COMBATING FAKE NEWS, MISINFORMATION, AND MACHINE LEARNING GENERATED FAKES: INSIGHTS FROM THE ISLAMIC ETHICAL TRADITION

Talat Zubair*
Amana Raquib**
Junaid Qadir***

Abstract: The growing trend of sharing and acquiring news through social media platforms and the World Wide Web has impacted individuals as well as societies, spreading misinformation and disinformation. This trend—along with rapid developments in the field of machine learning, particularly with the emergence of techniques such as deep learning that can be used to generate data—has grave political, social, ethical, security, and privacy implications for society. This paper discusses the technologies that have led to the rise of problems such as fake news articles, filter bubbles, social media bots, and deep-fake videos, and their implications, while providing insights from the Islamic ethical tradition that can aid in mitigating them. We view these technologies and artifacts through the Islamic lens, concluding that they violate the commandment of spreading truth and countering falsehood. We present a set of guidelines, with reference to Qur‘anic and Prophetic teachings and the practices of the early Muslim scholars, on countering deception, putting forward ideas on developing these technologies while keeping Islamic ethics in perspective.

Keywords: Islamic ethics, machine learning, fake news, misinformation, disinformation.

Introduction

Deception can be described as an act of “intentionally causing another person to have or continue to have a false belief that is truly believed to be false by the person intentionally causing the false belief by bringing about evidence on the basis of which the other person has or continues to have that false belief.”1 Deception has always existed in some form or another; rumour, partisan news and manipulative content are nothing new. However, in the present day, deception has a far greater range of influence and manipulation than ever before,
facilitated by the ease of sharing information on social media platforms, the rapid accessibility of uploaded information on the World Wide Web, and the rapidly progressing fields of Artificial Intelligence (AI) and Machine Learning (ML). As a result, various new ethical issues have arisen.2

The information we consume impacts upon our perception of the truth, altering our views about politics, economics, relationships, needs and wants. In the present, the impact of misinformation, disinformation and artificially generated content has increased to the extent that we find ourselves questioning if our opinions or views are actually ours or the result of the influence of deceitful content viewed on the web. As a result, it is now necessary to create strategies for countering deceptive artifacts.

Experts distinguish between misinformation and disinformation. A piece of information is labelled as misinformation when it is false or misleading—e.g. sharing a fraudulent online article without verifying the authenticity of its source would be misinformation. A piece of information is labelled disinformation, on the other hand, when it is deliberately falsified to obscure the truth—e.g. spreading false news with the intention to harm an individual’s reputation constitutes disinformation.

In this paper, we highlight the problems that arise due to misinformation, disinformation, and the generation of fake digital artifacts. We review the relevant background and introduce the problems arising from misinformation, disinformation, and fake news in Section II. In Section III, we provide guidelines from the Islamic tradition on facing these issues and their various associated ethical problems. In Section IV, we provide a sketch of a technical solution for verifying reports using the science of Hadith. We conclude the paper in Section V.

Background

In this section, the advancements of technology and the impact they can have with regards to deceptively altering an individual’s perception and influencing their decisions and choices is discussed.

A. Problems of Disinformation and Misinformation

Jonathon Swift, the eighteenth-century Anglo-Irish author, pithily remarked that, “falsehood flies, and the truth comes limping after it.”3 This phenomenon has not only subsisted through time, but has greatly increased in intensity during the present with the emergence of online social networks on the Internet. The use of fake accounts operated by software bots also contributes to the spread of
misinformation on social media through the interaction of fake accounts with genuine accounts, exploiting social connections. Since social connections are trusted, the accounts are able to post content that can manipulate readers into believing spurious news items.\(^4\)

The impact of the spread of fake news has been witnessed globally. In 2018, five people were killed by a mob in India, with a wave of violence following, after untrue messages about child abductions were forwarded as WhatsApp messages.\(^5\) In 2016 the Pakistani defense minister, Khawaja Muhammad Asif, threatened the Israeli defense minister with Pakistan’s nuclear weapons after reading a false news report.\(^6\) Moreover, before the Spanish elections of 2019, WhatsApp messages were circulated among Spaniards spreading fake news and hateful comments about the Prime Minister, Pedro Sánchez, with more than one in four Spaniards reportedly having received false and misleading posts.\(^7\) In France, following the fire at Notre Dame Cathedral in Paris, fake videos and posts claiming the involvement of Muslims became viral, giving rise to Islamophobic reactions.\(^8\) An MIT study conducted in 2018 reported that false news stories spread more rapidly on social media networks than authentic ones, concluding that they are 70 per cent more likely to be shared on Twitter and that authentic stories take six times longer to reach 1,500 people compared to the time taken by fake news.\(^9\)

Fake news can have a strong impact on listeners and readers. Repeated exposure to a piece of information makes it familiar, until it is eventually perceived as acceptable and valid. This phenomenon is known as the “illusory truth” effect.\(^{10}\) A possible illustration of this effect took place during the US elections of 2016; Donald Trump’s victory has been attributed to various pro-Trump false news reports extensively shared on Facebook prior to the elections.\(^{11}\) The illusory truth effect is also commonly used in marketing, where the repetition of potentially deceptive and/or exaggerated claims about the marketed product or service can be used to sway buyers.

### B. The Problem of Filter Bubbles

Social media websites, including Facebook, Google Plus, Twitter and LinkedIn, construct ‘filter bubbles’ that allow users to only view content they agree with or that aligns with their pre-existing beliefs.\(^{12}\) The term ‘filter bubble’ was introduced by Eli Pariser, who alerted readers about the associated dangers in his book *The Filter Bubble: What the Internet Is Hiding from You*.\(^{13}\) According to Pariser, a filter bubble is an individual’s personal and unique online space, the nature of which depends on their online identity. The individual has no say regarding what penetrates into the space and what gets edited out. Furthermore,
social media users are isolated in their own cultural or socio-political ‘cocoon’ where they are shielded from uncomfortable news items and views that contradict their beliefs. Pariser writes that social media websites, through creating filter bubbles, are “fundamentally altering the way we encounter ideas and information.”

C. Weapons of Mass Deception

Recently developed technology like generative ML models, including Generative Adversarial Networks (GAN), have blurred the line between genuine and fake digital artifacts, thereby exposing content consumers to a commoditised form of mass deception. For example, the technology now exists for creating realistic-looking fake videos in which individuals appear to be speaking or doing something they never actually did. In a study by researchers from the University of Washington, a GAN-based ML model was trained on videos of presidential addresses by former United States president Barack Obama in order to create fake videos of him making an arbitrary statement, with his lip movements realistically synchronised. With such technology, creating deep-fake videos is as easy as editing the text of a video transcript in order to change the speaker’s words. In other recent work, researchers from Samsung have developed a new ML algorithm that takes still images of individuals as input, producing as output a video of them speaking. Moreover, the OpenAI lab in San Francisco, in collaboration with the Allen Institute for Artificial Intelligence in Seattle, have built an artificial “fake-text” generator that mimics human writing, thereby producing machine-composed writing that could pass as genuine. The technology poses such a serious threat that OpenAI considered the technology “too dangerous to be released into the real world.” Disturbingly, deep-fake technology has been used to create fake pornographic videos by swapping the faces of real women with pornographic actresses, while DeepNude is an application that uses the same technology to remove clothing from the images of women, generating genuine-looking photos of the person naked.

The availability of such technology raises many ethical issues. It has the potential to make any individual appear engaged in an act they did not commit, such as showing a political leader making a racist remark, a candidate being involved in criminal activities before elections, or a public officer committing an adulterous act. It can affect an individual’s control over their reputation, while being a threat to human dignity, national security, and the well-being of human societies. Tackling the relevant issues requires the urgent attention of researchers from multiple disciplines, including the humanities, in addition to researchers from technological fields.
D. Technical Solutions

Several technical solutions have been proposed for identifying and mitigating fake news.23

1. **Initiatives by Major Internet Companies:** Major Internet companies, such as Google and Facebook, are working to stem the problem of fake news. Facebook is fighting the spread of fake news in multiple ways, including, among other things, by improving its ranking system for posts shown in the News Feed and expediting the reporting of misleading content. The Facebook Journalism Project aims to collaborate with news organisations and journalism experts to improve the quality of information shared on the platform.24 A research report from 2018 concluded that Facebook’s role in the dissemination of fake news has been reduced.25 Facebook-owned WhatsApp launched a fact-checking service for Indians ahead of national elections, where users can verify if the shared information is “true, false, misleading or disputed.”26

   Google has presented a white paper outlining measures to prevent the spread of disinformation through Google Search, Google News, YouTube and Google advertising products. The company has mentioned attempts to enhance its search result ranking system by ranking content on the basis of expertise, trustworthiness and authority. It has also stated that it aims to take strict action against misleading websites, spamming and impersonation.27

2. **Blockchain:** A widely discussed and debated solution to the creation and spread of fake content is the implementation of blockchain technology. Blockchain is a digital ledger that can maintain a record of content shared online, in addition to any alterations made to it, making it possible for all alterations to be traced back to their initiator. In this way, the authenticity of news items can be verified by tracing them back to their origin. Chen et al. have presented a solution for preventing the spreading of rumours on social media networks using blockchain.28 To protect the privacy of media shared on social networks, their suggestion is to cryptographically store the media on blockchain and thereby ensure that every interaction involving the content can be detected.29 Another approach is to make the content stored on the blockchain immutable, to make sure the saved content cannot be altered, changed or deleted.30 Blockchain technology can be used to detect deep-fake videos; one proposed solution is to trace back different versions of the video to the original. If a video cannot be traced to the original creator, then it is identified as fake.31
3. **Ethics Initiatives**: The IEEE Global Initiative has published a report entitled “Ethically Aligned Design”, intended to guide and encourage the development of autonomous systems that are primarily focused on human well-being and protecting human rights through the prevention of the misuse of AI, thereby ensuring system transparency and the development of a framework for developer accountability. The report is a collective effort produced by several hundred individuals from various countries across the globe, all of whom possess expertise in science and technology, law, governance, policy-making and civil society.

The AI Foundation, whose stated mission is to build tools to protect against the risks of AI, has developed the intelligent software, “Reality Defender”, to detect potentially fake media in the digital world. The program scans images, videos and other media to identify known fakes, runs AI-driven analysis techniques to detect alterations in the original content, and allows for reporting of suspected fakes.

4. **Other Proposed Solutions**: Another solution for preventing rumours from spreading on social media platforms is to develop a rumour classification system. This would involve a 4-step process that starts by identifying if a piece of information is a rumour or a non-rumour using a binary classifier. Then a rumour tracking component would be introduced that gathers posts related to and that discuss the rumour while outputting a collection of these posts. The third step would involve stance classification, which uses the output from the rumour tracking component and labels it by determining the stance of each post on the rumour; these labels would include Supporting, Denying, Quarrying and Commenting. The process then ends by assigning a truth value to the rumour through veracity classification, labeling a rumour as true, false or unverified. While the scope of the study only covers the detection and resolution of rumours (which can be true or false), the same work can be done for fake news and hoaxes (which are always fake).

Kumar and Shah present a method to detect opinion-based (individual opinion, whether honest or dishonest, e.g. product reviews) and fact-based (statements contradicting, fabricating or agreeing with a true statement, e.g. rumours) false information on both social media and the web. The paper first outlines the characteristics of each type of false information, outlining that fake opinion-based statements (e-commerce reviews in this case) tend to be shorter and more exaggerated, with extreme ratings and short inter-arrival times between subsequent ratings, whereas fake fact-based information (news articles in this case) usually have an incoherent body and headline, are posted by newly registered accounts, and tend to spread faster.
and quicker on multiple platforms. The second part of the paper discusses the detection of false information using feature-based, graph-based and modeling-based algorithms.

A detailed survey of various techniques proposed for detecting fake news is provided in Zhou and Zafarani. A piece of news is analysed for its authenticity from four perspectives: (1) knowledge-based (knowledge that a news story provides); (2) style-based (intention behind the news); (3) propagation-based (how the news is spread); and (4) credibility-based (assessing the credibility of the news spreaders and creators). Each part of the process from news creation to publication and dissemination is evaluated from these four perspectives.

Islamic Guidelines for the New Online World

A. Islamic Guidelines on Information Dissemination and Sharing

The Islamic ethics of disseminating information requires information literacy, which is the set of skills needed to discover, evaluate, interpret, and use information properly and truthfully. With reference to these skills, information dissemination can be broken down into three steps: acquiring, evaluating and sharing. The first phase of acquiring information dictates that a Muslim verify the credibility of their sources. Reporters and news transmitters need to thoroughly check the reliability of their sources before they accept or deny what they are told. A guideline to verification for consumers of news, however, includes knowing the author of the article and the sponsor of the medium, determining if there is an agenda underlying the content, what kind of page it is, and for online sources whether the URL is relevant to the content, along with ensuring the information is complete, accurate and current. With the growth of media sources, determining the credibility of every source can be a challenge, hence, if authentication is not possible, it is advisable to suspend judgment. The Qur’an says (17:36): “Do not follow that of which you have no knowledge. Surely the hearing, the sight, the heart - each of these shall be called to account.” Muslims should take inspiration from the hadith scholars, who were very cautious about accepting narrations from unknown individuals, or individuals whose identity and character were doubtful. In addition to this, the Prophet (pbuh) also commented: “Indeed among the excellence of a person’s Islam is that he leaves what does not concern him.” Following these teachings, Muslims are advised to ignore news that does not affect them, whether in this world or in the Hereafter. This includes celebrity gossip, scandals and other news that may seem interesting but is of no benefit.
The second phase is evaluating the information—or, where Islam guides Muslims to seek clarification. Upon receiving a piece of information, an individual should investigate it before acting upon it.44 This command is given in the Qur’an, in the following verse: “O you who have believed, if there comes to you a disobedient one with information, investigate, lest you harm a people out of ignorance and become, over what you have done, regretful.”45 When judging news, it is advisable to do so without prejudice.

After the initial two phases of acquiring and assessing information, the third phase demands that news be shared in line with Islamic ethics. It is an Islamic belief that human intellect, wisdom, understanding and reason are a trust (amanah) from Allah and should not be used to harm another human soul. Hence, before spreading any information, Muslims have a responsibility to consider the impact it could have on an individual. The Prophet (pbuh) is reported to have said: “A Muslim is the one from whose tongue and hand the people are safe, and the believer is the one from whom the people’s lives and wealth are safe,”46 and “He who believes in Allah and the Last Day must either speak good or remain silent.”47 Muslims, therefore, need to be cautious when clicking on the share button or forwarding a message that can be insulting, hurtful or offensive to a sect, religion or individual. This cautiousness also applies to spreading misinformation in the form of spam messages, chain mails, conspiracy theories, or fake news articles, as these can be damaging to society.48 By controlling the virality of fake news, its impact and acceptability can be reduced.

B. Responsible Journalism and the Dissemination of News

“Say, ‘Bring your proof, if you are truthful.”’ Qur’an 2:111.

In Islam, a lot of emphasis is placed on the verification of truth, the authenticity of sources, and the proper management of knowledge. This is evident from the efforts Muslim scholars have gone to over the centuries to preserve and authenticate the sayings of the Prophet (pbuh).49 The Qur’an commands Muslims to verify the source of all information they receive before believing it to be true: “O you who have believed, if there comes to you a disobedient one with information, investigate, lest you harm a people out of ignorance and become, over what you have done, regretful.”50 This command also applies to frequently shared pieces of information found on social media, news channels and newspapers. A Muslim is not supposed to act on an unexamined source; all information should be authenticated.
Muchtar et al. have reported on journalistic roles in twelve Muslim majority countries, while also describing the normative Islamic principles regarding news reporting and journalism based on four principles: (1) truth-telling (haqq); (2) pedagogy and communication (tabligh); (3) seeking public interest (maslahah); and (4) moderation (wasatiyyah). In another work, Pintak identified the additional concepts of justice (‘adl), independence and sincerity (nasihah), balance (iʿtidal), and prevention of evil attitude and public accounting (hisbah) as being critical to Islamic journalism.

Having considered the impact inauthentic information can have at the political and social levels, the responsibility falls on Muslims to take action. The Qur’an says: “And let there be [arising] from you a nation inviting to [all that is] good, enjoining what is right and forbidding what is wrong, and those will be the successful.” The command at the beginning of this verse, “let there be [arising] from you a nation”, implies a collective obligation on the part of the community. A relevant hadith also directs Muslims to take action against evil: “Whoever among you sees an evil, let him change it with his hand; if he cannot, then with his tongue; if he cannot, then with his heart—and that is the weakest of Faith.” A suggested approach to strengthen these principles is to establish a centre for Islamic journalism as a platform for providing media services that strictly adhere to an Islamic framework for the dissemination of news, while also encouraging positive interaction between practitioners and scholars. The Islamic obligation to enjoin what is right and forbid what is wrong includes institutions of social communication, including the press, radio, television, and cinema. It can also be extended to social media platforms. Furthermore, there is a need to train Muslim journalists in editorial tasks, circulation, distribution, advertisement and effective use of new communication technologies within the framework provided by Islam.

Social media posts or website articles receive more attention when they are accompanied by a huge number of likes, shares, or comments. Islam prohibits Muslims from sharing unverified information. Allah Almighty says in the Qur’an: “When you received it with your tongues and said with your mouths that of which you had no knowledge and thought it was insignificant while it was, in the sight of Allah, tremendous.” The Prophet (pbuh) has also stressed avoiding spreading false news. He said: “The person who goes about with calumnies will never enter Paradise.” The Companions of the Prophet and the second Caliph, Umar ibn al-Khattab, also said: “It is enough of a lie for a man that he narrates everything he hears.”
C. Islamic Scholars on the Verification of Information

1. Ibn Khaldun: the prominent learned figure who wrote *Muqaddimah*. An excerpt from this text reads: “…to establish the truth and soundness of information about factual happenings, a requirement to consider is the conformity (or lack of conformity of the reported information with the general conditions). Therefore, it is necessary to investigate whether it is possible that the (reported facts) could have happened.” Ibn Khaldun writes that partisanship results in the spreading of misinformation: “If the soul is impartial in receiving information, it devotes to that information the share of critical investigation the information deserves, and its truth or untruth thus becomes clear. However, if the soul is infected with partisanship for a sect or an opinion, it accepts without a moment’s hesitation the information that is agreeable to it. Prejudice and partisanship obscure the critical faculty and preclude critical investigation. The result is that falsehoods are accepted and transmitted.” Since the specialisation of Ibn Khaldun was historiography, he describes the complex psychology of humans as a reason for the misrepresentation and falsification of history, saying: “…prejudice and partisanship for a particular opinion and sect...too much unquestioning reliance on past historian[s]...ignorance of how physical conditions conform with reality...unnecessarily lavishing praise on high-ranking personalities. Human souls long for praise and [historians] go for the position and wealth it offers.” Numbers are often easy to fudge towards these ends, as Ibn Khaldun notes: “When historians do not probe with the yardstick of knowledge of the nature of things, they stray away from the truth. This is specially the case with figures either of sums of money or soldiers. These offer a good opportunity for false information and constitute a vehicle for nonsensical statements.”

2. Ibn Al-Haytham: a mathematician, physicist, and astronomer of the Islamic Golden Age (eighth to fourteenth centuries) who wrote that: “the duty of the man who investigates the writings of scientists, if learning the truth is his goal, is to make himself an enemy of all that he reads, and... attack it from every side. He should also suspect himself as he performs his critical examination of it, so that he may avoid falling into either prejudice or leniency.” The message is to verify writings and explanations instead of taking them at face value, to think critically and demand proof.
D. Miscellaneous

1. **Islamic Guidelines on Deep-fake Videos:** To counter indistinguishably realistic deep-fakes that rapidly spread false news, we refer to Islamic rules and ethics to ensure our perceptions are influenced only by honest and trustworthy content, and that we do not become victims to manipulation and deceit. According to Islamic principles, the creation of deep-fake videos raises two major ethical concerns: firstly, the original videos or images used to make the fake videos are often taken without the owner’s consent, which violates their right to privacy, a sacred human right; secondly, the purpose behind the creation of these videos is to show falsehood and conceal the truth, which is strictly against Islamic law.

   According to the Islamic tradition, those who promote and spread falsehood remain deprived of Allah’s guidance and are worthy of severe punishment. Allah says in the Qur’an: “Indeed, Allah does not guide one who is a transgressor and a liar” and “And do not mix the truth with falsehood or conceal the truth while you know [it]”, which emphatically proscribes Muslims from producing such content and sharing it with others.

   The Prophet (pbuh) guided Muslims towards hiding the faults of others. The Prophet (pbuh) is reported to have said: “Whoever covers (the sin of) a Muslim, Allah will cover him (his sin) in this world and in the Hereafter.” Thus, videos that reveal an individual’s crime, whether real or purported, should not be disseminated. Deep-fake videos can be used to accuse public figures of crimes and immoral acts that they did not commit, to defame their character. In this regard, Allah has warned against spreading immorality, saying, “Indeed, those who like that immorality should be spread [or publicised] among those who have believed will have a painful punishment in this world and the Hereafter. And Allah knows and you do not know.”

   We have seen that deep-fake videos have the potential to make an individual accountable for speaking words they did not utter and performing actions they did not perform. In this regard, it is important to refer to the teachings of the Qur’an about believing such accusations and holding assumptions against the supposed perpetrators. The Qur’an says: “Why do not the believing men and women, whenever such [a rumour] is heard, think the best of one another and say, ‘This is an obvious falsehood’?” and “O you who have believed, avoid much [negative] assumption. Indeed, some assumption is sin. And do not spy or backbite each other. Would one of you like to eat the flesh of his brother when dead? You would detest it. And fear Allah; indeed, Allah is accepting of repentance and Merciful.”
Not only should Muslims not be involved in creating deceitful content, but they also need to devise ways to counter deep-fake videos. They should participate in and develop ethical initiatives (such as IEEE’s Ethically Aligned Design) in which Islamic teachings are used to develop solutions to vexing modern technical ethical issues and pave the way for human well-being. One approach could be to explore the development of algorithms that can detect and report discrimination and deceptive content and identities, for this can open the way for combating ML-generated fakes.74

E. A Maqasid al-Shari‘ah Approach

As stated by Imam al-Ghazali, the objectives of Islamic law (maqasid al-shariah) encourage the individual, societal, and communal well-being of humanity through the guarding of religion, life, intellect, dignity and property.75 If these objectives are not met, chaos and disorder will prevail.76 The acts that fulfill these objectives are desirable, whereas the acts that do not should be avoided.77 In this section, we evaluate the ethical issues surrounding deceptive content with reference to each of these objectives.

1. Religion: The protection of religion signifies preserving the faith of individuals from all forms of violence and destruction.78 Dissemination and consumption of content that portrays a falsely negative image of Islam (like the Notre Dame fire incident, see Section II-A) or misrepresents Islamic teachings will negatively affect an individual’s religious beliefs, misleading and deceiving them.

2. Life: The intent to protect life is another objective of Islamic law.79 Fake news that motivates harm cannot therefore be considered permissible in Islam. In Section II-A, we discussed the violent incident (mob lynching) in India, which provides evidence that dissemination of fake news can hurt human life.

3. Intellect: Considering the ‘illusory truth’ effect discussed in Section II-A, false information alters one’s beliefs and affects perceptions of truth. Fake content that appears too authentic to be false makes one skeptical of the truth. In addition to this, articles that share partisan or biased news can manipulate and misinform people: Along with the filter bubbles that restrict information on the web, these can distort one’s views.
4. **Dignity:** Deep-fake videos have the potential to tarnish and dishonour someone’s reputation by producing false content showing their involvement in dishonorable or illegal acts. Any attempt to defame an individual, either through machine-generated content or rumours in the form of web articles, attacks their dignity.

5. **Wealth and Property:** In Section II-C, the discussion about deep learning technology showed how an individual’s personal data (their body and voice) can be manipulated to produce fake content. This goes against the intent of protecting one’s property. Additionally, an immense amount of wealth is going into research for questionable technologies that corrupt the objective of wealth generation and spending of wealth for social benefit.

In order to promote individual and collective good, while preventing disorder in the world, one should adhere to the acts that are intended to preserve religion, life, intellect, dignity and property. Dissemination of false information is a threat to truth, individual privacy, integrity and societal security.

**Verifying Reports Using the Science of Hadith (‘ilm al-hadith)**

Hadith, or oral traditions that report the words and deeds of the Prophet Muhammad (pbuh), are an important tool in the Islamic tradition, both for understanding the Qur’an and for various matters relating to Islamic jurisprudence. Hadith contains two parts: (1) the text (al-matn), which refers to the actual narrative; and (2) the chain of narrators (isnad), which refers to the chronological list of narrators who transmitted the hadith.

During the lifetime of the Prophet (pbuh), the Companions meticulously conveyed what they had directly heard from him. However, the fabrication of hadith, whether intentionally or unintentionally, began around 40 A.H., following the assassination of the third caliph, Uthman, due to political unrest. These fake hadith emerged from various sources, such as: new immigrants from the former Persian and Roman empires; storytellers who invented their own hadiths; and mistakes in copying and transmission. Hadith scholars soon realised the imminent threat the fabrication of hadith posed and developed a rigorous science of hadith (‘ilm al-hadith) to protect the authenticity of the teachings of the Prophet (pbuh), and thereby shield the Muslims from falsehood. Each hadith was verified based on its narrators (rawi), the text of the hadith (matn), and the chain of narrators (sanad, pl. isnad). This isnad system is the great contribution of hadith scholars to Islam, through which the authenticity of hadiths was ascertained.
Hadith scholars use an indigenous science of criticism and praise (‘ilm al-jarh wa al-ta’dil) to scrutinise each link in the chain of narrators, thereby grading the overall authenticity of the hadith. This science of verification was developed to maintain the purity and sanctity of Islam, and protect the religion from the craftiness of liars. The objective of the science of ‘ilm al-jarh wa al-ta’dil was to assess the status and trustworthiness of a narrator in order to determine the authenticity of a hadith. The term al-jarh is used for accusation and denunciation, whereas al-ta’dil means to establish, enforce or attribute personal good. With reference to the validation of hadith, al-jarh focuses on the rejection of an individual, while al-ta’dil compliments the narrator.

A. Traditional Verification of Hadiths

1. Reliability of Narrators: the study of the transmitters of hadith (‘ilm ul rijal al-hadith) is another integral discipline in the science of hadith. It focuses on studying the biographies of each narrator in an isnad to determine their reliability and, hence, the validity of the hadith they transmitted. An narrator is accepted if they are Muslim, mature, possessed of a sound mind, and free from anything that could lead to open sinfulness. Moreover, for an hadith to be valid, the biographical accounts of the narrators should indicate that they were practicing Muslims and not involved in major forbidden acts. If a narrator was a known liar, they are classified as such (kadhaab) and all hadiths transmitted by them are classified as weak (da’if).

2. Authenticity of Narrations: The scholars of hadith have carefully studied the matn and isnad of the prophetic traditions in order to ensure their authenticity. To be considered authentic (sahih), an hadith’s chain of transmission must be continuous. If the chain is broken due to possible untruthfulness, the hadith is rejected. As for the text of a hadith, comparisons are made between the accounts of different students, between the statements of a single scholar over time, between oral and written documents, and between the hadith and a related verse of the Qur’an.

In his book Naqd al-manqul, Ibn Qayyim al-Jawziyyah presented guidelines for identifying false narrations. He argued that an hadith is probably fabricated if it clearly contradicts: (1) an established teaching in the Qur’an; (2) a widely established sunnah of the Prophet (pbuh) or any authenticated hadiths; (3) authenticated historical facts; or (4) the basics of Islamic law (shariah). Narrations with illogical or severe, aggravated or grievous, connotations are also considered false, and hence unacceptable.
along with hadiths that do not reflect the words of the Prophet (pbuh). 93, 94

Isnad criticism has priority over matn criticism; when the isnad of a hadith is rejected, evaluation of the matn is not required. However, the reliability of the isnad does not guarantee the authenticity of the narration. 95

3. Validation through the Act of Narrating: The method of narrating an hadith involves either directly hearing it from a teacher or hearing it from another transmitter. For both methods, a narration is considered weak if there is any uncertainty caused by concealing the identity of a teacher in an isnad. A person reporting from a teacher whom they have not heard from or seen, using well-known names while concealing the real name of a teacher, or omitting the name of a weak authority, renders an hadith weak. 96

B. Probability Theory Based Hadith Verification

Taking heed from the attempts of Islamic scholars to develop disciplines to differentiate between authentic and fake reports, similar efforts can be used to address the problems that have arisen due to fake news and misinformation on the web. The entire process of news dissemination can be analysed by assessing the reliability of each transmitter individually. The criticism of narrators can be applied to organisations and websites that share information; principles of authenticating narrations can be applicable to the information being shared, allowing the content author and act of narrating information to be critically examined. With no regulations being imposed on the content shared on the web, an attempt to critically evaluate the reliability of websites and organisations, or to maintain a directory of reliable and unreliable sources, is difficult. Rejecting all content shared through unreliable sources could, however, reduce the amount of fake news being transmitted.

Probabilistic hadith transmitter criticism uses probability theory to compute the reliability of information transmitted by hadith transmitters by combining all existing transmission chains present in hadith collections using two processes: vertical and horizontal combination. 97 Each transmitter in a chain is assigned a baseline reliability score of 0.6, with 1 representing total reliability and 0 no reliability. Due to the extreme vetting process hadith transmitters have undergone, it is assumed that each transmitter is likely to transmit information truthfully and accurately. Transmitters who are known to have had unreliable memories (mukhtalit, sayyi al-hifz), or whose truthfulness has been put into question (majruh, ghayr hujja, da ‘if, laysa bi-shay’, matruk) are assigned lower reliability scores. 98
When judging hadith transmitters, a source of inconsistency and moral hazard flows from the sectarian leanings of some hadith critics. Certain hadith critics tended to judge transmitters more harshly than others, and would even consider them entirely unreliable, due to their perceived unorthodox beliefs and practices—with orthodoxy being defined by the hadith scholars. The important early hadith collector Malik b. Anas (d.795 CE), for example, urged his students to disregard hadiths transmitted by proselytisers of “religious innovation.” Since it was left to each scholars’ own judgment, to determine orthodoxy, this created the possibility of hadith critics judging a transmitter weak despite their widely accepted truthfulness and ability to transmit information accurately. Additionally, hadith critics had an aversion to the transmission of information through writing, despite the obvious fact that written information carries greater integrity than that stored by human memory. This led to hadith critics considering certain transmitters unreliable merely for relying on written hadith materials. Another source of inconsistency in the hadith literature is the fact that hadith critics used to judge legal hadith in a far stricter manner than non-legal hadith, due to their belief that legal hadith would influence the way Muslims practiced Islam, while non-legal hadiths would not.

These observations point to the fact that human beliefs, biases, and social group membership can strongly influence judgment of the reliability of information transmitters, thereby representing a potential source of error that could cause empirically correct information to be erroneously judged unreliable, or empirically doubtful information to be judged reliable. Probabilistic hadith criticism represents a step forward in removing some of these biases by providing a consistent set of criteria to be used on all hadiths and, by being collection-agnostic, avoiding the common bias of considering hadiths within the more famous collections to be of higher quality without due regard for the hadiths’ transmitters. The problem of judging the reliability of transmitters remains, however, and requires further empiricalisation and standardisation; probabilistic hadith criticism can contribute to this area of research by providing a consistently-judged corpus of hadith, comprising all major hadith collections, to be used as test cases for judging the reliability of transmitters.

Falling victim to fake news is more a symptom of ‘lazy’ thinking than of biased computation. Fact-checking systems can therefore serve an essential role, enabling information consumers to delegate the task of information verification to systems designed to fight fake news. Lessons learned from our research in hadith criticism can contribute to the issue of designing systems to counter fake news. Consistently applied criteria for judging the reliability of information transmitters is essential for preventing the introduction of human error and bias into the verification process. Humans employ various techniques to
discount information that clashes with their beliefs, such as biased assimilation, relative weighting of evidence, and minimisation of impact.\textsuperscript{103, 104} Any system designed to fight the spread of false and inaccurate information can fall victim to the human biases of its designers and operators. It is therefore necessary to design and operate such systems with the express goal of minimising the impact of these factors through consistent and transparent criteria and procedures. Besides the benefit of increasing the system’s reliability, such consistency and transparency can enhance the reputation of the system. Such a reputation can determine whether users will trust the judgments issued by it.

Another aspect of hadith criticism that could be beneficial to fighting fake news is the comparative process by which the reliability of little-known transmitters is determined. Such transmitters are judged by the concordance of their transmitted information with the narrations of other transmitters. Persons known to mostly transmit narrations containing strange and unusual assertions (\textit{al-manakir}) are considered less reputable, while those whose narrations are largely in accord with the rest of the hadith literature are considered more trustworthy. A person whose career largely involves accurately transmitting information corroborated by other transmitters can be trusted, even in those few cases where the information they transmit is rare and uncorroborated, with their personal reputation serving to enhance the status of their narration.

Fake news-fighting systems, in the absence of knowledge about the reputation of a given information source, can monitor the track record of the source to determine the rate at which it has published unusual or uncorroborated assertions. Even in the complete absence of reputational information, an automated system can judge the reliability of an information source, such as a little-known news website, by determining the concordance rate of its contents with information found in reputable sources. A largely unknown news source can be judged trustworthy when it is determined that an overwhelming majority of its news stories are corroborated by other sources. An important benefit of such a system is its ability to judge a news piece trustworthy despite: (a) a lack of external knowledge about the reputation of the source, and (b) a lack of corroborating news pieces from other sources. In such a case the automatically computed concordance rate of previously published news pieces at the source with pieces found at other well-known sources serves as evidence enhancing the probability of the accuracy of the novel and uncorroborated news piece from the little-known source. Conversely, news sources suffering from a low concordance rate can be downgraded in the absence of further information about the source’s trustworthiness and reliability.

Probabilistic hadith criticism can also serve as an educational tool for teaching important aspects of critical thinking when it comes to judging the
reliability of information; inherently probabilistic nature of the reliability of transmitted information; influence of human error when judging the reliability of information; and importance of applying explicit and consistent criteria and procedures when judging information.

Conclusions

During the present era, technological breakthroughs are happening at an exponential rate, with fancy gadgets and applications being rapidly introduced. However, these advancements have raised serious ethical questions. The growth of social media platforms and the World Wide Web have led to the spread of misinformation and disinformation through fake news, along with deep-fake videos that produce highly realistic synthetic content. The potential of these technologies to cause harm to individuals, societies and governments is extremely alarming. In this paper, we highlighted instances where such technology has indeed posed a threat. Deceitful content can result in identity theft and extortion, sexual victimisation, fake criminal accusations and, more importantly, destruction of trust. When falsehood becomes too prevalent, it ignites skepticism, making it hard to believe real and authentic information. Considering the threat that fake news and synthetic content poses for humanity, we have consulted the Islamic guidance on falsehood—gleaned from the teachings of the Qur’an, sunnah of the Prophet (pbuh), and practices of the early Islamic scholars—and presented guidelines therefrom. We looked at deceptive content from multiple dimensions, evaluating each dimension in the light of Islamic principles. In addition, to reduce the impact of technology, we have also provided guidelines and alternatives for Muslims to apply when combating fake news, misinformation, and machine learning generated fakes.

Recommendations

• Muslims need to realise the negative societal, communal, and political impact of fake news and misinformation, and understand the Islamic stance on spreading falsehood. Muslims must be able to actively participate in initiatives that work towards fighting the spread of false content.

• In order to ensure false information is not consumed, individuals should practice assessing the credibility of sources, whether on social media or the World Wide Web, when acquiring information.
• Taking insights from the Muslim scholars of the past, present-day Muslims should critically and unbiasedly evaluate information. One should not shy away from demanding proof or investigating further, especially when it comes to synthetic content, lest we become the victims of manipulation.

• Refrain from spreading news whose validity is doubtful and ensure that content is shared with sufficient information about its producer and publisher.

• Journalistic organisations should establish platforms that adhere to the Islamic framework of news dissemination, while training journalists to ensure the practice of responsible journalism.

• By seeking guidance from the process of verifying hadith through probability theory, individuals can apply similar strategies to combat fake news.

Notes

* Talat Zubair is from the Institute of Business Administration, Karachi, Pakistan.

** Amana Raquib is an assistant professor at the Institute of Business Administration, Karachi, Pakistan.

*** Junaid Qadir is from Information Technology University (ITU), Lahore, Pakistan.

The authors would like to thank Ikram Hawramani of the Hawramani Institute for his helpful input, particularly in relation to the topic of probabilistic news verification.

3. The Examiner No. XIV (Thursday, November 9, 1710)
14. Ibid.
25. Andrew Guess, Jonathan Nagler and Joshua Tucker, ‘Less Than You Think:


42. Ahmed, ‘Fake News’, 61-84


44. Parrott, ‘Finding Truth’.


47. Sahih Muslim, Book 18, Hadith 1511.


49. Ibid.
53. Qur’an 3:104
55. Sunan an-Nasa’i, Vol. 6, Book 47, Hadith 5011
61. Qur’an 24:15
65. Ibid.
69. Qur’an 2:42.
71. Qur’an 24:19.
72. Qur’an 24:12.
73. Qur’an 49:12.
74. IEEE Global et al., ‘Ethically Aligned’.
77. Ibid.
79. Shompa et al., ‘Guiding Social Media’
85. Ibid.
86. Ibid.
90. Ibid.
93. Ibid.
94. Ismail, ‘The Matan’.
95. Ibid.
98. Ibid.