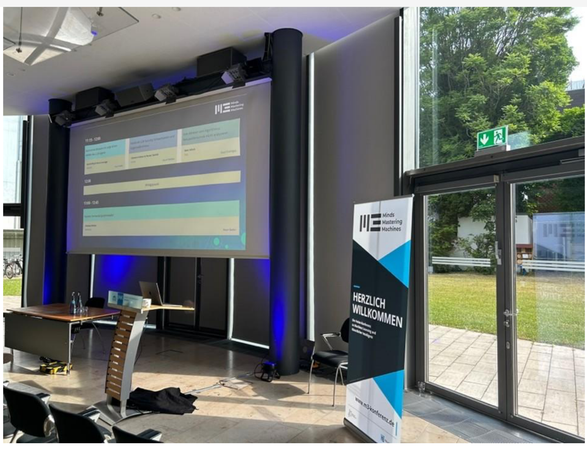

M3 KONFERENZ: MACHINE LEARNING UND KI IN DER PRAXIS

Vom 19. bis 21. Mai fand in Karlsruhe die achte Ausgabe der Minds Mastering Machines Konferenz, kurz M3, statt. Auf der üblichen Liste der Marktforschungs-Events steht sie nicht, doch die enorm gewachsene Relevanz von Machine Learning für die Marktforschung hat uns neugierig gemacht. Auf der Konferenz treffen sich Entwicklerinnen und Entwickler von IT-Dienstleistern und aus den IT-Abteilungen verschiedenster Unternehmen, um neue Entwicklungen zu besprechen.

Eine pragmatische Sicht

Es ist eine Konferenz der Pragmatiker: der Fokus liegt auf der praktischen Umsetzung von KI-Lösungen, statt auf luftigen Zukunftsvisionen. Das Erdbeer-Beispiel wurde viel zitiert: ein LLM soll zählen, wie häufig der Buchstabe "r" im englischen Wort "strawberry" vorkommt. Viele Modelle zählen falsch - trotz hoher Werte in verschiedenen Intelligenztests. Bei dieser ungewöhnlichen, aber doch trivialen Aufgabe zeigt sich, dass sie eben doch nicht denken, sondern es nur clever simulieren.
Nichtsdestotrotz ist eine verblüffend gute Simulation von Intelligenz nützlich. Unter anderem stellten die Techniker Krankenkasse und Kühne + Nagel Ihre ausgefeilten KI-Infrastrukturen vor. Sebastian Rabe rechnete vor, dass die 80.000 Kühne + Nagel Mitarbeitenden viel Zeit sparen werden. Statt manuell zu suchen, können Sie einen Chatbot befragen, der mittels Retrieval Augmented Generation (RAG) auf Unternehmensdaten zugreifen kann.

Evaluation: quantitativ und qualitativ

Die verlässliche und sichere Bereitstellung solcher RAG-Pipelines und Chatbots war ein Schwerpunkt der Konferenz. Wie wird gemessen, ob die richtigen Dokumente gefunden werden und ob die generierte Antwort den Fakten entspricht? Dazu stellte Tim Wüllner von Open Knowledge diverse Metriken vor. Weit verbreitet ist das Generieren von synthetischen Testdaten und eine Bewertung der Antworten durch ein zweites KI-System, das nennt sich LLM as Judge. Das ist skalierbar, erfordert aber das Vertrauen in eine zweite KI-Lösung. Tim Wüllner argumentiert daher, dass die Kollaboration mit Domain-Expertinnen und Experten essenziell ist. Sie haben das Fachwissen und den nötigen Praxisbezug. Besonders eindrucksvoll zeigten das Alina Döring und Robin Senge von inovex, die gemeinsam mit der Helios Klinik Wiesbaden eine KI für den OP-Saal entwickeln. Eine Benutzeroberfläche zum Klicken funktioniert nicht, wenn die operierende Ärztin keine Hand frei hat, um sie zu bedienen. Zentral bei ihrem Projekt sind semi-strukturierte Interviews - und so zeigt sich das Marktforschung und Machine Learning viel gemeinsam haben.

Zahlreiche Sicherheitsrisiken in KI-Systemen

Neben der Korrektheit war auch die Sicherheit von KI-Systemen ein Schwerpunkt. OWASP, eine führende Organisation für Softwaresicherheit hat im November eine Liste der Top 10 Verwundbarkeiten von LLM-basierter veröffentlicht. Dazu trugen Johann-Peter Hartmann von Mayflower und Clemens Hübner und Florian Teutsch von inovex vor. Sie zeigten an einfachen Beispielen, wie mittels Prompt Injection und anderen missbräuchliche Anweisungen an ein LLM geschickt werden können. Die Beispiele zeigten teils schockierende Sicherheitslücken, insbesondere wenn grundlegende Prinzipien, wie das Prinzip der geringsten Privilegien, im KI-Kontext missachtet werden.

Reasoning, Agenten und das Modell Context Protocol

Steve Haupt von andrena objects und Christian Winkler von datanizing stellten neue Entwicklungen bei Modellen und umliegenden Systemen vor. Mit dem Paradigma "Reasoning" formulieren Modelle vor ihrer endgültigen Antwort Gedanken. Damit erzielten die Modelle o1-preview von OpenAI und insbesondere das chinesische DeepSeek R1 Modell einen Durchbruch. Andere LLM-Schmieden zogen nach und hoben das Niveau an, insbesondere bei Logik-Aufgaben. Der Preis für intelligentere Antworten ist mehr Rechenaufwand. Doch auch bei der Effizienz tut sich viel. Stefan Kühn von Sporting Rock gab einen Ausblick auf Liquid Foundation Models, eine experimentelle, ressourcenschonende Alternative zur Transformer-Architektur.
Neben Fortschritten bei den Modellen tut sich auch viel bei den Software-Systemen, in denen sie eingebettet sind. Als Agenten werden Systeme verstanden, die LLMs mit Werkzeugen ausstatten, selbstständig in mehreren Schritten arbeiten lassen und sogar Sub-Agenten beauftragen können. Häufige Anwendungen sind Recherche und Programmieren. Bisher sind diese Systeme jedoch experimentell, da mit jedem Arbeitsschritt und jedem neuen Werkzeug auch neue Fehlerquellen hinzukommen.
Die Ausstattung der Agenten und auch von Chat-Anwendungen mit Werkzeugen wird mit dem im November 2024 von Anthropic herausgebrachten Model Context Protocol vereinfacht. Christian Winkler von datanizing stellte es auf der M3 vor. MCP standardisiert die Schnittstelle zwischen Modellen und Werkzeugen jeder Art, von Websuche bis zur Kontrolle eines Smart Homes. Die Webseite MCP.so listet bereits über 13.000 Integrationen.

Fazit: KI muss messbar und verlässlich sein

Von Tech Demos allein lässt sich bei der M3 kaum jemand beeindrucken. Die Währung ist Produktionsreife und messbarer Nutzen für das Unternehmen. Neue Entwicklungen werden aufmerksam verfolgt und ausprobiert, aber auch kritisch evaluiert. Dabei zeigt sich, dass die Einführung von KI und das Erwartungsmanagement ein Prozess für das ganze Unternehmen ist, und nicht nur ein Projekt der IT.

Autor: Paul Simmering