# Comparison of Automated English Editing Tools



Dora Alexopoulou Principal Research Associate University of Cambridge

June 11, 2022

## Summary of comparison and key findings

This report compares the performance of seven online digital writing tools, Paperpal, Grammarly, Instatext, AJE, Trinka, Writefull and Quillbot.

Three sample texts of around 500 words were extracted from three separate papers from https://www.preprints.org/. Each sample is from the categories of humanities, life sciences and physics. The evaluation focuses primarily on the analysis of the language edits each tool contributes with some comments on aspects of the overall tool effectiveness and usability.

The language analysis first considers the number of edits suggested by each tool, the number of edits accepted by the Human Editor (HE) and the number of overlaps between the tool edits and the ones suggested by the HE. A summary of results is presented in Table 1 and Figure 1.<sup>1</sup>

Document	TOOL	Corrections	Accepted	Overlap with HE
All docs	Human Editor •	48		
	Paperpal •	149	58 (38,9%)	24/48 (50%)
	Grammarly -	88	32 (37,5%)	21/48 (43,7%)
	InstaText ·	265	111 (41,8%)	33/48 (68,7%)
	Trinka •	89	26 (29,2%)	12/48 (25%)
	AJE ·	62	28 (45%)	14/48 (29,1%)
	QuillBot •	246	49 (19.9%)	
	Writefull •	117	27(23%)	17/48 (35%)

Table 1: Summary of edits per tool

<sup>&</sup>lt;sup>1</sup>As explained later, the extensive paraphrasing Quillbot provides rendered scoring of overlap with HE impossible.



Figure 1: Summary of edits per tool

The results suggest a trade-off between the volume of suggestions contributed by each tool and the rate of acceptance as well as the type of suggested corrections. The rate of acceptance of corrections affects the perception of the tool by the user as helpful or not. Intuitively, if most of your time as an author is spent rejecting unhelpful suggestions, then the tool cannot be perceived as particularly helpful. At the same time, the volume of corrections can also impact the perception of a tool as helpful. A tool that provides minimal corrections might be perceived as mainly a correcting/checking tool rather than a tool enhancing text readability.

Paperpal and Instatext stand out for contributing a high number of (accepted) improvements, thus, impacting on a text more than the other tools. In addition, these tools provide the highest number of alternative wordings and phrasings, which enhance the readability of the texts in a more concrete manner, as discussed in more detail in the next section. However, the higher impact on text readability does mean that authors need to consider a higher volume of suggestions. The two tools have comparable rates of acceptance, but InstaText makes 40-45% more suggestions than Paperpal, yielding a much higher number of improvements. The higher number of suggestions. The volume of suggestions can be an important consideration for authors since, in a typical journal manuscript of 10,000 words, an author would have to consider around 1000 suggestions from Paperpal against 1750 from InstaText.

AJE and Grammarly have high rates of acceptance but the volume of suggestions and improvements is small, lower than the number of edits provided by the Human Editor. Thus, while of generally high accuracy, these tools have a weaker impact on the text.

Finally, Writeful, Trinka and Quillbot have low acceptance rates which means authors need to reject more than 70% of the suggested edits.

We also analyse in detail the types of suggestions made by each tool, classifying them in spelling, punctuation, grammar and word choice and clarity suggestions. Paperpal and Instatext make the highest number of word choice and clarity suggestions, generally creating a stronger impression of enhancing the readability and clarity of a text. This impression is stronger for InstaText since two-thirds of InstaText's suggestions involve alternative wordings and phrasings in comparison to under half by Paperpal.

In conclusion, the overall impression is that digital tools can complement human editing and improve the writing of authors. This is particularly evident in the case of Paperpal and InstaText which contribute many more improvements (accepted edits) than the Human Editor. In this respect, these tools expand human editing. A significant development is edits involving alternative wordings and phrasings which can help authors clarify their language and use more effective words and phrasing. Accuracy, though, remains a challenge since even the best tools have more than have of their edits rejected.

#### **User interfaces**

There are two broad types of interfaces. The first type allows uploading a document (e.g. Word) and then downloading a revised/edited version where revisions can be considered by the author as tracked changes on Word (Paperpal, AJE), or as changes indicated in the main text (Trinka, Writefull). The second type handles editing on a specially developed online interface (Grammarly, InstaText, Quillbot). Though exports are possible, users need to copy the edited text and import it back to their editor/document. In general, the various online interfaces are better suited to handling corrections and editing, in particular when tools make a high volume of editing suggestions. Grammarly's and InstaText's interfaces stand out for ease of use. InstaText's split window allowing comparison between original and edited version is particularly helpful. On the other hand, downloading an edited Word document with tracked changes is convenient for authors and works well for tools making a small number of editing suggestions. Word exports with track changes are likely to work for the majority of authors. In this respect, Paperpal and AJE are likely to appeal to most authors.

## Methodology

### About the author of the report



The evaluation was carried out by Dr Dora Alexopoulou, Principal Investigator at the University of Cambridge. Dora is a linguist, specialising in grammar and second language learning with a focus on second language writing. You can visit Dora's website here https://www.mmll.cam.ac.uk/ta259

### The sample texts

Three recent papers were selected from https://www.preprints.org/ from the categories humanities, life sciences and physics. Pre-prints have not undergone the usual final stages of editing for journal publications and, therefore, are still in need of considerable polishing. I selected around 500 words from the discussion section. I copied the extracts from pdf files and pasted them in a Word document which created some formatting issues. I tidied up the texts but some issues remained which the digital writing tools could pick. The sample texts can be found in the Appendix.

### **Classifying language edits**

For each tool, I considered the number of suggested edits, the number of accepted edits and the number of edits overlapping with the edits of the HE.

To better understand the nature of suggestions made by each tool, the areas targeted and why the number of suggestions can vary so much between tools we classified suggestions in categories. Consider the sentences in (1) and (2) below.

- (1) a. ...graduates require basic sustainability literacy, and engagement with sociological and ethical engagements issues
  - b. SRD education requires not only appropriate learning principles and approaches, but also capacities to navigate obstacles ... and ... transition strategies [ which ] we describe below.
- (2) a. Datasets  $\rightarrow$  data sets
  - b. cause and effect  $\rightarrow$  cause-and-effect

Intuitively, the suggestions in (1) increase the clarity and readability of the text and are likely to be perceived as enhancing the text by an author. The suggestions in (2), while contributing to consistent spelling, are welcome by authors but are not perceived as enhancing the clarity and readability of a text.

I classified corrections in broad categories, avoiding fine grained distinctions so that numbers in each group would not be too small to interpret. Below I show examples of each category.

#### Spelling and typos

Any issue identified regarding the spelling of words, including capitalisations and punctuation inside words as in examples in (3).

- (3) a. re collapses  $\rightarrow$  recollapses
  - b. pair produces  $\rightarrow$  pair-produces
  - c.  $follow \rightarrow follow$
  - $d. \quad Duoverse \to duoverse$
  - e. ex perimental  $\rightarrow$  experimental

#### Punctuation

Standard punctuation corrections, commas, full stops etc. including inverted commas on individual words, e.g. 'happen' in Physics sample.

#### **Grammar corrections**

Corrections involving articles, verb agreement or preposition choice as in (4).

- (4) a. ... used for [the] evaluation of learning
  - b. ..evolving threshold concepts ... adds greater robustness
  - c. ...samples derived [from the] binding database

#### **Word Choice**

Generally lexical alternatives as in (5); I also classified as word choice corrections involving verb tense or noun plurality.

- (5) a. utilized  $\rightarrow$  used
  - b. engagement  $\rightarrow$  exposure
  - c. pointed  $\rightarrow$  points
  - d. in taught class  $\rightarrow$  in taught classes

#### Style

Suggestions to turn passive voice sentences into active and one suggestion to rewrite number 5 in letters (as five).

#### Clarity

A wide range of suggestions targeting the clarity of a sentence, including alternative wordings and/or word order. Some suggestive examples in (6).

- (6) a. SRD requires... transition strategies [which] we describe below
  - b. in supporting encounters  $\rightarrow$  to support encounters
  - c. knowledge traditionally generated in design education oriented toward aestehetic concerns  $\rightarrow$  traditionally generated in design education, which focuses on esthetic concerns

### Scoring

I labelled each suggestion made by the HE and digital tools in one of the above categories and scored them as accepted or not (except for HE). I was generally conservative as an Editor and accepted only suggestions that, in my view, did not deviate from the intended meaning. It is possible that, in some cases, authors would have been more accepting, in particular, in the case word choice and clarity suggestions. I also respected the authors' decision regarding the use of capitals for key terms in their texts. For example, Universe and Duoverse were corrected by many tools to 'universe' and 'duoverse'; I rejected such suggestions as it is common in scientific papers to use capitals for terms that are defined in a technical sense in a paper. If a word involved two corrections, I count it as one. E.g. AJE

Duoverses  $\rightarrow$  duoverse, correction from plural to singular and replacement of initial capital with small 'd' was counted as one Spelling error.

To consider the overlap between the corrections of a tool and the corrections made by myself (HE), I identified in each edited text those issues that had been picked by the HE but not by the digital tool. I calculated all other issues as correctly recalled. This approach was taken because in many cases clarity corrections meant that the issue initially identified by the HE was not relevant in the corrected version, so I counted it as addressed. Consider example 4a above. Some tools might have offered the acceptable correction 'used to evaluate learning'. In this case, I considered the grammatical error picked by the HE as correctly addressed, even if the tool correction did not include a grammatical error label per se.

## How tools target different types of language edits

We first consider word choice and clarity corrections. They are crucial to our analysis because they distinguish the digital editing tools from the usual spell and grammar checkers accompanying most editing programs. As can be seen in Table 2, Paperpal and InstaText have comparable acceptance rates (with InstaText higher). At the same time, InstaText makes roughly three times more suggestions than Paperpal leading to three times the improvements Paperpal contributes. AJE contributes only 6 improvements in this area. (Though QuillBot contributes a high number of improvements here the overall acceptance rate is too low and, as discussed later the overall quality of the suggestions is poor.)

Document	TOOL	WC/Clarity Corrections	Accepted	Notes
All docs	Human Editor •	13		
	Paperpal •	58	22 (37,9%)	
	Grammarly -	31	3	Too small numbers for percentages
	InstaText •	178	74 (41,5%)	
	Trinka •	19	7	Too small numbers
	AJE -	21	6	Too small numbers
	QuillBot •	227	39 (17,1%)	
	Writefull •	54	7 (12,9%)	-

Table 2: Summary of word choice and clarity edits per tool

Table 3 summarises spelling errors and typos. The numbers are too small to be interpreted reliably in terms of percentages; some tools (e.g. Papepal and Trinka) tend to suggest unnecessary capitalisations. Spelling errors is an area of strength for Grammarly. Also note that many spelling and punctuation corrections, are often about various conventions e.g. between British and American spellings, hyphens etc. which might partly explain differences between tools. Nevertheless, it is worth noting that AJE contributes the same number of improvements in Spelling errors (7) as in word choice and clarity corrections (6). This contrasts with Paperpal which contributes 22 word choice and clarity corrections in comparison to 8 Spelling errors. Even starker is the contrast with InstaText, which contributes 20 improvements in spelling errors in comparison to 74 word choice and clarity corrections.

Document	TOOL	Spelling/Typos	Accepted	Notes
All Docs	Human Editor •	18		
	Paperpal •	31	8	
	Grammarly -	28	22	
	InstaText ·	55	20	
	Trinka -	44	7	
	AJE ·	16	7	
	QuillBot -	8	3	
	Writefull •	33	14	

Table 3: Summary of spelling edits per tool

Table 4 summarises punctuation suggestions. This is generally an area of strength for all tools and for Paperpal in particular. Writefull is an exception to this. What is noticeable here is that AJE and InstaText contribute similar numbers of accepted improvements; in the case of AJE it is 10 punctuation corrections out of a total of 28 improvements while for InstaText it is 11 punctuation improvements out of a total 111 improvements. Paperpal contributes a higher number of punctuation improvements: 21 out of a total of 58 improvements are punctuation edits, meaning that punctuation is a sizeable part of the overall Paperpal improvements.

Document	TOOL	Punctuation	Accepted	Notes
All Docs	Human Editor •	12		
	Paperpal -	22	21	
	Grammarly •	8	5	
	InstaText •	14	11	
	Trinka ·	10	9	
	AJE -	13	10	
	QuillBot •	3	3	
	Writefull •	11	3	

 Table 4: Summary of punctuation edits per tool

Document	TOOL	Grammatical Errors	Accepted	Notes
All Docs	Human Editor •	5		
V	Paperpal •	39	6	
	Grammarly -	14	1	
	InstaText •	17	5	
	Trinka •	16	3	
	AJE ·	12	5	
	QuillBot •	8	4	
	Writefull •	19	3	

### Table 5: Summary of grammar edits per tool

Table 5 shows grammatical errors. This is a problematic area for many tools including Paperpal. For Paperpal, out of 91 rejected suggestions (see Table 1), 33 involve suggested grammatical errors. This means that grammatical errors account for around 36% of the total of rejected edits. Most cases involve the articles, erroneously introducing a definite article when an articless noun is acceptable or preferred. AJE and InstaText are doing well with grammatical errors.

To summarise, the quantitative data suggest that different tools target different types of corrections and show varying performance across the different categories. These quantitative differences reflect the user experience. Paperpal makes many helpful suggestions regarding word choices and clarity, but these are just under half of the total of suggestions; so the user may feel distracted by the number of spelling and grammatical error corrections, many of which are unnecessary. By contrast, two-thirds of InstaText's corrections involve word choice and clarity. This creates a stronger impression that the tool enhances the readability of the text. AJE's corrections engage with the clarity and readability of the text in a very minimal way so that the tool feels more focused on correcting and checking language rather than on enhancing the text in terms of clarity and readability. Note also that AJE's overlap with HE corrections is relatively low.

It is worth noting that accepted corrections for Paperpal and InstaText outnumber the HE corrections, suggesting that these tools can complement and strengthen human editing.

## Scores per tool

## **Human Editor**

Editor	Doc	Corrections	Number of corrections	Accepted	Notes
Epxert Editor	Humanities	Spelling/typos ·	0		
		Punctuation +	2		
		Grammar (e.g •	4		
		Word Choice 🕞	2		
		Style (e.g. pass ·	0		
		Clarity (e.g. re •	6		
		ALL •	14		
	Life Sciences	Spelling/typos ·	7		V
		Punctuation -	1		
		Grammar (e.g +	1		
		Word Choice 🔹	0		
		Style (e.g. pass +	0		
		Clarity (e.g. re *	2		
		ALL -	11		
	Physics	Spelling/typos •	11		Broken words
		Punctuation •	9		inverted commas
		Grammar (e.g +	0		
		Word Choice -	1		
		Style (e.g. pass +	0		
		Clarity (e.g. re •	2		
		ALL	23		
	ALLDOCS	ALL	48		
		Spelling/typos ·	18		
		Punctuation ·	12		
		Grammar (e.g +	5		
		Word Choice 🔹	3		
		Style (e.g. pass	0		
		Clarity (e.g. re *	10		
		WC+Clarity ·	13 (27%)		

Table 6: Summary of edits by the Human Editor

The Physics doc had a lot of spelling errors due to broken words; there were issues with punctuation and a couple of points where I suggested more explicit language for clarity.

The Life Sciences was generally well written and clear. There was one long sentence that would need to be rephrased for clarity. The style was more reporting findings rather than developing an argument, so the language followed a template.

The Humanities doc had most clarity issues, primarily due to long and complex sentences; a couple of grammar errors were due to the complex language.

#### Grammarly

Grammarly is doing very well with spelling and punctuation but suggestions for style and clarity are generally not helpful. One persistent issue is flagging article errors and making erroneous suggestions.

The overall impression is of a tool that is doing very well with spelling and punctuation but does not enhance the readability of the text. The style errors predominantly identify sentences in passive voice e.g. 'be ordered', 'be invoked' (physics sample) and suggest rewording. This particular suggestion reduced the credibility of the tool because of the indiscriminate way in which the passiveto-active suggestion was applied. Grammarly did pick some of the problematic sentences in particular in the humanities text but the suggestions made were rejected.

The interface was easy to use with a very pleasant layout. The corrections are easy to see and handle. The system also asks to specify the type of document I submitted, e.g. academic, business etc. whether it was an essay or report. Grammarly allows export of a detailed report on errors (though not clear this information is useful).

Editor	Doc	Corrections	Number of corrections	Accepted	Notes	Overlap with HE
Grammarly	Humanities	Spelling/typos -	1	1		0/1
		Punctuation •	2	0		0/2
		Grammar (e.g. p •	6	1	articles	1/4
		Word Choice ·	4	0		0/2
		Style (e.g. passive) +	2	0		
		Clarity (e.g. rewo *	5	3		0/6
		ALL	20	5	25%	1/14
	Life Sci	Spelling/typos -	10	8		7/7
		Punctuation *	0			0/1
		Grammar (e.g. p •	1	0		0/1
		Word Choice ·	3	0		0/0
		Style (e.g. passive) +	2	1		0/0
		Clarity (e.g. rewo *	2	0		0/2
		ALL -	18	9	50%	7/11
	Physics	Spelling/typos *	17	13		9/11
		Punctuation -	6	5		4/9
		Grammar (e.g. p •	7	0	Mostly articles	0/0
		Word Choice -	2	0		0/1
		Style (e.g. passive) +	3	0		0/0
		Clarity (e.g. rewo *	15	0		0/2
		ALL ·	50	18	36%	13/23
	3-Docs	ALL •	88	32	37.5%	21/48 (43,7%)
		Spelling/typos ·	28	22	78%	16/19
		Punctuation •	8	5	62%	4/12
		Grammar (e.g. p +	14	1		1/5
		Word Choice ·	9	0	0	0
		Style (e.g. passive) +	7	1		
		Clarity (e.g. rewo *	22	3		0
		WC+Clarity -	31	3		0

Table 7: Summary of edits by Grammarly

### Paperpal

Paperpal makes a high number of editing suggestions (149 in total). The acceptance score is reasonably high at 38.9% as is the overlap with HE. The tool is doing very well with punctuation. The acceptance score for spelling errors is lower mainly due to the over-correction of capital letters in special terms used by authors and unnecessary capitalisation of words at the beginning of (broken) lines. Grammar errors also have a low acceptance score primarily due to difficulty distinguishing when an article is needed and when not. Clarity and word choice suggestions account for 30% of all suggestions.

The interface is nice and easy to use; it was very easy to upload the papers. Edited papers can then be downloaded with the corrections showing on Word as tracked changes. The Word output would be appreciated by users of Word. Paperpal provides a relatively high number of corrections which are displayed as tracked changes on a Word Document. Paperpal makes many helpful clarity/word choice suggestions that can enhance the readability of the texts and which set Paperpal apart from the more regular grammar and spellcheckers. But this positive impression is countered by a very high number of unnecessary spelling and grammar corrections.

Editor	Doc	Corrections	Number of	Accepted	Notes	Overlap with HE
			corrections			
Paperpal	Humanities	Spelling/typos •	7	2		0/0
		Punctuation *	10	9		1/2
		Grammar (e.g. preposition, +	15	2	articles	3/4
		Word Choice ·	10	4		1/2
		Style (e.g. passive) -	0	0		
		Clarity (e.g. rewording) *	12	6		1/6
		ALL	54	23	42% acceptance	6/14
	Life Sci	Spelling/typos *	10	3		4/7
		Punctuation *	6	6		0/1
		Grammar (e.g. preposition, +	16	4	articles	1/1
		Word Choice -	9	3		
		Style (e.g. passive) +	1	1		
		Clarity (e.g. rewording) *	11	2		0/2
		ALL •	51	19	37%	5/11
	Physics	Spelling/typos ·	14	3		6/11
		Punctuation -	6	6	Ψ.	6/9
		Grammar (e.g. preposition, $+$	8	0		
		Word Choice *	10	2		0/1
		Style (e.g. passive) *	0	0		
		Clarity (e.g. rewording) *	6	5		1/2
		ALL ·	44 (without S=30)	16 (without S=13		13/23
	3-docs	ALL -	149	58	38,9%	24/48 50%
		Spelling/typos •	31	8		
		Punctuation ·	22	21		
		Grammar (e.g. preposition, *	39	6		
		Word Choice ·	29	9		
		Style (e.g. passive) 🔹	1	1		
		Clarity (e.g. rewording) •	29	13		
		WC+Clarity -	58 (=30%)	22	37,9%	

Table 8: Summary of edits by Paperpal

## Writefull

Editor	Doc	Corrections	Number of corrections	Accepted	Notes	Overlap with HE
Writefull	Humanities	Spelling/typos -	3	3		0/0
		Punctuation -	3	0		0/2
		Grammar (e.g. preposit •	4	1		2/4
		Word Choice ·	11	1		1/2
		Style (e.g. passive) *	0			0/0
		Clarity (e.g. rewording) +	12	1		1/6
		ALL •	33	6	18%	4/14
	Life Sciences	Spelling/typos -	14	1		2/7
		Punctuation -	4	0		0/1
		Grammar (e.g. preposit *	10	2		1/1
		Word Choice ·	15	2		
		Style (e.g. passive) -	1	1		
		Clarity (e.g. rewording) +	7	2		0/2
		ALL •	50	7	14%	3/11
	Physics	Spelling/typos -	16	10		5/11
		Punctuation -	4	3		5/9
		Grammar (e.g. preposit +	5	0		0/0
		Word Choice ·	7	0		0/1
		Style (e.g. passive) -				0/0
		Clarity (e.g. rewording) ·	2	1		0/2
		ALL	34	14	41%	10/23
	ALL DOCS		117	27	23%	17/48 (35%)
		WC+Clarity -	54	7		
		Spelling/typos ·	33	14		
		Punctuation -	11	3		
		Grammar (e.g. preposit +	19	3		

Table 9: Summary of edits by Writefull

The quality of corrections is overall not very high. It was not always easy to understand what the intended correction was.

The website is nice but it took me a while to find how to use 'writefull revise' and avoid uploading the app (the app is not compatible with Ubuntu, which is the operating system on my laptop). Writefull accepts Word docs but also Tex files so authors who use Latex can use it, in particular, if they use Overleaf. I uploaded Word docs on 'writefull revise' and then downloaded the edited files.

### AJE

Relatively small number of corrections of generally good accuracy. Across all three docs acceptance is at 45%. A third of all corrections are word choice/clarity errors (21/62=33.8%). The general impression is of a tool that is rather 'conservative' in the suggested corrections in comparison to other tools, with some helpful suggestions but overall not creating a strong impression that it can enhance the readability of a text. Note it misses roughly 2 out 3 issues picked by the HE.

The interface is simple to use and ideally suited for users of Word as one can upload a Word document and then download an edited version with all corrections as tracked changes in the Word document. Given the relatively small number of corrections, tracked changes on a Word document work well.

Editor	Doc	Corrections	Number of Corrections	Accepted	Notes	Overlap with HE
AJE	Humanities	Spelling/typos •	4	2		0/0
		Punctuation •	2	1		0/2
		Grammar (e.g *	3	2		2/4
		Word Choice *	1	0		0/2
		Style (e.g. pass *	0			0/0
		Clarity (e.g. re *	1	1		1/6 🔍
		ALL -	11	6		3/14
	Life Sciences	Spelling/typos +	2	1		2/7
		Punctuation +	5	3		0/1
		Grammar (e.g *	5	3		1/1
		Word Choice *	10	1		
		Style (e.g. pass *	0	0		0/0
		Clarity (e.g. re	2	1		0/2
		ALL -	24	9	37.5%	3/11
	Physics	Spelling/typos *	10	4		1/11
		Punctuation *	6	6		6/9
		Grammar (e.g *	4	0		-
		Word Choice -	4	1		1/1
		Style (e.g. pass •	0	0		-
		Clarity (e.g. re *	3	2		0/2
		ALL	27	13	48%	8/23
	ALL DOCS	ALL •	62	28	45%	14/47 29,1%
		Spelling/typos *	16	7		
		Punctuation ·	13	10		
		Grammar (e.g	12	5		
		WC+Clarity +	21 (33,8%)	6	28,5%	

Table 10: Summary of edits by AJE

## InstaText

Editor	Doc	Corrections	Number of corrections	Accepted	Notes	Overlap with HE
InstaText	Humanities	Spelling/typos ·	16	10		0/0
		Punctuation ·	5	3		0/2
		Grammar (e.g 🔹	3	2		4/4
		Word Choice ·	47	18		2/2
		Style (e.g. pass ·				
		Clarity (e.g. re *	22	13		4/6
		ALL	93	46	49%	10/14
	Life Sciences	Spelling/typos ·	4	1		7/7
		Punctuation ·	1	1		1/1
		Grammar (e.g •	9	3		1/1
		Word Choice ·	33	18		0/0
		Style (e.g. pass	1	1		0/0
		Clarity (e.g. re +	20	12		0/2
		ALL -	68	36	52.9%	9/11
	Physics	Spelling/typos ·	35	9		7/11
		Punctuation ·	8	7		6/9
		Grammar (e.g •	5	0		0/0 👻
		Word Choice ·	41	8		1/1
		Style (e.g. pass ·	0	0		0/0
		Clarity (e.g. re •	15	5		0/2
		ALL ·	104	29	27.8%	14/23
	ALL DOCS	ALL -	265	111	41,8%	33/48 (68,7%)
		Spelling/typos ·	55	20		
		Punctuation ·	14	11		
		Grammar (e.g +	17	5		
		Word Choice -	121	44		
		Style (e.g. pass +	1	1		
		Clarity (e.g. re •	57	30		
		WC+Clarity -	178 (67%)	74	41,5%	

Table 11: Summary of edits by InstaText

High acceptance score and a very high overlap with HE. A good range of word/phrasing alternatives, around half of which very helpful. Low rate of false positives for grammatical errors. What stands out is the very high number of word choice and clarity suggestions (67% of all suggestions). Moreover, in particular in the humanities text, many of the rejected word choice and clarity suggestions are plausible alternatives. This contrasts with other tools where suggestions of alternative wording/phrasing are implausible, often distorting rather than enhancing the intended message. The performance on the physics text is noticeably lower. This is due to a high number of correcting capital letters in technical terms introduced by the authors. In addition, the physics sample is a more scientific text and the alternative phrasings could not capture the precise meaning intended by the authors. For example only 19% of word choice suggestions were accepted in the physics text, in comparison to 54% in the Life Sci text and 38% in the humanities text.

InstaText makes a high number of suggestions that would mean 1500 to 2000 for a full paper. Some authors might find this number of corrections excessive, given that more than half are usually rejected.

The interface is very easy to use and very helpful. One cannot upload a file but needs to paste a document. The main window is split in two with the original document on the left and the document with the suggested corrections on the right. The interface on the right allows the user to view the corrected version but also the suggested corrections with a choice appearing on top of each word allowing the author to accept or reject a suggestion. Given the high number of corrections InstaText makes, this interface is crucial to allow comparison with the original text and easy handling of corrections. Once an author has completed their editing they can copy-and-paste their revised text to their editor. Many authors might find useful an export function with an edited paper in Word.

## QuillBot

Editor	Doc	Corrections	Number of corrections	Accepted	Notes
QuillBot	Humanities	Spelling/typos ·			
		Punctuation -			
		Grammar (e.g 🔹	4	3	
		Word Choice ·	63	6	
		Style (e.g. pass +			
		Clarity (e.g. re	39	13	
		ALL •	106	22 (20%)	
	Life Sciences	Spelling/typos ·	3	0	
		Punctuation ·	0		
		Grammar (e.g *	2	1	
		Word Choice +	34	3	
		Style (e.g. pass *			
		Clarity (e.g. re •	38	6	
		ALL •	77	10	
	Physics	Spelling/typos -	5	3	
		Punctuation •	3	3	
		Grammar (e.g 🔹	2	0	
		Word Choice +	25	5	
		Style (e.g. pass +			
		Clarity (e.g. re *	28	6	
		ALL •	63	17 (26,9)	
	ALL DOCS	ALL	246	49 (19,9%)	
		Spelling/typos ·	8	3	
		Punctuation -	3	3	
		Grammar (e.g +	8	4	
		WC+Clarity -	227	39	

Table 12: Summary of edits by QuillBot

The interface looks beta; it took a while to register and subscribe. The exported file did not have any corrections, unlike the online interface. The drop-down menu with list of alternative words/phrasings is an interesting feature. I have only scored the suggested corrections (rather than the full range of choice).

It seems that the goal of the tool is to offer alternative wordings/phrasings to authors. Thus, each alternative word on the online interface is related to a list of other potential choices. The paraphrasing is on a line by line basis which means that some suggestions lead to incoherence or no linking to the next line. It was hard to score, as the paraphrasing is often significant and it is hard to record individual changes. There were some critical errors, e.g. suggesting 'fundamental' as an alternative 'foundational' when the context is undergraduate curriculum. Some sentence reorganisations for clarity purposes are very good but there many incoherent suggestions, deviating from the original text significantly and in some cases distorting completely the meaning as shown in the examples below.

- (7) a. Original: Thus, electric charge would be the counterpart of quantum spin where quantum sin is intrinsic spin around the spatial axes.
  - b. QuillBot: Thus, electric charge is the polar opposite of quantum spin, which is inherent spin around the spatial axes.
- (8) a. Original: ... we could benefit from analysing the effects of data quality and quantity
  - b. QuillBot: ...we could benefit from evaluating this data because it is very important for predicting protein-ligand affinity the implications of data quantity and quality
- (9) a. Original: Such a dramatic effect is similar to observations from previous studies using different datasets.
  - b. QuillBot: Observations from prior studies utilising other datasets have found a strong influence.

The overall acceptance score is low at 28,4%. It became pointless to calculate the overlap with the HE due to the very high rate of paraphrasing, which would have led to an artificially high overlap.

## Trinka

Editor	Doc	Corrections	Number of corrections	Accepted	Notes	Overlap with HE
Trinka	Humanities	Spelling/typos •	23	4	Many incorrect capitalisations at line breaks	
		Punctuation -	2	2		0/2
		Grammar (e.g *	8	2		1/4
		Word Choice +	1	1		0/2
		Style (e.g. pass *				
		Clarity (e.g. re *	1	1		1/6
		ALL	35	10	28%	2/14
	Life Sciences	Spelling/typos •	15	2		3/7
		Punctuation *	1	1		0/1
		Grammar (e.g +	3	1		0/1
		Word Choice +	7	3		0/0
		Style (e.g. pass *	0	o		0/0
		Clarity (e.g. re *	1	0		0/2
		ALL	27	7	25,9%	3/11
	Physics	Spelling/typos -	6	1		2/11
		Punctuation -	7	6		5/9
		Grammar (e.g *	5	0		0/0
		Word Choice *	4	0		0/1
		Style (e.g. pass *	0	0		0/0
		Clarity (e.g. re +	5	2		0/2
		ALL •	27	9	34,6%	7/23
	All DOcs	ALL	89	26	29,2%	12/48 25%
		Spelling/typos •	44	7		
		Punctuation *	10	9		
		Grammar (e.g *	16	3		
		WC+Clarity ·	19	7		

Table 13: Summary of edits by Trinka

This tool feels very much like a spell-checking tool. A relatively small overall number of suggestions. Only a fifth of suggestions involve word choice and clarity

edits with the rest on spelling, punctuation and grammar. Both acceptance rate and overlap with HE are low, below 30

The interface is very nice allowing uploading of documents which can then be downloaded with edits. Edits are shown directly on the text (rather than as tracked changes). This set up is helpful for Word users. In addition Trinka provides a summary of suggested revisions.

## Appendix

Below I include the first paragraph from the Physics extract as edited from the various tools to indicate the nature and extent of edits.

The authors introduce a technical term Duoverse; various tools correct the capital letter, a correction I rejected treating Duoverse as a term introduced and defined by the authors. Trinka missed 're collapses' while IstaText corrected to 'collapses again' which would have been acceptable in a non-scientific document, but recollapses is a more suitable choice for a technical text. Many tools corrected 'which is discrete' to 'that is discrete'. It is generally ture that 'that' is the most frequent relativiser in English, but in this case 'which' is preferable. Various tools introduced a 'that' after 'this means', which I accepted as a suggestion (though I did not make myself). I introduced a repetition of 'which cycle' in the middle of the second paragraph to clarify the meaning of the sentence that no tool overlapped with. InstaText has made some radical rephrasings some of which are improving the text. QuillBot has made the most radical rephrasings leading though to incoherent sentences.

#### **Human Editor**

#### XVII. THE MANY WORLDS

The Duoverse described thus far contains all the events in the Universe and anti-Universe for a single expansion from beginning to end. However, the Duoverse then re collapses-S, annihilates, and pair produces-S(hyphen) a brand new Duoverse. Therefore, we can think of each successive expansion and contraction of the Duoverse as happening along another dimension.P-comma which is discrete. This dimension essentially labels the different countably infinite random set of Duoverse.

Since each Duoverse begins with annihilation, this means each Duoverse begins with a random configuration after annihilation. Therefore, there is no cause and effect relationship between Duoverses from cycle to cycle. This means the cycles cannot be ordered sequentially.Pcomma because there is no way to know which cycle preceded or which cycle CL will folow the current cycle. If we cannot order the cycles in a sequence, then we can think of them all as being --> as being all-CL parallel to each other. While events within a cycle can have cause and effect relationships (i.e. the events 'happen' -P at given times), the various cycles themselves do not 'happen' P they just exist along side all other cycles. Thus P comma we can think of the annihilation events as being a single event from which infinite Duoverses emerge and to which they return. This implies that finding ourselves in a particular Duoverse is completely probabilistic Pcomma where the probability that we find ourselves in a Duoverse with a particular configuration depends on how likely that configuration

is across all possible configurations. This gives us the many worlds that have been invoked to explain quantum probability in the Everett many worlds interpretation of QM.

Extract from the physics document: edits by the HE

#### Paperpal

	1	Paperpal	Deleted: the
XVII. THE MANY WORLDS		Paperpal	Deleted: U
The Duoverse described thus far contains all events in the universe and anti-universe for a single expansion		Paperpal	Deleted: U
from beginning to end. However, the duoverse then recollapses, annihilates, and the pair produces a brand new duoverse. Therefore, we can think of each successive expansion and contraction of the duoverse as		Paperpal	Deleted: D
occurring along another dimension that is discrete. This dimension essentially labels the countably infinite		Paperpal	Deleted:
Because each Duoverse begins with annihilation, each Duoverse begins with a random configuration after	×.	Paperpal	Deleted: D
annihilation. Therefore, there is no cause and effect relationship between duoverses from cycle to cycle. This means that the cycles cannot be ordered sequentially, because there is no way to know which cycle precedes	1	Papernal	Deleted: D
or follows the current cycle. If we cannot order the cycles in a sequence, then we can think of them as parallel	N	Panamal	Deleted: happening
given times), the various cycles themselves do not 'happen, they just exist along the side of all other cycles.	$\mathbb{N}$	r aperpai	Deleted. happening
Thus, we can think of the annihilation events as being a single event from which infinite duoverses emerge	A.	Paperpal	Deleted: Which
where the probability that we find ourselves in a duoverse with a particular configuration depends on how likely	M	Paperpal	Deleted: different
that configuration is across all possible configurations. This gives us the many worlds that have been invoked to explain guantum	$\mathcal{N}^{1}$	Paperpal	Deleted: D
probability in the Everett many worlds interpretation of QM.	24	Paperpal	Deleted: Since
的现在分词	λŇ		N 1 / 1 /kis

Extract from the physics document: edits by Paperpal

### Grammarly

Sample2-physics	33 All suggestions	HIDE ASSISTANT >>	
XVII. THE MANY WORLDS The Duoverse described thus far contains all the events in the Universe and anti-Universe for a single expansion from beginning to end. However, the Duoverse then re collapses, annihilates, and pair produces a brand new Duoverse. Therefore, we can think of each successive expansion and contraction of the Duoverse as happening along another dimension which is discrete. This dimension essentially labels the different countably infinite random set of Duoverse begins with an andom configuration after annihilation. Therefore, there is no cause and effect relationship between Duoverses from cycle to cycle. This means the cycles cannot be ordered sequentially because there is no way to know which cycle preceded or will follow the current cycle. If	Consistence  Cons	73 Overall score See performance > Goals A address > A address Apa > All suggestions Correctness Is alerts Clarity A bit unclear Engagement	
we cannot order the cycles in a sequence, then we can think of them all	• pair - Remove the phrase	Very engaging	
as being parallel to each other. While events within a cycle can have cause and effect relationships (i.e. the events 'happen' at given times), the various cycles themselves do not 'happen', they just exist alongside	another dimension which is dis · Remove wordiness	Delivery Slightly off	
all other cycles. Thus we can think of the annihilation events as being a	• set - Fix the agreement mistake	Style guide	
single event from which infinite Duoverses emerge and to which they return. This implies that finding ourselves in a particular Duoverse is completely onobabilistic where the norbability that we find ourselves in a	• this means - Remove the phrase		
Duoverse with a particular configuration depends on how likely that	This · Rewrite the sentence		
configuration			
is across all possible configurations. This gives us the many worlds that	<ul> <li>be ordered - Rewrite the sentence</li> </ul>		
have been invoked to explain quantum probability in the Everett many			
worlds interpretation of QM.	<ul> <li>sequentially - Rewrite the sentence</li> </ul>		

## Extract from the physics document: edits by Grammarly

### AJE

	Editor	Your document has been modified usi
KVII. THE MANY WORLD	Editor 2	Deleted: WORLDS
The Duoverse described thus far contains all the events in the Universe and anti-Universe for a single	Editor 2	Deleted: which
expansion from beginning to end. However, the <u>Duoverse</u> then re collapses, annihilates, and pair produces a brand new Duoverse. Therefore, we can think of each successive expansion and contraction of the Duoverse	Editor 2	Deleted: Duoverses
as happening along another dimension that is discrete. This dimension essentially labels the different countably infinite random set of duoverses.	Editor 2	Deleted: Duoverse
Since each <u>duoverse</u> begins with annihilation, each <u>duoverse</u> begins with a random configuration after	Editor 2	Deleted: this means
means that the cycles cannot be ordered sequentially because there is no way to know which cycle to cycle. This way to know which cycle preceded	Editor 2	Deleted: Duoverse
being parallel to each other. While events within a cycle can have cause and effect relationships (i.e., the	Editor 2	Deleted:
events <u>nappen</u> at given times), the various cycles themselves do not <u>nappen</u> , they just exist <u>along side</u> all other cycles. Thus, we can think of the annihilation events as being a <i>single</i> event from which infinite <u>duovers</u>	Editor 2	Deleted:
emerge and to which they return. This implies that finding ourselves in a particular Duoverse is completely probabilistic, where the probability that we find ourselves in a Duoverse with a particular configuration depends	Editor 2	Deleted: Duoverses
on how likely that configuration is across all possible configurations. This gives us the many worlds that have been invoked to explain quantum	Editor 2	Deleted: folow
probability in the Everett many worlds interpretation of QM.	Editor 2	Deleted: 'happen'
		··· · · · · · · · · · · · · · · · · ·

Extract from the physics document: edits by AJE

#### InstaText

XVII. THE MANY WORLD	
The Dupueron departiesd thus for contains all the quarte in the Universe and anti-Universe for a single	
The binderse described that is an contains an use events in the converse and altu-converse for a single	100
brand new Duoverse. Therefore, we can think of each successive expansion and contraction of the Duoverse	and the second
as happening along another dimension that is discrete. This dimension essentially labels the different	
countably infinite random set of duoverses.	
Since each duoverse begins with annihilation, each duoverse begins with a random configuration after	
annihilation. Therefore, there is no cause-and-effect relationship between duovers from cycle to cycle. This	Sec. 22.
means that the cycles cannot be ordered sequentially because there is no way to know which cycle preceded	Sec.
or will follow the current cycle. If we cannot order the cycles in a sequence, then we can think of them all as	
being parallel to each other. While events within a cycle can have cause and effect relationships (i.e., the	N. N.N.
events nappen at given times), the various cycles themselves do not nappen, they just exist along side all	$( \langle \cdot \rangle )$
order cycles. Thus, we can mink of the annihilation events as being a single event from which minite <u>utovers</u> is completely.	1. 16
emerge and to which they retain. This implies that infinity outserves in a particular configuration depended	1.22
probabilisely that configuration	N.N.
is across all possible configurations. This gives us the many worlds that have been invoked to explain quantum	

Deleted: WORLDS Deleted: which Deleted: Duoverses Deleted: Duoverse Deleted: this means Deleted: Duoverse Deleted: Deleted: Deleted: Duoverses Deleted: folow Deleted: 'happen'

Your document has been modified using

### Writefull

		Writefull	Deleted: Duoverse
¢∨⊪.	THE MANY WORLDS	Writefull	Deleted: re collapses,
Th	e Duoverse described thus far contains all the events in the Universe and anti-Universe for a single	Writefull	Deleted: Duoverse.
expa bran	nsion from beginning to end. However, the <u>duoverse</u> then re-collapses, annihilates, and pair produces a d new <u>duoverse</u> . Therefore, we can think of each successive expansion and contraction of the <u>Duoverse</u>	Writefull	Deleted: dimension
as h coun	appening along another dimension, which is discrete. This dimension essentially labels the different tably infinite random sets of Duoverses.	Writefull	Deleted: set
Si	nce each Duoverse begins with annihilation, this means that each Duoverse begins with a random	Writefull	Deleted: means
cycle	to cycle. This means the cycles cannot be ordered sequentially because there is no way to know which	Writefull	Deleted: cause and effect
of th	em all as being parallel to each other. Although events within a cycle can have cause-and-effect	Writefull	Deleted: folow
exist	along with all other cycles. Thus, we can think of the annihilation events as being a single event from	Writefull	Deleted: While
Which Duoy	n infinite Duoverses emerge and to which they return. This implies that finding ourselves in a particular erse is completely probabilistic where the probability that we find ourselves in a Duoverse with a particular	Writefull	Deleted: cause and effect
confi is in a	guration depends on how likely that configuration is. all possible configurations. This gives us many worlds that have been invoked to explain quantum probability	Writefull	Deleted: (i.e.
in the	Everett many worlds interpretation of QM.	Writefull	Deleted: side
	AVA A A		

Extract from the physics document: edits by Writefull

Extract from the physics document: edits by InstaText

28

Editor Editor 2

Editor 2 Editor 2

Editor 2

Editor 2

Editor 2

Editor 2

Editor 2

Editor 2

Editor 2

Editor 2

----

## QuillBot

= 🖸 QuillBot	PREMIUM	Paraphraser		🟶 🕻 т
Paraphraser			Basharase to come we the	
Grammar Checker		Modes: Standard Fluency Formal Simple Creative Expand Shorten	ence in different modes	103
Plagiarism Checker	NEW		Rephrase 5 🗅 🏳	*
Co Miritor	_	XVII. THE MANY WORLDS	THE MANY WORLDS (XVII)	
Co-writer		The Duoverse described thus far contains all the events in the Universe and	All of the occurrences in the Universe and anti-Universe for a single expansion are	11.
😑 Summarizer		anti-Universe for a single expansion from beginning to end. However, the	anti-Universe for a single expansion from beginning to end. However, the contained in the Duoverse described thus far. The Duoverse, on the other hand,	
Citation Generator		Duoverse then re collapses, annihilates, and pair produces a brand new collapses, annihilates, and pairs to create a new Duoverse. As a result, each		
-		Duoverse. Therefore, we can think of each successive expansion and	subsequent expansion and contraction of the Duoverse can be thought of as	
Orrome Extension		contraction of the Duoverse as happening along another dimension which is	occurring along a separate dimension. The separate countably infinite random set	
Docs Extension		discrete. This dimension essentially labels the different countably infinite	of Duoverses is labelled by this dimension.	
Word Extension	12	random set of Duoverses.	Because each Duoverse begins with destruction, after annihilation, each Duoverse	
		Since each Duoverse begins with annihilation, this means each Duoverse	begins with a random configuration. As a result, from cycle to cycle, there is no	
⑦ Help Center		begins with a random configuration after annihilation. Therefore, there is no	cause and effect link between Duoverses. This means that the cycles can't be	
Contact Us		cause and effect relationship between Duoverses from cycle to cycle. This	sorted in any particular sequence because there's no way of knowing which cycle	
		means the cycles cannot be ordered sequentially because there is no way to	came before or will come after the present one. We can think of the cycles as being	
		know which cycle preceded or will folow the current cycle. If we cannot order	parallel to each other if we can't put them in a sequence. While events inside a	
		the cycles in a sequence, then we can think of them all as being parallel to	cycle can have cause and effect links (i.e., events 'happen' at specific periods), the	
		each other. While events within a cycle can have cause and effect	various cycles do not 'happen,' they just exist alongside other cycles. As a result,	
		relationships (i.e. the events 'happen' at given times), the various cycles	we can consider annihilation events to be a single event from which infinite	
		themselves do not 'happen', they just exist along side all other cycles. Thus we	Duoverses emerge and return. This means that our chances of landing in a specific	
		can think of the annihilation events as being a single event from which infinite	Duoverse are entirely probabilistic, with the likelihood of landing in a Duoverse	
		Duoverses emerge and to which they return. This implies that finding	with a specific configuration based on how likely that configuration is among all	
		ourselves in a particular Duoverse is completely probabilistic where the	potential configurations. In the Everett many worlds interpretation of QM, this	
		probability that we find ourselves in a Duoverse with a particular	gives us the many worlds that have been used to explain quantum probability.	
		configuration depends on how likely that configuration		
		is across all possible configurations. This gives us the many worlds that have		
		been invoked to explain quantum probability in the Everett many worlds		
		interpretation of QM.		
		265 Words Rephrase	^ ∨ 1/13 Sentences • 262 Words 🗉 👌 🗋	

Extract from the physics document: edits by QuillBot

### Trinka

		Trinka	Deleted: pair
	VII. THE MANY WORLDS	Trinka	Deleted: dimension which is discrete
	The Duoverse described thus far contains all the events in the Universe and anti-Universe for a single	Trinka	Deleted: set
1	a brand new Duoverse. Therefore, we can think of each successive expansion and contraction of the Duoverse	Trinka	Deleted: each
	infinite random sets of Duoverses.	Trinka	Deleted: folow
1	configuration after annihilation. Therefore, there is no cause and effect relationship between Duoverses from	Trinka	Deleted: then
1	cycle to cycle. This means the cycles cannot be ordered sequentially because there is no way to know which cycle preceded or will follow the current cycle. If we cannot order the cycles in a sequence, we can think of	Trinka	Deleted: do
1	them all as being parallel to each other. While events within a cycle can have cause and effect relationships (i.e. the events 'happen' at given times), the various cycles themselves are not happen', they just exist along	Trinka	Deleted: not '
	side all other cycles. Thus, we can think of the annihilation events as being a <i>single</i> event from which infinite Duoverses emerge and to which they return. This implies that finding ourselves in a particular Duoverse is	Trinka	Deleted: we
	completely probabilistic, where the probability that we find ourselves in a Duoverse with a particular configuration depends on how likely that configuration	Trinka	Deleted: where
	Is across all possible configurations. This gives us the many worlds invoked to explain quantum probability in the Everett many worlds interpretation of OM	Trinka	Deleted: is
1	are Everen many wonds interpretation of give.	Trinks	Deleted- that have been

Extract from the physics document: edits by Trinka