

Title: Inauthentic text

URL: [https://en.wikipedia.org/wiki/Inauthentic\\_text](https://en.wikipedia.org/wiki/Inauthentic_text)

PageID: 5008963

Categories: Category:Fraud, Category:Machine learning, Category:Machine learning stubs, Category:Scientific misconduct, Category:Semantics, Category:Semantics stubs

Source: Wikipedia (CC BY-SA 4.0).

-----

An inauthentic text is a computer-generated expository document meant to appear as genuine, but which is actually meaningless. Frequently they are created in order to be intermixed with genuine documents and thus manipulate the results of search engines, as with Spam blogs . They are also carried along in email in order to fool spam filters by giving the spam the superficial characteristics of legitimate text.

Sometimes nonsensical documents are created with computer assistance for humorous effect, as with Dissociated press or Flarf poetry . They have also been used to challenge the veracity of a publication— MIT students submitted papers generated by a computer program called SCIfgen to a conference, where they were initially accepted. This led the students to claim that the bar for submissions was too low.

With the amount of computer generated text outpacing the ability of people to humans to curate it, there needs some means of distinguishing between the two. Yet automated approaches to determining absolutely whether a text is authentic or not face intrinsic challenges of semantics. Noam Chomsky coined the phrase " Colorless green ideas sleep furiously " giving an example of grammatically-correct, but semantically incoherent sentence; some will point out that in certain contexts one could give this sentence (or any phrase) meaning.

The first group to use the expression in this regard can be found below from Indiana University . Their work explains in detail an attempt to detect inauthentic texts and identify pernicious problems of inauthentic texts in cyberspace. The site has a means of submitting text that assesses, based on supervised learning, whether a corpus is inauthentic or not. Many users have submitted incorrect types of data and have correspondingly commented on the scores. This application is meant for a specific kind of data; therefore, submitting, say, an email, will not return a meaningful score.

See also

Scraper site

Spamdexing

Stochastic parrot

External links

An Inauthentic Paper Detector from Indiana University School of Informatics

This semantics article is a stub . You can help Wikipedia by expanding it .

v

t

e

This machine learning -related article is a stub . You can help Wikipedia by expanding it .

v

t

e