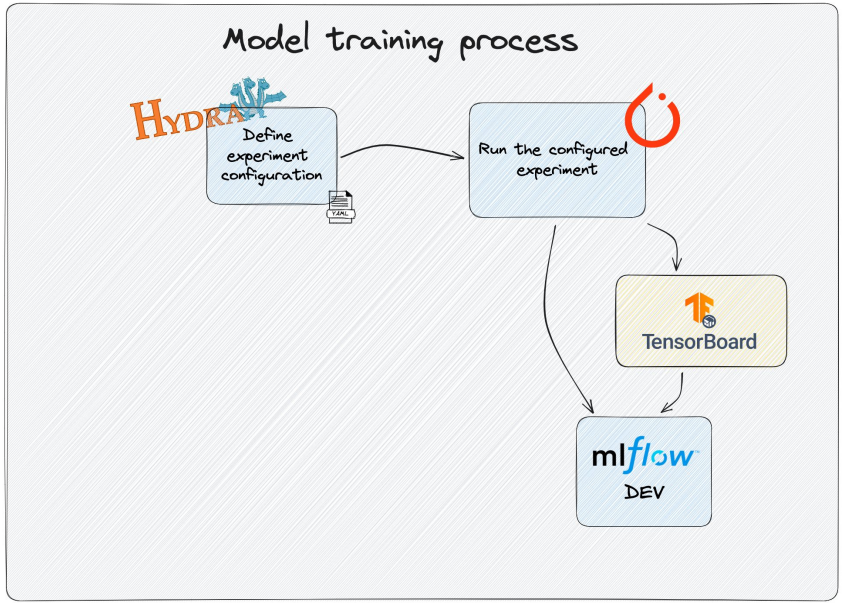
Scalable building blocks for experiments with deep learning models



Configurable pipeline that combines **pre-existing solutions** with a few **new components** suitable for experiments with neural networks

Experimentation Framework

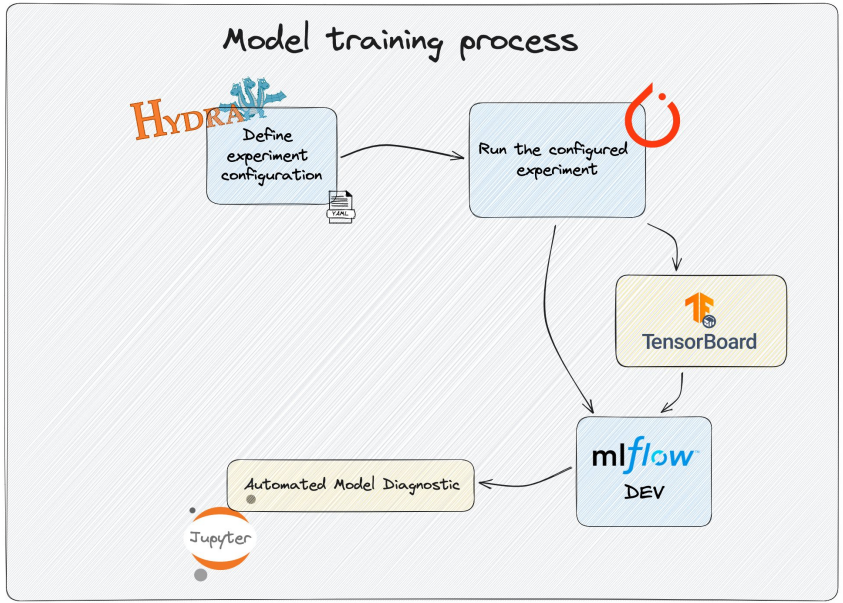
Scalable building blocks for experiments with deep learning models



Configurable pipeline that combines **pre-existing solutions** with a few **new components** suitable for experiments with neural networks

Experimentation Framework

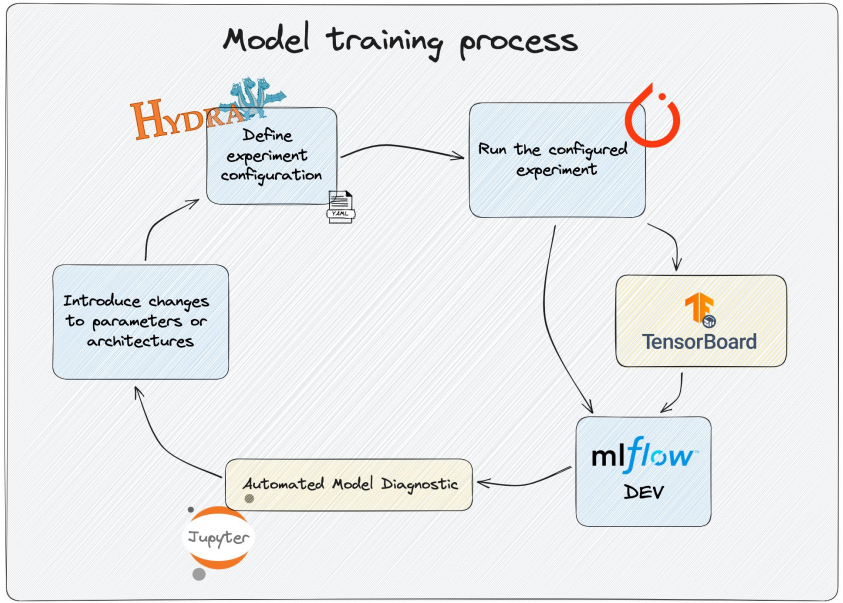
Scalable building blocks for experiments with deep learning models



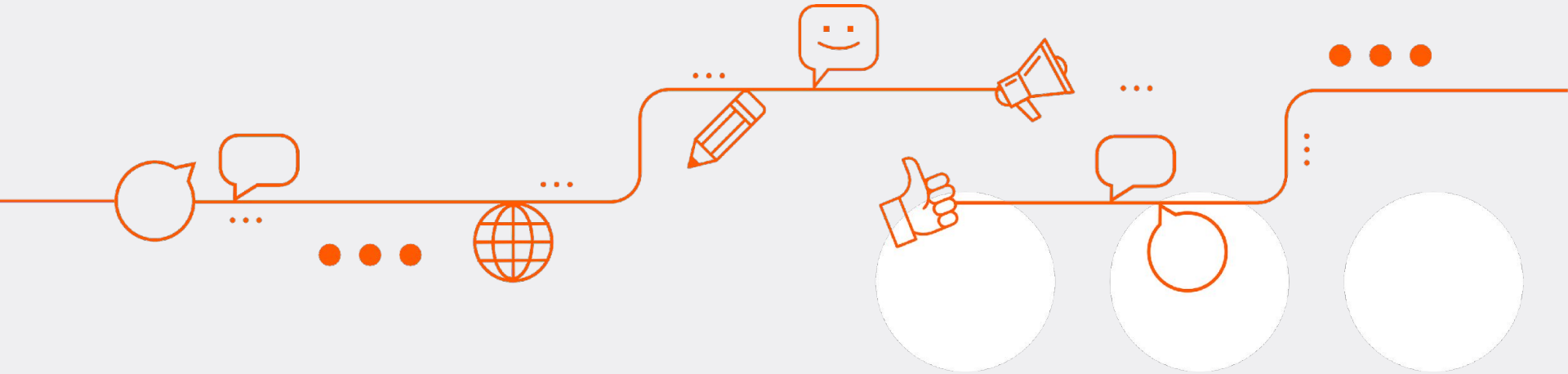
Configurable pipeline that combines **pre-existing solutions** with a few **new components** suitable for experiments with neural networks

Experimentation Framework

Scalable building blocks for experiments with deep learning models

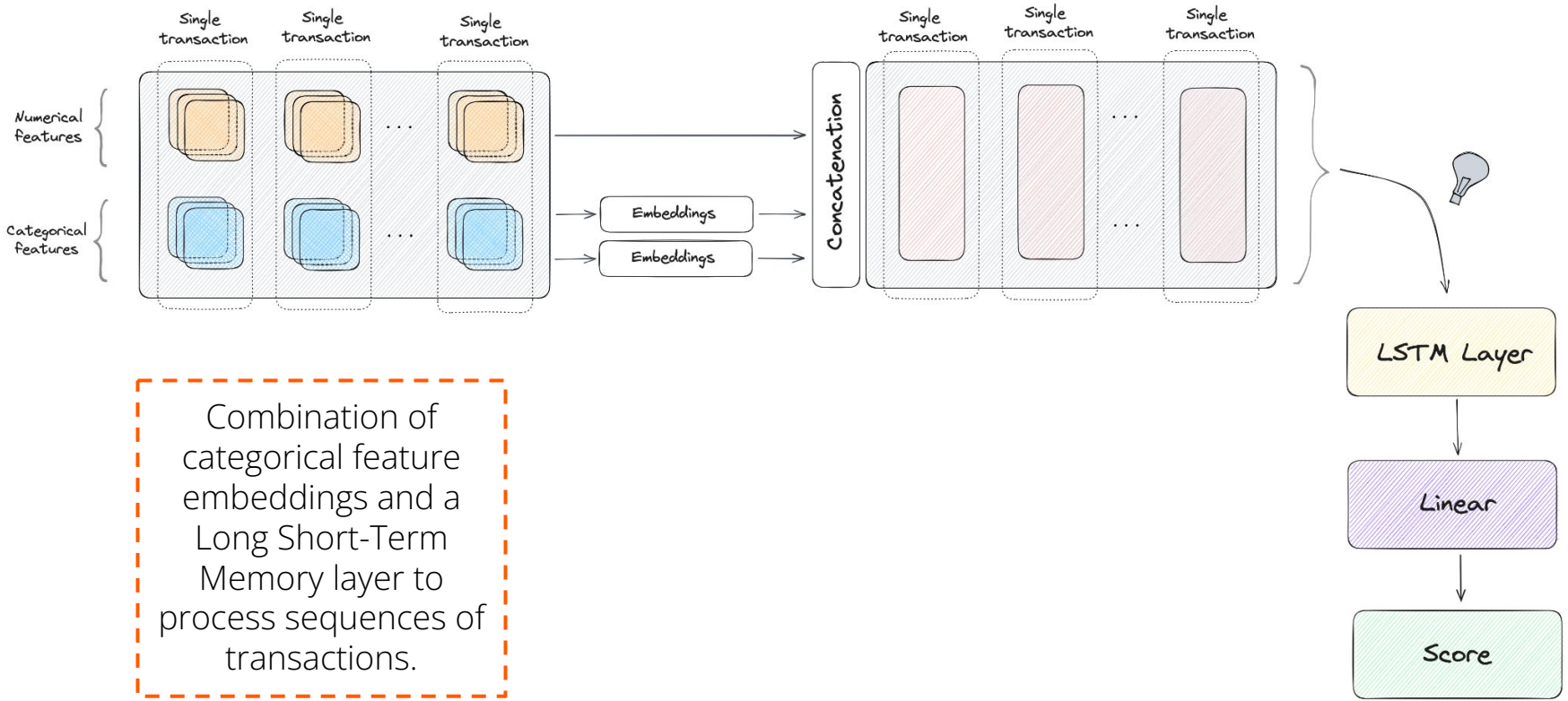


Configurable pipeline that combines **pre-existing solutions** with a few **new components** suitable for experiments with neural networks



Architecture and Results

Architecture



Combination of categorical feature embeddings and a Long Short-Term Memory layer to process sequences of transactions.

Results

		All Clients	Group HA	Group LA
AUC	**Benchmark I**	0.7949*	0.7591*	0.6005
	Benchmark II	0.7695	0.7429	0.6414
	Embedding-LSTM	0.6889	0.7383	0.6936*
mAP	**Benchmark I**	0.0961*	0.025*	0.04
	Benchmark II	0.0556	0.0198	0.0478
	Embedding-LSTM	0.0415	0.02	0.0701*

Group HA stands for a subset of users with high activity, while Group LA stands for a subset of users with high activity
Benchmark I is the best model on all available features. Benchmark II contains only features reflected in the Embedding-LSTM model.

Next steps

01 Add new features!

02 Experiment with new architectures

03 Experiment with larger datasets

Our solution is still in the development phase, but we definitely want to **deploy it to production**

allegro

What if...

we could do great things together?

jobs.allegro.eu

Thank you!

Got any questions?

maciej.wysocki@allegro.com