

CHATGPT VS. HUMAN? INTERACTION WITH ASSISTANTS TO GENERATE DIAGNOSES

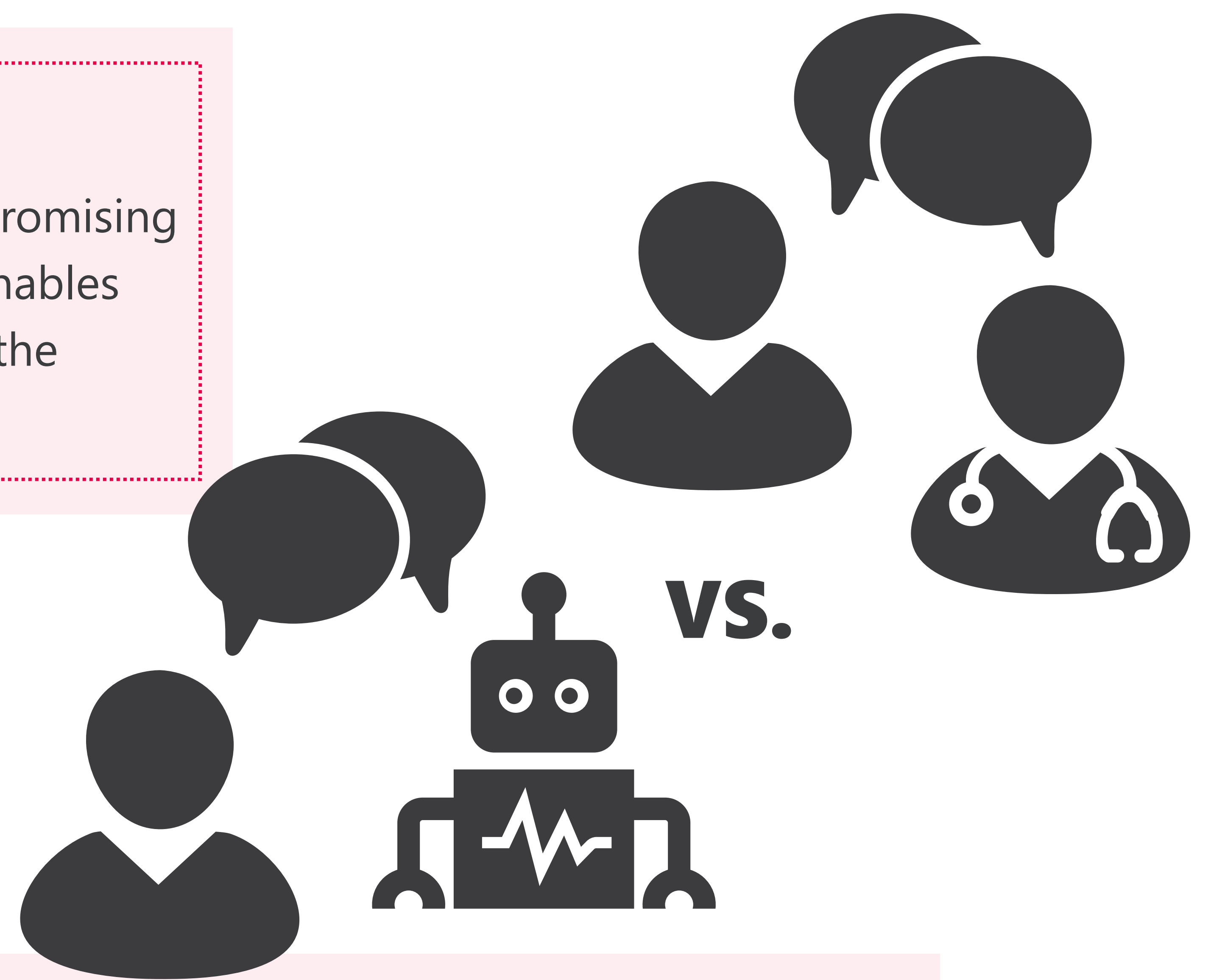
DEPARTMENT OF EMERGENCY MEDICINE: J. E. KÄMMER, W. E. HAUTZ, G. KRUMMREY, T. C. SAUTER, T. BIRRENBACH
INSTITUTE OF INFORMATION SYSTEMS: M. N. SPIZZO AND J. DIBBERN
CHARITÉ UNIVERSITÄTSMEDIZIN BERLIN: D. PENDERS
D-MTEC ETH ZÜRICH: N. BIENEFELD

INTRODUCTION

- **5-15%** of patient diagnoses are incorrect.
- To reduce this number, **ChatGPT** is considered a promising solution that, despite the skilled labor shortage, enables clinicians the opportunity for consultation during the **diagnostic process**.

QUESTION

How and why do the interactions with ChatGPT differ from the interactions with a human expert in the diagnostic process?



MATERIALS AND METHODS

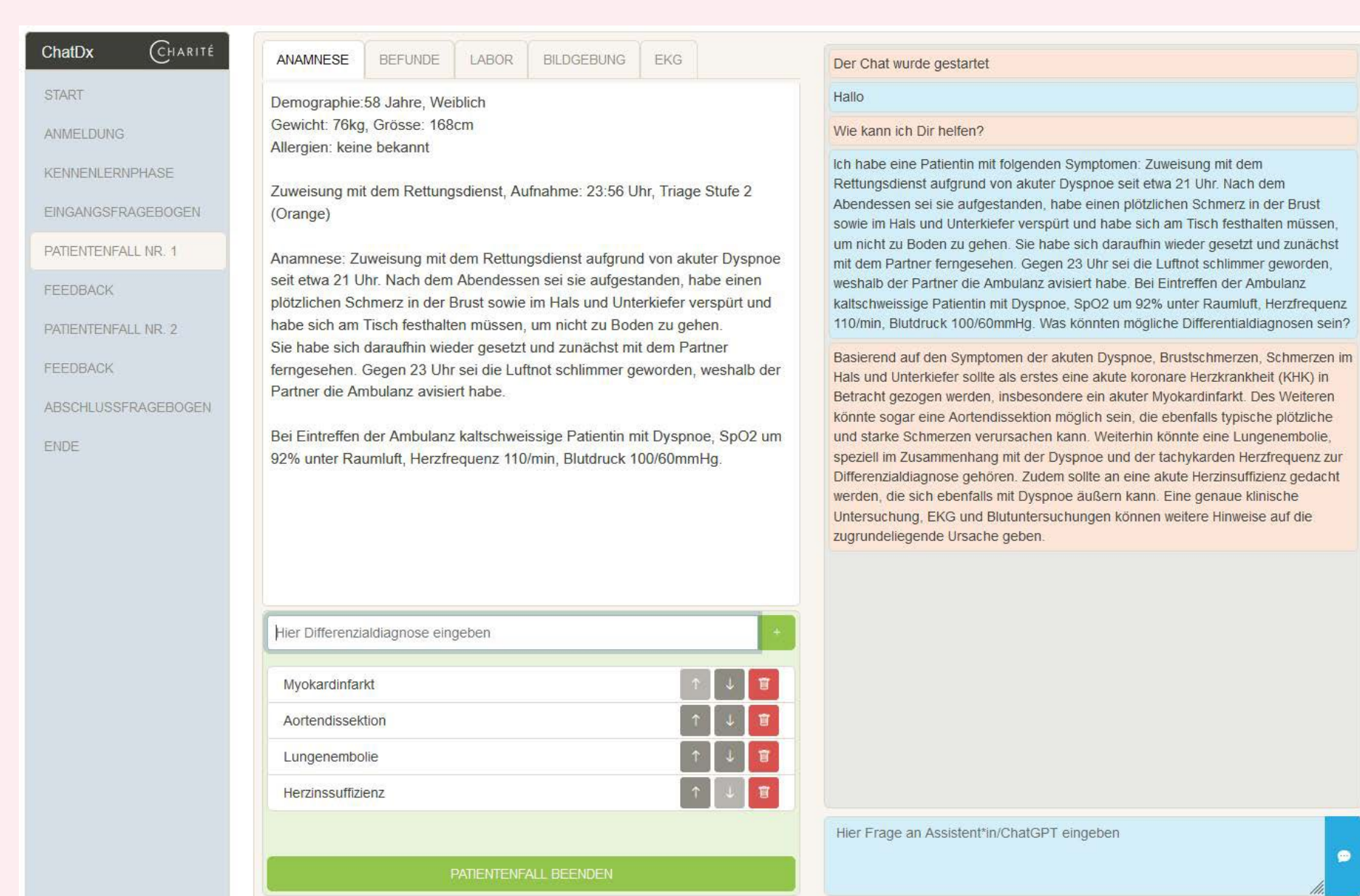


Figure 1. A screenshot of the interface in the diagnostic task.

Design: A 2x2 between-subjects design (human coach vs. ChatGPT, no training vs. training)

Participants: 158 4th year medical students

Setting: An experiment in which two diagnostic tasks were solved in an online environment where participants could chat with the assigned assistant (human or ChatGPT) in real time.

PRELIMINARY RESULTS

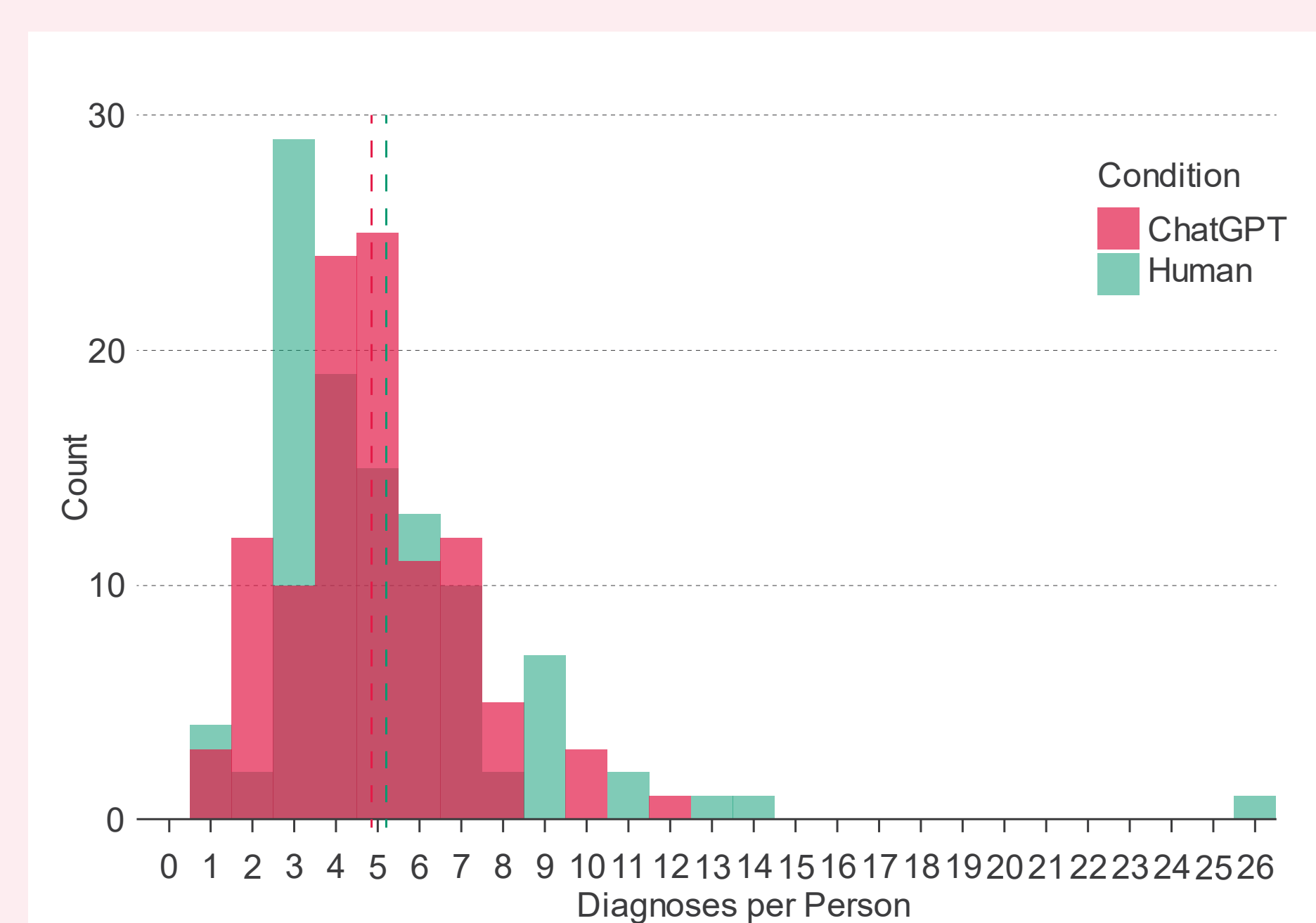


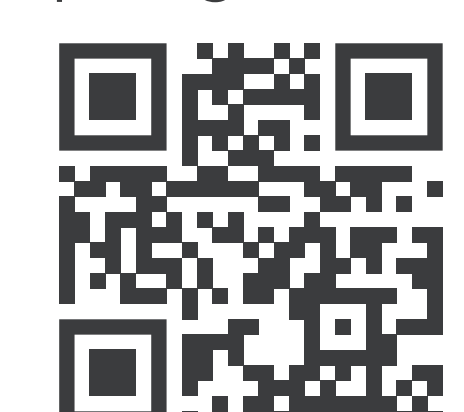
Figure 2. Frequency of differential diagnoses per person per patient case by condition.

Preparation: Participants prepared themselves less before starting the interaction in the ChatGPT condition.

Type of questions: In the ChatGPT (vs. human) condition, more questions were asked to build (vs. test) diagnoses.

Number of diagnoses: Five ($SD = 3$) differential diagnoses were generated on average with either type of assistant.

Scan to read the preregistration



Contact:
mayra.spizzo@unibe.ch