News in an Online World: The Need for an "Automatic Crap Detector"

Yimin Chen, Niall J. Conroy, and Victoria L. Rubin
Language and Information Technology Research Lab (LIT.RL)
Faculty of Information and Media Studies
University of Western Ontario, London, Ontario, CANADA
vrubin@uwo.ca, ychen582@uwo.ca, nconroy1@uwo.ca

ABSTRACT

Widespread adoption of internet technologies has changed the way that news is created and consumed. The current online news environment is one that incentivizes speed and spectacle in reporting, at the cost of fact-checking and verification. The line between user generated content and traditional news has also become increasingly blurred. This poster reviews some of the professional and cultural issues surrounding online news and argues for a two-pronged approach inspired by Hemingway's "automatic crap detector" (Manning, 1965) in order to address these problems: a) proactive public engagement by educators, librarians, and information specialists to promote digital literacy practices; b) the development of automated tools and technologies to assist journalists in vetting, verifying, and fact-checking, and to assist news readers by filtering and flagging dubious information.

Keywords

News verification; deception detection; fake news detection; credibility assessment; journalism practices; online news; media commercialization; natural language processing.

INTRODUCTION

The current state of online information is one of overload – with so many competing sources of varying quality, it becomes difficult for readers to evaluate the credibility and trustworthiness of what they see on in the internet. LIS researchers have been interested in issues of online credibility for quite some time (Fritch & Cromwell, 2001; Rubin & Liddy, 2006), with recent studies noting a difference in how readers perceive User Generated Content (UGC) and Traditional Media Content (TMC) (Rieh, 2014). While research suggests that internet users tend to hold TMC in higher regard than UGC (Rieh, 2014), journalists and editors have called into question whether or not the average internet content consumer can accurately differentiate between the two (Dzieza, 2014). This blurring of lines between TMC and UGC is particularly relevant to news organizations whose credibility has traditionally been linked to their reputation.

The popularity of social media platforms, such as Facebook and Twitter, has created an environment where information is pushed onto consumers either through sponsored messages or via "shares" from friends (Mitchell & Page, 2015). Unless the reader pays special attention, the article becomes decontextualized from its source and fact mixes freely with fiction. From satirical stories to intentional misinformation to plain shoddy reporting, unverified and even outright false news can quickly spread far and wide on the internet.

This paper argues the need for two complementary efforts to counter the spread of rumors, hoaxes, and fake news online: a) an increase in awareness of online misinformation and proactive public engagement by educators, librarians, and information specialists to promote good digital literacy practices; b) the development of automated tools and technologies to assist journalists in vetting, verifying, and fact-checking, and to assist news readers by filtering and flagging dubious information.

NEWS IN CRISIS

For over a decade, journalism has grappled with a cultural shift as more and more of the news moved online (Pavlik, 2000). With the rise of ubiquitous internet access, consumers are now accustomed to having information and news updates on demand. Non-traditional news sources like blogs have grown to cater to these needs (Downie Jr. & Schudson, 2009). However, these sites are generally run by amateurs, hobbyists, and advertisers whose reporting can often be highly opinionated, sensationalized, misinformed, misleading, unverified, or otherwise unreliable (*Ibid.*).

The current state of online media is one that heavily incentivizes speed and spectacle over restraint and verification. The online media economy is largely based around monetization of "views" through advertising revenue. As such, clicks and page views translate directly to a dollar amount – annual digital ad revenue grew to $50.7 billion in the USA this year (Mitchell & Page, 2015). On the other hand, the value of responsible reporting is extremely hard to quantify. Furthermore, news aggregators

ASIST 2015, November 6-10, 2015, St. Louis, MO, USA.

and social media sites, where an increasing number of news consumers encounter stories incidentally (Mitchell & Page, 2014), have further eroded the value on institutional reputation by decontextualizing news from its source (Dzieza, 2014). This removes much of the benefit of building trust, as users arriving at news sites via indirect means show far less loyalty than those who visit directly (Mitchell & Page, 2014). All these factors have created an environment where being first and flashiest is preferable to being accurate, degrading the overall quality of the news (Hallin, 2008, p.55). These trends have caught traditional news organizations in a culture war between old-school journalistic ethics and profitability as they struggle to adapt the dying print model to the internet (Fisher, 2014).

One problematic news practice in particular is that journalists sometimes write about rumors as if they were verified fact (Silverman, 2015 p.8). Two examples of this are headline/body dissonance and "questioning" headlines. The first refers to the use of headlines that introduce an unverified story as if it were fact (see Fig. 1), while the second refers to headlines phrased as a question to lend credence to an extraordinary or unsubstantiated claim (see Fig. 2). Since 55% of readers who click on an article link don't end up reading the actual article (Haile, 2014), these sorts of headlines can be considered misleading "clickbait"[1] and play a significant role in spreading misinformation online (for further discussion of fake news, see Rubin, Chen & Conroy, 2015).

Another issue affecting modern journalism is a controversy over what the role of the journalist should be in the internet age. As a response to competitive pressures, some online publishers and bloggers have suggested that the work of a journalist is to collect and curate items of interest to present to their readers (Fisher, 2014). From this perspective, the job of the journalist is to sift through the immense amount of information flowing through the internet to pick out the interesting bits. Here, the journalist simply "points" at information, without making any truth claims (Silverman, 2015, p.61). On the other hand, what we might call the "legacy" perspective maintains that the value of journalism lies beyond the mere reporting of information – journalists also verify, contextualize, and interpret the information that they report. In this view, journalists play an important role in fostering an informed citizenry, and so, should have a professional obligation to verify that their reporting is accurate and responsible (Downie Jr. & Schudson, 2009).

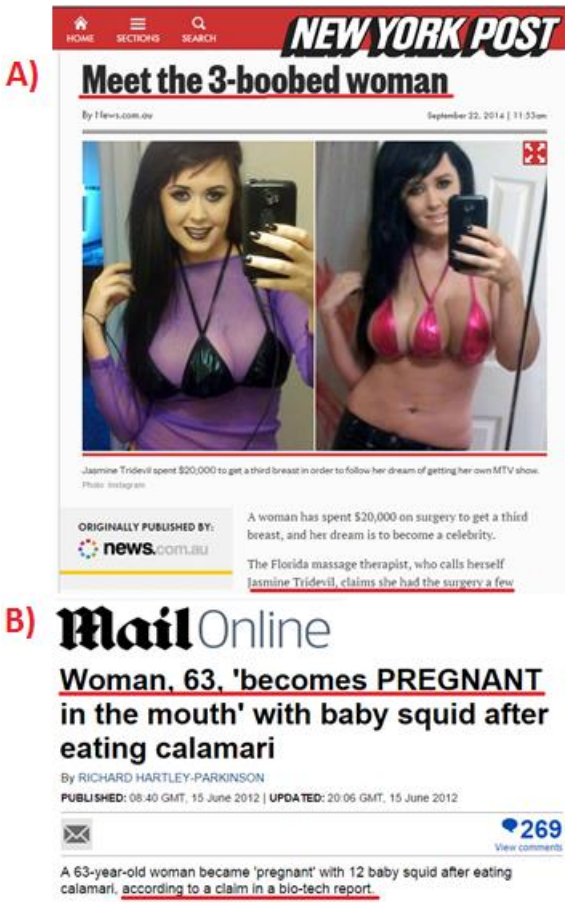

Figure 1 – Headline/Body Dissonance
Both headlines present the story as fact, while the article bodies simply report (unverified) "claims."

Unfortunately, it appears that the "pointing" viewpoint is winning out over the traditional one. "The blogosphere and older media have become increasingly symbiotic" (Downie Jr. & Schudson, 2009), but this symbiosis is not necessarily a positive one. Social media can be a valuable source of breaking news and eyewitness accounts, but are also full of unverified reports and even fake news. Rather than verify their sources, reporters often rush these stories to print and hand off fact-checking responsibilities to sites like Snopes and PolitiFact. "Too good to check" used to be a warning to newspaper editors not to jump on bullshit stories. Now it's a business model" (Chittum, 2013). Even if new information contradicts or invalidates a previous report, the correction is rarely given its own article. Instead, an update or correction may be appended to the original article, but the incorrect headline is often kept in its original form (Silverman, 2015, p.107). Furthermore, there is a pronounced overlap between digital news and public relations. This practice of "native advertising," where content is paid for by commercial advertisers, but written by staff, is often not clearly differentiated from a site's regular content (Mitchell & Page, 2014).

[1] "Clickbait, [mass noun] informal: (On the Internet) content whose main purpose is to attract attention and encourage visitors to click on a link to a particular web page" http://www.oxforddictionaries.com/definition/english/clickbait

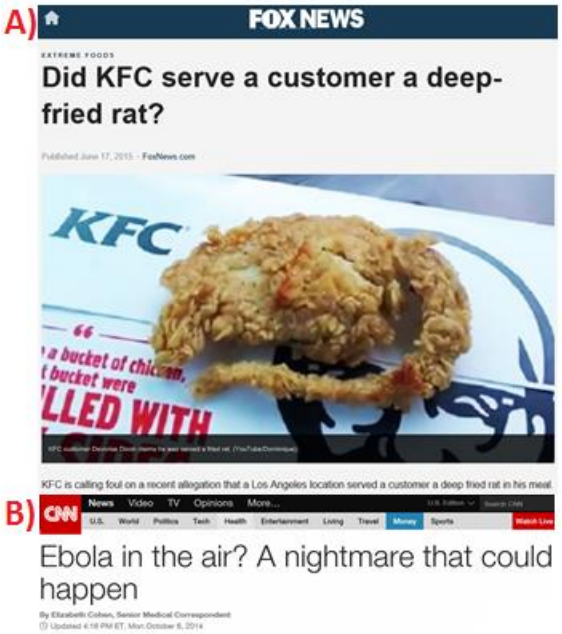

Figure 2 – Questioning Headlines
Using a question as a headline creates the impression that the claim is credible, though it may not be true.

THE NEED FOR DIGITAL LITERACY

Ernest Hemingway once said that "every man should have a built-in automatic crap detector operating inside him" (Manning, 1965). In our digital age of shams, scams, and spam, the ability to critically judge and evaluate the quality of information is an essential skill, but one that is lacking in a large segment of the population. 49% of adult Canadians, for example, are considered to have low literacy skills (Hango, 2014) and even "digital natives" show significant gaps in their digital literacy (Ng, 2012). Clearly, we shouldn't believe all the news that we read online, but how do we decide whether something is credible or not? Schools and libraries are uniquely situated to provide digital literacy initiatives to the public, but face numerous challenges when it comes to funding and expertise. While 100% of American libraries now offer public internet access, only 49% offer formal technology training classes (iPAC, 2014). There is both a need and an opportunity for digital literacy instruction, but a lack of resources and support.

THE NEED FOR AUTOMATION

What seems clear at this point is that many voices within the field of journalism are critical of the current state of news media and would like to see a greater effort towards fact-checking and verification. It is also clear that economic factors do not reward time spent on verification over speed and spectacle. The desire is there, but not the time; these circumstances present an ideal situation in which to apply an automated fact-checking and verification process. Such an automated news verification system based on natural language processing techniques (Rubin, Conroy & Chen, 2015; Zhou et al., 2015) could be used to quickly check for common characteristics of false or misleading information to assist journalists, editors, and other content creators in ensuring that their reporting is accurate and responsibly presented (for further discussion of automation methods, see Conroy, Rubin & Chen, 2015). This tool could also help educators and information specialists teach critical content evaluation skills and reduce information overload for news consumers by filtering and flagging suspicious stories.

CONCLUSION

As the internet continues to break down barriers to content creation and dissemination, the line between user generated content and traditional media content has become increasingly blurred. When information is decontextualized from its source, alternative methods are required to verify and judge the trustworthiness of news reporting. This paper argues the need for an automated assistive tool to aid both content creators and consumers in evaluating the credibility of online news. We also advocate for an increased awareness of potential misinformation in digital online environments and public engagement by educators, librarians, and information specialists to promote good digital literacy practices.

ACKNOWLEDGMENTS

This research has been funded by the Government of Canada Social Sciences and Humanities Research Council (SSHRC) Insight Grant (#435-2015-0065) awarded to Dr. Rubin for the project entitled *Digital Deception Detection: Identifying Deliberate Misinformation in Online News.*

REFERENCES

Chittum, R. (2013, Dec 6). Audit Notes: BS as business model, Larry Kudlow, Third Way revealed. *Columbia Review of Journalism*. http://www.cjr.org/the_audit/audit_notes_bs_as_business_mod.php

Conroy, N. J., Chen, Y., & Rubin, V. L. (2015). Automatic Deception Detection: Methods for Finding Fake News. In *The Proceedings of the Association for Information Science and Technology Annual Meeting (ASIST2015)*, Nov. 6-10, St. Louis.

Downie Jr., L & Schudson, M. (2009, Oct 19). The Reconstruction of American Journalism. *Columbia Review of Journalism*. http://www.cjr.org/reconstruction/the_reconstruction_of_american.php

DiFonzo, N. (2008, Oct 20). *To Rumor Is Human*. *Psychology Today*. www.psychologytoday.com/blog/around-the-watercooler/200810/rumor-is-human

Dzieza, J. (2014, Oct 22). Fake News Sites Are Using Facebook to Spread Ebola Panic. *The Verge*. www.theverge.com/2014/10/22/7028983/fake-news-sites-are-using-facebook-to-spread-ebola-panic

Fisher, M. (2014, Mar 3). Who Cares if it's True? *Columbia Review of Journalism*. http://www.cjr.org/cover_story/who_cares_if_its_true.php

Fritch, J. W., & Cromwell, R. L. (2001). Evaluating Internet resources: Identity, affiliation, and cognitive authority in a networked world. *Journal of the American Society for Information science and Technology*, *52*(6), 499-507.

Haile, T. (2014, Mar 9). What You Think You Know About the Web is Wrong. *Time*. http://time.com/12933/what-you-think-you-know-about-the-web-is-wrong

Hallin, D. (2008) Neoliberalism, Social Movements and Change in Media Systems in the Late Twentieth Century. In Hesmondhalgh, D. & Toynbee, J. (Eds.), *The Media and Social Theory* (43-58). New York : Routledge.

Hango, D. (2014). *University graduates with lower levels of literacy and numeracy skills*. http://www.statcan.gc.ca/pub/75-006-x/2014001/article/14094-eng.htm

iPAC, (2014). *Public Libraries & Digital Literacy*. http://digitalinclusion.umd.edu/sites/default/files/DigitalLiteracyIssueBrief2014.pdf

Mitchell, A. & Page, D. (2014). *State of the News Media 2014*. http://www.journalism.org/packages/state-of-the-news-media-2014

Mitchell, A. & Page, D. (2015). *State of the News Media 2015*. http://www.journalism.org/2015/04/29/state-of-the-news-media-2015

Manning, R. (1965, August). Hemingway in Cuba. *The Atlantic.* http://www.theatlantic.com/past/docs/issues/65aug/6508manning.htm

Ng, W. (2012). Can we teach digital natives digital literacy?. *Computers & Education*, *59*(3), 1065-1078.

Pavlik, J. (2000). The impact of technology on journalism. *Journalism Studies*,*1*(2), 229-237.

Rieh, S. Y. (2014). Credibility assessment of online information in context. *Journal of Information Science Theory and Practice*, 2(3), 6-17.

Rubin, V.L., Conroy, N., & Chen, Y. (2015). Towards News Verification: Deception Detection Methods for News Discourse. The Rapid Screening Technologies, Deception Detection and Credibility Assessment Symposium, *Hawaii International Conference on System Sciences (HICSS48)*, January 2015.

Rubin, V. L., Chen, Y., and Conroy, N. (2015). Deception Detection for News: Three Types of Fakes. In *The Proceedings of the Association for Information Science and Technology Annual Meeting (ASIST2015)*, Nov. 6-10, St. Louis.

Silverman, C. (2015). *Lies, Damn Lies, and Viral Content: How News Websites Spread (and Debunk) Online Rumors, Unverified Claims and Misinformation*. Tow Center for Digital Journalism.

Xing Z., Juan C., Zhiwei J., Fei X., Yu S., Dafeng C. & Xuehui C. (2015). Real-Time News Certification System on Sina Weibo. *Rumors and Deception in Social Media: Detection, Tracking, and Visualization Workshop*. May 19, 2015, Florence, Italy www.pheme.eu/events/rdsm2015