

# Writefull – Advanced academic language feedback

Writefull provides advanced academic and technical writing software that goes well beyond the tools provided by other grammar checkers. Using highly complex algorithms, it offers a full array of language feedback – covering everything from wording, syntax, and grammar – catered specifically to the academic community. Research Outreach caught up with Dr Hilde van Zeeland, Applied Linguist at Writefull, to learn more about how these algorithms work. We also spoke broadly about natural language processing (NLP) and Deep Learning algorithms as they pertain to academic publishing.

Writefull is unique in that its algorithms are trained by and focused upon technical scientific content, meaning its proofreading is tailored specifically to scientists and academics. This focus does not limit its application across disciplines, however, as the software is regularly tested on papers covering a vast array of topics. Writefull has recently taken its software to a higher level, with the launch of Full Edit. This fully AI-based feedback mode is capable of rigorous and thorough editing, ensuring all papers – regardless of the author's native tongue or language proficiency – reach the highest level of technical specificity and style. Dr Hilde van Zeeland told us more.

*Could you tell us about the history of Writefull, and about the expertise that went into its creation?*

Two of our team members developed the first version of Writefull around ten years ago, during our PhDs. Being junior researchers and non-native English speakers, we would often resort to Google while writing: we would search for a word or phrase, and use the number of results and the snippets as a sanity check. As we found out that many of our peers did the same, we decided to develop an app that would directly fetch this information from Google Scholar. At the time, this was more of a fun side-project than anything else. Our PhDs were in Computer Science (Artificial Intelligence) and Applied Linguistics, and we just enjoyed combining our field knowledge to develop an app – it was a welcome break from academia, too.

Researchers loved the app. But we also realised that, to fully support



Dr Hilde van Zeeland.

authors, it would be best if Writefull would automatically proofread their texts. This is where we moved to the current, AI-based version of Writefull. The development of this new version required a lot of technical and linguistic expertise. Parsing and cleaning the data, then training, testing, and retraining the models – the first 'new' version took months to develop, and the first years of it were mostly about manually checking the models' results, tracking the errors, and tweaking and retraining the models.

## Writefull's Full Edit fixes and rewrites even the longest and most complex sentences.

*Writefull uses Artificial Intelligence and Deep Learning processes to ensure its algorithms provide a thorough language editing software. Could you tell us more about these advanced algorithms?*

The output of any AI-driven model depends on the data it's been trained on. In Writefull's case, we trained the algorithms on scientific and technical content only; the training data consists of peer-reviewed Open Access articles. Thanks to this focused dataset, the models now give language feedback that is appropriate to scientific writing. You notice this in the sense that its

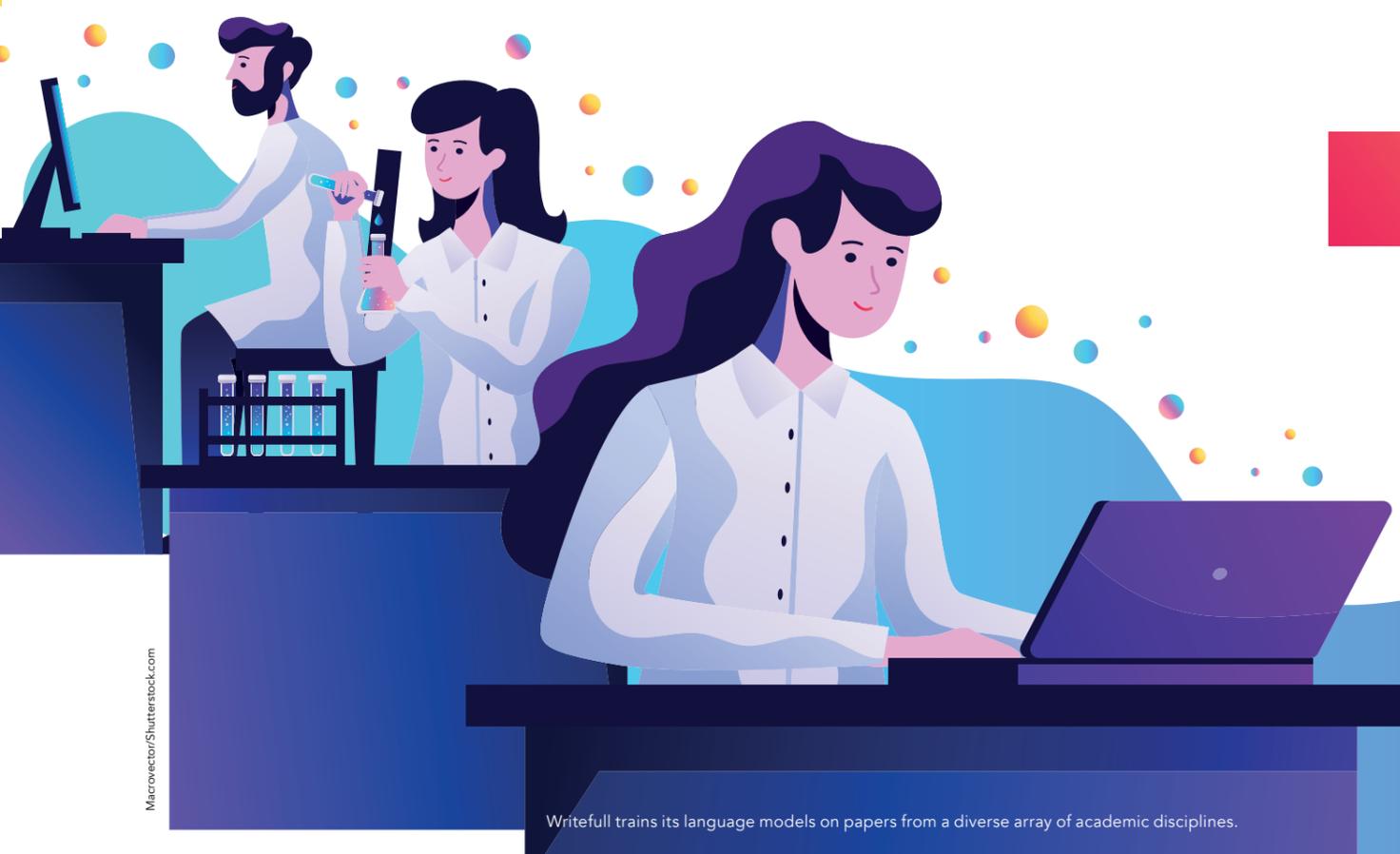
suggestions make a text sound more scientific – for example, it might suggest to replace 'a couple of' with 'several' or 'big' with 'key'. But you also notice it in other, more implicit ways. For example, it does not flag disciplinary vocabulary as misspellings as the models have seen these words in training, and it keeps functioning when sentences get complex thanks to the training data consisting mostly of complex sentences.

We've built the algorithms in such a way that they're flexible, and we can fine-tune them to new data with little effort. For example, given a dataset of physics

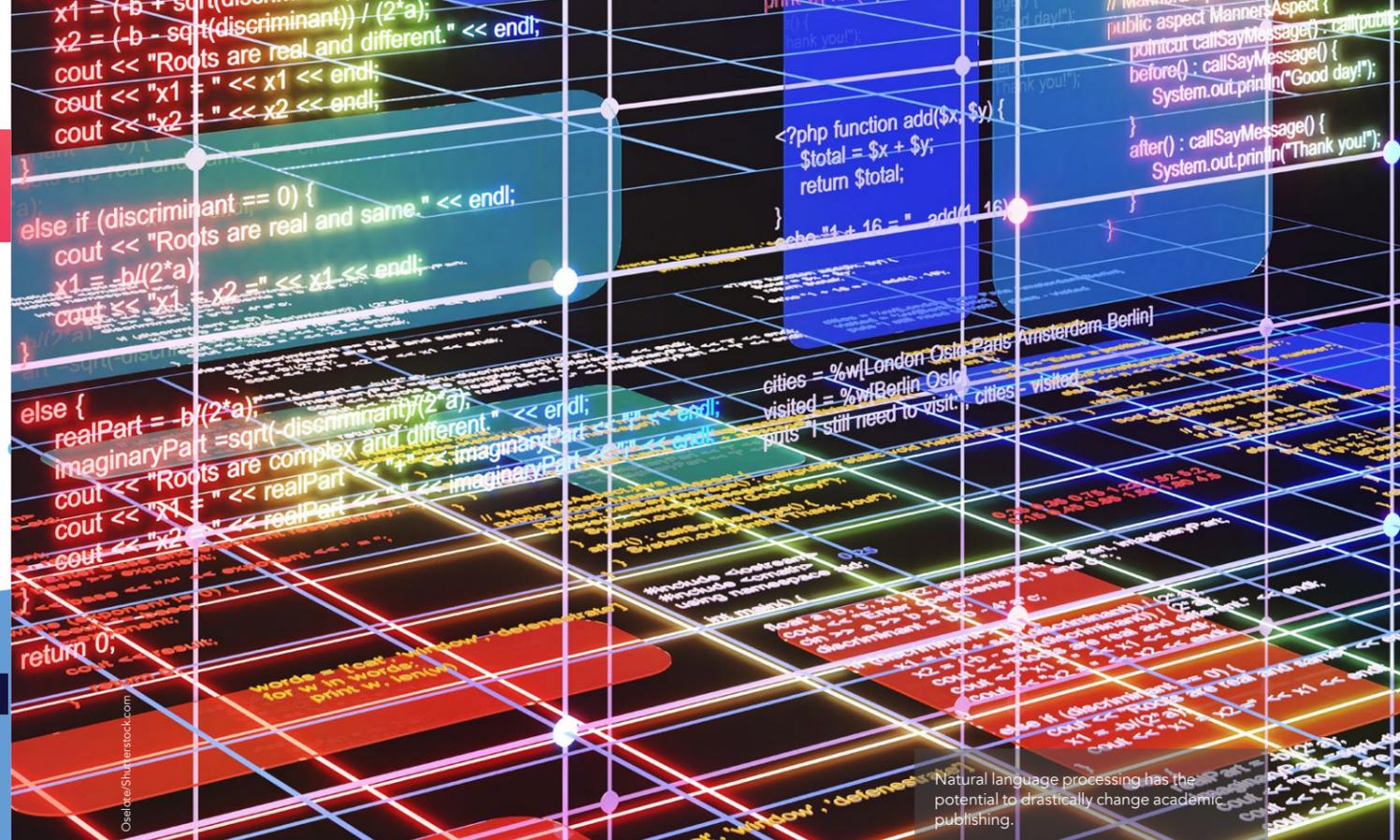
papers, we could make the models also pick up on inconsistencies or deviations within physics-specific sentences, such as around formulas.

*How does Full Edit – recently launched by Writefull – go beyond other tools in its language feedback capabilities?*

The main difference with other language tools is that Full Edit checks and corrects a lot more. Other tools are often limited to a basic grammar screening and a set of vocabulary checks. Writefull's Full Edit goes far beyond this, fixing and rewriting even the longest and most complex sentences. This difference



Writefull trains its language models on papers from a diverse array of academic disciplines.



Natural language processing has the potential to drastically change academic publishing.

comes from the fact that Writefull is fully AI-based, whereas other tools use little or no AI. They use predefined rules instead, of the type *the researches* becomes *the research* and *singular subject + plural verb* becomes *singular subject + singular verb*. Such replacements work, but only for as long as an author writes exactly what needs to be replaced so that it maps onto the predefined rules. This means that it works only for sentences that follow predictable grammar. I recently came across a tool changing 'The results of this survey reveal...' for 'The results

of this survey reveals...': the grammar rule fetched survey as the subject here. To the contrary, if you use AI that's advanced enough, like Full Edit, those errors won't occur. Full Edit's level is really impressive. If you use it with Track Changes activated in Word, it's almost scary – as if a human has been editing your document.

*Is Writefull aimed at academics from across a broad array of disciplines?*

Yes, it is. As we trained our models on Open Access content from a wide range of disciplines, the models perform well

on texts from any field. We regularly test Writefull on different types of texts: manuscripts from different fields or written by authors with different first languages or English-proficiency levels. We see its performance is consistent. Writefull is also used by publishers from a range of disciplines. For example, Hindawi and CUP both offer Writefull to their authors before submission. Writefull is also used by scientific copy-editing companies – all of which use the same language models for all of the disciplines they proofread.

In some cases, publishers ask us to fine-tune our models to their own data, so that they also pick up on highly disciplinary language characteristics. For example, we work with a publisher with a lot of field-specific language norms around style, such as how numbers are presented or figures are described. Once Writefull's models were fine-tuned on this publisher's content, they could make stylistic improvements that the editors would otherwise have to make manually.

*How are you hoping to improve Writefull in the coming months and years?*

The first steps are to bring some functionality that we have ready in the

**We've built the algorithms in such a way that they're flexible, and we can fine-tune them to new data with little effort.**

backend into the apps. For example, the option to select either British or American English. This check sounds simple, and it can be simple, too. But we want it to do more than only check basic spelling (-s or -z) and vocabulary (lift or elevator) – it should check for more nuanced differences, too. We have also recently launched a few AI-based writing widgets that we might add to our apps. The first is a title generator that generates a manuscript title based on an abstract. This is really tailored to research writing and may feel a bit niche, but we've found that it's especially those quirky things that people appreciate. The second is an automated paraphraser, also trained on scientific texts only, that helps students and researchers discover alternative words and phrases. Of course, we will keep improving the language models. This is something we're constantly working on and it's Writefull's main strength.

*Do you feel that the rapidly advancing capabilities of natural language*

*processing (NLP) and Deep Learning algorithms will lead to broader changes within scientific publishing?*

Yes, definitely. While language revision is key to publishing, it's one of many things that NLP can help with. From the author's perspective, NLP could help with many tasks during manuscript preparation: it could offer them relevant articles and auto-write digests of those, it could show them relevant journals based on manuscript content, it could screen their manuscripts for completeness and consistency, it could automatically proofread and revise their text, and it could even write part of their text. For the publisher, besides automated proofreading before submission, NLP can assist in automatically triaging manuscripts based on language quality, automatically copy-editing texts, and performing structural checks on manuscripts – think of checking whether all references are added to appropriate sentences. Many of these tasks and checks are semantic in nature: they

require an understanding of a text. We do see developments in all of these domains, and while not all are applied to research writing yet, this is likely only a matter of time. With more tasks becoming automated with the help of NLP, manuscripts will likely be published much quicker in the future.

*Where can our readers learn more about Writefull?*

Please visit our website ([www.writefull.com/](http://www.writefull.com/)) for more information about our products and company. Follow us on Twitter ([twitter.com/Writefullapp](https://twitter.com/Writefullapp)) to stay up to date!



**W:** <https://www.writefull.com/>  
**T:** <https://twitter.com/Writefullapp>

