Artificial intelligence for medical writing, part two: transparency, flexibility and control

AI should be a partnership between person and machine, says Dr Barry Drees, senior partner at Trilogy Writing & Consulting.

In an earlier article about artificial intelligence (AI) for medical writing, I stressed that AI tools would assist, accelerate and augment the process of clinical documentation. I also argued that the key to the third aspect, augmentation, was the use and future acceptance of standard texts. This, of course, is going to be the hardest to implement and where most of the pushback occurs.

Everyone is in favour of making the preparation of clinical documents easier and faster, but when you start talking about standardising the text, the acceptance drops precipitously. In addition, many writers feel uncomfortable with a tool that seems to be taking their thinking away from them and produces a document that they don’t understand or feel connected to. So how do we address these challenges to bring the huge potential power of AI to the writing of clinical documents? I believe that the keys are transparency, flexibility and control.

Let’s start with transparency. A common complaint about many AI systems is the ‘black box’ nature of how the system makes decisions. The AI is run on some data set and an answer is produced. But the reasons or logic behind the decision is not immediately clear; in fact, many AI programmers have to admit they have no idea exactly how the AI system came to its solution. This is an inherent weakness of ‘big data’ AI systems that have been developed by feeding thousands (in our case) of documents into the system and allowing it to look for commonalities. Although they can produce amazing and plausible results, one never really knows how the solution was achieved and thus what other possible solutions were passed over or ignored.

Some medical writers I have spoken with say they have tried AI tools and found them very unsatisfactory: you feed the protocol and data tables into the AI tool and some kind of output results. But you have no idea what the thinking was that went into the decisions, so it’s very hard to know whether you agree with the decisions or not.

However, unlike using big data, a rules-based approach (where the computer is instructed to follow specific rules) can avoid this problem in that the AI tool can be instructed to indicate in the standard text what was decided and how. For example, usually some sort of summarisation is used in the CSR text when describing a lengthy data table, like, for example, the list of treatment-emergent adverse events (TEAEs), which can be many pages long. This is frequently done by creating a small summary table within the text displaying only a selected portion of all of the events. However, this often leads the reader to wonder how the data were selected from the original TFL table. The events listed are usually chosen according to some kind of cut-off value in frequency (eg all events that occurred in at least 5% of participants) and an AI tool can be instructed to give the cut-off that was selected in the summarising table title and text, as in “The most common (>5% of all participants) TEAEs were as follows”. This can be used for all of the many poorly defined comparative words like “most common”, “well-balanced”, “relevant difference”, “slightly higher”, etc, that currently plague clinical study reports and about which the regulatory authorities frequently complain. Thus, transparency will be a critical aspect of any successful AI medical writing tool and can also improve the consistency of clinical documentation across studies.

The great need for this is illustrated by a recent experience I had with a submission team and a discussion about what to say about the subgroups in a study for a clinical study report. For safety, should the events be compared between the treatment groups within a subgroup or between the subgroup and its opposite (men versus women or elderly versus...
non-elderly)? The best choice is the former, as the latter will just show events typical for the subgroup (more pregnancies in the women, more deaths in the elderly, etc.). Standardised comparisons noting exactly what was done should eliminate the need for these discussions.

This brings us to the next key: flexibility. Another common complaint about AI systems based on big data is that they are often very dependent on the format of the data used to train the tool. There are cases where an AI tool trained for radiological diagnosis on one set of X-rays was then unable to be used by other clinics that used different X-ray machines and where the output had minor differences³. This problem led one radiologist to state, “I would want a human physician no matter what – even if a machine hums alongside.”³³ Training and creating an AI tool when all of the data is in an identical format is relatively straightforward; the trouble arises when the form of the output changes (sometimes for reasons totally unrelated to the AI tool). It is tempting for companies developing an AI tool for their internal use to design it specifically for their formats, but this can make it of very limited usefulness over time or across organisations.

In contrast, not being a large pharmaceutical company that could demand a consistent format but rather a service organisation with multiple clients, my company, Trilogy Writing & Consulting, did not have this luxury and wanted our tool to be as flexible as possible. As we would discover, this would present some serious programming challenges in dealing with different templates, data inputs and document styles. This was in addition to the flexibility necessary for the full range of clinical documents (just in terms of CSRs, we wanted to be able to handle phases I, II and III; comparative and non-comparative studies; and pivotal efficacy, massive safety and small first-in-human studies). Such flexibility is definitely a challenge in developing an AI tool, but if the goal is for AI to be truly transformational, then I believe it will be necessary.

The final key is control. Unfortunately, the word ‘control’ has acquired a rather unsavoury reputation of late, and many people immediately think of its abuse (and concepts like ‘control freak’, for example). In this case, however, I again would refer to the feedback that I have received from medical writers working with early versions of AI for clinical documentation. Many were frustrated by the fact that after running the tool they felt they had lost control of what was produced. In subsequent review cycles, they felt it was difficult to discuss the draft with clients or team members, as they were forced to defend writing or data selection decisions with which they disagreed. Thus, a useful AI tool will need to be interactive with the writer, allowing them to make decisions about the document before or as it is created as well as being able to easily edit the document that is produced to avoid the “It would have been easier if I had just done it myself from the beginning” sentiment common to many writers using current AI tools.

This is one of the biggest challenges facing any AI tool, in medical writing or elsewhere: acceptance of the tool by the people currently employed in whatever activity the tool performs. In medical writing, this is particularly true, as writers are encouraged to feel ownership of the documents they produce to ensure quality. Giving the writer a measure of input and control into the process will be essential for overcoming the acceptance problem. It also improves the final product, for, while there are things that computers can do so much better than people in generating text, there are other things that are just simpler for people, who tend to be better at ‘fuzzy logic’ than computers. The true potential of AI will not be with an exclusively computerised process, but rather a hybrid process – a partnership between person and machine.

Bringing AI abilities to the writing of clinical documents will not be like flicking a switch; it will require attention in the design and use of the AI tool to make it transparent, flexible and without losing control, in order to help make the approval process for new medicines faster and more transparent. There is great potential here, but it will require attention to the process and integration of the machine and the medical writer for the optimal outcome.

References
2. Allen, B. A Road Map for Translational Research on Artificial Intelligence in Medical Imaging: from the 2018 National Institutes of Health/RSNA/ACR/The Academy Workshop. JACR 2019; 16:

To find out more about Trilogy Writing & Consulting, visit www.trilogywriting.com

Outsourcing in Clinical Trials Handbook | 11