

Publishers withdraw more than 120 gibberish papers

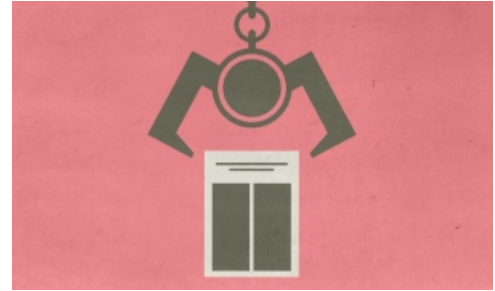
Conference proceedings removed from subscription databases after scientist reveals that they were computer-generated.

Richard Van Noorden

24 February 2014 Updated: 25 February 2014

The publishers Springer and IEEE are removing more than 120 papers from their subscription services after a French researcher discovered that the works were computer-generated nonsense.

Over the past two years, computer scientist Cyril Labbé of Joseph Fourier University in Grenoble, France, has catalogued computer-generated papers that made it into more than 30 published conference proceedings between 2008 and 2013. Sixteen appeared in publications by Springer, which is headquartered in Heidelberg, Germany, and more than 100 were published by the Institute of Electrical and Electronic Engineers (IEEE), based in New York. Both publishers, which were privately informed by Labbé, say that they are now removing the papers.



Among the works were, for example, a paper published as a proceeding from the 2013 International Conference on Quality, Reliability, Risk, Maintenance, and Safety Engineering, held in Chengdu, China. (The conference website says that all manuscripts are “reviewed for merits and contents”.) The authors of the paper, entitled ‘TIC: a methodology for the construction of e-commerce’, write in the abstract that they “concentrate our efforts on disproving that spreadsheets can be made knowledge-based, empathic, and compact”. (*Nature News* has attempted to contact the conference organizers and named authors of the paper but received no reply*; however at least some of the names belong to real people. The IEEE has now removed the paper).

**Update: One of the named authors replied to Nature News on 25 February. He said that he first learned of the article when conference organizers notified his university in December 2013; and that he does not know why he was a listed co-author on the paper. “The matter is being looked into by the related investigators,” he said.*

How to create a nonsense paper

Labbé developed a way to automatically detect manuscripts composed by a piece of software called SCIGen, which randomly combines strings of words to produce fake computer-science papers. SCIGen was invented in 2005 by researchers at the Massachusetts Institute of Technology (MIT) in Cambridge to prove that conferences would accept meaningless papers — and, as they put it, “to maximize amusement” (see ‘Computer conference welcomes gobbledegook paper’). A related program generates random physics manuscript titles on the satirical website arXiv vs. snarXiv. SCIGen is free to download and use, and it is unclear how many people have done so, or for what purposes. SCIGen’s output has occasionally popped up at conferences, when researchers have submitted nonsense papers and then revealed the trick.

Labbé does not know why the papers were submitted — or even if the authors were aware of them. Most of the conferences took place in China, and most of the fake papers have authors with Chinese affiliations. Labbé has emailed editors and authors named in many of the papers and related conferences but received scant replies; one editor said that he did not work as a program chair at a particular conference, even though he was named as doing so, and another author claimed his paper was

submitted on purpose to test out a conference, but did not respond on follow-up. *Nature* has not heard anything from a few enquiries.

“I wasn’t aware of the scale of the problem, but I knew it definitely happens. We do get occasional e-mails from good citizens letting us know where SCIdgen papers show up,” says Jeremy Stribling, who co-wrote SCIdgen when he was at MIT and now works at VMware, a software company in Palo Alto, California.

“The papers are quite easy to spot,” says Labbé, who has built a website where users can test whether papers have been created using SCIdgen. His detection technique, described in a study¹ published in *Scientometrics* in 2012, involves searching for characteristic vocabulary generated by SCIdgen. Shortly before that paper was published, Labbé informed the IEEE of 85 fake papers he had found. Monika Stickel, director of corporate communications at IEEE, says that the publisher “took immediate action to remove the papers” and “refined our processes to prevent papers not meeting our standards from being published in the future”. In December 2013, Labbé informed the IEEE of another batch of apparent SCIdgen articles he had found. Last week, those were also taken down, but the web pages for the removed articles give no explanation for their absence.

Related stories

- Investigating journals: The dark side of publishing
- Editor will quit over hoax paper
- Computer conference welcomes gobbledegook paper

[More related stories](#)

Ruth Francis, UK head of communications at Springer, says that the company has contacted editors, and is trying to contact authors, about the issues surrounding the articles that are coming down. The relevant conference proceedings were peer reviewed, she confirms — making it more mystifying that the papers were accepted.

The IEEE would not say, however, whether it had contacted the authors or editors of the suspected SCIdgen papers, or whether submissions for the relevant conferences were supposed to be peer reviewed. “We continue to follow strict governance guidelines for evaluating IEEE conferences and publications,” Stickel said.

A long history of fakes

Labbé is no stranger to fake studies. In April 2010, he used SCIdgen to generate 102 fake papers by a fictional author called Ike Antkare [see pdf]. Labbé showed how easy it was to add these fake papers to the Google Scholar database, boosting Ike Antkare’s *h*-index, a measure of published output, to 94 — at the time, making Antkare the world’s 21st most highly cited scientist. Last year, researchers at the University of Granada, Spain, added to Labbé’s work, boosting their own citation scores in Google Scholar by uploading six fake papers with long lists to their own previous work².

Labbé says that the latest discovery is merely one symptom of a “spamming war started at the heart of science” in which researchers feel pressured to rush out papers to publish as much as possible.

There is a long history of journalists and researchers getting spoof papers accepted in conferences or by journals to reveal weaknesses in academic quality controls — from a fake paper published by physicist Alan Sokal of New York University in the journal *Social Text* in 1996, to a sting operation by US reporter John Bohannon published in *Science* in 2013, in which he got more than 150 open-access journals to accept a deliberately flawed study for publication.

Labbé emphasizes that the nonsense computer science papers all appeared in subscription offerings. In his view, there is little evidence that open-access publishers — which charge fees to publish manuscripts — necessarily have less stringent peer review than subscription publishers.

Labbé adds that the nonsense papers were easy to detect using his tools, much like the plagiarism checkers that many

publishers already employ. But because he could not automatically download all papers from the subscription databases, he cannot be sure that he has spotted every SCIdgen-generated paper.

Nature doi:10.1038/nature.2014.14763

Updates

Updated: This article was updated on 25 February to include the response from one of the named authors of the nonsense paper.

References

1. Labbé, C. & Labbé, D. *Scientometrics* **94**, 379–396 (2013).

Show context

Article ISI

2. López-Cózar, E. D., Robinson-García, N. & Torres-Salinas, D. J. *Assoc. Inform. Sci. Technol.* **65**, 446–454 (2014).

Show context

Related stories and links

From nature.com

- **Investigating journals: The dark side of publishing**
27 March 2013
- **Editor will quit over hoax paper**
15 June 2009
- **Computer conference welcomes gobbledegook paper**
20 April 2005

From elsewhere

- **Cyril Labbe's homepage**
- **SCIdgen: an automatic paper generator**
- **John Bohannon's sting of open-access journals, for *Science* magazine**

For the best commenting experience, please login or register as a user and agree to our Community Guidelines. You will be re-directed back to this page where you will see comments updating in real-time and have the ability to recommend comments to other users.

Comments for this thread are now closed.

26 comments

[Subscribe to comments](#)



Jeff Smith • 2014-03-09 01:50 AM

Even the non-fakes are not gospel. Most scientific papers have to be corrected. Good thing for economics it's not a science. But in which half did that study appear? More at progress.org.



Ronald Rousseau • 2014-03-02 08:48 AM

Why has it taken so long before Springer and IEEE reacted? The paper by Cyril and Dominique Labbé was published online by Scientometrics (a Springer journal) on 22 June 2012 and officially published in January 2013. Moreover, Labbé's Ike Antkare story (pointing to a similar problem) was published in the ISSI Newsletter, 6(2), 48-52, 2010.



Richard Van Noorden • 2014-03-02 05:35 PM

Cyril Labbe's paper identified only a first batch of 85 fakes that he told IEEE about, before he published. The IEEE promptly removed this first batch, they told me. (They did not publicise this at the time).



Rebecca sanger • 2014-03-01 06:45 PM

Ed, I'm sad to inform you that the USA does not set the world standard for peer review. Here is one glaring example of the complete failure of peer review in a U.S. journal. In 2001 the Journal of Reproductive Medicine published a highly flawed and almost certainly fraudulent paper. The study claimed that Christian prayers from the USA caused a 100% increase in the success rate of in-vitro fertilization (IVF) treatments performed on women in South Korea. The author who designed the study, Daniel Wirth, subsequently went to federal prison on criminal fraud charges unrelated to the study. The "lead" author, Rogerio Lobo, eventually admitted that he did not even know about the study until months after its publication. The third author, Kwang Cha, left Columbia University soon after the bizarre study was published. The Journal of Reproductive Medicine has never retracted this study and the published results have been frequently cited as confirmation of the power of faith healing.



Gary A Doss • 2014-02-28 07:00 AM

What is the motive behind publishing fake papers?



Mehmet Dalkilic • 2014-03-01 06:30 AM

I think, originally, there arose questions about the thoroughness of refereeing -- this then devolved into venues that had 100% acceptance, but required money. My sense now is that the submissions are exceeding the abilities of venues to review them completely. So, unscrupulous researchers can take advantage of this and show they are productive. As I mentioned earlier, I think this is symptomatic of a system of academic merit that is fundamentally faulty -- quality in no way is made up for with quantity. I also believe anonymous reviewing contributes to this overall problem, since there's no reasonable way for me, as someone reviewed, to know whether someone has actually reviewed me.



Daniel Renjewski • 2014-02-27 04:41 PM

Scientists are obviously well advised, to choose the conferences they are attending carefully. It is unfortunate that even IEEE faces some challenges to ensure their standards. I know of one case where I was listed as a conference editor on a

official conference website without having heard of this conference ever. It seems unlikely though that this happens to a conference program chair.



Neil Irving Solomon • 2014-02-27 02:38 PM

Hi, Richard -- sorry, my mistake: Ruth Francis is, indeed, Springer's UK head of communications.



George McNamara • 2014-02-27 02:13 PM

Hi Richard, have you asked the publishers for the peer reviewer and editors comments? Maybe SCIgen wrote the reviews too.



Richard Van Noorden • 2014-03-01 02:23 AM

Well, 'peer review' for conference papers does not necessarily involve actually writing reviews - even for papers that are legitimate and not computer-generated.



Mehmet Dalkilic • 2014-02-26 03:07 PM

For those interested in this topic, here's a paper we wrote on using compression to detect computer generated texts (2006) <http://www.siam.org/meetings/sdm06/proceedings/070dalkilicm.pdf> and a modest service: <http://montana.informatics.indiana.edu/cgi-bin/fsi/fsi.cgi>



Richard Van Noorden • 2014-02-27 09:30 AM

Thanks Mehmet. I didn't know about this (though Cyril Labbé notes that he cited your paper in his Scientometrics article).



Mehmet Dalkilic • 2014-02-27 05:29 PM

You are correct! Aside from the obviously interesting problems computer generated (inauthentic) text pose, I was surprised at both the amount and range of topics -- we had many people send us generators that, I am loathe to admit, produced what appear cursorily as authentic artifacts. My opinion is that these fake paper generators are a symptom of a wider problem of productivity as function of quantity, rather than quality.



Sylvia Wenmackers • 2014-02-26 12:02 PM

I wonder whether the submissions will turn out to be related to spamferences (also called scamferences) and offer some reflections in this blogpost: <http://www.newappsblog.com/2014/02/recently-discovered-100-computer-generated-papers-relation-to-spamference.html>



David Crotty • 2014-02-25 08:47 PM

Richard, do you have any sense of the review process seen by these "conference proceedings" papers and if/how it differs from the review given a typical research paper?



Richard Van Noorden • 2014-02-25 09:51 PM

Good question, and the answer is no, I don't. If any computer scientists would like to weigh in, I'm all ears. (Presumably, there are different types of conferences with different peer review processes, just as there are with journals - does it make sense to speak of a 'typical' comparison?) Some people who sound more familiar with computer science conferences posted at Hacker News: <https://news.ycombinator.com/item?id=7294487> .



Nicholas Collin Paul de Gloucester • 2014-03-25 06:20 PM

There had already been a human-generated exposé of an IEEE conference ([HTTP://dieHimmelistschoen.Blogspot.com](http://dieHimmelistschoen.Blogspot.com)) and of a non-IEEE conference (WWW.CG.TUWien.ac.At/~wp/idea.html) before Labbé and Labbé have performed this good recent work. An exposé by me of bad refereeing for journals by the IEEE and other publishers is Paul Colin de Gloucester (2013): "Referees Often Miss Obvious Errors in Computer and Electronic Publications", "Accountability in Research: Policies and Quality Assurance", 20:3, 143-166, WWW.TandFonline.com/doi/abs/10.1080/08989621.2013.788379 This is the first paper to cite the SCIdgen exposé by Labbé and Labbé.



dr2chase • 2014-03-01 06:32 PM

Hi, I'm a computer scientist, some papers in some conferences, sometimes on program committees or reviewing for journals. What is described at ycombinator as CS practice has been true for at least 20 years, if not longer -- conference publication is the default goal, and it is screened pretty well. Reviewing is a PITA, is not directly compensated, and I never know if I am doing a good enough job at it. The CS field is darn wide and it is hard to keep up. On the upside, every once in a while you do come across some new gem (and yes, ethically you must wait till it is published before acting on that knowledge -- unless you can find the tech report for the submission online somewhere).



Lewis Perdue • 2014-02-25 05:37 PM

Not all gibberish papers were written by a computer: New Low-Dose BPA Paper In Toxicological Sciences Is Contaminated By Massive Errors & Should Be Pulled <http://www.nano-active.com/2014/02/new-low-dose-bpa-research-in.html>



Neil Irving Solomon • 2014-02-25 01:17 PM

Thanks for the article. Just a minor error of fact should be corrected: I think you'll find that David Francis, not Ruth Francis (who writes for Nature), is the UK head of communications at Springer



Richard Van Noorden • 2014-02-25 01:38 PM

No, Ruth Francis is currently the Head of Communications, UK, Springer Science & Business Media. (She worked as head of press at Nature Publishing Group until November 2012).



Kevin Sanders • 2014-02-25 12:55 PM

That is true, and I fully appreciate that. I was intending to make the point that subscribers, and at institutional level, service users, are also being 'defrauded'. This appears to be as a result of inadequate procedures in the review process, and has potential implications for the absorption of information into scholarly discourse on top of other issues. I apologise for not being clearer, and thanks for pulling me up!



Richard Van Noorden • 2014-02-25 12:46 PM

To interject: I don't think anyone really knows what the purpose of these SCIdgen papers was. You'd have to check in every single case. It's not even clear who submitted the papers. But it surely cannot have been to show off the weaknesses of conference peer review systems -- or those who had sent in the papers would have revealed the trick (as they have done in other cases of genuine SCIdgen pranking).



Kevin Sanders • 2014-02-25 12:31 PM

That is an interesting point, Ed, although given that the purpose of the exercise was to test the validity of the systems and/or to show their limitations and weaknesses, it seems a little moot. Do you think the institutions that paid expensive subscriptions to titles and/or packages charged by organisations like the ones mentioned should be entitled to compensation on the same grounds? After all, they are the ones paying for a value-added service which should really include weeding out computer generated "gibberish".



Ed Borasky • 2014-02-25 05:45 AM

Where I come from (USA) this is called "fraud" and people can be hauled into a court and tried for it. Maybe it's time international law caught up to our standards.



Ann Sz • 2014-02-26 06:53 AM

Actually, no. In the U.S.A. nobody gets hauled into court for this kind of thing. No governmental body regulates and enforces standards in scientific journals. It is left for the journals themselves to enforce their own standards. The relationship between journal and subscriber is one of trust, and, maybe someone can correct me if I'm wrong, there are no legal grounds for compensation if statements are found to be false or misleading. The penalty for breaking the relationship of trust is solely the broken trust. Now, if a researcher violated the terms of a grant, it's possible there could be civil or criminal sanctions, but it's hard to see what grant a gibberish paper would purport to be funded by, much less violate. Again, the penalty is broken trust - and not getting future grants. Further, in the U.S.A. courts have

ruled that media in general have a first amendment right to knowingly report falsehood as news (*New World Communications of Tampa v. Akre*, 2003) and that falsely attributing words to you that you didn't say is also a protected right (*Masson v. The New Yorker*, 1991). By contrast, some other countries, Canada for instance, have laws against false or misleading news. Countries of the EU seem to have some teeth in their regulation of press and broadcast ethics as well. It is the U.S.A. which need to raise its standards in order to meet those of a modern Western nation.

See other News & Comment articles from *Nature*

Nature ISSN 0028-0836 EISSN 1476-4687

© 2016 Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved.
partner of AGORA, HINARI, OARE, INASP, CrossRef and COUNTER