

26.06.25



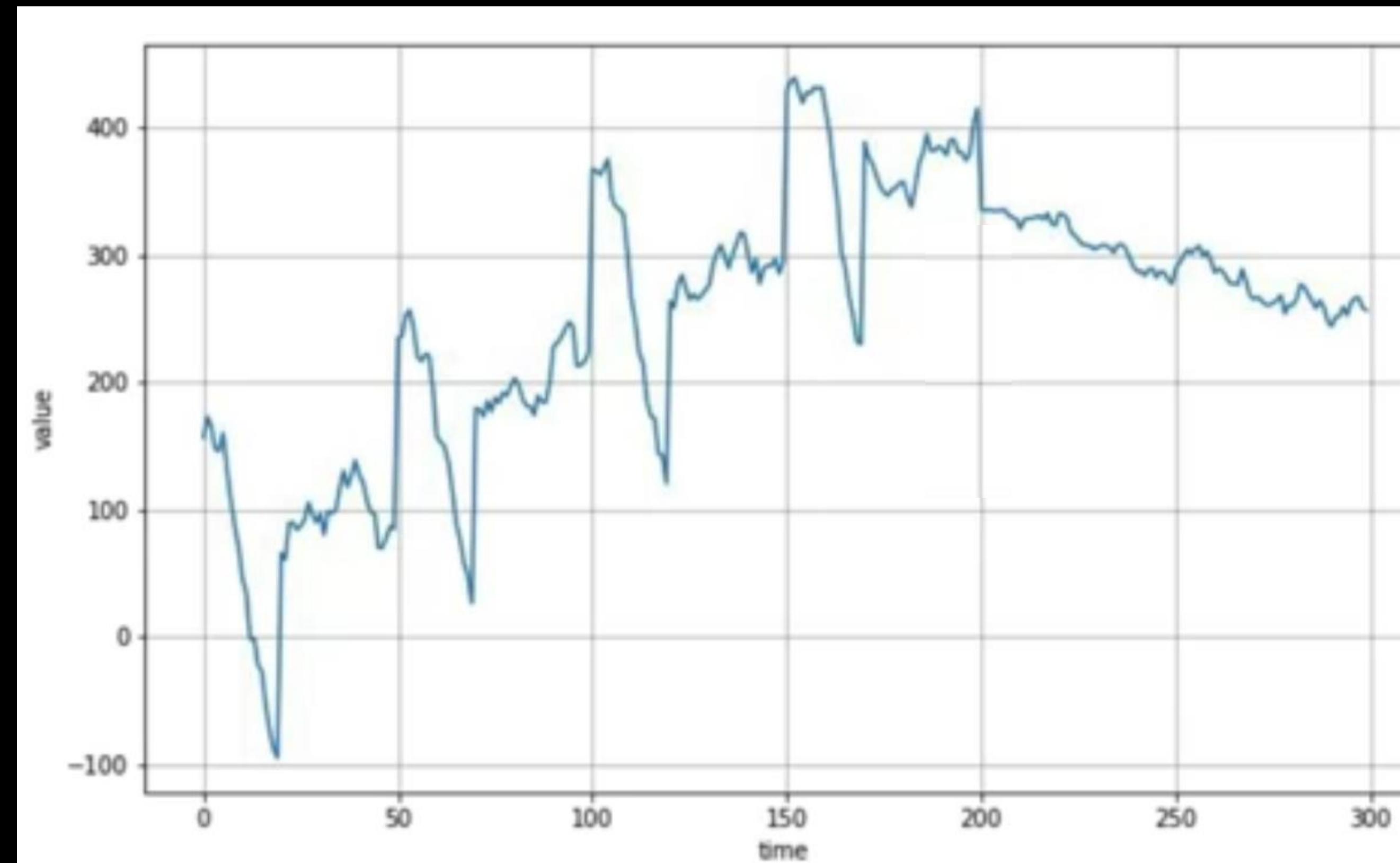
# Introduction to Data Science and Machine Learning

## TIME SERIES ANALYSES

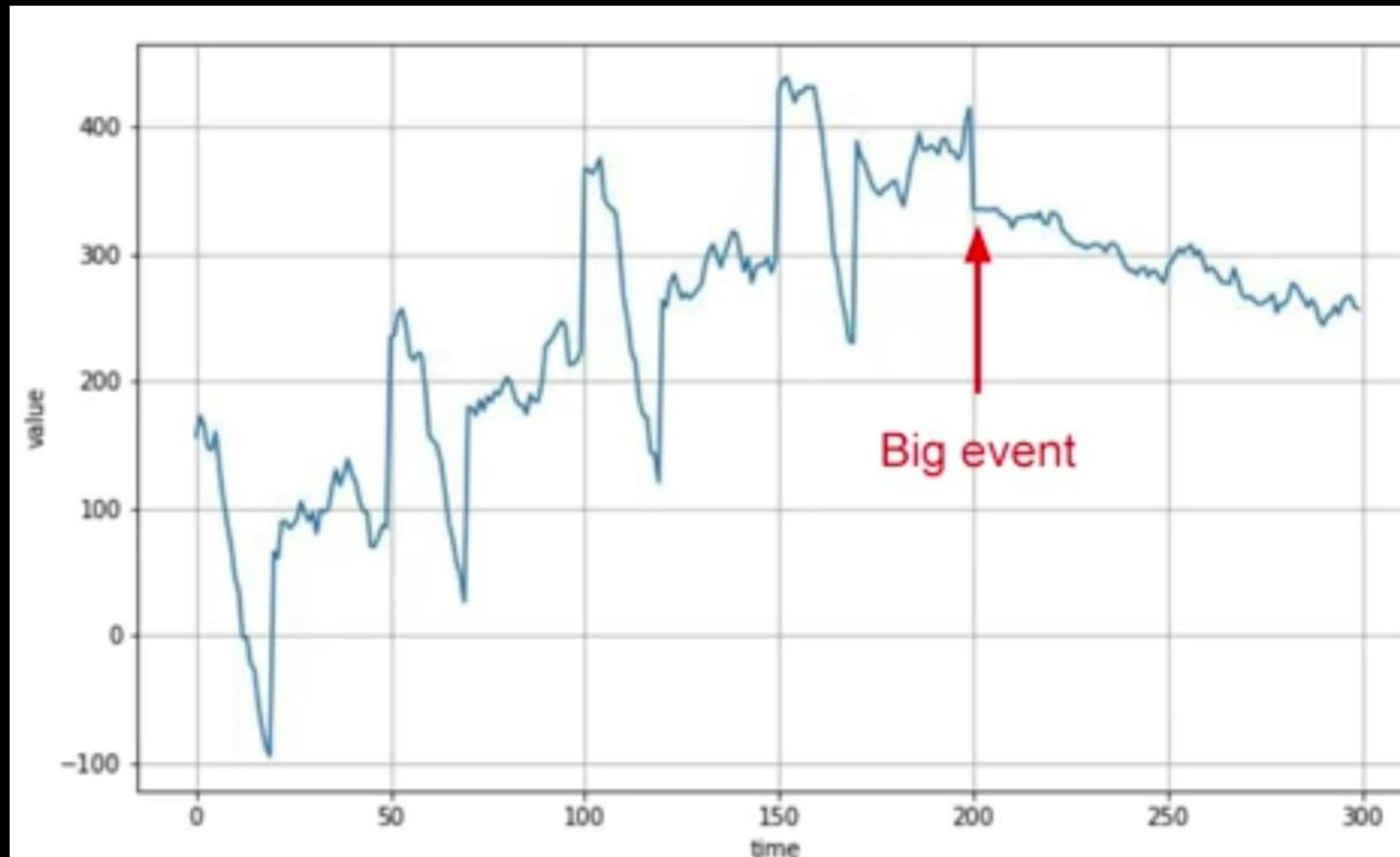
- Patterns in Time Series
- Non-Stationarity
- Baseline Models und Naïve Forecasting
- Project Presentation
- *And now?*

# DISCUSSION

What different types of patterns can be identified in the displayed figure?



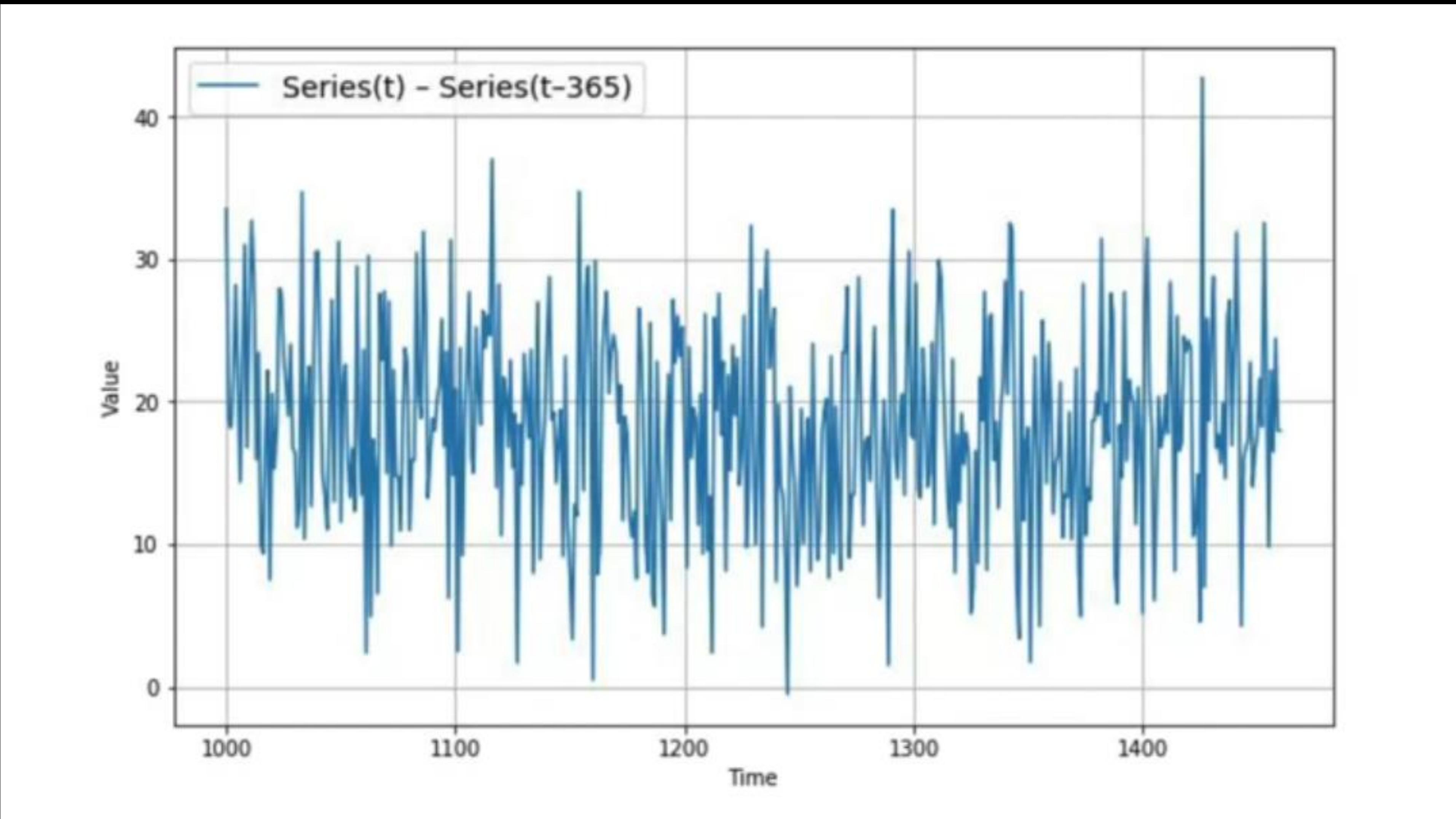
# NON-STATIONARITY



# PATTERNS IN TIME SERIES

- Trends
- Seasonality
- Noise
- Autocorrelation (serial correlation)

# DIFFERENCING



# DIFFERENCING

- Subtracting the previous observation from the current observation
- A method for removing trends, seasonality, or other potential autocorrelations in time series
- Helps to reduce autocorrelation and obtain stationary time series

# „LAG“

- **The difference between consecutive observations is called the Lag-1 difference.**
- **More generally, a Lag-k difference refers to the difference between an observation and the observation k time steps earlier.**
- **The lag value should be chosen based on the temporal structure of the data.**
- **For time series with a seasonal component, the lag is typically set to match the length of the seasonal cycle (e.g., lag = 12 for monthly data with yearly seasonality).**

# BASELINE MODELS

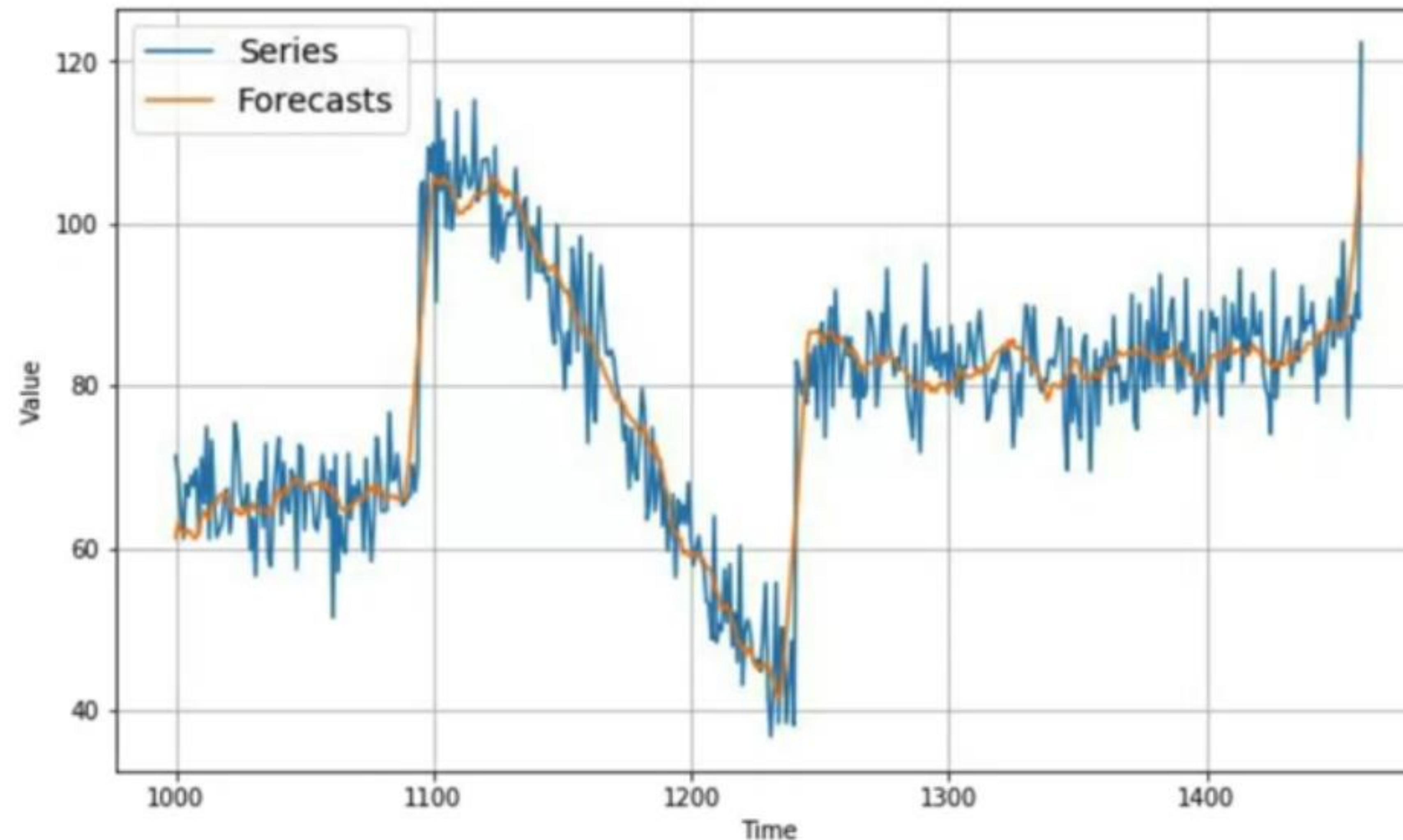
- Metrics like MAPE or RMSE alone often don't reveal how well a model performs.
- Always compare results to a baseline for meaningful evaluation.

## Typical Baselines:

- Previous models on the same dataset
- Models on similar datasets
- For time series: Naïve Forecasting

# NAÏVE FORECASTING

- **Simple baseline model for time series analysis**
- **Forecast = last observed value**
- **Seasonal Naïve Forecasting:**  
**Forecast = last observed value from the same seasonal period**



Forecasts = trailing moving average of differenced series + centered moving average of past series ( $t - 365$ )

# DISCUSSION

**In practice, the focus is often on forecasting the next day, not the whole year.**

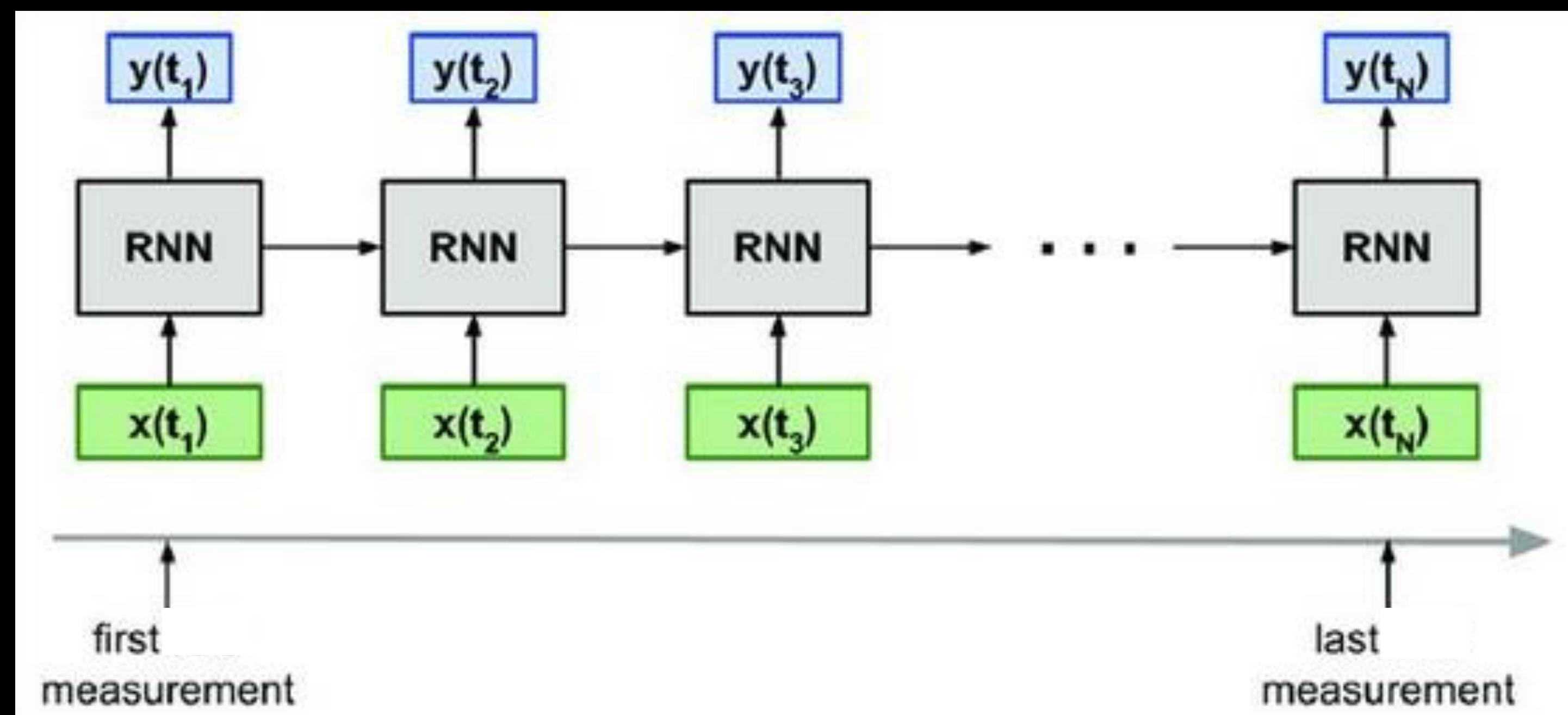
- **Which time series features help most with short-term forecasts?**

# BREAKOUT

- Discuss options to incorporate time series information in your dataset
- Try implementing it

# RECURRENT NEURAL NET (RNN)

- In an RNN layer, the nodes within a layer are interconnected (recurrently connected).



# PROJECT PRESENTATION

## Presentation (Powerpoint, Keybote or similar)

Prepare an 8 to 10-minute presentation including:

- Your team members' names on the title slide
- List and brief description of self-created variables
- Bar charts with confidence intervals for two self-created variables
- Linear model optimization: model equation and adjusted  $R^2$
- Type of missing value imputation used
- Neural network optimization:
  - Source code defining the neural network
  - Loss function plots for training and validation sets
  - MAPE scores for the overall validation set and each product group
- Highlight "Worst Fail" and "Best Improvement" cases

Each team member should have a part in the presentation!

## ***Important Notes***

- **Presentation duration: approx. 8 minutes per team**
- **Every team member should have a part**
- **Use PowerPoint, Keynote, or similar**

## ***Deadlines***

- **By 4 PM next week:**  
**Generate predictions for the Kaggle competition test dataset using your best model and upload them**
- **By July 31, 2025:**
  - **Add your presentation to the repository**
  - **Set the repository to public**
  - **Complete the repository as described in the READMEs**
  - **One team member uploads the main README to the EduHub platform as instructed**

# TASKS

- **Using your best model, generate predictions for the Kaggle competition test dataset and upload them there.**
- **Prepare your final presentation (see the guidelines for Week 10 [here](#)).**

**HOW CAN I MOVE ON IN MY MACHINE  
LEARNING JOURNEY**

=



KI-Forschung

27. Dezember 2024

## Deepseek-V3: Ein 5,6-Millionen-Dollar-Wunder aus China mischt die KI-Elite auf



### THE DECODER Newsletter

Die wichtigen KI-News direkt ins E-Mail-Postfach.

- ✓ 1x wöchentlich
- ✓ kostenlos
- ✓ jederzeit kündbar

E-Mail-Adresse \*

Kostenlos abonnieren

# Stay up to date in Data Science.

Get the Data Elixir newsletter for a weekly dose of the top data science picks from around the web. Covering machine learning, data visualization, analytics, and strategy.

Sign up for Free

and join more than **55,000**  
data lovers today.

Email

SIGN UP

No spam, ever.

What readers say...



Caitlin Hudon  
@beeonaposy



Replies to @beeonaposy and @joecrobak  
Best wide-ranging sampler: Data Elixer

@lonriesberg compiles a broad look at what's happening in data, with titles like 'Modeling coronavirus. How to put R in production. Ray tips & tricks. Mathematics for ML. Finding your way in ML. Distrusting data.'



Julia Evans  
@b0rk · Nov 11, 2018



do you subscribe to a mailing list you think is really useful? what is it?



Laurence Watson  
@LaurenceWWatson

I find @dataelixir by @lonriesberg excellent; not too much and quality stuff



4



See Laurence Watson's other Tweets



Datenschutzerklärung ·  
Nutzungsbedingungen

Explore topics

Machine Learning

Artificial Intelligence

Technology

Data Science

Deep Learning

AI

Programming

Python

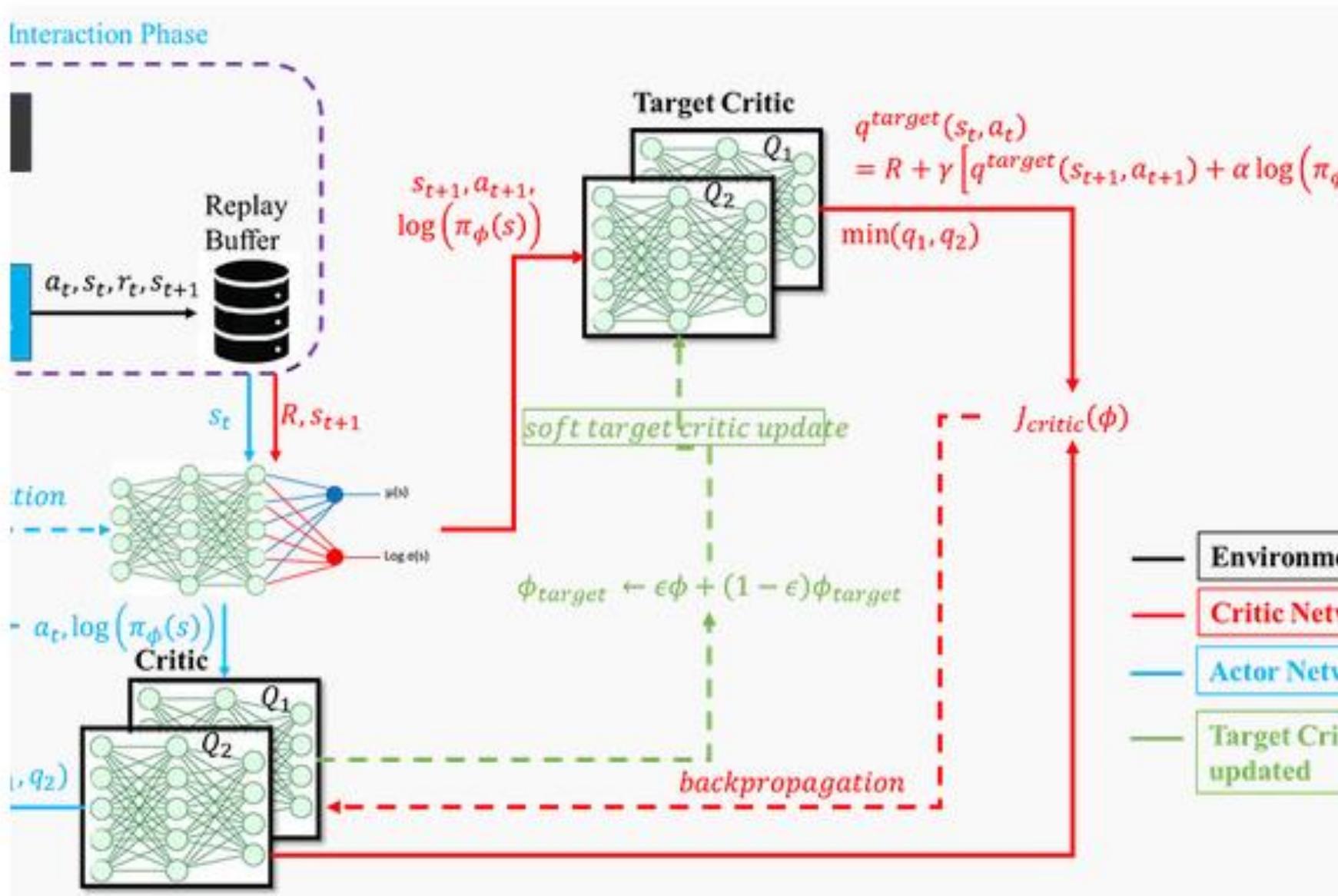
Re >

# Machine Learning

Topic · 5.1M Followers · 361K Stories

Follow

## Recommended stories



In AI Advances by Najib Sharifi, Ph.D.

Mathematical Foundation Underpinning



In The Generator by Jim the AI Whisperer

People considered “at risk” are more likely to get



## Stay on top of the latest in AI

Get a 5-minute email summary of the latest breakthrough news, models, research, and repos.

Your best email...

GET MY SUMMARY

[Read the latest email](#)



Join 202.021 world-class developers from:



- Home
- Beliebt
- Erkunden
- Alle

INDIVIDUELLE FEEDS

- + Individuellen Feed erstellen

Zuletzt besucht

- r/MachineLearning

Communities

- + Erstelle eine Community
- r/MachineLearning ★
- r/OpenAI ★
- r/nlpclass ★

MEHR WISSEN

- Über Reddit
- Werben
- Hilfe
- Blog
- Karriere
- Presse
- Communities

# r/MachineLearning

Heiß ▾ ▾

Community-Highlights

[D] Simple Questions Thread  
0 Votes • 4 Kommentare

[D] Monthly Who's Hiring and Who wants to be Hired?  
31 Votes • 2 Kommentare

u/alexsh1 vor 2 Std.

## [R] Numerical features with factorization machines

Happy to share our recent [TMLR paper](#), "Function Basis Encoding of Numerical Features in Factorization Machines", by Alex Shtoff, Elie Abboud, Rotem Stram, and Oren Somekh. This paper proposes an interesting insight into the interplay between Factorization Machines (FMs), and feature encoding using basis functions, in the context of recommender systems. The same interplay with linear models is an old classic, and most of us have learned in our ML 101 courses. Polynomial regression is one of them - we encode a feature  $x$  using the standard polynomial basis  $\{1, x, x^2, \dots\}$ . FMs are family of models that model a quadratic polynomial...

Research

↑ 9 ↓ 0 Teilen

u/stonedoubt vor 12 Std.

## [R]AST+Shorthand+HybridRag

Research

TEELINEAST: EFFICIENT CODE-SPECIFIC RAG THROUGH STRUCTURAL COMPRESSION AND MULTI-MODAL RETRIEVAL

Shawn McAllister  
Engineered Automated Systems for Artificial Intelligence  
Kannapolis, NC 28081  
(Shawn McAllister) [essaifounder@icloud.com](mailto:essaifounder@icloud.com)

December 29, 2024

ABSTRACT

Current Retrieval-Augmented Generation (RAG) systems treat source code as natural language text, leading to inefficient storage, imprecise retrieval, and loss of structural information. We present TeelineAST, a novel approach that leverages Abstract Syntax Tree compression with multi-modal storage and retrieval mechanisms. Our theoretical framework demonstrates potential improvements of up to 90% in storage efficiency, 5x in retrieval speed, and 98% in accuracy compared to traditional approaches. TeelineAST introduces a paradigm shift in code-specific RAG systems by preserving

**Machine Learning**  
Beginners -> /r/mlquestions , AGI -> /r/singularity, career advices -> /r/cscareerquestions, datasets -> r/datasets

Erstellt am 29. Juli 2009

Öffentlich

**2,9 Mio. 128 Oberste 1 %**  
Mitglieder Online Nach Größe sortieren

COMMUNITY-ERFOLGE

Langzeitmitglied

1 freigeschaltet

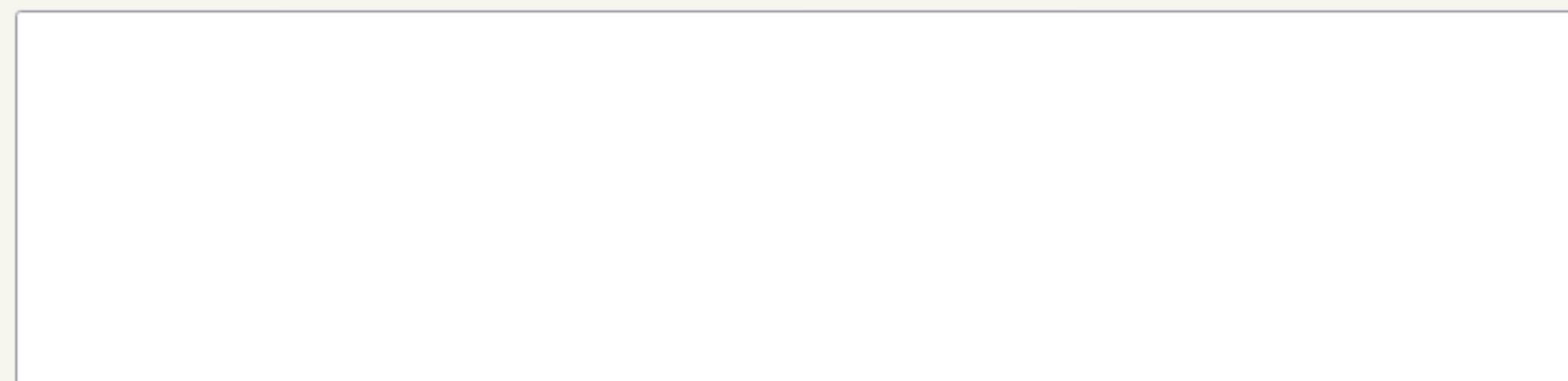
Alle anzeigen

REGELN

- 1 Be nice: no offensive behavior, insults or attacks
- 2 Make your post clear and comprehensive
- 3 Posts without appropriate tag in title will be removed
- 4 Beginner or career related questions go elsewhere
- 5 Non-arxiv link posts only allowed on weekends (must be demos)\*
- 6 Beginner's tutorials and projects go elsewhere
- 7 Quality Contribution
- 8 Limit self-promotion

▲ Safe Superintelligence Inc. (ssi.inc)

1023 points by nick\_pou 16 hours ago | hide | past | favorite | 860 comments



[add comment](#)

▲ insane\_dreamer 15 hours ago | next [-]

I understand the concern that a "superintelligence" will emerge that will escape its bounds and threaten humanity. That is a risk.

My bigger, and more pressing worry, is that a "superintelligence" will emerge that does not escape its bounds, and the question will be which humans control it. Look no further than history to see what happens when humans acquire great power. The "cold war" nuclear arms race, which brought the world to the brink of (at least partial) annihilation, is a good recent example.

Quis custodiet ipsos custodes? -- That is my biggest concern.

Update: I'm not as worried about Ilya et al as commercial companies (including formerly "open" OpenAI) discovering AGI.

[reply](#)

▲ benreesman 5 hours ago | parent | next [-]

It's just clearly military R&D at this point.

And it's not even a little bit controversial that cutting edge military R&D is classified in general and to an extreme in wartime.

The new thing is the lie that it's a consumer offering. What's new is giving the helm to shady failed social network founders with no accountability.

These people aren't retired generals with combat experience. They aren't tenured professors at Princeton IAS on a Nobel shortlist and encumbered by TS clearance.

They're godawful almost ran psychos who never built anything that wasn't extractive and owe their position in the world to pg's partisanship 15 fucking years ago.

[reply](#)

▲ yayr 0 minutes ago | root | parent | next [-]

most technology is dual or multiple use, starting with a rock or knife...

so it is up to the fabric of our society and everyone involved in dealing with the technology, how the rules and boundaries are set.

that there will be military use is obvious. However, it is naive to think one can avoid military use by others by not enabling oneself for it.

[reply](#)

▲ sgregnt 3 hours ago | root | parent | prev | next [-]

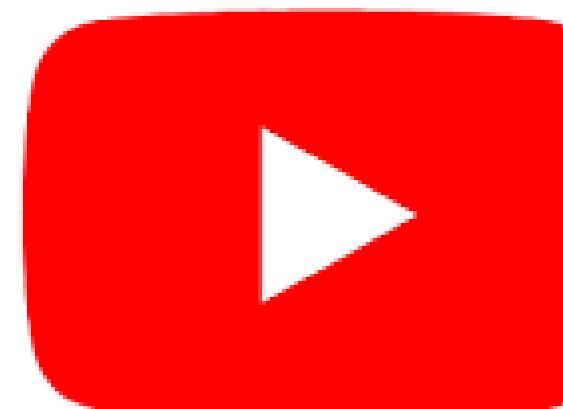
To me it is not clear at all, can you please elaborate why you make such a strong claim?

# LINKEDIN

- **Philipp Schmid**
- **Lior Sinclair**
- **Hugging Face**
- **NVIDIA**
- ...



Udemy



YouTube

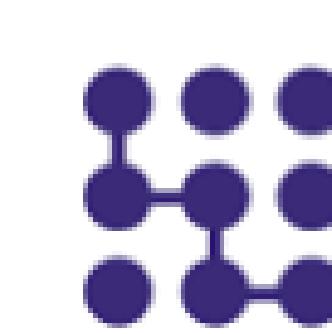
coursera

UDACITY

OPEN



Hasso  
Plattner  
Institut



KI-Campus

Die Lernplattform  
für Künstliche Intelligenz

# All Courses

Grow your AI career with foundational specializations and skill-specific short courses taught by leaders in the field.

 Search

Filters

Clear All

Course Type ⓘ

 Short Course

Most Popular

Top Rated



Short Course

**AI Python for Beginners**

Learn Python programming with AI assistance. Gain skills writing, testing, and



Short Co

ChatGP

Developer

Learn



# The AI community building the future.

The platform where the machine learning community collaborates on models, datasets, and applications.

[Explore AI Apps](#)[or Browse 1M+ models](#)[Tasks](#) [Libraries](#) [Datasets](#) [Languages](#) [Licenses](#) [Other](#)[Filter Tasks by name](#)

Multimodal

[Text-to-Image](#) [Image-to-Text](#)  
[Text-to-Video](#) [Visual Question Answering](#)[Document Question Answering](#) [Graph Machine Learning](#)

Computer Vision

[Depth Estimation](#) [Image Classification](#)  
[Object Detection](#) [Image Segmentation](#)  
[Image-to-Image](#) [Unconditional Image Generation](#)  
[Video Classification](#) [Zero-Shot Image Classification](#)

Natural Language Processing

[Text Classification](#) [Token Classification](#)  
[Table Question Answering](#) [Question Answering](#)  
[Zero-Shot Classification](#) [Translation](#)  
[Summarization](#) [Conversational](#)  
[Text Generation](#) [Text2Text Generation](#)  
[Sentence Similarity](#)

Audio

[Text-to-Speech](#) [Automatic Speech Recognition](#)  
[Audio-to-Audio](#) [Audio Classification](#)  
[Voice Activity Detection](#)

Tabular

[Tabular Classification](#) [Tabular Regression](#)

Reinforcement Learning

[Reinforcement Learning](#) [Robotics](#)

Models 469,541

[Filter by name](#)

meta-llama/Llama-2-70b

Text Generation • Updated 4 days ago • ± 25.2k • ❤ 54

stabilityai/stable-diffusion-xl-base-0.9

Updated 6 days ago • ± 2.01k • ❤ 393

openchat/openchat

Text Generation • Updated 2 days ago • ± 1.3k • ❤ 136

illyasviel/ControlNet-v1-1

Updated Apr 26 • ❤ 1.87k

cerespense/zeroscope\_v2\_XL

Updated 3 days ago • ± 2.66k • ❤ 334

meta-llama/Llama-2-13b

Text Generation • Updated 4 days ago • ± 328 • ❤ 64

tiiuae/falcon-40b-instruct

Text Generation • Updated 27 days ago • ± 288k • ❤ 899

WizardLM/WizardCoder-15B-V1.0

Text Generation • Updated 3 days ago • ± 12.5k • ❤ 332

CompVis/stable-diffusion-v1-4

Text-to-Image • Updated about 17 hours ago • ± 448k • ❤ 5.72k

stabilityai/stable-diffusion-2-1

Text-to-Image • Updated about 17 hours ago • ± 782k • ❤ 2.81k

Salesforce/xgen-7b-8k-inst

Text Generation • Updated 4 days ago • ± 6.18k • ❤ 57

# MODELS

The screenshot shows the Hugging Face Models page. The top navigation bar includes the Hugging Face logo, a search bar, and links for Models, Datasets, Spaces, Docs, Solutions, Pricing, and a user profile. On the left, a sidebar lists categories: Tasks (Tasks, Libraries, Datasets, Languages, Licenses, Other), Multimodal (Feature Extraction, Text-to-Image, Image-to-Text, Text-to-Video, Visual Question Answering, Document Question Answering, Graph Machine Learning), Computer Vision (Depth Estimation, Image Classification, Object Detection, Image Segmentation, Image-to-Image, Unconditional Image Generation, Video Classification, Zero-Shot Image Classification), and Natural Language Processing.

The main content area displays a list of 221,071 models, sorted by Most Downloads. Each model entry includes the owner's profile picture, the model name, a brief description, and its last update date, size, and download count.

- facebook/dino-vitb16**  
• Updated 16 days ago • 95.9M • 85
- jonatasgrosman/wav2vec2-large-xlsr-53-english**  
• Updated Mar 25 • 64M • 140
- bert-base-uncased**  
• Updated 10 days ago • 55M • 886
- xlm-roberta-base**  
• Updated Apr 7 • 20.3M • 317
- gpt2**  
• Updated Dec 16, 2022 • 20.1M • 1.13k
- xlm-roberta-large**  
• Updated Apr 6 • 19.7M • 147
- microsoft/resnet-50**  
• Updated Mar 10 • 17.1M • 91

# DATASETS

The screenshot shows the Hugging Face Datasets interface. At the top, there is a navigation bar with the Hugging Face logo, a search bar, and links for Models, Datasets, Spaces, Docs, Solutions, Pricing, and a user profile. Below the navigation bar, there is a sidebar with filters for Tasks, Sizes, Sub-tasks, Languages, Licenses, and Other, along with a "Filter by" button. A search bar for "Filter Tasks by name" is also present. The main content area displays a list of datasets, each with a thumbnail, name, preview link, update date, size, and download count. The datasets listed are:

- openwebtext: Preview • Updated Apr 5 • 1.11M • 149
- glue: Preview • Updated 6 days ago • 889k • 189
- bigcode/the-stack-dedup: Preview • Updated May 4 • 728k • 137
- shunk031/JGLUE: Preview • Updated Mar 3 • 556k • 11
- pica: Preview • Updated Jan 25 • 538k • 30
- sciq: Preview • Updated 1 day ago • 512k • 30
- EleutherAI/lambada\_openai: Preview • Updated Dec 16, 2022 • 497k • 26

The sidebar also includes sections for Multimodal tasks (Feature Extraction, Text-to-Image, Image-to-Text, Text-to-Video, Visual Question Answering, Graph Machine Learning) and Computer Vision tasks (Depth Estimation, Image Classification, Object Detection, Image Segmentation, Image-to-Image, Unconditional Image Generation, Video Classification, Zero-Shot Image Classification). There is also a section for Natural Language Processing tasks (Text Classification, Token Classification, Table Question Answering, Question Answering).

# MODEL CARD

Hugging Face  Models Datasets Spaces Docs Solutions Pricing |

**gpt2** 1.13k

Text Generation PyTorch TensorFlow JAX TF Lite Rust Safetensors Transformers English doi:10.57967/hf/0039 gpt2 exbert License: mit

Model card Files and versions Community 46 Edit model card

**GPT-2**

Test the whole generation capabilities here: <https://transformer.huggingface.co/doc/gpt2-large>

Pretrained model on English language using a causal language modeling (CLM) objective. It was introduced in [this paper](#) and first released at [this page](#).

Disclaimer: The team releasing GPT-2 also wrote a [model card](#) for their model. Content from this model card has been written by the Hugging Face team to complete the information they provided and give specific examples of bias.

**Model description**

GPT-2 is a transformers model pretrained on a very large corpus of English data in a self-supervised fashion. This means it was pretrained on the raw texts only, with no humans labelling them in any way (which is why it can use lots of publicly available data) with an automatic process to generate inputs and labels from those texts. More

Downloads last month  
**20,136,184**

Safetensors Model size 137M params Tensor type F32

**Hosted inference API**

Text Generation Examples

Once upon a time, the game became all about money. In order to make money, most players had to work for it. It was the ultimate money game.

The first game it all started in was Magic: The Gathering. It was

Compute 0,3

Computation time on Intel Xeon 3rd Gen Scalable cpu: cached

JSON Output

# PIPELINE FUNCTION

```
from transformers import pipeline

classifier = pipeline("sentiment-analysis")
classifier("I've been waiting for a HuggingFace course my whole life.")
```

```
[{'label': 'POSITIVE', 'score': 0.9598047137260437}]
```

# TASK IDENTIFIER

`task( str ) -`

The task defining which pipeline will be returned. Currently accepted tasks are:

- "feature-extraction": will return a [FeatureExtractionPipeline](#)
- "sentiment-analysis": will return a [TextClassificationPipeline](#)
- "ner": will return a [TokenClassificationPipeline](#)
- "question-answering": will return a [QuestionAnsweringPipeline](#)
- "fill-mask": will return a [FillMaskPipeline](#)
- "summarization": will return a [SummarizationPipeline](#)
- "translation\_xx\_to\_yy": will return a [TranslationPipeline](#)
- "text-generation": will return a [TextGenerationPipeline](#)

# TASKS

The screenshot shows the Hugging Face Tasks page with a dark theme. At the top, there is a navigation bar with the Hugging Face logo, a search bar, and links for Models, Datasets, Spaces, Docs, Solutions, Pricing, and a user profile icon.

## Tasks

Hugging Face is the home for all Machine Learning tasks. Here you can find what you need to get started with a task: demos, use cases, models, datasets, and more!

### Computer Vision

- Depth Estimation** (66 models)
- Image Classification** (4,008 models)
- Image Segmentation** (219 models)
- Image-to-Image** (137 models)
- Object Detection** (443 models)
- Video Classification** (149 models)
- Unconditional Image Generation** (689 models)
- Zero-Shot Image Classification** (137 models)

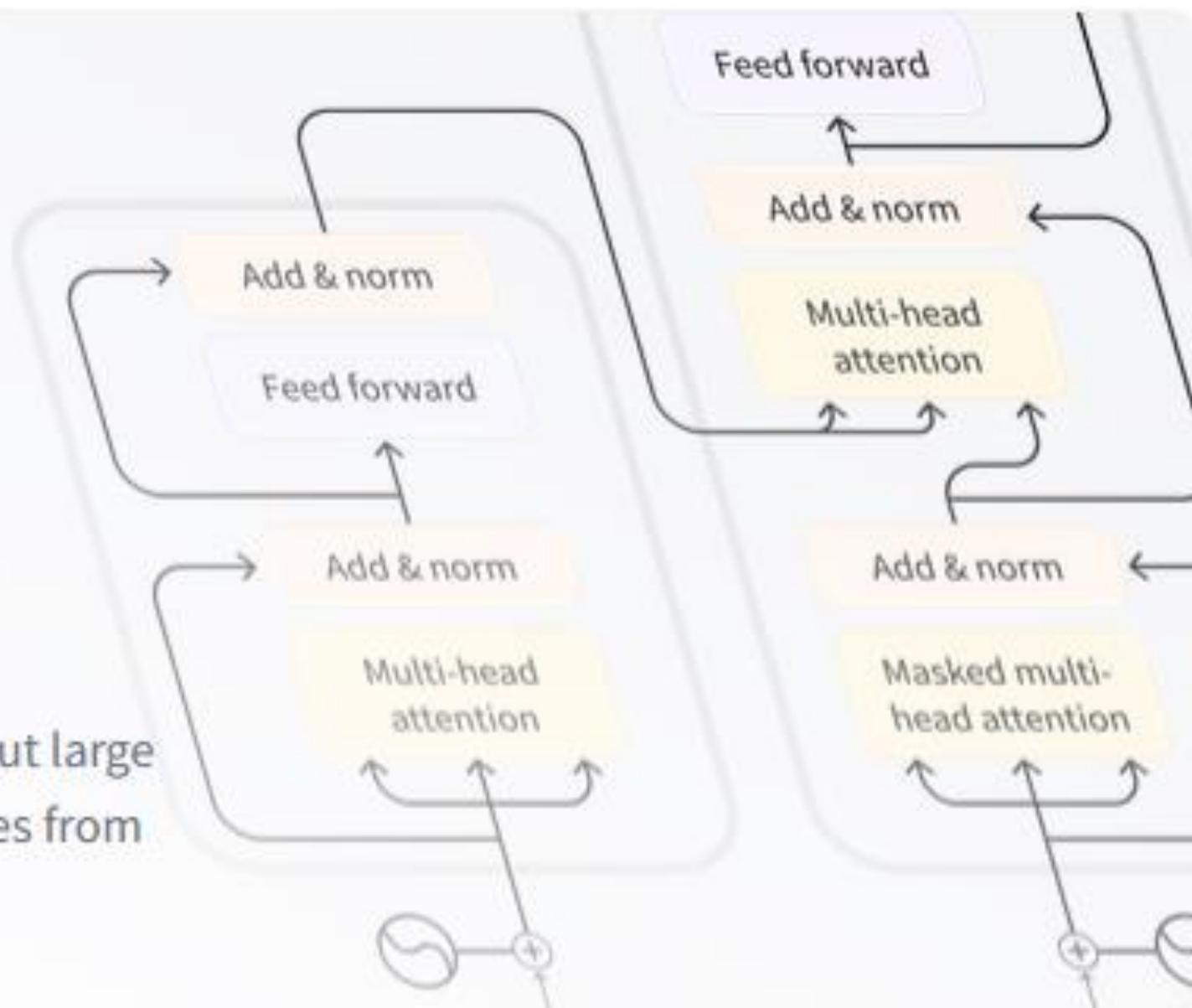
### Natural Language Processing

- Conversational** (2,194 models)
- Fill-Mask** (7,113 models)
- Question Answering** (5,081 models)
- Sentence Similarity** (1,979 models)
- Summarization** (1,149 models)
- Table Question Answering** (62 models)
- Text Classification** (24,294 models)
- Text Generation** (12,936 models)
- Token Classification** (9,153 models)

- Translation** (2,136 models)
- Zero-Shot Classification** (141 models)



## Learn



## LLM Course

This course will teach you about large language models using libraries from the HF ecosystem



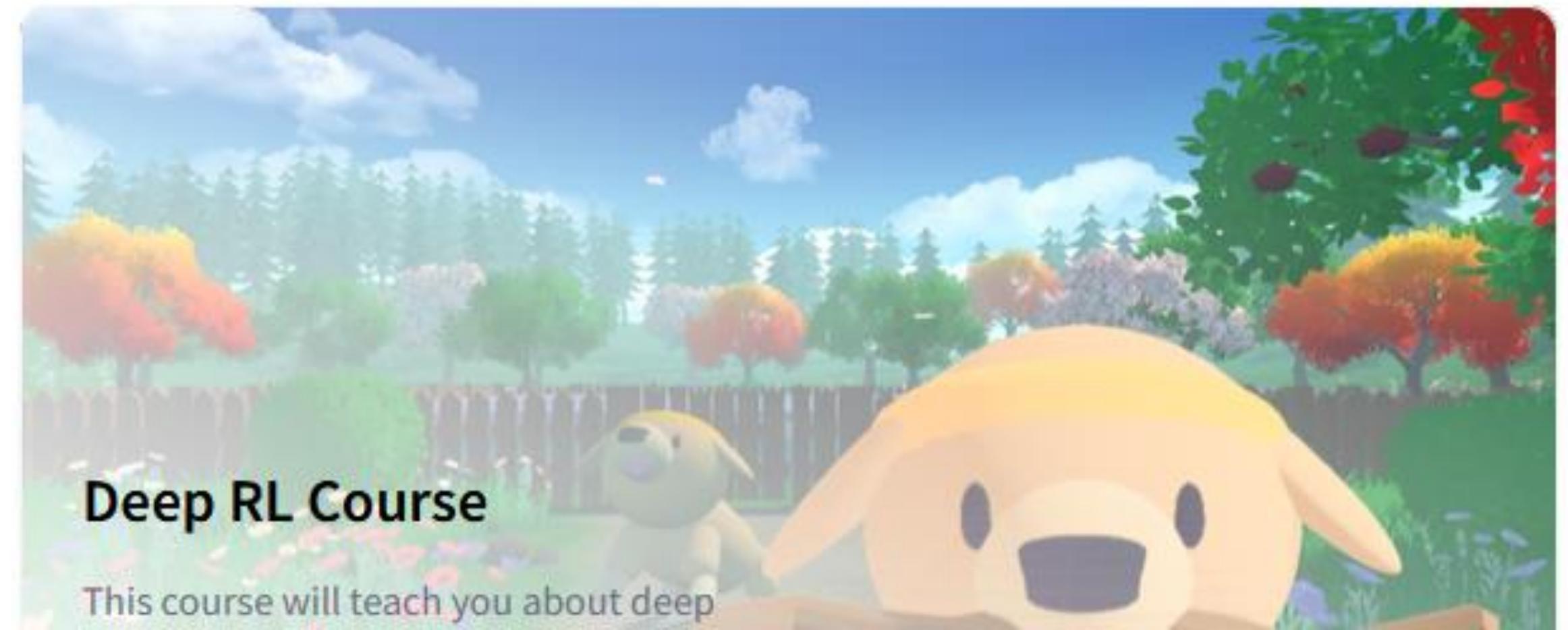
## MCP Course

This course will teach you about Model Context Protocol



## Agents Course

<https://huggingface.co/learn/llm-course>



Deep RL Course

This course will teach you about deep



## Machine Learning with TensorFlow

DONNERSTAG 18:15 - 20:00

Get hands-on experience in applying machine learning techniques with TensorFlow.

Die Bewerbungsfrist ist leider abgelaufen. Bleibe immer auf dem Laufenden und verpasste nicht unser nächstes Kursprogramm, Events, Jobs und was es sonst noch Neues gibt: [ZUM NEWSLETTER](#)

### Du wirst lernen

- ✓ Mastering best practices for TensorFlow, a popular open-source machine learning framework, to train neural networks
- ✓ Handling real-world image data and exploring strategies to avoid overfitting, including techniques like augmentation and dropout
- ✓ Building a system capable of processing natural language
- ✓ Applying Recurrent Neural Networks (RNNs), Gated Recurrent Units (GRUs), and Long Short-Term Memory (LSTMs) networks to train models using text and time-series data

DONNERSTAG

18:15 -  
20:00

ECTS

5



HEIDE +  
KIEL +  
ONLINE



ENGLISCH



## Machine Learning Degree

After completing this program you will have a solid understanding of machine learning and will be able to implement your own state of the art machine learning projects.

Die Bewerbungsfrist ist leider abgelaufen. Bleibe immer auf dem Laufenden und verpasste nicht unser nächstes Kursprogramm, Events, Jobs und was es sonst noch Neues gibt: [ZUM NEWSLETTER](#)

### Aktuelle Angebote im Rahmen des Degrees

- [Advanced Time Series Prediction](#)
- [Intermediate Machine Learning](#)
- [AI Builder's Arena](#)
- [Machine Learning with TensorFlow](#)
- [Einführung in Data Science & maschinelles Lernen](#)
- [TinyML: Embedded Machine Learning](#)

ECTS

12,5

KIEL +  
ONLINE

ENGLISCH



Mehr



# CODING.WATERKANT

Coding Camp  
from July 7-11, 2025

 [Looking for the 2024 program? View it here →](#)





# Coding.Waterkant

★★★★★ (128) 2

📍 Kiel, Deutschland

👤 1.003 Mitglieder · Öffentliche Gruppe i

👤 Organisiert von [opencampus.sh](#) and **6 others**

✉️ [Mitglieder kontaktieren](#)

Teilen: [n](#) [f](#) [t](#) [in](#) [✉](#)

Info

Events

Mitglieder

Fotos

Diskussionen

Mehr

Event erstellen ▼

Gruppe verwalten ▼

## Über uns

Our meetup is organized with support of [opencampus.sh](#) and the Digitale Wirtschaft Schleswig-Holstein (DiWiSH).

[Mehr lesen](#)



## Organizers



[opencampus.sh and 6 others](#)

[✉ Nachricht](#)

## Members (1.003)

[Alles ansehen](#)

