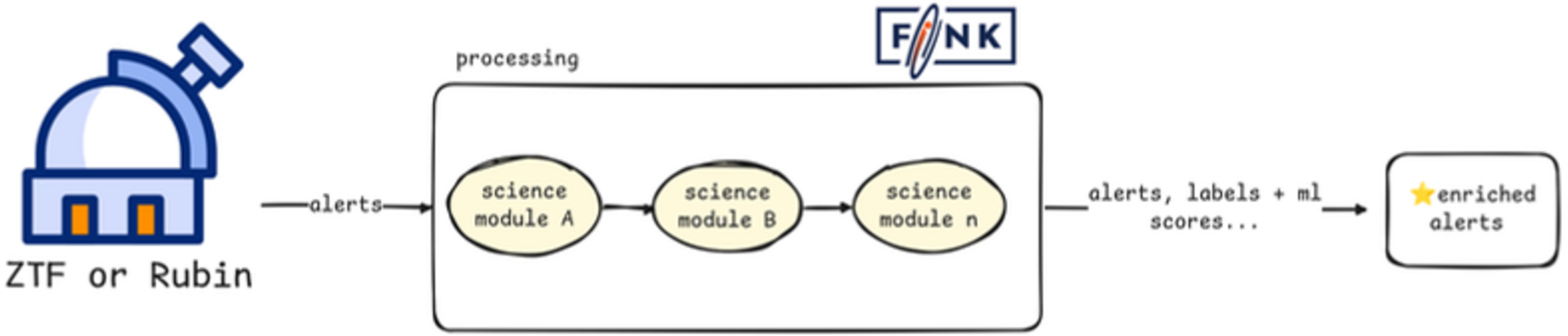
Running Your Own Machine Learning Models on Real Data

Farid MAMAN

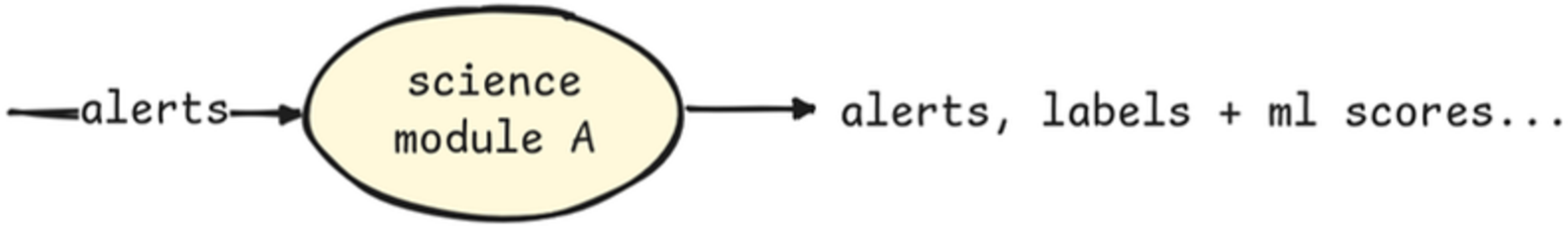
How are alerts processed?



⭐ **Enriched alerts:**

- Saved in the database
- Filters are applied

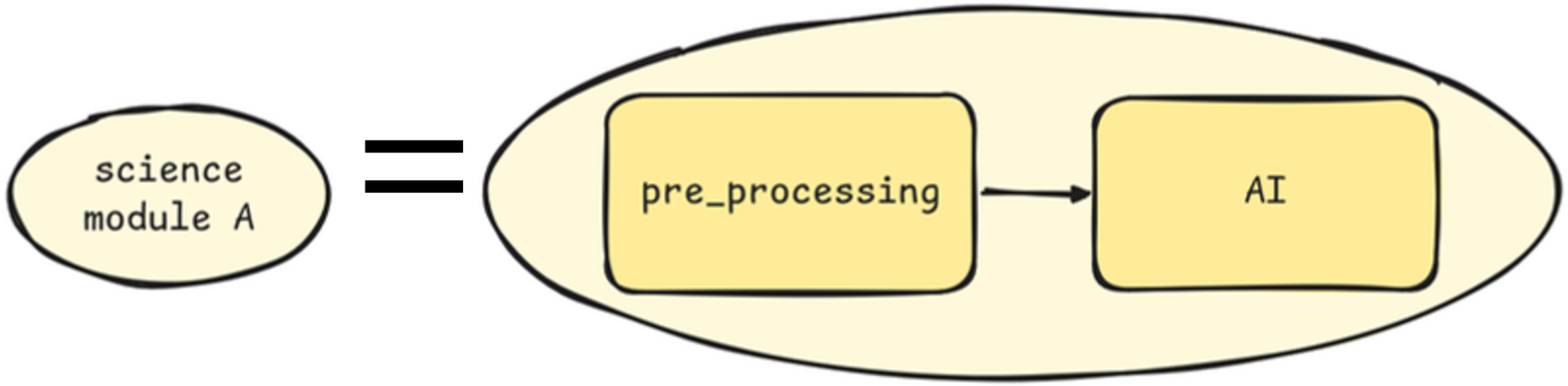
Science modules



- Each module takes an alert as input
- Returns a score attached to the alert

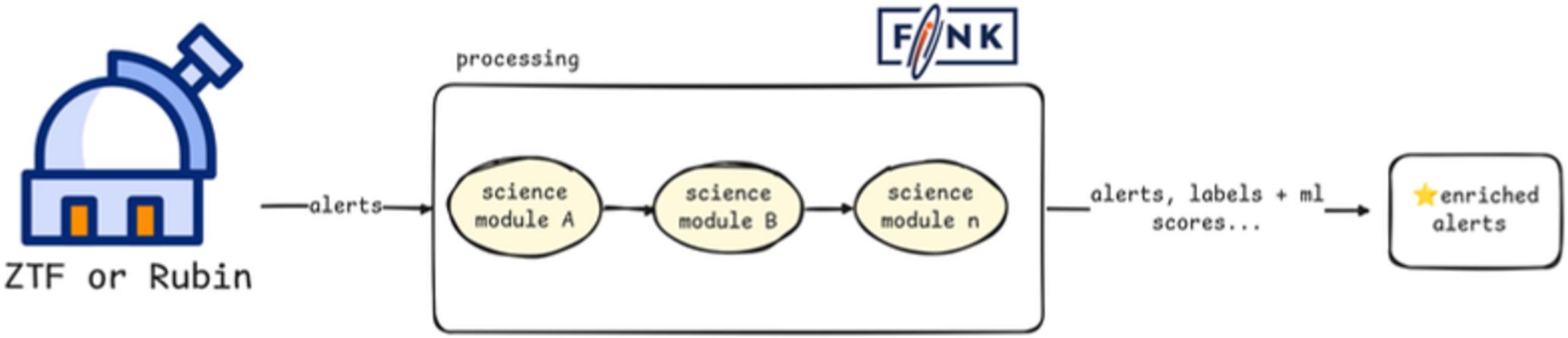
Science modules

Each science module is an AI model dedicated to a specific classification task.

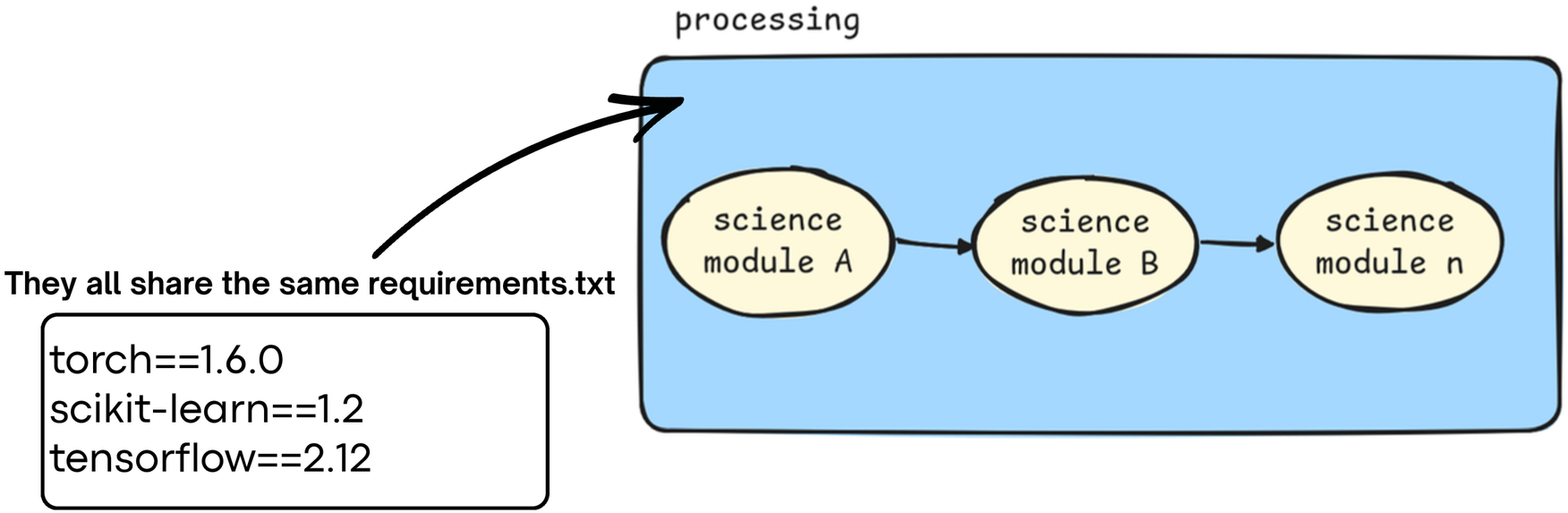


- **pre_processing**: cleans and formats the alert data before feeding it to the model
- **AI model** outputs a score reflecting the probability

What is the problem with this infrastructure?



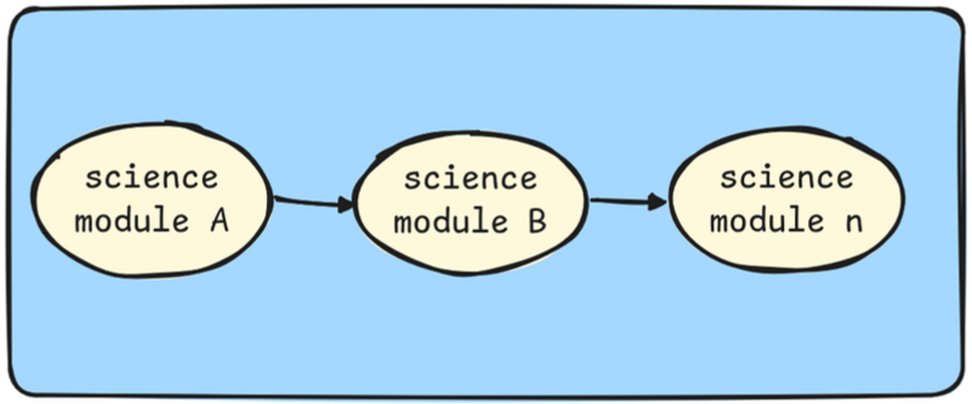
Shared execution environment



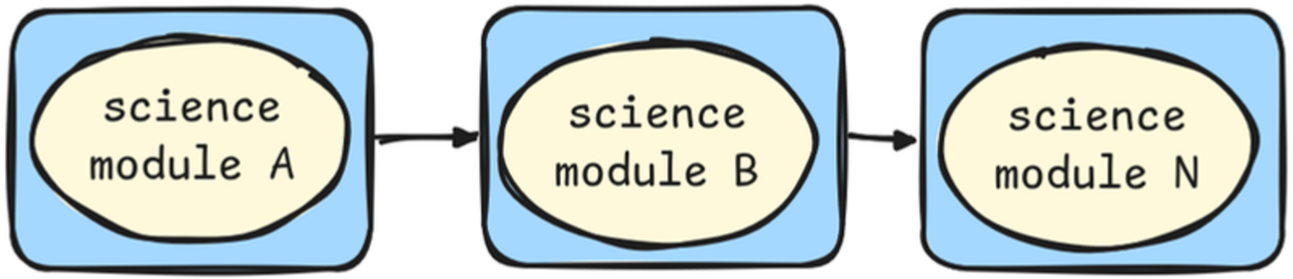
- Every science module uses the same packages and versions
- Need a different version? → Conflicts
- New dependency? => you need to open a PR

The solutions is to wrapper evreything

From this



To this



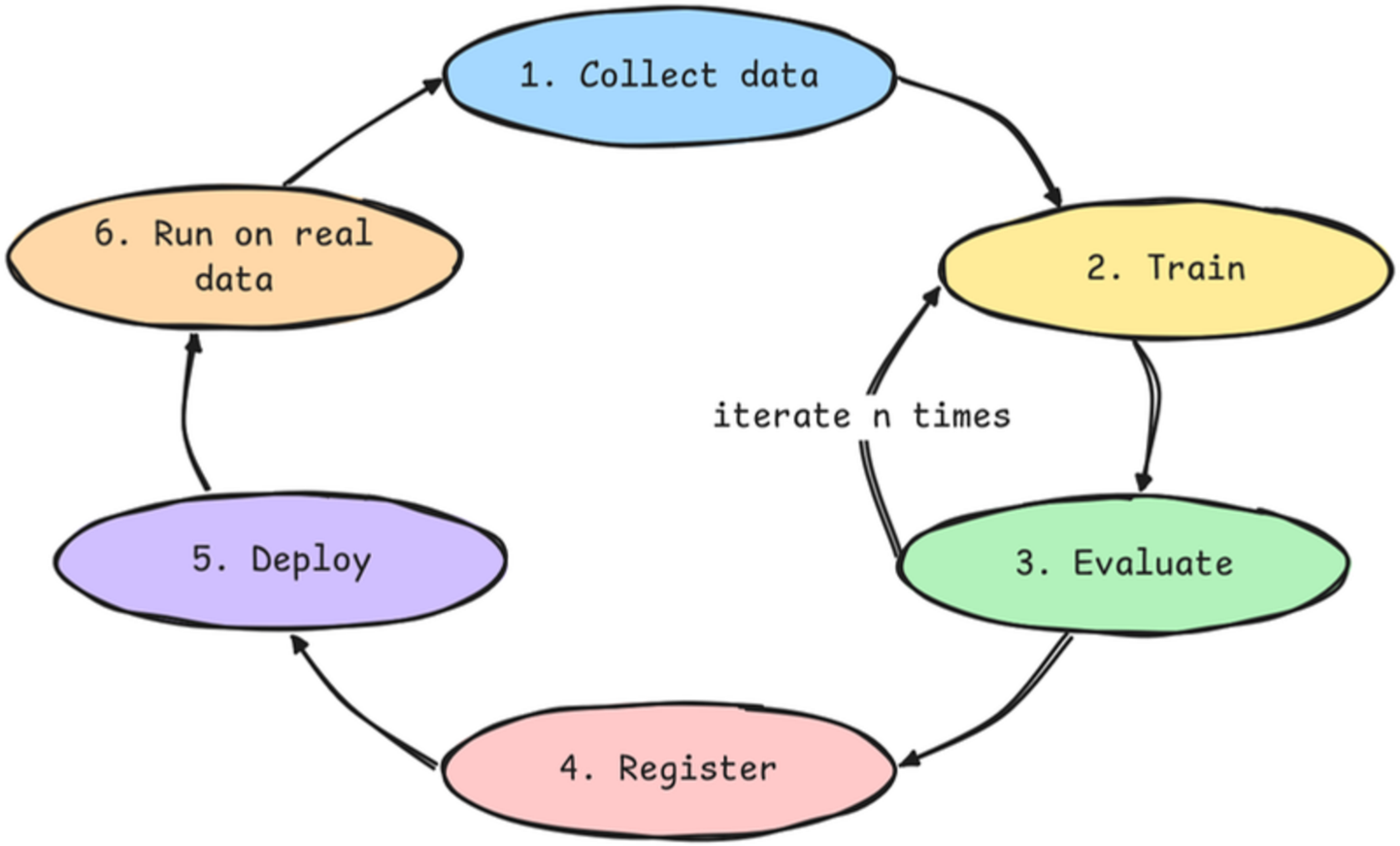
How ?

MLflow

MLflow is an open-source platform to manage the full lifecycle of a machine learning model.

Think of it as: "**Git, but for trained models.**"

MLflow | ML lifecycle



- **Tracks your parameters, metrics and results**

MLFlow | Key concepts

Experiment

- **Groups all your training runs**
- **A run = one training session**

Experiment

runs

Traks :
-parameter
-metrics

Artifacts

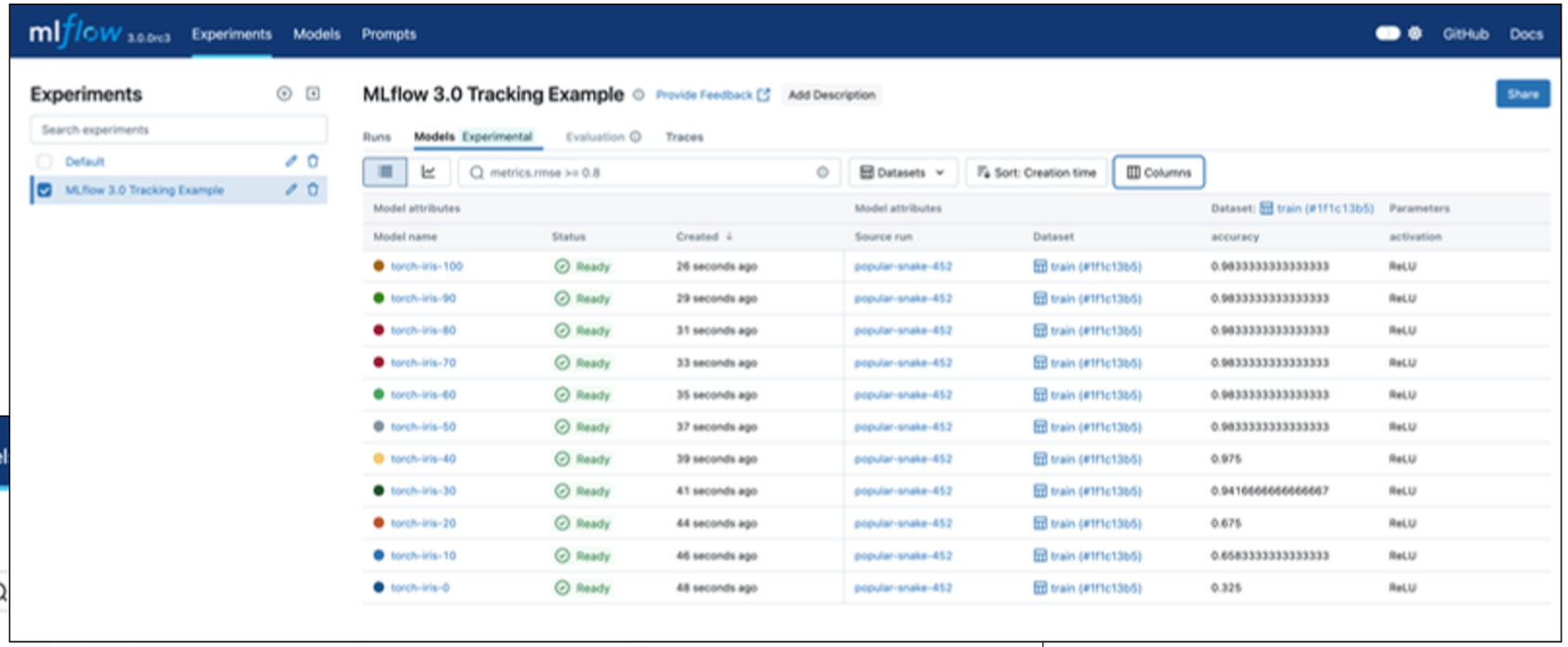
Artifactcs = files stored in mlflow

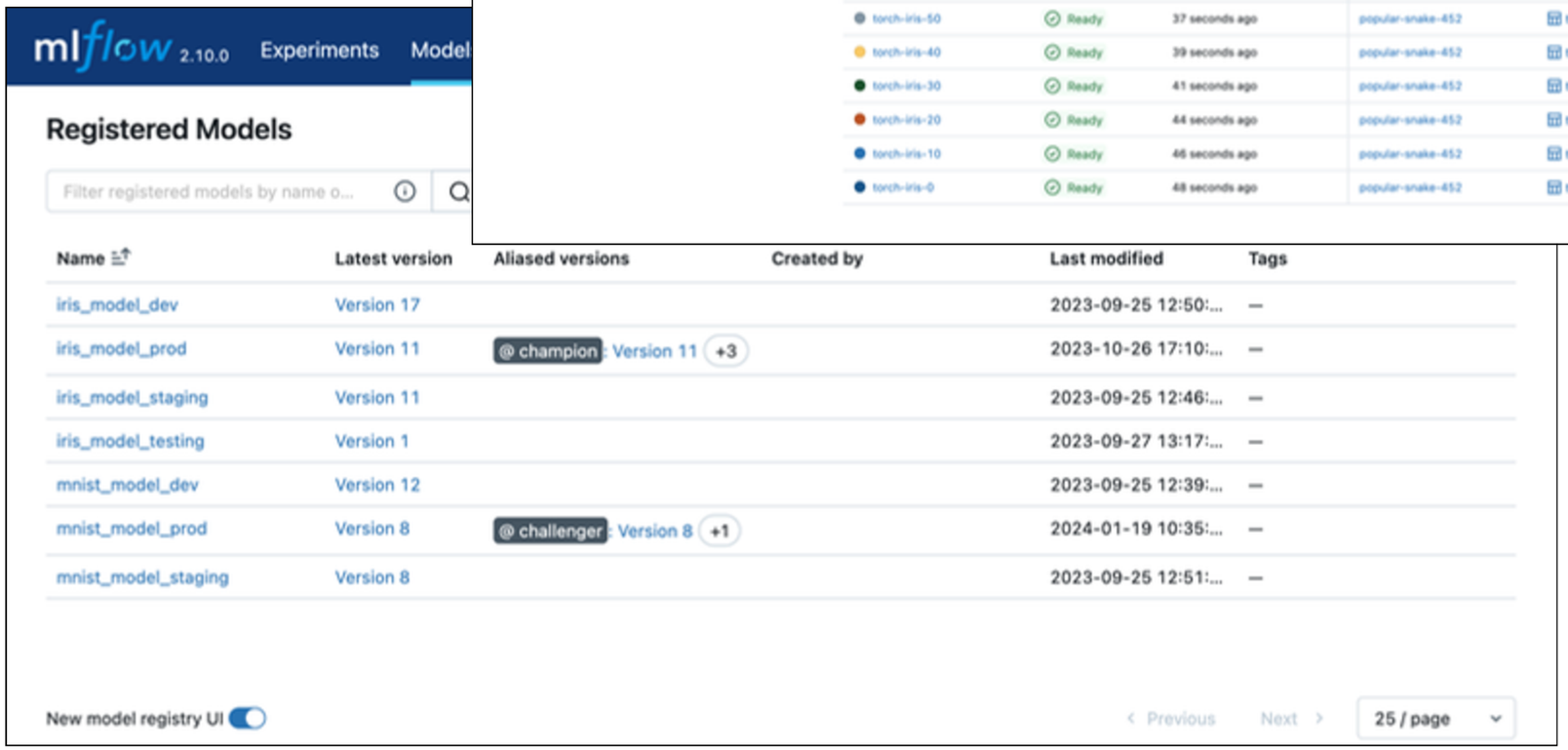
Registry

Registry = is a catalogue where models are stored

Demonstration ?

- [MLflow Fink]
- [MLflow tutorial]






TACK
ARIGATO
GRACIAS
RAHMET
DANKE
TACK
THANK YOU
MERCI
KIITOS
谢谢
GRAZIE
HVALA
TAKK
BARKA
DANK JE