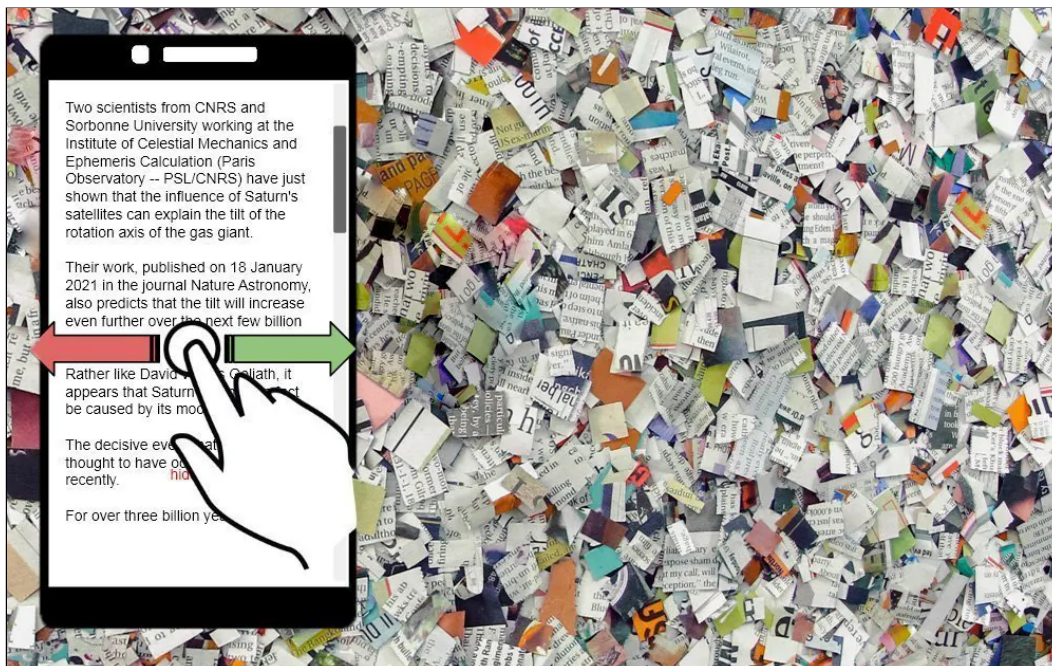


A Machine Learning System to Rewrite an Article While You Read It

Originally published at <https://www.unite.ai/a-machine-learning-system-to-rewrite-an-article-while-you-read-it/>



New research from Canada proposes a method to automatically rewrite an article as you read it, based on Tinder-style 'swiping', or on passive observation of the reader's interaction with the various kinds of content that the article contains.

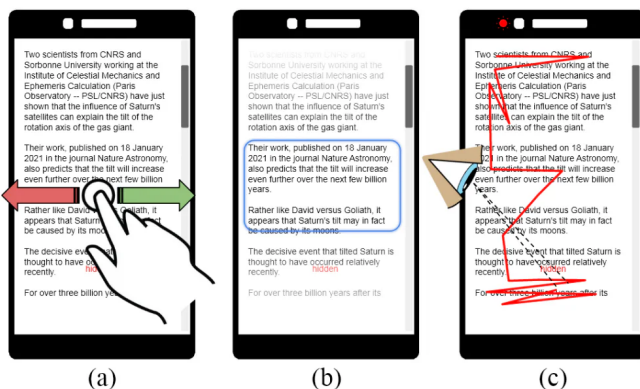
The system, titled Hone As You Read (HARE), is presented in a [paper](#) from Western University at Ontario, Canada, with corresponding Python code [at GitHub](#).

The central idea of the project is that an article may contain various kinds of content, evolving (much like this one) from the headline down to further details. Later parts of an article may contain different kinds of supporting material, use cases, or hypotheses or conjecture about the ramifications of the news.

Under HARE, if you don't like that kind of material, you can vote it away on a paragraph-by-paragraph basis while the system learns your preferences, so that by the time you scroll down, content similar to the material that you 'downvoted' has already been removed or rewritten. If you don't want to actively participate in training the system, HARE can deduce your choices by observing your passive interactions with the document.

Tinder-Style Voting For Displeasing Sentences

In the image below, we see three possible types of inferred categorization for HARE, based on the user's explicit or implicit behavior. In the first case (left), the user actively 'swipes left' (or right), in a Tinder-style voting gesture expressing approbation or displeasure in the content of the paragraph or sentence, or in its style, complexity or tone.

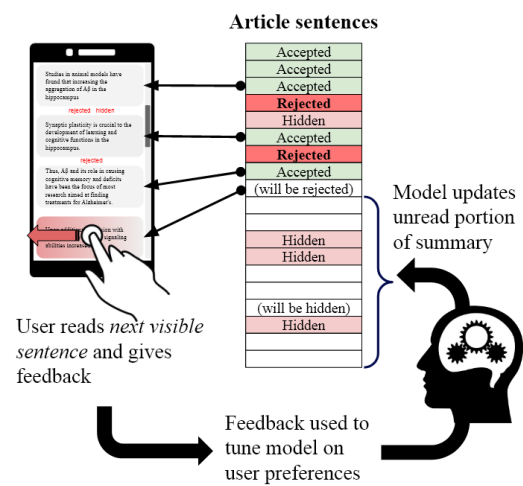


Source: <https://arxiv.org/pdf/2105.02923.pdf>

In the second case (center), the system uses dwell-time as a metric of user interest, based on the positioning and duration of scroll pause.

In the third case (right), HARE is using the smartphone camera to estimate the path and dwell-time of the viewer’s gaze location across the paragraphs of the visible documents.

The researchers contend that increased dwell time on any one paragraph can indicate increased user interest, though logically this may not be the case where the viewer is trying to assimilate text that may be complicated or just poorly-written.



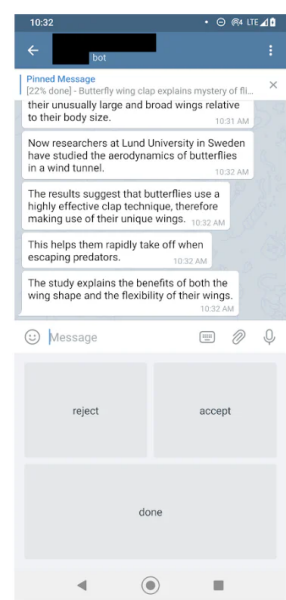
User feedback effectively edits, rewrites or entirely erases as-yet unseen portions of the article.

Pre-Processing Content To The User’s Preferences

The paper deals with the user experience of HARE on a per-article basis, but clearly the user’s historical interaction with documents allows for the customization of future reading experiences, by consistently recognizing types of content and applying templated user preferences to new articles, so that the need for interaction diminishes as the user sees less and less ‘unwanted’ content.

HARE is characterized as a summarizer algorithm, allowing for unseen content further down the page to be rewritten in terms of style or concision before the user arrives at it; but the paper makes clear that it can also preemptively remove content based on user feedback.

For testing purposes, the system utilized a [corpus](#) of 11,222 articles from the UK’s *Daily Mail* newspaper, and was evaluated via a test deployment on the Telegram chat app. Articles with less than ten paragraphs were discarded for trial purposes.



The Telegram HARE app in a test phase with users.

The researchers’ methodology uses [K-Means clustering](#) on [SBERT](#) sentence embeddings in the articles, with initially random weights for concepts dealt with.

Among a wide group of algorithms and approaches, HARE features three comparison models, the first of which (ORACLEGREEDY) has access to prior user preferences, indicating the intent that the algorithm could pre-process articles on load, rather than interactively.

The other models, ORACLESORTED and ORACLEUNIFORM, selects sentences based on interest level or randomly throughout the article, respectively.

Content Removal And Rewriting

Surprisingly, ORACLEUNIFORM outperformed the control set, even though it has no access to prior user interests. The researchers contend that this is because it deals with the entire article in one sweep, 'choosing only the most interesting sentences'. The researchers admit that this may restrict the available content to those sentences that deal solely with the most important concept, logically removing other text that may deal with ramifications or evaluation of the concept.

The extractive summarizers used in HARE are LexRank, [SumBasic](#), and [TextRank](#).

HARE was tested on 13 volunteers over the course of 70 trials and varying algorithmic approaches, and was able to update summaries (rewritten/excised text) at somewhere between 1.3 milliseconds and 100ms on a consumer grade laptop, depending on the model being trialed. Results found that the models which removed most text did not perform well, mainly because this can affect the coherence of the remaining text.

Ethical Implications Of Dynamic Article Rewriting

The researchers acknowledge ethical concerns around technologies of this nature:

'The HARE task is intended for the design of future user-facing applications. By design, these applications have the ability to control what a user reads from a given article. It is possible that, when deployed without sufficient care, these tools could exacerbate the "echochamber" effect already produced by automated news feeds, search results, and online communities.'

However, they also note that such a system could be used in future applications to mitigate the echo chamber effect by injecting text that proposes alternative viewpoints that may not have been initially present in the article. They observe: 'The weighting of this factor could be tuned to provide both an engaging reading experience and exposure to a diversity of ideas.'

Those likely to benefit from such a system, according to the researchers, are readers that want to save time in taking in information, and content publishers.

Writer on machine learning, domain specialist in human image synthesis. Former head of research content at Metaphysic.ai, until its dissolution into DNEG's Brahma.ai.

Portfolio site: martinanderson.ai

Contact: martin@martinanderson.ai



Discover More