1 Appendix A: Questionnaire development

2 The questionnaire was written in Chinese. The initial step involved systematic literature retrieval 3 to gather information on implementing ethics in social experiments based on medical artificial 4 intelligence from guidelines, expert consensus, practice standards, and norms. A librarian working 5 in the hospital library provided valuable assistance during this process (see Appendix B). Then, a 6 focus group interview, consisting of 10 experts (two medical ethics professors, one sociology 7 professor, three artificial intelligence professors, and four medical professors proficient in medical 8 artificial intelligence social experiments), was conducted. In the interview meeting, all experts were 9 encouraged to express their opinions on the following questions: (1) What is your understanding of 10 implementing ethics in medical artificial intelligence (MAI) social experiments? (2) What 11 knowledge should medical staff, involved in MAI social experiments, master to facilitate 12 implementing ethics? (3) What are your attitudes regarding implementing ethics in MAI social 13 experiments? Eventually, relevant contents from literature and interviews were extracted, and then 14 they were classified according to the Knowledge-Attitude-Practice model. Based on the model, 15 knowledge is comprised of scientific knowledge, local knowledge, tacit knowledge, and self-16 reflective knowledge [1]. Attitude refers to a positive or negative option of objective evaluation [2]. 17 After item generation, item deletion and modification were made according to experts' opinions 18 through three rounds of Delphi expert consultation. Ten experts, including medical ethics professors, 19 sociology professors, artificial intelligence professors, and medical professors, were invited. 20 Eventually, the draft questionnaire was developed. 21 Before the formal survey started, eight individuals (according to the cognitive debriefing

22 guidelines provided by the PROMIS Translation Director from our previous study), including

23	artificial intelligence researchers and healthcare workers, joined the cognitive debriefing. The
24	purpose of the cognitive debriefing is to confirm that all items are understood by the target
25	participants as intended. First, they were invited to complete the draft questionnaires, and after that
26	they were interviewed with the following questions: (1) Was each item clearly expressed without
27	ambiguity? If not, please identify the unclear or ambiguous expressions; (2) Were any items
28	challenging to understand? If yes, please specify the difficulties, and if not, please try to explain
29	each item in your own words; (3) What were your reasons for each of your answers? (4) What else
30	is needed to be added? All participants were able to correctly explain the meaning of the item and
31	respond logically in their own words on 17 items. Language readability modification of the other 4
32	items was made according to the participants' comments. The final questionnaire consisted of two
33	dimensions, comprising 21 items. In the context of the knowledge dimension, respondents were
34	asked to express their familiarity with various aspects, including the progress, ethical issues related
35	to conducting MAI social experiments, and ethics governance according to norms and principles for
36	such experiments. The response options ranged from 'familiar', 'uncertain' to 'unfamiliar'. In the
37	attitude dimension, respondents were asked to express their agreement with behavioral statements
38	concerning implementing ethics in MAI social experiments. The response options ranged from
39	'disagree', 'neutral' to 'agree'. The questionnaire was pilot-tested through convenience sampling from
40	June 6, 2023, to June 17, 2023. A survey was conducted with a sample of 52 individuals to assess
41	the face validity, focusing on determining the relevance and accuracy of the 21 items, as well as
42	examining their reliability according to the Consensus-based Standards for the Selection of Health
43	Measurement Instruments (COSMIN) methodology for assessing the content validity and reliability
44	of Patient-reported Outcome Measures (PROMs)[3,4]. The items were found to be reliable, with

- 45 an acceptable Cronbach's alpha coefficient of 0.727[5]. Item-content validity index(I-CVI) and
- 46 scale-content validity index(S-CVI) were 0.791 and 0.877, respectively [6].
- 47 **References:**
- 48 1. Oliva A, Grassi S, Vetrugno G, Rossi R, Della Morte G, Pinchi V, Caputo M. Management of
- 49 Medico-Legal Risks in Digital Health Era: A Scoping Review. Front Med (Lausanne). 2022 Jan
- 50 11;8:821756. Doi: 10.3389/fmed.2021.821756. PMID: 35087854.
- 51 2. Adam AM. Sample size determination in survey research. JSRR, 2020,26(5):90-97. DOI:
- 52 10.9734/JSRR/2020/v26i530263
- 53 3. Terwee CB, Prinsen CAC, Chiarotto A, Westerman MJ, Patrick DL, Alonso J, Bouter LM, de Vet
- 54 HCW, Mokkink LB. COSMIN methodology for evaluating the content validity of patient-reported
- outcome measures: a Delphi study. Qual Life Res. 2018 May;27(5):1159-1170.
- 56 4. Prinsen CAC, Mokkink LB, Bouter LM, Alonso J, Patrick DL, de Vet HCW, Terwee CB.
- 57 COSMIN guideline for systematic reviews of patient-reported outcome measures. Qual Life Res.
- 58 2018 May;27(5):1147-1157.
- 59 5. Suresh K, Chandrashekara S. Sample size estimation and power analysis for clinical research
- 60 studies. J Hum Reprod Sci. 2012 Jan;5(1):7-13. doi: 10.4103/0974-1208.97779. Retraction in: J
- 61 Hum Reprod Sci. 2015 Jul-Sep;8(3):186. PMID: 22870008; PMCID: PMC3409926.
- 62 6. Wu Minglong. Questionnaire statistical analysis practice -SPSS operation and application [M].
- 63 Chongqing: Chongqing University Press, 2018: 158-193.