

https://formulare.htwk-leipzig.de/db/1/app/16/start/wf/15/dt/20/form?COM_ID=1&AttChoose2=2211021005



Oberseminar

Vertrauenswürdige Künstliche Intelligenz

Prof. Dr. Sibylle Schwarz
HTWK Leipzig, Fakultät IM
Gustav-Freytag-Str. 42a, 04277 Leipzig
Zimmer Z 411 (Zuse-Bau)
<https://informatik.htwk-leipzig.de/schwarz>
sibylle.schwarz@htwk-leipzig.de

Wintersemester 2021/22

Vertrauenswürdigkeit

Wann ist ein Agent / System vertrauenswürdig?

- ▶ Zuverlässigkeit: tut, was es soll
- ▶ Sicherheit: richtet keinen Schaden an
- ▶ Robustheit gegenüber Angriffen
- ▶ Schutz persönlicher Daten
- ▶ Verfügbarkeit
- ▶ einfache Nutzbarkeit

Speziell für KI-Systeme

- ▶ Genauigkeit bei unbekanntem Eingaben
- ▶ Stabilität gegenüber geringen Änderungen der Eingaben
- ▶ Fairness, kein Bias
- ▶ Klarheit der Verantwortung für Lösungen (Haftbarkeit)
- ▶ Transparenz / Verständlichkeit der Funktionalität und Ausgaben
- ▶ Erklärung / Begründung der Lösungen
- ▶ Ethische Verantwortung bei Daten-Sammlung und -Verwendung

Vortragsthemen

zu den Themengebieten:

- ▶ Probleme
- ▶ Ansätze
- ▶ Explainable AI

jeweils mit Link im Wiki im Opal-Kurs (Themen)

<https://bildungsportal.sachsen.de/opal/auth/RepositoryEntry/32061358101>

dort (Termine) bitte auch Zuordnung
Termin - Thema - Vortragender eintragen

zwei Artikel sind lang (markiert) und können geteilt werden

Organisation

12 Teilnehmer und 1 + 13 Termine

also meist je Woche ein Vortrag
bei langen Beiträgen 2 Vorträge je Termin

jeder Vortrag 45-60 min (+ 30-45 min Diskussion)

Jeder muss seinen Vortrag bei Bedarf schon eine Woche vor seinem geplanten Termin halten können.

rechtzeitige Vorbereitung notwendig

Vortrag

Der Vortrag muss alle folgenden Fragen klar und kompakt beantworten:

- ▶ Welche Aufgabe wird gelöst?
Eingaben, Ausgaben, Qualitätskriterien
- ▶ Welche Verfahren / Werkzeuge werden vorgestellt?
Grundidee, Funktionsweise
- ▶ Welche aus Ihrem Studium bekannten Ideen werden dabei verwendet?
- ▶ Warum lösen die vorgestellten Verfahren die Aufgabe?
- ▶ Wie gut lösen die vorgestellten Verfahren die Aufgabe?
- ▶ Wie lassen sich die vorgestellten Verfahren verbessern / ergänzen?

Nachbereitung:

- ▶ Folien (ggf. korrigiert) in OPAL zur Verfügung stellen
Der nächste Vortragende verweist i.A. darauf.

Bewertungskriterien Vortrag

- ▶ Inhalt
- ▶ Struktur
- ▶ Anspruch, Wissenschaftlichkeit
- ▶ Einordnung in die Gesamt-LV
- ▶ Verständnis
- ▶ Verständlichkeit (inhaltlich, akustisch)
- ▶ sinnvolle Beispiele
- ▶ Präsentation (Folien)
 - ▶ logischer Aufbau
 - ▶ Lesbarkeit (hinreichend große Schrift, auch in Diagrammen)
 - ▶ Rechtschreibung / Grammatik / Ausdruck
 - ▶ Quellenangaben
- ▶ Länge des Vortrags (ca. 45-60 min einschließlich Aufbauen, Beispielen, Demos)
- ▶ Reaktionen in der Diskussion

Bewertungskriterien außerdem

- ▶ regelmäßige aktive Teilnahme an Vorträgen der anderen
sinnvolle Fragen, kein störendes Tastatur-Klappern
- ▶ Zeitmanagement:
rechtzeitige Vorbereitung
(jeder muss seinen Vortrag bei Bedarf schon eine Woche vor
seinem Termin halten können)
zügiges Hochladen der Präsentation nach dem Vortrag
- ▶ Konsultation:
ca. 2 Wochen vor Vortragsdatum
rechtzeitige Terminanfrage (per Email an mich)
aussagefähige Vorversion der Präsentation wenigstens 2 Tage
vor Konsultationstermin (per Email an mich)
Hinweise aus der Konsultation beachten