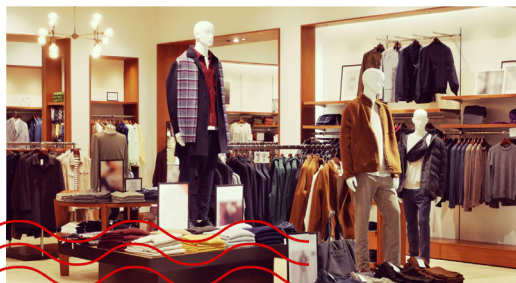




# Optimizing Search With Large Language Model Relevance Judgement



E. Breuninger GmbH & Co • Germany  
[www.breuninger.com](http://www.breuninger.com)



TRKKN supported Breuninger in developing a large language model based tool for quantifying search result quality. This tool automates the assessment of product list relevance, significantly improving efficiency and accuracy.

## The Challenge

Breuninger faced challenges in accurately and efficiently evaluating the quality of search results for their product listings. They needed a solution to automate the process, provide objective quality scores, and allow for frequent monitoring of thousands of keywords.

## The Solution

TRKKN partnered with Breuninger and delivered a custom tool to assess the relevance of search results based on Gemini and Cloud Run. This tool analyses lists of search results and their respective search term, providing a quantitative score for each search result. The solution leverages the speed and efficiency to evaluate multiple products simultaneously, significantly reducing assessment time.

## The Result

The large language model-based tool achieved total agreement rate with human expert judgments well above 80%, with partial agreement rates even higher for images and text attributes, respectively, demonstrating its accuracy and reliability, which allowed Breuninger to now efficiently monitor and assess the quality of thousands of keywords weekly, using a cost-effective solution.

The benefits for us are considerable, as we can regularly screen and evaluate the most important search terms. In addition, the solution enables us to evaluate changes in the search settings based not only on KPI's, but also including support from Large “Language Models”

— Johannes von Krane, Leitung Consumer Service / E.Breuninger GmbH & Co.

