AI-POWERED DIAGNOSTICS

Development Of An Automated System For The Analysis Of Large-scale Patient Care Records

Markus Antwerpen Bundeswehr Institute of Microbiology **Michael Seidel** Bundeswehr Institute of Microbiology **Robin Lea Woelfel** Bundeswehr Institute of Microbiology **Jakob Fuchs** Bundeswehr Institute of Microbiology **Roman Woelfel** Bundeswehr Institute of Microbiology

The Bundeswehr Institute of Microbiology is a specialised military medical facility for defence against highly dangerous infectious diseases. It analyses and evaluates medical specimens from military and civilian patients from Germany and around the world. Fast and reliable processing of the sample material and all associated data is essential. Medical samples are regularly collected along with a range of metadata to further describe the sample and provide context for the required analyses. However, it is not uncommon for data, such as a patient's travel history, to be collected and transmitted in non-standardised free-form text. Such unstructured data cannot be used directly for automated processing. It requires manual intervention and control. In order to systematically analyse the analytical data, e.g. to group it according to certain characteristics, entities and relationships in such free-form text, it must be extracted and stored appropriately.

In light of recent advances in natural language processing and large language models, we have evaluated the applicability of different algorithmic approaches. To this end, we have developed two functional proof-of-concept systems for querying large, publicly available free-form medical texts. Both approaches are designed to achieve reliability, reproducibility, completeness and correctness of the identified data sets. Furthermore, given the highly sensitive nature of the data processed in the intended use of such a tool, both solutions place a strong emphasis on data security, for example through local processing on site.

Initial results show that the pool of medical data, which is only available in unstructured free text, can be opened up and made accessible to both scientists and treating physicians for further analyses. Additional (meta) data, which previously required careful manual curation, is now available to researchers on a large scale, enabling broader and/or more detailed findings. Clinicians have a powerful interface to search and identify previous case reports based on even an approximate description. In both cases, the quality of diagnoses and study results, as well as the time it takes for this data to become available, can be significantly reduced.