Author’s response to reviews

Title: Customization Scenarios for De-identification of Clinical Notes

Authors:

Tzvika Hartman (tzvika@google.com)
Michael Howell (mdhowell@google.com)
Jeff Dean (jeff@google.com)
Shlomo Hoory (hoorys@google.com)
Ronit Slyper (rslyper@google.com)
Itay Laish (itaylaish@google.com)
Oren Gilon (ogilon@google.com)
Danny Vainstein (dannyvain@google.com)
Greg Corrado (gcorrado@google.com)
Katherine Chou (kic@google.com)
Ming Jack Po (mingp@google.com)
Jutta Williams (juttawms@google.com)
Scott Ellis (scellis@google.com)
Gavin Bee (gavinbee@google.com)
Avinatan Hassidim (avinatan@google.com)
Rony Amira (ronya@google.com)
Genady Beryozkin (genady@google.com)
Idan Szpektor (szpektor@google.com)
Yossi Matias (yossi@google.com)

Version: 3 Date: 12 Jan 2020

Author’s response to reviews:
General Comments

Q1: Please clarify whether any formal approval was required to access the MIMIC-III, i2b2-200626 and i2b2-20149,27 and datasets of de-identified clinical records and include a statement in the 'Ethics approval and consent to participate' section.

R1: Added: "Formal approval was obtained from each respective dataset owner."

Q2: In the Funding section of the Declarations please indicate the role of the funding body in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript. If no specific funding was received for this study, please clearly indicate this in the Funding section.

R2: Added a sentence clarifying that all of the above tasks were carried out by Google employees.

Q3: If you wish to acknowledge someone by name in the Acknowledgements, please ensure you have obtained permission from them to so do.

R3: Done

Q4: At this stage, please upload your proofread manuscript as a single, final, clean version that does not contain any tracked changes, comments, highlights, strikethrough or text in different colours. All relevant tables and figures should also be clean versions. Figures (and additional files) should remain uploaded as separate files. Should you wish to respond to these revision requests, please include the information in the designated input box only.

R4: Done

Specific comments from the editor

Q1. Response to Reviewer 4, Q2 should be moving the section beginning "Our first scenario is a fully customized system..." to paragraph beginning "Our last scenario is partial...". The descriptions of the scenarios used in this study should be in the Methods.

R1: We sent an email to the editor, with the following suggestion:

We already addressed it by moving the technical parts of the scenario description to the methods as requested.

However, we thought it would be more readable to leave a high level overview in the introduction. Examining several examples of papers published at BMC Medical Informatics and Decision Making, we noticed that the section names are not necessarily abstract, background, method, results, ..., and that some variation is acceptable.

Is it ok if we leave the high level overview where it is and fix the section names to read: Abstract, Introduction (with the high level scenario description as it is now), Related work, Methods (subsections: Data Sources, Text De-Identification System Architecture, and Experimental Setup for Deployment Scenarios), Results, ...

Q2: 2. "Although we are generalizing results from the available datasets and thus cannot provide performance guarantees" is a limitation of this study and other deidentification methods should be
discussed here.

R2: Removed that sentence, and added a paragraph (the one before the last) to the Conclusions.