studia medioznawcze
2019, Vol. 20, No. 4 (79)
ISSN 2451-1617

A Comparison of Fake News Detecting and Fact-Checking AI Based Solutions*

Andrej Školkay
School of Communication and Media, Bratislava
ORCID: 0000-0002-8445-0580

Juraj Filin
School of Communication and Media, Bratislava
ORCID: 0000-0001-5902-6190

ABSTRACT

Scientific objective of this paper is to analyse how advanced are Artificial Intelligence (AI) tools to fight successfully information disorder. More specifically, this is an overview and ranking on existing tools based on AI in this specific area. **Research method** is comparative analytics. We compare the most developed and publicly available fake-news detecting and fact-checking AI based solutions (intelligent machines). The comparison is based on two key parameters: accuracy and comprehensiveness. **Results and conclusions:** Analyse show that a third of the examined AI systems are, in terms of comprehensiveness, in the top category, while the majority are in the medium category. As far as accuracy is concerned, very few AI machine developers are interested in providing further details about their products and functionalities for studies such as ours which raises suspicions about their actual performance. Surprisingly, one of the most discussed AI systems among EU leaders seems to actually belong to the least developed. **Cognitive value:** There is a need for a larger and more detailed study with involvement of AI specialists who would be able, and allowed, to test all available AI machines with their key features and functionalities.

KEYWORDS

artificial intelligence, comparison, fact-checking, fake news, testing

* This is partial deliverable of the H2020 CSA Project: COMPACT: From Research To Policy Through Raising Awareness of the State of the Art on Social Media and Convergence, Project Number 762128. The Commission is not responsible for any claims made in this article.

There are about 50 fake news detecting and fact-checking organisations in Europe, and double that number in other parts of the world (Funke, 2018). Fake news detecting, fact-checking and debunking organisations and initiatives rely almost exclusively on manual tracking of fake news systems (information disorder), and only rarely employ semi-automated tracking systems (Pavleska, Školkay, Zankova, Ribeiro, & Bechmann, 2018). This is a costly, inefficient, error-prone and slow process of making sense of information disorder (which includes deliberately and accidentally or unintentionally misleading information, unexpected offensive outcomes, hoaxes, and conspiracy theories) in both online and offline environments. Measured by volume, only about 0.25 percent of total content delivered by Google contains offensive or clearly misleading content, but this fraction is still considered to be potentially damaging to society.[1] A possible solution appears to be the use of AI powered news and social discourse analysis for such a purpose. Obviously, AI can be used for the same (negative) purpose as a digital weapon in cyber wars using bots. It may be that several AI applications, such as algorithmic journalism, identifiaction of target-groups for specific disinformation campaigns, or the maintenance of user networks, may play a role in fake news distribution.

Nonetheless, this article aims at exploring the most recent advances in this strategic research, focused only on the positive side of the use of AI tools in order to provide up-to-date knowledge and the first comparative assessment of state-of-the-art of AI solutions aimed at detecting and debunking fake news and carrying out fact-checking. Our comparison does not claim to be comprehensive, but is rather a contribution to the debate. In spite of some scepticism about the potential of AI (as we discuss below), including some contradictory gloomy forecasting of the AI negative impact (e.g. Shotter, 1997, and perhaps the most well-known Hawkins, see e.g. Cellan-Jones, 2014), the exploration of AI seems to be highly relevant to the current scientific discourse. For example, 40% of calls (100 of out of 250) for conferences published on the 'easychair' portal in March 2018 included AI among their key words. Yet only about 10 of these actually tackled fake news and/or social media as a major topic and, moreover, there is not a single paper that tackles the role of AI within information disorder in general and the effectiveness of AI tools using a comparative method in particular. Although one can agree with Chinnappa's (2017) and Craft, Ashley and Maks's (2017) arguments that the best way to combat the problem of fake news is to support the development and identification of high-quality online content, promoting media literacy, restricting the flow of money to deliberately misleading content, and ensuring that reporting and feedback tools are as effective as they can be, nevertheless, the AI contribution within this context can, and should be, explored in more detail. There is an important contribution to this debate but it is almost exclusively from experts within the AI – i.e. technology – field (e.g. Vlachos and Riedel 2016; Popat, Mukherjee, Strötgen, & Weikum, 2016; Hassan, Li, & Tremayne 2015; Zhao, Resnick, & Mei, 2015). There also is a paper by Özgöbek and Gullain (2017) in which they offer a brief state of the art overview of the automatic detection of fake news. However, they do not present any AI tools. Therefore, as highlighted by Babakar and Moy (2016, 19): 'There is an urgent need for a thorough literature review of work on automated checking, including work outside academia.' There is a very brief overview of the landscape of automated fact-checking initiatives and research by Graves (2018). It covers only few examples from our sample but brings additional

[1] https://blog.google/products/search/our-latest-quality-improvements-search/

ones. Anyway, Graves (2018, p. 7) concludes that:..." the potential for automated responses to online misinformation that work at scale and don't require human supervision remains sharply limited today." A bit more comprehensive study by Alaphilippe, Gizikis, Hanot and Bontcheva (2019) also summarises state-of-the-art technological approaches to fighting online misinformation. The authors conclude that:..." current automated solutions are not sufficiently effective." (Alaphilippe, Gizikis, Hanot and Bontcheva, 2019, p. 42).

First, we introduce the concept and role of AI within the information disorder context, and we then present general strategies used, or suggested for, fighting information disorder; we also present methodologies for the assessment of AI based detecting and debunking tools. In our key section, we present the first comparison of the more developed and publicly accessible AI machine-learning tools. This comparison is based on a social science approach and is thus limited by the availability of sources, reports and technical pilot testing studies. Nevertheless, such a first-ever study should be of interest to social scientists and policy makers.

AI and Information Disorder

Artificial Intelligence (AI) is the name given to a computer system that attempts to imitate mechanisms of the human intelligence and (in advanced versions) to process human-like learning. However, it is difficult to find a universally satisfying definition for AI because the definition of intelligence itself conjures up fundamental questions of human consciousness that have not yet been resolved by natural and social sciences (Wood, 2016). Even the Association for the Advancement of Artifical Intelligence (AAAI) defines AI quite broadly as: 'he scientific understanding of the mechanisms underlying thought and intelligent behaviour and their embodiment in machines'

AI is typically divided into two groups – strong (broad) AI and weak (narrow) AI. This is the most often used categorisation. Sometimes, one can find division into three broad categories of AI: narrow AI, Artificial General Intelligence (AGI) – (hypothetical) and Superintelligence – (hypothetical) (e.g. Carriço, 2018).

AI is based on the designing of intelligent machines to be capable of acting and thinking with great intellectual competence. However, in our paper it is the capability to learn and process information that is important both for general AI development and for the purpose we discuss in our paper. It is believed that AI has the ability to transform various aspects of people's lives (Joshi, 2017). On the other hand, some argue (e.g. Orlowski, 2017) that while AI is not entirely useless, it is vastly overhyped. Others argue that 'it seems self-evident that the growing capabilities of AI are leading to an increased potential for impact on human society' (Russell, Dewey, & Tegmark 2015, p. 112). Thus, clearly, there is a large expert gap in the assessment of AI Currently, AI is not able to evaluate more complicated and normative statements and cannot disentangle the simplest ambiguities in sentences, e.g. those which cannot be quantified. Identifying manipulated (deepfake) photos and videos is even more challenging.

Chart 1 shows various AI applications and where they are in the current research and development cycle (as of July 2017).

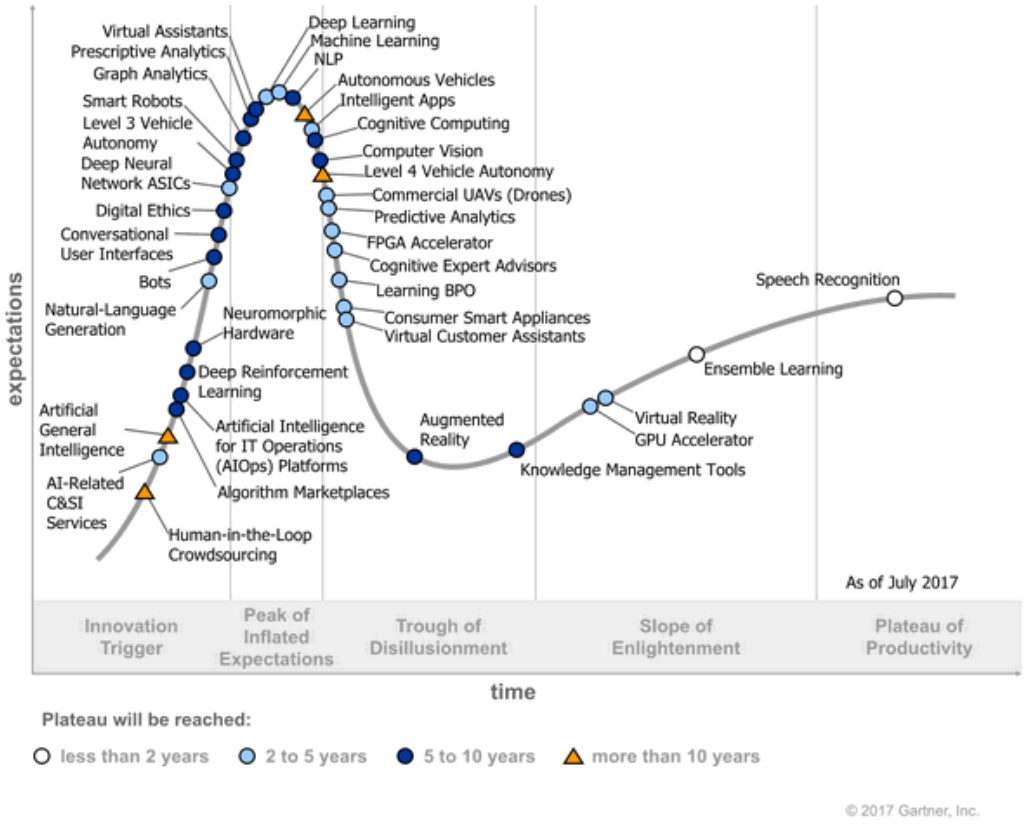

Chart 1. Phases of AI Development

Source: Gartner (published with permission)

These great innovations have been favored not only by the greater availability of data that have made it possible to train computers, but also by advances in cloud computing and new machine learning techniques such as deep learning (Joshi, 2017).

The use of AI is likely to experience social and political challenges (Brundage et al, 2018). So far, there is a very inadequate power of computation since AI may require a high level of calculations, and hence, a lot of power is used for processing. There is a small number of organizations that are ready to invest in the growth and development of artificial intelligence skills (Marr, 2017). To what extent, however, can AI be currently used for detecting and fighting fake news and hoaxes, or various types of disinformation? As Babakar and Moy (2016:1) note, there are many automated fact checking projects worldwide, but they are fragmented and not coordinated.

Strategies for Fact-checking, Detecting and Debunking Fake News with the Help of AI

Till recent years most of the work on identifying fake news was done manually without the use of automated tools (eg. politifact http://www.politifact.com). The procedure of composing a document feature matrix and using it to train a classificator is the traditional ML-approach which is used in Naive Bayes, Regression, or Support Vector Machines. Recently, the natural language processing (NLP) scientific community has turned its attention to creating automatic

tools to identify fake news. These tools are based on creating mathematical models which will classify a script as fake or not, or they will classify a script by some proposed levels of truthfulness (how true or fake an event in news is). One of the most important goals of these models is to not train them only on word occurrences, but also to train them to understand the semantic relations of words (context) in a way which is the same as, or close to, human understanding.

To develop an AI methodology based on mathematical modeling, we need to create a matrix (feature space) in which each column will be a chosen feature and each row is a record. For classifying news as fake or not we need to have not only features based on word occurrences and word relations (both semantic and syntactic), but also features based on how humans check the facts. So, first we need to study human behavior in the process of manual detection of fake news. Humans check if the facts support the story, facts such as people, places or items of interest, such as who was involved, where the event took place, etc. All these facts can be used as features in the above-mentioned matrix. These mathematical models need the feature space in order to be trained. The more records in the feature space, the better the mathematical model will be; this means it will be increasingly close to human accuracy. Most of the feature space is composed by automated text analysis, part of speech tagging, semantic networks, and grammar parsing. Crowdsourcing is required in tagging reference material, not in the extraction of features.

First, the scientists will create the first instance of the feature space, which will contain enough records to be able to train a mathematical model to pass certain baselines, such as a majority baseline or a random baseline, and come close to human performance. Nevertheless, the training of the model does not end here. Eventually, the feature space of the models will need to be updated and more recent records will need to be added. This can be achieved by engaging humans in the process. First, the human flags a news or article as fake. The program will then do a feature extraction, to extract the data needed to fill the feature of each new record in the feature space. In this process, the user first flags a news or article as fake and then a new record in the feature space is created. The mathematical model is then re-trained, to gain more information on how to accurately identify fake news.

The advantage of AI-text (or image) recognition, however, is the lack of this step. These deep learning systems do not rely on manually prepared feature lists for texts but generate their own feature lists, networks, and decision trees from the available material. The input for AI, therefore, is not a matrix but the annotated material itself. The system autonomously extracts features that discriminate between the categories.

As Ghafourifar (2017) reminds us, if we want to build a powerful, intelligent AI-based tool that can detect fake news, we will also need to overcome our own biases, we will have to exercise scepticism about what we read, share and write on social media platforms and on the internet. The advantage of the machines is that they are able to analyze large volumes of content thoroughly, unlike a person.

For more specific AI approaches (e.g. stylometric, semi-supervised learning and hybrid convolutional neural network see e.g. Wang, 2017).

In the meantime, reference approaches and, in a slightly different domain, contextual approaches seem to be closest to delivering real products for fact-checkers (Babakar & Moy, 2016, pp. 18–19).

Comparison of AI Machines for Tackling Information Disorder

In general, for the use and testing of AI machine systems we need to understand what kind of proper data and what proper amount of data is required to train an AI solution. When determining the track record of the product we need to look for proof of use, and preferably case studies (Faggella, 2018). For example, the Fake News Challenge 2017 evaluation was based on a weighted, two-level scoring system.[2] We have followed this approach. In addition to presenting summaries of available case studies (pilot testing), in this section we present a review and definition of possible indicators/metrics and criteria for indicator/metric-choice. This is necessary due to the lack of case-studies for all AI solutions identified, and also because it may contribute to an additional or alternative analytical assessment angle.

On the basis of this literature review we developed indicators for the chosen metric (comprehensiveness) in the context of information disorder (or fake news). Thus, we use both a meta-analytical approach, i.e. a systematic review that summarizes the body of research-based evidence on a specific research question (if there are results available from pilot testing) and also a set of indicators based on defining unique features (functionalities) of each AI solution, developed by ourselves. In particular, our eligibility criteria for including a case (AI-driven software based solutions) in our sample include all AI-based solutions that are publicly available in English and other European languages and are at least at the testing phase. Altogether 23 disinformation-fighting and fact-checking projects were eventually closely scrutinized from these nine were selected as being relevant for preparing systemic calculations (Table 1). In order to illustrate and further specify this task, we mention the key strengths and weaknesses of each AI-based software solution at a certain point of development.

Furthemore, we identified two key indicators for assessing the usefulness of AI-based solutions in fighting information disorder. These are seen as complementary rather than mutually exclusive criteria, as we explain below.

The first key indicator is (grand) accuracy. By accuracy we mean how precise an AI solution is in detecting and analysing/identifying fake news and hoaxes. The generally accepted principle here is based on the elementary recognition test, the numerical results of which distribute themselves into four groups: true positive (tp), false positive (fp), true negative (tn) and false negative (fn). We can calculate the parameters: precision, recall, F1 (f-score) and accuracy itself ,as follows:

$$\text{Precision} = \frac{tp}{tp + fp}$$

$$\text{Recall} = \frac{tp}{tp + fn}$$

$$\text{F1} = \frac{tp}{tp + \frac{1}{2}(fp + fn)}$$

$$\text{Accuracy} = \frac{tp + tn}{tp + tn + fp + fn}$$

[2] http://www.fakenewschallenge.org/

Table 1. Fake News Detecting and Fact-checking AI Tools

	Name of the solution/ Vendor/ Reference	Objective of the system/ solution	User (target) group	Technology and methodology employed	Methods employed, including those supported by AI	Strengths/ Weaknesses
1.	Automated Fact-Checking for Real-Time Validation of Emerging Claims on the Web AIPHES Research Group Darmstadt http://www.k4all.org/wp-cotent/uploads/2017/09/WPOC2017_paper_6.pdf	Fact-checking and validation of news on the web at large in real time	Both casual and professional consumers of web news	Methods for evidence extraction, stance classification, and claim validation. The machine learning methods are trained on a corpus, which was constructed by crawling the snopes.com website. For stance detection, a feature-based multilayer perception was used (one of the best performing models in the Fake News Challenge 2017). For the claim validation, different LSTM network structures were applied.	machine learning methods, Long short-term memory (LSTM/ BiLSTM) Support vector machines (SVM)	A very clear objective and a bold build-up of a complex system for automated fact-checking by using and testing promising AI methods. On the other hand, as the validation process is very challenging, the objective of the authors is not to develop a fully automated system, but 'a system, which is able to assist a fact-checker in the valida-tion process in order to speed up the procedure rather than taking over the job entirely.' The development of the system is currently in progress (2017).
2.	Fully Automated Fact Checking System (Using Ext. Sources) Sofia University Qatar Computing Research Institute, HBKU https://www.researchgate.net/publication/306260513_In_Search_of_Credible_News	Automatically distinguishing false rumors from factually true claims.	Online users, researchers, journalists	The framework of the system uses a deep neural network with LSTM text encoding to combine semantic kernels with task-specific embeddings that encode a claim together with pieces of potentially relevant text fragments from the Web, taking the source reliability into account. The system works fully automatically. It does not use any heavy feature engineering and can be easily used in combination with task-specific approaches as well, as a core subsystem. It combines the representational strength of recurrent neural networks with kernel-based classification.	Neural networks LSTM SVM Natural language processing	The combination of the representational power of neural networks with the classification of kernel-based methods has proven to be crucial for making balanced predictions and obtaining good results. Overall, the strong performance of the model across two different fact checking tasks confirms its generality and potential applicability for different domains and for different fact-checking task formulations. The evaluation results show good performance on two different tasks and datasets: (i) rumor detection and (ii) fact checking of the answers to a question in community question answering forums. At the moment the method is 'lightweight' in terms of features and can be very efficient because it shows good performance by only using the snippets provided by the search engines.

3.	ClaimBuster: The First-ever End-to-end Fact-checking System Team of 13 scholars from Universities of Mississippi and Texas http://idir-server2.uta.edu/claimbuster/ http://www.vldb.org/pvldb/vol10/p1945-li.pdf	Unveiling fake news and fact-checking claims published in media. 'The challenge is that the human fact-checkers cannot keep up with the amount of misinformation and the speed at which it spreads. This creates an opportunity for automated fact-checking systems";	General public, journalists, scholars	End-to-end system that uses machine learning, natural language processing, and database query techniques to aid in the process of fact-checking. It monitors political discourses (e.g., interviews, speeches and debates), social media/platforms, and news to identify factual claims, detect matches with a curated repository of fact-checks from professionals. Its made up of work from human fact-checkers at places including PolitiFact and The Washington Post. (The 2016 U.S. presidential election debates were used for testing, too.) The system quantifies for the claims the probability of being false in a range of 0 – 1.	Machine learning, natural language processing, database query techniques	ClaimBuster can quickly extract and order sentences in ways that will aid in the identification of important factual claims. Discrepancies between human checkers and the machine are still considerable. The algorithm shows some specific shortfalls, according to a review[1]: - Some bold claims can be neglected if a clear subject is missing in the sentence; - Does not weigh more-important words over nonspecific words – this leads to mixing of topics to some extent.

4.	DiversiNews & iDiversiNews: Surfacing Diversity in Online News https://www.researchgate.net/publication/292501422_DiversiNews_Surfacing_Diversity_in_Online_News	To help readers orientate themselves various, often contradictory claims/ opinions about topics published on the internet.	Both casual and professional consumers of web news	The software system collects news articles by crawling the Internet, groups them into stories (that is, clusters of articles reporting on the same event or issue), and presents them through a novel user interface that helps readers discover contrasting perspectives on the news. The central screen of the application, focusing on a single story, presents an overview of the contributing articles: what aspects of the story they emphasize, where in the world they were written and whether they view the story in a positive or negative light. The user can reorder the articles based on any combination of the modalities (i.e. subtopic or aspect, geography, sentiment) to highlight a specific point of view. The summary also changes to reflect the new focus of interest.	Natural language processing	Acoording to a review[2]: Delivers value to the user who needs no special external dependencies or assumptions; the review cites 'the extremely useful feedback... collected from the raters'. On a conceptual level, users can find that making diverse news more accessible is important; on a practical level, they appreciate the summary-based interface and being in control of the criteria by which the news is organized and presented. The implemented summary-centric approach is very appealing for users, as it reduces information overhead while making it possible to grasp different opinions by reading just a few sentences offered via the interface. However, the evaluation showed rather low relatedness assessment of the summaries, that were caused partly, probably, by subjective factors.
5.	BaitBuster: Destined to Save You Some Clicks Team of 3 scholars from Universities of Mississippi and Oklahoma https://www.researchgate.net/publication/320288079_BaitBuster_Destined_to_Save_You_Some_Clicks	Automated clickbait detection	General (readers´) public, scholars	System adopts deep learning techniques, not requiring feature engineering. Distributed subword embeddings transform words into 300 dimensional embeddings that are used to map sentences into vectors over which a softmax function is applied as a classifier. The solution provides explanations of why a headline is a clickbait. Part of it this system is a social bot that regularly publishes automatically generated reports about contemporary clickbait articles. The objective of this bot is to fight against the rising number of malicious bots which parasite on clickbait, listicle and fake contents.	Machine deep learning	Authors: BaitBuster uniquely provides deep learning powered classification and supplements it with explanation and summary by leveraging the headline-body relation. The classification model outperforms existing methods in terms of accuracy.

6.	FiB (student project at hackathon) https://devpost.com/software/fib http://projectfib.azurewebsites.net/	Verifying the authenticity of posts on Facebook	Facebook users	The Chrome-extension system goes through a Facebook feed in real time as the user browses it and verifies the authenticity of posts including status updates, images and links. The backend AI checks the facts using image recognition, keyword extraction, and source verification. This includes a twitter search to verify screenshots. The posts are visually tagged directly on the open FB page. The chatbot inside the system checks every new item.	Backend AI – not precisely specified (Natural Language Processing)	The system has resolved the very actual challenge of verifying claims in the Facebook feed in real time. It does this by the extension of functionality of standard search software. It is designed to recognize and check both text and images and also to examine external links. The system is still narrowly focused on the Facebook environment. It is the result of a one-shot quick product of a student team at a hackathon in 2016 and it has not visibly evolved since that time. There is a question of compliance with Facebook rules and technology about functional and visual intervention in the composition of posts and pages.
7.	FightHoax: AI-Powered News Analysis FightHoax company http://fighthoax.com https://medium.com/fighthoax	To empower news analysis and data journalism with Artificial Intelligence and Big Data methods	Journalists, founders of news, social and data startups, tech-appreciating people	Using the power of IBM Watson to enhance every news article with Natural Language Understanding technologies. Using Google as a „database" so FightHoax evolves as the news story evolves. Scanning the world's news sources and blogs as if it were a database. Many NLP techniques are being in order to assess whether a news article contains legit and trusted information. Algorithm understands the content of each news article like humans do, then, it performs logical steps that human fact-checkers perform by doing comparisons. In addition, the algorithm analyzes the language used, the author of the article and other factors to calculate the outcome.	Natural Language Processing text-mining sentiment mining	It understands different aspects of the article like the topic, the sentiment of each sentence, taxonomy, also tiny parts of speech. It can provide information on the source of the article and background of the author. It can decide if the article is an opinion article, a clickbait article, and whether it includes propaganda or hate speech. It does not evaluate the trustworthiness of an opinion article. During an independent test, FightHoax overall performed with reasonable accuracy, especially in the true positive range, but experienced several inconsistencies in identifying some of fresh news.[3]

8.	FakeRank (AdVerif.ai) adverifai.com	Verification of advertisements Fighting spam, malware and inappropriate content	Advertisers publishers advertising agencies	FakeRank is like PageRank for Fake News detection, only that instead of links between web pages, the network consists of facts and supporting evidence. It leverages knowledge from the Web with Deep Learning and Natural Language Processing techniques to understand the meaning of a news story and verify that it is supported by facts. It uses a spectrum of AI tools – from machine vision for image manipulation detection to natural language processing for psycholinguistic feature analysis, and data pipelines for deep learning.	Natural Language Processing machine vision	Strength – propitiatory data and methods pertaining to deep learning and natural language processing. Itis designed for fake news detection, rather than as a fact-checking tool. Lacks the ability to assess the accuracy of purported facts within an article (does not have a database of common facts). (from the review: David Cox at NBC News)[4]
9.	Search Quality Rater Helping search algorithms eliminate misleading content. By Google https://blog.google/products/search/our-latest-quality-improvements-search/	Providing users with access to reliable sources, i.e. identifying such sources and preventing the spread of misleading content.	General users	Developing changes to Search involves a process of experimentation that includes human evaluators. Recent updates have improved the system´s ability to flag misleading, offensive and unsupported conspiracy content. This has begun to help algorithms in demoting such low-quality content.		A practically implemented and ever- improving system aimed at maximising effectiveness in an immense digital environment. The automated part of the Google quality rater system is still relying to a substantial extent on the human element – evaluating and data supervising by humans.

[1] Review: Brooke Borel at Popula Science: Can AI solve the internet's fake news problem? A fact-checker investigates. Retrieved from https://www.popsci.com/can-artificial-intelligence-solve- internets-fake-news-problem

[2] Review: Daniele Pighin, Enrique Alfonseca, Felix Leif Keppmann, Mitja Trampus: Evaluation of the DiversiNews diversified news service (Technical report). Retrieved July 2014 from https://arxiv.org/ftp/arxiv/papers/1407/1407.4454.pdf

[3] Review: Demetrios Pogkas at GitHub. Retrieved from https://github.com/demetriospogkas/FightHoax-Artificial-Intelligence-Fact-Checking-Tests

[4] Review: David Cox at NBC News. Retrieved from https://www.nbcnews.com/mach/science/fake-news-still-problem-ai-solution-ncna848276

Source: own study

For some of the examined AI systems, the creators published numerical values for some of the above-mentioned parameters related to grand accuracy. In some cases, the reviewers did so. However, there is no unified view on this question, i. e. which of the parameters would best describe the abilities of a respective system and what methodology should be applied. Moreover, in the given phase and conditions, there could be doubts about the objectivity of the accuracy measurements in some cases. Several systems are still in development aimed at improving recognition reliability. It was not the primary intention of the researchers to minutely measure "physical" performance of the systems, but rather to assess their design and elaboration potential.

Those authors of AI systems who released accuracy-related data have indicated that the figure for the accuracy is rather high – between 89 and 98.3 %. They were; FightHoax (89%), FakeRank (90%) and BaitBuster (98.3%). The creators of ClaimBuster put their parameters for both precision and recall at between 74 and 79%. The AIPHES research group indicates that the F1 score is 55% for its system. It also cites the evaluation metrics for Fake News Challenge at 82.7%. There could be a topic issue here for future research projects to find and apply a suitable universal metric to test, measure and fairly compare the achieved performance of fake news detecting AI systems.

The second key aspect is comprehensiveness. By comprehensiveness we mean how complex the AI solution is, i.e. how broadly it covers various aspects of the problem with its functionalities. While accuracy can be very high when focused on a narrow sample, comprehensiveness can be very low. Indeed Su, Zhang, Chen, Yi, Chen, Gao, (2018) revealed a tradeoff in accuracy and robustness These researchers are worried about gap in well-trained deep neural networks versus adversarial examples. In other words, it is more related to security issues.

Thus, it is necessary to combine both accuracy and comprehensiveness. However, there is a methodological challenge here. The narrower the scope, the more likely the AI fake news checking project is to provide practical tools for factcheckers. The more ambitious the scope of the project (aiming at achieving as many as possible goals), the closer it is likely to be to pure research and not practically usable one (Babakar & Moy, 2016, p. 21).

Considering this caveat, we still think that our overview may be useful. Comprehensiveness is assessed independently by both the authors of this study and three external assessors, based on the available description of the AI solution. It should be mentioned here that Alaphilippe, Gizikis, Hanot and Bontcheva (2019, p. 42) seem to consider accuracy and effectiveness of misinformation technology as the most relevant criteria for assessment. Moreover, they suggest that:.." there is also strong need to look beyond "simply" evaluating the and also consider how succeptible to abuse are current algorithms." (Alaphilippe, Gizikis, Hanot & Bontcheva, 2019, p. 42). This latter issue is related to security parameters.

For the purpose of this research we have decomposed (broken down) the content of the term 'comprehensiveness' with the aim of identifying, designating and restructuring a set of components that allow its 'volume' to be quantified as achieved by the respective AI systems. Altogether 20 basic-level categories were selected, describing various features, qualities and functionalities of the systems. These categories/indicators were extrapolated from available descriptions of AI tools. Arguably, the total of categories/indicators identified can be considered as the current maximum level of comprehensiveness of AI tool in this category. The categories were initially assessed and rated separately, and the results were then aggregated according to three main indicators ('evaluation pillars', listed below) and then further numerically processed at the indicator level up to calculation of the final numeric value. In the first two steps, the values of both 'elementary' categories (accuracy and comprehensiveness) and the pre-composed

indicators were weighted using selected proportions. There is an element of subjectivity in setting the weighting parameters that can be discussed in the future. However, in creating the weighting structure we respected the logic of the topic and research objectives. The above-mentioned three pillars are as follows:

A. recognition of the VERACITY (weighting 70%);

B. detection of the MANIPULATION OF FACTS (20%);

C. added value/useful special functionality of the system (10%).

The contributing categories were weighted within respective indicators at various levels from 5% to 70%. Justification of justify these weights can be seen in the line above and in the tables below, Their values were based on collective discussion of researchers, considering overall aim of these AI tools. We obviously included the irrelevant indicators, to provide a rather complex overview of each AI system. Moreover, we could simply underestimate importance of a particular indicator, thus it was fair to include them all.

The pattern of the evaluation, together with assigned category weighting, can be seen in Table 2, where the example is the ClaimBuster system.

The table is composed of assessments as provided by five evaluators within a simple range: Yes – Questionable – No. Only 'Yes' and 'No' evaluations are shown. The votes of the evaluators are weighted, too, as they are variously disposed towards the point of view of the research topics. For a 'Yes' answer there is a full point rating, for the question mark just a half. The totals for the A, B and C indicators are weighted, too, and the sum of the three percentage rates creates the overall rating as a percentage. The table composition ensures that the resulting total (the last number on the right down) cannot exceed 100.

The evaluators had to examine categories of the systems' features by descriptions provided by their creators, as well as occasional external reviewers (e.g. available peer reviews). This does not offer quite sufficient possibilities for rating the practical performance of every system, but it rather delivers an informed view on the system functionality in terms of basic features, also taking into account the system's ambitions for the future. Some of the projects seem to be relatively short-lived or halted at the moment; however, they were chosen for calculating the rating in the same way as the others, as they are relevant in relation to the research objectives. There was also one system with a very low availability of information and data – Google´s Search Quality Rater´s extension to the fields of Artificial Intelligence and fake news detection. It is reasonable to assume that the company is employing part of its extensive capacity in this direction, particularly since the Google contribution to AI is known to be very strong and active. However, lack of data and information about the outcomes leaves to the evaluators of the non-transparent AI system little opportunity to provide optimistic ratings.

An analogical table (Table 2) is provided for every examined system. The results, together with particular results for indicators A, B and C, are shown in Table 3. The nine systems are sorted according to the calculated score. However, numeric differences between some of them are very small and it was necessary, as suggested above to also take into account the subjective features of the methodology. The grading taxology of existing AI systems and differentiating them into 'High', 'Medium' and 'Low' levels for comprehensiveness would also be logically of some subjective uncertainty,. An overall view of the evaluation results shows a grouping of three items around the 60 mark, there is then a group of achievers in between 44 and 54, and then, the Google system.. Taking into account these empirical valuations, we can for the current purpose assign the "High", 'Medium' and 'Low' grade of comprehensiveness to the three parts on the vertical axes, with formal limits arbitrarily (but considering above mentioned emerging parameters) selected, 35 and 55 percent.

Table 2. Indicators and partial categories of comprehensiveness: assessments by evaluators (ClaimBuster example)

	Category/summary	Indicator weight		Assessments of evaluators					Σ (pts)	Composed	Σ (%)
	Evaluator			E1	E2	E3	E4	E5			
	Evaluators´ weight			*30*	*20*	*20*	*15*	*15*			
Indicator A – recognition of VERACITY	**Provides algorithmic and computational tools to assist lay people and/or professionals in checking and vetting claims**	**60**	assessment	Yes		Yes	Yes	Yes			
			points	30		20	15	15	80	48	
	Able to analyze content on internet and social media newsfeeds	**10**	assessment	Yes		Yes	Yes	Yes			
			points	30		20	15	15	80	8	
	Monitor (live) discourse on social media	**10**	assessment	Yes			Yes	Yes			
			points	30			15	15	60	6	
	Able to analyze content on SM in general	**10**	assessment	Yes		Yes		Yes			
			points	30		20		15	65	6.5	
	Detecting manipulated content or context of images	**5**	assessment				Yes				
			points				15		15	1.5	
	Able to analyze audiovisual content	**5**	assessment			Yes	Yes				
			points			20	15		35	3.5	
	Σ for indicator A (points, %)		(weight 70%)							73.5	**51.45%**
Indicator B – detection of manipulation of facts	**Decides whether language of the article promotes propaganda or if there is an intention to influence opinion**	**20**	assessment		Yes						
			points		20				20	4	
	Suggests that headlines aim to entice readers to click (clickbait) = detecting false connection – the story does not support the headline	**70**	assessment								
			points					7.5	7.5	5.25	
	Detects patterns associated with spam and inappropriate content	**10**	assessment		Yes		Yes	Yes			
			points		20		15	15	50	5	

	Σ for indicator B (points, %)		(weight 20%)							14.25	**2.85%**
Indicator C – extra added value/ special functionalities	**Understands different aspects about the article such as topic, sentiment, taxonomy, entities**	**20**	assessment			Yes	Yes	Yes			
			points			20	15	15	50	10	
	Indicates whether the article is an opinion article	**10**	assessment		?						
			points		10				10	1	
	Arguments about the quality of the article	**10**	assessment				?	Yes			
			points				7.5	15	22.5	2.25	
	Offers information about author's past activities	**10**	assessment								
			points								
	Provides links to articles on the same topic from different sources (and political views)	**10**	assessment				Yes				
			points				15		15	1.5	
	Is able to analyze sources of the article	**10**	assessment		Yes	Yes	Yes	Yes			
			points		20	20	15	15	70	7	
	Detects disguising impersonation of sources	**10**	assessment				Yes	Yes			
			points				15	15	30	3	
	Delivers matches instantly to the audience	**10**	assessment	Yes		Yes		Yes			
			points	30		20	7.5	15	72.5	7.25	
	Automatically translates claims if not found in original language	**10**	assessment	Yes			Yes	Yes			
			points	30			15	15	60	6	
	Relies on results of work of "manual" fact-checking entities *(informational)*	0	assessment					Yes			
			points					15	15		
	Includes knowledge bases/repositories *(informational)*	0	assessment			Yes	Yes	Yes			
			points			20	15	15	50		
	Σ for indicator C (points, %)		(weight 10%)							38	**3.8%**
	Σ for the system A+B+C (points)										**58.1%**

Source: own study

Table 3. Assessment of fake news detecting and fact-checking AI tools in terms of comprehensiveness*

	System	**Veracity evaluation → Fake news detection**		**Detection of manipulation of facts**		**Useful extra functionalities**		**Σ for the system**
		= Indicator A		= Indicator B		= Indicator C		
	Comprehensiveness	(weight 70%)		(weight 20%)		(weight 10%)		
		resultant?	weighted	resultant	weighted	resultant	weighted	
	High							
1	**AIPHES**	73.5	51.45%	26.25	5.25%	41.25	4.125%	**60.825%**
2	**Sofia – Qatar**	75.5	52.85%	18	3.6%	39.25	3.925%	**60.375%**
3	**ClaimBuster**	73.5	51.45%	14.25	2.85%	38	3.8%	**58.1%**
	Medium							
4	**DiversiNews**	64.875	45.412%	12.75	2.55%	62.25	6.225%	**54.187%**
5	**BaitBuster**	45,25	31.675%	76	15.2%	44	4.4%	**51.275%**
6	**FiB**	55.375	38.762%	34.25	6.85%	34.5	3.45%	**49.062%**
7	**FightHoax**	35	24.5 %	52.25	10.45%	52.5	5.25%	**44.662%**
8	**FakeRank**	39.25	27,47%	69	13.8%	29.25	2.925%	**44.2%**
	Low							
9	**Search Qual. Rater**	20.75	14.525%	39.75	7.95%	30.75	3.075%	**25.55%**

* according to the resulting values in %

Source: own study

The overall results indicate that a third of the examined AI systems are, in terms of comprehensiveness, in the top category, while the majority are in the medium category.

Disproportions can also be seen between the evaluation results for the systems by researchers on one side and the creators on the other side. We tried to acquire from the creators´ teams their own evaluation; the most compact evaluation was provided by the AdVerify company which delivers the FakeRank AI machine. Surprisingly, two sets of major qualities and properties of this system, as seen by its creators, versus independent researchers match just loosely. The creators´ rating actually comes out as lower than the researchers, as is showed in Table 4. Specifically, the creators had rated better special functionalities that were not clearly visible in the systems´ descriptions; on the other hand they did not rate too highly the potential abilities of the system in the better weighted categories that describe the potential for directly revealing disinformation in general, as well as detecting clickbaits. (Note to the methodology: the evaluation by creators has a standard category structure adapted to just one evaluator with a vote weight of 100%.)

Table 4. Assessment of system qualities – researchers vs. creators

Case: FakeRank

	System	**Veracity evaluation → Fake news detection**		**Detection of manipulation of facts**		**Useful extra functionalities**		**Σ for the system**
		= Indicator A		= Indicator B		= Indicator C		
			(weight 70%)		(weight 20%)		(weight 10%)	
		resultant	weighted	resultant	weighted	resultant	weighted	
8	**FakeRank – by research**	39.25	27.47%	69	13.8%	29.25	2.925%	**44.2%**
8	**FakeRank – by AdVerify**	40	28%	10	2%	35	3.5	**33.5%**

Source: own study

Conclusion

Although it is unlikely that AI will play a key role in the few next years, it can still contribute partially, but nevertheless significantly, to detecting and debunking fake news within the context of fighting information disorder. This contribution of AI can be even more relevant if there is involvement of additional AI features in the current, only partially automated, fact-checking and fake news detecting systems. Our survey has brought together a first comprehensive, but still only tentative, overview of some prototypes focused on detecting and debunking fake news and fact-checking with AI features. However, only a few of them appear to have been independently tested, and sometimes these pilot tests show large discrepancies between claims by the producers and the testers´ findings. Moreover, very few AI machine developers are interested in providing further details about their products and functionalities for studies such as ours. This raises suspicions about their actual performance. We have stated below the sources that communicated and thus co-operated with us, although some of them did not explain to us all the issues. In some cases it appears that there are only abandoned early versions of AI-backed prototypes. There is a need for a larger and more detailed study with involvement of AI specialists who would be able, and allowed, to test all available AI machines with their key features and functionalities. The most promising AI machines should be further supported and developed. In general, there is a need to pool human and financial resources and to develop and/or to test further the most promising AI machines that could help us to tackle information disorder as soon as possible. There appears to be a prevailing consensus that this task requires a few more years at least.

Bibliography

Alaphilippe, A., Gizikis, A., Hanot, C., & Bontcheva, K. (2019). Automated Tackling of Disinformation. Retrieved from http://www.europarl.europa.eu/RegData/etudes/STUD/2019/624278/EPRS_STU(2019)624278_EN.pdf

Babakar, M., & Moy, W. (2016). *The State of Automated Factchecking. How to Make Factchecking Dramatically More Effective with Technology We Have Now. Full Fact.* Retrieved from https://fullfact.org/media/uploads/full_fact-the_state_of_automated_factchecking_aug_2016.pdf

Carriço, G. (2018). The EU and Artificial Intelligence: A Human-Centred Perspective. *European View*, 2018, *17*(1): 29–36. DOI: 10.1177/1781685818764821

Cellan-Jones, R. (2014, December 2). *Stephen Hawking Warns Artificial Intelligence Could End Mankind*, BBC. Retrieved from http://www.bbc.com/news/technology-30290540

Chinnappa, M. (2017). *We Are All in This Together*, *British Journalism Review*, 28(3), 50-55. DOI: 10.1177/0956474817730769

Craft, S., Ashley, S. & Maksl, A. (2017). News Media Literacy and Conspiracy Theory Endorsement. *Communication and the Public*, *2*(4): 388–401. DOI: 10.1177/2057047317725539

Faggella, D. (2018, May 14). *How to Assess an Artificial Intelligence Product or Solution (Even if You're Not an AI Expert)*. Retrieved from https://www.techemergence.com/how-to-assess-an-artificial-intelligence-product-or-solution-for-non-experts/

Funke, D. (2018, February 23). *Report: There are 149 Fact-Checking Projects in 53 Countries. That's a New High*. Retrieved from https://www.poynter.org/news/report-there-are-149-fact-checking-projects-53-countries-thats-new-high

Ghafourifar, A. (2017). *How AI is Winning the War Against Fake News*. Retrieved from https://venturebeat.com/2017/06/11/how-ai-is-winning-the-war-against-fake-news/

Graves, L. (2018, February). *Understanding the Promise and Limits of Automated Fact-Checking*. Factsheet, Reuters Institute for the Study of Journalism. Retrieved from https://reutersinstitute.politics.ox.ac.uk/sites/default/files/2018-02/graves_factsheet_180226%20FINAL.pdf

Hassan, N., Li, Ch., & Tremayne, M. (2015). Detecting Check-worthy Factual Claims in Presidential Debates. *Proceedings of the 24th ACM International Conference on Information and Knowledge Management*: 1835–1838.

Hassan, N., Zhang, G., Arslan, F., Caraballo, J., Jimenez, D., Gawsane, S.,… Shohedul, H. (2017). *ClaimBuster: The First-ever End-to-end Fact-checking System*. Proceedings of the VLDB Endowment, *10(*12): 1945–1948.

Marr, B. (2017). *The Biggest Challenges Facing Artificial Intelligence (AI) In Business And Society*. Retrieved from https://www.forbes.com/sites/bernardmarr/2017/07/13/the-biggest-challenges-facing-artificial-intelligence-ai-in-business-and-society/3/#37119b357b6

Özgöbek, Ö., & Gullain, J.A. (2017). *Towards an Understanding of Fake News*. Norwegian Big Data Symposium NOBIDS:35-42.

Orlowski, A. (2017, January 2). *Artificial Intelligence' Was 2016's Fake News*. Retrieved from http://www.theregister.co.uk/2017/01/02/ai_was_the_fake_news_of_2016/?page=1

Pavleska, T., Školkay, A., Zankova, B., Ribeiro, N., & Bechmann, A. (2018, February). *Performance Analysis of Fact-Checking Organizations and Initiatives in Europe: A Critical Overview of Online Platforms Fighting Fake News.* Retrieved from http://compact-media.eu/fake-news/

Popat, K., Mukherjee, S., Strötgen, J., & Weikum, G. (2016). *Credibility Assessment of Textual Claims on the Web.* Proceedings of the 25th ACM International on Conference on Information and Knowledge Management, 2173–2178. *ACM*. DOI: 10.1145/2983323.2983661

Prateek, J. (2017). *Artificial Intelligence with Python.* Packt Publishing.

Rony, M.M.U., Hassan, N., & Yousuf, M. (2018). *BaitBuster: A Clickbait Identification Framework.* Proceedings of the 32nd AAAI Conference on Artificial Intelligence. (AAAI–18).

Rony, M.M.U., Hassan, N., & Yousuf, M. (2017). *BaitBuster: Destined to Save You Some Clicks.* Proceedings of the 2017 Computation+Journalism Symposium.

Rony, M.M.U., Hassan, N., & Yousuf, M. (2017). *Diving Deep into Clickbaits: Who Use Them to What Extents in Which Topics with What Effects?*. Proceedings of 2017 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining.

Russell, S., Dewey, D., & Tegmark, M. (2015). *Research Priorities for Robust and Beneficial Artificial Intelligence*. Association for the Advancement of Artificial Intelligence: 105–114.

Shotter, J. (1997, May). *Artificial Intelligence and the Dialogical*, American Behavioral Scientists, 40(6): 813–827.

Shu, K., Sliva, A., Wang, S., Tang, J., & Liu, H. (2017, September 3). *Fake News Detection on Social Media: A Data Mining Perspective*. Retrieved from https://arxiv.org/pdf/1708.01967.pdf

Su, D., Zhang, H., Chen, H., Yi, J., Chen, Pin-Yu., & Gao, Y. (2018). *Is Robustness the Cost of Accuracy? A Comprehensive Study on the Robustness of 18 Deep Image Classification Models.* Retrieved from http://openaccess.thecvf.com/content_ECCV_2018/html/Dong_Su_Is_Robustness_the_ECCV_2018_paper.html

Vlachos, A., & Riedel, S. (2014). *Fact Checking: Task Definition and Dataset Construction*. Proceedings of the ACL 2014 Workshop on Language Technologies and Computational Social Science:18–22.

Wang, W.Y. (2017, May 1). *Liar, Liar Pants on Fire: A New Benchmark Dataset for Fake News Detection.* Retrieved from https://arxiv.org/pdf/1705.00648.pdf

Wood, C. (2016, August 1). *What Is Artificial Intelligence? Government Technology*. Retrieved from http://www.govtech.com/computing/What-Is-Artificial-Intelligence.html

Zhao, Z., Resnick, P., & Mei, Q. (2015). *Enquiring Minds: Early Detection of Rumors in Social Media from Enquiry Posts*. Proceedings of the 24th International Conference on World Wide Web: 1395–1405.