

# LLM BENCHMARKING AND BEST-OF-BREED: How Insiders identify the right AI solution for you

## Insurance Processes of the Future

Max Mustermann has moved and needs to update his address with his insurance company – preferably via app or email. Address changes like this are standard processes that occur thousands of times a week, both at his insurer and others, and they come with high manual effort and a significant potential for errors. The solution? Intelligent process automation from Insiders powered by LLMs!

Large Language Models (LLMs) such as Claude, Llama, and GPT are popping up everywhere. Each is suited to specific use cases and brings a variety of advantages and disadvantages. But which LLM is the best fit for your challenges...?

Given the rapid developments in the LLM market, companies are faced with a tough choice, right? Not Insiders' customers! With our Best-of-Breed approach, customers always benefit from the best LLM for their individual needs in all Insiders products.

This is made possible by consistent and continuous LLM benchmarking and monitoring of various models in the global technology market. Insights into these key results are provided by Insiders' LLM Benchmarking.

Insiders' LLM Benchmarking not only highlights the current top performers among LLMs but also offers a clear overview of all relevant models in the Data to Process space. However, the best solution isn't always about relying solely on the most powerful models. Depending on your specific needs, a model with special characteristics, like the Insiders Private LLM, which focuses on the highest data privacy standards, may be a better fit for your business.

## LLM Benchmarking – The Foundation of Innovation Power

What does LLM benchmarking at Insiders mean? Our benchmarking is based on a specialized IDP benchmark, built from our extensive experience as a successful AI and software company. As a spin-off of DFKI, we have developed comprehensive expertise, particularly in the insurance and financial sectors. The standardized Insiders test data are selected to cover a broad range of typical business processes in the insurance and financial industries. This allows for an objective evaluation of the overall performance of an LLM in the most relevant areas. Among the test data, for example, are address and name changes, premium invoices, damage reports, SEPA mandates, and medical documents, all of which form the basis for a variety of common business processes.

The Insiders LLM benchmarking is an ongoing process that drives the Best-of-Breed approach. This ensures that we are always informed about the performance of the latest LLMs and guarantees that customers can always use the optimal solution for their individual needs and requirements, thanks to the flexible LLM integration in the Insiders OvAltion Engine. Our AI experts continuously monitor the most powerful technologies and adjust the LLM integration across all products in response to market and technological developments.

With the Best-of-Breed approach, Insiders gives its customers the opportunity to create hybrid architectures that combine the best of external LLMs and Insiders' proven, technology-leading AI solutions – all without requiring customers to invest time in the rapidly evolving and complex developments of the global LLM market.

As of: 29.01.2025

Through the Insiders OvAltion Engine, multiple LLMs are integrated in parallel, ensuring that the optimal models are used for each specific use case. Additional integrated features, such as Green Voting, help validate the results of different models, further increasing the dark processing rate for Insiders' customers. This approach can drastically reduce manual processing efforts.

### An Overview of the Current Results

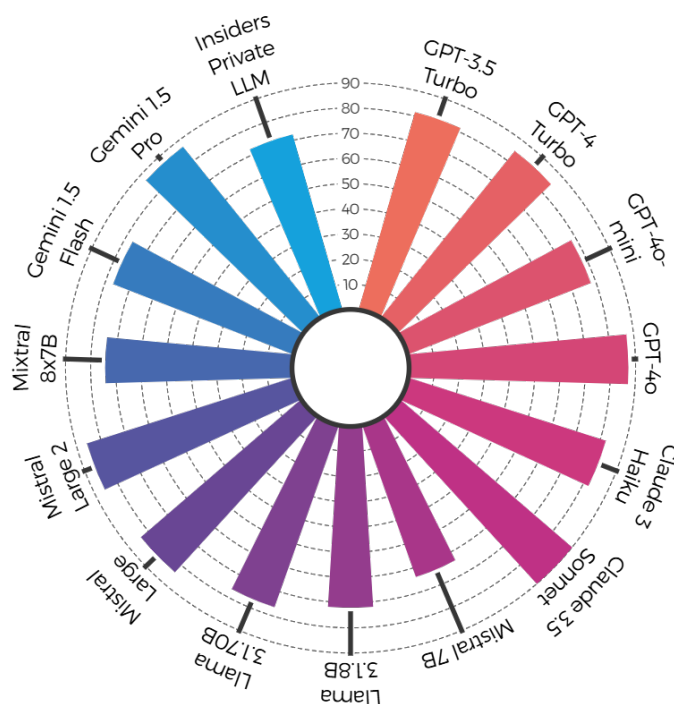
The Insiders LLM Benchmarking regularly assesses the overall performance of various LLMs. In this standardized performance check, those LLMs that have been pre-selected by Insiders' AI experts as particularly suitable for use in the Data to Process field are tested. This includes evaluating their performance in tasks like information classification and extraction. In the current Insiders LLM Benchmarking, 15 different models were tested, including well-known names such as GPT-4 Turbo, Claude 3.5 Sonnet, Mistral Large, and the Insiders Private LLM. The ranking of overall performance, which takes into account both extraction and classification performance, enables a comparison of the LLMs based on their respective capabilities and the remaining post-processing efforts required.

The latest Insiders LLM Benchmarking shows that the well-established US-based LLM providers, due to their immense computing power and vast amounts of training data, lead the pack as expected. Taking the top spot in the overall ranking is Anthropic's Claude 3.5 Sonnet, with a score of 90.10, closely followed by the media sensation GPT-4o from OpenAI with 87.11. In third place is Google's Gemini 1.5 Pro with 86.98.

Compared to the LLMs from the US tech giants, the Insiders Private LLM achieves a lower overall performance of 72.33, but stands out for its unparalleled data privacy standards. This is a conscious tradeoff on Insiders' part, especially critical for information-sensitive industries like insurance and finance. In these sectors, sensitive data such as identity cards, SEPA mandates, or medical information must be processed quickly and reliably, without compromising on privacy.

The Insiders Private LLM is operated independently within our Insiders Cloud and adheres to the well-known high data privacy standards. While large external LLMs are often affected by factors such as throttling or rate limits, and their data flow is not always transparently traceable, the Insiders Private LLM offers complete independence and scalability tailored to individual custo-

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



As of: 29.01.2025

mer needs. With an ISO 27001-certified infrastructure and seamless integration into existing systems, the Insiders Private LLM guarantees maximum security and full control for customers with the highest data privacy requirements.

### What does „the best LLM“ mean?

The latest Insiders LLM Benchmarking shows that Insiders, as an AI expert, is keeping a close eye on the LLM market and successfully managing the balancing act between performance and security for its customers with the Best-of-Breed approach.

The Best-of-Breed approach means that Insiders identifies and flexibly integrates the most powerful LLMs available on the market into all of its products. When a new model is released, it is tested in the current Insiders LLM Benchmarking and compared with other models. The insights gained are immediately incorporated into product development, ensuring consistently high quality for Insiders' customers. Based on the current overall performance, in January 2025, Anthropic's Claude 3.5 Sonnet is leading the pack.

However, the question of „the best“ isn't just about raw power. Particularly in the insurance and finance sectors, companies often have to balance performance and security. Additionally, there are individual company requirements and preferences regarding cost, speed, and dark processing.

Through a combination of state-of-the-art deep learning techniques, deep LLM expertise, and over 25 years of industry experience in the financial and insurance markets, Insiders offers its customers powerful automation solutions tailored to the needs of security-critical use cases while also delivering the highest possible performance. This ensures that AI can be used meaningfully and securely within companies.

Depending on individual requirements in the areas of performance, latency, dark processing, and cost, Insiders, through the integration of third-party LLMs, provides its customers with convenient and flexible access to exactly the LLM they need. Insiders' customers don't have to choose between performance and security; they get both, tailored to their specific needs, while staying flexible to leverage the diverse applications of LLMs for their business.

With an ISO-certified infrastructure and seamless integration through the Insiders OvAltion Engine, Insiders offers OmniA, an automation platform that is not only secure but also future-proof.

The Insiders LLM Benchmarking is a reliable reference point to stay on top of the rapidly evolving LLM market.

*For individual benchmarkings with your own use cases, the Insiders AI experts offer comprehensive consulting for your company. Simply reach out to our Insiders AI experts for personalized benchmarking.*