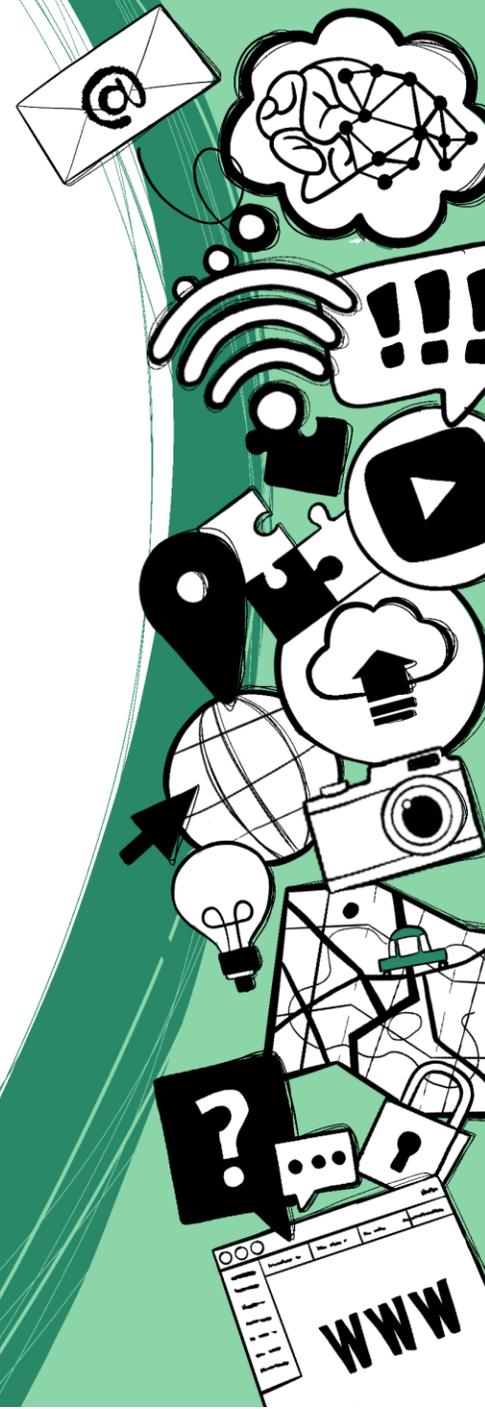


# Supervised Learning Vertrauen auf Daten



# Overfitting (Überanpassung)



# Overfitting (Überanpassung)

## Charakteristika

- Das Modell schneidet **im Training** sehr gut ab



Hund ✓

Katze ✓

**99% Genauigkeit**



# Overfitting

## Charakteristika

- Das Modell schneidet **im Training** sehr gut ab



Hund ✓  
Katze ✓

**99% Genauigkeit**

- Das Modell schneidet bei **verschiedenen Daten** sehr schlecht ab



Katze ✗  
Hund ✗

**30% Genauigkeit**



# Overfitting

## Mögliche Gründe

- Das Modell hat nicht **möglichst allgemeine Eigenschaften** gelernt
- Das Modell konzentriert sich auf die **falschen Merkmale**



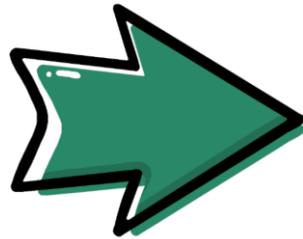
## Beispiel 1

Die meisten Bilder von Hunden im Trainingsatz enthalten Gras, daher **wird jedes Bild mit Gras als Hund klassifiziert.**



## Beispiel 2

Alle Bilder sind **Nahaufnahmen** von Katzen, daher wird ein Bild einer Katze auf einem Feld nicht erkannt.





# Overfitting

Welche Beispiele fallen **dir** ein?

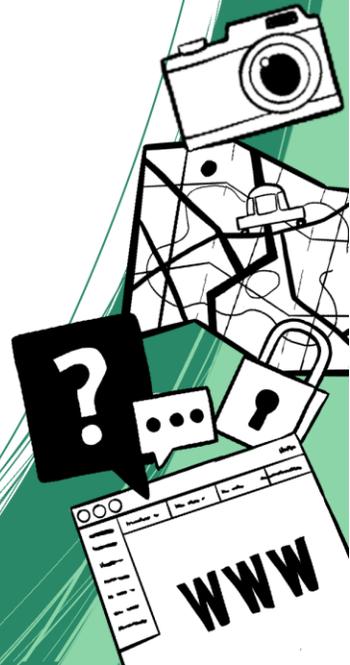


# Overfitting

Welche Beispiele fallen **dir** ein?

- Fokus auf der Hand, die das Objekt hält
- Fokus auf Person oder Objekt im Hintergrund
- Fokus auf Körperhaltung statt Mimik
- Fokus auf Haare und Schmuck statt Gesicht
- Fokus auf Handhaltung statt Gesten
- Alle Trainingsbilder wurden bei Tageslicht aufgenommen

...



# Overfitting

## Mögliche Lösungen

- **Mehr** Daten
- Besser **verallgemeinerte** Daten
- Bearbeitung von Daten zur **Reduzierung unbedeutender Unterschiede**
- **Vielfalt steigern durch:**



spiegeln



skalieren



drehen



Weichzeichnen



entsättigen

...



# Underfitting



# Underfitting

## Charakteristika

- das Modell liefert einfach **keine brauchbaren Ergebnisse**



Katze   
Katze 

50% Genauigkeit



# Underfitting

## Mögliche Gründe

- Zu **wenige** Proben
- Trainingsdaten sind **zu allgemein**
- (Trainingsparameter sind **unpassend gewählt**)



cat



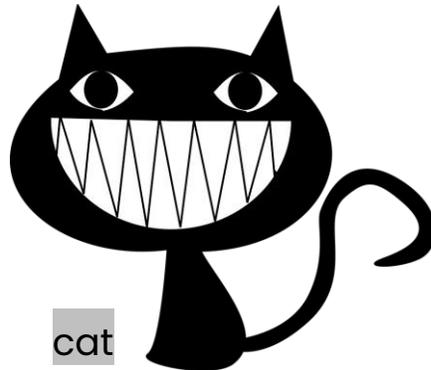
dog

CAT

cat



dog



cat



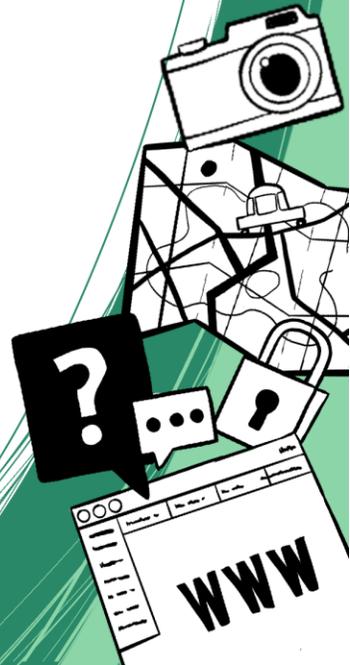
dog



# Underfitting

## Mögliche Lösungen

- **mehr/bessere** Stichproben für Trainingsdaten zu verwenden
- Trainingsparameter **anpassen** (mehr Iterationen, ...)

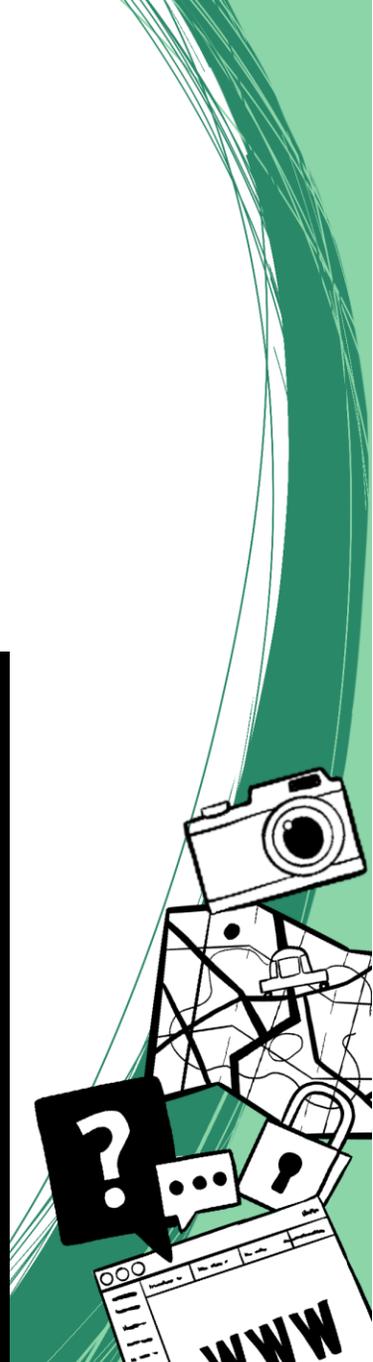


# Biases (Vorurteile) und Fairness



# Biases and Fairness

Stell dir vor, du erstellst einen Chatbot, der sich wie ein normaler Mensch verhalten soll. Deshalb nutzt du Live-Chats von Streaming-Plattformen wie Twitch oder YouTube zum Training.

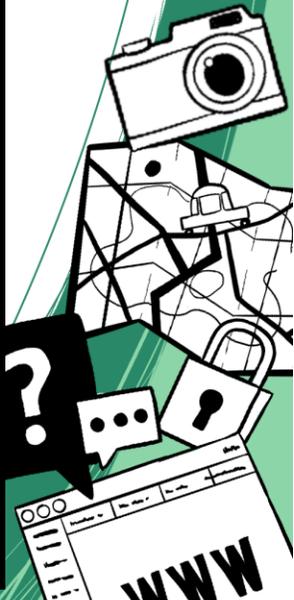


# Biases and Fairness

Stell dir vor, du erstellst einen Chatbot, der sich wie ein normaler Mensch verhalten soll. Deshalb nutzt du Live-Chats von Streaming-Plattformen wie Twitch oder YouTube zum Training.

Was glaubst du, wie gut dieser Bot funktionieren würde?

Siehst du irgendwelche Probleme, die auftreten könnten?



# Biases and Fairness

- Das Netzwerk **repliziert**, was es **gesehen/gelernt** hat
  - Rassismus, Sexismus, Vulgarität, ...
  - Tendenz zu Streamern mit lautem/aktivem Publikum
  - Voreingenommenheit gegenüber Namen und Normen, die zum Zeitpunkt des Trainings relevant sind...

Es ist schwierig, unvoreingenommene Algorithmen zu erstellen!

- Ethische Frage, wo solche Algorithmen eingesetzt werden sollen
  - Wo sollen Maschinen Entscheidungen treffen dürfen?
  - Wie können wir ein faires und unvoreingenommenes Ergebnis garantieren?



# Trainingszeit und Transfer Lernen



# Trainingszeit

- Das Training eines **guten** Modells von Grund auf kann Wochen dauern
- Wenn du die Suche nach den richtigen Parametern mit einbeziehst, kann es sogar Monate oder Jahre dauern...



# Transfer Lernen

- Verwendet ein **bereits trainiertes Modell** als Ausgangspunkt
- Benötigt viel **weniger** Trainingsdaten
- brauchbare Ergebnisse oft mit ein **paar Dutzend Bildern**
- Benötigt viel **weniger Zeit**
- normalerweise Sekunden bis Minuten
- Viele vortrainierte Modelle sind **frei verfügbar**
- insbesondere für die Bilderkennung





# Als nächstes

Trainiere ein **reales Bildklassifizierungsmodell** und verwende es, um ein **Snake-Spiel** zu steuern!

