Context Aggregation and Analysis: A Tool for User-Generated Video Verification

Olga Papadopoulou
CERTH-ITI
Thessaloniki, Greece
olgapapa@iti.gr

Dimitrios Giomelakis*
School of Journalism and Mass Communications, Aristotle University
Thessaloniki, Greece
dgiomela@jour.auth.gr

Lazaros Apostolidis
CERTH-ITI
Thessaloniki, Greece
laaposto@iti.gr

Symeon Papadopoulos
CERTH-ITI
Thessaloniki, Greece
papadop@iti.gr

Yiannis Kompatsiaris
CERTH-ITI
Thessaloniki, Greece
ikom@iti.gr

ABSTRACT

The uncontrolled dissemination of User-Generated Content (UGC) through social media and video platforms raises increasing concerns about the intentional or unintentional spread of misleading information. As a result, people who are turning to the Internet for their daily news, need tools that help them distinguish between reliable and unreliable content. Here we present the Context Aggregation and Analysis tool, with the aim to facilitate the investigation of the veracity of User-Generated videos (UGVs). The tool collects and calculates a set of verification cues based on the video context, that is the information surrounding the video rather than the video itself, and then creates a verification report. The cues include information about the video and user that posted it, as well as the activity of other users surrounding it (what we call "wisdom of the crowd"), cross-checking with previous cases of fakes ("wisdom of the past"), and employing machine learning systems trained on past cases of real and fake videos ("wisdom of the machine"). We evaluate the tool in two ways: i) we carry out a user study where end users are manually assessing the tool's features on a set of UGVs from a real-world dataset of news-related videos, and ii) we quantitatively evaluate the automatic verification component of the tool. The tool assisted successfully with the debunking of 132 out of 200 fake videos, the verification of 142 out of 180 real videos and the performance of the classifiers reached an F-score of 0.72.

KEYWORDS

video verification, user-generated content, verification tool, machine learning, online disinformation

1 INTRODUCTION

The spread of Internet-connected devices (smartphones, tables, laptops) enables bystanders to share content about newsworthy events as they unfold. During critical events (e.g., elections, natural disasters, etc.), a massive amount of UGC is created and spreads on the Web. Information in the form of text, images and videos circulates through social media (e.g., Facebook, Twitter) and well-known image and video sharing platforms (e.g., Instagram, YouTube) with minimal control, providing a source of information for breaking news that otherwise would be inaccessible to news organizations and other users, but at the same time it poses risks of mis- and disinformation. Several cases of unreliable content appeared in recent years with grave consequences. For example, a large-scale spread of disinformation during the 2016 US presidential election could have influence the election result [1, 28], while incidents of lynchings in India were caused by rumours over WhatsApp and led to the loss of human life [10, 20]. Several instances of disinformation are included in a recently published dataset of debunked and verified UGVs called the Fake Video Corpus (FVC-2018) [21].

Online tools have been proposed over the last few years to help users verify online content in an automatic or semi-automatic way. However, these tools (cf. Section 2) do not offer adequate support for coping with the ever-growing amount of information that needs examination, which makes necessary the development of novel features for verification. To this end, we present the Context Aggregation and Analysis tool[1] with the aim to assist journalists and citizens to verify whether a video on YouTube, Facebook or Twitter is credible or not. The features that it collects are based on the context that surrounds the video and derive directly from the platform API and Twitter shares (tweets sharing the video URL). The collected information is fed to the tool and a verification report is created, that leverages the "wisdom of the crowd" (e.g. user comments that could be helpful for verification), " wisdom of the past" (whether the video matches a known case) and "wisdom of the machine" ("fakeness" score produced by trained model), and presented to the end user.

The tool has been evaluated through a small user study where the majority ($> 70\%$) of the examined videos were successfully debunked (in about three minutes per video) or verified leveraging one or a combination of the provided verification features. Additionally, a set of experiments has been conducted to quantitatively evaluate the trained classifiers reaching an F-score of 72%.

*Also with CERTH-ITI, Thessaloniki, Greece.

[1] https://caa.iti.gr/

ROME 2019, July 25, 2019, Paris, France

2 RELATED WORK

There are several types of disinformation making the problem of multimedia verification very diverse. In some cases we are dealing with tampered content, in which case multimedia forensics algorithms are used to solve the problem [6, 31, 32]. In other cases the content is genuine but is published with false contextual information. In these cases we have to rely on external knowledge and try to find inconsistencies in the contextual characteristics of the post, or -in the case of reposting old unrelated content- to locate the original post.

Due to the significance of the problem, several public challenges dealing with disinformation have been organized and have attracted interest and participation, leading to the development of a multitude of verification methods. The 'Verifying Multimedia Use' benchmark task, which took place in MediaEval 2015 [3] and 2016 [4], focused on the automatic classification of multimedia Twitter posts into credible or misleading. Moreover, a grassroots effort of over 100 volunteers and 71 teams from academia and industry organized the Fake News Challenge[2] in 2017. Finally, the International Workshop on Semantic Evaluation (SemEval) has introduced the SemEval-2017 Task 8 'RumourEval: Determining rumour veracity and support for rumours' [12] and the SemEval-2019 Task 4 'Hyperpartisan News Detection' [15].

Reference guides for verification aim to provide a structured framework consisting of checklists and tutorials attempting to standardize the verification process. The Verification Handbook by the European Journalism Centre [24] provides the tools, techniques and step-by-step guidelines on how to deal with UGC during emergencies and is available in English, Greek, Spanish, Arabic, and other languages. Google News Lab has announced its Fact Check feature [17] along with a set of tutorials on how to use Google tools for verification [11]. The Bellingcat investigative organization provides its advanced guide on verifying video content [26] and suggests the Amnesty International's YouTube DataViewer [14] as an easy tool to conduct easy reverse search on videos using thumbnails. Besides Google reverse image search[3], other similar tools exist, such as TinEye[4].

Verification tools leverage the content and metadata of images attached to news articles or social media posts to help users make a decision. A semi-automatic approach [8] uses image and text clustering techniques for verifying images and through them the corresponding online news stories. Fakebox [27] is a tool developed for verifying news articles by analysing the title, content and domain of the article. Relevant Tweet verification tools using contextual information include TruthNest [2] and the Tweet Verification Assistant[5]. The InVID verification plugin [25] is a browser extension which aggregates verification-related information and creates shortcuts to multimedia analysis components in order to assist journalists in their verification efforts. Concerning rumour analysis, Hoaxy [23] focuses on the social dynamics of online news sharing, while RumourFlow [7] tries to expose rumour content and the activity of the rumour participants by sophisticated visualizations.

[2]http://www.fakenewschallenge.org/
[3]https://www.google.com/imghp
[4]https://tineye.com/
[5]http://reveal-mklab.iti.gr/reveal/fake/

With regards to automatic verification approaches, proposed algorithms rely on extracting characteristics of the text surrounding the multimedia item and the information about the poster of the item [13, 33]. The work of [5] proposed two types of feature, tweet-based and user-based, which are used to classify tweets as real or fake. A comprehensive survey of approaches dealing with false information is presented in [18]. The authors group the works based on the platforms they study and the characteristics and features they rely on. They analyse the challenge from different perspectives i.e. the types of mechanisms that are used for spreading false information, the intent and knowledge content of the false information, etc. Finally, recent works rely on more sophisticated models such as the work of [29] where a hybrid Convolutional Neural Network (CNN) is proposed to integrate metadata with text, showing promising results on the problem of fake news detection.

3 CONTEXT AGGREGATION AND ANALYSIS

The proposed tool's verification pipeline is illustrated in Figure 1. The starting point for the analysis is a YouTube, Facebook or Twitter video[6]. Then, the tool produces a verification report, which is progressively made available to the end users giving them the opportunity to start the investigation without waiting for the whole process to complete. To this end, the tool connects to the Platform APIs and requests publicly available data for the target video and the account who posted it. With respect to Facebook, three types of account (User, Page and Group) can share videos but only videos posted by a public Facebook Page can be analysed by the CAA tool due to API restrictions. The amount of data returned by the APIs for a given video is often large. This raised the need for conducting a careful analysis in order to keep only the information that is helpful for verification. The collected data pass through the tool's internal components, as shown in Figure 1, resulting in the verification report, which is structured in sections to make the presentation of results easier to digest.

General. This section features information that is collected directly from the APIs and is selected by the CAA tool, keeping only information that is considered to be relevant for verification. Figure 2 illustrates the indicators that constitute part of the verification report and refer to the characteristics of the video (a) and the user/channel (b) that posted it. The indicators are grouped based on whether they are available across all three platforms or only in one of them.

Among others, a significant indicator refers to the time when the channel/user was created compared to the time that the video in question was posted. Recently created channels posting sensational videos create doubts about the authenticity and credibility of their videos. For instance, a viral video of a girl being chased by a bear while snowboarding was posted five days after the channel was created (#3 of Table 2). The video gained millions of views before it was debunked [9].

Apart from the verification cues, which derive directly from the social media APIs, a significant contribution to the verification report comes from indicators that are computed by the tool. The *Average number of videos per month uploaded by the channel* is a

[6]For Twitter both Native Twitter videos and tweets containing a link to a YouTube or a Facebook video are supported.

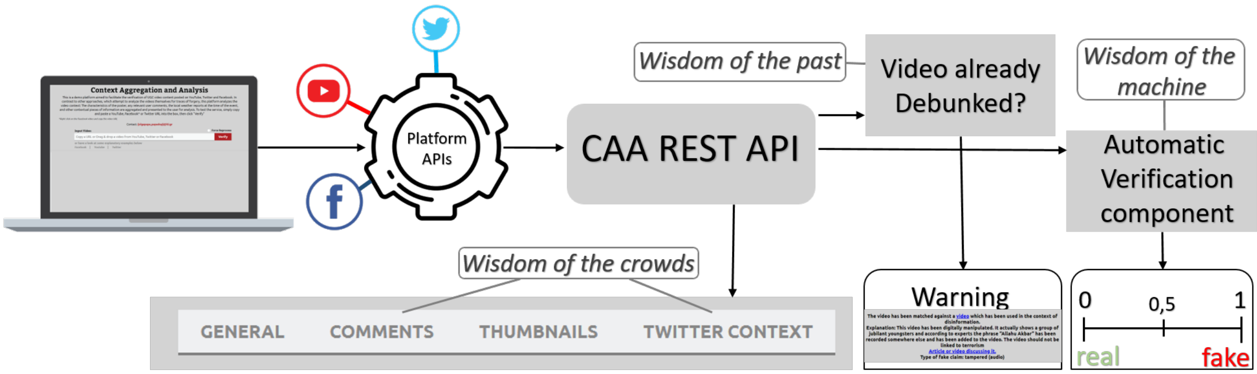
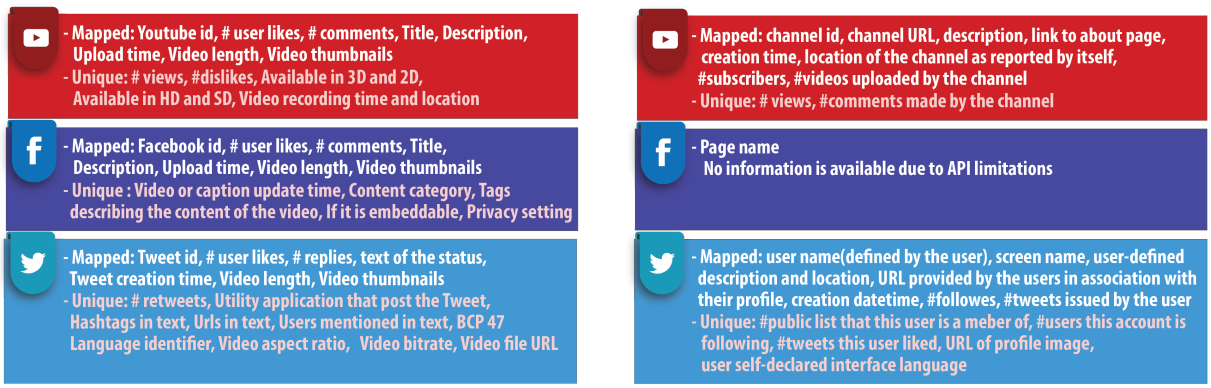


Figure 1: Context Aggregation and Analysis service pipeline.



(a) Video verification indicators

(b) User/Channel verification indicators

Figure 2: A number of indicators are mapped to all three platforms and the rest are unique in one of them.

feature capturing the activity volume of the channel and is calculated by dividing the total number of posted videos to the channel age (number of months that the channel is online). The average number of videos per month for real videos (0.018) is considerably larger than that of fake ones (0.0019)[7].

Locations mentioned: Posting a video and claiming that it was captured at a location other than the actual one is a common case of disinformation. CAA submits a query to Recognyze [30] with the title and description of the video in question and it presents the returned location-related named entities. Recognyze searches and aligns locations with established knowledge bases such as GeoNames and DBpedia, and refines them by exploiting structure and context to solve abbreviations and ambiguities.

Twitter search URL: This is an automatically generated query URL that can be submitted to Twitter search in order to retrieve tweets that contain a link to the submitted YouTube or Facebook video.

Tweet verification. A feature that is applied only for Twitter videos is a verification label (fake/real), extracted by the Tweet Verification Assistant API [5] indicating the veracity of the tweet based on the tweet text and the user that posted the tweet.

Wisdom of the crowd. This set of verification cues refer to the external knowledge around the event derived by user comments and tweets. Users often leave *comments* below videos and such comments often contain useful cues for verification. They may express their personal opinion or experience in relation to what the video shows, convey public statements about the event shown in the video, and often challenge or support the credibility of the video. Under the *comments* section, a tab contains all comments (replies in case of Twitter) on the video showing the comment text, author and date of creation[8]. To assist with the verification, a subset of the comments, called *verification comments*, is automatically retrieved by filtering them with a list of predefined verification-related keywords currently in six languages (German, Greek, Arabic, French, Spanish and Farsi)[9]. For example, the video in Figure 3 claims that a young Syrian boy is rescuing a girl amid gunfire; a verification-related comment retrieved by the tool explains that the video is staged [19].

The verification comments have proven very useful for the verification process, but there are cases where user-defined keywords

[7] The statistics are based on analysis of the 380 videos of the FVC-2018 [21].

[8] For Facebook, only the comment text is available.

[9] In English, the keywords are 'fake', 'false', 'lie', 'lying', 'liar', 'misleading', 'propaganda', 'wrong', 'incorrect', 'confirm', 'where', 'location'.

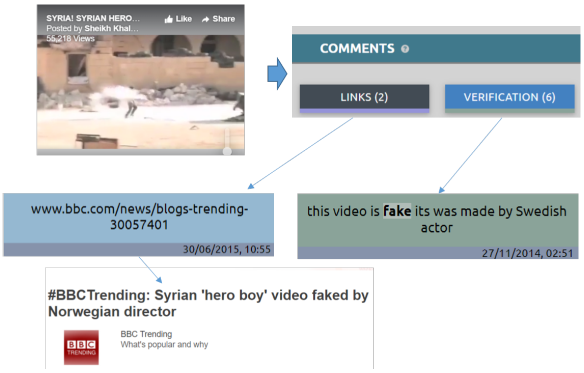


Figure 3: Exploiting the 'links' and 'verification-related comments' features for debunking a staged fake video.

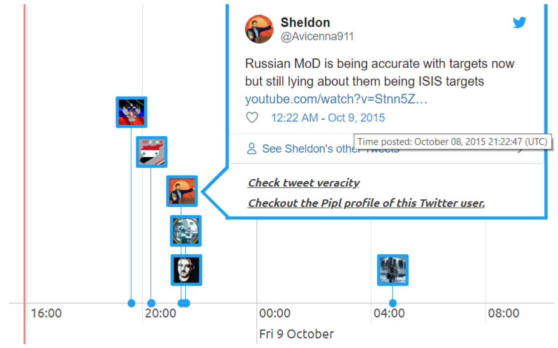


Figure 4: Twitter timeline. A tweet is posted couple of hours after the video was shared on YouTube (red line) explaining that the claim of ISIS being the target of the bombing is false.

may lead to additional valuable cues. The tool enables the user to create a new subset of comments filtered by keywords of their choice that are combined with boolean AND/OR operators. Finally, comments often contain *links* to other sources (videos, articles, social media posts), which sometimes provide valuable information, such as articles that debunk the video in question. In the above example of Figure 3, a comment containing a link to a reputable news source's article is provided below the 'Links' tab and points to an article explaining that the video was shot by a professional film maker.

With respect to *Twitter Context*, an additional aggregation step is triggered for each submitted video to collect the tweets that contain a link to the input video and use them to generate a Twitter timeline. The tweets are presented in temporal order (oldest to newest) as illustrated in Figure 4 and they can result in further useful indicators using existing online tools for tweet verification, such as the Tweet Verification Assistant [5]. With respect to Twitter videos, the retweets of the submitted tweet are similarly used.

Video Thumbnails. The video thumbnails are retrieved directly by each platform API. With respect to YouTube and Twitter, the number of keyframes is fixed, while for Facebook it varies. Below the thumbnails, there are buttons labelled 'Google' and 'Yandex', which trigger a query to the corresponding Image search engine with the video thumbnail leading to a page with the reverse image search results. In cases where the video under consideration is a repost of a previously published video, but someone is claiming that it was captured during an unfolding event, reverse image search makes it possible to retrieve the original video and debunk the reposted instance. Moreover, articles or other videos debunking the video may appear in the results, which could also offer valuable clues. Additionally, the tool checks whether the main video thumbnail (the one that is shown when the video appears in lists and search results) exists within the video. This is done by applying a near-duplicate detection algorithm [16] to compare the main thumbnail with the video frames. If it turns out that the thumbnail is not part of the video, the video is labelled as 'possibly clickbait', since this is a common practice among video publishers to draw the attention of video viewers and mislead them into clicking on the video.

Wisdom of the past. The tool includes a verification feature that prevents users from falling again for known cases of 'fake' videos, which were already debunked by reputable sources (*Video Already Debunked?* component in Figure 1). To this end, the FVC-2018 dataset is used as a *background collection* of debunked and verified videos. For a given video, the near-duplicate algorithm of [16] is used to search in the pool of FVC-2018 videos. If there is a match, the CAA returns a warning following some rules: i) if the matched video is unique then the video is directly returned, ii) if the matched video has near-duplicates, the tool selects the earliest video among all near-duplicates, iii) if the matched video is earlier than the submitted one but has been removed from the video source and is not available online, the video metadata of the removed video along with URLs of other near-duplicate instances (if they exist) are returned, and iv) if the matched video is later than the submitted one, a message that the submitted video is either the original one, which was later reused to mislead or it is a near-duplicate that retains the false claim but is not part of the background collection is presented.

Wisdom of the machine. A "fakeness" score is calculated by a model trained on the FVC-2018 and following the video-based verification approach of [21]. The score ranges from 0 to 1: the higher its value, the more likely the video is misleading or inaccurate. Section 4 includes a description of the automatic verification approach and evaluation on the FVC-2018 dataset.

Apart from the collected and calculated verification features, a challenge that reporters face when investigating a video is the multiple languages in which the content is shared. To help users save time by eliminating the need for using external translation services, we integrated an automatic translation feature[10], which gives the user the option to translate on demand the text of interest (title, description, comments) into English.

4 EVALUATION

4.1 User Study

We carried out a small user study to evaluate the CAA tool on the tasks of debunking the 200 fake and verifying the 180 real videos

[10] https://cloud.google.com/translate/

Table 1: Number of fake videos per verification outcome and time needed for the debunking task.

Is Debunked	# videos	Time (seconds)
True	132	208
False	46	272
Uncertain	22	270

of the FVC-2018. Although this set of videos is modest in size, it reflects a wide variety of both misleading and accurate videos, which makes it an appropriate sample to evaluate verification tools. Moreover, collecting and annotating considerably more videos would require an overwhelming amount of time. The study was carried out by two of the authors of the paper, a male with journalistic background and a female with computer engineering background. Both users have knowledge and experience in video verification and specifically in debunking videos that disseminate disinformation through social media or video platforms. The users followed the following procedure and recorded the results: i) submit a video URL to the tool, ii) check and analyse the produced verification report, iii) decide about the video veracity among three labels (True, False, Uncertain), and iv) record the results and the time spent on the task. The 'Uncertain' label corresponds to cases where there are indicators that create doubts about the video credibility but there is no concrete evidence proving that the video is fake or real.

Table 1 presents the outcomes of the above debunking process and the time needed for the investigations. The majority of videos (~ 70%) were successfully debunked by consulting the verification features of the tool (#1, #2 and #3 of Table 2). Videos labelled as 'Uncertain', i.e. they could not be debunked, but the tool offered some cues that they might be fake (#4 of Table 2), are 22 (~ 10%), while those where the tool did not offer any helpful cues (#5 of Table 2) are 46 (~ 20%). The users needed on average more than three minutes for successfully debunking the videos and more than four minutes to decide that they could not debunk a video with the help of the tool.

For each fake video, one or a combination of verification features were taken into consideration to label the video as real or fake. Specifically, 57 out of 132 videos were debunked by consulting one verification feature and 75 videos needed a combination of two or more features for concluding to a verification label. Figure 5 presents the number of videos where each verification cue assisted during the debunking procedure. The Google reverse image search feature appears most frequently indicating that searching online sources with a keyframe extracted from the video returns useful information for verification (#1 and #3 of Table 2). As expected, this feature is especially helpful for cases where a video posted for a past event is reused for a breaking news event. The keywords of the verification-related list contribute to debunking half of the videos (#1, #2 and #3 of Table 2). However, there are 29 videos that were debunked relying on other comments, which highlights the need of a more sophisticated and effective approach to retrieve verification-related comments (e.g. #2 of Table 2 where comments containing links assisted the verification process).

We should also consider the fact that the videos of the FVC-2018 dataset relate to past events, which means that they have already

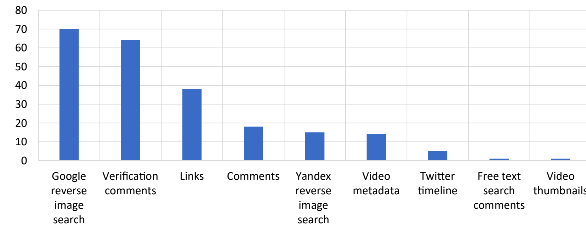


Figure 5: The number of videos for which each cue contributed. 132 of the 200 fake videos are considered, which were successfully debunked using the tool.

been extensively discussed online and, as a result, a large amount of comments and even articles around them already exists online. For videos that appear in the context of breaking news, such online information might be considerably more limited for some time (minutes to hours) after a video is posted.

The features concerning the match to the already debunked videos of the FVC-2018 and the score calculated by the automatic verification algorithm were not considered in the user study as the first would lead to the debunking of all videos (since FVC-2018 is used as the background collection of the tool) and the latter was possible to quantitatively evaluate in an automated way, as presented in the next section.

The challenge of spreading disinformation and how to assist users to recognize whether a video conveys false claims prompted us to focus on the detection of fake videos. However, the behavior of the verification features in real videos and the potential of false positives was also investigated in order to ensure that the tool does not suffer from a high false positive rate (i.e. leading users to flag a real video as fake). With respect to the 180 examined real videos of the FVC-2018, ~ 80% of them were successfully verified relying mostly on reverse image search, which pointed to trusted news sources (#6 of Table 2). The remaining videos, which we were unable to verify (#7 and #8 of Table 2), included videos where the tools could not provide enough information for a definite confirmation (~ 14%), and cases where the verification cues actually cast doubt on their authenticity (~ 6%).

4.2 Evaluation of Automatic Verification

We used the FVC-2018 dataset and the video-based approach of [22] to train a model which classifies UGVs to real or fake. In the dataset, videos are organized in cascades, each of which comprises a first instance of the video and a set of near-duplicates. For each such cascade, a Random Forest classifier was trained with the videos of the rest of the cascades and then used to classify each video of the target cascade. The frequency distribution of the classifiers' outputs is presented in Figure 6, showing that high probabilities of the 'fake' class are assigned mostly to fake videos. However, we note that there are several false positives indicating that videos conveying accurate content may be classified as fake by the tool. Aggregating the prediction scores across all cascades, an F-score of 0.72 (Precision 0.66 and Recall 0.81) is achieved demonstrating that the automatically produced verification class could be considered as a valuable additional cue for verification.

Table 2: Examples of debunking and verifying videos of the FVC-2018.

#	Claim	Truth	Verification feature	Explanation	Label
"Fake" Examples					
1	CNN: Donald Trump Rips Marine's Hat Off After Assaulting Him	The footage is actually being played in reverse.	Verification comments Google reverse image search	The real clip shows Donald Trump picking up the Marine's hat after it blew off in the wind. Several verification comments explain that the video is played in reverse. Google reverse image search points to the original video.	True (FAKE)
2	GoPro: Man Fights Off Great White Shark In Sydney Harbour	This video is staged and part of a viral experiment produced and promoted by RIOT production studio.	Links Verification comments	Comments contain links to trusted news sources which explain that the video is staged and produced as part of an experiment. Several verification-related comments create doubts about the video authenticity.	True (FAKE)
3	Snowboarder Chased By A Bear	This video is staged and part of a viral experiment produced and promoted by RIOT production studio.	Video metadata Verification comments Google reverse image search	The channel that posted the video was created five days before the video was uploaded online and the total number of videos posted by this channel is three. Several verification comments discuss the video with doubts. Reverse image search results point to debunking articles.	True (FAKE)
4	URGENT: ISIS executioner 'Jihadi John' is killed in air strike	The video is from previous event and shows completely irrelevant air strikes.	Comments Verification comments	Several comments and verification-related comments describe the video as fake but no evidence that the video is indeed fake is given.	Uncertain (True Negative)
5	Muslims burn Christmas tree in Belgium	It shows a group of jubilant youngsters and the phrase 'Allahu Akbar' was added to the video digitally.	No verification feature did contribute to the debunking	Comments are disabled for this video. Reverse search did not return any useful information.	False
"Real" Examples					
6	Bloody Syrian boy (5-year- old boy Omran Daqneesh) after an air raid in Aleppo	The footage is real.	Comments Google reverse image search	Comments talk about the video. Reverse image search returns articles from trusted news sources.	True (REAL)
7	Attacks in Paris - During Bataclan Theatre attack - GRAPHIC CONTENT	The footage is real.	Verification comments Google reverse image search	Some comments talk about hoax and false flag creating doubt. Reverse image search returns no trusted news source.	Uncertain (False Positive)
8	Japan earthquake makes skyscraper dancing	The footage is real.	No verification feature did contribute to the verifying	The verification cues did not provide any evidence for verifying the authenticity of the video.	Uncertain

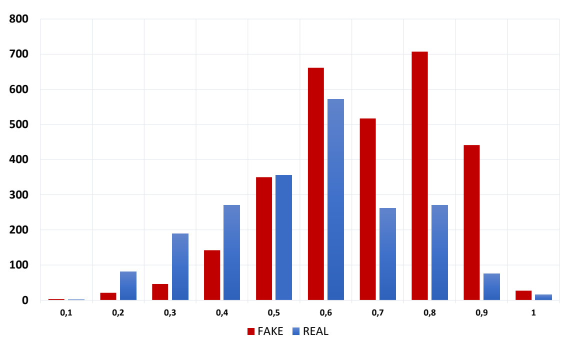


Figure 6: Frequency distribution of 'fake' class probability for the videos of the FVC-2018.

Tool usage. The presented tool has been publicly available from late 2017 until now as a standalone service and as a component of the InVID verification plugin [25]. Our record of usage statistics of the tool demonstrates that the tool is valuable for journalists and citizens, since more than 12,000 unique users, from all over the world (United States, France, India, Saudi Arabia and other countries), have used it to assess the veracity of more than 17,000 unique videos over a period of 15 months.

5 CONCLUSION

In this paper, we presented a tool that supports the verification of UGVs disseminated through social media platforms. The tool leverages information and data that surround the videos and are provided by the platform APIs, exploit public opinion and knowledge, past known cases of fake videos, and an automatic method for labeling the video as real or fake. The tool led to promising results, and several of its features proved to be valuable for the verification or debunking of videos according to a user study we performed on all videos of the FVC-2018 dataset.

During development, we faced the so-called *Walled Garden* issue, i.e. challenges in setting up and maintaining the data collection pipeline of the tool for the supported platforms. This is due to the fact that social media and networking platforms are in fact closed ecosystems, in which all operations are centrally controlled. The platform APIs impose great limitations on the information that is accessible programmatically even when this information is publicly available. The Walled Garden issue results in insurmountable challenges when trying to integrate more platforms and makes it difficult to develop new verification features.

In addition to limiting the potential usefulness and effectiveness of the tool for end users as a result of the limited available information, API limitations also contribute to long response times, especially in cases, where for instance a video is associated with a large number of comments (which are not conveniently available through a single API request).

In the future, we intend to i) conduct a larger user study with more users having different knowledge background, and ii) to improve the performance of the tool, making it easier to interpret by non-trained users.

ACKNOWLEDGMENTS

This work is supported by the WeVerify project, which is funded by the European Commission under contract number 825297.

REFERENCES

[1] Hunt Allcott and Matthew Gentzkow. 2017. Social media and fake news in the 2016 election. *Journal of economic perspectives* 31, 2 (2017), 211–36.

[2] ATC. 2018. TruthNest: Look beneath the surface with deep twitter analytics. https://www.truthnest.com/

[3] Christina Boididou, Katerina Andreadou, Symeon Papadopoulos, Duc-Tien Dang-Nguyen, Giulia Boato, Michael Riegler, Yiannis Kompatsiaris, et al. 2015. Verifying Multimedia Use at MediaEval 2015.. In *MediaEval*.

[4] Christina Boididou, Symeon Papadopoulos, Duc-Tien Dang-Nguyen, Giulia Boato, Michael Riegler, Stuart E. Middleton, Andreas Petlund, and Yiannis Kompatsiaris. 2016. Verifying Multimedia Use at MediaEval 2016. In *Working Notes Proceedings of the MediaEval 2016 Workshop, Hilversum, The Netherlands, October 20-21, 2016*, Vol. 1739. CEUR-WS.org.

[5] Christina Boididou, Symeon Papadopoulos, Markos Zampoglou, Lazaros Apostolidis, Olga Papadopoulou, and Yiannis Kompatsiaris. 2018. Detection and visualization of misleading content on Twitter. *International Journal of Multimedia Information Retrieval* 7, 1 (2018), 71–86.

[6] Sajjad Dadkhah, Azizah Abd Manaf, Yoshiaki Hori, Aboul Ella Hassanien, and Somayeh Sadeghi. 2014. An effective SVD-based image tampering detection and self-recovery using active watermarking. *Signal Processing: Image Communication* 29, 10 (2014), 1197–1210.

[7] Anh Dang, Abidalrahman Moh'd, Evangelos Milios, and Rosane Minghim. 2016. What is in a rumour: Combined visual analysis of rumour flow and user activity. In *Proceedings of the 33rd Computer Graphics International*. ACM, 17–20.

[8] Sarah Elkasrawi, Andreas Dengel, Ahmed Abdelsamad, and Syed Saqib Bukhari. 2016. What you see is what you get? Automatic Image Verification for Online News Content. In *2016 12th IAPR Workshop on Document Analysis Systems (DAS)*. IEEE, 114–119.

[9] Dan Evon. April 2016. Snowboarder Girl Chased by Bear. Retrieved May 6, 2019 from https://www.snopes.com/fact-check/snowboarder-girl-chased-by-bear/

[10] Gowhar Farooq. 2018. Politics of Fake News: How WhatsApp Became a Potent Propaganda Tool in India. (2018).

[11] Google. [n.d.]. Google News Initiative Training Center. Retrieved May 6, 2019 from https://newsinitiative.withgoogle.com/training/

[12] Genevieve Gorrell, Kalina Bontcheva, Leon Derczynski, Elena Kochkina, Maria Liakata, and Arkaitz Zubiaga. 2018. RumourEval 2019: Determining Rumour Veracity and Support for Rumours. *arXiv preprint arXiv:1809.06683* (2018).

[13] Aditi Gupta and Ponnurangam Kumaraguru. 2012. Credibility ranking of tweets during high impact events. In *Proceedings of the 1st workshop on privacy and security in online social media*. ACM, 2.

[14] Amnesty International. 2017. Youtube DataViewer. https://citizenevidence.amnestyusa.org/

[15] Johannes Kiesel, Maria Mestre, Rishabh Shukla, Emmanuel Vincent, Payam Adineh, David Corney, Benno Stein, and Martin Potthast. 2019. SemEval-2019 Task 4: Hyperpartisan News Detection. In *Proceedings of The 13th International Workshop on Semantic Evaluation (SemEval 2019)*. Association for Computational Linguistics.

[16] Giorgos Kordopatis-Zilos, Symeon Papadopoulos, Ioannis Patras, and Yiannis Kompatsiaris. 2017. Near-duplicate video retrieval with deep metric learning. In *Proceedings of the IEEE International Conference on Computer Vision*. 347–356.

[17] Justin Kosslyn and Cong Yu. 2017. Fact Check now available in Google Search and News around the world. *Google Blog* (2017).

[18] Srijan Kumar and Neil Shah. 2018. False information on web and social media: A survey. *arXiv preprint arXiv:1804.08559* (2018).

[19] David Mikkelson. November 2014. Syrian Hero Boy. Retrieved May 6, 2019 from https://www.snopes.com/fact-check/hero-ploy/

[20] Marcos Martinez BBC Monitoring. Novermber 2018. Burned to death because of a rumour on WhatsApp. Retrieved May 6, 2019 from https://www.bbc.com/news/world-latin-america-46145986

[21] Olga Papadopoulou, Markos Zampoglou, Symeon Papadopoulos, and Ioannis Kompatsiaris. 2019. A corpus of debunked and verified user-generated videos. *Online Information Review* 43, 1 (2019), 72–88.

[22] Olga Papadopoulou, Markos Zampoglou, Symeon Papadopoulos, and Yiannis Kompatsiaris. 2017. Web video verification using contextual cues. In *Proceedings of the 2nd International Workshop on Multimedia Forensics and Security*. ACM, 6–10.

[23] Chengcheng Shao, Giovanni Luca Ciampaglia, Alessandro Flammini, and Filippo Menczer. 2016. Hoaxy: A platform for tracking online misinformation. In *Proceedings of the 25th international conference companion on world wide web*. International World Wide Web Conferences Steering Committee, 745–750.

[24] Craig Silverman. 2013. Verification handbook. (2013).

[25] Denis Teyssou, Jean-Michel Leung, Evlampios Apostolidis, Konstantinos Apostolidis, Symeon Papadopoulos, Markos Zampoglou, Olga Papadopoulou, and Vasileios Mezaris. 2017. The InVID plug-in: web video verification on the browser. In *Proceedings of the First International Workshop on Multimedia Verification*. ACM, 23–30.

[26] Aric Toler. 2017. Advanced Guide on Verifying Video Content. Retrieved May 6, 2019 from https://newsinitiative.withgoogle.com/training/

[27] Veritone. 2017. Fakebox. https://machinebox.io/docs/fakebox

[28] Soroush Vosoughi, Deb Roy, and Sinan Aral. 2018. The spread of true and false news online. *Science* 359, 6380 (2018), 1146–1151.

[29] William Yang Wang. 2017. " liar, liar pants on fire": A new benchmark dataset for fake news detection. *arXiv preprint arXiv:1705.00648* (2017).

[30] Albert Weichselbraun, Philipp Kuntschik, and Adrian MP Braşoveanu. 2018. Mining and leveraging background knowledge for improving named entity linking. In *Proceedings of the 8th International Conference on Web Intelligence, Mining and Semantics*. ACM, 27.

[31] Markos Zampoglou, Symeon Papadopoulos, Yiannis Kompatsiaris, Ruben Bouwmeester, and Jochen Spangenberg. 2016. Web and social media image forensics for news professionals. In *Tenth International AAAI Conference on Web and Social Media*.

[32] Mohsen Zandi, Ahmad Mahmoudi-Aznaveh, and Alireza Talebpour. 2016. Iterative copy-move forgery detection based on a new interest point detector. *IEEE Transactions on Information Forensics and Security* 11, 11 (2016), 2499–2512.

[33] Arkaitz Zubiaga, Ahmet Aker, Kalina Bontcheva, Maria Liakata, and Rob Procter. 2018. Detection and resolution of rumours in social media: A survey. *ACM Computing Surveys (CSUR)* 51, 2 (2018), 32.