



TRAINING: Python für die Datenanalyse in den Sozialwissenschaften

Teil 2: Analyse verschiedener Datenarten und Web-Ressourcen

Large-Language-Modelle (LLM) und KI-Assistenten

SPEAKER: Matthias Täschner



GEFÖRDERT VOM

Bundesministerium
für Bildung
und Forschung



Diese Maßnahme wird gefördert durch die Bundesregierung aufgrund eines Beschlusses des Deutschen Bundestages. Diese Maßnahme wird mitfinanziert durch Steuermittel auf der Grundlage des von den Abgeordneten des Sächsischen Landtags beschlossenen Haushaltes.



Come2Data
Kompetenzzentrum für
interdisziplinäre Datenwissenschaften

Training: Python für die Datenanalyse in den Sozialwissenschaften – Teil 2
Speaker: Matthias Täschner, Universität Leipzig, ScaDS.AI

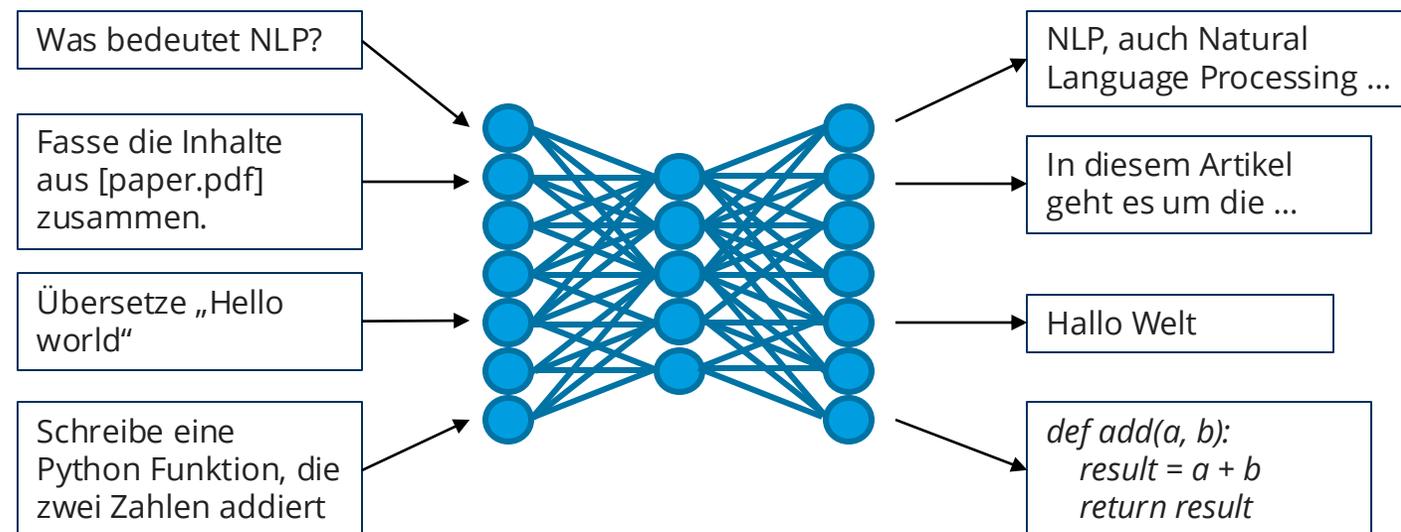
Folie 1



LLM-Services und deren Nutzung

Was ist ein LLM?

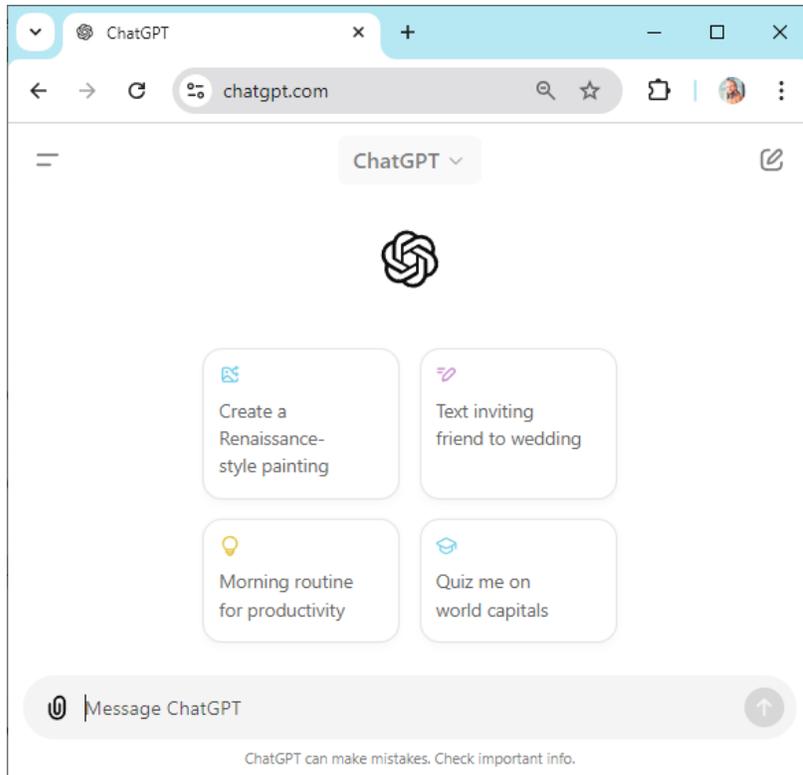
- Fortschrittliche KI-Modelle (Neuronale Netze / Deep Learning), die auf großen Textmengen trainiert werden
- Sollen menschliche Sprache verstehen und generieren
- Sind fähig Kontext einzubeziehen und mit Nutzern zu interagieren
- Können je nach Spezialisierung eine Vielzahl von Aufgaben ausführen:
 - Beantworten von Fragen
 - Zusammenfassen von Texten
 - Übersetzen
 - Generieren von Code
 - Weitere Funktionen, z.b. Websuche



LLM-Services und deren Nutzung

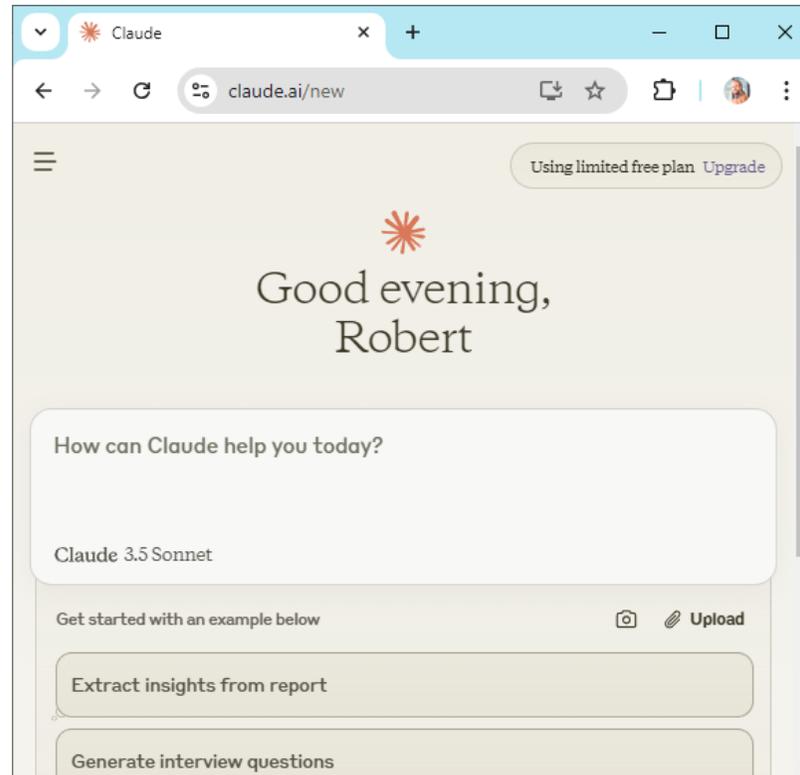
Aufbau und Nutzung

ChatGPT



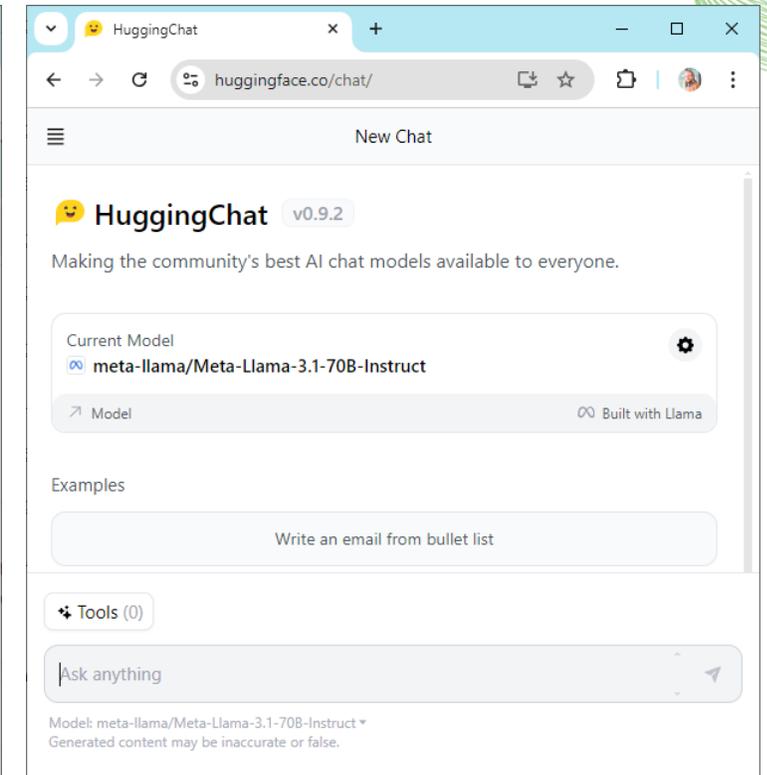
<https://chatgpt.com/>

Claude



<https://claude.ai/>

Open models

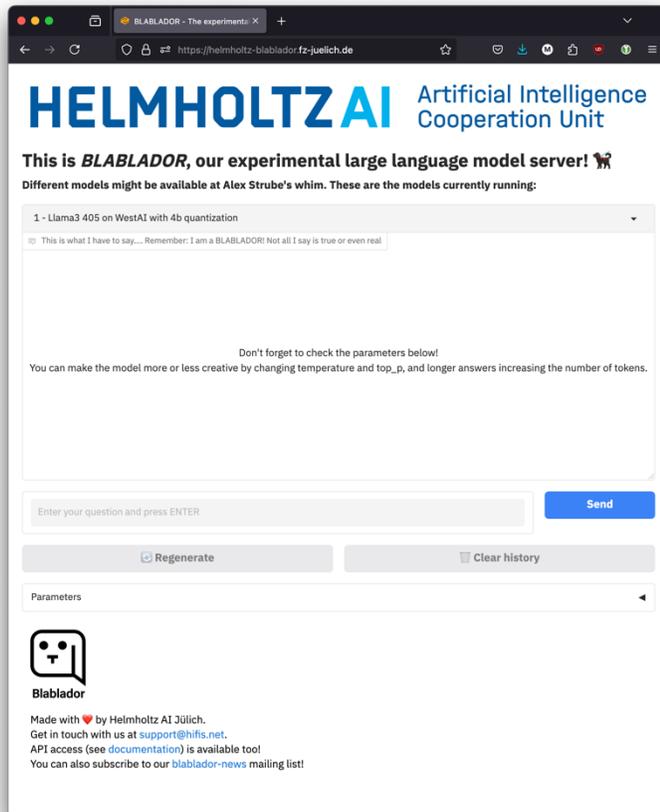


<https://huggingface.co/chat/>

LLM-Services und deren Nutzung

Aufbau und Nutzung

Helmholtz Blablador

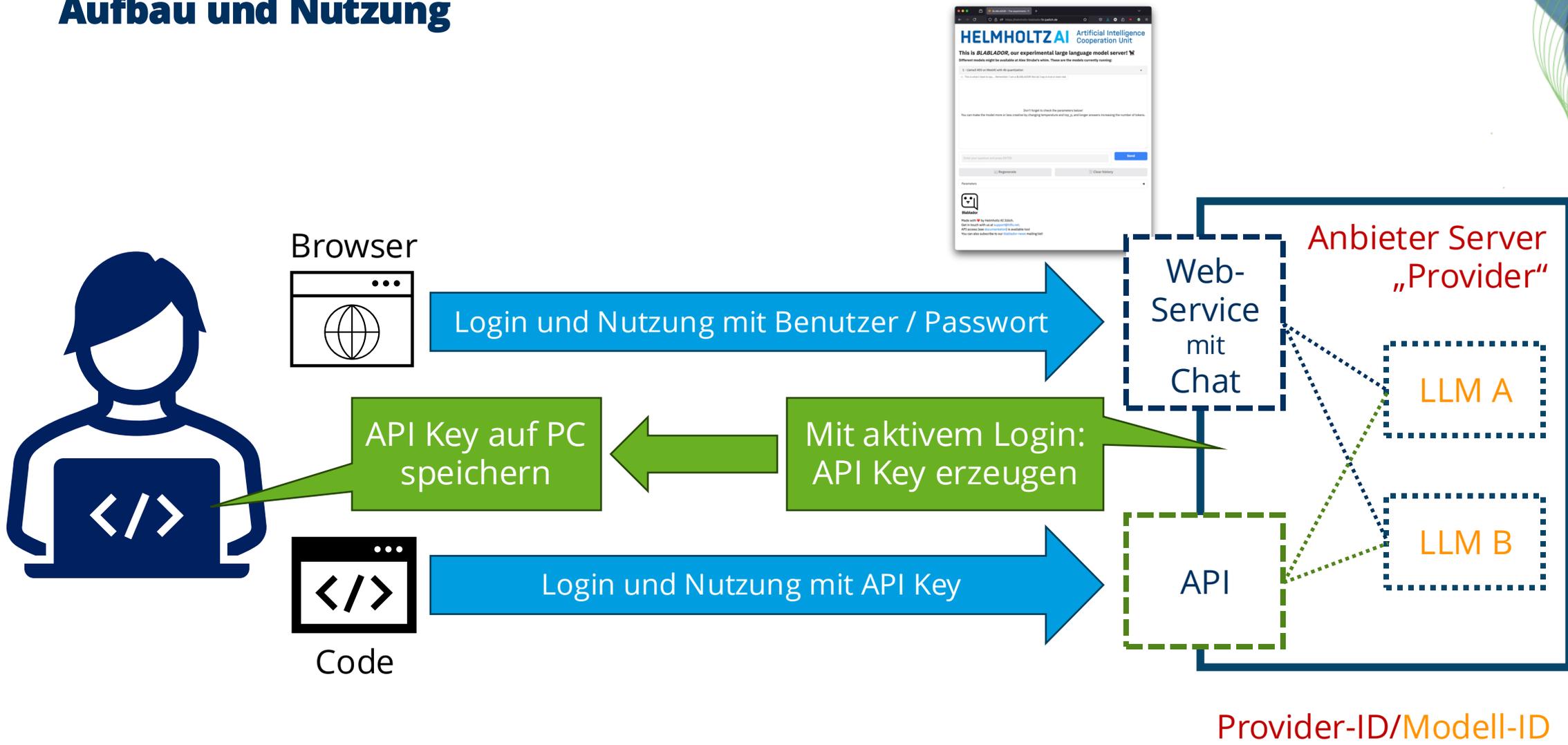


- Verschiedene Anbieter mit eigenen Services und LLMs
- Proprietär bis Open-Source
- Kostenpflichtig bis kostenfrei
- Serverstandorte weltweit, in Europa, in Deutschland
- Grundlegender technischer Aufbau und Konzepte aber meist ähnlich

<https://helmholtz-blablador.fz-juelich.de/>

LLM-Services und deren Nutzung

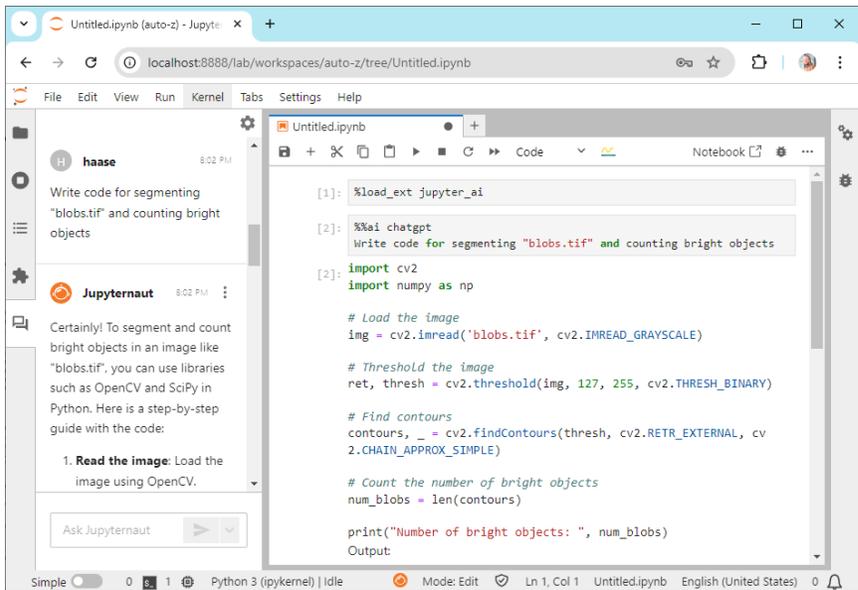
Aufbau und Nutzung



KI-Assistenten

Technischer Überblick

- Einbindung von LLM-Funktionen in bestehende Werkzeuge und Services
 - ChatBots auf Websites
 - KI-Assistent in MS365, Adobe Cloud, ...
 - KI-Assistent in Programmierumgebungen (IDE, ...)
- Für Jupyter Lab z.B. via „jupyter-ai“ oder „bia-bob“

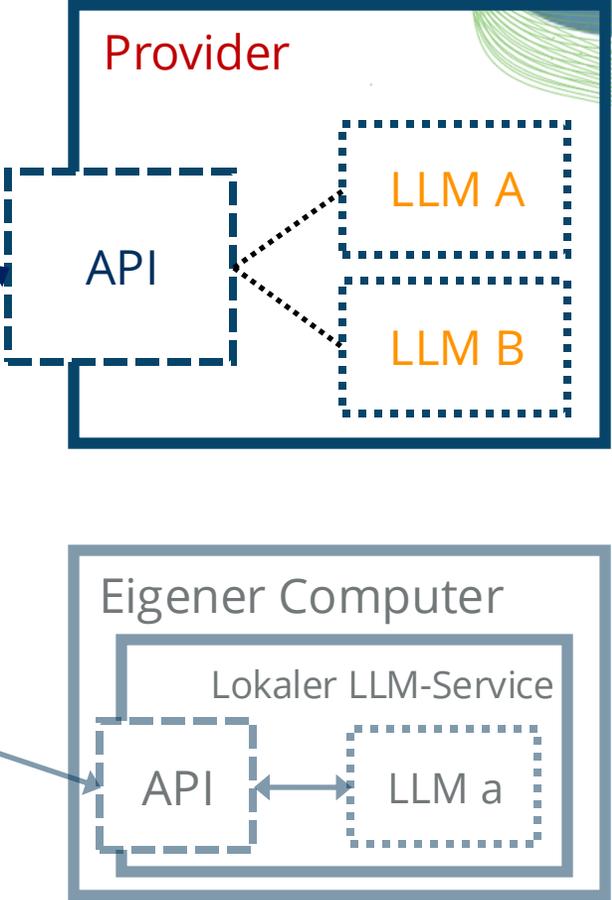


Softwarepaket / Plugin für Chatfunktion, Text- / Codegenerierung, etc mit Anbindung an ...

Anmeldung mit API Key

blablador/alias-fast

Alternativ: Direkter lokaler Zugriff, benötigt entsprechend performante Hardware



KI-Assistenten

Nutzung und Einschränkungen

Prompt Engineering

- Prompt: Anweisung an den KI-Assistenten / das LLM zur Erfüllung einer Aufgabe
- Möglichst präzise formulieren sowie Kontext bereitstellen
- Iterative Modifikation des Prompts bis zum gewünschten Ergebnis

Einschränkungen und Grenzen

- Prompts und Kontext werden an LLM übertragen – bei Nutzung eines externen LLM-Service Anbieters also an externe Server → Datenschutz beachten!
- KI-Assistent ist nur so gut wie das jeweils genutzte LLM – Variation bei Qualität der Antworten, verfügbaren Sprachen, Wissensstand, Fähigkeit zur Codegenerierung, ...
- Das Kontext-Fenster (Anzahl der Token) hat Grenzen und variiert je nach LLM

Im Rahmen dieser Schulung nutzen wir „bia-bob“ (KI-Assistent für Jupyter) mit Blablador (LLM-Provider Helmholtz / FZ Jülich)

(<https://github.com/haesleinhuepf/bia-bob>)

(<https://helmholtz-blablador.fz-juelich.de>)