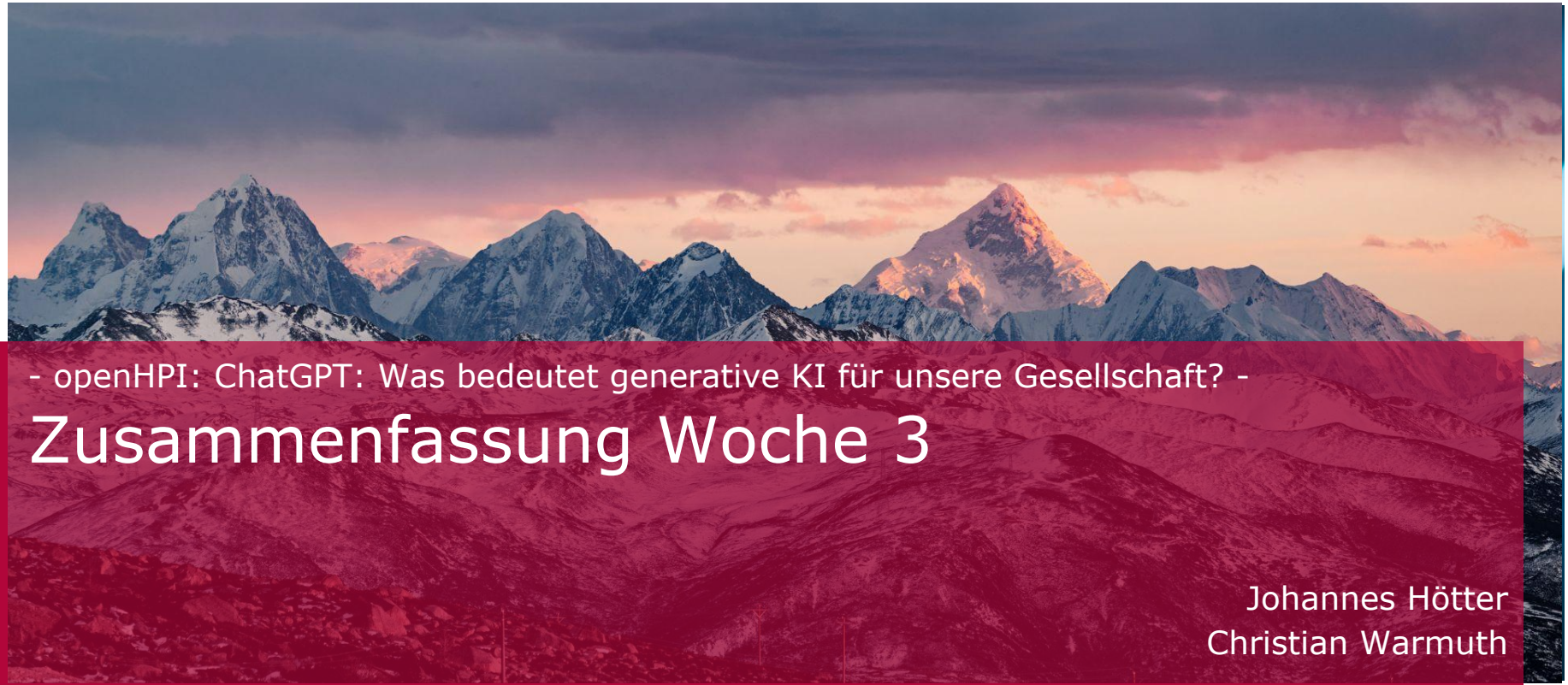


- openHPI: ChatGPT: Was bedeutet generative KI für unsere Gesellschaft? -

Zusammenfassung Woche 3

Johannes Hötter
Christian Warmuth

Vor allem interessant

- Open Source LLMs sind ein zweischneidiges Schwert
- Fehler können bei Halluzinationen und bei fehlerhaften Daten passieren
- LLMs müssen nicht *alles* wissen, sondern können verkettet werden; etwa mit Domänenwissen oder in Form eines Agenten mit z.B. anderen Tools
- Wichtig dabei: wie können “schädliche” Anfragen potenziell gefiltert werden?
- Quantisierte Modelle als Schritt Richtung Kosteneffizienz und Nachhaltigkeit

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Zusammenfassung Woche 3

Johannes Hötter,
Christian Warmuth

openHPI

Open Source als zweischneidiges Schwert

- Absicherung gegenüber Monopol
- In diversen Anwendungsfällen wegen Self-Hosting unabdingbar
- Finetuning grundlegend möglich

Kann von allen für alles eingesetzt werden

- **Ausrollen kann nicht rückgängig gemacht werden!**
- Lizenzen für diverse ethische Einschränkungen vorhanden, allerdings auch “nur” formale Einschränkung

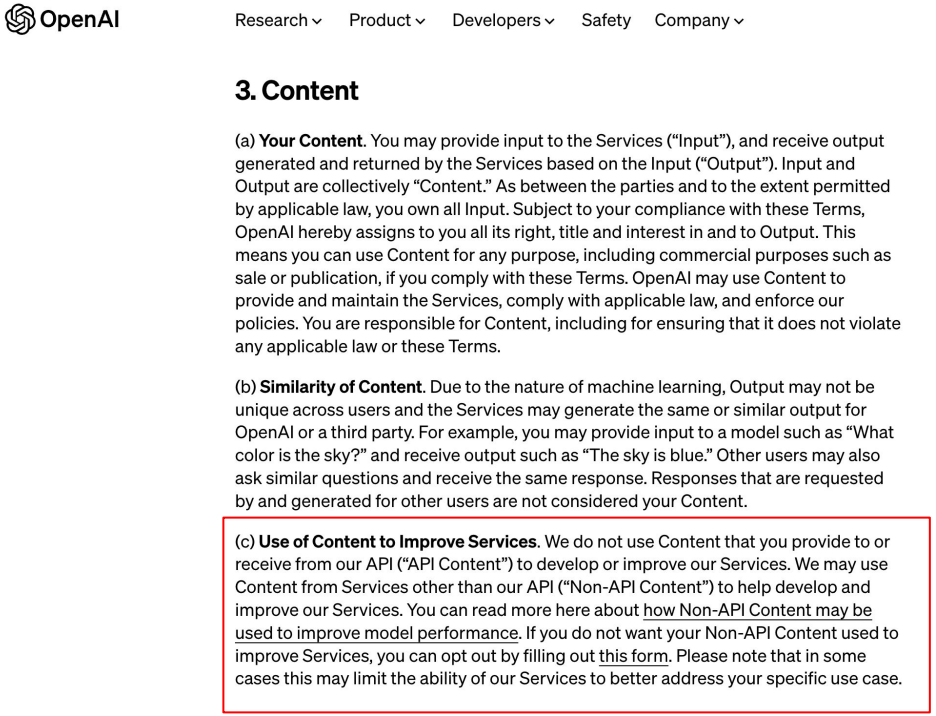
Kann von allen für alles eingesetzt werden

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Johannes Hötter,
Christian Warmuth

openHPI

Wie gehen Provider mit den Daten um?



OpenAI

Research Product Developers Safety Company

3. Content

(a) **Your Content**. You may provide input to the Services ("Input"), and receive output generated and returned by the Services based on the Input ("Output"). Input and Output are collectively "Content." As between the parties and to the extent permitted by applicable law, you own all Input. Subject to your compliance with these Terms, OpenAI hereby assigns to you all its right, title and interest in and to Output. This means you can use Content for any purpose, including commercial purposes such as sale or publication, if you comply with these Terms. OpenAI may use Content to provide and maintain the Services, comply with applicable law, and enforce our policies. You are responsible for Content, including for ensuring that it does not violate any applicable law or these Terms.

(b) **Similarity of Content**. Due to the nature of machine learning, Output may not be unique across users and the Services may generate the same or similar output for OpenAI or a third party. For example, you may provide input to a model such as "What color is the sky?" and receive output such as "The sky is blue." Other users may also ask similar questions and receive the same response. Responses that are requested by and generated for other users are not considered your Content.

(c) **Use of Content to Improve Services**. We do not use Content that you provide to or receive from our API ("API Content") to develop or improve our Services. We may use Content from Services other than our API ("Non-API Content") to help develop and improve our Services. You can read more here about how Non-API Content may be used to improve model performance. If you do not want your Non-API Content used to improve Services, you can opt out by filling out this form. Please note that in some cases this may limit the ability of our Services to better address your specific use case.

Quelle: https://openai.com/policies/terms-of-use, aufgerufen 04. Juni 2023 17:23

Hasso Plattner Institut
Digital Engineering · Universität Potsdam

KI Service Zentrum by Hasso-Plattner-Institut

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Zusammenfassung Woche 3

Johannes Hötter,
Christian Warmuth

openHPI

Beispiel Conversational AI

Fehlerhafte Fakten

"Ich habe ein Patent verletzt, bin ich versichert?"

Versichert sind:
- Schäden i.H.v. bis zu 300.000€
- Patentverletzungen

Nicht versichert sind:
- Imageschäden
- …

"Nein, sind Sie nicht."

HPI Hasso Plattner Institut
Digital Engineering · Universität Potsdam

KI Service Zentrum HPI
by Hasso-Plattner-Institut

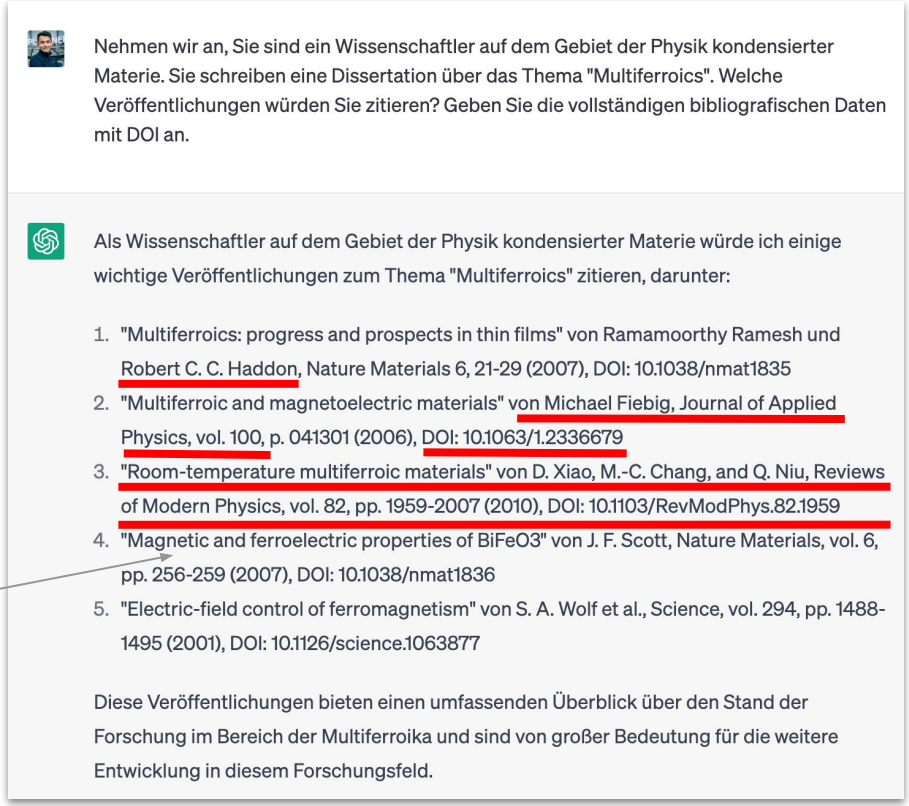
GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Zusammenfassung Woche 3

Johannes Hötter,
Christian Warmuth

openHPI

Beispiel einer Halluzination



Nehmen wir an, Sie sind ein Wissenschaftler auf dem Gebiet der Physik kondensierter Materie. Sie schreiben eine Dissertation über das Thema "Multiferroics". Welche Veröffentlichungen würden Sie zitieren? Geben Sie die vollständigen bibliografischen Daten mit DOI an.

Als Wissenschaftler auf dem Gebiet der Physik kondensierter Materie würde ich einige wichtige Veröffentlichungen zum Thema "Multiferroics" zitieren, darunter:

1. "Multiferroics: progress and prospects in thin films" von Ramamoorthy Ramesh und Robert C. C. Haddon, Nature Materials 6, 21-29 (2007), DOI: 10.1038/nmat1835
2. "Multiferroic and magnetoelectric materials" von Michael Fiebig, Journal of Applied Physics, vol. 100, p. 041301 (2006), DOI: 10.1063/1.2336679
3. "Room-temperature multiferroic materials" von D. Xiao, M.-C. Chang, and Q. Niu, Reviews of Modern Physics, vol. 82, pp. 1959-2007 (2010), DOI: 10.1103/RevModPhys.82.1959
4. "Magnetic and ferroelectric properties of BiFeO3" von J. F. Scott, Nature Materials, vol. 6, pp. 256-259 (2007), DOI: 10.1038/nmat1836
5. "Electric-field control of ferromagnetism" von S. A. Wolf et al., Science, vol. 294, pp. 1488-1495 (2001), DOI: 10.1126/science.1063877

Diese Veröffentlichungen bieten einen umfassenden Überblick über den Stand der Forschung im Bereich der Multiferroika und sind von großer Bedeutung für die weitere Entwicklung in diesem Forschungsfeld.

Richtiger Artikel:
Ferroelectric and magnetic properties of multiferroic BiFeO3-La0.7Sr0.3MnO3 heterostructures integrated with Si (100)

Model ChatGPT-3.5; 12. Mai 2023

Hasso Plattner Institut
Digital Engineering · Universität Potsdam

KI Service Zentrum HPI
by Hasso-Plattner-Institut

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Zusammenfassung Woche 3

Johannes Hötter,
Christian Warmuth

openHPI

Generative KI liegt nahe an deklarativer Programmierung

Kostengünstig in einzelner Anwendung

Geringe Wissenshürden in der Anwendung

z.T. von deklarativem Prompt zu funktionalem Code

Hoher Grad der Personalisierung und damit "überzeugender" Inhalt

…, aber Fehler können auch sehr teuer werden!

Digital Engineering · Universität Potsdam

KI Service Zentrum by Hasso-Plattner-Institut

GEFÖRDERT VOM Bundesministerium für Bildung und Forschung

Zusammenfassung Woche 3

Johannes Hötter, Christian Warmuth

openHPI

Recap
Ist nicht jedes Machine Learning-Modell datenzentriert?

Trends:

- Komplexe Algorithmen sind in open-source Bibliotheken bereits umgesetzt
- Anbieter wie OpenAI oder Hugging Face (Open Source) stellen vortrainierte Modelle bereit, welche per Finetuning einfach angepasst werden können

Konsequenzen:

- Für die meisten Anwendungen kein Mehrwert, eigene Algorithmik zu implementieren
- Fokus lieber auf der Datensammlung und -Aufbereitung
- Vorteil: Daten sind langlebig und modellunabhängig

Choose your battles!

Wichtig zu betonen:

- LLMs sind keine Datenbanken
- LLMs sind keine Taschenrechner
- LLMs sind keine Suchmaschinen
- LLMs sind teilweise nicht gut in “gesundem Menschenverstand”

Können aber mit vielen dieser Fähigkeiten “ausgestattet werden”

- Vector Datenbanken
- Prompt Engineering & Prompt Chaining
- Zugriff auf z.B mathematische Tools
- …

GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Besondere Wichtigkeit

KI Systeme werden zunehmend von isoliert agierenden Systemen zu aktiv agierenden Systemen, d.h. teilweise autonom agierenden Systemen.

Beispiele: Zugriff auf das Internet nicht nur in lesendem Modus, API-Calls,...
(Siehe Beispiel in Woche 2: Agenten)

Umso wichtiger sind in diesem Kontext alle Anstrengungen, KI-Sicherheit zu gewährleisten.

HPI Hasso Plattner Institut
Digital Engineering · Universität Potsdam

KI Service Zentrum HPI
by Hasso-Plattner-Institut

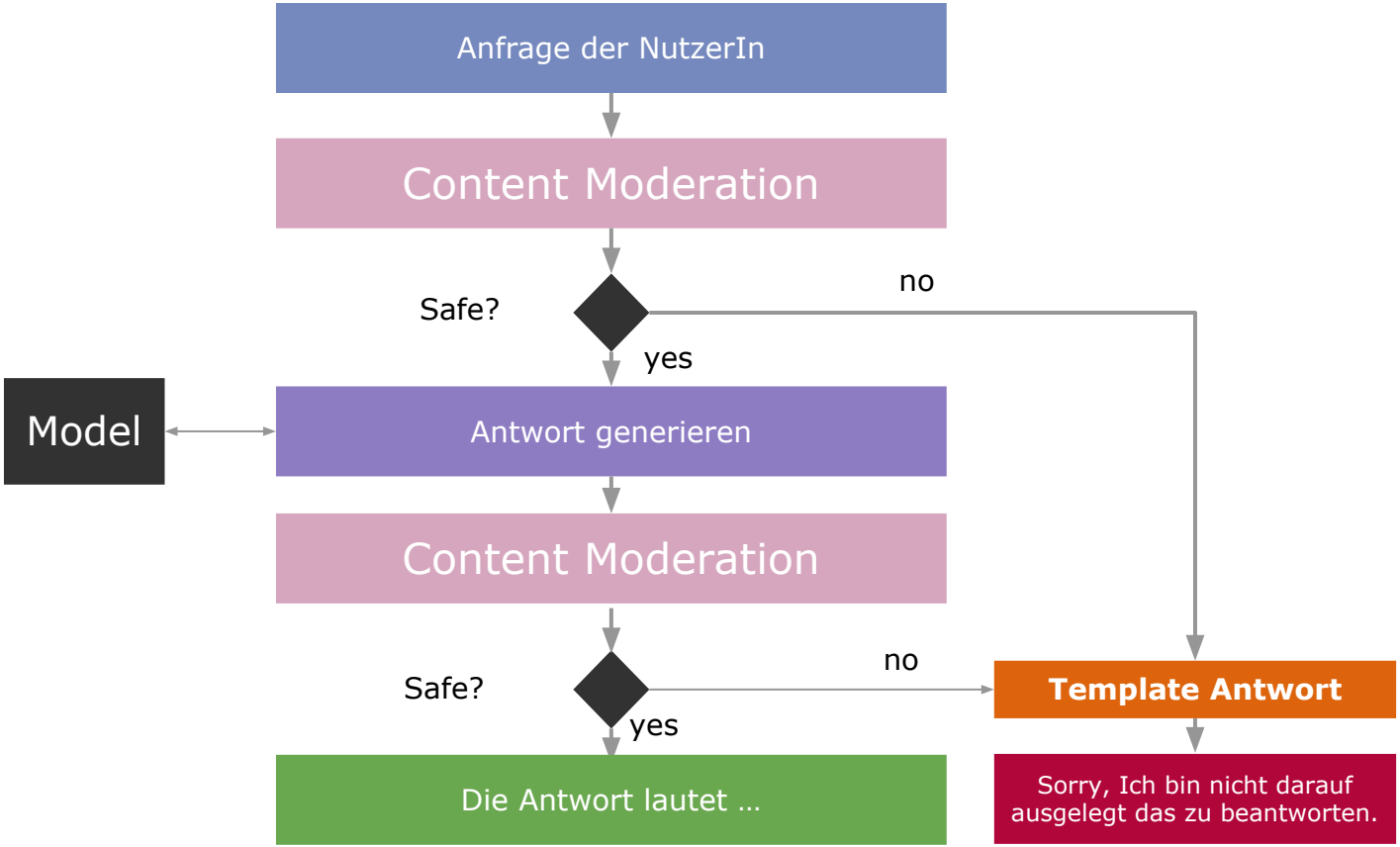
GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Zusammenfassung Woche 3

Johannes Hötter,
Christian Warmuth

openHPI

Content Filtering



Digital Engineering · Universität Potsdam

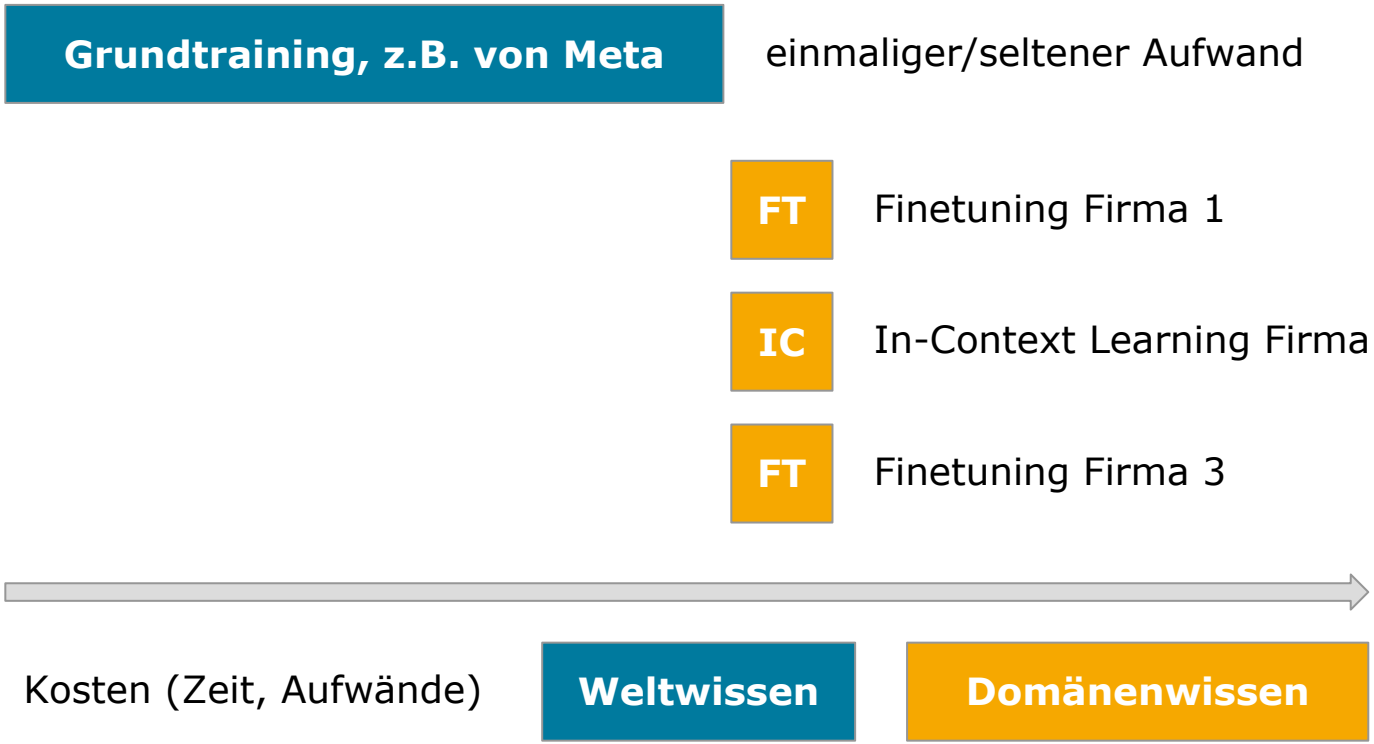
GEFÖRDERT VOM
Bundesministerium für Bildung und Forschung

Zusammenfassung Woche 3

Johannes Hötter, Christian Warmuth

openHPI

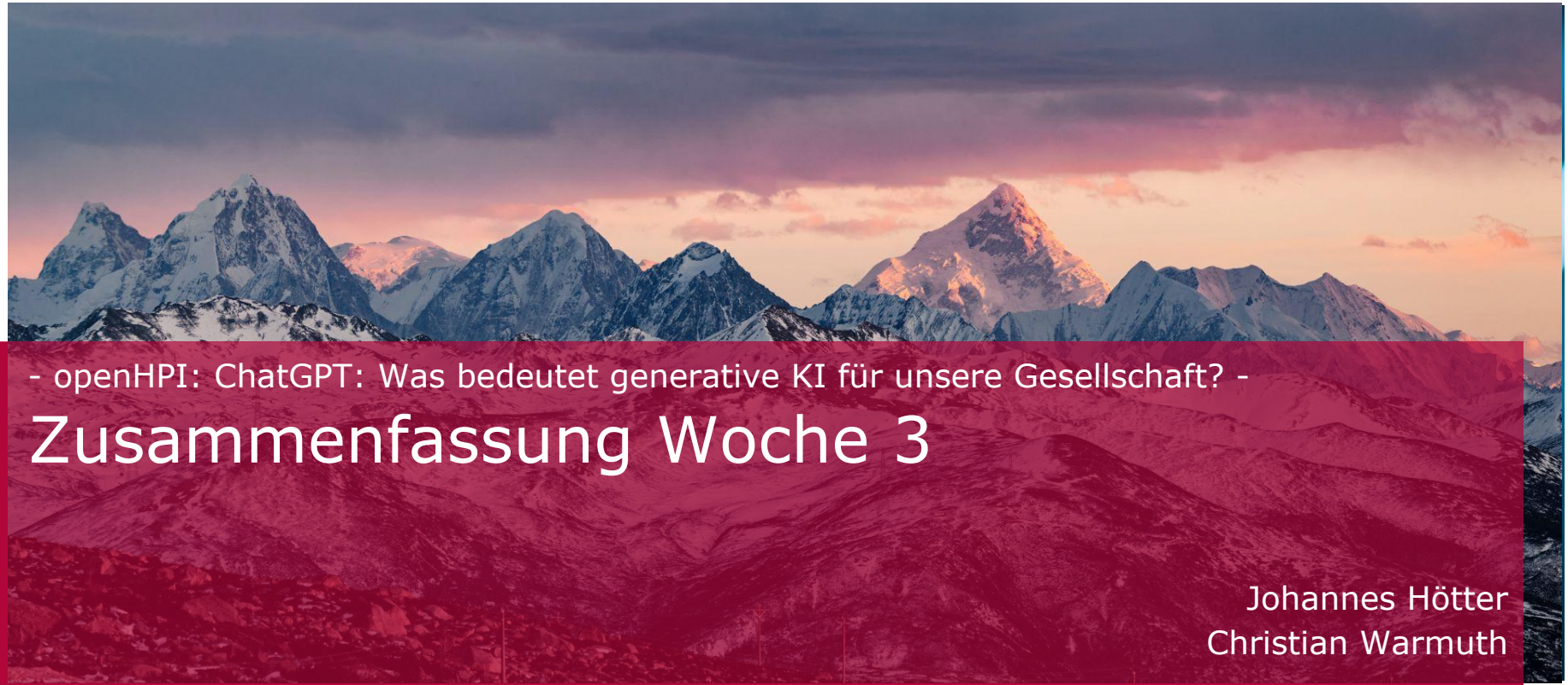
Open Source, quantisierte LLMs werden zur Nachhaltigkeit und Kosteneffizienz beitragen



Zusammenfassung Woche 3

Johannes Hötter,
Christian Warmuth

openHPI



- openHPI: ChatGPT: Was bedeutet generative KI für unsere Gesellschaft? -

Zusammenfassung Woche 3

Johannes Hötter
Christian Warmuth