

# From dating profiles to Brexit—how to spot an online lie

June 29 2016, by Tom Van Laer, City University London

---



Credit: Unsplash/CC0 Public Domain

There are three things you can be sure of in life: death, taxes – and lying. The latter certainly appears to have been borne out by the UK's recent Brexit referendum, with a [number](#) of the [Leave campaign's](#) pledges

[looking more](#) like [porkie pies](#) than solid truths.

But from internet advertising, visa applications and academic articles to political blogs, insurance claims and dating profiles, there are countless places we can tell digital lies. So how can one go about spotting these online fibs? Well, Stephan Ludwig from the University of Westminster, Ko de Ruyter from City University London's Cass Business School, Mike Friedman of the Catholic University of Louvain, and yours truly have developed a digital lie detector – and it can uncover a whole host of internet untruths.

[In our new research](#), we used linguistic cues to compare tens of thousands of emails pre-identified as lies with those known to be truthful. And from this comparison, we developed a text analytic algorithm that can detect deception. It works on three levels.

## **1. Word use**

Keyword searches can be a reasonable approach when dealing with large amounts of digital data. So, we first uncovered differences in word usage between the two document sets. These differences identify text that is likely to contain a lie. We found that individuals who lie generally use fewer personal pronouns, such as I, you, and he/she, and more adjectives, such as brilliant, fearless, and sublime. They also use fewer first-person singular pronouns, such as I, me, mine, with discrepancy words, such as could, should, would, as well as more second-person pronouns (you, your) with achievement words (earn, hero, win).

Fewer personal pronouns indicate an author's attempt to dissociate themselves from their words, while using more adjectives is an attempt to distract from the lie through a flurry of superfluous descriptions. Fewer first-person singular pronouns combined with discrepancy words indicate a lack of subtlety and a positive self-image, while more second-

person pronouns combined with achievement words indicate an attempt to flatter recipients. We therefore included these combinations of search terms in our algorithm.

## 2. Structure scrutiny

Another part of the solution lay in analysing the variance of cognitive process words, such as cause, because, know and ought – and we identified a relationship between structure words and lies.



Credit: AI-generated image ([disclaimer](#))

Liars cannot generate deceptive emails from actual memory so they avoid spontaneity to evade detection. That does not mean that liars use

more cognitive process words overall than people who are telling the truth, but they do include these words more consistently. For example, they tend to connect every sentence to the next – "we know this happened because of this, because this ought to be the case". Our algorithm detects such usage of process words in communications.

### **3. Cross-email approach**

We also studied the ways in which a sender of an email alters their linguistic style while exchanging a number of emails with someone else. This part of the study revealed that as the exchange went on, the more the sender tended to use the function words that the receiver was using.

Function words are [words](#) that contribute to the syntax, or structure, rather than the meaning of a sentence – for example an, am, to. And senders revised the linguistic style of their messages to match that of the receiver. As a consequence, our algorithm identifies and collects such matching.

### **Exciting applications**

Consumer watchdogs can use this technology to assign a "possibly lying" score to advertisements of a dubious nature. Security companies and national border forces can use the algorithm to assess documents, such as visa applications and landing cards, to better monitor compliance with access and entry rules and regulations. Secretaries of higher education exam committees and editors of academic journals can improve their proofing tools for automatically checking student theses and academic articles for plagiarism.

In fact, the potential applications go on and on. Political blogs can successfully monitor their social media interactions for textual

anomalies, while dating and review sites can classify messages submitted by users on the basis of their "possibly lying" score. Insurance companies can make better use of their time and resources available for claim auditing. Accountants, tax advisers, and forensic specialists can investigate financial statements and tax claims and find deceptive smoking guns through our algorithm.

Humans are startlingly bad at consciously detecting deception. Indeed, [human accuracy when it comes to spotting a lie is just 54%](#), hardly better than chance. Our digital lie detector, meanwhile, is 70% accurate. It can be put to work to fight fraud wherever it occurs in computerised content and as the technology evolves, its Pinocchio warnings can be wholly automated and its accuracy will increase even further. Just as Pinocchio's nose reflexively signalled falsehood, so does our digital lie detector. Fibbers beware.

*This article was originally published on [The Conversation](#). Read the [original article](#).*

Source: The Conversation

Citation: From dating profiles to Brexit—how to spot an online lie (2016, June 29) retrieved 11 May 2024 from <https://phys.org/news/2016-06-dating-profiles-brexithow-online.html>

<p>This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.</p>
--