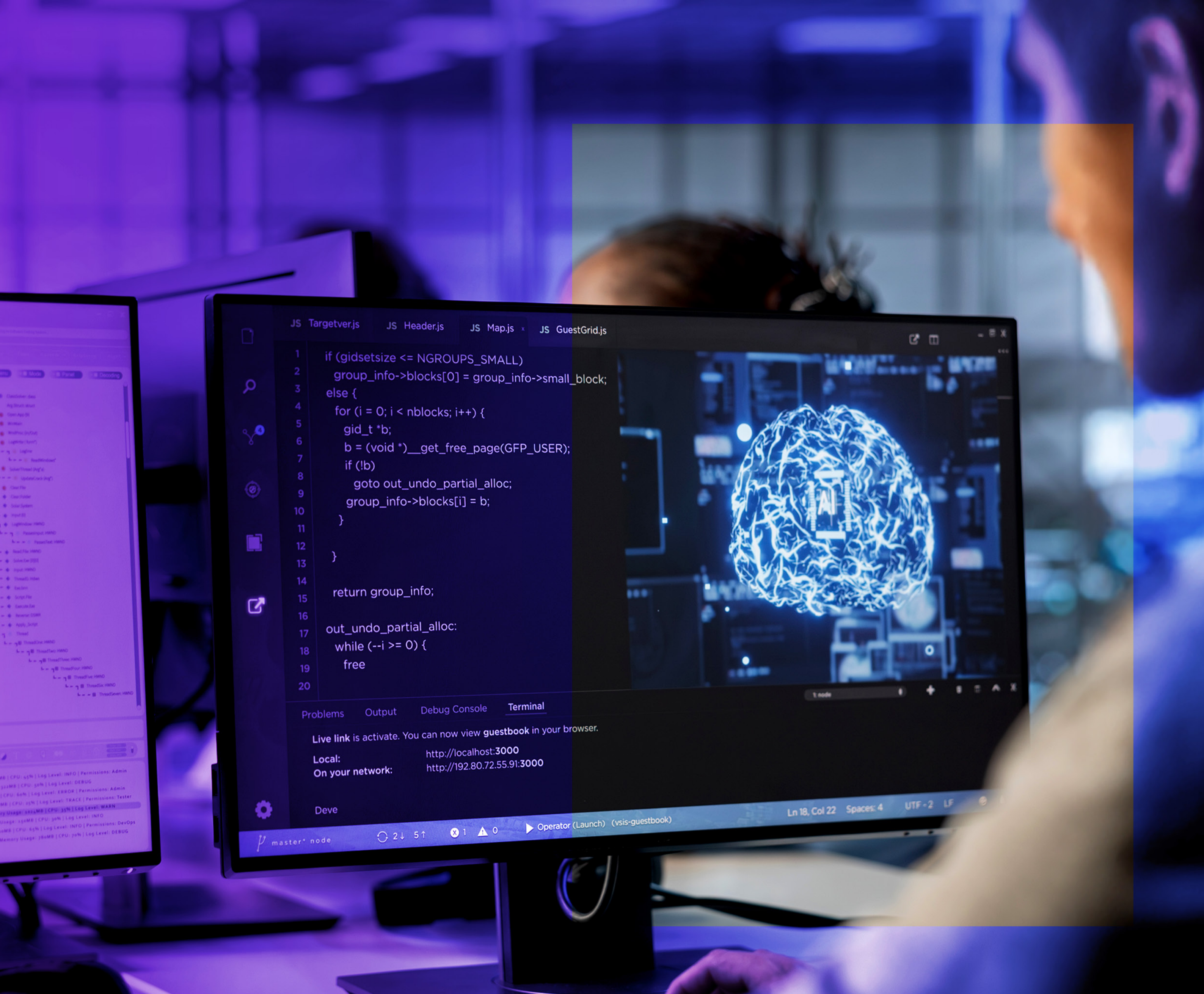
Evaluation of Generative AI Using the Example of Insurance Companies

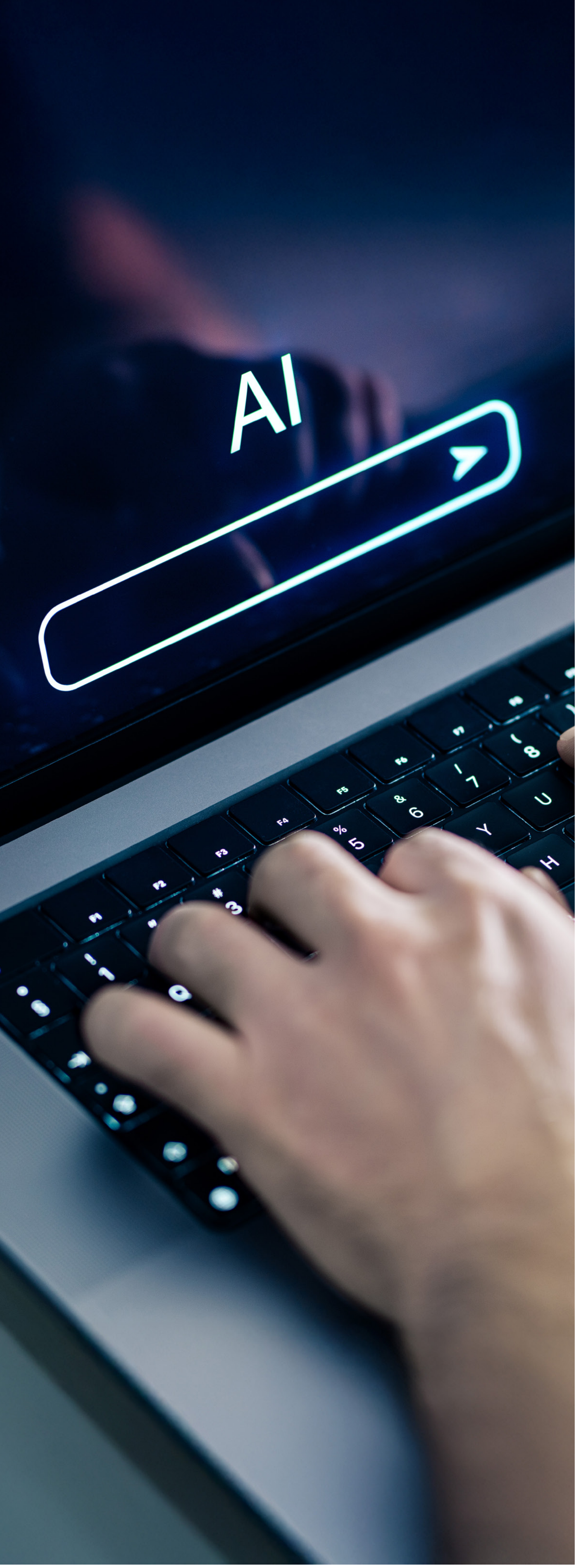
How the quality of generative AI can be objectively and quantitatively assessed and improved

Short study



```
if (gidsetsize <= NGROUPS_SMALL)
  group_info->blocks[0] = group_info->small_block;
else {
  for (i = 0; i < nblocks; i++) {
    gid_t *b;
    b = (void *)__get_free_page(GFP_USER);
    if (!b)
      goto out_undo_partial_alloc;
    group_info->blocks[i] = b;
  }

}

return group_info;

out_undo_partial_alloc:
  while (--i >= 0) {
    free
```

```
Live link is activate. You can now view guestbook in your browser.

Local:                 http://localhost:3000
On your network:       http://192.80.72.55.91:3000

Deve
```



Introduction

Currently, generative artificial intelligence (generative AI) is revolutionizing numerous industries, including the insurance sector, its actuarial departments, accounting, and risk management. Generative AI promises to make complex processes, data analyses, and forecasts more efficient and precise. However, despite its impressive capabilities, the question remains: How reliable are the results produced by generative AI? Hallucinations and erroneous results can discourage users and hinder or even prevent the use of generative AI.

To answer this question, we conducted a comprehensive study comparing texts generated by generative AI with those created by human experts. Our investigation focuses on the insurance industry, where precise data analyses and informed decisions are of paramount importance. The study aims to evaluate the quality and accuracy of AI-generated texts and understand how they align with content created by experienced experts. Additionally, it provides a methodology for quality assurance in the implementation of generative AI applications. The results of our study were developed using the example of the insurance industry but are transferable to banks, asset managers, and other industries.

Methodology

Our study examines the quality of texts written by generative AI using a deviation measure that compares these texts with reference texts from public reports by experts in the insurance industry. Since the texts are based on the same data foundation, a direct comparison is possible, allowing for systematic investigation of quantitative deviations. Our methodology follows Kryscinski et al. (2020), which we use as a tool to evaluate the content and factual accuracy of texts generated by generative AI compared to those created by humans (referred to as the accuracy measure). We also tested other alternative measures, which did not meet our quality and consistency requirements. Unlike traditional studies evaluating such accuracies, our study focuses on the areas of actuarial science, accounting, and risk management. These areas are of particular interest because they have high potential for technology, while also requiring high accuracy with the use of uncommon technical terms.

In addition to the objective evaluation using our accuracy measure, we employ a control loop with experts as a quality assurance measure. Actuaries, risk managers, and auditors from KPMG have reviewed the texts generated by generative AI for content accuracy. This ensures that the results of our study are both statistically and professionally sound. Through this multi-step process, the study can make quality-assured statements about the results of generative AI and simultaneously quantitatively and objectively demonstrate which measures lead to significant improvements. The following measures are to be evaluated here:

- Prompt engineering, and
- Meta-prompting.

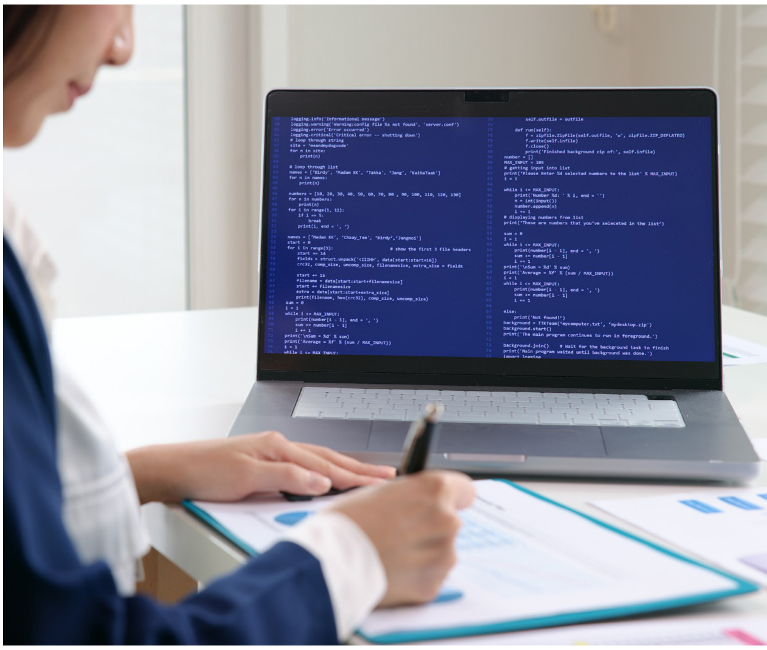
Prompt Engineering

Prompt engineering refers to the technique of providing specific inputs or instructions to generative AI to achieve desired results. Through targeted prompting, AI can be guided to generate more precise and contextually relevant texts. In this study, a prompting database from KPMG is used to optimize inputs and improve the quality of generated content.

Meta-Prompting

Meta-prompting extends prompt engineering with more complex and layered inputs, allowing AI to better process the structure and context of desired results. The meta-prompting database from KPMG provides AI with more comprehensive and strategic instructions, leading to higher accuracy and coherence of generated texts.

For this study, over 200 texts from the insurance industry were evaluated. All texts focused on the following areas: actuarial science, accounting, ESG or credit risk management, as well as general areas of risk management. The study deliberately includes technical terms, where the correct application by generative AI is fundamentally questioned. This can be explained by their rare use in everyday language and, accordingly, in the training data of the generative AI model.



[1] Wojciech Kryscinski, Bryan McCann, Caiming Xiong, and Richard Socher. 2020. Evaluating the Factual Consistency of Abstractive Text Summarization. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 9332–9346, Online. Association for Computational Linguistics.

Results

The evaluation of the accuracy measure in the following table shows that generative AI does not readily deliver high accuracy and factual fidelity. However, quality can be significantly improved. An appropriate accuracy of more than 95 percent is only achieved through the combined use of prompt engineering and meta-prompting. Thus, this study shows that the meaningful use of generative AI in complex situations must be professionally planned and implemented to create real value.

Measure	Accuracy Measure	Deviation from Previous Line (p-Value)	Verification of Correctness by Experts
Simple Prompting	57%	n/a	60%
Prompt Engineering	76%	19%-Points (0.1)	78%
Prompting Engineering incl. Meta-Prompts	98%	**22%-Points*** (8.54e-05)	99%

*Values rounded: for $p<0.05$, for $p<0.01$, and for $p<0.001$

Conclusion

These results demonstrate that the use of generative AI can achieve accuracies that meet the high demands of a company. However, to achieve this quality, preparatory work is necessary and must be tailored to the specific application case. This applies above all to industries and areas that frequently use technical terms in their work.

To create added value, companies should consider the following when implementing and using generative AI: A consistent focus on quality, accuracy, and reliability is crucial to achieving the desired results and building trust in AI solutions. Specialist knowledge and expertise in business processes should support the effective implementation of generative AI in order to make the best use of the technology.

When used in a targeted manner, processes can be optimized and automated, which can lead to efficiency gains and cost savings. It is important that all strategies and methods are regularly reviewed and adjusted in order to derive maximum benefit from generative AI.

One solution could be to build your own libraries that provide specific (meta) prompts. For example, this study used the following libraries:

- KPMG Prompt Library for prompt engineering
- KPMG Meta-Prompt Library for meta-prompting.

Outlook

For use cases that go beyond public data, a RAG approach (Retrieval-Augmented Generation) is recommended to improve quality by using internal information. RAG is a technique that supplies generative AI with additional information from supplementary data sources. This method combines the ability of AI to generate texts with the ability to retrieve relevant information from a database, thereby increasing the factual accuracy and relevance of generated content. RAG can be used, for example, to ensure that generated texts are not only creative but also factually correct.

To be successful, companies should pay attention to various aspects when implementing AI strategies: This includes developing a clear strategic direction for the use of AI and the development and prioritization of suitable use cases that can be implemented in business processes. It is important to integrate AI applications specifically into business processes to promote efficiency and innovation. Additionally, AI governance should be established, taking into account legal requirements such as the EU AI Act. Finally, quality assurance, including methods for objectively evaluating content generated by generative AI, is of great importance. Companies should ensure that these aspects are carefully planned and implemented to maximize the benefits of AI applications.

Contact

KPMG AG
Wirtschaftsprüfungsgesellschaft

Dr. Fabian Bohnert
Director, Financial Services
T +49 170 7016615
fbohnert@kpmg.com

Some or all of the services described herein may not be permissible for KPMG audit clients and their affiliates or related entities.

www.kpmg.de

www.kpmg.de/socialmedia

The information contained herein is of a general nature and is not intended to address the circumstances of any particular individual or entity. Although we endeavor to provide accurate and timely information, there can be no guarantee that such information is accurate as of the date it is received or that it will continue to be accurate in the future. No one should act on such information without appropriate professional advice after a thorough examination of the particular situation.