

PROGRAMM AM 15. OKTOBER 2026

09:00 - 09:15

Begrüßung

09:15 - 10:00

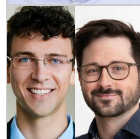


Software für lokale LLMs

Christian Winkler

datanizing

10:15 - 11:00

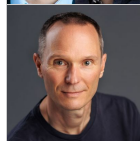


Wie viel GPU braucht der Chatbot?

Benjamin Merkel & Lukas Rammelmüller

TNG

11:15 - 12:00



Performance-Optimierung

Rainer Stropek

software architects

12:15 - 13:15

Mittagspause

13:15 - 14:00

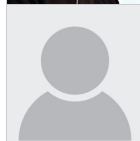


LLMs und Embedding-Modelle, selbstgebaut

Johann-Peter Hartmann

Mayflower

14:15 - 15:00



Coding-Agenten im Enterprise-Einsatz bewerten: das System, nicht die Funktion

Jan Jikeli

mgm

15:15 - 15:30

Kaffeepause

15:30 - 16:15



GPU oder API? Die tatsächlichen Kosten von KI-Inferenz und wie man flexibel bleibt

Christopher Haar & Johannes Koch

upbound, FICO

WORKSHOP: 22. OKTOBER 2026

09:00 - 16:00



Eigene Sprachmodelle nutzen und feintunen

Christian Winkler

datanizing

WORKSHOP: 18. NOVEMBER 2026

09:00 - 16:00



Lokale Sprachmodelle effizient betreiben

Christian Winkler

datanizing