

# **Benchmarking-on-demand/Benchmarking-as-a-service: New concepts for numerical benchmarking in the age of Chatgpt and other AI tools**

*Thursday 17 October 2024 14:00 (30 minutes)*

Nowadays, it is more or less standard that newly proposed numerical algorithms and software tools are validated and evaluated by known and community-accepted benchmark results. This typically requires presenting corresponding numerical results for (at least) three different “grid sizes” (in terms of mesh widths in space and time) so that comparisons can be made with the corresponding reference results found in the literature. However, in the age of Chatgpt and similar AI tools, it seems increasingly possible to automatically provide corresponding numerical results that mimic the expected (asymptotic) behavior of the underlying methods in a way that makes it difficult even for specialists to adequately assess the quality of the newly proposed methods.

As an alternative, we want to discuss the concept of “benchmarking-on-demand” (resp. “benchmarking-as-a-service”), i.e. fully automated benchmark results for specific applications that are not known before publication, so that a more rigorous (and reliable) evaluation of new approaches becomes possible. However, this concept requires a network of participating “trusted” partners that can be certified to act as appropriate “benchmark centers” for various specific benchmarking cases. We illustrate the underlying concepts in detail with some CFD benchmarks that are commonly used and might be candidates for such specific and new benchmarking scenarios, among other cases.

**Author:** TUREK, Stefan (TU Dortmund)

**Presenter:** TUREK, Stefan (TU Dortmund)

**Session Classification:** Contributed