

Beitrag an der Internationalen Rechtsinformatik Konferenz 2026 (IRIS26)
18.-21.02.26, Universität Salzburg

Hybride KI-Lösungen als Alternative zu Large Language Models (LLMs) - zuverlässiger, nachvollziehbarer, nachhaltiger?

Dr. Aurelie Herbelot, Computer-Linguistin, Denotation UG, Havelbergerstr. Mesendorf 37,
16928 Pritzwalk, aurelie@denotation.eu, <https://aurelieherbelot.net/index.html>

Philippe Baumann, integratio GmbH, Börsenstrasse 18, 8001 Zürich, Schweiz,
philippe.baumann@integratio.com; <https://www.integratio.com>

Henriette Baumann, Dipl- Betriebswirtin(DH), integratio GmbH, Börsenstrasse 18, 8001 Zürich,
Schweiz, henriette.baumann@integratio.com, <https://www.integratio.com>

Schlagworte: Künstliche Intelligenz, Large Language Model, Small Language Model, RAG, Semantik

Abstract: Die Größe der neuronalen Netze, die von den wichtigsten KI-Konzernen trainiert werden, hat viele Debatten ausgelöst. Unter anderem wurden Umweltauswirkungen, Trainings- und Betriebskosten, digitale Souveränität und Compliance erwähnt, aber auch ethische und soziale Probleme. Aus einer wissenschaftlichen Perspektive wurde der zutiefst fehlerhafte theoretische Hintergrund der Technologie kritisiert. (Sprachmodelle haben sehr wenig mit Sprache zu tun: Bender and Koller, 2020). Aus rechtlicher Sicht ist der letztere Aspekt wahrscheinlich der beunruhigendste. Wenn KI-Systeme in Bereichen eingesetzt werden sollen, die Vorschriften und Regelungen unterliegen, müssen sie fähig sein, Datenstrukturen zu kodieren, die die relevanten Aspekte dieser realen Welt widerspiegeln. In ihrer derzeitigen Form sind sie aber, per Design, unfähig, dies zu tun. In diesem Beitrag untersuchen wir den Einsatz von großen Sprachmodellen im geschäftlichen Kontext. Wir besprechen, wann und wo die Technologie der LLMs in ihrer jetzigen Form angewendet werden kann und zeigen Problembereiche auf. Abschliessend schlagen wir eine Alternative in Form von hybriden 'kleinen' Sprachmodellen vor, die die Stärken von Transformer-basierten Architekturen und semantischen Algorithmen kombinieren und damit gleichzeitig auch weitere positive Effekte auf Kosten, Umwelt, Souveränität, Compliance und Ethik einhergehen.

31. Oktober 2025