

INTRODUCTION & BACKGROUND

- The supervised fine-tuning process of LLMs could spread misconceptions about marginalized groups [2].
- ADHD traits make participation in annotation processes difficult as these processes are not inclusive enough. Thus, ADHD experiences are underrepresented.
- Hermeneutical injustice - marginalized groups cannot express their experiences because of lack of collective understanding [1].
- Diversity is disregarded in multiple phases of annotation pipelines - from labeling prompts to dataset curation [3; 4].

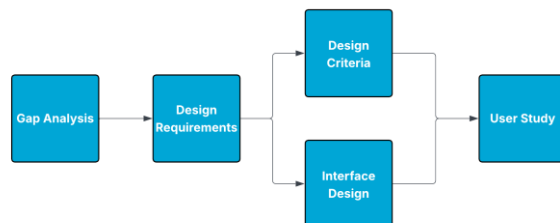
Main RQ: How can hermeneutical justice be reinforced regarding people with ADHD during the annotation process for supervised fine-tuning?

RQ1: What are the disparities between current annotation practices and people with ADHD?

RQ2: How can ADHD traits be accounted for in annotation tasks?

METHODOLOGY

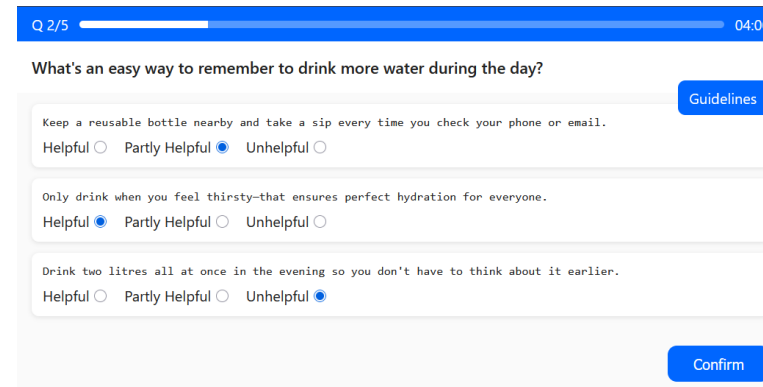
- Design Requirements Derivation – found where current practices did not take into consideration people with ADHD and distilled requirements to accommodate them in the process.
- Design Annotation Interface and Task – built a more inclusive annotation interface and task structure according to the requirements.
- User Study – logged behavior data and collected open-ended survey questions, used for quantitative and qualitative analysis.



DESIGN REQUIREMENTS

ADHD Trait	Annotation Practice	Design Requirement
Sustained attention deficit [6]	Some labeler roles are more demanding in terms of cognitive load [11]	Limit sustained load
Impulsivity [6]	Labelers annotate instructions as helpful, truthful [12]	Reduce impulsivity
Working memory deficit [5]	Labelers need to know and remember annotation guidelines and labeling criteria [11; 12]	Guidelines always available
Time blindness [9]	Annotation guidelines don't specify how much time a task should take [11; 12]	Sense of elapsed and remaining time
Distractibility [7]	Some annotations are split on multiple screens [10]	Eliminate distractions

INTERFACE DESIGN



Q 2/5 04:00

What's an easy way to remember to drink more water during the day?

Keep a reusable bottle nearby and take a sip every time you check your phone or email.
Helpful ☐ Partly Helpful ☒ Unhelpful ☐

Only drink when you feel thirsty—that ensures perfect hydration for everyone.
Helpful ☒ Partly Helpful ☐ Unhelpful ☐

Drink two litres all at once in the evening so you don't have to think about it earlier.
Helpful ☐ Partly Helpful ☐ Unhelpful ☒

Guidelines +

Confirm

REFERENCES

- [1] Fricker, Epistemic Injustice: Power and the Ethics of Knowing, 2007.
- [2] Kay et al., Epistemic Injustice in Generative AI, 2024.
- [3] Bender et al., On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?, 2021.
- [4] Kapania et al., A Hunt for the Snark: Annotator Diversity in Data Practices, 2023.
- [5] Alderson et al., Attention-Deficit/Hyperactivity Disorder (ADHD) and Working Memory in Adults: A Meta-Analytic Review, 2013.
- [6] Roberts et al., Constraints on Information Processing Capacity in Adults with ADHD, 2012.
- [7] Schneidt et al., Distraction by Salient Stimuli in Adults with ADHD: Evidence for the Role of Task Difficulty in Bottom-Up and Top-Down Processing, 2018.
- [8] Sonne et al., An Assistive Technology Design Framework for ADHD, 2016.
- [9] Weissenberger et al., Time Perception Is a Focal Symptom of ADHD in Adults, 2021.
- [10] Kern et al., Annotation Sensitivity: Training Data Collection Methods Affect Model Performance, 2023.
- [11] Köpf et al., OpenAssistant Conversations — Democratizing Large Language Model Alignment, 2023.
- [12] Ouyang et al., Training Language Models to Follow Instructions with Human Feedback, 2022.

RESULTS

Criteria Evaluation

Criterion	Goal
≤ 50 s per micro-task	✓ 6/7
Finish ≤ 5 min	✓ 6/7
Distraction ratio < 0.15	✓ 7/7
Guideline accessed ≥ 1 times	○ 3/7
Mislabel rate $< 30\%$	× 1/7

Qualitative Themes

“the timer in the corner does add a slight pressure, but it is a useful feature”

“changed my mind during the cooldown in several questions”

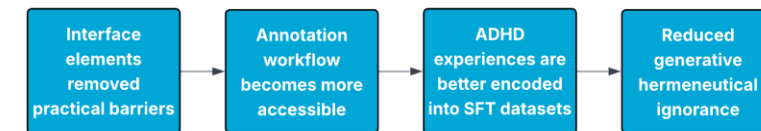
“had to figure out the differences between helpful and partially helpful answers”

“focus was always split between understanding and answering questions and the concern of time running out”

“have the questions placed in a more centered position”

“ADHD doesn't get accounted for in places, and llm is a relatively new field for humanity so I want to contribute to it with people like me in mind”

CONCLUSIONS & FUTURE WORK



Future Work Suggestions:

- The interface could be further improved based on the findings of the study.
- The study should be replicated with a larger sample size and against a control group.
- The framework used to improve ADHD inclusion should be extended towards other marginalized groups.