

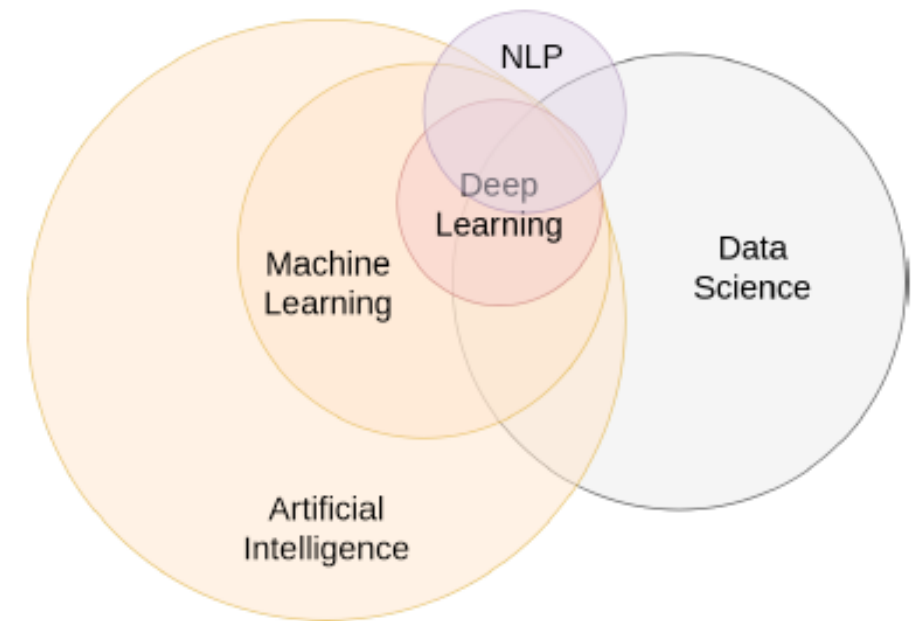


# Was ist NLP

- Natural Language Processing = Natursprachverarbeitung
- Auch
  - Computational Linguistic = Computerlinguistik

## NLP:

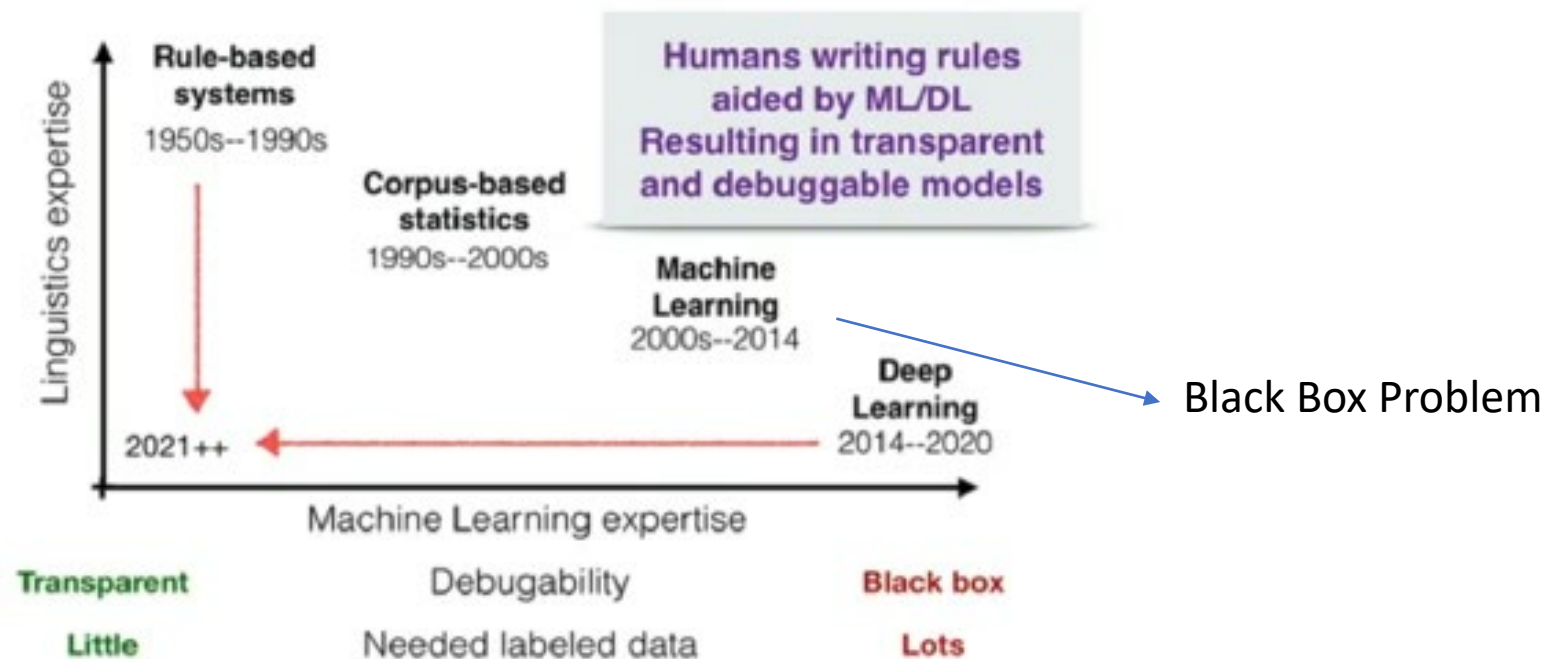
- Rasante Entwicklung
- Industrie orientiert.
- Große Nachfrage an Spezialisten



# NLP Aufgaben

- Text Processing
- Text Classification
- Verbal Reasoning
- Language Generation
- Multimodal Data Processing
- Computational Social Science

# NLP Tomorrow



Yoav Goldberg: The missing elements in NLP (spaCy IRL 2019)

# Installierung und Nutzung von Jupyter Notebook



# Was ist ein Jupyter Notebook?

- Kombination von **Text, Code, Outputs, Visualisierungen...**
- Ein Dokument
- Sehr populär in Data Science
- Gut Workflow and Kommunikation
- Einfach, um online zu teilen.

# Installierung



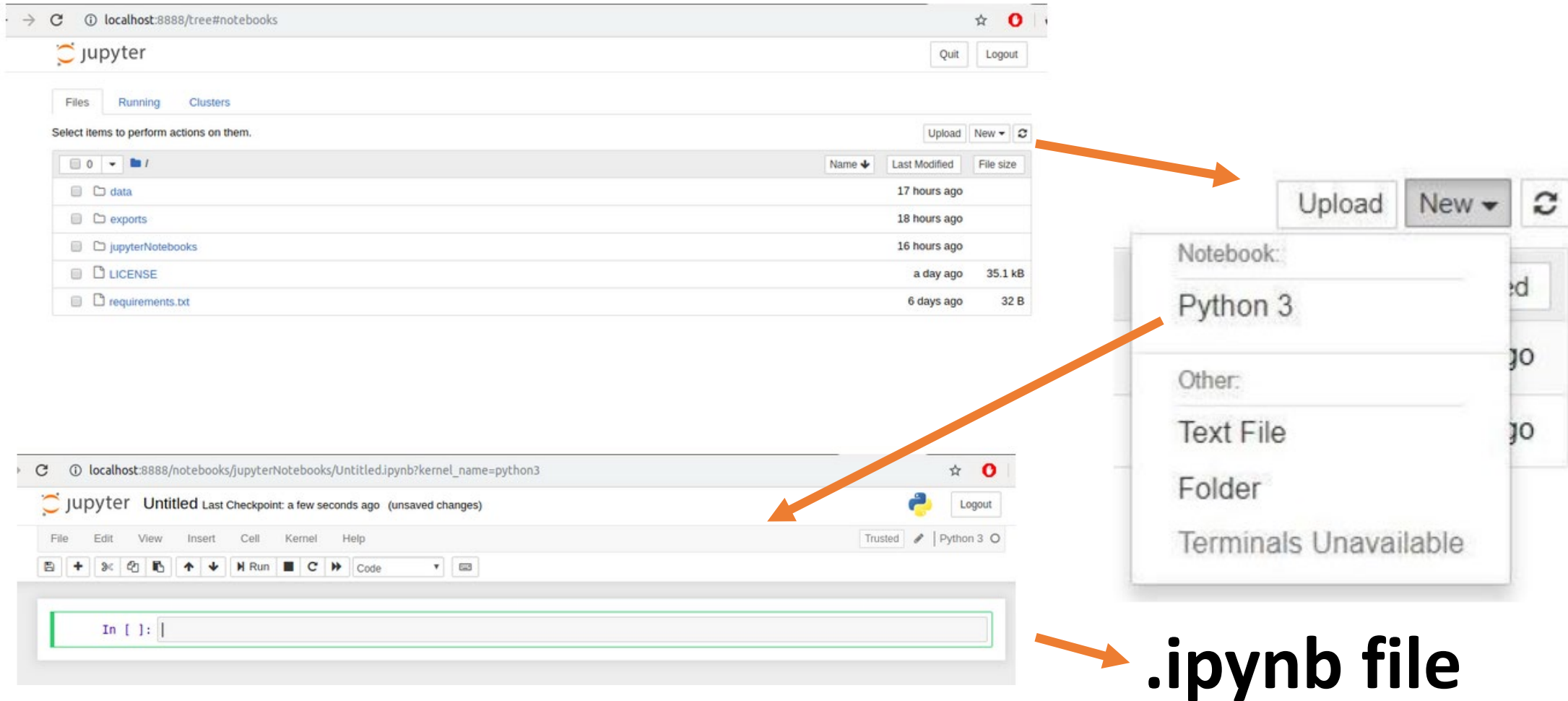
**ANACONDA®**

- Zuerst Anaconda installieren

<https://docs.jupyter.org/en/latest/install.html>

Wenn du Anaconda installieren, auch: Python3 und Jupyter

# Interface



Quelle: [www.medium.com/velotio-perspectives/the-ultimate-beginners-guide-to-jupyter-notebooks-6b00846ed2af](https://www.medium.com/velotio-perspectives/the-ultimate-beginners-guide-to-jupyter-notebooks-6b00846ed2af):



# Interface

- **Cells:**

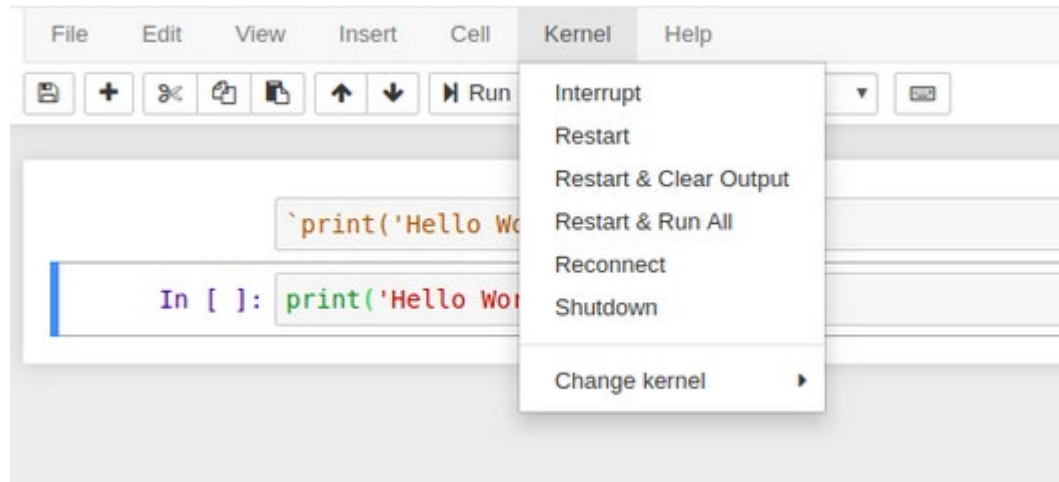
- Code
- Markdown (Text)
- ...

- **Kernel:**

- läuft die Code
- Variabel eines Cells ist verfügbar für das nächstes.

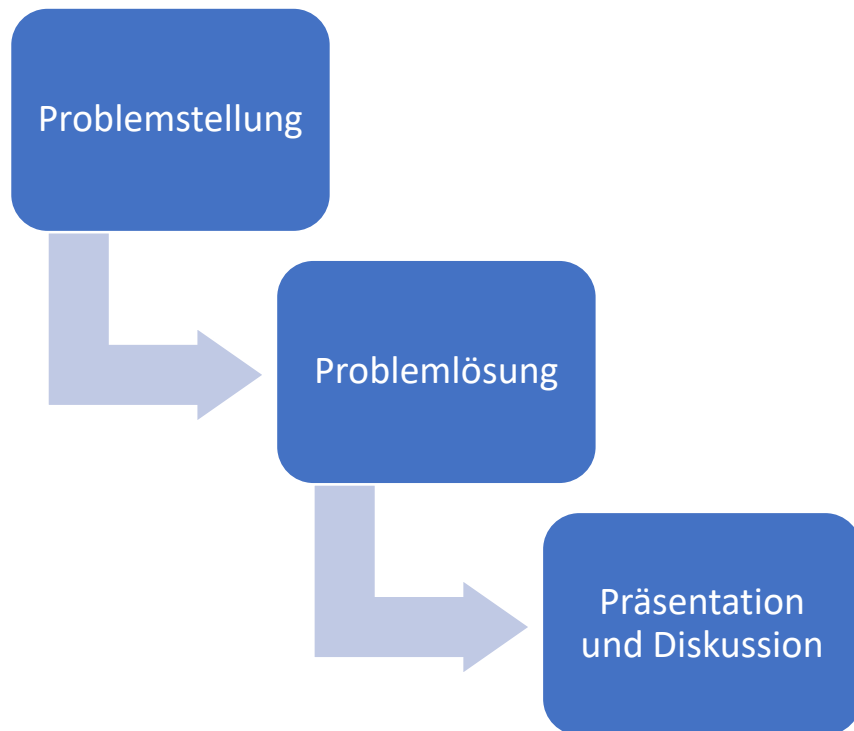
# Einige Befehlen

- Ctrl + Shift + P:
  - Ctrl + Enter
  - A
  - B
  - D



# (Mini)Problembasiertes Lernen

- Aktiv lernen!
- Teamarbeit!
- Arbeit nicht nur machen, aber auch zeigen!



**Jede Gruppe muss mindestens einmal präsentieren**

# Kurs Überblick

- **Aufgabe 1:** Texte und Kodierung: Mit Kodierungsproblem umgehen und Python Grundlage aktivieren.
- **Aufgabe 2:** Reguläre Ausdrücke: Bestimmte Muster von Daten zu extrahieren.
- **Aufgabe 3:** Vorverarbeitung von Texten: Tokenisierung und Wörter zählen. Mögliche Erweiterung: Wichtigkeit von Wort zu bestimmen.
- **Aufgabe 4:** Text-Klassifikation: Klassifikator zu trainieren. Text-Kategorie zu bestimmen.
- **Aufgabe 5:** Sentiment-Analyse: Analyse zu evaluieren.
- Vielleicht:
  - Named Entities
  - Sprachbestimmung

# Bücher Empfehlung (English)

Beide Bücher sind online verfügbar und sind eine gute Referenz für NLP:

- *Natural Language Processing with Python - Steven Bird, Ewan Klein, and Edward Loper*

Kommentar: einführend und praktisch. Etwas zu fokussiert auf NLTK Packet.

<https://www.nltk.org/book/>

- *Speech and Language Processing - Dan Jurafsky and James H. Martin*

Kommentar: Fortgeschritten und etwa theoretischer

<https://web.stanford.edu/~jurafsky/slp3/>