

# Monitor Training W&B and MLflow

Simo Tuomisto, Hossein Firooz

# Does this look familiar?

- Too many experiments
- Forgot which hyperparameters worked
- Hard to reproduce results
- Collaboration nightmares
- `esults_v2_final_final_really_final.csv`

# Why Track Training Runs?

- **Reproducibility:** know exactly how each result was produced
- **Comparison:** evaluate multiple models and configs
- **Accountability:** store data, code, and metrics
- **Collaboration:** share and review results easily

# What Should You Log?

- **Code & Environment: commit, conda, etc**
- **Data version and preprocessing**
- **Model parameters, loss, optimizer**
- **Hyperparameters: lr, batch size, etc.**
- **Metrics: loss, accuracy**
- **Artifacts: plots, checkpoints**

# Tracking Tools

- **Tensorboard**
  - Best for Metric Visualization during training
- **W&B**
  - Cloud-base
  - Quick and easy setup
  - Hyperparameter search and experiments
- **MLflow**
  - Privacy first
  - Manual setup
  - CI/CD and deployment

# MLflow Demo