

# LLM BENCHMARKING UND BEST-OF-BREED: Wie Insiders die passende KI-Lösung für Sie findet

## Versicherungsprozesse der Zukunft

Max Mustermann ist umgezogen und muss bei seiner Versicherung seine Adresse ändern – am liebsten per App oder Email. Adressänderungen wie diese sind ein Standardprozess, der sowohl bei seinem als auch bei anderen Versicherern zig 1.000 Mal in der Woche vorkommt – und dabei hohe manuelle Aufwände und ein großes Fehlerpotenzial mit sich bringt. Die Lösung? Intelligente Prozessautomatisierung von Insiders mit der Power von LLMs!

Large Language Models (LLMs) wie Claude, Llama und GPT sprießen wie Pilze aus dem Boden, sind alle unterschiedlich gut für spezifische Use Cases geeignet und bringen wiederum eine Vielzahl an Vor- und Nachteilen mit sich. Aber welches LLM ist nun das Beste für Ihre Herausforderungen...?

Angesichts der rasanten Entwicklungen auf dem LLM-Markt stehen Unternehmen vor der Qual der Wahl, oder? Nicht die Insiders-Kunden! Mit unserem Best-of-Breed-Ansatz nutzen Kunden in allen Insiders Produkten immer das aktuell beste LLM für ihre individuellen Bedürfnisse.

Möglich wird das durch ein konsequentes und kontinuierliches LLM Benchmarking und ein damit verbundenes Monitoring der unterschiedlichen Modelle auf dem globalen Technologiemarkt. Einblicke in diese zentralen Ergebnisse bietet das Insiders LLM Benchmarking.

Das Insiders LLM Benchmarking zeigt nicht nur die aktuellen Top-Performer unter den LLMs auf, sondern bietet auch einen anschaulichen Überblick über alle im Data to Process-Bereich relevanten Modelle. Dabei ist es nicht immer die beste Lösung, ausschließlich auf die leistungsstärksten Modelle zu setzen. Je nach Ihren individuellen Anforderungen kann ein Modell mit besonderen Eigenschaften, wie das Insiders Private LLM mit seinem Fokus auf höchsten Datenschutz, besser zu Ihrem Unternehmen passen.

## LLM Benchmarking – die Grundlage der Innovationskraft

Was bedeutet ein LLM Benchmarking bei Insiders? Unser Benchmarking orientiert sich an einem spezialisierten IDP-Benchmark, der aus unserer langjährigen Erfahrung als erfolgreiches KI- und Softwareunternehmen hervorgeht. Als DFKI-Spin-Off haben wir besonders in der Versicherungs- und Finanzbranche umfassendes Know-how aufgebaut. Die standardisierten Insiders Testdaten werden so ausgewählt, dass sie ein breites Feld der typischen Geschäftsvorgänge in der Versicherungs- und Finanzbranche abdecken. Sie ermöglichen so eine objektive Bewertung der Gesamtleistung eines LLMs in den gefragten Bereichen. So sind in der Testmenge unter anderem Adress- und Namensänderungen, Prämienrechnungen, Schadensberichte, SEPA-Mandate und medizinische Dokumente enthalten, die die Grundlage für eine Vielzahl an gängigen Geschäftsprozessen bilden.

Das Insiders LLM Benchmarking ist ein kontinuierlicher Prozess, der den Best-of-Breed-Ansatz antreibt. So bleiben wir stets über die Performance der neuesten LLMs informiert und gewährleisten, dass Kunden dank der flexiblen LLM-Integration der Insiders OvAltion Engine immer die optimale Lösung für ihre individuellen Bedürfnisse und Anforderungen nutzen können. Die Insiders KI-Experten überwachen laufend die leistungsfähigsten Technologien und passen die LLM-Integration in allen Produkten entsprechend an die Markt- und Technologieentwicklungen an.

Mit dem Best-of-Breed-Ansatz gibt Insiders seinen Kunden somit die Möglichkeit, hybride Architekturen zu schaffen, die das Beste aus der Welt der externen LLMs und der bewährten und technologisch führenden Insiders-KI-Technologien vereinen – und zwar ohne dass sich die Kunden selbst zeitintensiv mit den rasanten und komplexen Entwicklungen auf dem globalen LLM-Markt beschäftigen müssen.

Stand: 29.01.2025

Über die Insiders OvAltion Engine sind zudem gleich mehrere LLMs parallel eingebunden, sodass für jeden Anwendungsfall die optimalen Modelle genutzt werden. Weitere integrierte Funktionen wie das Green Voting sorgen dafür, dass die Ergebnisse verschiedener Modelle validiert und somit die Dunkelverarbeitungsquote für Insiders Kunden weiter gesteigert werden kann. Manuelle Bearbeitungsaufwände können mit diesem Ansatz drastisch gesenkt werden.

### Ein Blick auf die aktuellen Ergebnisse

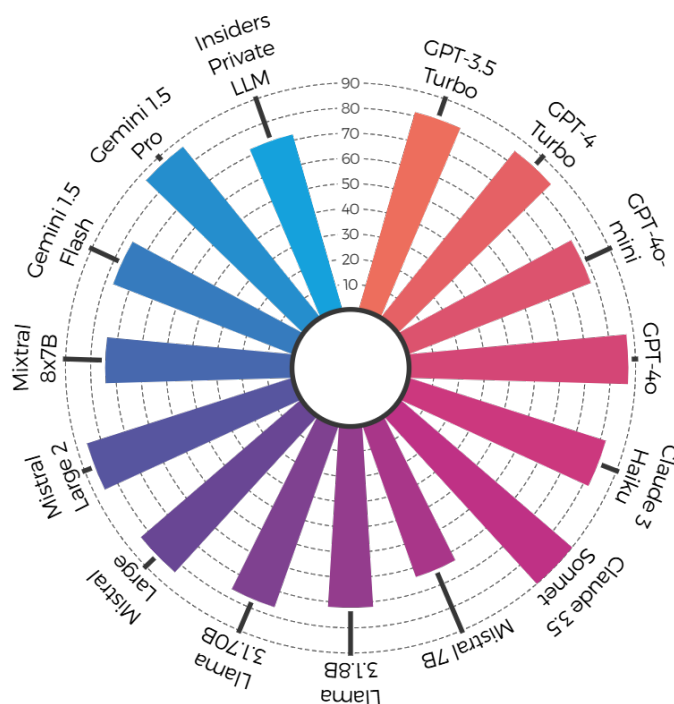
Das Insiders LLM Benchmarking bildet regelmäßig die Gesamtperformance verschiedener LLMs ab. In diesem standardisierten Performancecheck werden diejenigen LLMs getestet, die von den Insiders KI-Experten in einer Vorauswahl als besonders geeignet für den Einsatz im Data to Process-Bereich identifiziert wurden. Das umfasst u.a. ihre jeweilige Performance in den Arbeitsfeldern Informationsklassifikation und -extraktion. Im aktuellen Insiders LLM Benchmarking wurden 15 verschiedene Modelle getestet, darunter bekannte Namen wie GPT-4 Turbo, Claude 3.5 Sonnet und Mistral Large sowie das Insiders Private LLM. Das Ranking der Gesamtperformance, in die sowohl die Leistung im Bereich Extraktion als auch Klassifikation einfließen, er-

möglicht dadurch einen Vergleich der LLMs in Bezug auf die jeweilige Leistungsfähigkeit in Relation zu den verbleibenden Nachbearbeitungsaufwänden.

Das aktuelle Insiders LLM Benchmarking zeigt, dass die bekannten US-amerikanischen LLM-Anbieter aufgrund ihrer immensen Rechenpower und der massiven Menge an Trainingsdaten wie erwartet die Nase vorn haben. Auf Platz 1 im Gesamtranking landet das Modell Claude 3.5 Sonnet von Anthropic mit einem Score von 90,10, dicht gefolgt vom Medienstar GPT-4o von OpenAI mit 87,11. Den dritten Platz belegt Gemini 1.5 Pro von Google mit 86,98.

Im Vergleich mit den LLMs der US-Tech-Giganten erreicht das Insiders Private LLM zwar eine geringere Gesamtperformance von insgesamt 72,33, punktet jedoch durch seine unübertroffenen Datenschutzstandards. Dies ist seitens Insiders ein ganz bewusster Tradeoff, der speziell für informationsensible Branchen wie Versicherungen und Finanzen entscheidend ist. Hier müssen sensible Daten wie Personalausweise, SEPA-Mandate oder medizinische Informationen schnell und zuverlässig verarbeitet werden, ohne dabei Kompromisse beim Datenschutz einzugehen.

Model	Performance
Claude 3.5 Sonnet	90,10
GPT-4o	87,11
Gemini 1.5 Pro	86,98
Mistral Large 2	85,64
GPT-4 Turbo	84,83
Mistral Large	83,78
Claude 3 Haiku	82,21
GPT-3.5 Turbo	81,55
GPT-4o-mini	77,99
Gemini 1.5 Flash	76,76
Llama 3.1 70B	76,46
Mixtral 8x7B	74,22
Insiders Private LLM	72,33
Llama 3.1 8b	72,01
Mistral 7B	63,73



Stand: 29.01.2025

Das Insiders Private LLM wird in unserer Insiders Cloud eigenständig betrieben und unterliegt dem bekannten, hohen Datenschutzlevel. Während große externe LLMs zudem oft von Faktoren wie Drosselungen oder Rate Limits betroffen sind und der Datenfluss nicht immer transparent nachvollziehbar ist, bietet das Insiders Private LLM volle Unabhängigkeit und Skalierbarkeit nach den individuellen Kundenbedürfnissen. Mit einer ISO 27001-zertifizierten Infrastruktur und der nahtlosen Integration in bestehende Systeme garantiert das Insiders Private LLM maximale Sicherheit und volle Kontrolle für Kunden mit höchsten Datenschutzansprüchen.

### Was bedeutet also „das beste LLM“?

Das aktuelle Insiders LLM Benchmarking zeigt, dass Insiders als KI-Experte den LLM-Markt genau im Blick hat und den Balanceakt zwischen Performance und Sicherheit für seine Kunden mit dem Best-of-Breed -Ansatz meistert.

Der Best-of-Breed-Ansatz bedeutet, dass Insiders in all seinen Produkten die leistungsfähigsten LLMs auf dem Markt identifiziert und flexibel integriert. Erscheint ein neues Modell, wird dieses im aktuellen Insiders LLM Benchmarking auf die Probe gestellt und mit den anderen LLMs verglichen. Die gewonnen Erkenntnisse fließen umgehend in die Produktentwicklung ein und stellen eine dauerhaft hohe Qualität für die Insiders Kunden sicher. Basierend auf der aktuellen Gesamtleistung liegt dabei im Januar 2025 Claude 3.5 Sonnet von Anthropic ganz vorne.

Allerdings geht es bei der Frage nach „dem Besten“ nicht nur um die reine Power. Gerade im Versicherungs- und Finanzbereich müssen Unternehmen oft zwischen Leistung und Sicherheit abwägen. Hinzu kommen individuelle Vorgaben und Präferenzen der Unternehmen in Bezug auf Kosten, Geschwindigkeit und Dunkelverarbeitung.

Durch die Kombination aus state-of-the-art Deep Learning-Verfahren, tiefgehender LLM-Expertise und über 25 Jahren Branchenerfahrung im Finanz- und Versicherungsmarkt bietet Insiders seinen Kunden leistungsfähige Automatisierungslösungen, die speziell auf die Bedürfnisse sicherheitskritischer Anwendungsfälle zugeschnitten sind und gleichzeitig höchstmögliche Performance liefern. So kann KI im Unternehmen sinnvoll und sicher eingesetzt werden.

Je nach individuellen Anforderungen im Spannungsfeld von Performance, Latenz, Dunkelverarbeitung und Kosten ermöglicht Insiders durch die Integration von LLMs von Drittanbietern seinen Kunden einen bequemen und variablen Zugang zu genau dem LLM, das sie wirklich brauchen. Insiders Kunden müssen sich also nicht zwischen Performance und Sicherheit entscheiden. Sie erhalten beides, angepasst an ihre individuellen Bedürfnisse – und bleiben dabei flexibel, um die vielfältigen Einsatzmöglichkeiten von LLMs für Ihr Unternehmen zu erobern.

Mit der ISO-zertifizierten Infrastruktur und der nahtlosen Integration über die Insiders OvAltion Engine bietet Insiders mit OmnIA eine Automatisierungs-Plattform, die nicht nur sicher, sondern auch zukunftssicher ist.

Das Insiders LLM Benchmarking ist dabei der verlässliche Referenzpunkt, um im schnelllebigen LLM-Markt den Durchblick zu behalten.

*Für individuelle Benchmarkings mit eigenen Use Cases bieten die Insiders KI-Experten eine fundierte Beratung für Ihr Unternehmen an. Kommen Sie für ein individuelles Benchmarking einfach auf unsere Insiders KI-Experten zu.*