The 2020 State of

ENGINEERING PRACTICES for MACHINE LEARNING

Image by Will Menegas, Uchida Lab, Department of Molecular and Cellular Biology

SE4ML

se-ml.github.io

Key findings

Our **global survey** among teams that build software with machine learning components revealed which **engineering practices**

- are most adopted,
- by whom, and
- to which effect.

Read on for details …

Tech companies are leading in adoption of ML engineering best practices.

Larger and more **experienced teams** tend to adopt more practices.

General **software engineering** practices enjoy slightly lower adoption than specific **machine learning** practices.

Best practices for **feature management** are the least well adopted.

Desired outcomes such as **traceability**, **agility**, **effectiveness**, and **quality** are each related to specific sets of practices.

About the Survey

We ran a **global survey** among ML practitioners using an online questionnaire.

Between January and May 2020, we collected **350** responses.

Participants took on average **7 minutes** to answer about **40 questions**.

The goal was to determine the current state of the art regarding the adoption of software engineering practices by **teams** that develop software that has Machine Learning components.

The practices included in the questionnaire were determined through an **extensive review** of scientific articles and practitioner blogs.

Cite as: The 2020 State of Engineering Practices for Machine Learning by Alex Serban, Koen van der Blom, Holger Hoos, and Joost Visser.
More: For more information, and to stay up to date, visit the website of the SE4ML research project: https://se-ml.github.io

SE4ML
se-ml.github.io

Engineering Practices for Machine Learning

We extracted 29 recommended engineering practices from scientific articles and practitioner blogs. See our online catalogue for a more extensive description of each practice.

Data

Use sanity checks for all external data sources

Check that input data is complete, balanced and well distributed

Write reusable scripts for data cleaning and merging

Ensure data labelling is performed in a strictly controlled process

Make data sets available on shared infrastructure (private or public)

Coding

Run automated regression tests

Use continuous integration

Use static analysis to check code quality

Deployment

Automate model deployment

Continuously monitor the behaviour of deployed models

Enable shadow deployment

Perform checks to detect skews between models

Enable automatic roll backs for production models

Log production predictions with the model's version and input data

Team

Use a collaborative development platform

Work against a shared backlog

Communicate, align, and collaborate with multidisciplinary team members

Governance

Ensure fairness and privacy

Training

Share a clearly defined training objective within the team

Capture the training objective in a metric that is easy to measure and understand

Test all feature extraction code

Assign an owner to each feature and document its rationale

Actively remove or archive features that are not used

Peer review training scripts

Enable parallel training experiments

Automate hyper-parameter optimisation and model selection

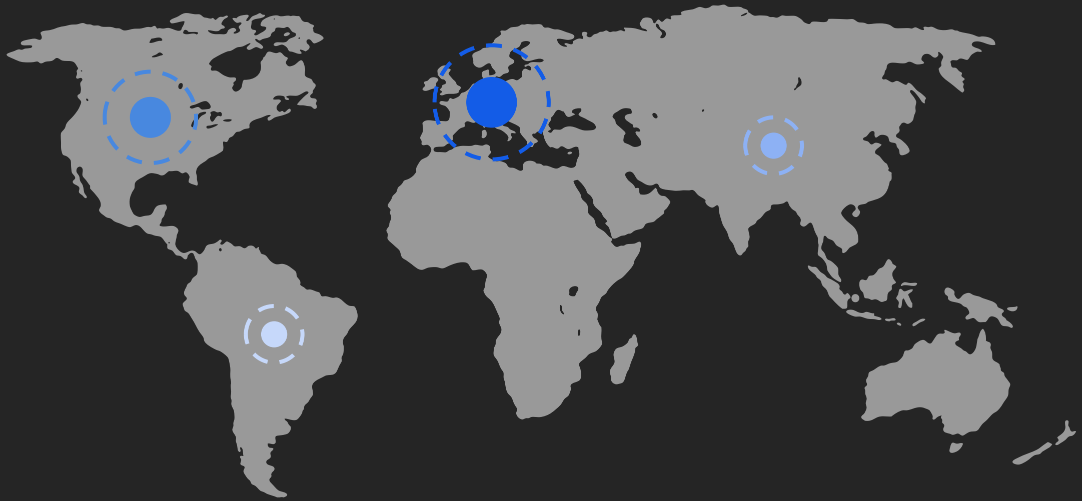
Continuously measure model quality and performance

Share status and outcomes of experiments within the team

Use versioning for data, model, configurations and training scripts

Machine learning teams around the globe

Response distribution
by region



01

Europe
55%

02

North America
23%

03

Asia and Oceania
12%

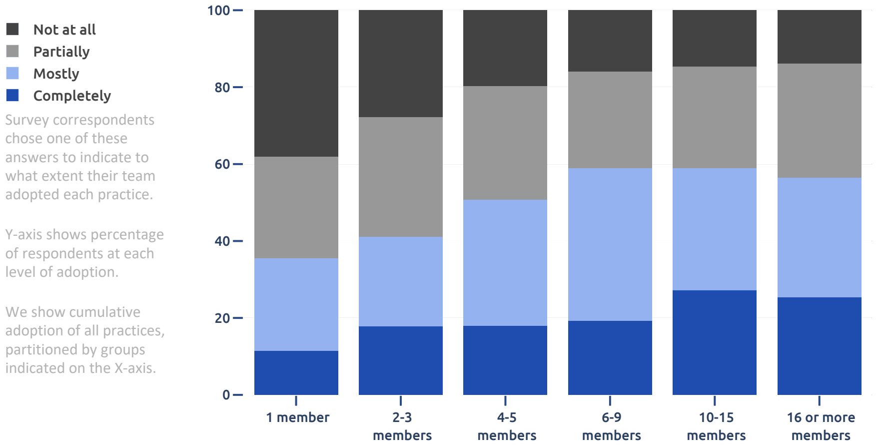
04

South America
10%

We did not receive any responses from ML teams in Africa.

Take the survey to help us improve geographical representation!

Team Size



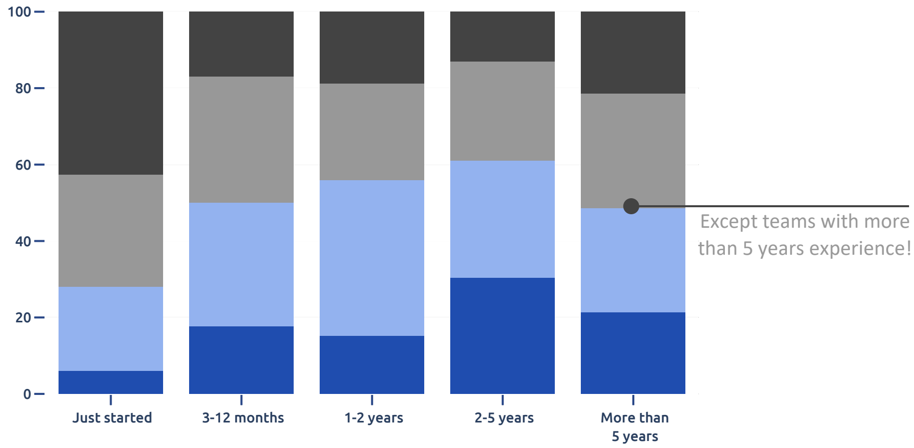
Survey correspondents chose one of these answers to indicate to what extent their team adopted each practice.

Y-axis shows percentage of respondents at each level of adoption.

We show cumulative adoption of all practices, partitioned by groups indicated on the X-axis.

Larger teams tend to adopt more practices.

Team Experience

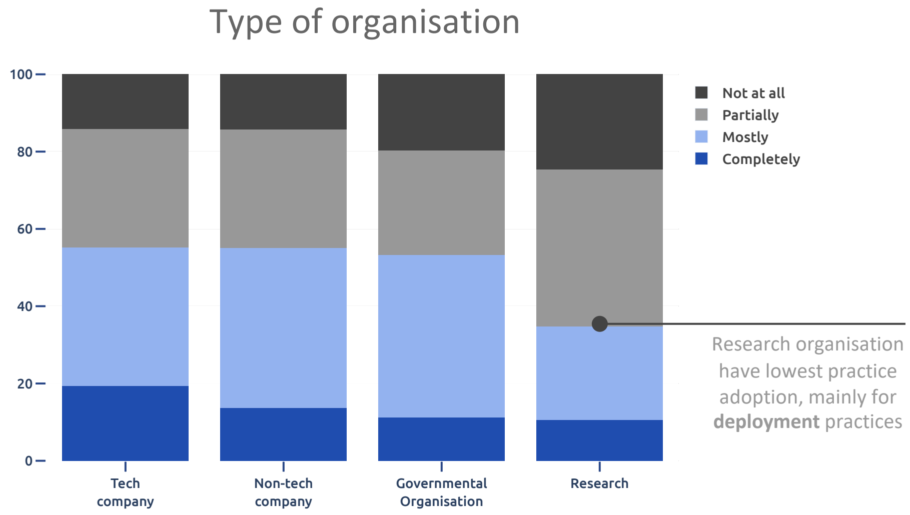


More experienced teams tend to adopt more practices.

Practice adoption increases with team size and experience

Tech companies lead practice adoption

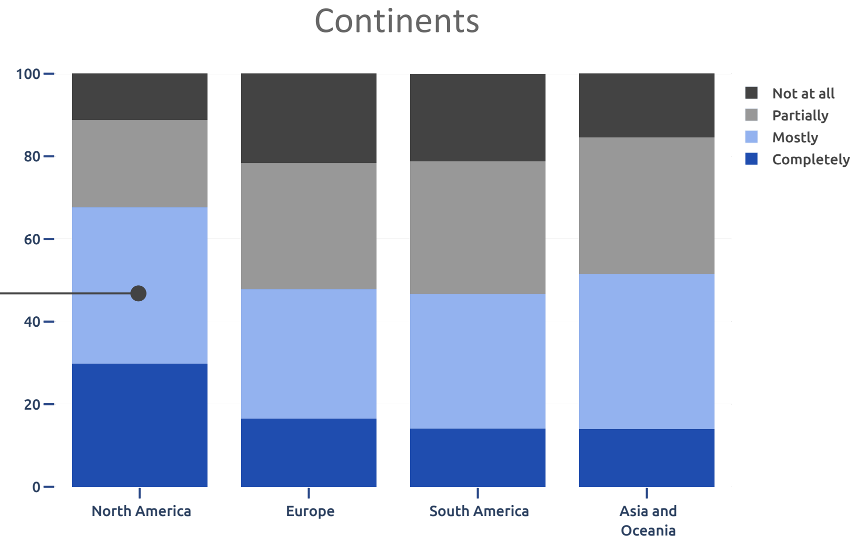
The adoption of best practices by tech companies is higher than by non-tech companies, governmental organisations, and research labs.



Tech companies adopt Continuous Integration 15% more often than non-tech companies.

Practice adoption around the world

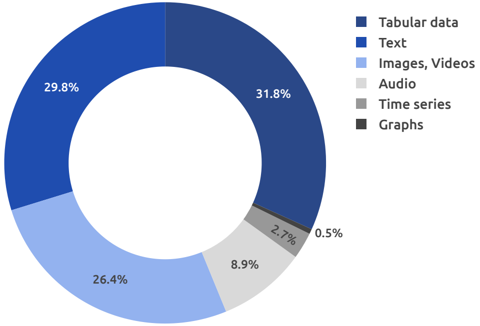
The adoption of best practices is similar across continents, except North-America, where adoption is markedly higher.



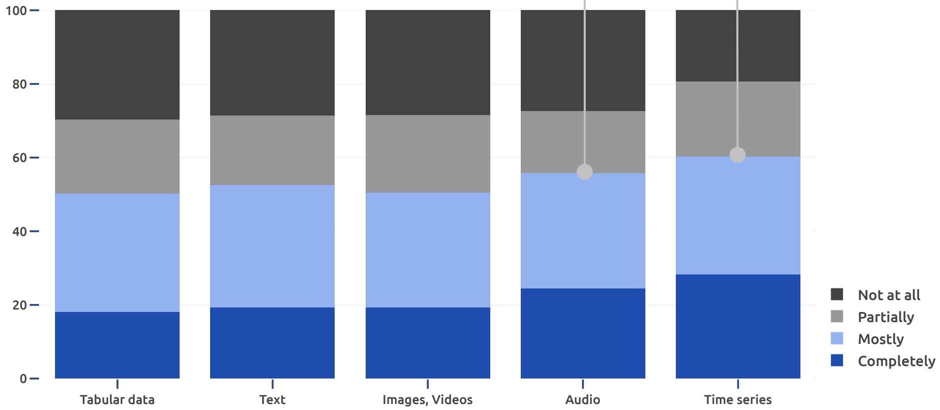
North-America is ahead of other continents

Los Angeles is the city with the highest practice adoption.

Types of data



Teams that work with Audio and Time Series data (11.4% of respondents) show higher practice adoption



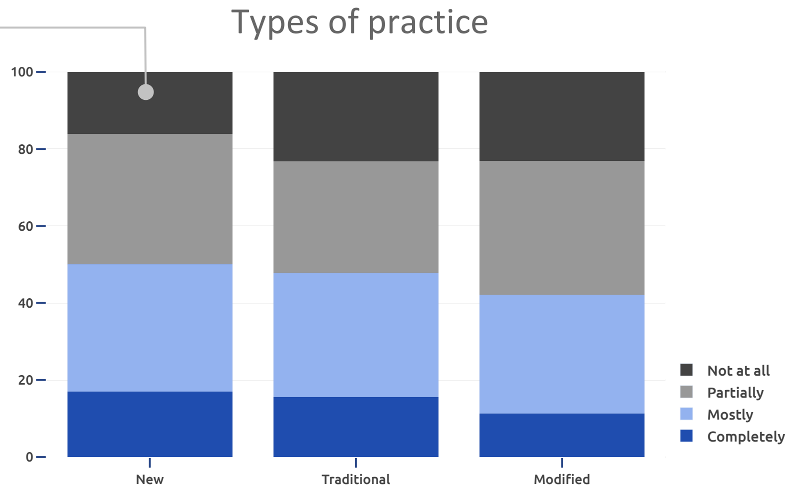
The adoption of best practices is largely independent of the type of data that is being processed.

Different data, same practices

Automatic hyper-parameter optimisation is adopted 10% more for tabular data than for other data types.

ML-specific practices are adopted slightly more than general Software Engineering practices

ML-specific practices enjoy the highest degree of adoption



Among ML teams, the adoption of ML-specific practices is highest, followed by general Software Engineering (SE) practices and SE practices adapted to ML.

SE practices related to code quality (static analysis and regression testing) have lowest adoption.

Most adopted practices

Practices related to **measurement** and **versioning** are widely adopted.

The top 4 adopted practices are all related to **model training.**

Top 5

1. Capture the training objective in a metric that is easy to measure and understand
2. Share a clearly defined training objective within the team
3. Use versioning for data, model, configurations and training scripts
4. Continuously measure model quality and performance
5. Write reusable scripts for data cleaning and merging

Least adopted practices

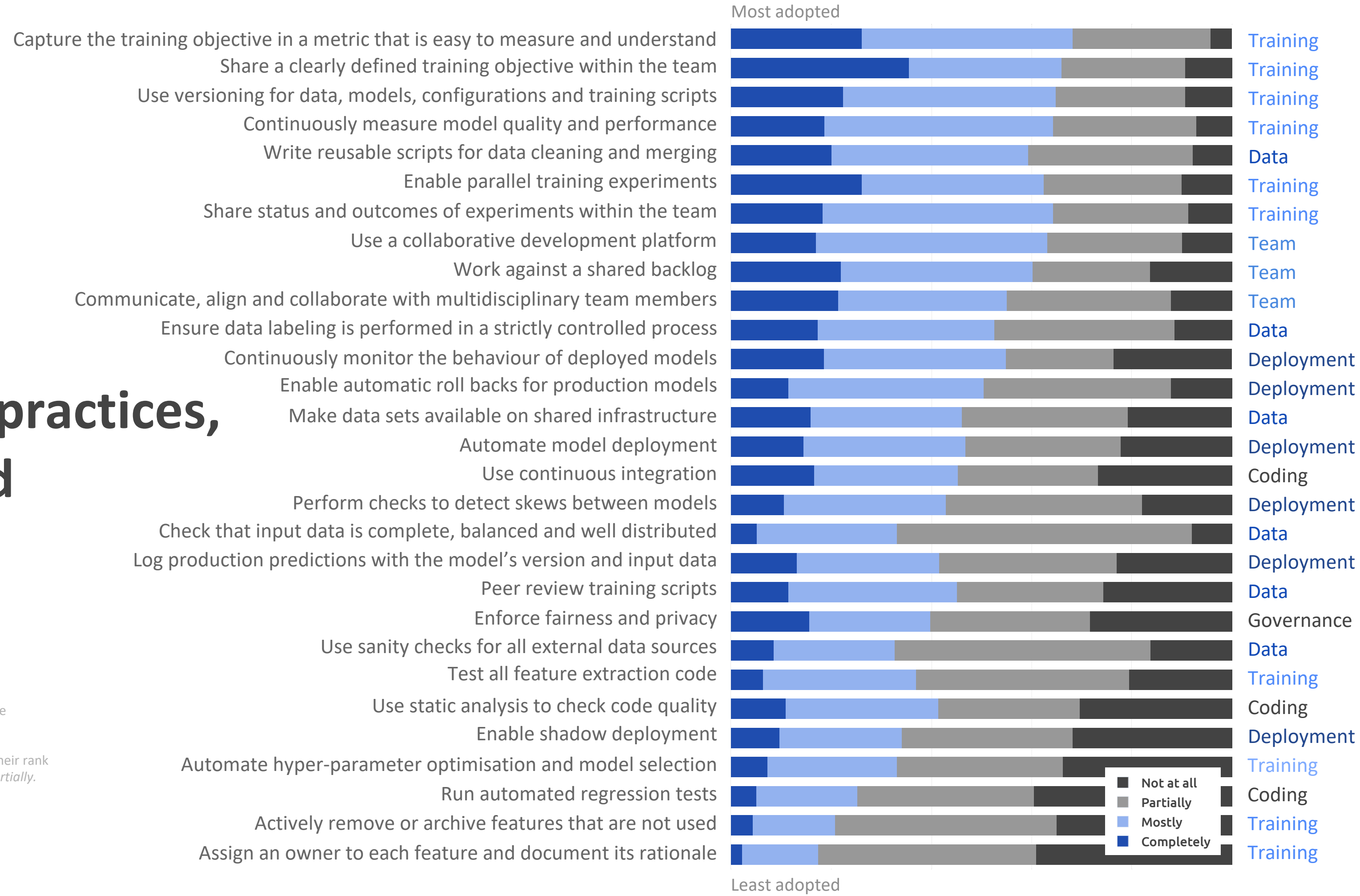
The two most neglected practices are related to **feature management**.

Outside research, **Automated ML** through automated optimisation of hyper-parameters and model selection, is not (yet) widely applied.

Bottom 5

1. Assign an owner to each feature and document its rationale
2. Actively remove or archive features that are not used
3. Run automated regression tests
4. Automate hyper-parameter optimisation and Model Selection
5. Enable shadow deployment

All 29 practices, ranked



Practices are ranked by the average of: their rank on *Completely*, their rank on *Completely+Mostly*, and their rank on *Completely+Mostly+Partially.*

Different practices, different outcomes

Analysis of survey responses shows that desired outcomes such as **traceability**, **agility**, team **effectiveness**, and software **quality** are each related to specific sets of practices.

Per desired outcome, we list the three practices with the largest influence.

Agility

1. Automate model deployment
2. Communicate, align, and collaborate with multidisciplinary team members
3. Enable parallel training experiments

Traceability

1. Log production predictions with the model's version and input data
2. Continuously monitor the behaviour of deployed models
3. Use versioning for data, model, configurations and training scripts

Team Effectiveness

1. Work against a shared backlog
2. Use a collaborative development platform
3. Share a clearly defined training objective within the team

Software Quality

1. Use continuous integration
2. Run automated regression tests
3. Use static analysis to check code quality

Learn more

Reading list

We reviewed scientific and popular literature to identify recommended practices. Check out this Awesome List with relevant literature.

Catalogue

The best practices that we identified are describe in more detail in this Catalogue of ML Engineering Best Practices.

Preprint

Full details of the methodology behind our survey are described in a scientific article. Read the preprint here.

se-ml.github.io

Visit our project website for more details, to take the survey yourself, and to stay up-to-date with our latest results.

Alex Serban

Koen van der Blom

Holger Hoos

Joost Visser

Team

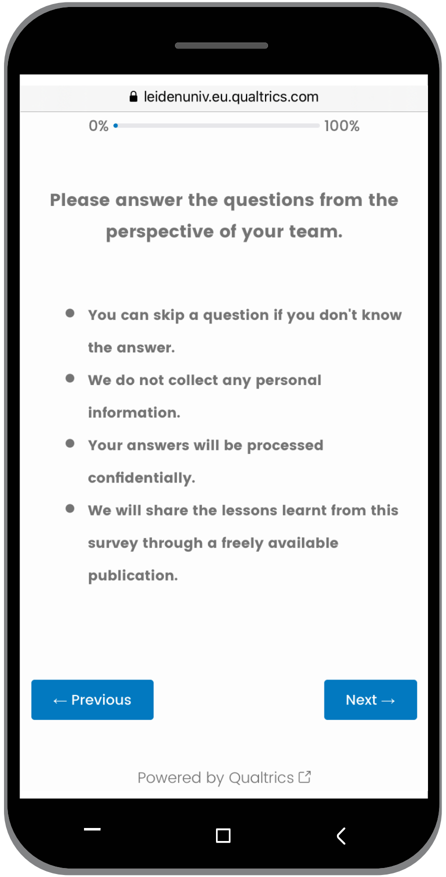
https://se-ml.github.io/members/

LIACS, Leiden University, The Netherlands

ICIS, Radboud University, The Netherlands

University of British Columbia, Canada

You can help



Take the Survey

If you have not done so yet,
please take our 10-min survey!

We will use your answers for our next report on the State of Engineering Practices for Machine Learning.