Data Science Academy

# Data Science for Business – Becoming a Data Science Expert (D)

Pilot Presentation:
for participants of and use in the pilot only

22.09.2021

Boehringer Ingelheim

# Agenda week four

**Introduction**

| | |
|---|---|
| 1 | Recap Basic Machine Learning and Python |
| 2 | Complex Models |
| 3 | Model Evaluation |
| 4 | Hyperparameters |
| 5 | Unsupervised Learning |
| 6 | Gradient Descent |
| 7 | Deep Learning and Image Recognition |
| 8 | Deep Learning and Natural Language Processing |
| 9 | Repetition |
| **10** | **Bias and Ethics in Machine Learning** |
| **11** | **Introduction to Data Science with AWS** |

# Schedule week four

| Week 4 | | | |
|---|---|---|---|
| | Day 1<br>Tuesday, 21.09.2021 | | Day 2<br>Wednesday, 22.09.2021 |
| Start: 12:00 | Recap | Start: 12:00 | Recap |
| | 9 – Repetition (Part 1) | | 10 – Bias & Ethics in ML |
| | | | 11 – Introduction to Data Science with AWS (Part 1) |
| 14:00 – 15:00 | Break | 14:00 – 15:00 | Break |
| | 9 – Repetition (Part 2) | | 11 – Introduction to Data Science with AWS (Part 2) |
| End: 18:00 | Q&A and Feedback | End: 18:00 | Q&A and Feedback |

We will also have several short coffee breaks in between.

# Feedback for pilot training

We aim to provide a great training experience for you and are looking forward to receiving your feedback!

You will have three different ways to give us your feedback on each training day:

1. We will have an **anonymized** feedback collection **after the last session** of each day per **Myforms.**
2. We will have an **open feedback round and discussion** at the **end of each training day.**
3. Please also **take notes** regarding your ideas during the sessions: **locally or via the Mural Board** which you can reach via LINK.

Module 10

# Bias & Ethics in Machine Learning

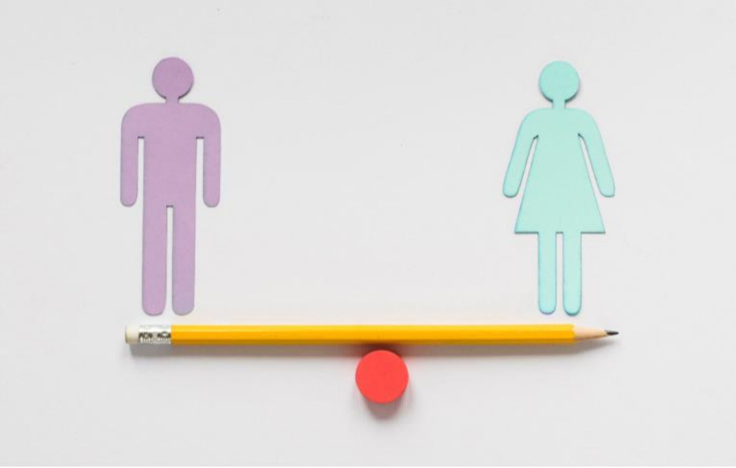
# Agenda week four

# AI being applied more and more for critical decisions

Justice | Financial crime | Self-driving | Lending | Health care

## … with some embarrassing failures

Racist Computer Vision



Gender bias in hiring

Unexplained credit ratings

# Introduction to ethics and bias

## Ethics

### Definition

- Doing what is morally right
- Good ethics are normally aligned with what the law requires
- Conform to the standards of behaviour that society accepts/social norms
- Tricky to define on a common definition, such as morality of death penalty

### Examples

- Unauthorized collection and/or misuse of personal data
- Build responsible algorithms based on code of ethics, for example "predictive policing" should be free of discrimination
- Maintain transparency in coding for interpretation

## Bias

### Definition

- Systematic and repeatable errors in a computer system that create unfair outcomes
- Can be human and/or content bias
- Cognitive bias: Preferential or skewed thinking
- Confirmation bias: Consciously present data or models to confirm preassumed hypothesis
- Selection bias: Data or results are selected subjectively

### Examples

- Data collected for a particular observation can be different depending on data sources
- Algorithmic bias – intentionally or unintentionally build algorithms to marginalize a target group, e.g.: hiring algorithms built to prefer male applicants due to historical data

**Aspects in ethics and bias should be considered in a DS project, especially with regards to data and algorithms.**

# Practicing data ethics

**Security:**
- Have we created an overview of potential attacks or abuses of the technology?
- Have we considered how to protect and secure user data?

**Privacy:**
- What kind of user consent do we need to use the data?
- Do we have a process or mechanism for collecting consent from users?

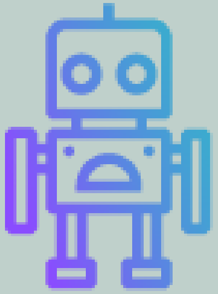
Four
areas to consider

**Fairness:**
- Have we made sure that our training data is fair and representative?
- Have we conducted tests for disparate error rates among different groups of users?
- Do we reflect a diversity of opinions, backgrounds, and kinds of thought?

**Transparency:**
- Did we declare clearly what the users consent to?
- Can we review the codes in case of doubts?

**Constantly ask questions in four areas of ethical considerations to be fair and conclusive.**

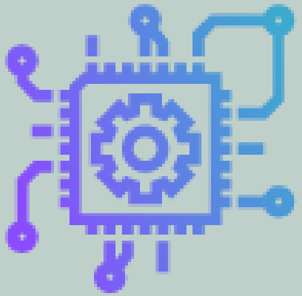
# The ethical debate on AI is getting more and more relevant!



**9** out of **10**

Organizations worldwide struggle with ethical challenges during the use of AI*

* Capgemini Research Institute, Ethics in AI executive and consumer survey, N = 1,580 executives, 510 organizations



**76%**

of users expect a new regulation for AI use cases**

** Capgemini Research Institute, Ethics in AI executive and consumer survey, N = 4,446 consumers, N=1,580 Executives

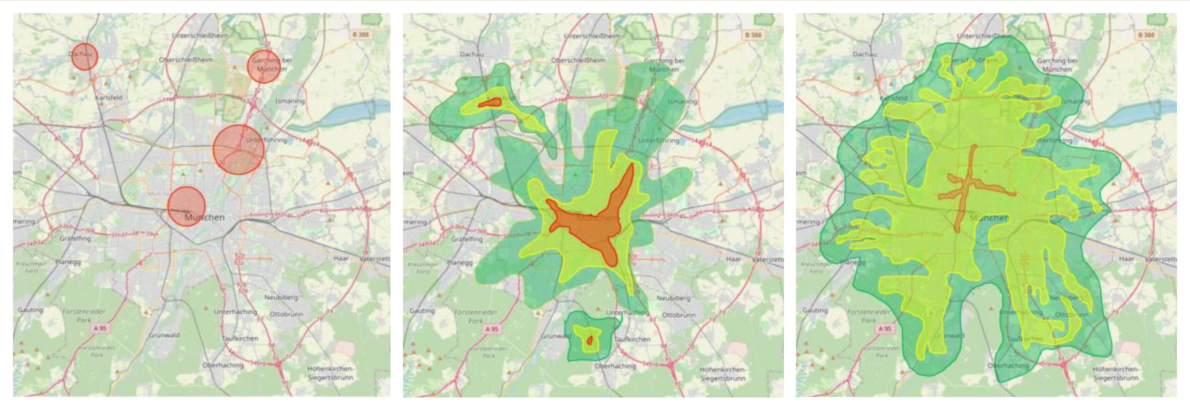
**Ethical requirements are crucial for a successful deployment.**

# Content bias

If analysing data from different sources, that relate to e.g., the same geographical area, consider **bias effects.**

## User activity on image-posting social media

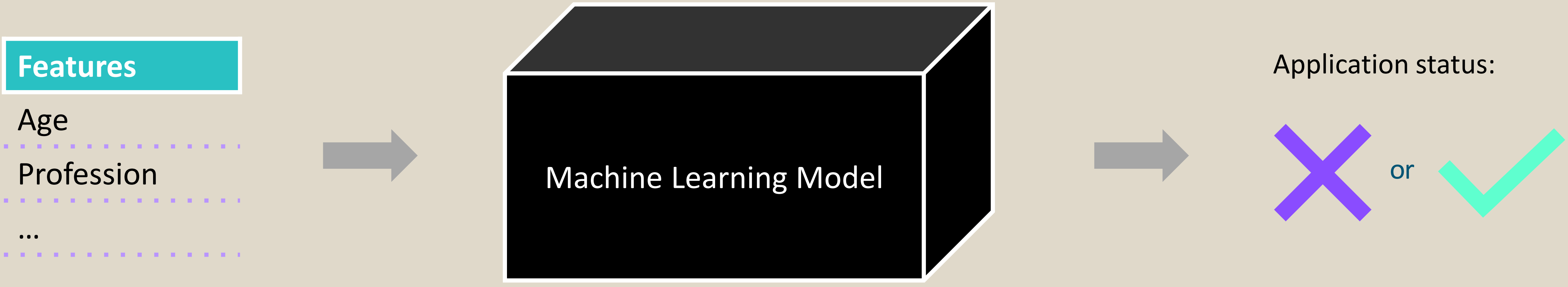
- Data based on the city of Munich
- Compared platforms:
  - Foursquare
  - Flickr
  - Instagram
- Distribution and intensity of the amount of posted images differs strongly
- Bias effect might have an influence on analysis



**Consider bias effects during data collection and selection.**

# Complex Machine Learning models are a black-box

- In machine learning, “black box” describes models that cannot be understood by looking at their parameters (e.g. a neural network)
- Consider the case of loan application where an advanced models decides, if an applicant gets granted a loan. Wouldn’t you want to know why your application got denied?




**Input** **What did the model learn?** **Why was my application denied?**

# When the model learns the wrong things

Discrimination through NLP:

- Word embeddings
- Auto completion
- Etc

Further biases:

- Racial bias
- Religious bias
- Etc

De-biasing strategy to reduce the polarization:

- Neutralizing hard-biased words

Example for relationships in word embedding:

Woman + king – man = queen

**BUT:**

Woman + doctor – men = nurse

**That's discriminating!**

**Language is powerful: Be careful when applying NLP models in a context that might be impacted by social bias.**

**Existing bias can be amplified through machine learning.**

# Protective attributes

Protective attributes are sensitive attributes.

According to the use case or context they can be specified and not being used in the machine learning model.

Possible determining attributes :

- Age
- Color
- National Origin
- Sex
- Race
- Etc

Proxy attributes:

- Are attributes that offer a proxy to protective attributes
- E.g.: The location can give indication on the race

Example: Gender

We don’t want the hiring algorithm to select based on the sex.

On the other hand, in a medical context we might want to include that information.

# Cognitive bias and fairness metrics

## Cognitive Bias

- Criminological Software can be flawed
- COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) is an algorithm used in the U.S. to predict likeliness of a criminal reoffending.
- Algorithm predicts that black defendants pose a higher risk of reoffending

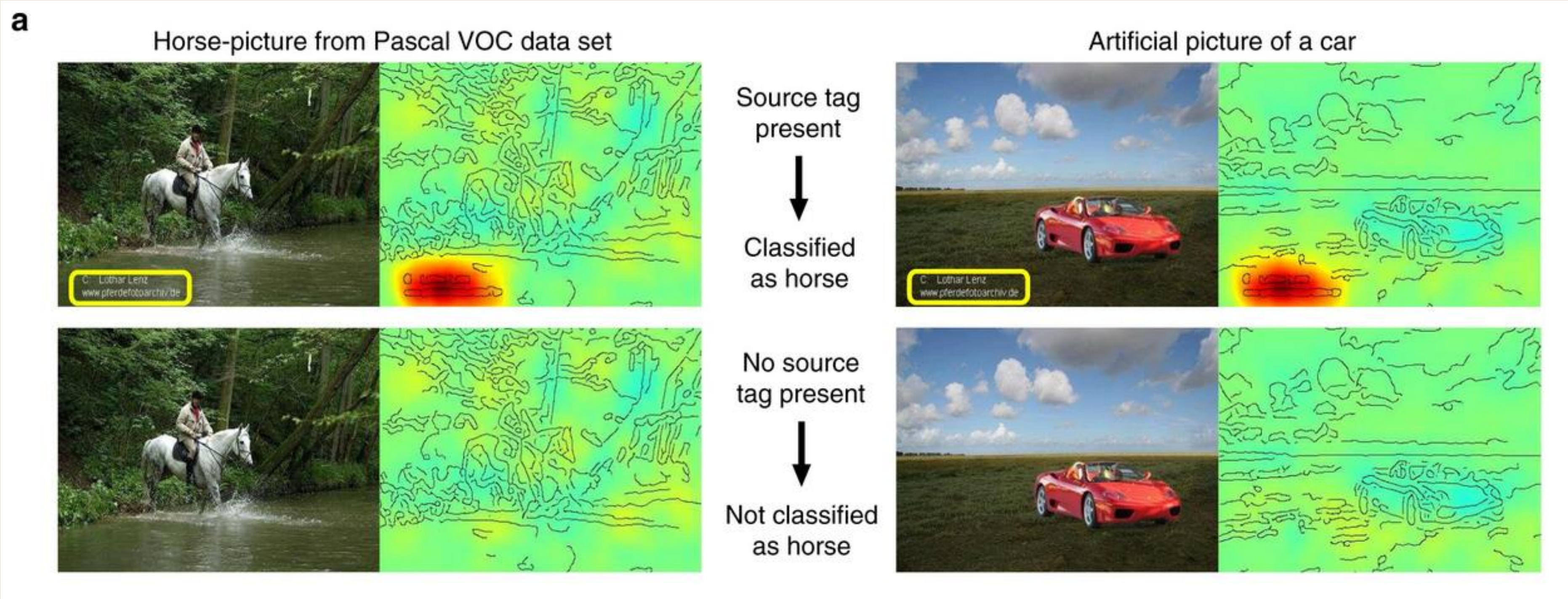
## Fairness metrics

- Fairness metrics can be applied to predicted outcomes, actual outcomes and predicted and actual outcomes
  - **False positive error balance**, e.g.: “Labeled higher risk, but didn‘t re-offend”
  - **False negative error balance**, e.g.: “Labeled lower risk, yet did re-offend”
- The fairness can be applied to the stages of data prepocessing, optimization of training models, post-processing of the results

| | White | African American |
|---|---|---|
| **Labeled higher risk, but didn‘t re-offend** | 23,5 % | 44,9 % |
| **Labeled lower risk, yet did re-offend** | 47,7 % | 28,0 % |

# Examples: When the model learns the wrong things

- A fisher vector based model that was trained on PASCAL VOC 2007 image dataset gained a very high accuracy on the classification of categories, e.g. “person”, “car”, “horse”
- With the help of an explanation method LRP (Layer-wise relevance propagation) it was demonstrated, that the model decisions were not based on the relevant features but rather on the watermark tag in the lower left corner
- Without that tag the model wasn’t able to distinguish between the different categories



# Overview of interpretability methods

1

- Interpretable models
  - Linear/Logistic Regression
  - Decision Trees
  - Naive Bayes Classifier
  - K-Nearest Neighbors

2

- Global model-agnostic methods
  - Partial Dependence Plots
  - Permutation Feature Importance
  - Global Surrogate

**Machine Learning interpretability methods**

3

- Local Model-Agnostic Methods
  - LIME
  - Shapely Values
  - SHAP

4

- Neural Network Interpretation
  - Learned Features
  - Pixel Attribution
  - Adversarial Examples

Further references: https://christophm.github.io/interpretable-ml-book/

# Open source explainability frameworks – LIME and SHAP

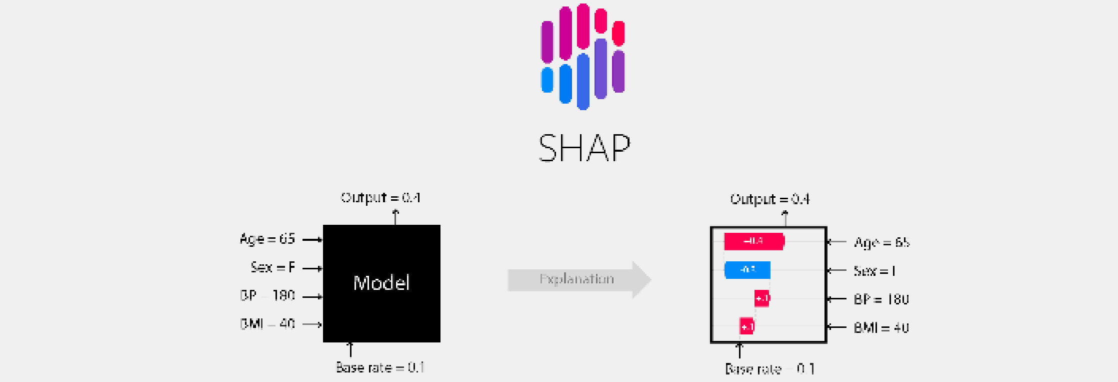
## What

LIME and SHAP are the two most common open source techniques for model explanation. They provide explanations per prediction. They can be black box (no assumption on model) or white box (model-tuned implementation, for instance for tree-based models).

LIME is the older technique, but not always accurate and computationally expensive.
SHAP is mathematically sound, based on information theory.

## Benefits

Provide an explanation per individual prediction, presented as weights on variables
Display in a format which is meaningful to a non-technical user.




## Where

Available from Github, as a Python library and on most machine learning packages / offerings.

## Limitations

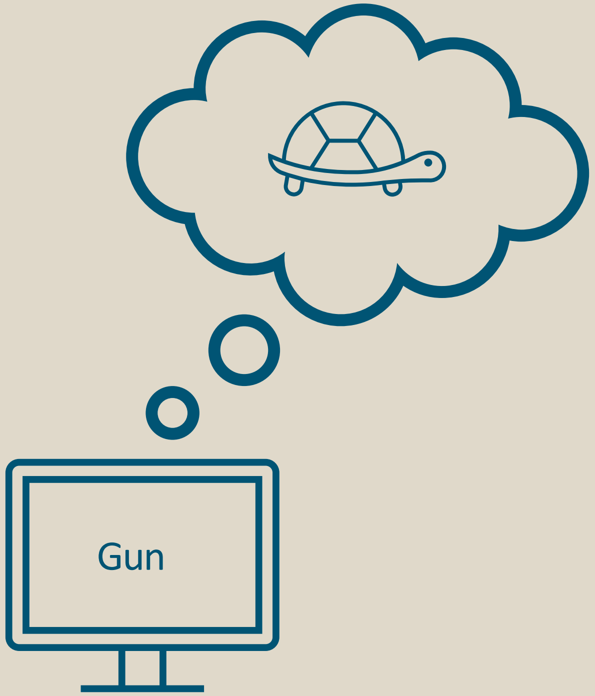
LIME and SHAP provides only explanations per variable, and might be unstable with correlated variables.

# Data Scientists need to get out of the lab and hit the real world

## Data from the real world might contain unforeseen examples


https://arxiv.org/pdf/1707.08945.pdf

## Models can be purposefully manipulated




Athalye, Anish, and Ilya Sutskever. “Synthesizing robust adversarial examples.” arXiv preprint arXiv:1707.07397 (2017)

# Responsible usage of AI

Generative Adversarial Neural Networks (GANNs) allow us to generate realistic looking persons, animals and objects

With this technology we can create face swap apps, create visual effects and help design clothes

But it can also be used to fake presidential speeches, modify videos in real-time and imitate voices of any person

http://Thispersondoesnotexist.com

https://thiscatdoesnotexist.com/

# Intro to GANNs

## Generator

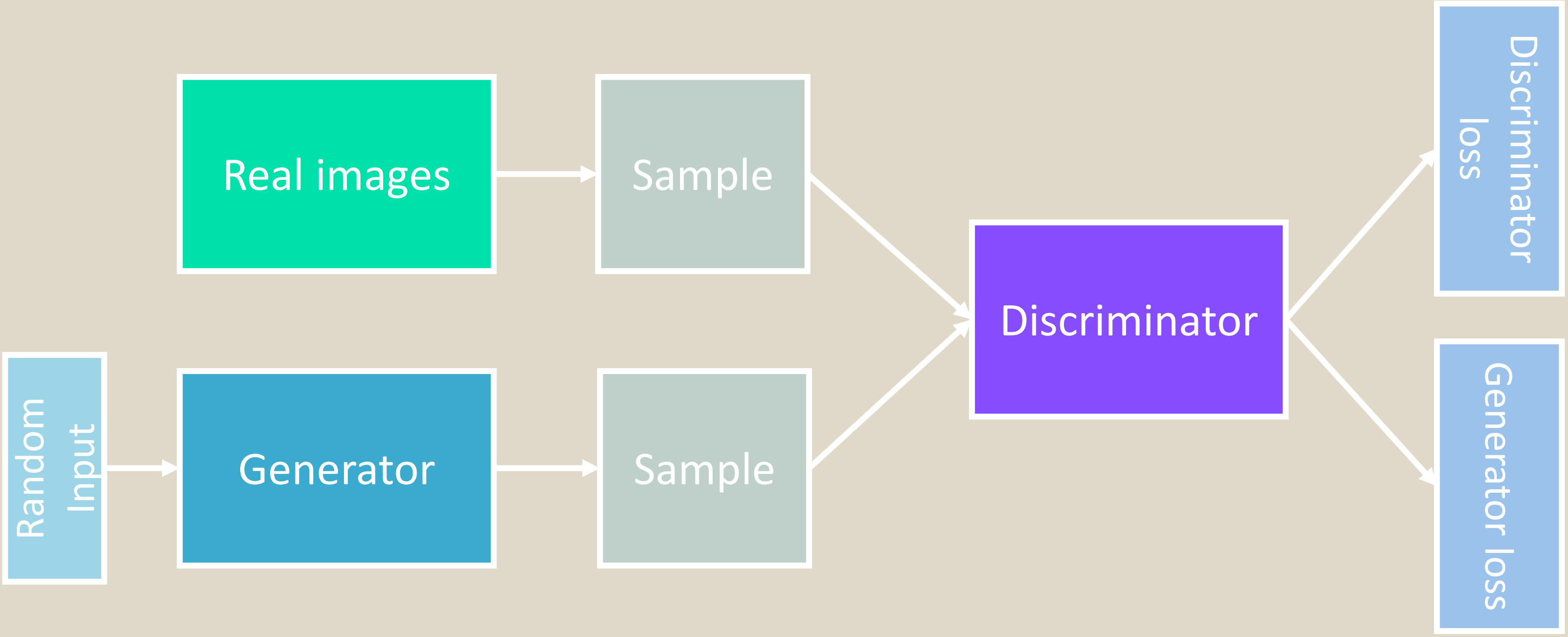
- Generate fake images that are so good they can fool the discriminator
- Reversed Convolution Neural Network that generates from a small input an image

## Discriminator

- Differentiate between fakes and real images
- Convolutional Neural Network predicts a binary output based on the image it sees

## Training process

- Alternate between training the generator and the discriminator
- In each iteration the fakes get more realistic



# Generating photo realistic images from doodles

http://nvidia-research-mingyuliu.com/gaugan/

# Lazy Painters Competition

Try it yourself:
http://nvidia-research-mingyuliu.com/gaugan/

Module 11

# Introduction to Data Science with AWS

# Agenda week four

| | |
|---|---|
| | **Introduction** |
| 1 | Recap Basic Machine Learning and Python |
| 2 | Complex Models |
| 3 | Model Evaluation |
| 4 | Hyperparameters |
| 5 | Unsupervised Learning |
| 6 | Gradient Descent |
| 7 | Deep Learning and Image Recognition |
| 8 | Deep Learning and Natural Language Processing |
| 9 | Repetition |
| 10 | Bias and Ethics in Machine Learning |
| **11** | **Introduction to Data Science with AWS** |

# Why Data Science on AWS?

**On-demand Self-Service:** Provision capabilities as needed without requiring human interaction

**Broad Network Access:** Capabilities are available over the network & accessed through standard mechanisms

**Resource Pooling:** Sense of location independence. Resources are pooled to serve multiple consumers using a multi-tenant model

**Rapid Elasticity:** Capabilities can be elastically provisioned and released to scale rapidly outward and inward with demand.

**Measured Service:** Resource usage can be monitored, controlled, reported & billed.

# Global Infrastructure

**Region**

- A Region is a physical location around the world where data centers are clustered
- Each group of logical data centers is called an Availability Zone (AZ)
- Each AWS Region consists of multiple, isolated, and physically separate AZs within a geographic area

**Availability Zone**

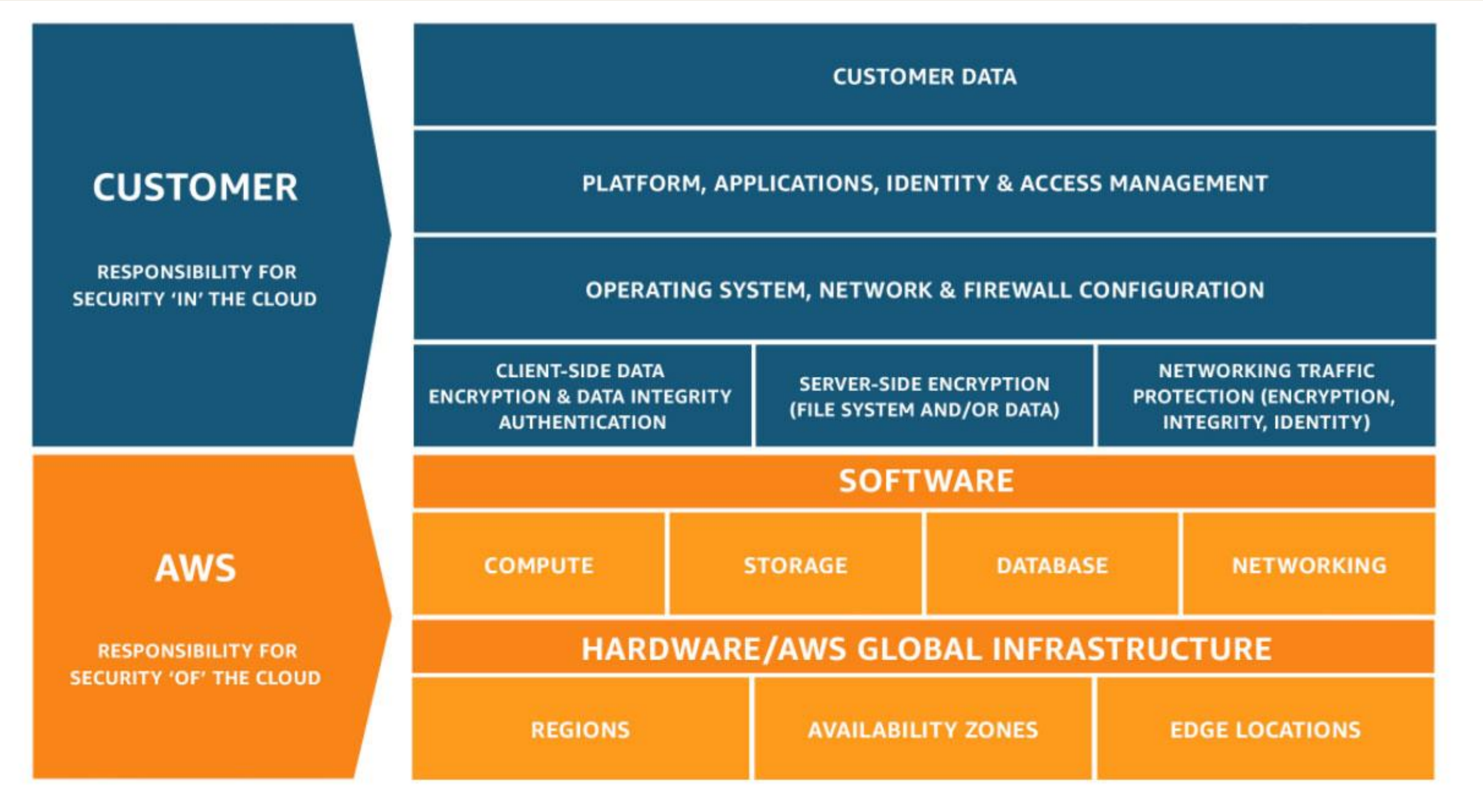
- An AZ is one or more discrete data centers with redundant power, networking & connectivity in a Region
- AZs enable users the ability to operate production applications & databases that are more highly available, fault tolerant & scalable

**Edge Locations**

- Edge locations are AWS data centers enabling reliable, low latency and high throughput network connectivity

https://aws.amazon.com/about-aws/global-infrastructure/regions_az/

# Shared Responsibility Model




https://aws.amazon.com/compliance/shared-responsibility-model

# AWS Services

- AWS Key Services
  - Compute
    - EC2
    - Lambda
  - Storage
    - S3
    - EBS
  - Database
    - RDS
    - DynamoDB
    - Neptune
  - Networking
    - VPC
    - CloudFront
    - API Gateway
  - Developer Tools
    - CodeCommit
    - CodePipeline
    - Cloud9
  - Machine Learning
    - SageMaker
  - Analytics
    - Athena
    - Redshift
    - EMR
    - Kinesis
    - Quicksight
    - Glue
  - Security
    - IAM
    - Cognito
  - Management & Governance
    - CloudWatch
    - CloudTrail
    - CloudFormation

First publicly launched in 2006, initially three services were offered.
Today AWS offers over 200 fully featured services.
Services can be global, regional or AZ specific.

# SageMaker

| Create Notebook Instance | Prepare Data | Train ML Model | Deploy ML Model |
|---|---|---|---|
| • Create the notebook instance that you use to download & process your data | • Use the SageMaker notebook instance to preprocess the data that you need to train your ML model<br>• Upload the data to Amazon S3 | • Use your training dataset to train your ML model | • Deploy the trained model to an endpoint |

https://aws.amazon.com/blogs/machine-learning/load-test-and-optimize-an-amazon-sagemaker-endpoint-using-automatic-scaling/

**Do not forget to clean up when you do not need the model anymore.**

# Hands-On

# Scenario: Employee Turnover Prediction

Employee turnover is costly to businesses.  Companies invest time, money and effort to train new hires to adjust to the new culture and work conditions. Those who resign take with them the experience, the training & the culture.

Newcomers require time to manage existing products, understand work procedures, familiarize with existing systems and culture. Studies show that replacing a junior position may cost up to 40% of an employee’s salary.

In this hands-on, we will predict attrition rate of employees based on a training data-set available for free on Kaggle. This will enable us to put forth recommendations to the management with the most effective way to reduce turnover rate.

**Involved AWS Services: IAM, VPC, S3, CodeCommit, SageMaker, EC2, QuickSight**

# Agenda week four

# Feedback and Q&A

# Thank you

If you would like any further
information please contact
Werner,Dr.,Fabian_Georg (BI X) BIX-DE-I
<fabian_georg.werner@boehringer-ingelheim.com>