SYNERGISTIC MACHINE LEARNING FOR DETECTING FAKE NEWS APPROACHES

Kowsalya[1], Krishnan Nallaperumal[2]*

Abstract:

The widespread availability of tools to produce and disseminate fake information poses a growing danger to people, businesses, and government agencies worldwide. The Internet allows for the creation of an "alternative" reality where false allegations and subsequent apologies have little to no effect on the situation. A significant amount of digital media information is created and circulated every day; at the moment, the primary facilitators of fake news circulation are social media networks. Quite easily, in this "flood," material may be manipulated to influence its users. This highlights the critical need of working on efficient countermeasures. This study proposes and describes the architecture of a system for detecting fake news, which is currently being developed as part of the Identification of fake news on social media platforms (FNSMP's) project, in light of the above. Its primary purpose is to secure digital media files, such as movies, TV shows, and music, and it employs several different methods, including digital fake news, signal processing, and machine learning.

Keywords: widespread, fake news on social media platforms, TV shows, and digital media files, digital fake news, signal processing, and machine learning

[1]Research Scholar, Centre for Information Technology and Engineering, Manonmaniam Sundaranar University, Abisheka patti, Tirunelveli-627012, Tamil Nadu, India
Email: kowsmalni@yahoo.co.in@gmail.com
[2]*Senior Professor, Centre for Information Technology and Engineering, Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli-627012, Tamil Nadu, India Email: krishnan17563@gmail.com

***Corresponding Author:** Krishnan Nallaperumal
***Senior Professor, Centre for Information Technology and Engineering, Manonmaniam Sundaranar University, Abishekapatti, Tirunelveli-627012, Tamil Nadu, India Email: krishnan17563@gmail.com

DOI: -10.48047/ecb/2023.12.si5a.0474

Introductions:

Even before the invention of the Internet, individuals were able to distribute different sorts of misinformation via fake news and hoaxes to affect specific groups or whole civilizations. The term "fake news" is sometimes used interchangeably with the term "yellow journalism," with the common understanding being that false news includes elements of actual news that may be hoaxes. As was alluded to before, such measures are often used to enforce specific beliefs so that they are seen publically as genuine ones.

Rapid information dissemination has never previously been possible in human history, thanks to the widespread use of social media platforms in recent years. People now have more opportunities than ever before to produce and distribute content thanks to social media platforms. There are some intentionally misleading reports here. Because of this, in today's highly digitized world, online media, particularly social media platforms are the primary vectors for spreading false information. It's not only text that's being tampered with; it's also digital photos, movies, and audio files.

The proliferation of false news through social media platforms has already influenced events in actual, non-digital life. Take the 2016 US presidential election as an example: during that campaign, numerous forms of false news regarding the candidates were extensively distributed through social networks. It has been claimed in [1] that this may have had a major impact on the actual election process, as more than 41.8% of the false news data flow in the election was projected to come from online social networks. Compared to more conventional methods, this had a far greater impact and reach (i.e, TV, radio, or printed media). The current SARS-CoV-2 corona virus epidemic is another current example. In recent days, misinformation campaigns about the virus's dangers, source, possible transmission routes, treatment options, and, lastly, vaccine, have begun to spread at a quicker rate than the actual virus itself. The latter case demonstrates that the spread of false information may have devastating consequences, including the loss of life.

There are two forms of false news that people need to be aware of. Some false news, for example, may have its origins in the information that is real but has been maliciously modified, such as by switching out the audio track of a video clip or by generating a deep fake from an actual video. Second, we can locate bogus news that is fabricated from whole cloth rather than by tampering with existing material. Both sorts of disinformation will be targeted by this effort, as will become clear in the next sections.

The Detection of fake news on social media platforms (FNSMP's) work was launched in my work to solve the aforementioned challenges. Its goal is to provide content producers with tools that would enable them to readily identify any modifications to their work by adding watermarks. Furthermore, this would make it possible for users of online social media to employ tools based on cutting-edge signal processing and machine learning (ML) techniques to identify fraudulent information. In FNSMP's, we will develop models to identify fraudulent digital media material with an emphasis on the distortions caused by signal processing procedures and recording apparatuses. Users will be able to distinguish between authentic and false multimedia material without the requirement for evaluation and control from a centralized service thanks to the combination of fake news and ML- based detection technologies.

A cross-cultural user experience design approach will be used by the FNSMP's project in addition to technological advancements to establish consequences for the design [3] at all stages of the process. This project will carry out thorough user experience research to create and build technologies that are useable, beneficial, and attractive. Last but not least, it is important to note that the collaboration of three partners from Japan, Poland, and Spain is essentialto complement their technical background and expertise and to expand on participants from various geographic locations, cultural backgrounds, and life stages in the user experience study, which enables consideration of the values of potential end users in various contexts. The longer version of our conference article [4] in this essay. It must be highlighted that the unique scientific contribution presented in this paper may be summed up as follows, in contrast to this earlier work:

- We outline the architecture of the proposed false news detection system in detail, with a particular emphasis on how several methodologies, such as signal processing, digital fake news, and machine learning, interact to create a functioning system.
- We describe the test-bed for the first project, which consists of specialized, bespoke software utilized for Internet collection of real-world multimedia material. This would enable the evaluation of several false news detection techniques that have been developed during the research and are known from cutting-edge

publications.

With this approach, we want to increase understanding of the misinformation problem among researchers working on data concealing and machine learning. We also design the FNSMP's system's architecture in a flexible and expandable way. Thus, we expect to draw fresh research that will eventually be included in the suggested remedy. Therefore, by using a decentralized method that doesn't result in censorship or choices that are biased and motivated by personal interests, we help to reduce the impact of false news. The research will provide the first set of technologies that can be linked with social media platforms for the identification of false news, integrating digital fake news and machine learning methods. The prototype will allow for the addition and removal of various elements, such as machine learning models or digital fake news methods. As a result, we want to provide the platform with a way to include technology from other contributors.

The rest of the essay is organized as follows. The writings that are most pertinent to the subject of this essay are presented in Section 2. After that, in Section 3, we go through the essentials required to comprehend the idea behind the suggested structure. The general concept of the FNSMP's project is then presented in Section 4. The project's primary research stages are described in Section 5. The envisioned general project architecture is defined in Section 6 along with the detection process's key phases and modules, and the intended assessment platform is introduced in Section 7. The project's anticipated effect is then shown in Section 8. Section 9 brings our study to a close and suggests some next research trajectories.

Literature Review:

Mass-self-communication (MSC) describes horizontal communication networks, such as social network sites (SNS), where users may send and receive messages [5]. In MSC, stories are delivered uniquely since several voices may be present and users may contribute in a variety of ways to the story's development [6]. In contrast to hierarchical media, "a combination of top-down and bottom-up factors decide how the content is disseminated in significantly more participative (and messier) ways" [6]. With digital media, the concept of prosumers is strengthened [7], since users may create, consume, and circulate material. People use media in ways that go beyond one-to-one communication or one-to-many spectator paradigms when they consume, share, reinterpret, mix, and create it. This transition from distribution to circulation relies on the participatory culture that grew with digital media while also opening up new channels for the production and spread of false information. Early warnings about the possibility of some people becoming trapped in "filter bubbles" or "echo chambers," where they only receive information that supports their preconceptions, were raised in the early 2000s as a result of the diversification of sources used to provide online news and the use of social networking sites to filter news [8]. Online opinion leaders may shape online communities and contribute to the spread of false information [9]. Furthermore, in an "information psychological war," the use of big data and artificial intelligence enables influencing people based on their projected ideologies as well as their predicted phobias and anxieties [10]. As a result, the proliferation of false news [12] and rumors [11] was aided by the growing significance of SNS and the growing acceptance of participatory culture [13].

The rise of false news has forced the news industry to make greater efforts to demonstrate its commitment to honesty. Different initiatives evolved in this setting. On the one hand, fact-checking initiatives designed to expose hoaxes have appeared in several nations, including the United States (FactCheck.org), Spain (Maldita.es), Poland (Demagogue.org.pl), and Japan (Fact Check Initiative), and have been largely embraced by consumers. These websites often display the metadata of audiovisual material to defend the media's authenticity. Metadata, however, needs further scrutiny since it is simple to alter and does not reveal content changes.

The Trust project, on the other hand, offers a protocol that consists of eight trust indicators to "amplify journalism's commitment to openness, truth, inclusiveness, and fairness so that the public may make educated news choices"[14].Numerous news websites throughout the globe have embraced the rules, including well-known media organizations like the BBC, South China Morning Post, and Bay Area News Group. One of the metrics focuses on the reader's capacity to recognize the journalist's level of competence. Who created this? Although these initiatives have gotten positive feedback, they lack the technology tools needed to automate the processes.

Regarding copyright protection, content authentication, tamper detection, and other issues, digital fake news is acknowledged as a promising technology [15]. In certain fake news applications, each copy of the transmitted material has a unique fingerprint that identifies the receiver of the multimedia content. By allowing the content owner to track the origin of the redistributed copy, this application discourages unlawful redistribution [16, 17].

Identification and tracking of false news is a prospective area where the use of digital fake news methods may be advantageous. Since this idea has not yet been explored in the literature, it might be seen as fresh and intriguing. While there have been some efforts to combat deep fakes [18] or fake news in photographs, no comprehensive solution connected with social media platforms has yet been developed or examined for other categories of digital information. As a result, we may say that the suggested strategy has the potential to be creative. Regarding detection methods, most efforts in spotting false films focus on seeing subtle details that are present in the fabricated footage. The approach suggested in [19], for instance, is based on the detection of eye blinking, a physiological signal that is poorly shown in synthetic false films. The technique in [20] displayed CNN layers and filters and found that the eyes and mouth are crucial for spotting faces that have been altered using deep fake software tools [21, 22].

The development of automated detection techniques is crucial in both academic and industry settings now that false video issues are understood. The quantity and characteristics of the hundreds of thousands of movies posted to the Internet or social media platforms vary, even if digital forensics specialists have created unique solutions for certain minor situations. The Partnership on AI's Media Integrity Steering Committee, Amazon Web Services (AWS), Facebook, Microsoft, and academics created and publicly distributed the Deep Fake Identification Challenge (DFDC) data set [23] to speed up progress in the detection of fraudulent media. The competition aims to inspire academics worldwide to develop novel new tools that may assist in identifying deep fakes and altered media.

In contrast to the previous research, the FNSMP'S project focuses on the artificial signals produced by signal processing operations and recording devices to develop models to identify fraudulent digital media material. While the phony materials are constructed with no evidence of change or editing, the signals are purposefully included as watermarks. Users will be able to distinguish between authentic and false information with ease thanks to the combination of fake news and ML-based detection technologies, eliminating the need for evaluation and control from a centralized service.

Background:

This section provides a short description of the methodologies that will be used throughout the project, including user experience, machine learning for the identification of false news, and digital fake news (UX).

Digital fake news

A group of methods known as digital fake news, a subset of data hiding [24], includes embedding data referred to as a mark or a water mark into a digital object or carrier while generally retaining the item's perceived quality. Multimedia materials including photos, music, and video, but also text and even network protocols, are the typical bearers of fake news. The cover object has a connection to the embedded mark, and the cover object is valued (often more valuable than the watermark). The watermark may be used, for instance, to show who has the copyright to the material or who has been granted permission to see it. Copyright protection, content authentication, broadcast monitoring, transaction tracking, and copy control are just a few examples of traditional uses for digital fake news.

Steganography is another well-known subfield that seeks to transmit confidential information between two communicating parties. Steganography, as opposed to cryptography, aims to keep the transmission of information itself secret by concealing the communication in a seemingly harmless carrier rather than rendering a piece of information inaccessible to those who are not permitted. In steganography, the secret message is the thing that has tobe secured; the cover object is often seen as worthless. It is crucial to note that digital fake news is not a specific kind of steganography but rather another data-hiding branch that has some similarities to steganography but distinct characteristics and uses.

The five fundamental characteristics of digital fake news schemes—capacity, robustness, transparency (or imperceptibility), blind or informed detection (or extraction), and security—are often examined.

The quantity of data that the indicated item is capable of transporting is referred to as its capacity or data payload. The quantity of data is often expressed in bits per unit, with the unit varying depending on the kind of item. Bits per second and bits per pixel are two examples.

The necessity or absence of the original (cover) item while conducting the detection (or extraction) of the concealed watermark is referred to as blind or informed detection or extraction. The detector/extractor is referred to be non- blind or informed if the cover item is necessary at the detection/extraction end. A non-blind fake news strategy or aninformed fake news scheme is the general phrase used to describe the whole plan. The detector/extractor is referred to as blind and the fake news method is blind fake news when the original cover object is not utilized by the extraction or detection algorithm.

The perceptual quality of the marked item in comparison to the cover (original) object determines its transparency or imperceptibility. The embedded information, for instance, will cause certain pixels of the marked item, such as an image, to vary from the corresponding pixels in the cover picture. The degree of "perceptual noise" that the embedding procedure introduces into the picture determines imperceptibility. Therefore, "perceptual similarity between the cover and the designated items" may be used to determine imperceptibility.

When a marked item is modified using common signal processing techniques like filtering, lossy compression, or geometric modification, robustness refers to the capacity of the fake news system to identify or extract the embedded watermark. Robust fake news is a kind of fake news that can withstand signal processing assaults. On the other hand, certain applications demand that the watermarks be eliminated when the watermarked item is subjected to all or partial modifications. Fragile fake news (no transformations permitted) or semi-fragile fake news (some transformations permitted) are the terms used in this situation.

Security is a fake news system's capacity to fend off unwanted attempts. In this instance, two types of attacks maybe distinguished: those that target the embedded watermark and those that target the (secret) keys used in the fake news technique. Unauthorized removal, embedding, and detection are the three categories into which attacks of the first kind may be divided (or extracted). The possibility of hiding or masking the embedded watermark so that it cannot be seen or extracted is known as "unauthorized removal." In terms of unlawful embedding (or forgeries), the attacker's goal is to include a watermark into a piece of work to give the contents a false sense of authenticity. Unauthorized detection can be considered a passive attack, in contrast to unauthorized removal and embedding, which are both active methods (and extraction).

The fake news keys utilized in the technique are connected to the second category of security threats. Many fake news techniques, like cryptography, involve the usage of secret (typically symmetric) keys that are used for embedding, detection, and extraction. However, in fake news schemes, some keys that are close to but not the same as the ones used by the embedded may result in the successful extraction or removal of the watermark without authorization. Fake news algorithms often use randomization as a security technique in some areas.

Additionally, the embedded message or watermark may be encrypted independently before embedding, requiring decryption at the receiving end to retrieve the embedded data. Most data-concealing apps use this combination of fake news and encryption, each with its unique key (fake news and crypto/cipher keys, respectively). Other characteristics of digital fake news include computational cost, false positive rate, modification, multiple watermarks, and embedding effectiveness.

Owner identification/proof of ownership, content authentication/tampering detection, and transaction monitoring arethe most pertinent uses of digital fake news that may serve as the foundation for the FNSMP'S project. A watermark can be used in owner identification/proof of ownership applications [25] to provide contact details for the owner or source of a specific work. Since the watermark is undetectable and integral to the tagged item, it goes beyond the typical textual copyright notifications included in these works. The purpose of the watermark in the event of evidence of ownership is not only to identify the owner but also to demonstrate that person is the rightful owner of the work, even in court. The embedded watermarks may also be employed in content authentication/tampering detection or localization applications [26, 27, 28] to determine if the marked work has been altered by an adversary. Some methods enable particular forgeries to be localized in the material in addition to a yes/no detection (tampering detection) (tampering localization). In contrast to the robust techniques that are used in many other applications, fragile or semi-fragile fake news systems are needed for this application. Applications for transaction monitoring and fingerprinting [29, 30, 31, 32] embed various watermarks (also known as fingerprints) in material that is made available to various users. Each copy of the material has a unique fingerprint stored in it that may be used to identifythe user if the user wishes to share the information again (tracing).

Digital fake news will need to be used in the area of fake news detection using a variety of methods, perhaps including robust watermarks for transaction monitoring and evidence of ownership and fragile or semi-fragile watermarks for tampering detection and localization [33]. This is a difficult application case since various kinds of watermarks have a variety of requirements, which will increase the complexity of the embedding and detection procedures. The system also has to be made to function with various forms of multimedia material (image, audio, and video).

Digital forensics on media

Steganography allows a malevolent actor to carry out covert conversations over a public channel without raising suspicion from potential listeners. The analysis of secret messages in multimedia information, also known as segno analysis, has undergone extensive research as a steganography countermeasure for categorizing content with/withouthidden messages.

The irregularity of multimedia information has been assessed from a forensics perspective, driven by the research of steganalysis. A group of scientific methods has lately been put out for the analysis of multimedia information, such as audio, video, and photographs, to extract evidence from them. Such technologies try to expose the content's history in particular:

- The identification of the data-producing acquisition equipment,
- The verification of the content's integrity,
- The information extraction from the signals used to create the content.

It is assumed for source identification that acquisition equipment leaves distinctive traces because of its inherent qualities (such as sensor noise, lens distortion, and others), which result from hardware-oriented distortion. Similarto this, altering multimedia material results in software-oriented distortion, which is created by signal processing procedures (such as lossy compression, filtering, and others). Identifiable traces produced by various processing methods may differ, and scene properties may change as a result of manipulation. We can identify these distortions and categorize malicious altering traces in multimedia material with the use of deep learning algorithms. These multimedia forensics methods will make it possible for us to identify false material with great precision.

ML-based surveillance:

The mapping from input to output is learned by ML algorithms. In classification problems, the algorithm learns a function called the decision boundary that divides a job into two (and sometimes more) groups. A particular data point's membership in a positive or negative class may be ascertained with the use of the decision boundary. To categorize bogus material, we first extract characteristics from a target piece of information, and then a machine learning algorithm determines whether those features belong to a positive or negative class by calculating a metric. The two stages of the ML model that should be taken into account here are feature extraction and feature selection. In feature extraction, the features needed to complete a job are extracted. On the other side, in feature selection, we choose the crucial elements that enhance an ML model's functionality.

In general, building the feature extraction function manually from a picture takes time and requires in-depth topic and domain expertise. The tuning of such a function may be computed automatically using deep learning methods. Recurrent neural networks (RNNs) and convolutional neural networks (CNNs), two subsets of deep learning methods, serve as the foundation for many others. When producing predictions, RNNs take into account the sequential information found in the input data, i.e., the relationship between the words in the text. CNN's, on the other hand, extract spatial information from a picture. The arrangement of pixels and their interactions within a picture is referred to as spatial characteristics. They enable precise item identification, object placement, and object relationship within a picture.

By replacing a person's face with that of another and synthesizing the movement of the face following the modified audio voice, a deep learning-based approach allows us to produce false content. Entire Face Synthesis, Identity Swap, Attribute Manipulation, and Expression Swap are four categories into which the face alterations may be divided [22].

There are several researches to categorize whether faces are genuine or artificially made as a defense against such fraudulent stuff. The early research in this field concentrated on the audiovisual artifacts included in the original fraudulent films. [21] Analyzes the discrepancies between lip movements and auditory speech. In [34], the Long

Short-Term Memory (LSTM), which is based on RNN, is used to enhance the detection capabilities. Using a logisticregression model and a multilayer perceptron (MLP) [36], several straightforward visual characteristics, such as eye color, missing reflections, and missing features in the eye and teeth regions, have been used in [35] for the categorization of false contents. A detection technique based on both head motions and facial expressions is suggested in [37], and the final classification is performed using a Support Vector Machine (SVM) [38]. [39].

A detection method [19, 40] based on CNN, such as VGG16, ResNet50, ResNet101, and ResNet152, identifies the existence of artifacts caused by the difference in resolution between the identified face regions and the surrounding areas. [20], which closely examines a CNN-based system, and proposes methods based on mesoscopic and steganalysis characteristics. Comparing several options, the detection system built on the Xception Net architecture produced the greatest results in this study.

Project planning
This project can be broken down into two main sections: the creation of technological tools to help users tell the difference between authentic and fake content in multimedia publications; and (ii) a case study to take into account the cultural aspect and focus on the development on the needs and behaviors of actual users. We give more information on the aforementioned topics in the subsections that follow.

Tools for fake news and detection:
A collection of technological tools will be made available as part of the initiative to help users distinguish between authentic and fake material in multimedia documents (i.e., sound, images, video). The FNSMP’S project will offer two different kinds of tools for that:

- Fake news equipment In this project, we'll create a system that uses fake news to embed data in media files before they're shared on social media sites. Once they are made public, it will be straightforward to determine whether a media document originates from a reliable source or if it has been modified to produce fake news. To accomplish this, a variety of digital fake news techniques will be used to incorporate authenticating watermarks into the original file that are difficult to remove without changing the content. These watermarks can be undetectable, visible, or audible. Any alteration that is done should achieve this goal by being instantly and immediately recognizable. As a result, the created system will be able to safeguard any type of digital media content (such as photos, video, and audio) and effectively alert consumers when they get fake content when incorporated within online social networking platforms. In many instances, it will also be feasible to pinpoint the source of the spread of fraudulent content. Taking into account the discussion above, the following are the primary goals connected to this research area: The development of the proof-of-concept implementation of the entire system follows the development of the following components: I selection of appropriate digital fake news techniques; (ii) design of the proposed system's architecture; (iii) design and development of the automatic and user-friendly digital content verification mechanism with a user warning feature if fake news is detected; (iv) design and development of the mechanism for identifying the source of the propagation of fake content; and (v) design and development of the entire

Comparing digital fake news to other detection techniques, such as ML solutions, has various benefits. It starts by not requiring a training set, which prevents issues like overfitting. A watermark's presence or absence can be used to distinguish between genuine and fake content. Additionally, the combination of various watermark kinds, such as robust, fragile, audio, and video watermarks, can be particularly effective for detecting forgeries. Additionally, fake news can be used for data tracing (data provenance), which makes it possible to determine the origin of falsified content.

- Detection equipment Fake news is frequently spread through direct messaging services rather than the established online social media. As a result, embedding data of any kind is not always successful. Furthermore, we cannot anticipate that all media outlets and social media platforms will publish media files using the proposed (or any other) fake news technique. Therefore, it's critical to offer technologies that can determine whether or not certain media content that hasn't been labeled as the fake is legitimate. The project will also offer a machine learning (ML) tool to analyze minute abnormal signals brought on by the creation of bogus content to address this difficulty. To achieve this, this project will develop a tool that combines two methodologies with a focus on I device-oriented distortions introduced by variations in the recording devices, which can be classified as hardware-dependent

characteristics, and
(ii) distortions caused by signal processing operations, such as ML tools, which can be classified as software- dependent characteristics. The following are the primary goals of this field of study: the development of an artificial neural network (ANN)-based architecture for classification; (ii) the design of the feature extraction from suspicious multimedia contents; (iii) the collection of datasets for training the proposed system; (iv) the development of a compact and efficient implementation; and (v) the development of the proof-of-concept implementation of the entiresystem.

Research of user experience

A user experience (UX) research, which is based on the experiences of several prospective users from the global north, will provide implications for the design, implementation, integration, and assessment of the tools. Spain, Poland, and Japan will all participate in the research. The three nations are from Asia and Europe and are classified as high-income nations by the World Bank [41]. They do, however, reflect various cultural and historical contexts that could affect the ingestion of false information. Fake news is also commonly spoken in national tongues. However, false news in Spanish may originate from a variety of backgrounds, while Japanese and Polish are mostly constrained by the geographical environment, making the three of them three distinct and intriguing situations to investigate.

- Identify cultural aspects that should be considered while designing the system. To do this, it is necessary to examine how readers respond to news in general and, more specifically, how they respond to the impartiality, reliability, and correctness of the reporting. The research will examine the pertinent aspects of news that raise or lower its credibility on various topics. We will also undertake a thorough examination into how the suggested technologies might help raise the subject's awareness of fake material and (ii) reduce the dissemination of false information.
- The research will examine the optimal methods for providing tools to the participants from the perspective of human-computer interaction. For example, we'll investigate the most effective kind of watermark, the information that has to be included in the various file formats, the forms of bogus information that subjects can recognize, and more. Additionally, the investigation will concentrate on the ideal methods for interacting with the suggested instruments and for disseminating the outcomes of the ML and fake news analyzers.
- Delivering findings that are relevant to the three nations involved in this project (Japan, Poland, and Spain) will enable cultural comparative study. The case study will be modified for this reason to reflect the cultural backgroundsof each of the aforementioned nations.
- Participate in the iterative prototyping process used to create and implement the suggested tools. This case study is planned to begin from the very beginning of the project to include the users' perspectives and the cultural information amassed around them as soon as feasible in the creation of the tools.

The following are the primary goals for this study area: I identify the cultural elements that need to be taken into account when designing the tools; (ii) make thoughtful judgments about the tool design; and (iii) assess the system's possible effects.

Superiority in science

Based on three primary research efforts, this project suggests an innovative strategy to counteract false news in multimedia content:

- A thorough assessment of the user's experience to place them at the center of this investigation. Aspects of culture and behavior that are crucial for creating useful tools will be examined in this research. The three partners will take part in the research to provide pertinent information that will be used to modify the tools and assess their effectiveness across three distinct nations.
- The development of fake news tools that use a cutting-edge methodology to instantly spot modifications in multimedia documents as part of a comprehensive system that can be connected to social media platforms.
- Create detection tools based on cutting-edge ML and signal processing techniques to let consumer's spot manipulation in multimedia assets without watermarks.

The integration of these three research initiatives, which will provide more generalizable findings than doing the study independently, achieve exploitable results, and produce an integrated and user-centric prototype, is the FNSMP's project's major strength.

Complementary benefits of working together across borders:

The following are only some of the many reasons why multilateral collaboration between institutions in Japan, Poland, and Spain is so crucial:

- People of various cultures and backgrounds may react differently to the effects of false news and other forms of deception. Due to the complexity of the issue, it is essential to conduct a case study of this project with three partners, including two nations with extremely distinct cultural backgrounds (i.e., Poland and Spain), as well as the Japanese viewpoint. The tools will be created with the end-user in mind. The target audience will be brought in earlyon in the development process. When considering human-computer interaction from a broader perspective than the national or regional viewpoint of a single partner, multilateral collaboration is essential.

- All partners and subjects from those three countries will participate in evaluations of the suggested instruments.
- Multilateral collaboration is crucial from a scientific perspective to address the FNSMP'S project's three primary foci. Spanish partner provides a multidisciplinary team with technical competence in digital fake news and social scientists that specialize in conducting case studies and human-computer interaction assessments; Polish institution has fake news specialists; Japan has signal processing and ML experts;

Five levels of study.

The FNSMP's platform provides three distinct technological components—(i) forensics, (ii) fake news, and (iii) ML-based detection tools—to shield people from false news. The goal of forensics tools is to ascertain whether or not the digital media created by content producers that seek to improve the online social media platform do so by presenting modified material. Then, the digital "stamp" on the newly presented material is provided by the fake news instruments. Even more so, consumers can ascertain whether or not the material they have received is authentic thanks to detection technologies based on machine learning.

Earlier, we established that the findings from the international user experience survey will inform the design and development of the FNSMP's tools. Please take note that Section 4 explains in detail the proposed framework's many parts.

In conclusion, the FNSMP's project's research efforts may be broken down into three distinct phases:Phase 1: Development and deployment of fake news infrastructure;

Phase 2: Tools for ML detection development and deployment constitute; andPhase 3: Case Analysis of the User Experience.

To that end, we detail the research approach that will be used at each stage of the project.

Phase 1: Development and deployment of fake news infrastructure.

At this stage of development, we will use a hybrid strategy that integrates qualitative and quantitative techniques. In particular, we will conduct experimental evaluations of the performance and whether or not the developed digital fake news system satisfies the predicted criteria by using proof-of-concept implementations of the components. Likewise, the designed system's proof-of-concept implementation will be tested in the lab for quality assurance.

Our focus is on the many uses of digital fake news, a technique for inserting data into multimedia files. For example,the material's recipients could be able to track down its source, and the detector might utilize the embedded markings to determine which users generated the content that was spread.

There are two basic types of fake news methods: robust and fragile. A robust watermark cannot be altered in any way. As opposed to a robust watermark, which would be unaffected by any changes to the content, a fragile watermark would be readily altered. The second method may be used to identify deliberate alterations. In this work, we integrate these two fake news methods to detect false material.

Phase 2: Tools for ML detection development and deployment constitute.

Multimedia material is essentially produced with the use of recording tools like digital cameras and microphones. Torecord high-quality multimedia, it is necessary to control for hardware-oriented distortions brought on by the FNSMP' sity between optical and sensor devices. Forensic multimedia is created by splicing together footage from many sources. During this procedure, there must be some distortions, such as the introduction of garbled or otherwise abnormal signals, or the introduction of noise with unexpected properties. Using signal processing and ML methods, we will investigate the anomalous signals associated with false materials.

Both physical device features and software operation may contribute to the generation of artificial signals. It will lead us to the material's original creators and reveal any signs of content recapturing. Meanwhile, nonlinear procedures like rounding and lossy compression are the major sources of distortion in software. The artificial signals will enable us to identify the bogus information even if it was created by a malevolent party utilizing deep learning methods.

Time and frequency domain consistency of attributes is hard to regulate in artificially manufactured and altered material. Similarities exist between steganalysis (the analysis of steganography) and the method used to analyze anomalous signals in the multimedia information.

In steganalysis, we investigate methods, such as machine learning methods [42, 43], that may reveal whether or not the concealed message is there. Similar to how bogus news is distorted during production, A methodology similar to steganalysis may be used to decipher the contents. To create an effective Deep Neural Network (DNN)-based classifier, it is necessary to conduct in-depth research into the analysis of distorted signalsand their features.

Many scientists have spent the last few years studying how to identify bogus information using ML in conjunction with traditional signal processing methods. The advancement of the detector, however, encourages other researchers to create more natural-looking false contents that are challenging to classify with the aid of Generative Adversarial Network (GAN) architecture. Some research also suggests that detectors may be attacked by injecting them with adversarial noise designed to lead to false positives [44]. Since adversarial noise may trick DNN-based classifiers without severely degrading the contents, it is acknowledged as a danger to ML systems [45]. An effective fake content classifier must also be resistant to adversarial noise.

Phase 3: Case Analysis of the User Experience.

To learn about and quantify how cultural factors affect the user experience in different regions, we propose to use a cross-cultural user-centered design method [2]. In the first stage of the study, we will use focus groups to learn more about the cultural elements that contribute to the spread of false news. Tools and the project's potential effect will be evaluated via iterative prototyping as well (B). as through heuristic assessment and usability testing, both of which mix qualitative and quantitative data. Since Japanese, Polish, and Spanish are all being considered potential target languages for the case study, linguistic concerns are naturally a major consideration at this stage. Comparison between envisioned and actual buildings:

As was said before, FNSMPs will use several various technologies to aid people in identifying phony news. If you want to know how all these technologies can be put together to (attempt to) spot false news, you need to know the system's architecture. Also included is a comparison of similar designs.

Proposed building designs

The planned design of the FNSMP's framework is shown at a high level in Fig. 1. A situation is imagined in which content creators and content consumers may be distinguished. The former may produce either real or fraudulent news, which can then be distributed through social media websites. They are the usual users of social media platforms who are targeted by bad actors that spread false information to them.

Aside from that, the suggested detection procedure is split into two parts, each of which has two subparts:

1, To make sure your sources and material are legit. In this step, we check to see whether the online content (audio/voice, video, or picture) we're looking at has been altered and if its source is legitimate. Fig.2 depicts a flowchart of this process. The two components that make up this stage are readily apparent:

(A). Source code checker module: It employs a strong watermark detector (the original creator should have incorporated a strong watermark before sharing the material) or, if no watermark is discovered, an Internet search to find the original content. The content's reliability is determined by comparing it to a database of trusted sources.

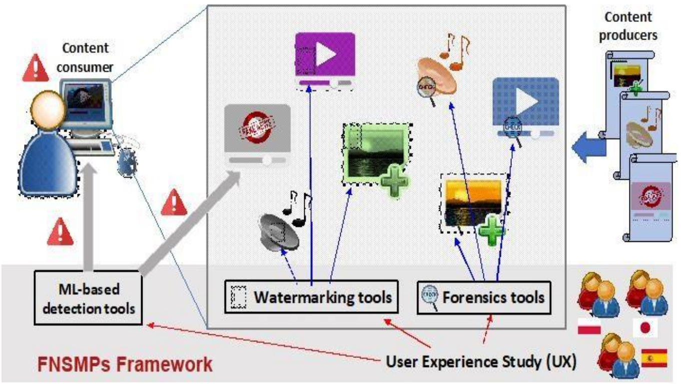


Fig 1: FNSMP's Framework architecture

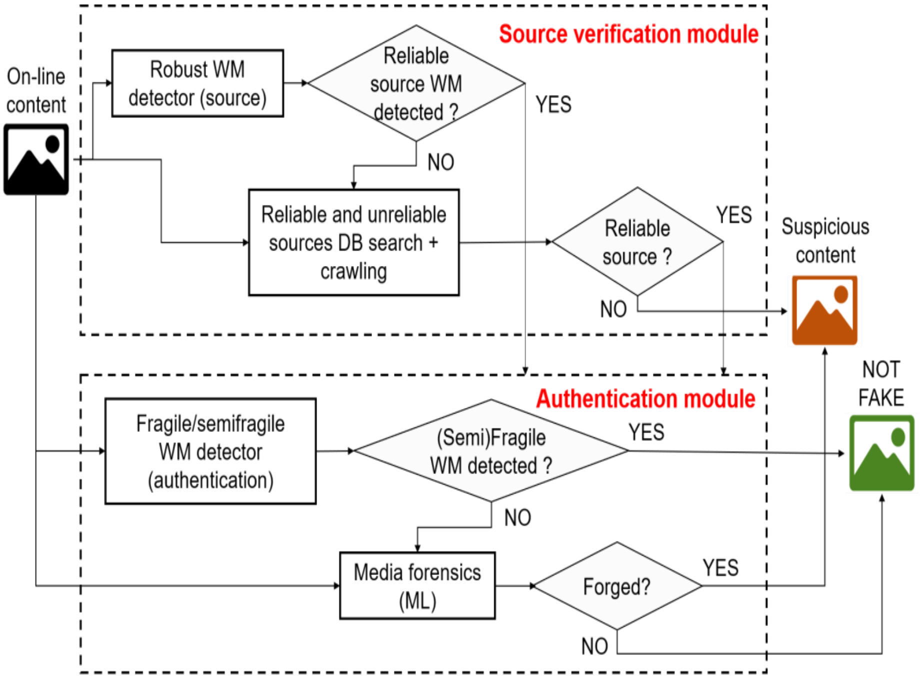

Source verification module
On-line content
Robust WM detector (source)
Reliable source WM detected ?
YES
NO
Reliable and unreliable sources DB search + crawling
Reliable source ?
YES
NO
Suspicious content
NOT FAKE
Authentication module
Fragile/semifragile WM detector (authentication)
(Semi)Fragile WM detected ?
YES
NO
Media forensics (ML)
Forged?
YES
NO

Fig 2: Phase 1: Authentication of the Source and Content

(C). Authentication Subsystem. Once a trustworthy origin for the material has been established, an authentication module may examine it for signs of fabrication. Media forensics techniques, usually based on ML algorithms, may be used to attempt to establish whether there has been any malicious change on the material if no fragile or semi-fragile fake news is employed at the source.

If the authentication module confirms that a piece of material came from a trustworthy source and has notbeen tampered with, the piece of content is marked as genuine (or "not false") and a message explaining the authentication process is given to the user. In all other cases, the material is classified as "suspect," necessitating additional investigation in Step 2.

2. Step 2 is to identify and track down any instances of fake news. After some information has been flagged as questionable, this stage attempts to sort it into one of three categories: false news; authentic; and undecided. Finally, if the information is identified as "fake news," a module attempts to determine where it originated so that it may be added to a database of discredited news outlets.

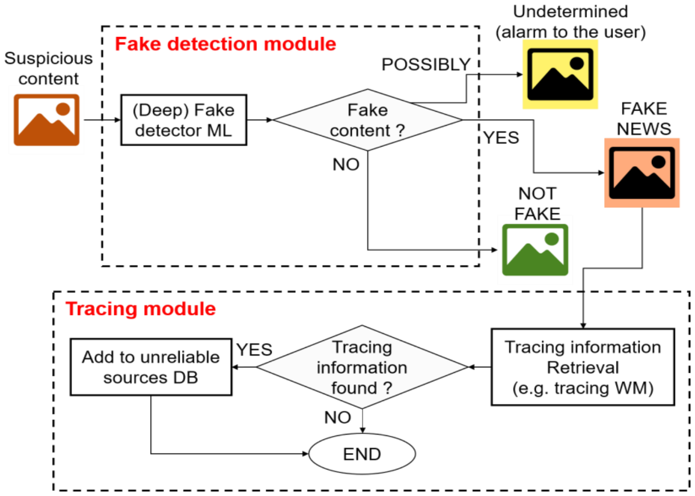

Undetermined (alarm to the user)
Suspicious content
Fake detection module
POSSIBLY
(Deep) Fake detector ML
Fake content ?
YES
NO
FAKE NEWS
NOT FAKE
Tracing module
Add to unreliable sources DB
YES
Tracing information found ?
NO
END
Tracing information Retrieval (e.g. tracing WM)

Fig 3: Step 2, Identifying and Tracking Fake NewsFigure 3 depicts a flowchart for this stage, outlining its two components:

a.) A fake detection module, a. Here, artificial intelligence technologies are used to assess if a pieceof material is authentic (i.e., not phony), fake news, or undecided. If the latter is true, the user willbe alerted and asked to evaluate whether or not to believe the data. In addition, if the user so chooses, his or her assessment of the material's trustworthiness will be recorded and utilized to assign trustworthiness rating to the content should another user submit it for review.

a.) Sub module for tracking. This module will attempt to track out the source of any information that has been flagged as false news, so that it may be included in a database of questionable resources.

This may be done by following digital fingerprints, or watermarks, or by collecting additional data (provenance data) that may be relevant for tracing purposes.

As was previously mentioned, the suggested architecture incorporates several different technologies, suchas digital fake news (including robust, fragile/semi-fragile, and digital fingerprints), forensic tools, and ML algorithms trained to detect false news from authentic materials. Some technologies (such as digital watermarks) are not required in this architecture, although they may greatly aid in the identification of false news if implemented.

There are two main categories of false news: those that start with real material but are altered in some way, and those that start with nothing but a malicious intent to mislead the public. This section outlines a potential architecture that may combat both types of false news. The following methods could be used in the detecting process:

Find out where the information came from. There are three possible consequences if this is possible: a) the content comes from a trusted and official source (in which case, manipulation can produce fake news of the first type); b) the content comes from a trusted but dishonest source (in which case, the user would be informed and could choose to ignore the information); and c) the content's origin cannot be established(this would the typical case for fake news of the second type).

Identifying manipulations from trusted, established sources (fake news of the first type).

Third, utilize machine learning to determine whether material from unidentified sources is real or not. In many situations, it would be difficult to distinguish only "lies," hence the accuracy of such categorization may be poor. It's possible that as the project develops, it will become important to determine whether or not a certain piece of news is also available from reputable sources. Indeed, this is the work of modern fact-checkers. However, owing to funding constraints, this is now outside the scope of the FNSMP'S project.

The goal of FNSMP'S project is to create a prototype of this architecture and make it open so that other solutions may be included. This will pave the way for a user-friendly, no-cost platform to aid in the identification of false news. The FNSMP'S platform is unique because it gives media consumers the flexibility to select whether and when they want to utilize the system to help them spot false news, avoid censorship or central control, and protect other basic rights, including the right to free speech online.

Contrastive Study:

Many academics have recently devoted time and energy to studying the issue of bogus material and developing methods to counter it. Traditional methods may be categorized essentially into two groups: active and passive. The differences between the current methodology and the proposed architecture are shown in Table 1. At the time of multimedia production, information is encoded using active approaches, such as the addition of a watermark [15]. You can tell whether the multimedia file has been altered by looking for the watermark. Additionally, the retrieved watermark may be used to identify altered target regions. Although digital watermarks have some limitations [51], this issue may be mitigated by using a blockchain [52, 33] to store an immutable record of watermarks and other content properties, making them impossible to alter. Fake news and blockchain technologies have been presented as proof-of-concept systems for detecting and preventing fraudulent video news by Alattar et al. [46].

To detect false information, passive methods examine telltale signs of modification. Artifacts created by Deep-Fakes are often indistinguishable to the naked eye but not to machines or forensic investigation. Examples of spatial artifacts include inconsistencies, backdrop abnormalities, and GAN fingerprints. Examples of temporal artifacts include the detection of behavioral and physiological oscillations, as well as coherence and the synchronization of video frames. One simple method for elucidating the differences between genuine and false is to work with pixels and leverage the correlations. There has been researching on DNN-based approaches to increase detection efficiency and generalization capability.

Active and passive methods have been combined to form the suggested design. Both the presence of a watermark and the verification of modification traces are used during source verification and content authentication to determine with high confidence whether or not the target material is legitimate. The suggested architecture not only classifies the information but also includes a tracing module that lets us trace it back to its source using a digital fingerprinting method.

Space for trying out ideas in development:

As was said above, a trustworthy assessment environment is necessary for newly developed false news detecting systems. Further, actual data collected directly from the Internet would be necessary during such an assessment process to demonstrate that the established methodologies would be efficient and successful in real-world settings. This will be accomplished by creating a tailored test-bed platform with purpose-built proprietary software as part of the FNSMP'S initiative. For the time being, it is true to say that the design of the software platform has begun.

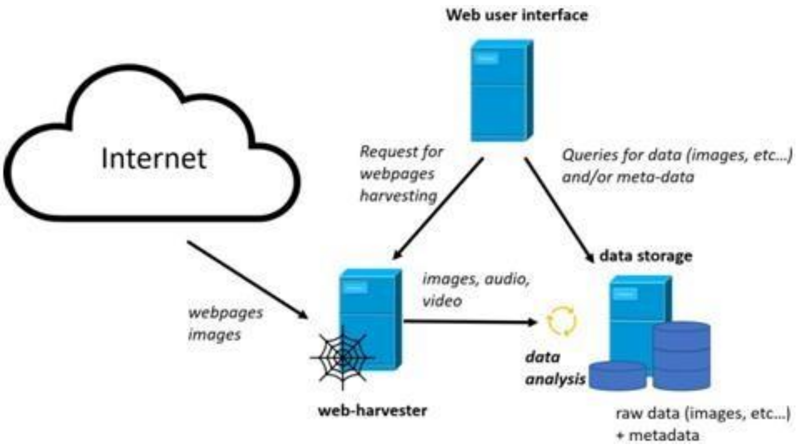

Web user interface
Internet
Request for webpages harvesting
Queries for data (images, etc...) and/or meta-data
data storage
images, audio, video
webpages images
data analysis
web-harvester
raw data (images, etc...) + metadata

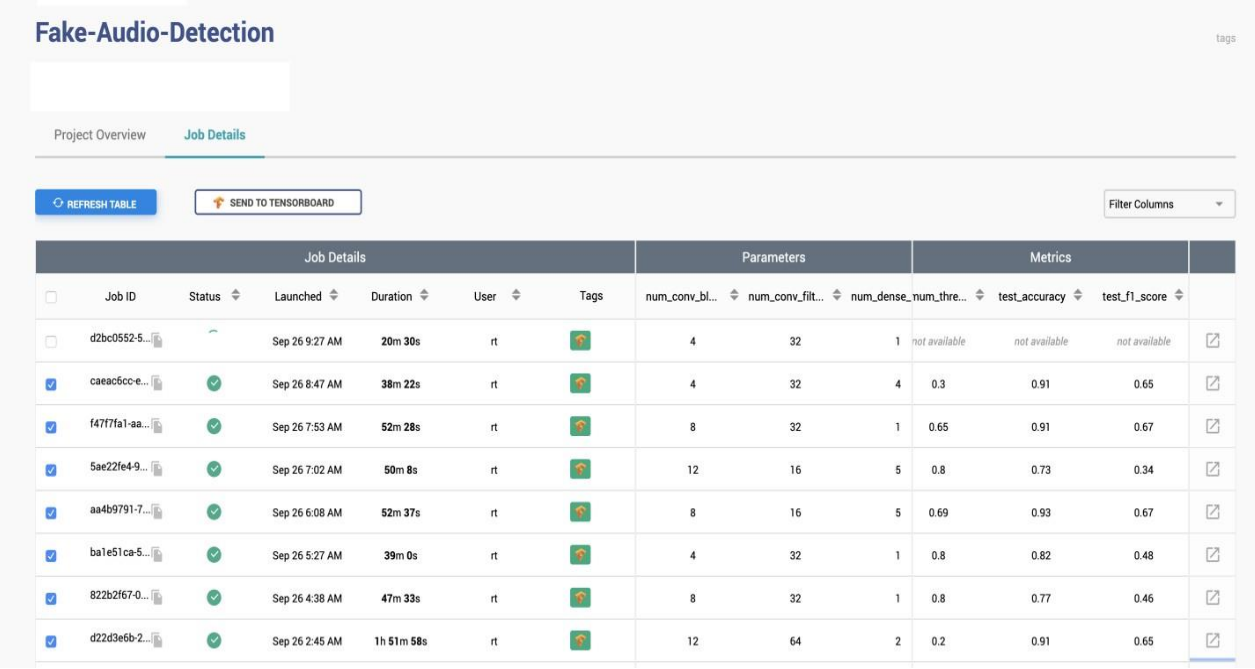
Fig 4: Warehouse of Cloud data center

Key features of the developed program are shown in Fig. 4. The website harvester is the system's primary "workhorse." This part of the system is in charge of accessing certain websites and collecting relevant multimedia items, such as digital photos, sound recordings, and videos, for later analysis. All data obtained through downloads are saved to disk and may be immediately evaluated with the help of the offered plug-in (containing, e.g., proposed detection schemes). New analytical plugging, such as those that detect embedded watermarks, identify the possibility of media manipulation for false news generation, or add steganographic information for secretly delivering instructions to infected PCs, may be easily added to the established system. During development, the project's collaborators may readily provide a specialized plug-in for trying out new algorithms and techniques. Whenever an analytical plug- in generates data on a specific media file (which is saved to disk), that data is recorded as metadata and saved in the appropriate database. To emphasize how important this database is to the whole system, it is shown in Fig. 4 as one of the three main components. The built platform is incomplete without a web- based user interface. It may be used to control the gathering of multimedia information from the web. In addition, the system database's metadata could be searched using it.

FNSMP will make use of the established system during its duration. In the early stages of the project, it will be utilized to compile information from the Internet's actual users. The proposed detection algorithms, such as those that would be able to identify the alteration of multimedia files for false news, may be added as plugins and assessed on real-world data as the project develops. We described the architecture of the system we've been testing so far in the previous paragraphs. Because the study involves collecting information from web servers all over the world, various experimental considerations need to be made.

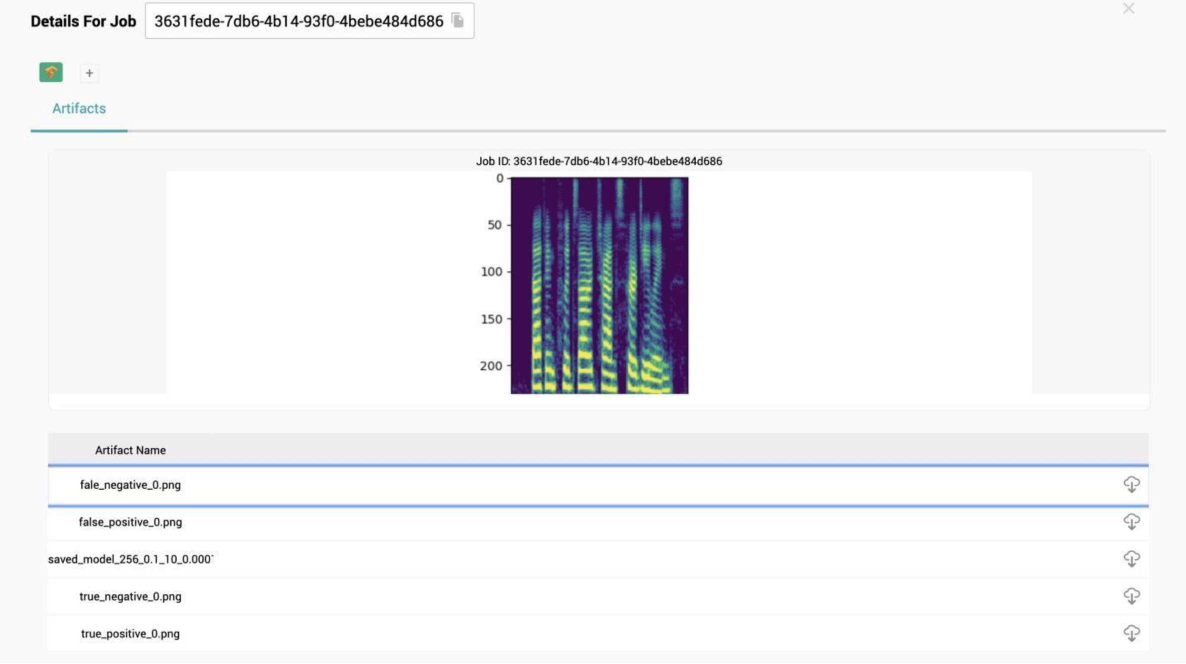
The first one is connected to how quickly subsequent web pages download. To get the most out of the experiment, the presumptive list of websites must be retrieved as quickly as possible. Unfortunately, such actions could harm collected public web servers. Such behavior may be seen as a Denial of Service (DoS) assault from the perspective of the web server administrator. In addition, some administrators set up an automated protection mechanism that prevents serving any further web pages to the address suspected of being part of a Denial of Service assault. However, the rapidity with which harvesting occurs is only one of many factors that can make it impossible to download certain online sites. While crawling through huge numbers of pages, we may end up on dangerous sites. Black holes are often used to eliminate threats from hostile domains. The Domain Name System (DNS) may be used to do blackholing by making sure that inaccessible domains (such as 127.0.0.1) respond with an unreachable address. The preliminary tests verified the reality of this phenomenon.

The second most important factor is informing web host providers that the harvesting we do is not an attack but rather part of the FNSMP'S research effort. Naturally, there has to be a way for web server owners to have their IP addresses removed from the list we use for testing reasons permanently. We looked at several channels for communicating with web hosters about our studies and related efforts. The first one acts as a web server from the same IP address that is used for harvesting. The hosted webpage details our study and provides website administrators with a means to get their IP addresses removed from our databases permanently. The second option makes use of certain special HTTP headers to provide specifics about the study done concerning FNSMP and links to relevant websites. We haven't settled on the best course of action yet, so either one might be used in the final analysis is see fig Fig 5 to 11.
Results and Discussions:


Fake-Audio-Detection

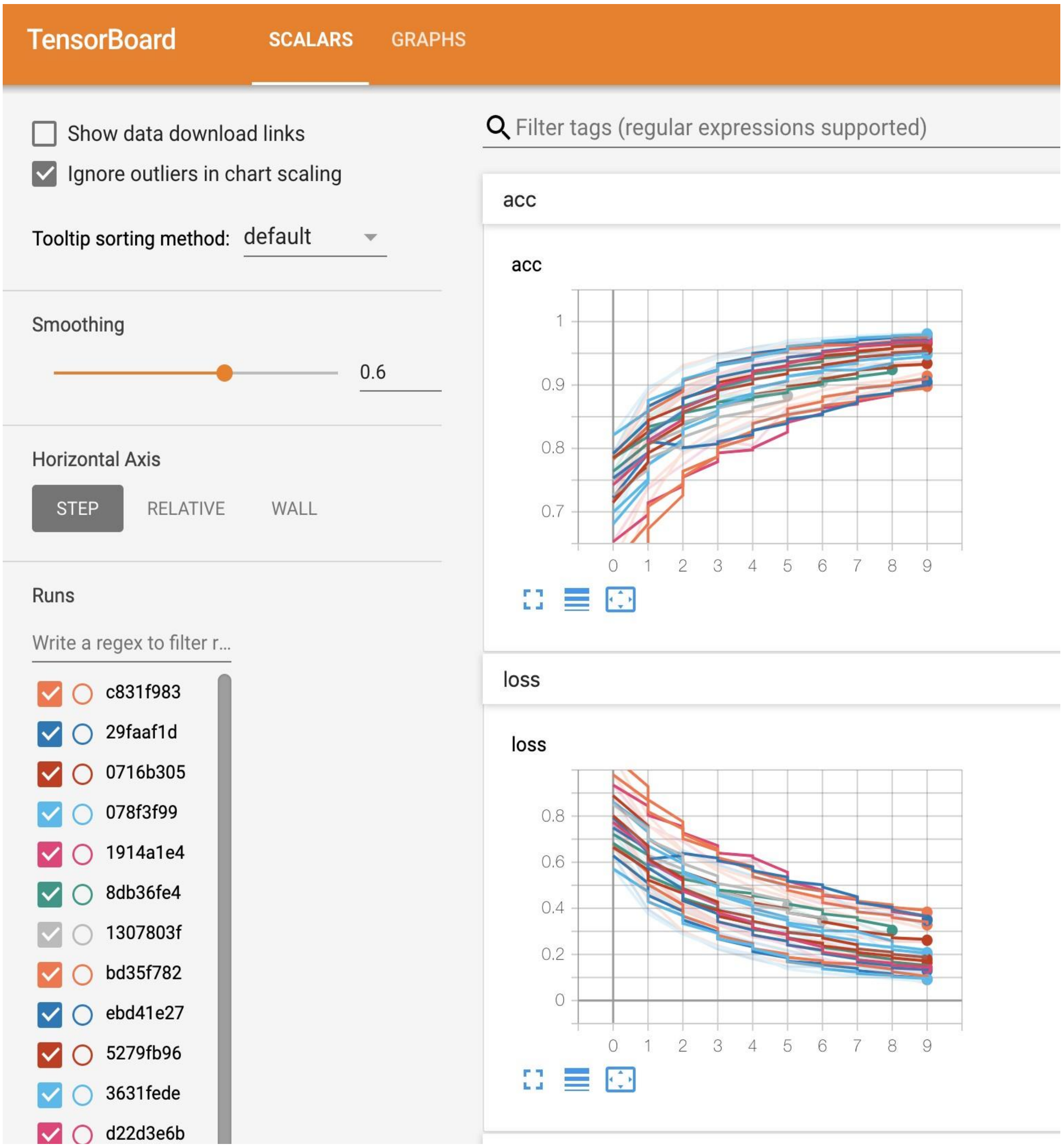
Job Details								Parameters			Metrics			
☐	Job ID	Status	Launched	Duration	User	Tags		num_conv_bl...	num_conv_filt...	num_dense_	num_thre...	test_accuracy	test_f1_score	
☐	d2bc0552-5...		Sep 26 9:27 AM	20m 30s	rt			4	32	1	not available	not available	not available	
☑	caeac6cc-e...	✓	Sep 26 8:47 AM	38m 22s	rt			4	32	4	0.3	0.91	0.65	
☑	f47f7fa1-aa...	✓	Sep 26 7:53 AM	52m 28s	rt			8	32	1	0.65	0.91	0.67	
☑	5ae22fe4-9...	✓	Sep 26 7:02 AM	50m 8s	rt			12	16	5	0.8	0.73	0.34	
☑	aa4b9791-7...	✓	Sep 26 6:08 AM	52m 37s	rt			8	16	5	0.69	0.93	0.67	
☑	ba1e51ca-5...	✓	Sep 26 5:27 AM	39m 0s	rt			4	32	1	0.8	0.82	0.48	
☑	822b2f67-0...	✓	Sep 26 4:38 AM	47m 33s	rt			8	32	1	0.8	0.77	0.46	
☑	d22d3e6b-2...	✓	Sep 26 2:45 AM	1h 51m 58s	rt			12	64	2	0.2	0.91	0.65	

Fig 5: Fake Audio Detection


Details For Job 3631fede-7db6-4b14-93f0-4bebe484d686

Artifact Name
fale_negative_0.png
false_positive_0.png
saved_model_256_0.1_10_0.000'
true_negative_0.png
true_positive_0.png

Fig 6: Audio Dataset



TensorBoard SCALARS GRAPHS

□ Show data download links

☑ Ignore outliers in chart scaling

Tooltip sorting method: default

Smoothing 0.6

Horizontal Axis: STEP RELATIVE WALL

Runs

Write a regex to filter r...

- ☑ ○ c831f983
- ☑ ○ 29faaf1d
- ☑ ○ 0716b305
- ☑ ○ 078f3f99
- ☑ ○ 1914a1e4
- ☑ ○ 8db36fe4
- ☑ ○ 1307803f
- ☑ ○ bd35f782
- ☑ ○ ebd41e27
- ☑ ○ 5279fb96
- ☑ ○ 3631fede
- ☑ ○ d22d3e6b

Filter tags (regular expressions supported)

Fig 7: Accuracy and Loss Audio Fake news DetectionThe FNSMP’S initiative is projected to have a threefold effect:

• This project will help social media users recognize whether multimedia material has been changed from its original form. These are intended to reduce the harm caused by false news while bolstering the standing of real news and unique content. Users of social media platforms will have a higher level of awareness thanks to the proposed digital fake news-based system, which will enable the detection of bogus news and its origin. In addition, it will provide readers with resources for counteracting the potentially negative effects of false news. The suggested detection techniques would aid in the identification of deep fakes, which are particularly difficult to discover and very destructive for persons participating in such multimedia materials, with the same overarching purpose of raising user awareness. Even if a GAN can trick certain false detection systems, the process of extracting relevant characteristics from bogus information will be secure. Active signal processing methods will be used alongside passive ones in our approach. Active approaches, such as data concealing in multimedia material, when combined with a multimodal solution, will be a remarkable new way of detecting fraudulent content.

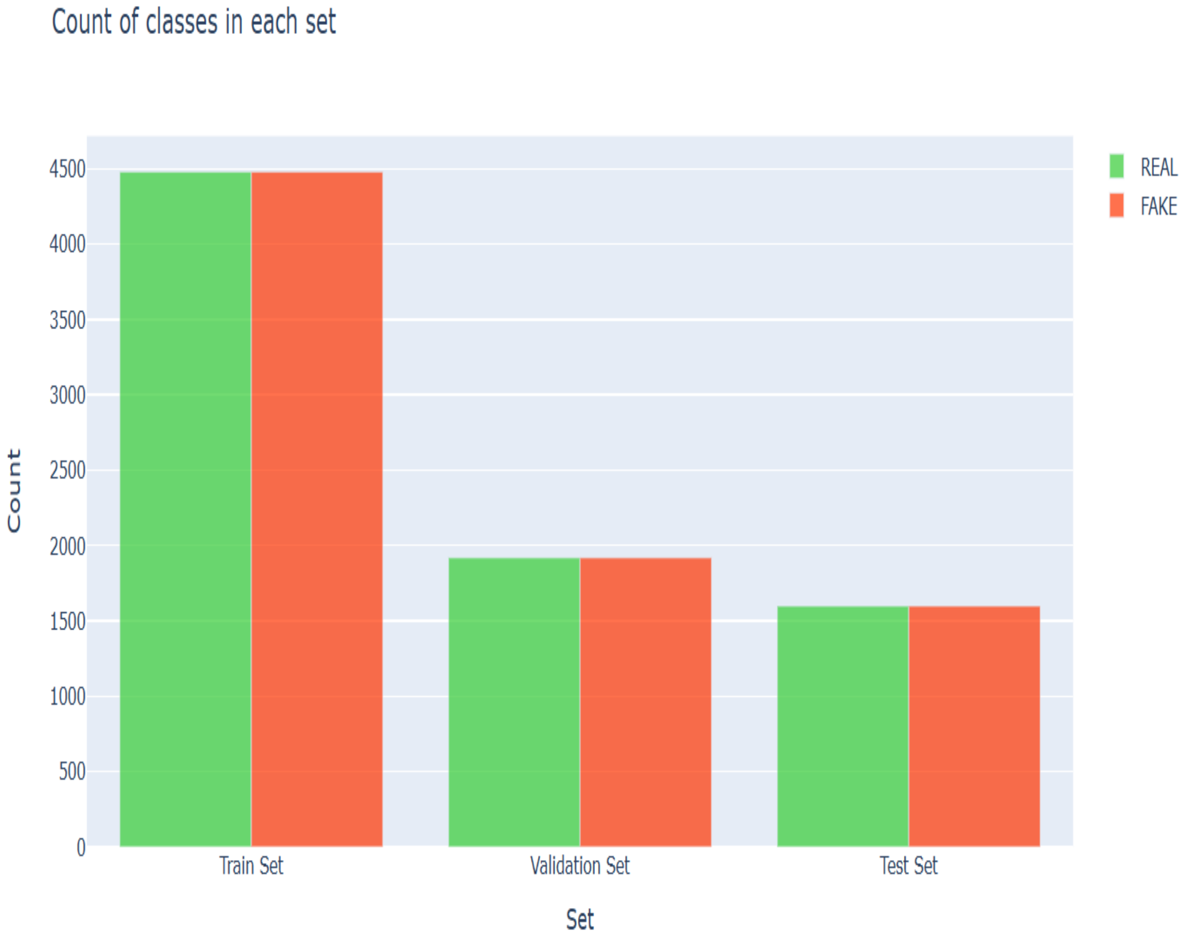


Fig 8: Audio and Video Train, Test and Validation

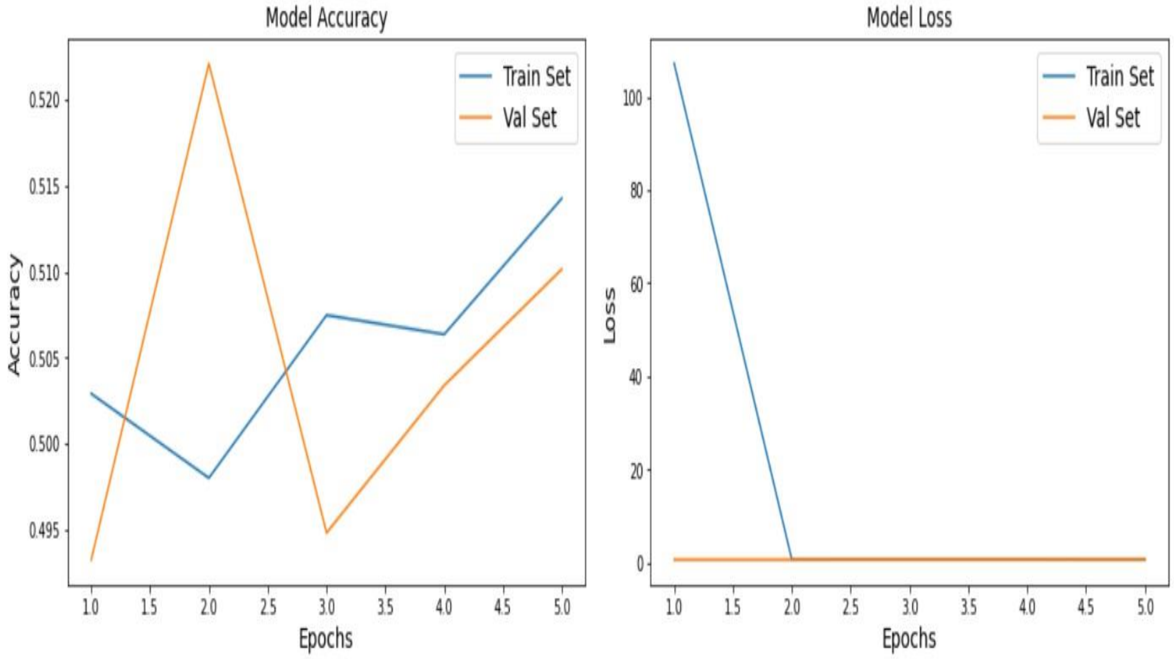


Fig 9: Model Audio and Video Accuracy and Loss

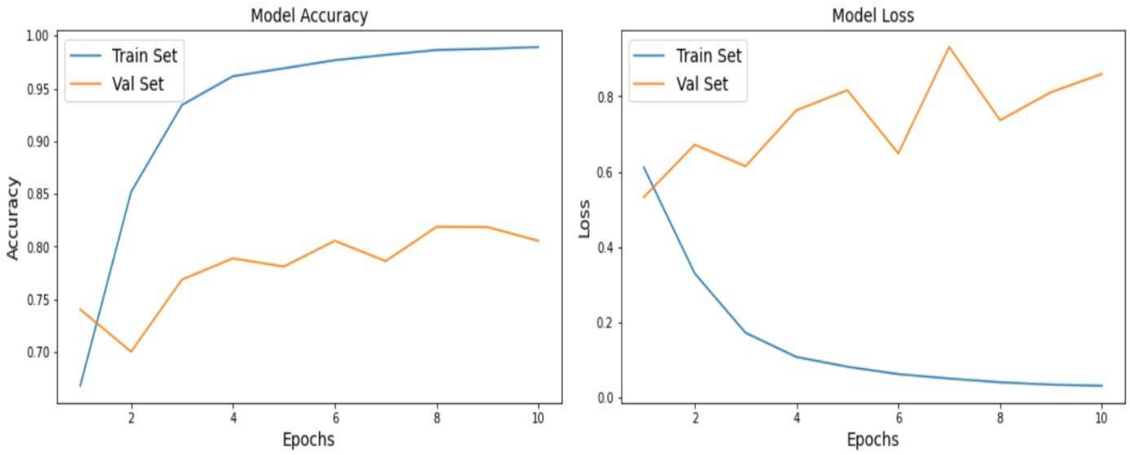


Fig 10: Model Image Accuracy and Loss

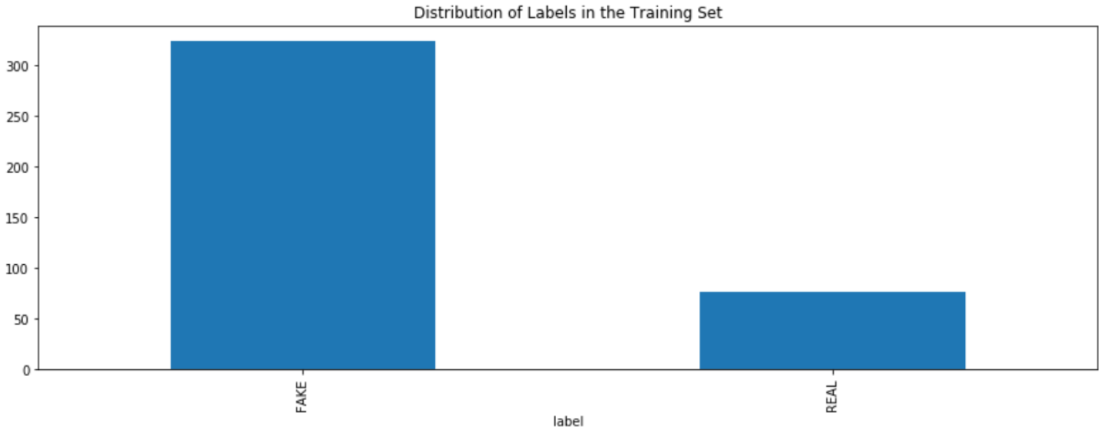


Fig 11: Fake and Real Comparison

- From the perspective of content creators, digital fake news will serve to counter the connection between untraceable content production or modification and an authorship factor by monitoring authorship in the creation (and, perhaps, transformation) of digital material. The use of watermarks might protect a content creator's reputation and, by extension, the integrity of digital networks from the damage caused by falsely credited information.
- From a research standpoint, this project will result in several papers that will advance research in a variety of technical areas (i.e., digital fake news, ML, and signal processing). Publications based on user experience research will also shed light on the societal factors that contribute to the effect of false news and its dissemination.

Conclusions and Future Work:

The spread of false information via social media is a growing concern for all communities, and there is evidence to suggest it may have devastating consequences. So, in this research, we provide a system for identifying bogus news in multimedia. Three companies from three different countries—Spain, Poland, and Japan—are working together on this technology as part of the FNSMP project, which will run through June 2021.

FNSMP'S ultimate objective is to develop methods for detecting bogus news in online communities by combining digital fake news, machine learning, and signal processing. In addition, we want to conduct user experience research to learn how different users (in terms of region, gender, age, etc.) interact with the tools we create. This will help us make informed decisions about how to best design, implement, integrate, and evaluate the tools we create. Keep in mind that this initiative is meant to raise awareness about the problem of false news on social media and to encourage the greater study of this phenomenon from a variety of academic and professional sectors. It is important to note that FNSMP is an interdisciplinary approach that incorporates assessments of the social and cultural issues posed by false news together with information hiding tactics, machine learning techniques, and multimedia forensics.

This article serves as the first step toward the aforementioned objectives. The FNSMP'S framework's architecture was carefully characterized, as was the test-bed platform that will be used to compare different false news detection strategies.

To demonstrate that a platform for detecting false news might be built using a mix of data concealing, machine learning, and multimedia forensic approaches that would be more effective than partial solutions alone, we propose to develop a prototype called FNSMP'S. Although such a proof-of-concept prototype will be supplemented with the first set of user-centric digital fake news and machine learning algorithms, we intend to release it to the security community so that other researchers and developers can improve upon it and create even more effective solutions by modifying some of its components.

In the future, we want to develop the core features of the platform, including digital fake news methods, machine learning-based detection strategies, and multimedia forensics. Last but not least, we want to build an assessment framework that can consistently run trials on various detection algorithms using real- world data sets of multimedia material.

References:

1. A. Hunt and G. Matthew. Social media and fake news in the 2016 election. Journal of Economic Perspectives, 31(2):21–36, May 2017.
2. A. Marcus. Cross-cultural user-experience design. In Proc. of the SIGGRAPH Asia 2011 Courses(SA'11), Hong Kong, China, pages 1–201. ACM, December 2006.
3. P. Dourish. Implications for design. In Proc. of the SIGCHI Conference on Human Factors in Computing Systems (CHI'06), Montre´al, Que´bec, Canada, pages 541—550. ACM, April 2006.
4. D. Meg´ıas, M. Kuribayashi, A. Rosales, and W. Mazurczyk. Dissimilar: Towards fake news detection using information hiding, signal processing, and machine learning. In Proc. of the 16th International Conference on Availability, Reliability, and Security (ARES'21), Vienna, Austria, pages 1–9. ACM, August 2021.
5. M. Castells. Communication Power. Oxford University Press, 2009.
6. H. Jenkins, S. Ford, and J. Green. Spreadable Media, Creating Value and Meaning in a Networked Culture. New York University Press, 2013.
7. A. Toffler. The Third Wave. William Morrow & Company, 1980.
8. C. R. Sunstein. Echo Chambers: Bush V. Gore, Impeachment, and Beyond. Princeton Digital Books+, 2001.
9. L. Guo, J. A. Rohde, and H. D. Wu. Who is responsible for Twitter's echo chamber problem? Evidence from 2016 U.S. election networks. Information Communication and Society, 23(2):234–251, July 2020.
10. R. Colmenarejo. La digitalizacio´n como paradigma: retos e´ticos emergentes. In Dia´logos entree´tica y ciencias sociales. Teor´ıa e investigacio´n en el campo social. Editorial Universidad Icesi, 2021.
11. C. R. Sunstein. On rumors, How falsehoods spread, why we believe them, and what can be done. Princeton University Press, 2014.
12. A. Hunt and G. Matthew. Social media and fake news in the 2016 election. Journal of Economic Perspectives, 31(2):21–36, May 2017.
13. H. Jenkins, M. Ito, and D. Boyd. Participatory Culture in a Networked Era. Polity, 2016.
14. TrustProject. The trust project – news with integrity. https://thetrustproject.org/ [Online; accessedon March 10, 2022].
15. D. Meg´ıas. Data hiding: New opportunities for security and privacy? In Proc. of the European Interdisciplinary Cybersecurity Conference (EICC'20), Rennes, France, pages 1–6. ACM, November 2020.
16. G. R. Blakley, C. Meadows, and G. B. Purdy. Fingerprinting long forgiving messages. In Proc. of the Advances in Cryptology (CRYPTO'85), Santa Barbara, CA, USA, volume 218 of Lecture Notes in Computer Science, pages 180–189. Springer Berlin Heidelberg, August 1986.
17. D. Boneh and J. Shaw. Collusion-secure fingerprinting for digital data. IEEE Transactions on Information Theory, 44(5):1897–1905, September 1998.
18. P. Korus and N. Memon. Content authentication for neural imaging pipelines: End-to-end optimization of photo provenance in complex distribution channels. In Proc. of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR'19), Long Beach, CA, USA, pages 8613–8621. IEEE, June 2019.
19. Y. Li, M.-C. Chang, and S. Lyu. In ictu oculi: Exposing ai created fake videos by detecting eye blinking. In Proc. of the IEEE International Workshop on Information Forensics and Security (WIFS'18), Hong

Kong, China, pages 1–7. IEEE, December 2018.

20. D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen. Mesonet: a compact facial video forgery detection network. In 2018 IEEE International Workshop on Information Forensics and Security (WIFS'18), Hong Kong, China., pages 1–7. IEEE, December 2018.
21. P. Korshunov and S. Marcel. Deepfakes: a new threat to face recognition? assessment and detection. arXiv preprint, abs/1812. 08685:1–5, December 2018.
22. R. Tolosana, R. V.-Rodriguez, J. Fierrez, A. Morales, and J. O.-Garcia. Deepfakes and beyond: A survey of face manipulation and fake detection. Information Fusion, 64:131–148, December 2020.
23. B. Dolhansky, R. Howes, B. Pflaum, N. Baram, and C. C.-Ferrer. The deepfake detection challenge (DFDC) preview dataset. arXiv preprint, abs/1910.08854, October 2019.
24. I. Cox, M. Miller, J. Bloom, J. Fridrich, and T. Kalker. Digital Fake news and Steganography. Morgan Kaufmann Publishers Inc. San Francisco, CA, USA, 2008.
25. A. Adelsbach and A.-R. Sadeghi. Zero-knowledge watermark detection and proof of ownership. In Proc. of the International Workshop on Information Hiding (IH'01), Pittsburgh, PA, USA, volume 2137 of Lecture Notes in Computer Science, pages 273–288. Springer Berlin Heidelberg, October 2001.
26. J. Serra-Ruiz and D. Meg´ıas. A novel semi-fragile forensic fake news scheme for remote sensing images. International Journal of Remote Sensing, 32(19):5583–5606, August 2011.
27. O. Benrhouma, H. Hermassi, A.A. Abd El-Latif, and S. Belghith. Chaotic watermark for blind forgery detection in images. Multimedia Tools and Applications, 75(14):8695–8718, July 2016.
28. J. Serra-Ruiz, A. Qureshi, and D. Meg´ıas. Entropy-based semi-fragile fake news of remote sensing images in the wavelet domain. Entropy, 1–21(9):847, August 2019.
29. D. Meg´ıas and A. Qureshi. Collusion-resistant and privacy-preserving P2P multimedia distribution based on recombined fingerprinting. Expert Systems with Applications, 71:147–172, April 2017.
30. M. Kuribayashi and N. Funabiki. Fingerprinting for the multimedia content broadcasting system. Journal of Information Security and Applications, 41:52–61, August 2018.
31. M. Kuribayashi and N. Funabiki. Decentralized tracing protocol for the fingerprinting system. APSIPA Transactions on Signal and Information Processing, 8(1): e2, January 2019.
32. D. Meg´ıas, M. Kuribayashi, and A. Qureshi. Survey on decentralized fingerprinting solutions: Copyright protection through piracy tracing. Computers, 9(2):1–26, April 2020.
33. A. Qureshi, D. Meg´ıas, and M. Kuribayashi. Detecting deepfake videos using digital fake news. In Proc. of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference, (APSIPAASC'21), Tokyo, Japan, pages 1786–1793. IEEE, December 2021.
34. P. Korshunov and S. Marcel. Speaker inconsistency detection in tampered video. In Proc. of the 26th Eu- European Signal Processing Conference (EUSIPCO'18), Rome, Italy, pages 2375–2379. EURASIP, February 2018.
35. F. Matern, C. Riess, and M. Stamminger. Exploiting visual artifacts to expose deepfakes and face manipulations. In Proc. of the IEEE Winter Applications of Computer Vision Workshops (WACVW'19), Waikoloa, HI, USA, pages 83–92, January 2019.
36. I. Goodfellow, Y. Bengio, and A. Courville. Deep Learning. MIT Press, 2016.
37. S. Agarwal, H. Farid, Y. Gu, M. He, K. Nagano, and H. Li. Protecting world leaders against deep fakes. In Proc. of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR'19), California, USA, pages 38–45. IEEE, June 2019.
38. P. L. Shrestha, M. Hempel, F. Rezaei, and H. Sharif. A support vector machine-based framework for detection of covert timing channels. IEEE Transactions on Dependable and Secure Computing, 13(2):274–283, April 2015.
39. T. Jung, S. Kim, and K. Kim. Deepvision: Deepfakes detection using human eye blinking pattern. IEEE Access, 8:83144–83154, April 2020.
40. Y. Li and S. Lyu. Exposing deepfake videos by detecting face warping artifacts. In Proc. of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW'19), Long Beach, CA, USA, pages 46–52. IEEE, June 2019.
41. World Bank. World bank country and lending groups – world bank data help desk, July

2021. https://datahelpdesk.worldbank.org/knowledgebase/articles/ 906519-world-bank-country-and- lending-groups [Online; accessed on March 10, 2022].

42. D. Lerch-Hostalot and D. Meg´ıas. Unsupervised steganalysis based on artificial training sets. Engineering Applications of Artificial Intelligence, 50:45–59, April 2016.
43. D. Lerch-Hostalot and D. Meg´ıas. Detection of classifier inconsistencies in image steganalysis. In Proc. of the ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec'19), Paris, France, pages 222–229. ACM, July 2019.
44. G. Apurva and J. Shomik. Adversarial perturbations fool Deepfake detectors. In Proc. of the 2020 International Joint Conference on Neural Networks (IJCNN'20), Glasgow, United Kingdom, pages 1–8. IEEE, July 2020.
45. C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus. Intriguing properties of neural networks. arXiv preprint, abs/1312.6199:1–10, December 2013.
46. A. Alattar, R. Sharma, and J. Scriven. A system for mitigating the problem of deepfake news videos using fake news. Electronic Imaging, 2020(4):11701–11710, January 2020.
47. J. Fridrich and J. Kodovsky. Rich models for steganalysis of digital images. IEEE Transactionson Informa- tion Forensics and Security, 7(3):868–882, May 2012.
48. D. Cozzolino, G. Poggi, and L. Verdoliva. Recasting residual-based local descriptors as convolutional neural networks: An application to image forgery detection. In Proc. of the 5th ACM Pennsylvania, Philadelphia, USA, pages 159–164. ACM, June 2017.
49. I. Masi, A. Killekar, R.M. Mascarenhas, S.P. Gurudatt, and W. AbdAlmageed. Two-branch recurrent network for isolating deepfakes in videos. In Proc. of the 16th European Conference on Computer Vision (ECCV'20), Glasgow, UK, volume 12356 of Lecture Notes in Computer Science, pages667–684. Springer International Publishing, August 2020.
50. Q. Yuyang, Y. Guojun, S. Lu, C. Zixuan, and S. Jing. Thinking in frequency: Face forgery detection by mining frequency-aware clues. In Proc. of the 16th European Conference on Computer Vision (ECCV'20), Part XII, Glasgow, UK, volume 12357 of Lecture Notes in Computer Science, pages 86–103. Springer, Cham, August 2020.
51. A. Qureshi and D. Meg´ıas. Blockchain-based P2P multimedia content distribution using collusion-resistant fingerprinting. In Proc. of the Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPAASC'19), Lanzhou, China, pages 1606–1615. IEEE, November 2019.
52. S. Nakamoto. Bitcoin: A peer-to-peer electronic cash system, 2008. https://bitcoin.org/bitcoin.pdf[Online;accessd on March 10, 2022].