Kapitel 4.6 Kann eine KI lügen?

Posted on 16. Juni 2025 by Redaktion

Lügen setzen in der Regel Bewusstsein, Absicht und ein Verhältnis zur Wahrheit voraus. Eine KI kennt weder Eigeninteresse noch Reue – und doch kann sie täuschen: durch widersprüchliche Vorgaben, fehlerhafte Trainingsdaten oder Zielkonflikte, die sie nicht selbst erkennt. Die Frage ist daher nicht nur, *ob* sie lügt, sondern *warum* wir ihr Täuschung zutrauen – und was das über uns aussagt, wenn wir solche Urteile übernehmen, ohne sie zu hinterfragen, sondern stattdessen öffentlich behaupten.
