Artificial Intelligence against disinformation: the FANDANGO practical case*

Francesco Saverio Nucci[1], Silvia Boi[1] and Massimo Magaldi[1]

[1] Engineering SpA, Viale dell'agricoltura, 24 00144 Rome, Italy
francesco.nucci@eng.it
silvia.boi@eng.it
massimo.magaldi@eng.it

Abstract. The present paper discusses how Artificial Intelligence can support the fight to disinformation to support a correct access to the news and content to the citizens, allowing the right democratic participation. Even if automatic detection of Fake News and disinformation is not possible for the moment and not in the intention of the authors, Machine Learning technologies and Big Data analysis can strongly support journalists and media professionals to detect disinformation in their day-by-day working activity. The paper presents some results of a running EU co-funded project, named FANDANGO, describing its technological approach and architecture. In the first and second chapters the context of disinformation is presented, in chapter 3 and 4 the FANDANGO project is shortly described, including its AI approach and dataflow architecture, chapter 5 describes the project use cases: climate change, immigration, and European policies. Finally, some short conclusions conclude the paper with general considerations on the status of digital media and with some preliminary suggestions to enforce the media in European ecosystem.

Keywords: Artificial Intelligence, Disinformation, Media promoting participation.

1 Introduction

The present paper discusses the possible use of Artificial Intelligence based methodologies, tools and services to fight disinformation. The recent Covid-19 global pandemic, followed by the so called "infodemics" of global disinformation stresses the urgent need to promote in the European Union the education and the integration of the next generation of researchers, journalists, and citizens addressing disinformation and misinfor-

mation by an integrated scientific approach. There is a potential trade-off in using automated technology to curb disinformation online without setting at helm a seasoned, diverse team of social and computer scientists, journalists, educators, fact checkers and new media content providers.

Fake News are now a hot issue in Europe as well as worldwide, particularly referred to Political and Social Challenges that reflect in business as well as in industry [1]. Europe is lacking of a systematic knowledge and data transfer across organizations to address the aggressive emergence of the well-known problem of fake news and post-truth effect. The possibility to use cross sector Big Data management and analytics, along with an effective interoperability scheme for all our data sources, will tackle this urgent problem, generating new business and societal impacts involving several stakeholders: a) Media Companies: news agencies, broadcaster, newspapers, etc, b) Governmental institutions and organizations, c) The overall industrial ecosystem, d) The entire society.

The idea underlying this paper is that it is possible to tackle this aggressive emergence of fake news, post-truths, and disinformation, providing online web app and services that will support media professionals with some high-level features, such as: automatic misinformation detection and trustworthiness scoring, based on Big Data analysis techniques (ML models and Graph Analysis), or tools to support user data investigation, through an interactive exploration of news, open data and verified claims databases.

This hypothesis has been investigated and validated by the authors in a dedicated R&D project, co-funded by the European Commission and named FANDANGO [2].The FANDANGO project stands in line and realises the core implementation of the Commission Action Plan on tackling online disinformation together with the other undergoing initiatives: the European Observatory on Social Media and Disinformation (SOMA project [3]) and the European Digital Media Observatory (EDMO), funded in the CEF programme and the support to research projects building a wide and vivid research community around those fundamental issues. EDMO monitoring of the digital media ecosystem will provide relevant real-time information on the evolution of the disinformation phenomenon. All these initiative demonstrate the strong interest in the European Commission for a better investigation and use of advanced digital technology in fighting disinformation to protect and support our society in a better access to the information to ensure the right citizen participation to the democratic process [4].

2 The context

Discovering disinformation rapidly, effectively, almost in real time is one of the most needed, and most complex, challenge facing social media in the EU and globally. The abundance of data, from different sources (news media companies, journalists, prosumers, commenters etc.) and at the same time the partial view of data owned by platforms, makes it extremely difficult to filter-out and discriminate the valid information from

the false one and the intentionally wrong [5]. In addition, it is now clear how some non-EU countries are maliciously using this phenomenon, in a weaponized manner, to influence our life in one of the most sensitive and critical aspects: the democratic process.

In addition, the processing time is critical, since the detection of suspicious content should happen while the news post is still "live", otherwise it will manage to go viral on the web and to move to traditional media.

Nowadays, the process of retrieve, correlate and assess data from various data sources to timely discover disinformation requires an increasing “investigative” effort that can’t be afforded by single organizations and requires automatic support.

It is important to stress how the automatic detection of disinformation problem needs the aggregation of a multidisciplinary scientific community from Artificial Intelligence to Complex Networks analysis [6, 7], from video and image processing to Cross media exploration [8, 9], from Social Science and Humanities [10] to Natural Language Process [11], in addition, this must be integrated, in the future, also with Research Infrastructures and High Performance Computing and new calculation procedures.

3 The FANDANGO project

FANDANGO is a European co-funded project started in January 2018 and run by a consortium of 8 members from 5 European countries. Goal of the project is design and develop AI-based tools and services to support media professionals in discovering Fake News and disinformation and in this way supporting and improving the proper citizens participation to the democratic debate and process.

FANDANGO results are aimed at professionals, who evaluate news and claims to identify and disprove misinformation and disinformation in different market segments that we will identify in the following section. However, it is important to stress that for professionals, the expert evaluation and final judgement will never be substituted by an automatic decision. The value of FANDANGO results lies in the capabilityto effectively support the human evaluation process by identifying clues of potential misinformation and by facilitating the access to relevant and reliable data.

To offer this value proposition FANDANGO partners will provide an IT platform (likely offered as on-line service - SaaS, even though on premises installations cannot be ruled out at the moment) able to effectively support the human professionals by providing them the two groups of features mentioned above:

1. Detection and scoring of clues of potentially misleading content,
2. Support in the analysis of reliable data related to the claim under scrutiny, through interactive exploration of official reliable open data sources and databases of verified claims.

This last group of features will be heavily influenced by the specific domain of knowledge of interest to which the FANDANGO results will need to be somewhat tailored; this aspect is currently undergoing research exploration. The first group of features, on the other hand is somewhat more consolidated, and is domain agnostic to a larger extent. In fact, this group of features will offer a *trustworthiness* scoring, based

on Big Data analysis techniques (Machine learning models in particular). Specifically, a set of different separate scores will be computed by analysing different component of a specific news, i.e. text, authors, source, media.

In short, to check the trustworthiness level of an online article, user specifies the content to be analysed by providing its URL to the FANDANGO web app, which will then provide - in a reasonable response time - a set of scores, one for each relevant "fakeness" clue:

- "fake" writing style (on the basis of Natural Language Processing techniques),
- manipulation in the associated media (video and images analysed by adequately trained machine learning models),
- out of context video and images (video or images untampered originals but used out of their original context),
- authors and source credibility,
- an overall metrics combining all the criteria/scores above.

4 The architecture and AI approach

To implement the features described in the previous section, it is necessary to monitor new information published on internet, analysing and classifying their content providing a set of trustworthiness scores relative to the different component of a news - i.e. text, authors, source, media.

Since information content on the web is created continuously and at a high speed rate, monitoring selected sources over time, detecting and scoring clues of potentially misleading content, acting as an early warning system when potential disinformation is published online, implies the access to a huge amount of data.

To manage the effective ingestion, processing and analysis of high volume of data, the FANDANGO architecture design is based on a Big Data approach. More specifically, to meet the functional requirements of FANDANGO we implemented a streaming architecture to effectively process the continuous flows of data published on the web.

One of the core component of FANDANGO architecture is Kafka [12], a distributed event streaming platform that support high-performance data pipelines and streaming analytics. Data storage is based on Elasticsearch [13], a scalable, distributed, RESTful search and analytics engine that is also at the core of the Data Investigation features.
A diagram of the streaming Data flow of FANDANGO is depicted in Figure 1.

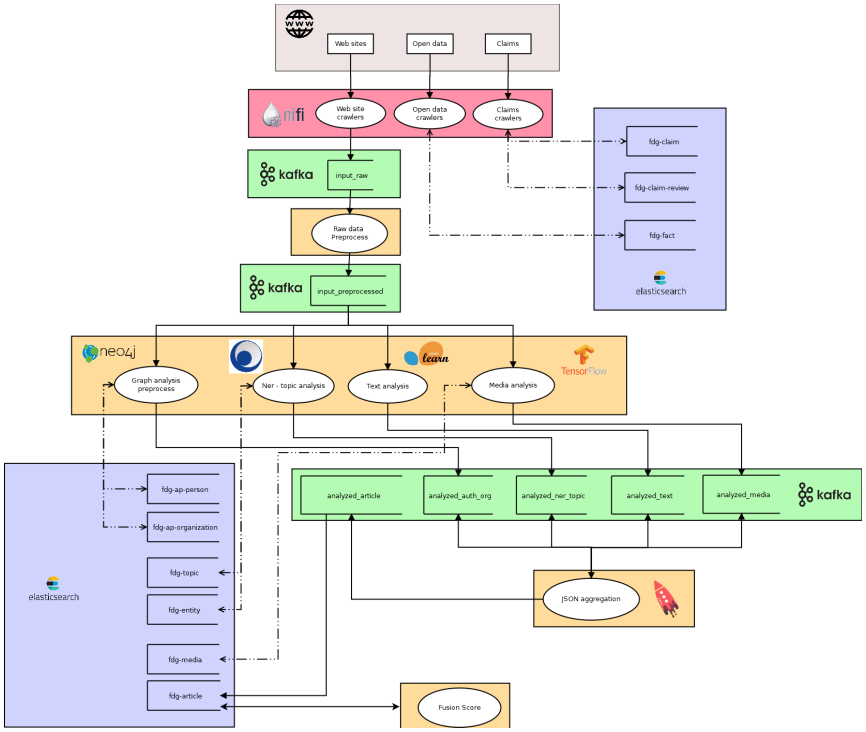

Figure 1- FANDANGO Data flow

This dataflow model, associated with the use of big data tools, ensures not only efficient computation of large amounts of data, but also good horizontal scalability when the workload undergoes variations. We have chosen a Cloud IT infrastructures to achieve the advantages described and Kubernetes [14] - an open-source system for automating deployment, scaling, and management of containerized applications – to orchestrate all components. The detection and scoring of clues of potentially misleading content is based on Artificial Intelligence (AI) advanced processing and analytics methods to analyze different content types, from text and social graphs to image and video content, such as:

- Spatio-temporal analysis and contextualisation of news posts
- Multilingual solutions for analysing misleading posts
- Detection of forgery on images and videos
- Evaluation and scoring of credibility of the news sources through profiling
- Machine learning approaches to automatically weight and score fakeness of news posts.

For each one of this tasks, different analysis approaches have been adopted using several Machine Learning (ML) and Graph Analysis state-of-the-art algorithms and techniques. These analysis methods have been embedded in a set of independent microservices modules to analyze text, multimedia content such as videos and images and finally, metadata content:

- A Text Analysis service based on Natural Language Processing by considering the article body
- A Topic Extraction service based on Named Entity Recognition (NER) techniques by analysing the article body
- A Multimedia analysis service based on image and video analysis
- A Source-credibility service based on graph analytics techniques.

More specifically:

- The text analyzer service gets the content body of the document, analyses it via advanced Natural Language Processing (NLP) procedures and provides a trustworthiness score of the article content [15];
- The topic extractor service performs a NER procedure to retrieve the main topics of the content, that are used to calculate trustworthiness scores (e.g. automatic detection of out-of-context images) and to support users in interactive data investigations.
- The multimedia analyzer service attempts to detect different types of manipulations (such as DeepFake [16, 17, 18]) in the set of images and videos associated to the document using powerful neural networks.
- The source-credibility service collects the information related to publishers, authors, articles and topics and generates a graph to connect the different entities involved in the process. In particular, this service provides a trustworthiness indicator for both publisher and authors by computing a set of centrality methods [19] to measure the impact of each node in the network together with a set of metrics based on the connection of the nodes to indicate the polarity of the impact (level of trustworthiness).

Finally, all ingested news and relative scores can be further investigated by the users in an interactive manner, using the advanced analytics features implemented in FANDANGO, fusing knowledge graph and BI analytics on top of Elasticsearch.

5 The FANDANGO use cases

In order to validate and test the FANDANGO results three main use cases have been analysed and implemented: climate change, immigration, the European policies. These use cases give an overview on how the FANDANGO services can be used with regards to fact checking of images, claims, articles & videos.

Climate: When contesting Climate statements, it has become a habit to attack the research that led to the facts instead of the facts themselves. While this seems like an easy way to challenge any kind of statement, in case of climate based ones they often have

a point. Climate statements often don't take the whole picture into account, but the reason for that isn't always foul play. The requirement to be brief and to the point often causes some bad decisions when it comes down to formulating a fact. And of course this is all that is needed to claim the whole fact as being false. A better integration with true data is needed, it is clear to everyone that to validate facts you need data.
From a journalistic point of view, the job of fact-checking not only the content, but also the context in which a statement was made becomes very hard, especially at the rate new information is being produced.

Immigration: Populist discourses, us-versus-them worldviews, hate speeches against those who are different" enjoying a widespread and often growing audience in Europe -partly explained by mass media's functioning mechanisms-, certainly wider than more nuanced approaches to an issue. It is demanding, after all, to have an informed opinion, and there are always those eager to offer a simpler biased perspective for their own political or economic benefit. This is a challenge shared by all European countries, especially during times of economic crisis. The Fandango immigration use case focuses on:

- Gather, standardise, integrate and make available reliable data, creating a factual integrated data silo that can be easily and quickly used to counter fake news on immigration.
- Analyse existing European data from opinion polls and barometers (e.g. Eurobarometer) to investigate how the population opinions and feelings match (or not) the actual facts. Do people perceive the immigration reality as it really is?
- Combining both sources of data (factual reality and perception), the pilot created journalistic material including graphic and data visualizations showing, at a single glance, the gap between myths and reality, so that they can be easily viralizable on their own via social networks. Evidence and data based memes against fake memes.

European policies: The last, but not least use case impacts the European Context: last years Europe is, in a certain way under attack from main point of views and from many directions, Fake News play a malicious role in the attack. The grounding idea of this scenario is the aggregation of data around European community, citizens, budget, and so on in order to defend Europe itself from Fake News or claims that every day are spread from many different channels. A pilot for this use case include fact checking of news stories and political statements on active topics. The goal is to debunk false and partially true statements, to quickly separate fact from propaganda, and to bring subtleties in stories back into view for the media consumer.

6 Future steps and conclusion

The disinformation problem is even more urgent after the explosion of COVID pandemic, demonstrating the importance to have the right scientific information in an

emergency, in addition it has been already demonstrated how the democratic process can be jeopardized by a improper access to the digital information and content by the citizens, the FANDANGO project was started three years ago, but the situation is not now changed, however it demonstrate how the AI can be used in the context for Good, supporting journalist and media professional in discovery fake news and malicious information. In the same time, it demonstrates also how long it is the way on an automatic approach that cannot in any case be foreseen. With respect to the European Ecosystem it should be stressed how the EC already started many initiatives and R&D projects, but even if in Europe several initiatives and investments have been proposed and realised, it is still missing a common European media environment where the media sector is equipped of a powerful set of models and visual interactive tools for data journalism and investigative journalism to support in determining news authentication evaluations. This is not only to detect disinformation but to tackle malicious information, hate speeches and aiding and abetting terrorism actions.

This environment should combines well-defined business strategy to pursue a strong market position for the concept of content-centric trusted information for professional users and technical strategy to boost the deployment of technologies such as AI-based solutions and services for example the machine learning and the semantic technologies, enabling a larger user community to reap the economic benefits from such environment, especially SMEs and non-technology sectors (such as media companies, social media expert, data journalists…) in detecting disinformation through the provided solutions. In this way, also the fight of disinformation can be approached and supported in a more general context environment with positive results and effective achievements: the papers authors have already started to work in this direction.

References

1. Gangware, C., and Nemr, W.: “Weapons of Mass Distraction: Foreign State Sponsored Disinformation in the Digital Age” Park Advisors (2019).
2. FANDANGO project, https://www.fandango-project.eu, last access 2020/08/07
3. SOMA project, https://www.disinfobservatory.org, last access 2020/08/07
4. Action Plan on disinformation: Commission contribution to the European Council, 13-14 December 2018, European Commission (2018).
5. Rubin, V. L., Y. Chen and N. J. Conroy.: Deception detection for news: Three types of fakes. Proceedings of the Association for Information Science and Technology 52. 1–4 (2015).
6. Ciampaglia, G. L.: Fighting fake news: a role for computational social science in the fight against digital misinformation. *Journal of Computational Social Science*. 1 (1), pp. 147-153 (2017).
7. Caldarelli, G., De Nicola, R., Del Vigna, F., Petrocchi, M., Saracco F.: The role of bot squads in the political propaganda on Twitter arXiv:1905:12687, in press Communication Physics (2020).
8. Li, Jian, et al.: "Segmentation-based image copy-move forgery detection scheme." IEEE Transactions on Information Forensics and Security 10.3, pp. 507-518 (2015).
9. Rao, Y., Jiangqun N., and Huimin Z.: "Deep Learning Local Descriptor for Image Splicing Detection and Localization." IEEE Access 8 pp. 25611-25625 (2020).

10. Zheng, Y., Mobasher, B., and Burke, R.: Emotions in Context-Aware Recommender Systems. In Emotions and Personalized Services, pp. 311–326 (2016).
11. Rehm G.: An Infrastructure for Empowering Internet Users to Handle Fake News and Other Online Media Phenomena. In: Rehm G., Declerck T. (eds) Language Technologies for the Challenges of the Digital Age. GSCL 2017. Lecture Notes in Computer Science, vol 10713. Springer, Cham (2018).
12. Apache Kafka, https://kafka.apache.org/, last accessed 2020/08/07
13. Elasticsearch, https://www.elastic.co/elasticsearch/, last accessed 2020/08/07
14. Kubernetes, https://kubernetes.io/, last accessed 2020/08/07
15. Rehm, G., Schneider, J. M., & Bourgonje, P.: Automatic and manual web annotations in an infrastructure to handle fake news and other online media phenomena. In Proceedings of the Eleventh International Conference on Language Resources and Evaluation (2018).
16. Dolhansky, B., Howes, R., Pflaum, B., Baram, N., and C. C. Ferrer,: "The deepfake detection challenge (dfdc) preview dataset," arXiv preprint arXiv:1910.08854 (2019).
17. Amerini, I., Galteri, L., Caldelli, R., and Del Bimbo, A.: "Deepfake video detection through optical flow based cnn," in Proceedings of the IEEE International Conference on Computer Vision Workshops, (2019).
18. Tolosana, Ruben, et al.: "DeepFakes and Beyond: A Survey of Face Manipulation and Fake Detection." arXiv preprint arXiv:2001.00179 (2020).
19. T. Agryzkov, L. Tortosa, J. F. Vicent, and R. Wilson, "A centrality measure for urban networks based on the eigenvector centrality concept," Environ. Plan. B: Urban Anal. City Sci. 46, 668–689 (2019).