

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

Halluzinationen

Johannes Hötter
Christian Warmuth

Halluzinationen

 Menü | 

Süddeutsche Zeitung[Abos](#) [Login](#)

Meine SZ | SZ Plus | Ukraine | Politik | Wirtschaft | Meinung | Panorama | Sport | München ▾ | Kultur | Medien | Bayern | Reportag >

Home > Wissen > Informatik > Wenn die KI lügt: Warum Chat-GPT so viel erfindet

Gesundheit Digital

Künstliche Intelligenz

Warum die KI so gerne lügt

28. März 2023, 10:54 Uhr | Lesezeit: 5 min



Halluzinationen

Johannes Hötter,
Christian Warmuth

openHPI

Was sind Halluzinationen

Halluzination: Generative KI (LLMs) generieren inkorrekte Ergebnisse oder Ergebnisse die nicht auf bereitgestellten Daten/Fakten basieren.

Konzept nicht anwendbar z.B. auf generative KI Modelle die Bilder erzeugen



Halluzinationen

Johannes Hötter,
Christian Warmuth

openHPI

Ein Beispiel



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:

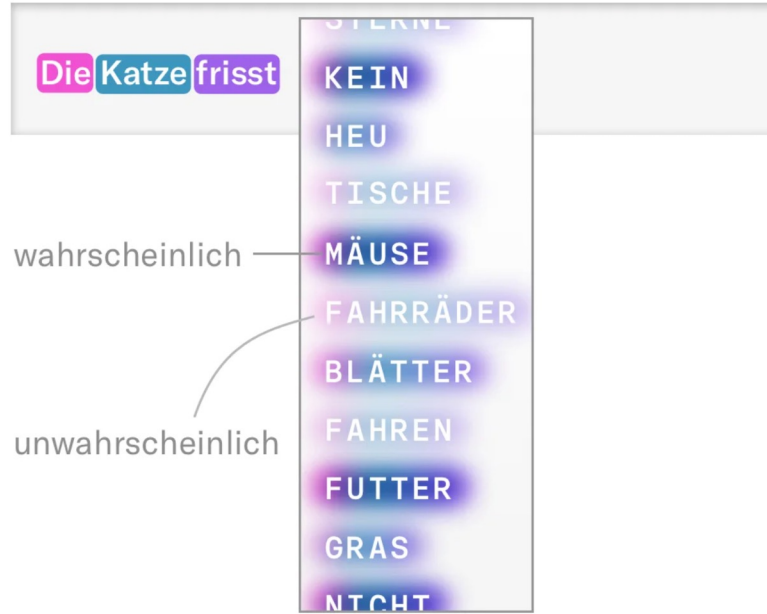
Ferroelectric and magnetic properties of multiferroic BiFeO₃-La_{0.7}Sr_{0.3}MnO₃ heterostructures integrated with Si (100)

Halluzinationen

Johannes Hötter,
Christian Warmuth

openHPI

Woher stammen Halluzinationen?



Quelle: NZZ



Modell ein "statistischer Papagei"
(engl. "stochastic parrot")

Halluzinationen

Johannes Hötter,
Christian Warmuth

openHPI

Was kann man dagegen machen?

- **Parameter für "Kreativität" anpassen**
(ChatGPT Parameter "Temperature" von 0-1)
- **Prompt Engineering:**
("Denke dir keine Quellen aus...")
- **Externe Quellen mit einbeziehen:**
(Vector Datenbanken,...)
- **Neue Modelle:**
(meist bereits deutlich besser & "reflektierter")

Halluzinationen

Johannes Hötter,
Christian Warmuth

openHPI

Ein Beispiel



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 KI verfüge ich nicht über die Fähigkeit, auf aktuelle Veröffentlichungen zuzugreifen oder DOI-Nummern bereitzustellen. Allerdings kann ich einige wichtige Veröffentlichungen im Bereich Multiferroika bis zum Jahr 2021 hervorheben, die in einer Dissertation zu diesem Thema zitiert werden könnten. Hier sind einige davon:

1. Fiebig, M., Lottermoser, T., Meier, D. & Trassin, M. (2016). The evolution of multiferroics. Nature Reviews Materials, 1, 16046.
2. Eerenstein, W., Mathur, N. D. & Scott, J. F. (2006). Multiferroic and magnetoelectric materials. *Nature*, 442, 759–765.
3. Wang, K. F., Liu, J.-M., & Ren, Z. F. (2009). Multiferroicity: the coupling between magnetic and polarization orders. *Advances in Physics*, 58(4), 321–448.
4. Khomskii, D. (2009). Classifying multiferroics: Mechanisms and effects. *Physics*, 2, 20.
5. Cheong, S.-W. & Mostovoy, M. (2007). Multiferroics: a magnetic twist for ferroelectricity. *Nature Materials*, 6, 13–20.

Bitte überprüfen Sie die genauen bibliografischen Daten und DOIs dieser Artikel in den entsprechenden Veröffentlichungsdatenbanken. Es ist auch wichtig, neuere Arbeiten zu berücksichtigen, die seit meinem letzten Training (September 2021) veröffentlicht wurden. Die oben genannten Artikel sind grundlegende Arbeiten zum Verständnis von Multiferroika, aber es gibt sicherlich auch neuere Arbeiten, die relevant sind. Es ist ratsam, Datenbanken wie Google Scholar, Web of Science oder PubMed für die aktuellsten Forschungsarbeiten zu durchsuchen.

Richtiger Artikel:

Fiebig, M., Lottermoser, T., Meier, D. *et al.* The evolution of multiferroics. *Nat Rev Mater* 1, 16046 (2016).
<https://doi.org/10.1038/natrevmats.2016.46>



Halluzinationen

Johannes Hötter,
Christian Warmuth

openHPI

Folie 7

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

Halluzinationen

Johannes Hötter
Christian Warmuth