

## *Industry watch*

### *NLP in a post-truth world*

ROBERT DALE

*Arria NLG plc*

*e-mail: Robert.Dale@arria.com*

*(Received 3 January 2017)*

---

#### **Abstract**

We live in a post-truth world. It now matters more whether people think something is true than whether something really is true. This is dangerous, and technology is at least partly to blame. So, as technologists, how can we help to fix this?

---

#### **1 The demise of factuality**

In the mid-1980s, I was working on a research project whose goal was to automate grammar and style checking, with an emphasis on supporting authors who had to write in a manner consistent with a specified house style. One key target group was newspaper and magazine journalists, and so I spent a lot of time talking to people in that profession about the kinds of functionality they'd like to see in editing tools. One wintry London afternoon I visited the offices of the Guardian, and showed a subeditor there a mockup of the application we intended to build. He listened patiently to my pitch as I explained how we'd implement the paper's style guide as a knowledge base in the system. Then he said: 'That's wonderful. But you know, what I really need is a fact checker'.

Fact-checking has always been part of the journalist's job, but never has the journalist needed more help than today. In November 2016, Oxford Dictionaries announced that their Word of the Year was to be 'post-truth'—a term that's been around since at least the early '90s, but whose usage increased significantly in coverage of Brexit in the United Kingdom and Donald Trump's rise to power in the United States.<sup>1</sup>

Oxford defines the term as 'relating to or denoting circumstances in which objective facts are less influential in shaping public opinion than appeals to emotion and personal belief'. A frequently cited example of the phenomenon is Trump's assertion that 81 per cent of murdered white people are killed by black people, when the statistics show that 84 per cent of murdered white people are murdered by white

<sup>1</sup> <https://www.oxforddictionaries.com/press/news/2016/11/17/WOTY-16>

people. During the presidential campaign in the United States, the fact-checking website PolitiFact concluded that 52 per cent of Trump's claims were either false or 'pants on fire' false (that last category being reserved for statements which make a ridiculous claim).

So what's new? Politicians have always lied, you might say. But never so brazenly; and at least in the past they still claimed to hold truthfulness to be important. Today, however, truthfulness has given way to 'truthiness': statements that, to the audience, feel true, even if they're not backed up by fact. And so, in a CNN interview, Newt Gingrich defended Trump's claims that violent crime was on the increase (when the FBI's statistics show a downward trend) by saying that Americans did not feel safer.<sup>2</sup>

Months earlier, a similar dismissal of the importance of truthfulness surfaced in the United Kingdom in the context of the Brexit vote. When the House of Commons Treasury Select Committee interrogated Leave Campaign Director Dominic Cummings about misinformation, his response was that 'accuracy is for snake-oil pussies'.<sup>3</sup>

How did we get here?

## 2 The role of technology in disseminating mistruth

We all have a tendency to believe what we want to believe, but the checks and balances that kept us honest in the past are being eroded. There was a time when information dissemination was primarily the task of the established media, and at least the more respectable publications felt a moral duty to check the accuracy of their reporting. But the established media is in decline, largely because classified advertising, its main source of revenue, has been replaced by online advertising. And just as the Internet has weakened the press, it also offers a young new thing to replace the dying old man: social media increasingly dominates the distribution of news, with nearly two-thirds of adults in the United States now accessing news via social media channels. The numbers are growing exponentially, according to a recent survey by the Pew Research Centre.<sup>4</sup>

But news delivery via social media is insidious. It has been widely observed that the algorithms used by social media sites show you the news they think you want to see, creating echo chambers where your beliefs are reinforced rather than being challenged. And into this world comes the phenomenon of fake news, where the truth of a story doesn't matter. What matters is whether you click on the headline to find out more, since that leads to advertising revenue for the fake news site that hosts the story. And of course social media makes it easy to share the story with

<sup>2</sup> <https://www.youtube.com/watch?v=xnhJWusyj4I>

<sup>3</sup> <https://www.theguardian.com/politics/2016/apr/20/accuracy-is-for-snake-oil-pussies-vote-leaves-campaign-director-defies-mps>

<sup>4</sup> <http://www.journalism.org/2016/05/26/news-use-across-social-media-platforms-2016/>

like-minded individuals, with the result that outrageous claims can, and do, spread like wildfire.<sup>5</sup>

One reaction to this phenomenon has been human fact-checking. According to the Duke Reporters' Lab, the number of active fact-checking websites, like PolitiFact, Full Reality, and FactCheck.org, has increased from 44 in 2014 to 119 today.<sup>6</sup> But these sites rely on human labour. The process is slow and expensive: by the time a human has carried out fact-checking, the erroneous story may have been shared and retweeted many times, and the damage done. Researchers at the College of Warwick and the College of Indiana determined that it can take more than twelve hours for a false claim to be debunked on-line, and even when that debunking is done, its impact is limited. In one instance from the now-defunct rumour detection website Emergent.info, a completely made-up article was shared 60,000 times, whereas its debunking was shared less than 2,000 times.<sup>7</sup>

As Jonathan Swift wrote over 200 years ago: 'Falsehood flies, and the Truth comes limping after it'.<sup>8</sup>

### 3 Technological fixes

The owners of the platforms that have contributed so much to this problem have—some would say rather belatedly—taken some steps to address it. After initially insisting that only a tiny amount of content posted on Facebook is fake news or hoaxes, Mark Zuckerberg has decided to deploy a new anti-clickbait algorithm, and has announced that the company will work with fact-checking organisations to flag fake news stories identified by users.<sup>9</sup> Google has announced a policy update that restricts adverts from being placed on fake news sites, thus diminishing the economic incentive that drives at least some of this content.<sup>10</sup>

To make significant progress in stemming the tide of fake news, however, we have to reduce our dependence on the resource bottleneck of human fact-checking labour. It just takes too long to monitor channels, identify what facts might need to be checked, assess their priority, and then carry out the laborious task of verification. And so inevitably we look for a technological fix: Can we provide multiplicative assistance to human fact-checkers, or even automate the fact-checking process? When my polite Guardian editor asked for this thirty years ago I considered it a pipe dream, but our technology has moved forward since then, and there is now a growing interest in what is becoming known as 'computational fact-checking'.

Much of the work in the area adopts a level of pragmatism that is entirely appropriate given the significance of the problem and the need to do something

<sup>5</sup> For an excellent analysis of the situation, see <http://www.economist.com/news/briefing/21706498-dishonesty-politics-nothing-new-manner-which-some-politicians-now-lie-and>.

<sup>6</sup> <http://reporterslab.org/fact-checkers-reach-keeps-growing/>

<sup>7</sup> <https://firstdraftnews.com/recent-research-reveals-false-rumours-really-do-travel-faster-and-further-than-the-truth/>

<sup>8</sup> <http://quoteinvestigator.com/2014/07/13/truth/>

<sup>9</sup> <http://newsroom.fb.com/news/2016/12/news-feed-fyi-addressing-hoaxes-and-fake-news/>

<sup>10</sup> <https://www.theguardian.com/technology/2016/nov/15/facebook-google-fake-news-sites-ad-networks>

about it now, rather than five or ten years down the road when more exploratory techniques might be ready for prime time. Full Fact, the UK's independent fact-checking charity, provides an excellent review of the state of the art and a roadmap with an emphasis on making progress in the near term: 'This is not the horizon of artificial intelligence; it is simply the application of existing technology to fact-checking.'<sup>11</sup> Full Fact categorises fact-checking technologies into three broad types: reference approaches, which look up a fact in some reference source; machine learning approaches, which attempt to learn signals for likelihood of truth; and contextual approaches, which assess likelihood of truth based on the how long stories survive in the marketplace of ideas. They suggest that the first and third approaches are more likely to show results in the short term, and argue for combining existing tools into a single automated fact-checking workflow that can be used today.

Indeed, it's easy to overlook the value of simple tools: Les Décodeurs, the fact-checking unit at *Le Monde*, has built an easy-to-use search interface that finds previously fact-checked claims.<sup>12</sup> This sounds trivial, but it's an invaluable time saver in a world of scarce fact-checking resources.

There are also a number of recent endeavours that aim to use state-of-the-art NLP technologies in automated fact checking. In particular, text mining techniques can support the assessment of certain kinds of claims that involve named entities and numerical expressions. Décodeurs is working with French data scientists on an automated fact-checker called ContentCheck.<sup>13</sup> 'If someone is searching for fact checks on unemployment, for instance, the tool would automatically extract the latest figures and plot a graph showing whether the indicator is rising or falling.' Similarly, Factmata aims to use numerical relation extraction to identify and check statistical claims like 'The number of unemployed in the US stands at 93 million' or '90% of all merchandise imported into the US is from China'.<sup>14</sup>

Machine learning approaches are also having some success. ClaimBuster, developed by Chengkai Li and colleagues at the University of Texas at Arlington, uses ML to determine the probability that a sentence contains a 'check-worthy claim' based on manually coded examples from past US presidential debates.<sup>15</sup> It also suggests an order of priority for tackling the claims identified. PHEME, an EU project that brings together a number of universities and commercial entities, combines NLP and social network analysis to identify four kinds of false claim in social media and on the web, in real time: rumours, disinformation, misinformation and speculation.<sup>16</sup> And looking further ahead, there is a growing community of researchers working on using structured knowledge networks as resources for fact checking.<sup>17</sup>

<sup>11</sup> <https://fullfact.org/blog/2016/aug/automated-factchecking/>

<sup>12</sup> <http://decodeurs.blog.lemonde.fr/>

<sup>13</sup> <https://team.inria.fr/cedar/contentcheck/>

<sup>14</sup> <https://medium.com/factmata/introducing-factmata-artificial-intelligence-for-political-fact-checking-db8acdbf4cf1.fvgh08vo>

<sup>15</sup> <http://cj2015.brown.columbia.edu/papers/automate-fact-checking.pdf>

<sup>16</sup> <https://www.pHEME.eu/>

<sup>17</sup> See for example <http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0128193sec001>.

These projects all use NLP techniques to determine whether a human-authored story is true. But it's entirely plausible that black hats will also aim to use NLP techniques to *create* believable fake news. It's certainly easy enough for technologies like text mining, document summarisation and natural language generation to make unintended mistakes; but the scope for automatically generated content that has deliberately malicious intent is truly frightening. There's a serious risk that attempts to defeat fake news could be swamped by machine-scale creation of the same, especially if the purveyors of machine-generated fake news actively seek out loopholes in automated fact-checking technology.

To avoid a spiralling arms race akin to what we see in the world of search engine optimisation, one response might be to pursue work that makes content generation technologies more trustworthy. There are already significant pushes in this direction in other areas. DARPA sees Explainable AI (XIA) as essential if future warfighters are to understand, appropriately trust and effectively manage an emerging generation of artificially intelligent machine partners.<sup>18</sup> European Union regulations on algorithmic decision-making and a 'right to explanation' will require machine learning systems to provide evidence for their decisions.<sup>19</sup> These concerns point to a need to build audit-trail mechanisms into our technology, so that machine-generated content comes ready to provide its own justification. If robot journalists start to write stories that are automatically instrumented for fact-checking, we might hope that those sources will become more trusted than those whose response to 'why should I believe you?' is simply 'because I said so'.

In 1942, with the aim of keeping humans safe in the face of increasingly capable machines, Isaac Asimov framed three Laws of Robotics: (1) a robot may not injure a human being or, through inaction, allow a human being to come to harm; (2) a robot must obey the orders given it by human beings except where such orders would conflict with the First Law; and (3) a robot must protect its own existence as long as such protection does not conflict with the First or Second Laws. In today's world of software bots, perhaps it's time to add a Fourth Law: 'A robot must not knowingly create or disseminate misinformation.' Or, in the snappier style of Google's 'Do No Evil', we might say 'Tell No Lies'.

#### 4 What can you do to help?

In the late 1960s, the eminent philosopher Michael Dummett (whose seminal and most important paper, as it happens, was titled 'Truth') famously put his philosophical career on hold for several years to devote himself to fighting racism. As a language technologist, you don't have to go quite so far to make a difference; you have the luxury of being able to make an impact without having to give up your day job. So at the very least, you might think what the Fourth Law of Robotics suggested above means for your own research. And if you're on the lookout for a grand challenge in NLP, fact-checking is certainly a worthy one.

<sup>18</sup> <http://www.darpa.mil/program/explainable-artificial-intelligence>

<sup>19</sup> <https://arxiv.org/pdf/1606.08813v3.pdf>

But if it's beyond your inclination or ability to get so directly involved, there's something very simple you can do: help keep serious journalism alive by subscribing to a quality newspaper. As Joni Mitchell put it, you don't know what you've got till it's gone.