



## Meldung einer Masterarbeit

### Studierender

---

Matrikelnummer: 12144056  
Familiennamen: Kerl  
Vorname: Tilman

### Studium

---

Typ: Master  
Studium: 066645 Data Science

### Arbeit

---

Titel DE: Evaluation of Sparse Autoencoder-based Refusal Features in LLMs: Dataset-dependence study  
Titel EN: Evaluation of Sparse Autoencoder-based Refusal Features in LLMs: Dataset-dependence study  
Institut: E194 - Institut für Information Systems Engineering

### Betreuung

---

Hauptbetreuung: Associate Prof. Dipl.-Ing. Dr.techn. Peter Knees  
Mitbetreuung: Univ.Ass. Ilya Lasy MSc

8.3.2025

Datum, Unterschrift Studierender

7.3.2025

Datum, Unterschrift Betreuer

-----  
Datum, Unterschrift Studiendekan\_in