Embracing a new friendship: Artificial intelligence and medical writers

Sonia Costa
Papyrus Scientific Communications

Correspondence to:
Sonia Costa
Papyrus Scientific Communications
Wissmannstr. 20
12049 Berlin, Germany
+49 (0)1522 279613
scosta@papyrus-scientific.com

Abstract
Artificial Intelligence (AI) and machine learning-driven software are evolving toward a technological advancement revolutionising the current global, social, and economic landscape. A major breakthrough in AI is the fast-growing natural language processing–based tools, where millions of euros are being invested in the development of software packages that automatically, and accurately, generate text. The impact of this technology in medical writing is immense. Will we be out of a job in the near future? In an optimistic (and perhaps realistic) point of view, one can argue in favour of a friendly interaction between medical writers and machines, with major advantages for the craft of medical writing.

Times are exciting! The prospective applications of artificial intelligence (AI) and machine learning (ML) technology in the medical and healthcare industries are revolutionary. For many years, AI/ML-based tools have been successfully used in radiology to improve the diagnosis and earlier detection of diseases.[1] The AI boom, in the recent years, has generated a plethora of platforms and applications that translate to new regulatory frameworks (both in the USA and EU). Some AI/ML-based technologies are categorised as medical devices (referred to as "software as medical devices" [SaMD] and thus are subject to strict evaluations to ensure patient safety.[2] Notably, AI has the potential to make inroads in a number of new avenues, from disease detection and improved diagnosis tools to the development of new (and personalised) therapeutics, and can therefore contribute to better clinical decision making.

AI redefines Pharma R&D
A growing number of the largest global pharma companies are investing millions of euros in acquiring or partnering with start-ups to develop and improve AI technology, with the goal of speeding up drug discovery. DeepMind Health (launched in 2016 by Google) and IBM Watson (particularly in the areas of health, drug discovery, and oncology) are pioneers in health-related AI applications and have already been transforming the long and costly process of bringing new therapeutics to the market.

AI-based tools are used to predict good candidates among thousands of small molecule compounds, through biomarker identification

and target discovery, increasing the numbers of potential drugs identified and shortening the period of pre-clinical study. Moreover, this technology is also being used to promote the discovery of new therapeutics for rare diseases (for more information, please refer to Bulgaru [2018][3] and Jiang [2019][4]). Interestingly, AI is also redefining the clinical trial landscape, with repercussions for scientists, physicians, clinical personnel, patients, and medical writers.

Natural language processing: A major breakthrough in AI

One of the characteristics of AI/ML-based tools is the fact that this technology relies on structured data to learn and improve its performance. A major breakthrough in the AI field is the fast improvement of natural language processing (NLP)-based tools, driven by the investment of titans, such as Microsoft, Google, and Amazon, which already have several trendy smart speakers integrated in the daily lives of many people. NLP is a rapidly expanding area of research and development. A major advantage of this technology is its capability to extract information from unstructured data, such as narrative text documents, which are otherwise incomprehensible for computer programs.[5] Consequently, a massive amount of information locked in scientific and healthcare text databases can be transformed into structured data, and this progress will undoubtedly redefine the medical and healthcare industries. AI-based writing assistance will be a reality, in the near future, within the reach of many of us, and will, inevitably, transform the medical writing field.

As NLP tools can read unstructured data, they can easily scan metadata, perform tailored literature searches, and accurately extract targeted information from vast unstructured databases. This is particularly important in an age where a growing number of scientific articles are available at PubMed and hundreds of clinical studies are accessible at ClinicalTrials.gov. This vast collection of data is of an immense value, however it is basically impossible for a human being to be up to date with the latest discoveries, and relate the information on a given study to other hundreds of studies already published. Thus, the application of NLP technology in clinical study design and pharmacovigilance is revolutionary, because skilled medical writers will be able to gather and analyse a surplus of information in a fast yet accurate mode. Many pharma companies are already investing in NLP-tools to perform complex data analysis and interpretation, study design, and generation of an array of documentation for each step in clinical research. Text mining (using NLP) performs linguistic analysis to the lexical level and is able to extract detailed information, reveal patterns across millions of documents, and automatically summarise loads of information.[5]

AI-based writing assistance will be a reality, in the near future, within the reach of many of us, and will, inevitably, transform the medical writing field.

NLP at the service of medical writers

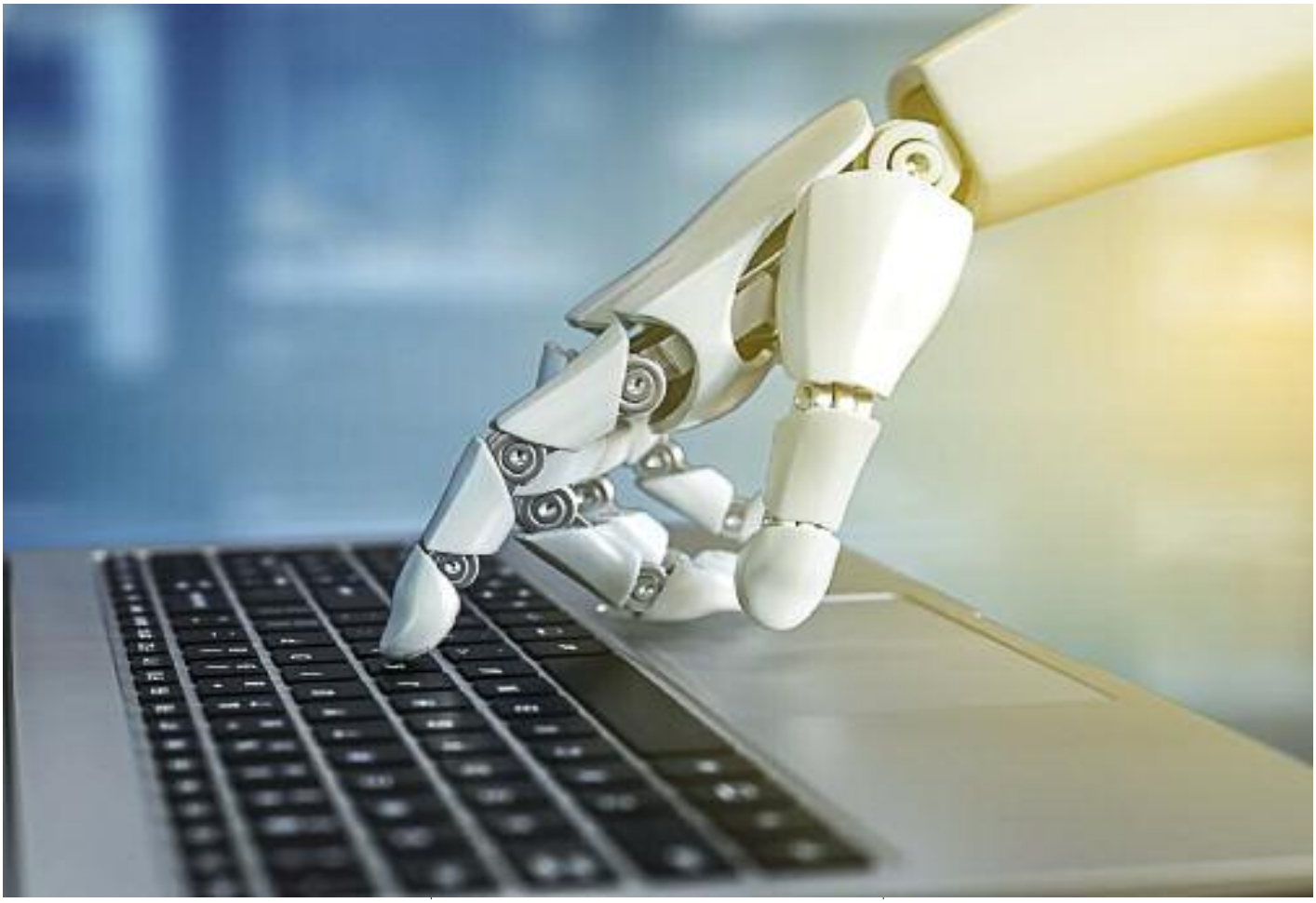
Aside from forming a basis for improving AI technology, how can this technology be useful to medical writers? A tremendous practical aspect of NLP tools is their ability to assist the medical writer in detecting errors during the anonymisation of patients enrolled in clinical studies, particularly in international multicentre clinical trials involving thousands of patients. Policies implemented by the European General Data Protection Regulation impose strict guidelines to ensure the de-identification of patients enrolled in clinical trials by pharmaceutical companies. To adhere to these policies, identifiers (information that may allow the identification of a patient) have to be redacted, which is a tedious, time-consuming work especially as they involve large databases. While several methods are used for anonymisation, it was recently shown that de-identification of large datasets is easily reversed by AI-based tools, i.e., these tools are able to accurately trace individuals in anonymised databanks.[6] A way to overcome a possible breach is to, ironically, use NLP-based technology to create robust anonymised documents.

Moreover, the automation of standard documents decreases human errors and the need for clinical staff and medical writers to certify that information has been accurately entered in large patient databases.[7] This favours a reduction of the early stages of clinical trials and also results in a better selection of patients taking part in a given study and promotes a high quality control of the entire process until the drug's market approval phase.

Interestingly, many content creators use AI writing assistant software that extracts information from the web and automatically produces a summarised text that is then edited by the content creator before publishing. The perspectives of the application of AI writing assistant software in clinical research are massive, especially if coupled with NLP translating tools to produce text documents disclosing complicated information to the general public. This is relevant to support medical writers in the production of documents complying with new strict European regulations, such as providing lay summaries (in each European language). Naturally, caution must be taken! Only specialised translators/writers should use this technology to generate documents. Pharmaceutical companies and medical writers are responsible for the accuracy of the information that is disclosed. The risk of emergent "fake health news" in wellness blogs and other social media should not be high, if the source information used by NLP-tools is correct. Nevertheless, policymakers need to take serious steps to protect the public from potential misinformation generated by a software (or its operator!) that is still in development.

Future challenges

In clinical research, the use of AI assistants specialised in regulatory writing is still at an early stage. Some biotech companies claim their software accurately produces 80% of a clinical study report, in less than 72 hours, still little official information on its efficacy is dislcosed by medical writing departments of pharma (or CRO) companies.[8] The improvement of AI-based tools for regulatory writing will inevitably redefine the role of medical writers. Medical writers will no longer be responsible for extracting information from different sources and for integrating it into a readable and comprehensible text document. AI-based tools can perform this task accurately in a short time, drastically reducing the load of preparation of documents. Unfortunately, due to strict confidentiality laws, scarce clinical data are available for use as sample data to develop and improve performance of NLP in the clinical field. Most available software uses historical data of clinical trials that pre-date the development of



AI-based tools for optimal performance.

In addition, little to no data exchange among different players, from pharma to academia and biotech companies, limits further progress of AI in clinical research. Remarkably, this culture of confidentiality is already diminishing through collaborations among the different players and the creation of repositories and platforms for sharing databases (e.g., clinical narratives with de-identified data). Another key step is the adoption of universal electronic health records, which contain an extensive and valuable coverage of clinical data of patients over time. Also, an important step towards AI-driven innovation is the implementation of dissemination centres that share resources among research communities, such as the Health Natural Language Processing Center (http://center.healthnlp.org) and European Language Resources Association (http://www.elra.info/en/).[9] Despite the current limitations of AI-driven technology, there is no doubt that clinical research is already being transformed with major benefits for the pharma industry and, ultimately, for the consumer.

Open-source AI software

For freelancers, open-source software is appealing, particularly if this technology is evolving as reported. If you have the chance to play around with this technology or if you are simply interested in getting familiar with what the future might look like, there are some options for user-friendly NLP-software. SemEHR, Apache cTAKES, GATE, and CLAMP, for example, are specialised in text analysis and extraction of target clinical information from electronic health records, unstructured clinical notes, and narrative patient reports. SemEHR also identifies contextualised mentions of biomedical concepts in clinical records, being able to pull out the cohort of relevant patients. You might gain some interesting insights on the array of possibilities that publicly available clinical databases and open-source NLP-tools can offer.

> A tremendous practical aspect of natural language processing tools is their ability to assist the medical writer in detecting errors during the anonymisation of patients enrolled in clinical studies, particularly in international multicentre clinical trials involving thousands of patients.

Conclusion

In a nutshell, the prospects of AI in the medical and healthcare industries translate into a significant reduction of time and costs required for the development of new drugs, an increase of new potential treatment options for rare diseases, and the development of personalised diagnostics and tailored therapeutics to the individual patient. The impact of AI-driven technology in the field of medical writing is immense. Any true lifelong friendship is not for the faintest of hearts and AI will, definitely, challenge the rise of a new breed of medical writers. We move at a fast pace, towards a future in which a medical writer, as a specialised craftsman, will use AI-based tools in a complementary fashion to enhance the writing process of complex documents. In the long run, AI will lessen repetitive tasks (e.g., extracting information from source and preparing document drafts) and give the medical writer more opportunities to apply their know-how in

producing complex scientific work. Remarkably, medical writers can use NLP tools to scan metadata across different databases and identify meaningful key information, otherwise inaccessible through classic keyword search. This has the potential to unravel new therapeutics options and is a rewarding feature of the career of medical writing. The intersection between medical writer and AI will have major implications across medical and healthcare industries, resulting in higher chances of disclosure of information to a wider audience and a cost-effective drug development and, therefore, more innovative treatments. Times are, indeed, exciting!

Conflicts of interest

The author declares no conflicts of interest.

References

1. Pesapane F, Codari M, Sardanelli F. Artificial Intelligence in medical imaging: threat or opportunity? Radiologists again at the forefront of innovation in medicine. Eur Radiol Exp. 2018;2(1):35.
2. US FDA. Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning (AI/ML)-Based Software as a Medical Device (SaMD) – Discussion Paper and Request for Feedback. [cited 2019 Aug 31]. Available from: https://www.fda.gov/media/122535/download.
3. Sennaar K. AI in pharma and biomedicine – analysis of the top 5 global drug companies. Emerj. 2019 [cited 2019 Aug 31]. Available from: https://emerj.com/ai-sector-overviews/ai-in-pharma-and-biomedicine/.
4. Bulgaru I. Pharma industry in the age of artificial intelligence: The future is bright. Healthcare Weekly. 2019 [cited 2019 Aug 31]. Available from: https://healthcareweekly.com/artificial-intelligence-in-pharmacology/.
5. Jiang F, Jiang Y, Zhi H, et al. Artificial intelligence in healthcare: past, present and future. Stroke Vasc Neurol. 2017;2(4):230–43.
6. Rocher L, Hendrickx J, de Montjoye Y. Estimating the success of re-identification in incomplete datasets using generative models. Nat Commun. 2019;10(1):3069.
7. Sennaar, K. AI and machine learning for clinical trials – Examining 3 current applications. Emerj. 2019 [cited 2019 Aug 29]. Available from https://emerj.com/ai-sector-overviews/ai-machine-learning-clinical-trials-examining-x-current-applications.
8. Fassbender M. Sanofi: NLP text mining and other AI tech to advance findings, improve data use. Outsourcing – Phama.com. 2019 [cited 2019 Aug 31]. Available from: https://www.outsourcing-pharma.com/Article/2019/01/30/Sanofi-NLP-text-mining-and-other-AI-tech-to- advance-findings-improve-data-use.
9. Velupillai S, Suominen H, Liakata M, et al. Using clinical natural language processing for health outcomes research: Overview and actionable suggestions for future advances. J Biomed Inform. 2018;88:11–9.

Author information

Sonia Costa, founder of Papyrus Scientific Communications, has a PhD in the field of bacterial infectious diseases and several years of experience in scientific communication. Since the beginning of 2019, she has been working as a freelance medical writer, and is currently involved in projects aiming for the improvement of AI software specific to medical writing.