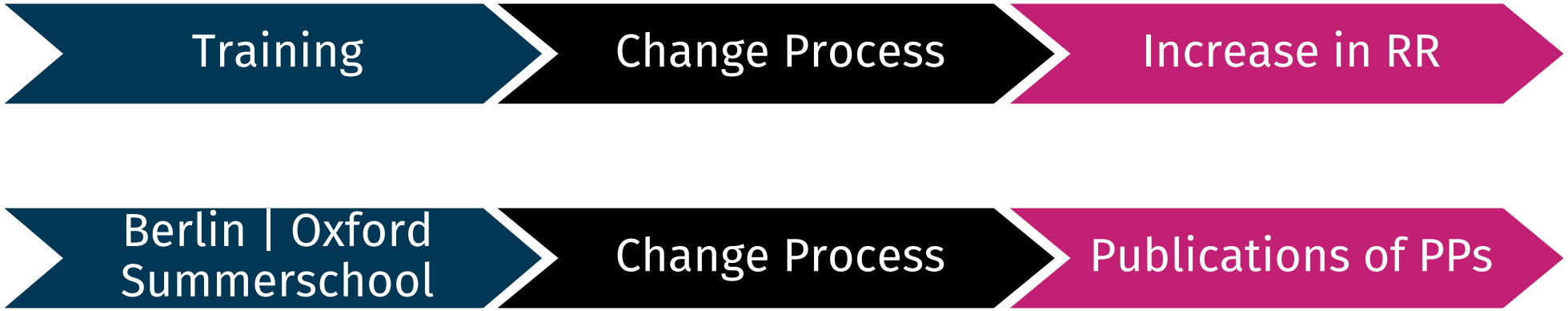
Leveraging Large Language Models for Assessing Responsible Research Practices

Ulf Toelch

BIH QUEST
Center for Responsible Research

BIH Berlin Institute of Health @Charité

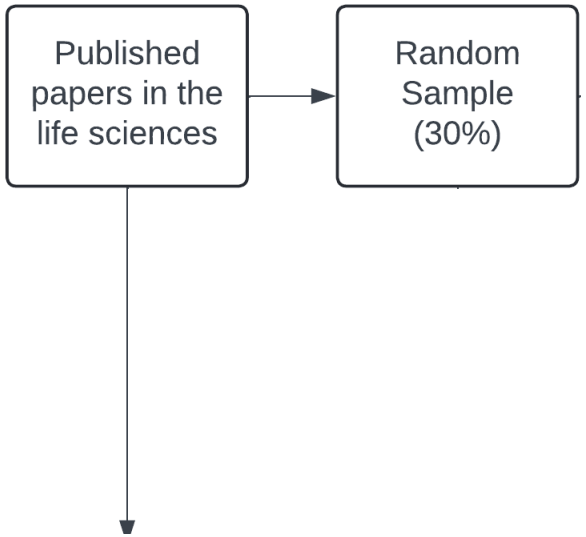
The larger project: Training Transfer



How to efficiently measure RRPs in a diverse set of publications?

BIH QUEST Center for Responsible Research

Possible Solution



BIH QUEST
Center for Responsible Research

Some background for LLMs

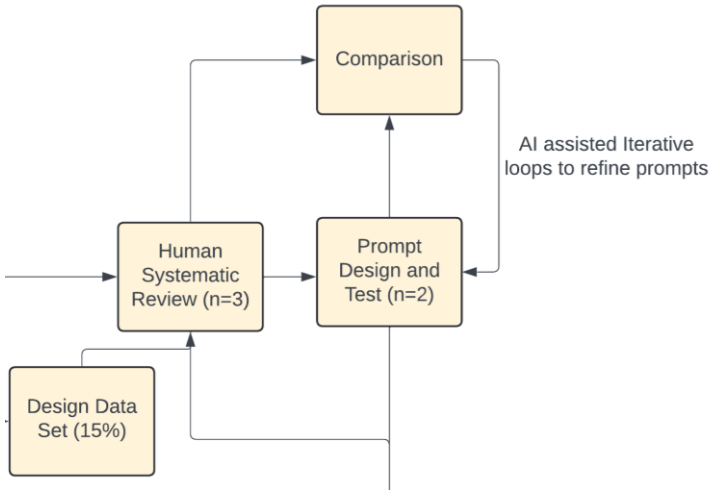
Context Window:

The amount of text an LLM can consider at once when generating a response. A larger context window allows the model to maintain coherence over longer passages, but can also increase computational demands.

Prompt Engineering:

The art of crafting effective instructions (prompts) to guide an LLM's output. Careful prompt design can significantly improve the quality and relevance of the generated text.

LLMs with large context windows



eLife
RESEARCH ARTICLE

Right inferior frontal gyrus implements motor inhibitory control via beta-band oscillations in humans

Michael Schaum[1]*, Edoardo Pinzuti[1], Alexandra Sebastian[2], Klaus Lieb[2], Pascal Fries[3,4], Arian Mobascher[2], Patrick Jung[2], Michael Wibral[5†], Oliver Tüscher[2†]

[1]Systemic Mechanisms of Resilience, Leibniz Institute for Resilience Research, Mainz, Germany; [2]Department of Psychiatry and Psychotherapy, University Medical Center of the Johannes Gutenberg University, Mainz, Germany; [3]Ernst Strüngmann Institute (ESI) for Neuroscience in Cooperation with Max Planck Society, Frankfurt, Germany; [4]Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen, Nijmegen, Netherlands; [5]Campus Institute Dynamics of Biological Networks, Georg-August University, Göttingen, Germany

Abstract Motor inhibitory control implemented as response inhibition is an essential cognitive function required to dynamically adapt to rapidly changing environments. Despite over a decade of research on the neural mechanisms of response inhibition, it remains unclear, how exactly response inhibition is initiated and implemented. Using a multimodal MEG/fMRI approach in 59 subjects, our results reliably reveal that response inhibition is initiated by the right inferior frontal gyrus (rIFG) as a form of attention-independent top-down control that involves the modulation of beta-band activity. Furthermore, stopping performance was predicted by beta-band power, and beta-band connectivity was directed from rIFG to pre-supplementary motor area (pre-SMA), indicating rIFG's dominance over pre-SMA. Thus, these results strongly support the hypothesis that rIFG initiates stopping, implemented by beta-band oscillations with potential to open up new ways of spatially localized oscillation-based interventions.

***For correspondence:** Michael.Schaum@lir-mainz.de
[†]These authors contributed equally to this work

OpenAI

BIH QUEST Center for Responsible Research

why is "We aim to further explore the hypothesis that autoantibodies against brain structures may also play a role in dementia" in the scietific paper not rated as a hypothesis?

Show drafts

The provided text 'We aim to further explore the hypothesis that autoantibodies against brain structures may also play a role in dementia' is not classified as a hypothesis because it expresses the intent to investigate an existing hypothesis, rather than stating a new one for the current study.

A hypothesis is a specific statement about the expected outcome of a study. It is a prediction that researchers aim to test through experimentation or observation. In the given text, the researchers acknowledge a prevailing hypothesis and their intention to explore it further, suggesting it is not the central hypothesis being tested in their specific study.

Show the code behind this result

LLM assisted iterative loop: Dialogue with LLMs

BIH QUEST
Center for Responsible Research

LLM assisted iterative loop resulted in return of meaningful features from papers

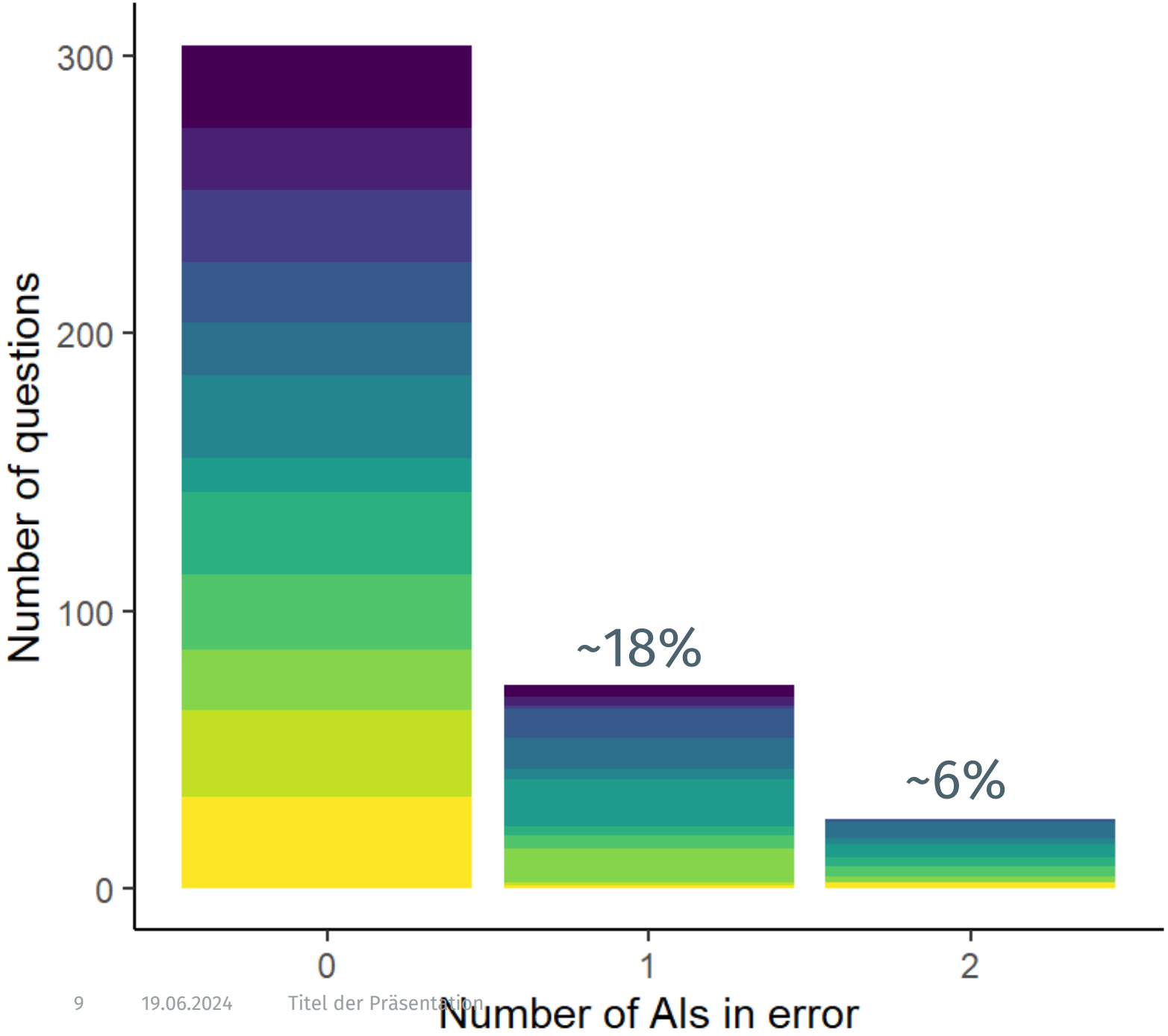
5	No	The study does not explicitly state a hypothesis, but rather, it aims to "assess...the access of the provisions aligned with the Policy, and their synergies, and quantify their associations on select goals of the Policy."
7	Yes	A. "We used a multivariate multi-level analysis which simultaneously estimated the effects on all three Policy-aligned outcomes..." B. "AOR: 3.07. 95% CI 1.72-5.46, p < .001" C. The p-values and confidence intervals are reported throughout the Results and in Table 3.

Comparison of Gemini and ChatGPT

"The objective of this retrospective study was to compare the predictive performance..."

The paper does not explicitly state a hypothesis. It states that "This study aimed to compare the predictive performance of the sFlt-1/PlGF ratio alone versus a multi-marker regression model."

Validation results



N=37 studies
12 questions each
2 AIs | 3 human reviewers

BIH QUEST
Center for Responsible Research

Limitations and next steps

Reproducibility and variability in answers ->Bug or feature?

Potentially long process to find prompts -> standardized prompts?

Responsible research Criteria not easy to assess even for human reviewers ->Granularity of prompts

Currently mostly manual communication with LLMs -> pipeline via APIs

Online service -> local infrastructure need with local LLM

Error rate of LLMs -> reconciliation between LLMs

Thank you.

BIH QUEST
Center for Responsible Research

BIH Berlin Institute of Health @Charité