Corporate Responsibility in Combating Online Misinformation

Fadi Safieddine  
Department of Management  
Information Systems  
College of Business  
American University of the Middle East  
Egaila, Kuwait

Wassim Masri  
Department of Technology  
College of Engineering and Technology  
American University of the Middle East  
Egaila, Kuwait

Pardis Pourghomi  
Department of Technology  
College of Engineering and Technology  
American University of the Middle East  
Egaila, Kuwait

Abstract—In the age of mass information and misinformation, the corporate duty of developers of browsers, social media, and search engines are falling short of the minimum standards of responsibility. The tools and technologies are already available to combat misinformation online but the desire to integrate these tools has not taken enough priority to warrant action. This paper presents an effective and practical method based on technologies already available that could be used for browsers and social media websites that would help combat misinformation presented in the form of photo evidence, video evidence, or textual evidence the authors have termed as the “Right-click Authenticate” every browser and social media website should have.

Keywords—Misinformation; Social Media; Browsers; Search Engines; Corporate; Ethical; Responsibility

I. INTRODUCTION

When it comes to authenticating information on social media, the responsibility seems to fall on the web users. An oligopoly of browsers, social media, and news agencies under the banner of freedom of expression have left users with a choice between being passive viewers allowed to either like, comment, or having to undergo long process of online searching in order to authenticate that piece of information. This apparently ethical approach to online freedom of expression seems to have backfired. It has allowed totally unprofessional content; developers bombard predominantly passive web content consumers with news, facts, and stories that cannot be easily challenged. Web users, and specifically online social media users, gradually filter pages, news, agencies, or even friends whom they disagree with their political, theological, and/or ethical predispositions. This has become a hotbed for conspiracy theories who are capitalizing on creating parallel realities to an audience unable to verify what is presented to them. Blocked away from the alternative facts or views, these web users become easy plucking to creating alternative realities that are alien from the truth.

Ethical and social responsibilities of browsers and social media dictate that they make it possible for any individual to authenticate the validity of the information presented.

Our goal is to provide a concept for minimizing the spread of misinformation across social media that is technically feasible and is accepted by all parties involved. The purpose of this paper is thus to provide a study describing the impact of misinformation on societies as well as lesson’s learned from the ways in which Wikipedia manages its contents’ validity. This is followed by the proposal of a solution, which the authors believe would introduce a simple and effective approach towards authenticity validation of information in social media. In so doing, we focus on the following research question:

In what forms is misinformation being shared online and how can browsers and social media websites combat the spread of them?

To this end, the paper will provide an evaluation of contents’ validation in the world of online information sharing and study the relative weaknesses of the current situation in order to propose our approach.

The forms of misinformation that have been identified in this paper are three:

1) Fake, edited, or misrepresented images: This happens when an image is presented as a fact to event, place or story that is untrue or inaccurate.

2) Fake, edited, or misrepresented videos: This happens when a video is presented as a fact to event, place or story that is untrue or inaccurate.

3) Fake, edited, or misrepresented texts: This happens when a story, quote, or news is presented as facts of events or places that are untrue or inaccurate.

These forms of misinformation can appear as political, medical, scientific, theological, race, gender, and others. Some of this misinformation can be dangerous and can have serious consequences. Misleading news in particular is dangerous as it can be used to instigate hatred, racism, theological intolerance, DIY medical treatments, and crime.

The problem with misinformation is that it spreads fast; it has short-term effect that can have long-term consequences. This paper proposes a method that would allow online users quickly, in as little as one click, to check the authenticity of the information. A one-click check can provide the means to discredit the source before it spreads or at the least make the job of unprofessional content developer much harder.
II. LITERATURE REVIEW

Social media cannot be underestimated as a major source of information for the masses. Online social networks have many benefits as a medium for fast, widespread information dissemination. They provide fast access to large scale news data, sometimes even before the mass media as in the case of the announcement of death of Michael Jackson [1]. They also serve as a medium to collectively achieve a social goal. While the ease of information propagation in social networks can be very beneficial, it can also have disruptive effects.

A. Text as form of misinformation

The research of the literature suggests two types of textual misinformation. One type observed during the shootings at Fort Hood, Texas, when a soldier inside the base sent out messages via Twitter as the event unfolded. Incorrect reports of multiple shooters and shooting locations quickly spread through the social network and reached the mass media where it was reported on television broadcasts [2]. This type could be best termed as breaking news and this type of misinformation would be extremely hard to validate as the events play out. Another example of this type is the spread of misinformation on swine flu on Twitter [3]. The spread of misinformation in this case reached a very large scale causing panic in the population. In the wake of the devastating attacks in Paris that left at least 129 people dead and hundreds more injured, social media sites like Twitter, Facebook and Instagram were flooded with updates on the coordinated attack. However, not all of that information was correct. More than 10.7 million tweets were posted about Paris between Friday and Saturday, NBC news reported at least 9 posts that contained misinformation [4].

The other type of textual misinformation shared online is in the form of reports and reflection on news. The difference noted here has to do with the time period which allows some element of reflection and spreading of rumors. This represents hoaxed for conspiracy theorist. For example, following the disappearance of Malaysia Airlines flight MH370, NBC news highlighted various false reports spreading on social media which alleged that the plane had made a safe landing [5].

Social media sites can be a more convenient way for people to source their information but it has been proven many times as inaccurate [6].

While the WEF data [7] showed the rapid spread of false information as a key trend for 2014, the 2013 Oxford Internet Survey [8] found that trust in the reliability of online information among British internet users has changed very little in the past 10 years. More worrying is that the same survey found a trend where web users identified the Internet as the most reliable source of information over television and radio (with a score of 3.6 on average, where 1 is unreliable and 5 is totally reliable) [5].

B. Images as form of misinformation

Various Web-based sharing and community services such as Facebook, Flickr, and YouTube have made a vast and rapidly growing amount of multimedia content available online. Uploaded by individual participants, these pools of content includes varied types of images accompanied by details such as source of image, any editing tools used, date creation, and/or descriptive textual information [9-10]. There is potential in this information that can help confirm the reality of the images users find online. Below, describes a few recent examples of image meta-data misuse on social media. In the past 10 years. More worrying is that the same survey found a trend where web users identified the Internet as the most reliable source of information over television and radio (with a score of 3.6 on average, where 1 is unreliable and 5 is totally reliable) [5].

1) Examples of misinformation by means of images

Misuse of images as a form of evidence can have more profound effect in getting misinformation believable. This form of misinformation has existed far before the arrival of the Internet.

Just as propaganda was not born alongside the Internet, images have been edited well before the advent of Twitter and Photoshop. As an example, the Soviet Union regularly erased disgraced political leaders from photos, even if the results appear woefully amateurish compared to the advanced photo-editing techniques [12].

![Fig. 2. Nicolai Yezhov and former Soviet leader Joseph Stalin during the 1930s [6]](image)

Technology has advanced significantly since and with it the means for content development. Disseminating false information with image evidence has become much easier for unprofessional content developers with the likes of image editing tools.

A Canadian Sikh’s bathroom selfie has gone viral after someone photoshopped the image and posted it on social media.
claiming that he was one of the terrorists behind the attacks in Paris [4].

The problem when images are used as form of misinformation is that they become harder to discredit. In this paper we shall look at a particular example of “Giant Skeleton Unearthed” [13].

Initially reported in 2004 by unprofessional bloggers in India it reached the news media. The Voice newspaper editor, who was first to report it, has since posted a retraction and an apology [13], however the image refused to go away and is still being shared almost 12 years later [14][15][16]. In fact, when searching for matches of this image and where it is being shared, the team unearthed over 800 occurrences. The image which was initially reported as a discovery in India linked to Mahabharata, a Hindu epic story from 200 B.C, has been recycled multiple times and always rebranded with new story of discovery in Portugal, El Salvador, Malaysia, South Africa, Greece, Dominican Republic, Egypt, Kenya, and Saudi Arabia. A key success factor in the wide spreading of misinformation via image is that the users ‘wanting to believe’ the image because it would confirm something they want to believe in which could be political, scientific, health (cures), prejudicial, or religious belief. This represents the framework that has the appearance of credibility.

2) Image meta-data

As digital photography becomes more prevalent, the number of digital images that are stored on photo-sharing sites is increasing dramatically, and the number of images will make it increasingly difficult to authenticate what users are looking at. Image Metadata provides some vital information that could only be available from the original image or subsequent editing of that image. Image metadata is sometimes copied by online photo-sharing sites and made available for views [17]. Image metadata could also assist in search, and is useful in retrieving desired images from a large collection of images [18]. For instance an image meta-data file can be a XML file containing meta-data from Flickr or other social media tools for all the retrieved photos (e.g. photo title, photo description, photo id, tags, Creative Common license type, number of posted comments, the URL link of the photo location, the photo owner’s name, user id, the number of times the photo has been displayed, etc.) [19]. However social media websites have been accused of removing metadata information from images as a part of the social media processing of images, and thus rendering the job of image forensics harder at validating images [20].

In 2013, the International Press Telecommunications Council (IPTC) published a study in the British Journal of Photography into the use of images by social media websites, finding that some of the most predominant ones, such as Facebook, Twitter and even Flickr, remove photographers’ metadata from images they host. The IPTC has tested 15 social media websites, looking at how image sharing, through upload and download, affects the integrity of embedded metadata as defined by the IPTC standards and the Exif standards. The results show that Facebook and Flickr are some of the worst offenders, with most of the metadata removed from the original files uploaded. Twitter has also been found to remove Exif and IPTC metadata from its files. Google+, however retained all types of metadata even when the pictures are embedded or downloaded from the social media site [20].

Although social networks are the main source of news for many people today, they are not considered reliable due to the concerns mentioned above. Clearly, in order for social networks to serve as a more reliable platform for disseminating critical information, it is necessary to have tools that limit or help in combating the effect of misinformation [21].

C. Wikipedia information management

Wikipedia has become an important source of information online, with more than 37 million articles published by November 2015 in more than 250 different languages. A study in 2005 [22] published by Nature concluded that the scientific articles published in Wikipedia and edited mainly by anonymous contributors came close to the level of accuracy of those published in Encyclopedia Britannica which are provided by renowned scholars and scientists. A key component in how Wikipedia manages to reduce misinformation and increase reliability of information is through their validation process and referencing.

Wikipedia still does not receive the same appraisal in academia due to the lack of the peer-review process and sometimes to the lack of proper referencing. Whether or not to trust information online will be largely debatable [23] and no consensus has been yet reached about this issue. Still Wikipedia has gained a huge success due to its strict authenticity of the information process. The system is mainly based on a letter scheme which reflects the quality of each article. The quality classes available are stub, start, class C, class B, Good Article, and Featured Article [24].

Good Articles are articles that passed the Good Articles criteria but not enough to make it to the Featured Articles category. Good articles should usually be [26] well written; verifiable, i.e. contain a list of all references, all inline citations are from reliable sources, and should contain no copyright violations or plagiarism; broad in its coverage, i.e. address main aspects and stay focused on the topic; neutral, stable, and illustrated. Featured Articles [27] are the best articles published on Wikipedia, distinguished by professional standards or
writing, and sourcing. Featured Articles should be well-written; comprehensive, i.e. should neglect no major facts or details; well-researched, where claims are verifiable against high-quality reliable sources and supported by inline citations; non-biased, stable, follow the Wikipedia style guidelines, contain enough media, and stay focused on the main topic without going into unnecessary details. Obviously, articles in other categories are less demanding in terms of quality, such as Class B articles which are mostly complete and without major problems, but require some further work to reach the Good articles category; Class C articles which are substantial, but still missing important content or contain much irrelevant material; Start articles which are developing, but quite incomplete with missing adequate reliable sources; and Stub articles which provide only a description of the topic. In the image below Figure 4, the evolution of an article in Wikipedia (“Atom”) is presented. It demonstrates how an article’s profile can develop through levels and time. For instance, it took the article “Atom” around 6 years to get from a Stub article to a Featured Article.

![Fig. 4. Development of the Article “Atom” through levels (24)](image)

The quality assessments of articles from stub to class B are mainly performed by members of WikiProjects [26], which are projects allowing a group of contributors to work together as a team to improve a certain topic area in Wikipedia. Good Articles and Featured Articles, on the other hand, are assessed by selecting an external panel. Candidates for these panels are nominated on nomination lists and then judged against well-defined criteria [24]. Judgments usually demand a certain consensus to be reached for an article to be categorized, and some projects have even assessment teams.

This thorough process, which is not immune to faults, has proven important and reasonably effective tool in Wikipedia fight to combat misinformation on its website. However, this process is deemed lengthy and time-consuming, making it more suitable for verifying the authenticity of static information rather than dynamic information such as news that might need a less rigorous but rapid method to verify it.

D. Current publications on combating misinformation

It came as no surprise that there is very limited literature on this subject. One work by Budak et al. presented a network algorithm that could be tested in case of two competing campaigns that would test the accuracy of the information. The paper, theoretical, relies on the design of the system itself and input of “influential” people to counter ‘bad’ campaign and limit misinformation. This could potentially be useful during time-sensitive political campaigns or breaking news events. The paper, however, does not suggest how the method in which ‘good’ campaign can participate in countering ‘bad’ campaign. No other academic papers could be found on this topic. The Observer-France 24 posted a guideline for verifying photos online, a process involves some 15 steps and thus confirming the challenges that web users have when authenticating images online [31][36].

III. Approach

The literature review on textual and image misinformation has demonstrated the issues and problems surrounding misinformation online and specifically on social media. However, the tools that can be used to review, rank, and identify misinformation are already found online but may have not been used combined together in a format that would help users in their quest for authentication check. The proposal here is a conceptualization of a quick and easy process that could be used to combat misinformation online. It should and could start with a right-click ‘Authenticate’ option as shown in Figure 5.

![Fig. 5. Conceptualizing a right-click ‘Authenticate’ option](image)

A. Images checks

Reverse image search [37] using Google Images [29], available via Chrome desktop browser as an add-on, is one tool that is underutilized. This is a completely different search engine to Google image keyword search that returns images based on the web user keywords using their standard search page. This search requires user to upload an image or copy the image’s web address to search for matches to that actual image online. The results reveal the sources and dates of the first appearances of that image online and content which appeared with that image. The Google Images search is refined to detect even modifications of the image including color tones changes, cropping, and writing, yet still be able to link it to the original image. Finding the earliest sources of an image is the first step to validating the image origin or the stories associated with it. Second layer is to validate any meta-data linked with the questioned image. Original, image metadata could reveal the device that was used to take the image, the creation date, what changes and on which parts of the image these meta-changes have taken place. Meta-data may also help detect if any image editing tools have been used [31] [32]. Finally, an editorial feedback in a similar format to how Wikipedia operates authentication of information, could be linked to an image. Image editorial feedback maybe combined with explanation based on the origin, date, meta-data, where it appears online, or article that dismisses or confirms that image. Finally a crowdsourcing of feedback could be added as final confirmation. These four sections could be identified as: Image Match, Image Metadata, Editorial, and Feedback respectively. The solution would be to bundle these four sections into one single right click option, see Figure 6.
Fig. 6. Conceptualization of the ‘Authenticate’ outcome as a separate page

So the right-click ‘Authenticate’ option would perform an image search to display early appearances, dates, and early text linked to the image; Display meta-data that shows creation dates, editing, and originality; editorial section with references; combined with crowdsourcing of feedback from visitors. Where an image is new and the authenticity of the image remains unanswered, this would be shown too although the attention could then be focused on the image metadata. Finally, using the same algorithm used for online search engines, an image that gets frequently selected as a match would get higher ranking than those images that do not get selected as a match. To demonstrate this concept, we have provided a conceptualization images in Figure 5, Figure 6, and Figure 7 of the giant skeleton identified in the literature review.

The Chrome browser in this case would take the lead with a right click that usually allows users to perform several other options, Figure 5 can now include an ‘Authenticate’ option. The output could be shown in a new tab, Figure 6 or as a layer over the current display, Figure 7. The information shown in figures 6 and 7 is genuine with the exception of Feedback Section. In the case of the giant skeleton, the Image Match section returned Google Images [29] results that almost immediately questioned the authenticity of the images; the Metadata section showed Adobe Photoshop 7.0 has been used on the photo with no information of camera or author [30]; the Editorial section is taken from National Geographic [13] but could have been easily linked to Wikipedia had it developed a section to authenticate images; and finally the Feedback section could have been the crowdsourcing of feedback allowing final confirmation on the quality of the editorial.

B. Text checks

The option to highlight a text and search for it online is already a well-established tool on many browsers [33]. The problem with such tools is that they only search for where the text appears and provides little or no further information on its authenticity. A right-click authenticate could select that sentence (or few sentences) and make specific online search following the criteria listed above: first appearance, origin, and editorial comments. Where there is dispute of its authenticity, this is would be clearly shown. What would make this option useful is that if it can again harness crowdsourcing to link such pieces of information to other pieces of information, which may be presented in different context or different wording. Turnitin [34], a tool used predominantly in academia to check the originality of students’ work is one of the tools that could be employed in this context.

IV. RESEARCH CHALLENGES AND LIMITATIONS

There are three aspects of research challenges and limitations the authors acknowledge: research sourcing, conceptualization, and implementation.

The team acknowledges that a large portion of the sources is that of online resources, but they owe much of these to the examples of sharing misinformation on social media, which could not be found or sourced from academic sources.

The second limitation has to do with the limitation of the concept. The authors acknowledge that the ‘authentication’ option would have little or no real impact at authenticating
breaking news. For example of misinformation that spread during the shootings at Fort Hood, Texas. Soldiers inside the base sent out messages via Twitter of multiple shooters and shooting locations that were incorrect. The misinformation quickly spread through the social networks and reached the mass media where it was reported on television broadcasts [2]. The right-click authenticate would not be able in such instances to give real answers or help.

Finally challenge would be the implementation, the authors attempted to recreate a working prototype, which is still in progress and will be subject to further research. However, early on the team has reported code obstacles in the way Google search is written. Early indications suggest that an authorized account is needed with the search engine and even then the codes would only allow text search and not image search [34]. Searching via http coding seems to have been disabled by Google.

V. FURTHER RESEARCH

As reported in the challenges, the authors will be working with developers, independent or corporate, to develop the concept into an actual working prototype. The authors also believe that this technique could eventually be extended further to authenticate videos, although complicated algorithms need to be designed and tested. At the time of publishing this paper, no accessible tools could be found to search for matches of videos based on content as opposed to titles. Many online video storage and social media allowed searching by keywords presented by the source and not by distinctive content of these videos.

VI. CONCLUSION

The work presented here remains at the early stage and is aimed at starting the debate on the importance and corporate responsibility of online companies and browsers towards their customers. The authors have showed that the tools and methods that could be used to authenticate text and images are available and achievable but may need cooperation between different corporations. The duty of regulatory bodies should be to set standards and timeline for these changes to come, in the format proposed or alternative formats. What cannot continue is inaction and with acceptance of improper use of a mass media at the scale of social media and the Internet.

REFERENCES


www.ijacsa.thesai.org

131 | P a g e


