Appendix 1

This research was approved by the Ethics Committee of the University of Tsukuba Hospital (permission number: H30-145). According to the Ethical Guidelines for Medical and Health Research Involving Human Subjects [77], our research was carried out with the opt-out consent of patients. Opt-in informed consent was obtained from medical doctors who engaged in the evaluations (please see Section 2.8).

The proofreading procedures used to ensure privacy protection in the experimental use of progress notes and other medical data from electronic health records were approved by the same committee (permission number as noted above).

Before the experiments commenced, patients who had rare diseases, participated in any clinical test, had criminal histories, or who were members of a sexual minority were excluded from the study, due to privacy concerns. If the patient's progress notes included descriptions about his or her faith in any religion or rare travel history, the patient was also excluded from this study, due to privacy concerns. Proper nouns, dates, numerical values (e.g., drug doses, results of laboratory tests, and values of the vital signs), patient histories, and family histories included in progress notes and other medical data were rewritten using approved procedures in a way that did not compromise the experiments. To be specific, for numerical values and dates, random numbers were added to or subtracted from the original numerical values and dates. The proper nouns were randomly replaced character by character. For patient histories and family histories, gender, and whether the diseased area was the right or left were randomly exchanged. After all data were preprocessed, three people (a medical doctor, a nurse, and a staff member working in the hospital) checked that the real patient could not be identified

based on the preprocessed notes. We then used the preprocessed actual progress notes (which we call "the real progress notes" in this paper) for the experiments.

In Step 1-b, since the real progress notes shown to crowd workers had been preprocessed, they sometimes contained unnatural proper nouns or numerical values for tests or medicines, etc. (e.g., see, in Fig 2, the drug "AQWE," or the dose "9,560 mg"). As a result, the imitated texts based on the real progress notes could also contain unnatural numerical values, proper nouns for tests or medicines, etc. We rewrote the imitated texts for their recomposition of the imitation texts in Step 1-c. First, for numerical values or dates, random numbers were added to or subtracted from the original numerical values and dates in the proofreading step. The same numbers were subtracted from or added to the imitated text numbers after Step 1-b. For example, in Fig 4, the proofreading step, 120 was added to the real systolic blood pressure value (130), and one was subtracted from the real diastolic blood pressure value (80); therefore, for the imitated text, 120 was subtracted from the systolic blood pressure value (240), and one was added to the diastolic blood pressure value (79) after Step 1-b. Second, the proper nouns for tests or medicines were reversed to the original nouns after Step 1-b. For example, in Fig 4, "BP" was preprocessed to "EP" in the proofreading step, and the crowd worker changed "EP" to "EF" in Step 1-b; "EF" was reverted to "BP" independent of the changing pattern of the crowd workers. Third, the proper nouns for places and people, patient histories, and family histories were also not reverted. The proofread nouns remain in the texts used for Step 1-c, meaning that the published progress notes contain nonexistent hospital names, medical doctor names, and so on, to ensure privacy

protection of the texts used for Step 1-c. After these procedures, we recomposed the imitated texts (Step 1-c).

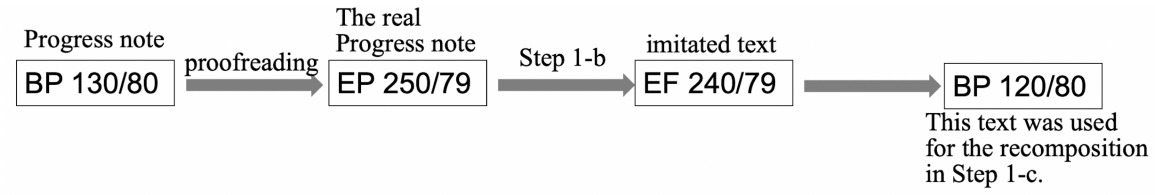

Fig. Example of the proofreading and inverse proofreading processes.