

SOTA literature review

ar005 · 1 July 2026

Miller (2024) - Adding Error Bars to Evals: A Statistical Approach to Language Model Evaluations

<https://arxiv.org/abs/2411.00640>

Summary: to be done by Jonas Relevance to testpath: to be done by Jonas

Summary of findings and relevance to testpath

Here goes integrated learnings from the Literature Reviews