

Making MLOps Uncool Again

David de la Iglesia

Software Engineer

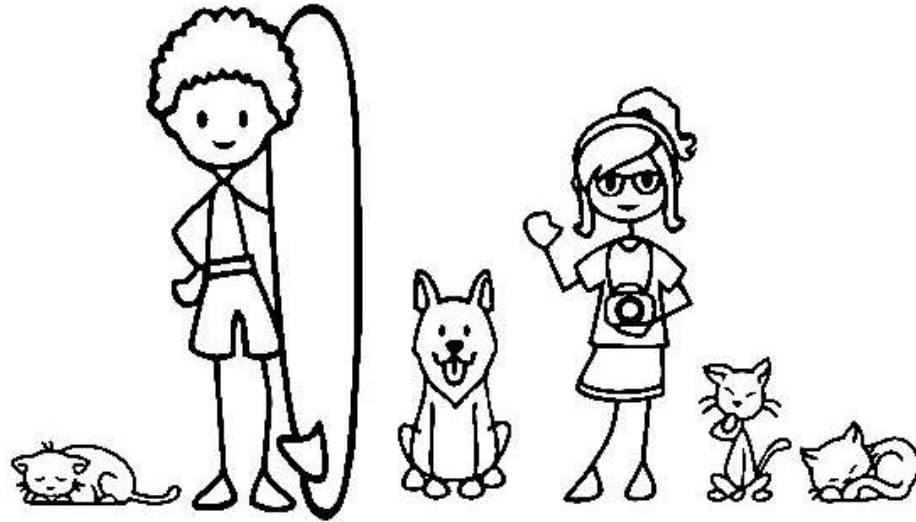
| <https://iterative.ai/about>

Tentative outline



- **Slides: ~15min**
- **Hands on: #NoEstimates**

About me

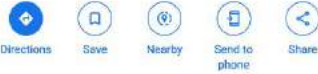


About me



Galicia

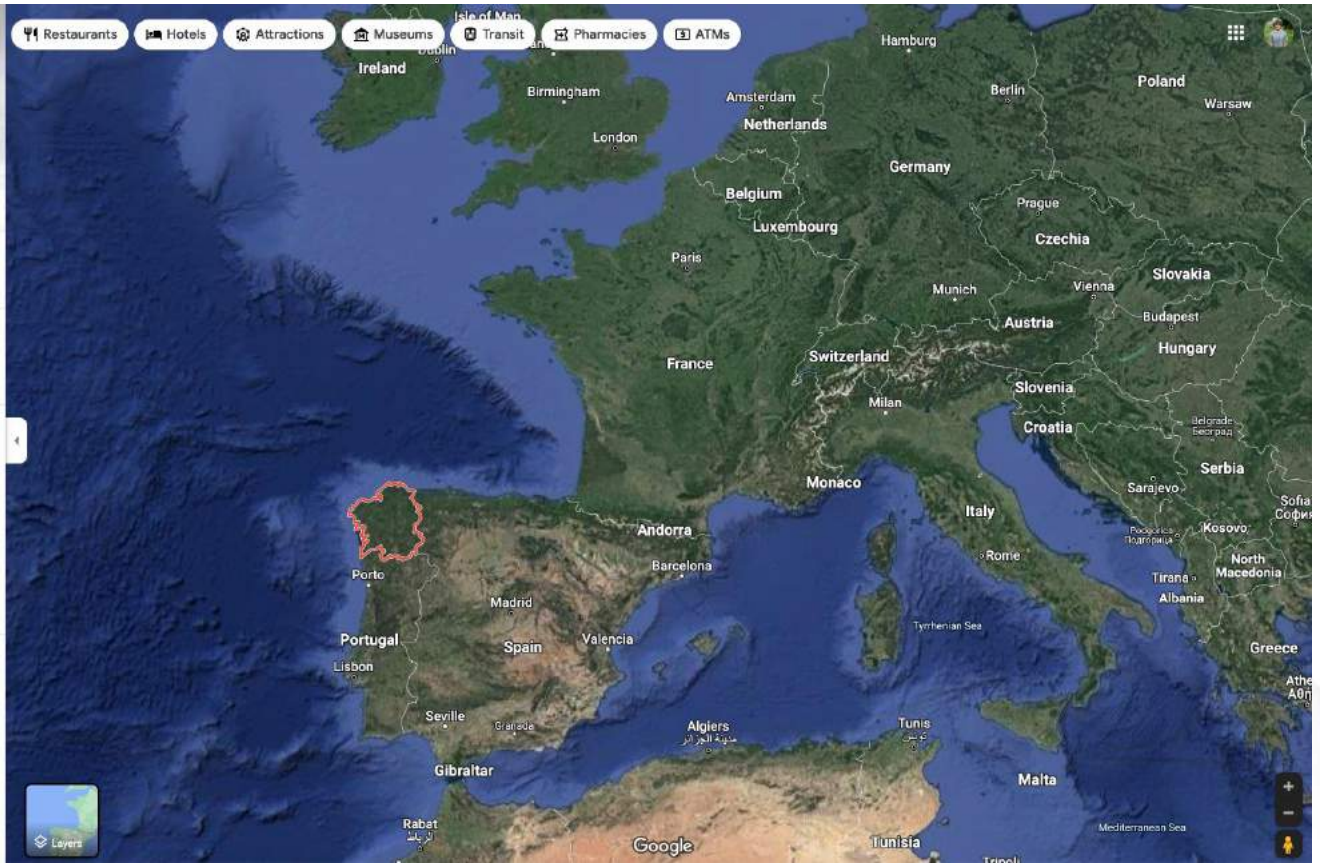
Spain



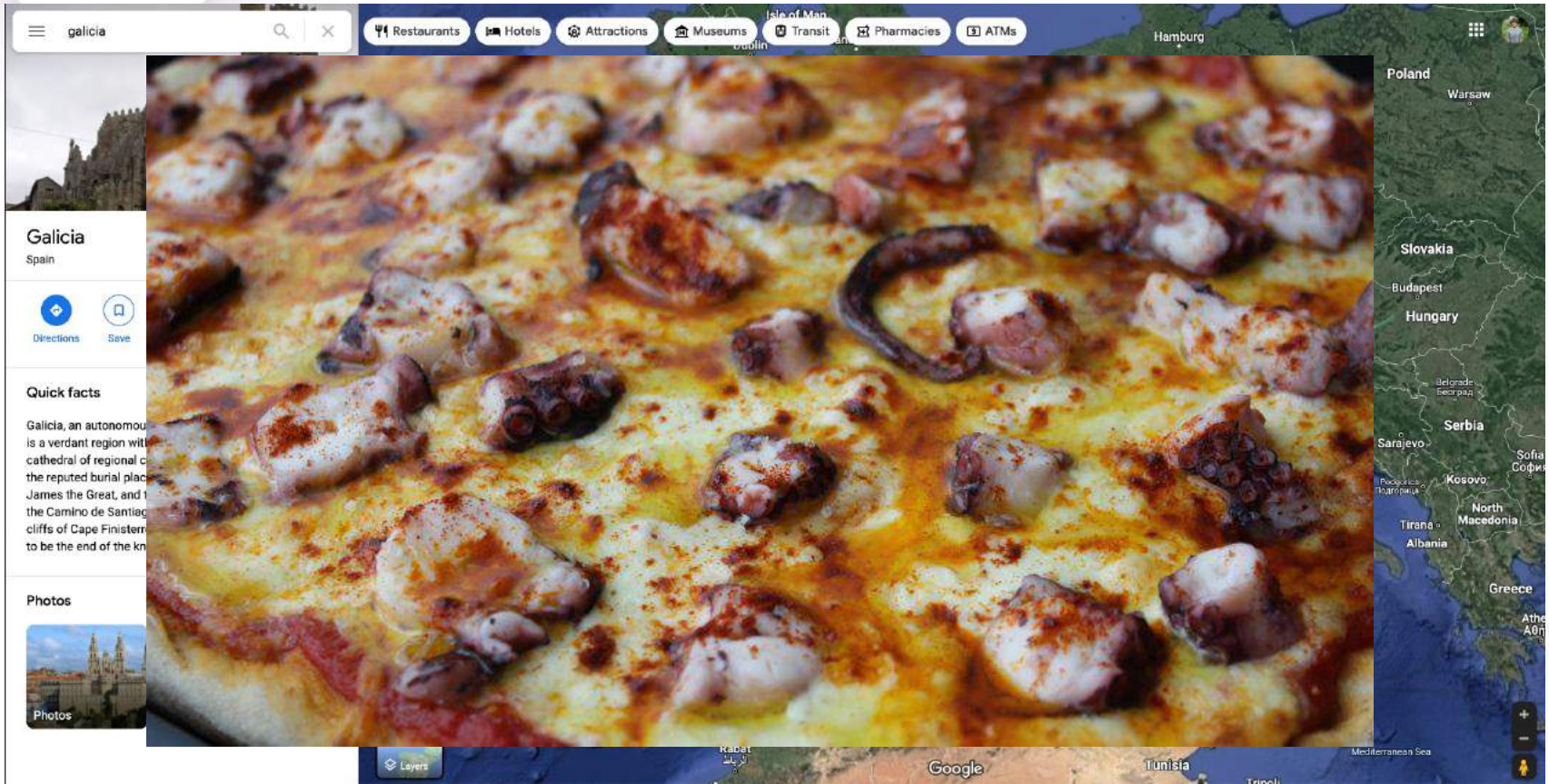
Quick facts

Galicia, an autonomous community in Spain's northwest, is a verdant region with an Atlantic coastline. The cathedral of regional capital Santiago de Compostela is the reputed burial place of the biblical apostle Saint James the Great, and the destination for those following the Camino de Santiago pilgrimage route. The western cliffs of Cape Finisterre were considered by the Romans to be the end of the known world.

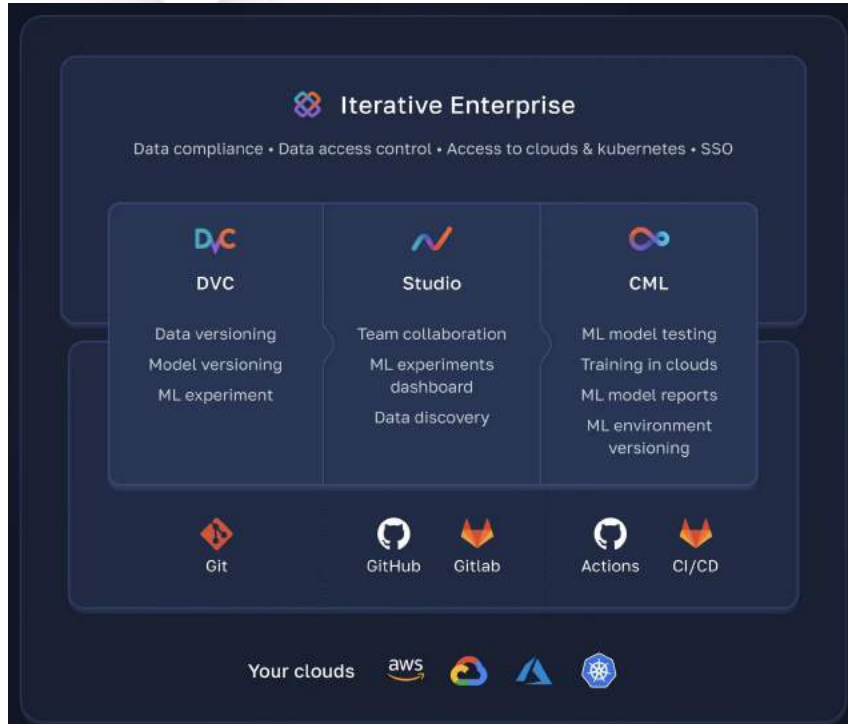
Photos



About me



About iterative.ai



The Team

At Iterative, we build [DVC](#), [CML](#), and other developer tools for machine learning. We're a well-funded, remote-first team on a mission to solve the complexity of managing datasets, ML infrastructure, and ML models lifecycle management.



About you

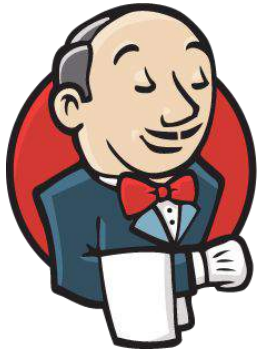


- **Data Science**
- **Machine Learning Engineering**
- **Other**

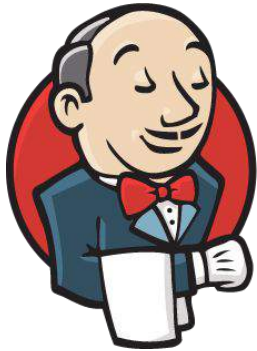


Uncool Again?

DevOps



DevOps



Buddy



DevOps



Concourse



circleci



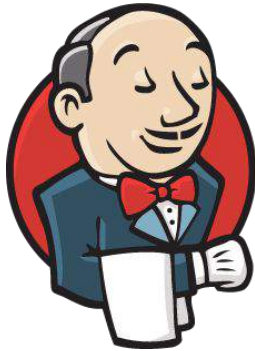
Buddy



CI/CD



GitHub Actions



DevOps



Concourse



circleci



Buddy



CI/CD



GitHub Actions

Uncool

Cool

Uncool



MLOps (IMO)

How I got into MLOps





MLOps Goals

Reproducibility



A researcher travels to a remote island ... and disappears!

The new researcher who takes over the project can easily access all details:

- ◇ Config
- ◇ Data
- ◇ Model
- ◇ Pipeline



Automation



A researcher travels to a remote island and ...

hard chilling

When new data arrives, the model automatically gets:

- ◇ Trained
- ◇ Evaluated
- ◇ Deployed



MLOps Tools

End to End ML platforms



- ◆ **Isolates** researchers from software engineering
- ◆ (Might be) Overkill

- ◆ Vendor lock-in



Open Source Collage



Linux Foundation AI Landscape

2020-11-01T21:39:53Z 83d88a

See the interactive landscape at lfaifoundation.org

Grayed logos are not open source

The screenshot displays the Linux Foundation AI Landscape dashboard, a comprehensive catalog of open-source AI and data tools. The dashboard is organized into a grid of categories, each with a set of logos representing the tools available in that space. The categories include:

- Machine Learning:** Frameworks (e.g., TensorFlow, PyTorch), Platforms (e.g., H2O, Databricks), Libraries (e.g., Scikit-learn, XGBoost), and Tools (e.g., MLflow, Kubeflow).
- Deep Learning:** Frameworks (e.g., PyTorch, TensorFlow), Platforms (e.g., H2O, Databricks), Libraries (e.g., Scikit-learn, XGBoost), and Tools (e.g., MLflow, Kubeflow).
- Reinforcement Learning:** Frameworks (e.g., OpenAI Gym, Stable Baselines), Platforms (e.g., H2O, Databricks), Libraries (e.g., Scikit-learn, XGBoost), and Tools (e.g., MLflow, Kubeflow).
- Programming:** Frameworks (e.g., Jupyter, RStudio), Platforms (e.g., H2O, Databricks), Libraries (e.g., Scikit-learn, XGBoost), and Tools (e.g., MLflow, Kubeflow).
- Notebook Environment:** Frameworks (e.g., Jupyter, RStudio), Platforms (e.g., H2O, Databricks), Libraries (e.g., Scikit-learn, XGBoost), and Tools (e.g., MLflow, Kubeflow).
- Versioning:** Frameworks (e.g., Jupyter, RStudio), Platforms (e.g., H2O, Databricks), Libraries (e.g., Scikit-learn, XGBoost), and Tools (e.g., MLflow, Kubeflow).
- Store & Format:** Frameworks (e.g., Jupyter, RStudio), Platforms (e.g., H2O, Databricks), Libraries (e.g., Scikit-learn, XGBoost), and Tools (e.g., MLflow, Kubeflow).
- Operations:** Frameworks (e.g., Jupyter, RStudio), Platforms (e.g., H2O, Databricks), Libraries (e.g., Scikit-learn, XGBoost), and Tools (e.g., MLflow, Kubeflow).
- Stream Processing:** Frameworks (e.g., Jupyter, RStudio), Platforms (e.g., H2O, Databricks), Libraries (e.g., Scikit-learn, XGBoost), and Tools (e.g., MLflow, Kubeflow).
- SQL Engine:** Frameworks (e.g., Jupyter, RStudio), Platforms (e.g., H2O, Databricks), Libraries (e.g., Scikit-learn, XGBoost), and Tools (e.g., MLflow, Kubeflow).
- Feature Engineering:** Frameworks (e.g., Jupyter, RStudio), Platforms (e.g., H2O, Databricks), Libraries (e.g., Scikit-learn, XGBoost), and Tools (e.g., MLflow, Kubeflow).
- Visualization:** Frameworks (e.g., Jupyter, RStudio), Platforms (e.g., H2O, Databricks), Libraries (e.g., Scikit-learn, XGBoost), and Tools (e.g., MLflow, Kubeflow).
- Pipeline Management:** Frameworks (e.g., Jupyter, RStudio), Platforms (e.g., H2O, Databricks), Libraries (e.g., Scikit-learn, XGBoost), and Tools (e.g., MLflow, Kubeflow).
- Labeling and Annotation:** Frameworks (e.g., Jupyter, RStudio), Platforms (e.g., H2O, Databricks), Libraries (e.g., Scikit-learn, XGBoost), and Tools (e.g., MLflow, Kubeflow).
- Governance:** Frameworks (e.g., Jupyter, RStudio), Platforms (e.g., H2O, Databricks), Libraries (e.g., Scikit-learn, XGBoost), and Tools (e.g., MLflow, Kubeflow).
- Benchmarking:** Frameworks (e.g., Jupyter, RStudio), Platforms (e.g., H2O, Databricks), Libraries (e.g., Scikit-learn, XGBoost), and Tools (e.g., MLflow, Kubeflow).
- Training:** Frameworks (e.g., Jupyter, RStudio), Platforms (e.g., H2O, Databricks), Libraries (e.g., Scikit-learn, XGBoost), and Tools (e.g., MLflow, Kubeflow).
- Parameter:** Frameworks (e.g., Jupyter, RStudio), Platforms (e.g., H2O, Databricks), Libraries (e.g., Scikit-learn, XGBoost), and Tools (e.g., MLflow, Kubeflow).
- Format & Interface:** Frameworks (e.g., Jupyter, RStudio), Platforms (e.g., H2O, Databricks), Libraries (e.g., Scikit-learn, XGBoost), and Tools (e.g., MLflow, Kubeflow).
- Marketplace:** Frameworks (e.g., Jupyter, RStudio), Platforms (e.g., H2O, Databricks), Libraries (e.g., Scikit-learn, XGBoost), and Tools (e.g., MLflow, Kubeflow).
- Workflow:** Frameworks (e.g., Jupyter, RStudio), Platforms (e.g., H2O, Databricks), Libraries (e.g., Scikit-learn, XGBoost), and Tools (e.g., MLflow, Kubeflow).
- Inference:** Frameworks (e.g., Jupyter, RStudio), Platforms (e.g., H2O, Databricks), Libraries (e.g., Scikit-learn, XGBoost), and Tools (e.g., MLflow, Kubeflow).
- Tool:** Frameworks (e.g., Jupyter, RStudio), Platforms (e.g., H2O, Databricks), Libraries (e.g., Scikit-learn, XGBoost), and Tools (e.g., MLflow, Kubeflow).
- Model:** Frameworks (e.g., Jupyter, RStudio), Platforms (e.g., H2O, Databricks), Libraries (e.g., Scikit-learn, XGBoost), and Tools (e.g., MLflow, Kubeflow).
- Computing & Management:** Frameworks (e.g., Jupyter, RStudio), Platforms (e.g., H2O, Databricks), Libraries (e.g., Scikit-learn, XGBoost), and Tools (e.g., MLflow, Kubeflow).
- Interface:** Frameworks (e.g., Jupyter, RStudio), Platforms (e.g., H2O, Databricks), Libraries (e.g., Scikit-learn, XGBoost), and Tools (e.g., MLflow, Kubeflow).
- Security & Privacy:** Frameworks (e.g., Jupyter, RStudio), Platforms (e.g., H2O, Databricks), Libraries (e.g., Scikit-learn, XGBoost), and Tools (e.g., MLflow, Kubeflow).
- Natural Language Processing:** Frameworks (e.g., Jupyter, RStudio), Platforms (e.g., H2O, Databricks), Libraries (e.g., Scikit-learn, XGBoost), and Tools (e.g., MLflow, Kubeflow).
- Education:** Frameworks (e.g., Jupyter, RStudio), Platforms (e.g., H2O, Databricks), Libraries (e.g., Scikit-learn, XGBoost), and Tools (e.g., MLflow, Kubeflow).

The dashboard also includes a QR code and a link to the DLFAI & DATA website.

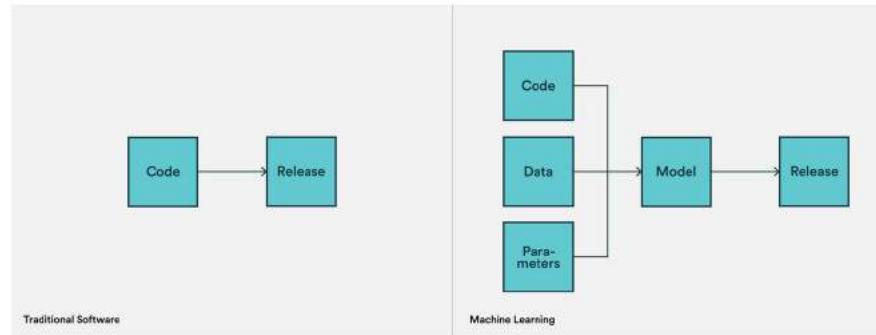


Just use DevOps?

What's missing on



Versioning for Machine Learning



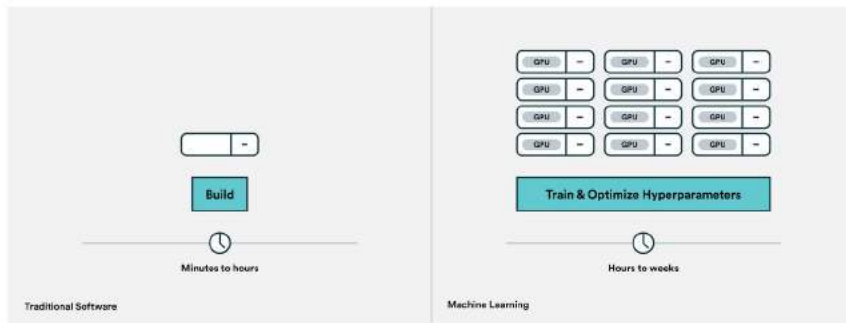
With DevOps, code version control is utilized to ensure clear documentation regarding any changes or adjustments made to the software being developed. With machine learning, however, the code isn't the only changing input. Data is the other critical input that'll need to be managed, as will parameters, metadata, logs, and finally, the model.

[What Is The Difference Between DevOps And MLOps?](#)

What's missing on



Hardware Required



CML

Training machine learning models, especially true for deep learning, tend to be very compute-intensive. For most software projects, the build time is entirely irrelevant, and thus the hardware on which it's done is also irrelevant. Larger models, however, can take anywhere from hours to weeks to train, even on most GPU machines cloud vendors offer, meaning that an MLOps setup needs to be much more sophisticated in what kind of machines it can manage.

[What Is The Difference Between DevOps And MLOps?](#)



Hands On



github.com/iterative/workshop-uncool-mlops

