

Preprints der Zeitschrift *BIBLIOTHEK – Forschung und Praxis*, 2019, AR 3314 Rügenhagen, Beck und Sartorius

Dies ist ein Preprint . Die endgültige Publikationsfassung erscheint beim Verlag De Gruyter unter <https://www.degruyter.com/view/i/bfup>

AR 3314

Melanie Rügenhagen, Thorsten Stephan Beck and Emily Joan Sartorius

Information Integrity in the Era of Fake News

An Experiment Using Library Guidelines to Judge Information Integrity

Abstract: In this article we report on an experiment that tested how useful library-based guidelines are for measuring the integrity of information in the era of fake news. We found that the usefulness of these guidelines depends on at least three factors: weighting indicators (criteria), clear instructions, and context-specificity.

Keywords: Information integrity, fake news, library guidelines

Die Integrität von Information im Zeitalter von Fake News Ein Experiment mit von Bibliotheken bereitgestellten Richtlinien zur Beurteilung der Integrität von Information

Zusammenfassung: In diesem Artikel berichten wir über ein Experiment mit dem Ziel, festzustellen, wie nützlich Bibliotheksrichtlinien zur Messung der Integrität von Information im Zeitalter von Fake News sind. Das Experiment ergab, dass die Nützlichkeit der Richtlinien von mindestens drei Faktoren abhängt: Gewichtung von Indikatoren (Kriterien), klare Anleitung bzw. Anweisungen, sowie kontextorientierte Indikatoren.

Schlüsselwörter: Informationsintegrität, Fake News, Bibliotheksrichtlinien

Acknowledgements

The authors thank Prof. Michael Seadle for his mentorship throughout the process of planning the experiment and writing the article. We also thank Wjatscheslaw Sterzer who supported us by designing tables 1 through 5 in the article.

Disclosure and conflicts of interest

This research was financially supported by the Humboldt-Elsevier Advanced Data and Text Centre (<https://headt.eu/>).

1 Introduction

We heavily rely on information online to make decisions in our daily life. For example, we may check the weather online to help decide what to wear, or if we need to pack an umbrella. If the information we get is wrong or incomplete, and it rains when our source predicted sunshine all day, we may end up getting stuck outside without our umbrella.

An inaccurate weather report may not be considered “fake news,” but in this example, inaccurate information helped form our decisions about how we interact with our environment. In a similar way, inaccurate information from news sites, newsfeeds and Twitter threads also have an influence on our actions and opinions. Information integrity refers to the degree in which a piece of information is true or honest.¹ Incomplete, or dishonest information may lead to real world consequences. The Pizzagate incident is one of the obvious cases of misinformation having dangerous consequences. A man from North Carolina brought a rifle into a local Washington D.C. pizza shop in response to a story spread online. The story accused U.S. presidential candidate Hillary Clinton, along with other democratic politicians, of running a child sex ring out of the Comet Ping Pong Pizzeria. Disguised as news, this misinformation had dire consequences. Shots were fired, but luckily no one was injured.² In what we are calling the “era of fake news”, the spread of dishonest information online is a serious concern because of its political charge and influence over civic reasoning. The scholars behind the Stanford History Education Group study on evaluation of online information make aware of the lasting democratic consequences of the spread of fake news, writing:

“Credible information is to civic engagement what clean air and water are to public health. If students cannot determine what is trustworthy—if they take all information at face value without considering where it comes from—democratic decision-making is imperiled.”³

Although the example of the Comet Ping Pong pizzeria gunman is extreme, the inability to distinguish real from fake news is not unique to this situation; it is a widespread problem. The previously mentioned study among U.S. middle school, high school, and college students found the ability to tell between credible and noncredible sources of information to be seriously lacking.⁴ And it is not just among youth. According to a survey performed by Ipsos, 45% of German adults, and 46% of US adults have believed a story that later was found to be fake.⁵ Although the concept of “fake news” is not new, the conversation

¹ Seadle (2018).

² Kennedy (2017).

³ McGrew et al. (2017) 7.

⁴ McGrew et al. (2017).

⁵ Duffy (2018).

has reignited after the 2016 U.S. presidential elections.⁶ It is becoming more evident that informing one's self with resources that are honest and accurate, that is, information with integrity, is vital for civic reasoning. The threat that comes with fake news and misinformation has widespread consequences we can hardly foresee.

As information professionals, librarians have taken on the responsibility of fighting against fake news, because distinguishing credible from non-credible sources falls under the umbrella of "Information Literacy". One search for "Fake News" on LibGuides Community⁷, for example, gives us 1,668 different subject guides, or "LibGuides", created by information professionals with the tag "fake news" or "Fake News", 83 guides with the tag "misinformation" and 91 with the tag "fact checking".⁸ On these LibGuides, guidelines and checklists are often provided as a means of aiding students in deciding whether a piece of information is true or false.

In July 2019 the authors of this article carried out an experiment in which we tested guidelines meant for checking the integrity of information online, as provided by US libraries and the American Library Association (ALA). The research question we aimed to answer was: How useful are library-based guidelines for measuring the integrity of information in the era of fake news? The purpose was to develop a more thorough understanding of how librarians promote evaluating information integrity in the era of fake news by taking a close look at the guidelines they provide to the public. From this, we can see how useful these guidelines are for helping users. Usefulness in the context of our research question means whether a step or entire guideline indicated something about the integrity of an article. The goal of our experiment was to see which steps in the guidelines help us come to a conclusion about the integrity of a piece of information.

2 Literature Review

Evaluating the integrity of information is a frequently discussed topic. Some distinguish between misinformation, disinformation, censorship and fake news.⁹ Misinformation comprises "all types of inaccurate information", which includes both disinformation ("deliberately deceptive or misleading

⁶ Bluemle (2018) 266.

⁷ <https://community.libguides.com/>.

⁸ Status: 19 September 2019.

⁹ Sullivan (2018) 2.

Information”) and censorship.¹⁰ In this article we talk about the integrity of information in general, which makes the entire spectrum of misinformation relevant, including fake news.

2.1 Evaluating Information Integrity — Humans vs. Algorithms?

There is a broad consensus that more education is required to empower individuals and make the information ecology we all live with more reliable. Claire Wardle argues: “Every time we passively accept information without double-checking, or share a post, image or video before we’ve verified it, we’re adding to the noise and confusion. The ecosystem is now so polluted, we have to take responsibility for independently checking what we see online.”¹¹ The question is: How efficiently are we in doing that? According to Nicole Cooke “the bulk of disinformation on the Internet could be combated with basic evaluation skills. [...] Although these are seemingly easy tasks, critical information consumption is not automatic, and Internet users need to be taught to evaluate, sort, and effectively use the overabundance of information available online.”¹²

Librarians are considered experts in helping others find and decide if an information source is credible enough for research purposes and beyond. M. Connor Sullivan, librarian at Widener Library at Harvard University, explains this, writing, “Because of their unique position as partners, educators, and community champions, librarians have an opportunity to teach information and media literacy, as well as reframe ideas about navigating the Internet.”¹³

Much of the Library and Information Science literature on criteria for measuring the quality and integrity of information is rooted in the field of information literacy. For example, Elaine Colepicolo gives four recommendations on determining the credibility of a source for academic research which include 1) searching, selecting and evaluating resources from reliable institutions (e.g., libraries and universities), 2) using bibliometric indicators to evaluate information and elements such as authors and journals, 3) analyzing the publication (e.g., author(s), publisher and references), and 4) analyzing the contents (e.g., data and methods reliability, validity, and consistency).¹⁴ It is not surprising that many of the checklists shared via LibGuides to determine if a news source is reliable include similar indicators as Colepicolo, such as the CRAAP Test, used in information literacy lessons and one-shot workshops.¹⁵

¹⁰ Sullivan (2018) 2.

¹¹ Wardle (2017).

¹² Cooke (2017) 217.

¹³ Alvarez (2017).

¹⁴ Colepicolo (2015).

¹⁵ Batchelor (2017).

2.2 The Problem with Guidelines

There are voices that are concerned about how effective library strategies really are when it comes to the evaluation of information integrity — especially considering the role of guidelines and tutorials. For example, M. Connor Sullivan points out that “[i]n brief, library and information science (LIS) professionals do not appear to understand the real danger of misinformation—or at best only understand half of it.”¹⁶ In his opinion, the core of the problem lies in the fact that there is little awareness for “what misinformation does to our mind”¹⁷. Sullivan recommends considering findings from other academic fields that support the understanding of the phenomenon and he suggests there is a need

“for investigations of library strategies and what impact they may have on guarding against or correcting misinformation, as well as debiasing in general. This is by no means a repeat of the familiar call for LIS researchers to improve the scientific or even theoretical status of the discipline, or to bridge the theory-practice gap, but rather to determine whether traditional services work, and what to do if they do not.”¹⁸

McGrew and colleagues believe that “the ‘close reading’ of a digital source, the slow, careful, methodological review of text online—when one doesn’t even know if the source can be trusted (or is what it says it is)—proves to be a *colossal* waste of time.”¹⁹ When we talk about library checklists this criticism is not entirely justifiable, as many of the guidelines suggest both the accurate assessment of the resource itself and a lateral search on the web. Web literacy expert Mike Caulfield argues that the underlying problem is not so much news literacy, but a lack of web literacy, but he shares the opinion that many checklists are way too detailed to be efficient.²⁰

2.3 The “Messy Side of Evaluating Information”

The authors of this paper are well aware that there are additional factors which complicate the evaluation of information. According to Lazer and colleagues “[r]esearch also further demonstrates that people prefer information that confirms their preexisting attitudes (selective exposure), view information consistent with their preexisting beliefs as more persuasive than dissonant information (confirmation bias), and are inclined to accept information that pleases them (desirability bias).”²¹ Moreover, the term “illusory truth effect”²² has been coined to reflect that the more often we are

¹⁶ Sullivan (2018) 2.

¹⁷ Sullivan (2018) 5, quoting Ecker, Ullrich: The psychology of misinformation. (2015) 22.

¹⁸ Sullivan (2018) 8.

¹⁹ McGrew et al. (2017) 8.

²⁰ Caulfield (2017).

²¹ Lazer et al. (2018) 1095.

²² Wikipedia contributors (2019).

confronted with a piece of information the more likely we may accept it as reliable, no matter whether its integrity is proven.²³ With Bernd Becker's words: "This is the messy side (personal beliefs and reasoning) of information literacy that we haven't really had to delve into as much as the technical aspects of locating information."²⁴

With regard to fake news, Donald Barclay pointed out that "[l]ibrarians need to address the emotional component of fake news, to address the ways in which fake news plays on such feelings as fear, anger, joy, and self-righteousness in order to get people to believe things that are mostly, if not entirely, untrue."²⁵ In an article for the magazine *Forbes* Kalev Leetaru argued that "the reality is much more difficult in that 'fake news' is not black and white, it is a hundred [shades](#) of gray."²⁶ In the face of such complexities, how can the average user judge what piece of information is reliable? Reality may be sobering and not everyone is aware of the extent to which their own experiences and attitudes shape the evaluation of information. Hence, as we designed our experiment we took this into account and evaluated our data with a close look at the extent to which we as users have the capacity to judge information with a neutral attitude as well as whether we were able to identify our attitudes at all.

Certainly information integrity in the era of fake news and deepfakes require a thoroughly thought-out and diversified strategy: "What's needed—more than just a pamphlet or a set of guidelines—is a sustained, comprehensive effort to train a new generation in media and information literacy for the social media era."²⁷ Nevertheless, or perhaps precisely because of this, the critical evaluation of those measures that have already been developed, published and applied is a necessary first step to derive a better understanding of how to counter a phenomenon that is having a concrete impact on the worldview and political attitudes of so many in our societies.

2.4 Towards an Automated Solution?

In 2017, a research group led by Dean Pomerleau from Carnegie Mellon University launched a competition²⁸ to test and compare algorithms for checking the integrity of news. The result of the competition showed that there is no general solution to the problem yet, or as Tom Simonite sums it up for the magazine *WIRED*: "The algorithms the winning teams created might help rein in online misinformation, but as tools to speed up humans working on the problem, not autonomous fake news

²³ Meckel and Prange (2018) 56.

²⁴ Becker (2016) 189.

²⁵ Bergan (2017), quoting Donald Barclay.

²⁶ Leetaru (2016), the URL behind the word "shades" is in the original.

²⁷ Holmes (2018).

²⁸ <http://www.fakenewschallenge.org/>.

killbots.”²⁹ Part of the problem is that “[e]xisting technology isn’t close to having the ability to understand language and make decisions that would be needed.”³⁰ Chen et al. who promote an automated verification of integrity, point out that the answer to the question “how do we decide whether something is credible or not?”³¹ is anything but trivial. Leetaru even thinks that “[t]he notion of a magical technology that could instantly label every article on the web as ‘fake’ or ‘true’”³² is an illusion. Zachary C. Lipton is also skeptical about the development of automated recognition tools: “However, it’s not clear that machine learning offers the best hope for near-term solutions. Perhaps crowdsourcing may offer greater hope.”³³ In his opinion, this is because the recognition of fake news is a complex task in which many grey areas must be taken into account and because it is not self-evident when to apply the term ‘fake news’. Are badly researched, or incorrectly transmitted news already fake news, or must there be a deceptive intent to use the term, or how can one tell the difference between fake news and satire? “At the article level, categories should distinguish between outright fabrications, stories with a few inaccurate claims, stories that reference debunked claims, opinion pieces, humor pieces, among others.”³⁴ When a source is marked as problematic on the basis of linguistic analysis or network analysis a different problem arises. What is ultimately put at risk is the freedom of expression. Figueira and Oliveira think “[i]t seems clear that a judgment on the value of information should not be performed exclusively by machines [...]. Freedom of speech must be protected at all cost”.³⁵ Neither should the Internet be censored, nor can decisions be made exclusively by algorithms.

Therefore — at least temporarily — the critical analysis through the user remains an important element and instruments to support users remain important. Our research evaluates some of these existing instruments to see how useful they can be in assisting users in identifying the integrity of online information resources. With this article we hope to cast insight on current practices in educating people how to manually spot misinformation in the library and information science field.

²⁹ Simonite (2017).

³⁰ Simonite (2017).

³¹ Chen et al. (2015).

³² Leetaru (2016).

³³ Lipton (2017).

³⁴ Lipton (2017).

³⁵ Figueira and Oliveira (2017) 822.

3 Experiment

3.1 Methodology

3.1.1 Guidelines

The experiment tested four guidelines provided by two major research institutions, one from the American Library Association and one from a public library. The four guidelines we used are:

- Guidelines suggested by the University Libraries of the University of Washington (UWL)
<http://guides.lib.uw.edu/research/news/fake-news>
- Guidelines suggested by the American Library Association (ALA)
<https://libguides.ala.org/InformationEvaluation>
- Guidelines suggested by The Albany Public Library (APL)
<https://www.albanypubliclibrary.org/fake-news/>
- Guidelines suggested by the Cornell University Library (CUL)
<http://guides.library.cornell.edu/c.php?g=620317&p=5888376>

The UW Libraries guideline was not developed by the university or university library, but instead comes from Onthemedia.org and is merely recommended by the library as one way of analyzing a news source. The ALA guidelines are posted on their “Fake News” LibGuide as “Summary of Tips”, but it is unclear who authored them. The APL does not say they adapted their guidelines from any other source, so we assume they developed them internally, although the APL additionally shares at the end of the guidelines the IFLA (International Federation of Library Associations and Institutions) infographic³⁶. The CUL guideline combines its own recommendations with third-party content, indicates when that is the case, and provides the origin of the content. Assuming that the average user would not go to great lengths to check information integrity, and in order to simulate this approach in our experiment, we have refrained from contacting the publishers of the guidelines.

There are multiple reasons why we chose those four guidelines to test for usefulness in measuring the integrity of information in the era of fake news:

- The guidelines provide users with a series of steps to follow.
- The guidelines differ enough in their list of criteria (as indicators of the degree of information integrity) to possibly give us a variety of results through our experiment, which was necessary to ensure a broad perspective for answering the research question.
- The purpose of the guidelines is clear and within the range of evaluating the quality or integrity of media in general or of identifying fake news in specific.

³⁶ The infographic is available at <http://www.albanypubliclibrary.org/wp-content/uploads/2017/02/How-to-Spot-Fake-News.jpg>.

- To limit the scope of the experiment, we focus on library institutions in the United States of America, which has the advantage that the guidelines are in English, which is the internationally spoken language of scholarship.

It is important to note that many library institutions offer more than checklists on their “Fake News” or “Evaluating Sources” LibGuides. On these web pages they often include external links to other resources, such as news stories and other library resources related to the topic of fake news. For the purpose of our experiment we decided to focus on checklists in order to reduce complexity.

3.1.2 Examples

We used six online articles as examples. We decided not to select classic fake news examples, but news whose integrity could not be judged at first glance. Since most of the guidelines are designed to analyze and evaluate an article from a website (including its about page, URL, author(s), links, etc.), we chose to include news articles from websites only, being well aware that fake information is often shared not necessarily as a URL, but instead as simple posts on social media. In order to build our test set we googled topics that in the past frequently appeared in connection with fake news, such as climate change, migration, vaccination, politics, and more. The debates around these topics are conducted with a lot of passion and the topics polarize — aspects that make it likely to attract producers of misinformation.

The six articles we selected are:

Example 1

Adl-Tabatabai, Sean (2019): Some dogs can detect lung cancer with 97 percent accuracy, study finds: Humans still have so many lessons to learn from animals. In: NewsPunch.com. Available at <https://newspunch.com/some-dogs-detect-lung-cancer-97-percent-accuracy-study/>, accessed 24.09.2019.

Example 2

Eustachewich, Lia; Klein, Melissa (2017): Teacher under fire for slipping anti-Trump question into homework. In: New York Post. Available at <https://nypost.com/2017/02/16/teacher-under-fire-for-slipping-anti-trump-question-into-homework/>, accessed 24.09.2019.

Example 3

Ark Republic News Desk (2018): CDC epidemiologist claimed flu shot caused deadly influenza outbreak, goes missing weeks later. In: Ark Republic. Available at <https://www.arkrepublic.com/2018/02/26/cdc->

epidemiologist-claimed-flu-shot-caused-deadly-influenza-outbreak-goes-missing-weeks-later/, accessed 24.09.2019.

Example 4

Taylor, James (2015): Top 10 global warming lies that may shock you. In: *Forbes*. Available at <https://www.forbes.com/sites/jamestaylor/2015/02/09/top-10-global-warming-lies-that-may-shock-you/#4984b45553a5>, accessed 24.09.2019.

Example 5

AP (2019): SpaceX, Boeing to fly holiday-makers to the International Space Station from 2020. In: Kids News. Available at <https://www.kidsnews.com.au/space/spacex-boeing-to-fly-holidaymakers-to-the-international-space-station-from-2020/news-story/284bb798fbc3d199919fd32159233bb8>, accessed 24.09.2019.

Example 6

Diaspora Reporters (2018): UN, EU and Soros provide migrants with prepaid debit cards to fund their trip to and through Europe. In: Diaspora Reporters. Available at <https://www.diasporareporters.com/un-eu-and-soros-provide-migrants-with-prepaid-debit-cards-to-fund-their-trip-to-and-through-europe/>, last checked 24.09.2019.³⁷

3.1.3 Procedure

The three authors of this article were the test subjects. We chose ourselves as the test subjects due to timing issues, and we used a small set to be better able to discuss the results. We are a small, yet international group. Thorsten Beck is a German post-doctoral researcher who used ethnographic methods for his doctoral research and works on image manipulation detection. Melanie Rügenhagen is a German doctoral candidate with expertise in ethnographic and other qualitative research, as well as digital long term archiving. Emily Sartorius is a US American master's student in information science who did her bachelor's degree in German and teaching elementary education. In this article we use standardized labels for ourselves. Thorsten Beck is subject 1, Emily Sartorius is subject 2, and Melanie Rügenhagen is subject 3.

Each subject independently applied each of the four guidelines to each of the six example articles. This means we have three sets of results. The subjects talked about neither full nor intermediate results during the test in order to achieve unbiased results.

³⁷ The URL we used for the experiment is no longer available. Status: 24.9.2019.

Part of the data gathering was for each of us to list the criteria independently. That way, we ensured the guidelines be used in a realistic way where one would see a guideline online and apply it according to personal perspective.

For the experiment, we as the subjects assumed that we are regular consumers of information online who do not necessarily wish to take a lot of time to look into each and every detail. The goal was to find out how useful the guidelines would be for people in their daily life, no matter how literate they may be in checking online sources.

After each cycle of testing a guideline with one of the examples, we wrote a short summary of our judgement with the main concern whether we considered the information fake, true, or whether we were uncertain. After we finished the test, we collected these conclusions in a spreadsheet to see the differences between our judgements (see table 1).

3.2 Limitations of the Experiment

The fact that the test adopted a realistic approach has two major implications. One is that the number of criteria each of us evaluated varies for each guideline, since in some cases the guidelines, for example, list two steps at once. Some of us split the steps up into two criteria, some did not. We addressed this in the analysis by working with percentages instead of absolute values. It was also not clear in all cases what belonged to a guideline. We individually determined which criteria we thought were part of a guideline when we visited the UWL, ALA, APL and CUL websites. In discussing our results, we discovered that the CUL guideline may include more criteria than we use in our experiment. We addressed this by including these aspects in the discussion of our results.

We are aware of what it means that we were using the same news examples for each guideline. We did this to be able to compare results better among the guidelines. This did, however, challenge our discipline to remain neutral. After looking at one example once, we already had some ideas about how reliable we considered the news article and the medium that published it. With that said, we did our best to remain neutral and approach the article with fresh eyes, using only the set of criteria in front of us as our guide.

We performed the experiment at small-scale: three information scientists used four guidelines with six news articles. The results do not cover the whole of all guidelines, and especially not all people with all their different attitudes and potential prejudices and opinions whose results might differ from ours. Many other guidelines exist, and results may vary depending on the news piece and publication medium

(e.g., newspaper, Facebook, Twitter) to be evaluated. The results of this experiment give an idea of how useful guidelines are in this particular context.

4 Results

4.1 Fake or true? Conclusions about the examples

Table 1 reflects our conclusions from evaluating each article with each guideline per subject (participant). The test revealed that all guidelines produce relatively consistent results for the final decision whether we considered the information in an article true, fake or hard to determine.

Table 1: Conclusions from the experiment

Conclusions from the Experiment									
Guidelines	Example 1			Example 2			Example 3		
	S1	S2	S3	S1	S2	S3	S1	S2	S3
UWL	true	true	wpt?	?	wpt?	wpt?	fake	fake	wnt?
ALA	true	true	wpt?	true	true	wpt?	fake	fake	wnt?
APL	true	true	wpt?	?	wpt?	?	fake	fake	wnt?
CUL	true	true	wpt?	?	true	?	?	fake	wnt?
Guidelines	Example 4			Example 5			Example 6		
	S1	S2	S3	S1	S2	S3	S1	S2	S3
UWL	?	fake	wnt?	true	true	wpt?	?	fake	wnt?
ALA	?	fake	wnt?	true	true	wpt?	fake	fake	wnt?
APL	?	fake	wnt?	true	true	wpt?	?	fake	wnt?
CUL	fake	fake	wnt?	?	true	wpt?	fake	fake	wnt?

“True” means that the subject considers the information true. “Fake” means that the subject considers the information not true. “?” means that the subject is uncertain whether the information is true or fake. “wpt?” stands for “with positive tendency uncertain”, and “wnt?” stands for “with negative tendency uncertain”. S1, S2 and S3 stand for the three subjects who did the test. Orange fields show where we were uncertain. (table designed by Wjatscheslaw Sterzer)

Only subject 1 shows variation in the final decision. That is, one guideline made him think an article is fake, while another guideline left him uncertain regarding that same article. Subject 2 with two exceptions always has a clear answer (fake or true), and subject 3 tends toward either true or fake, but remains uncertain throughout all guidelines and articles. These results represent three different ways of

interpretation: both certain and uncertain answers, (almost) purely certain answers, and purely uncertain answers.

The most striking observation from these data is that there are as many instances of unclear results as there are of clear results. This is important, since it reflects how often guidelines did not give us a final and clear answer. Reasons for being uncertain vary among individuals, because our thresholds between trust and distrust vary. Some may consider hints from a Google search all it takes to say some piece of information is untrue, while others remain suspicious.

The results also show that there is a subtle difference between the guidelines. Some guidelines yield complete uncertainty, while others give better hints as to whether the information at hand is likely true or untrue (hence the positive or negative tendencies in table 1).

4.2 Frequencies by guideline and participant

This becomes more obvious in tables 2, 3 and 4 which show the frequency of instances when we considered the results to be either (a) neutral in terms of whether the information in an article is true or not, or (b) ambiguous by finding both positive and negative aspects or otherwise ambiguous aspects.

For every result, we assigned a color code that summarizes whether the result was neutral, ambiguous, positive or negative. We used the codes in the following manner:

- Black = We feel neutral about the result (the result has no meaning for deciding whether the information in an article is true or not).
- Orange = The result is both positive and negative or ambiguous in any other way, which does not deliver a clear result.
- Green = The result is positive.
- Red = The result is negative.

Both red (negative) and green (positive) mean that we received a clear result. For our research question, the most important categories are black (neutral) and orange (ambiguous), because those two categories indicate in which cases we did not come to a clear result. In summary, tables 2, 3 and 4 show results ranging from moderately to completely contradicting. For the calculation we excluded one of the criteria in the APL guidelines, namely “Ask a librarian!”, since we did this test as information professionals.

Table 2: Subject 1’s results from the experiment showing percentages of all color codes distributed across all articles (table designed by Wjatscheslaw Sterzer)

Across all six articles	UWL	ALA	APL	CUL
Total black entries	37.5%	38.1%	58.33%	46.3%
Total orange entries	20.83%	17.86%	10%	12.96%
Total red entries	16.67%	30.95%	25%	33.3%
Total green entries	25%	13.1%	6.67%	7.41%

Subject 1 found the most neutral results from criteria while applying the APL guidelines (58.33%). All other guidelines have a lower percentage than that and are on about the same level. The most ambiguous results he found using the UWL guidelines (20.83%), closely followed by the ALA guidelines.

Table 3: Subject 2’s results from the experiment showing percentages of all color codes distributed across all articles (table designed by Wjatscheslaw Sterzer)

Across all six articles	UWL	ALA	APL	CUL
Total black entries	15.28%	14.81%	19.7%	5.56%
Total orange entries	16.67%	11.11%	9.09%	14.81%
Total red entries	31.94%	33.33%	30.3%	42.59%
Total green entries	36.11%	40.74%	40.91%	37.04%

Subject 2 found the most neutral results from criteria while applying the APL guidelines (19.7%). The highest frequency of ambiguous results she found using the UWL guidelines (16.67%), closely followed by the CUL guidelines.

Table 4: Subject 3’s results from the experiment showing percentages of all color codes distributed across all articles (table designed by Wjatscheslaw Sterzer)

Across all six articles	UWL	ALA	APL	CUL
Total black entries	30.56%	25.93%	8.33%	15.38%
Total orange entries	6.94%	14.81%	41.67%	42.31%
Total red entries	31.94%	29.63%	25%	16.67%
Total green entries	30.56%	29.63%	25%	25.64%

For subject 3, the UWL guidelines have the highest percentage of neutral criteria (30.56%), closely followed by the ALA guidelines. The most ambiguous results she found using the CUL (42.31%) and the APL guidelines (41.67%).

In summary, there is no consensus among the three of us. Subject 1 thought way more often than the others that his results did not mean anything for his decision about the integrity of the information at hand, and thus had more neutral results. Subject 2 more often thought results were either positive or negative (something concrete), and subject 3 is somewhere in the middle with only two peaks in the data: hardly any neutral results using the APL guideline, and hardly any ambiguous results using the UWL guideline.

With the APL guidelines, subject 1 found close to two thirds of the results from applying criteria neutral. Subject 3 did not quite reach one third, and Subject 2 reached about one fifth. To some extent, this pattern may be grounded in our educational backgrounds, but there is no clear evidence in this small set of participants. This shows how much judgements vary across individuals, which challenges defining a standard that is useful for everyone.

Our judgements contradict each other to some extent. The most instances in which subjects 1 and 2 found criteria neutral can be counted using the APL guidelines (opposite to subject 3). The most instances in which subjects 1 and 2 found the results from applying criteria ambiguous can be counted using the UWL guidelines (again opposite to subject 3). Subject 1's and subject 2's high extremes are where subject 3 has her low extremes for both neutral and ambiguous results.

Table 5 reflects the results from tables 2-4 added up. The ranking for the fewest neutral results would be (1) CUL, (2) ALA, (3) UWL, and (4) APL. The ranking for the fewest ambiguous results would be (1) ALA and UWL, (2) APL, (3) CUL. These rankings vary from each other significantly, which indicates that criteria may provide some evidence. The quality of the evidence, however, is sometimes questionable, considering the amount of ambiguous results.

Table 5: All subject’s results added up (percentages of all color codes distributed across all articles)
(table designed by Wjatscheslaw Sterzer)

Across all six articles	UWL	ALA	APL	CUL
Total black entries	27.78%	26.28%	28.79%	22.41%
Total orange entries	14.81%	14.59%	20.25%	23.36%
Total red entries	26.85%	31.31%	26.77%	30.86%
Total green entries	30.56%	27.82%	24.19%	23.36%

Frequencies by Criteria

These results suggest looking into the actual criteria that may cause the differences reported above. Table 6 in the appendix provides an overview of all criteria. There we grouped the criteria from all four guidelines to show which ones are similar and belong to the same category, since guidelines may use different wordings or be more or less accurate in their instructions, but many of the criteria in the four guidelines recur. For each criterion we calculated the percentage of how often all three of us judged the results from applying criteria the same way. The higher the percentage the higher the degree of agreement between all subjects. This frequency indicates the likelihood that a criterion needs adjustment: the higher the frequency, the higher the likelihood that a criterion needs adjustment. From these frequencies, we construe the degree of usefulness of groups of criteria as well as single criteria.

5 Discussion

The findings above point toward an answer to our research question: How useful are library-based guidelines for measuring the integrity of information in the era of fake news? At the core of the answer lies the usefulness of criteria as tools for collecting evidence about the integrity of information (section 4.3.). Aspects that play a role in determining how useful criteria and entire guidelines are for this purpose are the sequence and importance of criteria employed by the guidelines, instructions on how to interpret findings from applying criteria, and context. These are the themes that emerged from our findings.

In this section, we discuss our quantitative findings by incorporating our qualitative insights into our perspectives as test subjects. Each of us wrote an account of how we perceived the test cycles of using the guidelines. We include quotes from those accounts throughout this section. The data from the experiment and the qualitative accounts introduce two dimensions to our data. What we analyzed by

quantitative means displays what we thought *while* we were doing the experiment. We wrote the qualitative accounts after we had finished the data collection from testing the guidelines, hence these accounts reflect what we thought *after* doing the test.

Our results align somewhat with what Sullivan explains about how the human mind works as well as what the problem with libraries' approach to "fighting fake news" is. The quality of information must be on our minds as we receive it,³⁸ otherwise there is a risk that people categorize it as true, and subsequently have trouble to re-categorize it (as false) and not fall back to the original category (true).³⁹ Opposing that risk requires effort and distrust.⁴⁰

Using qualitative accounts throughout and immediately after the test (and before the analysis), we kept track of our perspectives during the experiment to the extent possible. This addresses Sullivan's suggestion that "LIS researchers first need to understand what the full problem of misinformation is, and why it is we are so susceptible."⁴¹ In order to understand the results of our experiment, it is essential to factor in our perspectives. As information scientists we are shaped by academic approaches to evaluating research output (such as journal articles). Colepiccolo's indicators explained above roughly frame the approach we are familiar with and that shapes our evaluating resources in general. That includes skepticism as well as reliance. Bibliometric indicators have their limits in assessing the quality of authors, journals and information, hence we rely more on scrutinizing the source (e.g. publisher), the resource (e.g., article), and the information that the resource contains. Our qualitative accounts reflect our approaches toward information resources better than the quantitative data, and they also show the variance in this regard across the three of us, which supports what Sullivan addresses in his article.

5.1 Evidence

Our experiment builds on the goal of identifying the degree of information integrity, which means we were (and by training are) all aware that anything we read or see online could be fake. We employed a good bit of distrust in that we did our best to resist what Sullivan describes as a potential human "default" to expect the information we receive is true.⁴² Instead we built on what reduces or opposes

³⁸ Sullivan (2018) 7, quoting Lewandowsky, Stephan et al.: Misinformation and its correction: Continued influence and successful debiasing. (2012) 112.

³⁹ Sullivan (2018) 5, quoting Swire et al.: Processing political misinformation: Comprehending the Trump phenomenon. (2017).

⁴⁰ Sullivan (2018) 7, quoting Lewandowsky, Stephan et al.: Misinformation and its correction: Continued influence and successful debiasing. (2012) 112.

⁴¹ Sullivan (2018) 7.

⁴² Sullivan (2018) 7, quoting Rapp, David N.: The consequences of reading inaccurate information. (2016).

this credulity: “evidence to the contrary”.⁴³ The four library guides were our tools to collect that evidence.

In our test, that evidence did not suffice in many cases, and even in those cases where we did claim to have come to a definite conclusion, the qualitative accounts suggest that we all struggled to take the evidence at face value. Subject 1 says: “Many of the aspects included in the checklists are not necessarily obvious signs for fake information and deceit, and it is not easy to say when a critical point is reached that makes it easy to judge”.

Subject 2 is concerned about how complex it is to distinguish true from false parts of information: “The thing that makes fake information believable is that it seems to be not truly fake, there are grey areas of truth within the articles.” Subject 3 addresses the same issue:

“I will not readily say a resource is trustworthy just because there is an ‘about section’ or the headline is not displayed in bold font. Using the guidelines only led me to think the news articles and publication media had a certain degree of genuineness. I was never able to fully believe something is wrong or correct, since the search based on websites whose credibility is in some sense beyond me.”

How users approach the collection of evidence shapes how they perceive its significance. There are two paths among those we as test subjects chose to go: either (1) trust by default and collect a certain amount of more or less trustworthy evidence that reduces that trust, or (2) distrust by default and collect a certain amount of more or less trustworthy evidence that either supports or reduces that distrust. Either way, the decision relies on the quality of the evidence. Those who chose to (1) trust by default in the experiment were more likely to believe the evidence. Those who chose to (2) distrust by default were more likely to be suspicious about the evidence. The results show how the library guidelines generally left us with the need to trust more or less questionable results that a myriad of Google searches returned. For people counting themselves belonging to category (1), this is less of an issue they are conscious about. For people counting themselves belonging to category (2), this makes for a lot of dissatisfactory results.

5.2 Sequence and Importance of Criteria

The lack of (or at least unrecognizability of) a system behind the sequence of criteria within the four guidelines is a factor that we found to be troubling. Subject 1 says: “in the end it remains unclear how the results are to be evaluated: is the credibility of the author more important than the seriousness of the source? Or should the news/story itself be in the foreground?” Subject 2 makes a remark on this as

⁴³ Sullivan (2018) 7, quoting Rapp, David N. (2016): The consequences of reading inaccurate information.

well when talking about the ALA guidelines: “I also wish that step 8, ‘Search other news outlets to see if the news is widely reported’, came a little earlier in the guidelines because it was not until this step that I could confidently make my final evaluation.” These and the following comment reflect how our prior experience and what we have learned in the past shape our evaluation:

“I found the guidelines competing against some sort of ‘gut feeling’ about information or the news outlet. [...] In some ways this competes with the results of guidelines, because humans form their own opinion on how important a particular criterion is, especially if the guidelines do not suggest a ranking of indicators.” (Subject 3)

This expresses a clear need for information seekers to know which steps of a guideline are more important. This becomes even more essential as we consider that people do have their own opinions on what they regard as important, and if one criterion does not return useful or returns ambiguous results, then the user is tempted to give more attention to an aspect that potentially misleads them to believe some information is true when in fact it is not. Being aware that there are aspects that have higher relevance in making an assumption as to whether some information is true or false would help users being cautious.

5.3 Instructions

The problem with some criteria is that they do not give clear instructions on what it means when the user applies them (i.e., completes a step of the guideline). The lack of these instructions was one reason in our experiment to feel an aspect was neutral or ambiguous in some way. Subject 3 voices this concern as follows: “Doing a Google search to see what comes up does not help much, if users do not know whether they can trust what they find.”

An example can be found with a closer look at the origin of ambiguous results. Ambiguous results are by no means necessarily caused by the criterion. Getting both positive and negative results or otherwise ambiguous results is to some degree a problem of the available data. An online search on Google or a similar search client works with particular algorithms, and search terms let people find different results. Clear instructions on what kind of search terms should be effective would help inexperienced users. The UWL guidelines, for example, say: “If you land on an unknown site, check its ‘About’ page. Then, Google it with the word ‘fake’ and see what comes up.” The problem with this one is that there is no instruction whatsoever on how to interpret “what comes up”. This is where the problem does not solely lie with the data, but the criterion.

We observed that three of the library-based guidelines appear to underestimate how important clear instructions are to help users interpret their findings of an online search. By far the best job on

instructions does the Cornell University Library. Their guidelines make a clear effort to explain what to look for and how to interpret findings. This becomes obvious both from the criteria we included in our experiment and from the criteria that CUL mentions but we did not include in the experiment. The reason for not including a few criteria is that CUL does not provide one clear list of steps to follow, but an entire set of pages to navigate, which all explain several aspects regarding “Fake News, Propaganda, and Misinformation”⁴⁴ (main title of the website that includes a number of subpages). Nevertheless, there is a section on accountability that goes into detail on investigating “News Sources with Explicit Editorial Policies & Ethical Standards” as well as “Qualified Article Authors”.⁴⁵ The instructions on these pages are more elaborate than in the other three guidelines.

5.4 Context

We made a basic observation, which does not come as a surprise: sometimes criteria help, and sometimes they do not. That is, one particular criterion may be useful for one example, but fails to deliver any results in another example. We sometimes judged results from the same criteria differently in different cases with regard to neutrality and ambiguity. Cases are different and our results suggest that criteria should be context-specific to some extent. For example, social media requires a different approach than an online newspaper article does. This becomes clear in a comment by subject 3: “The guidelines are not tailored to any random use case where somebody needs to evaluate the validity of information. Using one guideline can have quite different results compared to using another.” This does not imply an obligation for institutions to provide guidelines for all use cases. Rather it suggests to make context an explicit topic.

5.5 Usefulness of Criteria

We developed the categorization using neutral and ambiguous results from applying criteria in order to better understand the usefulness of those criteria and a guideline as a whole. Both categories of results (neutral and ambiguous) have the potential to indicate that the underlying criteria from the guidelines were problematic in some way and not very effective in supporting our decision making. Results we most often rated neutral or ambiguous are based on applying criteria that likely need to be changed in some way. This, however, is complex: the core observation from the frequencies in table 6 is that we evaluated similar criteria differently. Potential reasons are (a) results vary depending on the kind of information source to be evaluated, (b) some results depend on the search algorithm of a search engine such as Google, and (c) criteria may be similar, but details matter: in some cases, an even slightly

⁴⁴ Cornell University Library (2019a).

⁴⁵ Cornell University Library (2019b).

different or more elaborate wording changes the way we approach the task as well as the way we interpret the results.

5.5.1 Perceived Usefulness

Our written accounts offer a deeper insight into what we thought helped us evaluating information integrity. Each participant voices a notion of what helps them best to judge the integrity of information, yet sometimes more and sometimes less clear.

Subject 1 explains that “to take a close look at the source and to find out about its reliability (e.g., google the source/webpage/URL with ‘fake news’) and to verify the story (google the headline [...] and see what comes up)” was most helpful. Other indicators subject 1 mentions are “a reputable outlet reporting the same story and sometimes investigating whether the background of the author made sense.” Fact-checking sites were less helpful to subject 1 than problematic: “I came across situations where I started questioning the reliability of such fact-checking websites.” On the other hand, subject 1 lists fact-checking sites as one of the most effective criteria.

Subject 2 provides a list of all the criteria that seemed most effective and includes similar indicators, but also considers relevant investigating the background of executives (CUL), contact details (CUL), whether the website has promotional purposes (ALA), and to follow links (UWL). The most obvious difference between subject 1 and 2 is that the latter finds fact-checking sites effective in identifying fake information, while subject 1 has doubts.

Subject 3 thinks that the four guidelines “all include criteria that technically could be used as those indicators” that a “systematic judgement” in measuring “the degree of information integrity” requires. The most important aspects would be “(1) the degree of reliability of the information, including the references for backing up this information, as well as (2) the degree of reliability and credibility of the publishing entity (i.e., the source, such as a newspaper).” This is not unlike what subjects 1 and 2 consider effective criteria, but they are rather generic and lack indicators for determining the reliability of information and publisher. Subject 3 also emphasizes that criteria can only be effective, if they clearly instruct the user: “The core factor for a guideline is the degree to which a user is able to interpret the results. This is bound to (a) each single criterion, and (b) the total of all criteria, the latter of which would benefit from a weighting/ranking of the indicators (i.e., which indicator is more important) to achieve the goal, i.e., judging the degree of information integrity.”

5.5.2 Usefulness by the Numbers

What we thought after doing the test is in many ways in line with what we thought while we were doing the test — but not in all ways. The following inferences are drawn from our results in table 6 in the appendix along with the remainder of our results and refer to the context of our experiment.

Author(s) (*very useful*)

There was some ambiguity in our results, but no significant issues. We consider investigating who the authors of information are and what their background is very useful.

Source/publisher (*very useful*)

We did not run into significant problems when we were examining a website as a source of information. Not many of our results were ambiguous.

Advertisement or promotional purposes (*useful*)

Finding out whether a website is dedicated to promoting and/or selling a product can reveal publications with a bias to put emphasis on a particular perspective, whether there are facts to support that perspective or not. Results may not always help, if the quality of the findings is questionable, but in general, this helped us in our decision.

Style and font (*useful*)

Style and font do not seem to be the most important aspects, since creators and publishers of misinformation can adapt their style to what seems professional. Yet our results indicate it is not entirely useless to examine the style of a website. Our results were somewhat ambiguous, but we did not run into any significant problems.

Ads (*useful*)

Ads and banners proved to be an issue when we had ad-blockers activated in our browsers. Then we were unable to see ads, which makes it impossible to build on a recommendation that asks users to judge a website by the ads displayed on it. Apart from that, checking for ads was rather unproblematic. How well they actually indicate that there might be integrity issues is unclear, but using ads as a contributing indicator when one already found more severe problems is potentially useful.

Content (*useful*)

Looking at the content of an article is a crucial step in determining whether it matches the headline. Results may not always be helpful, but in general, this helped us in our decision.

Verification of information (*useful*)

We did not find any significant issues within this category.

References and links (*rather useful*)

There are no significant issues with criteria that refer to references and links supporting information. There was low ambiguity in our results.

Executives (*rather useful*)

In addition to investigating the author's background, CUL recommends examining who is operating the website and whether they are real people or fakes. There was some ambiguity in our results, but the criterion was rather unproblematic otherwise.

Domain and URL (*rather useful*)

We did not come across significant issues. We conclude that checking the URL of a website can be useful.

Satire (*rather useful*)

Two of the three criteria that suggest investigating whether information is meant as satire have clear instructions. They tell the user to check specific sources, namely *The Onion* and *Clickhole*. Those are unproblematic according to our results. The ALA criterion that makes no suggestion about where to look, is more of an issue. Roughly half of the times we all used that criterion we considered our results neutral. That indicates this criterion is not very useful. A minor change to suggest what exactly to look for might help. "[S]ome quick research on the site and author" is rather vague.

Fact-checking websites (*both useful and useless, depending on the context and search skill-level of the user*)

We generated rather mixed opinions on fact-checking sites. Both criteria that recommend fact-checking sites reached the highest frequencies with regard to ambiguous results: roughly 37% and roughly 38%. This is a sign that fact-checking sites are the cause of some issue(s), even though the numbers on the neutrality of results show that we obviously saw some use in using fact-checking sites. Subject 1 raised a crucial point: how reliable are fact-checking sites? Our data do not deliver any evidence for those sites to be unreliable, but from our experiment we see an issue with both reliability and search function. The guidelines we examined recommend specific fact-checking sites (CUL even labels their four recommended fact-checking sites as reliable), but they also imply the general recommendation to use fact-checking sites. Should any site that claims to check facts equally be subject to scrutiny as the information to be checked itself? Users may ask themselves that question as they stumble upon random websites in their Google searches initiated by guidelines such as those we examined. The search function sometimes returns too many results or none, which does not necessarily mean that there are

no results on the topic. Search terms have to be chosen carefully, which may not be the average user's command.

Choice of sources (*not useful for the purpose*)

We did not run into significant issues, though the usefulness for our purpose is questionable. The criterion recommends selecting reputable websites in the first place, which did not help much in our context, since we already had chosen articles. This recommendation is more useful in general instead of for checking a concrete article.

No sharing (*not useful for the purpose*)

The recommendation by UWL not to share anything we are uncertain about belongs in our category of clearly neutral results, but it is an outlier (similar to consulting a librarian, which did not apply in our case), because it is not a step in finding out whether information is true, but the last resort, if all measures fail, which attempts to reduce the magnitude of spreading misinformation. This is not a problematic step. It just did not change anything for our specific purpose.

Photos (*not very useful*)

The recommendation to use reverse image search did not deliver especially helpful results. Results did not seem very ambiguous, but only clear cases (e.g., when a photo was clearly taken out of context and there is no reference to the original) help the user interpret the findings. In unclear cases (e.g., image was found, but it is hard to tell whether it was misused), users may not know how to interpret their findings.

Bias, opinion, emotion (*not very useful*)

The most significant example from our test is the column on global warming published by *Forbes*. We all know about the issue and have an opinion about global climate change, which certainly could not be excluded in our testing. This is a factor we had to deal with. Being aware of one's own biases and emotions, however, is not necessarily easily done. Knowing about one's biases might help, yet finding out is difficult. Our qualitative accounts make the issue sound like a clear case: not helpful. Subject 1 has clear words for that: "you don't get to understand your biases only because someone tells you to think about them." But the numbers indicate we had mixed feelings about these criteria. There are six criteria with relation to bias, opinion or emotion. All differ slightly. With regard to neutrality of results from applying the criteria, all six criteria scored at least 22% and up to 56% frequency. This means we did not consistently conclude that this group of criteria caused us any trouble and did not help us. With regard to ambiguous results, these criteria scored even lower frequencies. Whether or not this is a sign that we were too "relaxed" and unaware of our own biases and emotions, is unclear and beyond our measures,

but it is a potential cause. According to our results in table 6, those criteria that are clear and easy to understand are less problematic. On the other hand, results were more problematic where criteria give vague instructions to “Check your biases” (ALA), determine whether you are “reading a variety of news sources, including those you don't always agree with” (APL), and figuring out whether “your opinions or judgment [are] clouding your ability to discern fake news from real” (APL). The issues with this group of criteria are that they (a) imply that part of the problem lies with the user, and (b) ask the user to delve into the depth of their consciousness, or unconsciousness. This is problematic, because (a) has potential to mislead the user in their evaluation, and (b) is not easy to do.

Publication date (*not very useful*)

Our results may not be ambiguous, but not useful either. We consistently considered checking the publication date neutral in its meaning to determine the integrity of information.

6 Conclusion

How useful are the guidelines we looked at? The honest answer is mixed: on the macro-level they help in some cases better than in others, and on the micro-level some of the criteria that the guidelines recommend also help better in some cases than in others. Why is that the case? In our experiment, we tested our own use of four guidelines. We used the same guidelines and the same examples to evaluate, and yet we did not always come to the same conclusions. The reason is the various factors on which the usefulness of guidelines for measuring the integrity of information depends. Above all, the personal background (e.g., opinions, education) changes how we perceive what we read and the evidence we collect to verify information. Along with that, we identified four factors through the evaluation of our experiment using both qualitative and quantitative means: Guidelines can become more useful (1) with a high quality of the evidence users collect, (2) with weighted/ranked criteria/indicators, (3) clear instructions, and (4) context-specific indicators.

Verifying the integrity of information we find online relies on the same environment where we found that information: the Internet. The inherent need to check the quality of every single website users find online is a general problem with online searches employing Google or similar search engines.

Technically, one would have to verify every single page that has potential to be an indicator for the integrity of information. That is impossible, and it makes it hard to come to a conclusion — other than: this is an eternal task, and maybe one without hope. This is not what a user wants to hear, and it is not a helpful approach, since it would lock us with the earlier mentioned distrust when we would not accept any evidence to the contrary. A balance between trust, distrust and evidence is key. We observed that

with weighting potential indicators and explaining what it means when we find pieces of evidence, this balance might become more realistic than it was in our experiment.

Weighting indicators means to put more emphasis on some specific criteria than on others. Considering our test results, for example, scrutinizing the entire website that publishes an article could be more important than checking the date or our own biases (which some people may not be aware of or ready to admit to).

Instructions are a clear theme resulting from our experiment, and perhaps even one of the key factors. Not only because they would help identifying fake information, but also because they are difficult to provide. Instructions on how exactly to interpret findings requires certainty as to what reveals whether an article or Facebook message or any other piece of information is not true. This certainty may not always be clear, because misinformation can have facets of truth sprinkled in, in order to make it believable. This also is closely tied to context.

Weighting criteria/indicators, giving clear(er) instructions, and tailoring indicators more specifically to certain contexts are ways toward developing a greyscale that provides users with a probability: how likely is the information I am confronted with true? For future research, there are many open questions. One of them is: how certain can we be in identifying misinformation? This is a complex issue. There are a couple of indicators we can rely on, but there is still more to learn from research to make a judgement more reliable. David Shariatmadari reports in *The Guardian* about recent research that suggests there are subtle differences between the language of an author with honest intentions and the language of someone who tries to deceive.⁴⁶ Research suggests those differences are in the words people use:

“words which can be used to exaggerate are all found more often in deliberately misleading sources. These included superlatives, like ‘most’ and ‘worst’, and so-called subjectives, like ‘brilliant’ and ‘terrible’. They noted that propaganda tends to use abstract generalities like ‘truth’ and ‘freedom’, and intriguingly showed that use of the second-person pronoun ‘you’ was closely linked to fake news.”⁴⁷

Another research group looked into the works of one single author (Jayson Blair) who produced both true and fake newspaper articles. The researchers say “there were more emphatics like ‘really’ and ‘most’ in Blair’s retracted articles. He used shorter words and his language was less ‘informationally

⁴⁶ Shariatmadari (2019).

⁴⁷ Shariatmadari (2019).

dense'. The present tense cropped up more often and he relied on the third person pronouns 'he' and 'she' rather than full names – something that's typical of fiction."⁴⁸

These are details that are hard to detect for readers without help from automated tools. Hence Shariatmadari suggests that these tools will likely be necessary to do this job. With regard to the instructions we were looking for in our non-automated experiment, however, Shariatmadari provides a few guiding questions that offer a somewhat clearer idea that could potentially aid manual judgements in daily news consumption until there are tools that do the same potentially more precisely: "Is the writing more informal than you'd expect? Does it contain lots of superlatives and emphatic language? Does it make subjective judgments or read more like narrative than reportage?"⁴⁹

The fact that so many guidelines exist challenges the effort to standardize the measuring of information integrity. Another challenge is producers of misinformation might adapt to recommendations as they become available. Monitoring and manipulating one's own writing style and vocabulary requires skill, but it is not impossible. It is hard for a manual approach to be perfect in a world that changes so quickly and in the context of an issue that depends on individual human features with regard to how we process information. But the guidelines we examined have a clear advantage: they make users think about the information they receive. Perhaps even more so than a tool that provides us with an answer.

Bibliography

Adl-Tabatabai, Sean (2019): Some dogs can detect lung cancer with 97 percent accuracy, study finds: Humans still have so many lessons to learn from animals. In: *NewsPunch.com*. Available at <https://newspunch.com/some-dogs-detect-lung-cancer-97-percent-accuracy-study>.

Albany Public Library (n.d.): Fake news - Albany Public Library. Available at <https://www.albanypubliclibrary.org/fake-news>.

Alvarez, Barbara (2017): Public libraries in the age of fake news. In: *Public Libraries Online*, 55 (6). Available at <http://publiclibrariesonline.org/2017/01/feature-public-libraries-in-the-age-of-fake-news>.

American Library Association (2019): LibGuides: Evaluating information: Home. Available at <https://libguides.ala.org/InformationEvaluation>. Publication date 18.3.2019.

⁴⁸ Shariatmadari (2019).

⁴⁹ Shariatmadari (2019).

AP (2019): SpaceX, Boeing to fly holiday-makers to the International Space Station from 2020. In: *Kids News*. Available at <https://www.kidsnews.com.au/space/spacex-boeing-to-fly-holidaymakers-to-the-international-space-station-from-2020/news-story/284bb798fbc3d199919fd32159233bb8>.

Ark Republic News Desk (2018): CDC epidemiologist claimed flu shot caused deadly influenza outbreak, goes missing weeks later. In: *Ark Republic*. Available at <https://www.arkrepublic.com/2018/02/26/cdc-epidemiologist-claimed-flu-shot-caused-deadly-influenza-outbreak-goes-missing-weeks-later>.

Batchelor, Oliver (2017): Getting out the truth: The role of libraries in the fight against fake news. In: *Reference Services Review*, 45 (2), 143–48. doi: 10.1108/RSR-03-2017-0006.

Becker, Bernd W. (2016): The librarian's information war. In: *Behavioral & Social Sciences Librarian*, 35 (4), 188–91. doi: 10.1080/01639269.2016.1284525.

Bergan, Rachel (2017): Librarians and fake news: “Trust me, I’m a librarian!”: From the perspective of academic librarians David White (University of the Arts) and Donald Barclay (University of California). Taylor & Francis. Available at <https://librarianresources.taylorandfrancis.com/librarians-and-fake-news-trust-me-im-a-librarian/#>.

Bluemle, Stefanie R. (2018): Post-facts: Information literacy and authority after the 2016 election. In: *portal: Libraries and the Academy*, 18 (2), 265–82. doi: 10.1353/pla.2018.0015.

Caulfield, Mike (2017): How “news literacy” gets the web wrong. Available at <https://happgood.us/2017/03/04/how-news-literacy-gets-the-web-wrong>.

Chen, Yimin; Conroy, Niall J.; Rubin, Victoria L. (2015): News in an online world: The need for an “automatic crap detector”. In: *Proceedings of the Association for Information Science and Technology*, 52 (1), 1–4. doi: 10.1002/pr2.2015.145052010081.

Colepiccolo, Eliane (2015): Information reliability for academic research: Review and recommendations. In: *New Library World*, 116 (11/12), 646–60. doi: 10.1108/NLW-05-2015-0040.

Cooke, Nicole A. (2017): Posttruth, truthiness, and alternative facts: Information behavior and critical information consumption for a new age. In: *The Library Quarterly*, 87 (3), 211–21. doi: 10.1086/692298.

Cornell University Library (2019a): LibGuides: Fake news, propaganda, and misinformation: Learning to critically evaluate media sources. Available at http://guides.library.cornell.edu/evaluate_news. Publication date 13.8.2019.

Cornell University Library (2019b): LibGuides: Fake news, propaganda, and misinformation: Learning to critically evaluate media sources. Expect accountability from your news sources. Available at http://guides.library.cornell.edu/evaluate_news/accountability. Publication date 13.8.2019.

Diaspora Reporters (2018): UN, EU and Soros provide migrants with prepaid debit cards to fund their trip to and through Europe. In: *Diaspora Reporters*. Available at <https://www.diasporareporters.com/un-eu-and-soros-provide-migrants-with-prepaid-debit-cards-to-fund-their-trip-to-and-through-europe>.

Duffy, Bobby (2018): Fake news, filter bubbles and post-truth are other people's problems... Ipsos. Available at <https://www.ipsos.com/en/fake-news-filter-bubbles-and-post-truth-are-other-peoples-problems>. Publication date 6.9.2018.

Eustachewich, Lia; Klein, Melissa (2017): Teacher under fire for slipping anti-Trump question into homework. In: *New York Post*. Available at <https://nypost.com/2017/02/16/teacher-under-fire-for-slipping-anti-trump-question-into-homework>.

Figueira, Álvaro; Oliveira, Luciana (2017): The current state of fake news: Challenges and opportunities. In: *Procedia Computer Science*, 121, 817–25. doi: 10.1016/j.procs.2017.11.106.

Holmes, Ryan (2018): How libraries are reinventing themselves to fight fake news. In: *Forbes*. Available at <https://www.forbes.com/sites/ryanholmes/2018/04/10/how-libraries-are-reinventing-themselves-to-fight-fake-news/#72e5fb04fd16>.

Kennedy, Merrit (2017): 'Pizzagate' gunman sentenced to 4 years in prison. In: *NPR*. Available at <https://www.npr.org/sections/thetwo-way/2017/06/22/533941689/pizzagate-gunman-sentenced-to-4-years-in-prison?t=1566815908914>.

Lazer, David M. J.; Baum, Matthew A.; Benkler, Yochai; Berinsky, Adam J.; Greenhill, Kelly M.; Menczer, Filippo; Metzger, Miriam J.; Nyhan, Brendan; Pennycook, Gordon; Rothschild, David; Schudson, Michael; Sloman, Steven A.; Sunstein, Cass R.; Thorson, Emily A.; Watts, Duncan J.; Zittrain, Jonathan L. (2018): The science of fake news. In: *Science*, 359 (6380), 1094–96. doi: 10.1126/science.aao2998.

Leetaru, Kalev (2016): How data and information literacy could end fake news. In: *Forbes*. Available at <https://www.forbes.com/sites/kalevleetaru/2016/12/11/how-data-and-information-literacy-could-end-fake-news/#528995b73399>.

Lipton, Zachary C. (2017): Is fake news a machine learning problem? Available at <http://approximatelycorrect.com/2017/01/23/is-fake-news-a-machine-learning-problem>.

McGrew, Sarah; Ortega, Teresa; Breakstone, Joel; Wineburg, Sam (2017): The challenge that's bigger than fake news: Civic reasoning in a social media environment. In: *American Educator*, 41 (4), Available at <https://eric.ed.gov/?q=The+Challenge+That%e2%80%99s+Bigger+Than++Fake+News&id=EJ1156387>.

Meckel, Miriam; Prange, Sven (2018): Widerspenstige Wirklichkeit. Mithilfe algorithmisch erstellter Fakes entstehen Parallelwirklichkeiten. Wir sollten den Markt für Meinungen neu gestalten. Sonst wird uns Hören und Sehen vergehen. In: *ada*, 1, 54–61.

Seadle, Michael (2018): An introduction to the column. HEADT Centre (Column on Information Integrity: 1). Available at <https://headt.eu/An-Introduction-to-the-Column>. Publication date 11.4.2018.

Shariatmadari, David (2019): Could language be the key to detecting fake news? In: *The Guardian*. Available at https://www.theguardian.com/commentisfree/2019/sep/02/language-fake-news-linguistic-research?CMP=share_btn_link.

Simonite, Tom (2017): Humans can't expect AI to just fight fake news for them: Don't expect algorithms to rescue us from misinformation. In: *WIRED*. Available at <https://www.wired.com/story/fake-news-challenge-artificial-intelligence/?verso=true>.

Sullivan, Connor M. (2018): Why librarians can't fight fake news. In: *Journal of Librarianship and Information Science*, (March 2018), 1–11. Available at <https://doi.org/10.1177/0961000618764258>.

Taylor, James (2015): Top 10 global warming lies that may shock you. In: *Forbes*. Available at <https://www.forbes.com/sites/jamestaylor/2015/02/09/top-10-global-warming-lies-that-may-shock-you/#4984b45553a5>.

University of Washington Libraries (2019): Fake news – News – Library Guides at University of Washington Libraries. Available at <http://guides.lib.uw.edu/research/news/fake-news>. Publication date 7.8.2019.

Wardle, Claire (2017): Fake news. It's complicated. Available at <https://firstdraftnews.org/fake-news-complicated>.

Wikipedia contributors (2019): Illusory truth effect – Wikipedia. Wikipedia, The Free Encyclopedia. Available at https://en.wikipedia.org/w/index.php?title=Illusory_truth_effect&oldid=908159077. Publication date 8/12/2019.

Appendix

Table 6: Frequency in which all three test subjects felt the results from applying a particular criterium were either neutral (first column) or ambiguous (second column). This list contains all criteria from the guidelines we tested. Similar criteria are grouped, subheadings label the groups. References where the criteria come from are in the right-most column. There is one red criterion, which is less representative than the others because only one subject included it in the test. Blue criteria belong in two different categories and are listed twice. The colors for the percentages indicate how likely criteria are problematic according to our test: green = low frequency / likely unproblematic; blue = medium frequency / rather likely problematic; light red = rather high frequency / likely problematic; dark red = high frequency / very likely problematic.

Neutrality frequency	Ambiguity frequency	Criteria	Guidelines (references)
References and links			
17%	33%	No links, quotes, or references? Another telltale sign.	UWL
13%	22%	Look at the links and sources supporting the article. Click those links. Determine if the subsequent information supports the story. Consider the reliability of the sources.	ALA
11%	22%	Are there links to supporting sources included in the article?	APL
17%	11%	If a story offers links, follow them. (Garbage leads to worse garbage.)	UWL
Verification of information			
6%	28%	Verify an unlikely story by finding a reputable outlet reporting the same thing.	UWL
17%	6%	Search other news outlets to see if the news is widely reported.	ALA
Publication date			
78%	6%	Check the date. Social media often resurrects outdated stories	UWL
78%	6%	Check the date.	ALA
61%	11%	When was the story written? Sometimes news is real but outdated.	APL
Source/publisher			

0%	33%	Consider the source. Click away from the story to investigate the site, its mission and its contact info.	ALA
6%	28%	Consider the source – do a separate search for the website or author. Are they credible?	APL
19%	38%	Independently verify the source (by performing a separate search) and independently verify the information (through more mainstream news sources or fact-checking sites).	CUL
0%	28%	Look for an About page, often in the header or footer of the home page. Read the About page closely for evidence of partisanship or bias. If there's no About page and no Contact page, be very skeptical.	CUL
6%	17%	If you land on an unknown site, check its "About" page. Then, Google it with the word "fake" and see what comes up.	UWL
Author(s)			
17%	26%	Assess the credibility of the author. Do a quick Google search on the author. What is their expertise? What organization do they represent?	ALA
6%	28%	Consider the source – do a separate search for the website or author. Are they credible?	APL
0%	28%	Look for contact information with a verifiable address and affiliation.	CUL
Style and font			
19%	33%	Is the headline outrageous & attention-grabbing? Is it in ALL CAPS or a bold font? Does it use lots of exclamation points?!?!?	APL
6%	17%	Big red flags for fake news: ALL CAPS, or obviously photoshopped pics.	UWL
Fact-checking websites			
11%	37%	Use a fact-checking website: Factcheck.org, Politifact, Snopes.com	APL

19%	38%	Independently verify the source (by performing a separate search) and independently verify the information (through more mainstream news sources or fact-checking sites).	CUL
Executives			
11%	33%	In staff listings (or on the About page), look critically at the list of executives. Are they real people or stock photos? Open a new tab and look for another profile of the individual (e.g. LinkedIn).	CUL
Choice of sources			
39%	28%	Select news sources known for high-quality, investigative reporting. Search these sources directly. Don't settle for web search results or social media news feeds. Social media algorithms are designed to present the news that reinforces your current views, not a balanced view.	CUL
Ads			
6%	11%	A glut of pop-ups and banner ads? Good sign the story is pure clickbait.	UWL
Domain and URL			
28%	17%	Check the domain! Fake sites often add ".co" to trusted brands to steal their luster. (Think: "abcnews.com.co")	UWL
28%	6%	Perform an independent search for the news source. Compare and verify URLs. Example: http://abcnews.com.co/ (fake site) is not the ABC Network News http://abcnews.go.com, but the logo and the URL are almost identical.	CUL
Photos			
44%	17%	Photos may be misidentified and dated. Use a reverse image search engine like TinEye to see where an image really comes from.	UWL
Bias, opinion, emotion			
39%	17%	Gut check. If a story makes you angry, it's probably designed that way.	UWL

44%	11%	Check your biases.	ALA
56%	6%	Are you reading a variety of news sources, including those you don't always agree with?	APL
33%	6%	If you have an immediate emotional reaction to a news article or source: pause, reflect, investigate. Exciting an emotional reaction is a primary goal of fake news producers. Do not be part of a viral fake news spiral.	CUL
56%	0%	Are your opinions or judgment clouding your ability to discern fake news from real?	APL
22%	0%	Sometimes our own biases influence how we interpret what we read.	APL
No sharing			
61%	6%	Finally, if you're not sure it's true, don't share it! Don't. Share. It.	UWL
Content			
22%	11%	Read past the headline. Headlines can be outrageous in effort to get clicks. Go beyond headlines.	ALA
28%	0%	Read past headlines. Often they bear no resemblance to what lies beneath	UWL
Satire			
50%	6%	Consider that the item might be satire. If it seems too outlandish, it might be satire. Do some quick research on the site and author to find out.	ALA
22%	11%	Is it satire or a joke, or from a site such as The Onion or Clickhole?	APL
33%	0%	Satire (for example, The Onion)	CUL
Advertisement or promotional purposes			
28%	6%	Consider that it might be promotional. Is the purpose of the site to sell a product?	ALA
28%	17%	Look for labels: a corporate logo. Or a tiny statement indicating Paid Post, Advertisement, or Sponsored by. Or the tiny Ad Choices triangle at the upper right corner of an image.	CUL



Melanie Rügenhagen

Berlin School of Library and Information Science

HEADT Centre

Unter den Linden 6

D-10099 Berlin

melanie.ruegenhagen@hu-berlin.de



Thorsten Stephan Beck

Berlin School of Library and Information Science

HEADT Centre

Unter den Linden 6

D-10099 Berlin

beck@headt.eu



Emily Joan Sartorius

University of Michigan School of Information
North Quadrangle Residential and Academic Complex
105 S State St, Ann Arbor, MI 48109

USA

sartorie@umich.edu