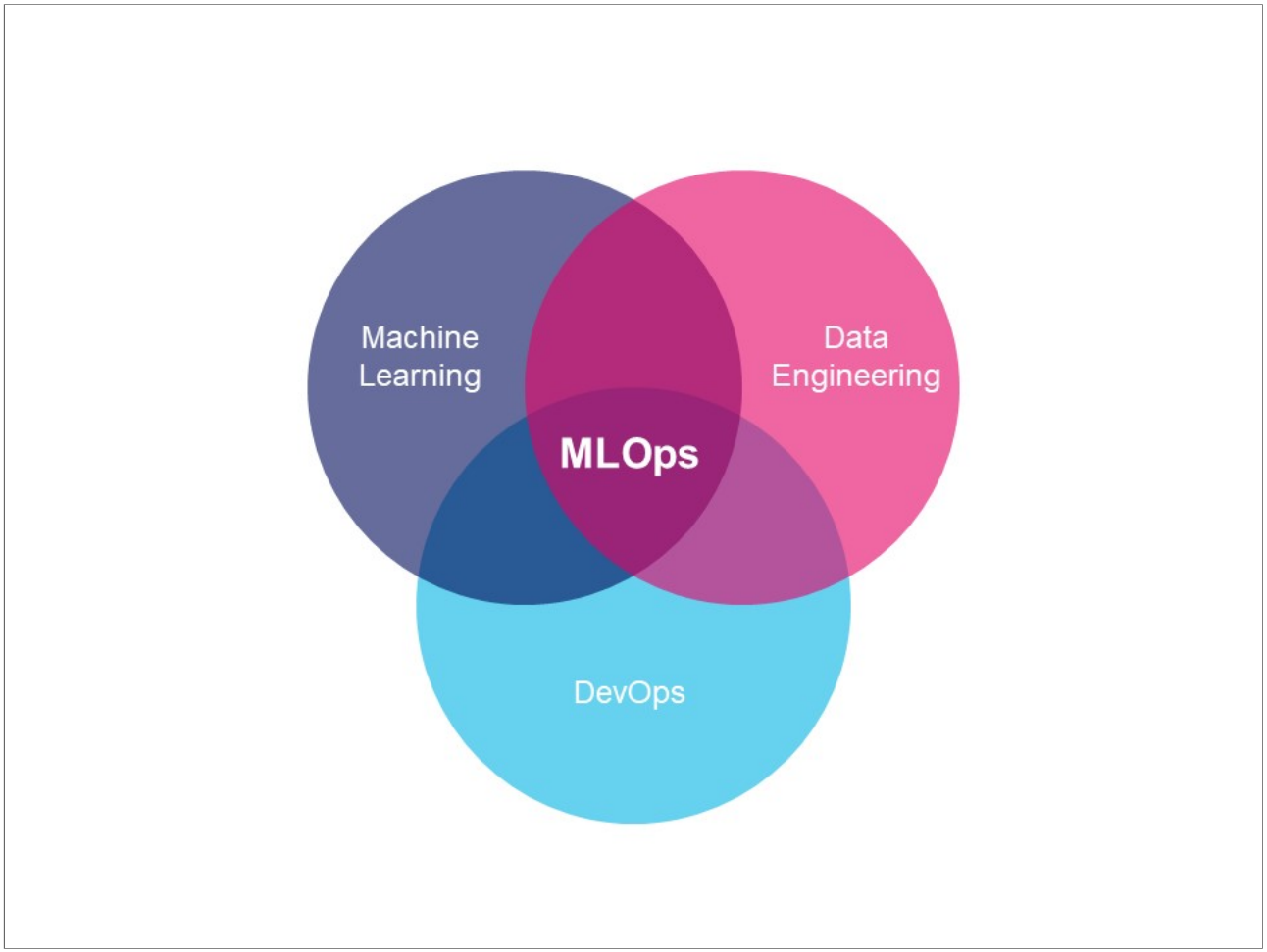
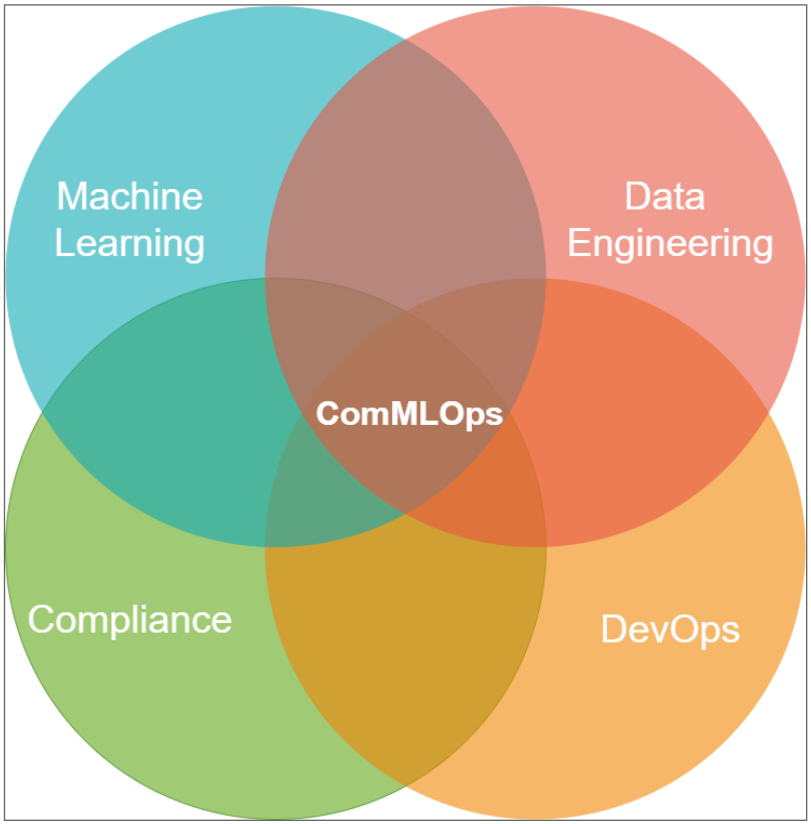
MLOps

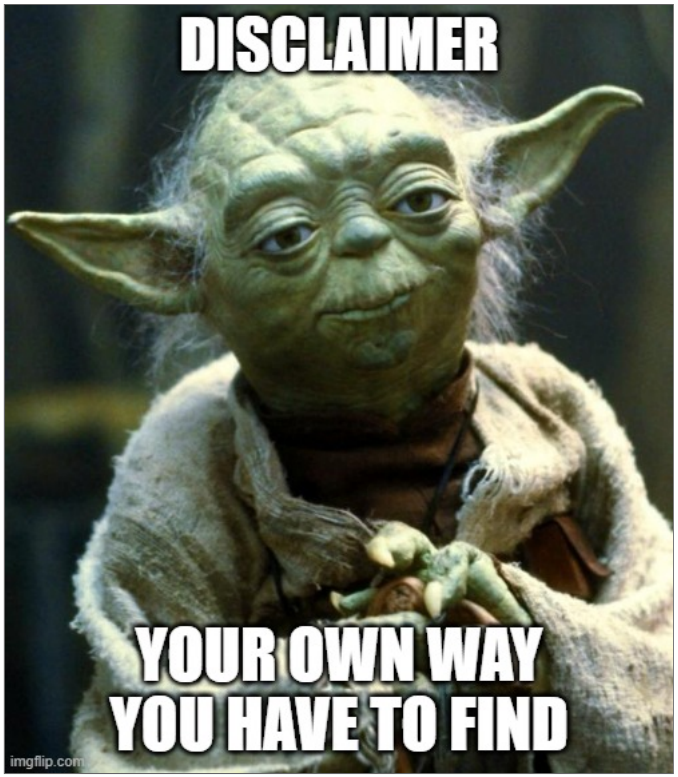


ComMLOps


Machine Learning
Data Engineering
ComMLOps
Compliance
DevOps

Disclaimer

- views are my own
- no claim to completeness of legal & regulatory requirements



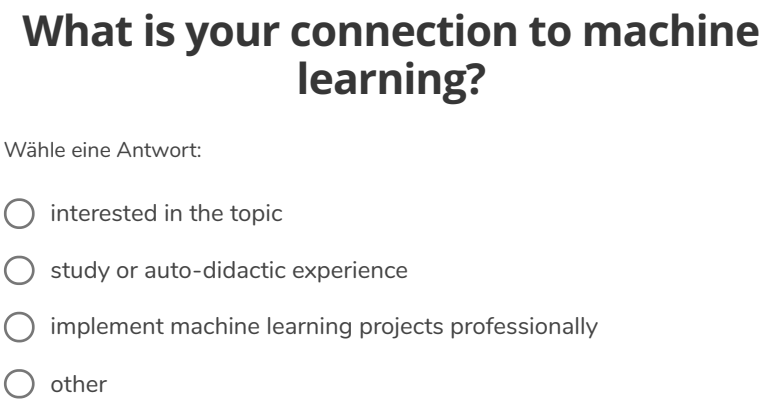
Poll

https://strawpoll.com/kf252h78r

In [1]:

```
from IPython.display import IFrame
IFrame('https://strawpoll.com/embed/kf252h78r', width=700, height=350)
```

Out[1]:



Before the project starts

Data Protection

Clarify legal basis for your purpose (**Art. 6 GDPR**):

- data subject has given consent to the processing
- fulfillment of a contract
- legal requirements
- vital interests of the data subjects
- performance of a task in the public interest
- legitimate interest

Data protection impact assessment usually required, as "automated processing" (Art 35(3) GDPR). The lists of the state data protection authorities must also be taken into account (e.g. **BayLDA: List of processing activities for which a DSFA must be performed**).

Processing of employee-related data

The workers' council must be informed about the processing of employee data within the framework of co-determination (**Section 87 (1) No. 6 German Works Constitution Act**).

Implementation of regulatory requirements in machine learning projects

https://github.com/mbunse/mlcomops

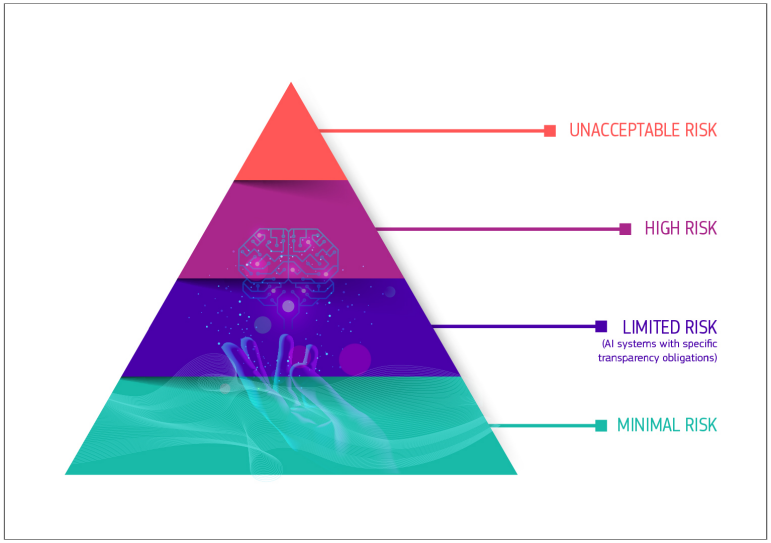
Requirements

Position paper of the Data Protection Conference on recommended technical and organizational measures for the development and operation of AI systems as of Nov. 2019.
Required in the position paper (excerpts):

- Documentation of the selection of the AI process (balancing traceability and required power).
- Preservation of availability of raw and training data.
- Prevention of unauthorized manipulation of AI components
- Possibility for data subjects to obtain information on how decisions and predictions were made
- Monitoring of the behavior of the AI component
- Regular testing of the AI component for discrimination and other undesirable behavior
- Regular testing of the quality of the AI system and its AI components on the basis of operational data.

EU regulation

Proposal for a Regulation laying down harmonised rules on artificial intelligence



High risk e.g.

- AI in road traffic
- Credit scoring

BaFin

15.06.2021 | Topic Digitalisierung

Big data and artificial intelligence: New paper published by BaFin to outline principles

On 15 June 2021, BaFin published supervisory principles for the use of algorithms in decision-making processes by financial institutions. Those principles are intended to promote the responsible use of big data and artificial intelligence (BDAI) and facilitate control of the associated risks. Financial market institutions are increasingly using technologies such as BDAI. In its 2018 study, " Big data meets artificial intelligence", BaFin noted that while BDAI applications would open up opportunities for institutions as well as for consumers, the risks that might be involved with them had to be kept in check (BaFin Perspectives Issue 1 | 2018)

Supervisory Principles for Big Data and AI from 6/15/2021

Aspects to be highlighted:

- **Reproducibility**: versioning of data and code:
    - Maintaining availability of raw and training data.
- **Experiment Tracking**:
    - Documentation of the selection of the AI procedure.
    - Evaluation of the selected AI procedure with respect to alternative, more explainable AI procedures.
- **Fairness**:
    - Periodic testing of the AI component for discrimination and other undesirable behavior.
- **Model Explainability**:
    - Ability to provide information to affected parties on how decisions and predictions were made.
- **Monitoring**:
    - monitoring of the behavior of the AI component

Poll

https://strawpoll.com/daprodsdy

In [2]:

```
from IPython.display import IFrame
IFrame('https://strawpoll.com/embed/daprodsdy', width=700, height=350)
```

Out[2]:

Which topic interests you the most?

Wähle eine Antwort:

- Reproducibility
- Experiment tracking
- Fairness
- Explainability
- Monitoring

Reproducibility

Data set

Load data

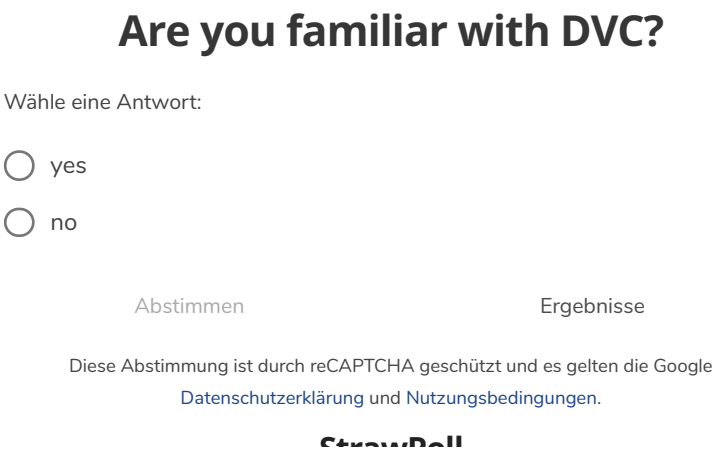
Poll

https://strawpoll.com/kc8pxhafz

In [3]:

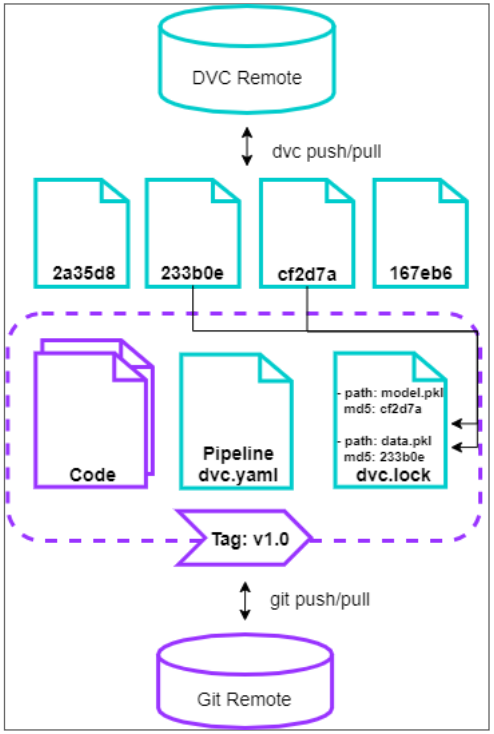
```
from IPython.display import IFrame
IFrame('https://strawpoll.com/embed/kc8pxhafz', width=700, height=350)
```

Out[3]:



Data Versioning,

e.g. with **DVC**



DVC Data

DVC Remote with Minio

In [ ]:

```
! cd .. & dvc pull
```

In [ ]:

```
! cd .. & dvc repro
```

DVC Data

In [ ]:

```
! cd .. & dvc push
```

Set all random seeds

cf. **"Confusion about R-value calculation of the RKI"**

> *Within this simulation, random numbers are drawn that will result slightly different each time the program is run and therefore cannot be exactly reproduced.*

Experiment Tracking

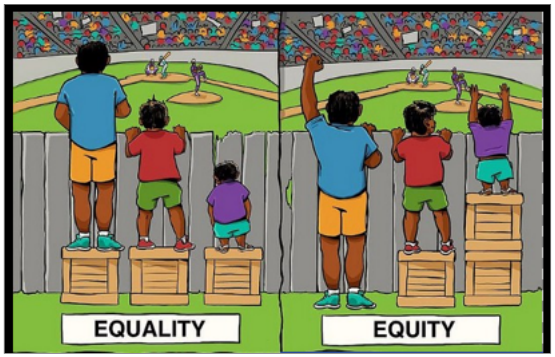
Experiment Tracking z.B. mit **MLflow**
Modell Training
MLFlow

Fairness in Machine Learning Projects

Prohibition of discrimination according to **§ 19 AGG**.

Model Training

https://towardsdatascience.com/real-life-examples-of-discriminating-artificial-intelligence-cae395a90070



Modell Training

Explainability

Explainer

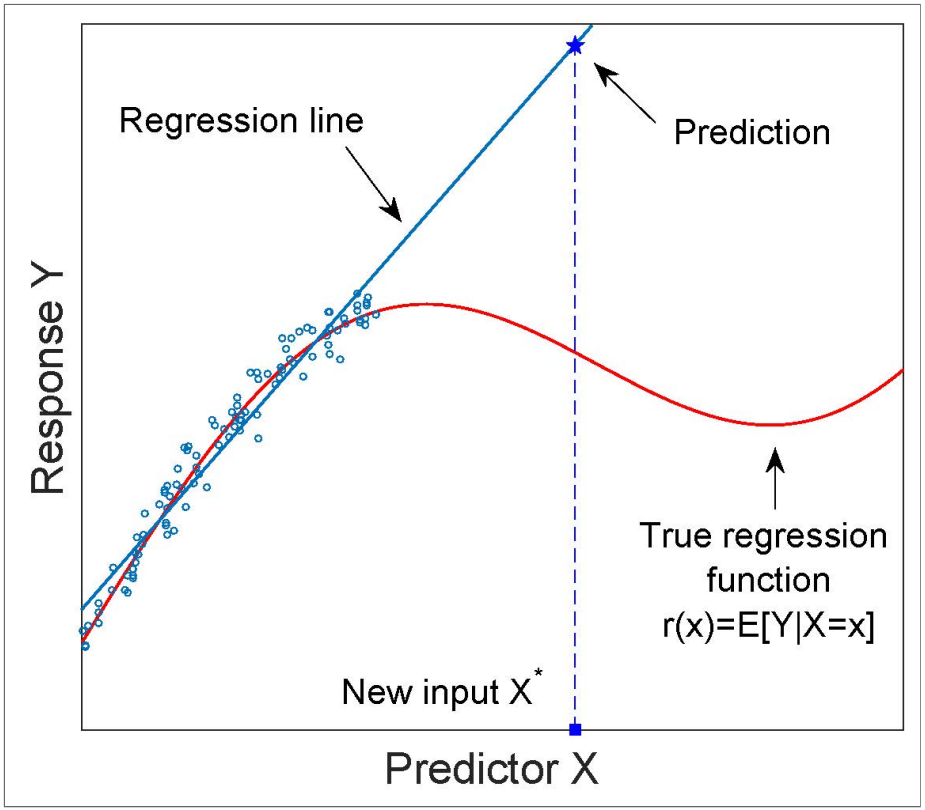
Monitoring

Model API
Metrics Endpoint
Grafana
Call API

Monitoring


General / Model Score
Model
Avg. Score in last 10 min
0.807
Label
1
0
22:08
22:09
22:10
22:11
22:12
0
50
105
Score Distribution
+Inf
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
22:08
22:09
22:10
22:11
22:12
0
20
40
60
74
Outlier Score
+Inf
1.7
1.5
1.3
1.1
0.9
0.7
0.5
0.3
0.1
22:08
22:09
22:10
22:11
22:12
0
20
40
60
68
Avg Outlier Score
1
0.500
0
-0.50
-1
22:08
22:09
22:10
22:11
22:12
outlier score

Outlier Detection



<u>Outlier Detector</u>

Drift detection

Drift detection

Outlier Detection

Call API with outliers

Automation

In [4]:

```
%%html
<blockquote class="twitter-tweet"><p lang="en" dir="ltr">Machine learning pipelines <a href="https://t.co/5FpG3HrdW0">pic.twitter.com/5FpG3HrdW0</a></p>&mdash; AI Memes for Ar
```

AI Memes for Artificially Intelligent Teens
@ai_memes

Machine learning pipelines

6:45 PM · Apr 14, 2021

What else?

- Data Science development environment
- Pull Requests
- Test Automation
    - Unit Tests
    - Integration tests
- Scaling (e.g. with Kubernetes)
- Staging
- CI/CD
- security
- ...