Blockchain for Social Good: Combating Misinformation on the Web with AI and Blockchain

Oshani Seneviratne
senevo@rpi.edu
Rensselaer Polytechnic Institute
Troy, New York, USA

ABSTRACT

The spread of deceptive or misleading information, commonly referred to as misinformation, poses a social, economic, and political threat. Such deceptive information spreads quickly and inexpensively. For example, with the hype around blockchain technologies, misinformation on "get rich quick" scams on the Web is rampant, as evidenced by sophisticated Twitter hacks of celebrities and many social media posts that bait unsuspecting users to visit phishing websites. Unfortunately, AI technologies have contributed to the growing pains of misinformation on the Web, with the advances in technologies such as generative adversarial deep learning techniques that can generate "deep fakes" for nefarious purposes. At the same time, researchers are working on a different set of AI technologies to combat misinformation, akin to "fighting fire with fire." As there is no clear way to win the online "cat-and-mouse" game against fake news generators and spreaders of misinformation, we believe social media platforms could be fortified with blockchain and AI technologies to mitigate the extent of misinformation propagation in various communities worldwide. Tamper-proof blockchain techniques can provide irrefutable evidence of what content is authentic, guaranteeing how the information has evolved with provenance trails. Various AI models that could be used for detecting fake news can be served on a blockchain for the effective and transparent utility of the model. Such a synergistic combination of AI and blockchain is a burgeoning area of research. This paper outlines a proposal for combining blockchain and AI techniques for handling misinformation on the Web and highlights some of the early ongoing work in this space.

KEYWORDS

misinformation, web, blockchain, AI, graph analytics

ACM Reference Format:
Oshani Seneviratne. 2022. Blockchain for Social Good: Combating Misinformation on the Web with AI and Blockchain. In *14th ACM Web Science Conference 2022 (WebSci '22), June 26–29, 2022, Barcelona, Spain.* ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3501247.3539016

WebSci '22, June 26–29, 2022, Barcelona, Spain
© 2022 Association for Computing Machinery.
ACM ISBN 978-1-4503-9191-7/22/06…$15.00
https://doi.org/10.1145/3501247.3539016

1 INTRODUCTION

The World Wide Web (WWW) was invented in 1989 by Tim Berners-Lee as a mechanism to connect the world's information [7]. Blockchain was invented in 2008 by Satoshi Nakamoto as a basis for a peer-to-peer electronic cash system [43]. Although these inventions are separated by 20 years, and the specific technology mechanisms appear to be very different on the surface, the Web and blockchain are both examples of decentralized systems, and they have demonstrated to grow scale-free [4, 52].

The Web has become very pervasive in our day-to-day lives. Similar to how the Web grew to what it is now from a simple matter of linking documents, blockchain technology is currently being explored in many different application areas beyond cryptocurrencies. At the same time, people are using both platforms to collectively collaborate on various tasks that have never before been possible. Web-based crowd-sourcing mechanisms, such as Amazon Mechanical Turk[1], have made it possible for people who have not met in the physical world to collaborate on a specific task and be compensated for their efforts [21]. At the same time, blockchain-based systems, through their permissionless innovation and scalable incentive mechanisms, have already resulted in several crowd-sourcing platform solutions [38, 39, 64, 66].

The goal of this paper is to propose how blockchain, coupled with AI and crowd-sourcing techniques, can be utilized for social good in combating misinformation on the Web. We first provide the background for this research and state of the art in utilizing blockchain to combat misinformation. We then introduce a specific misinformation topic–cryptocurrency scams–that is gaining in popularity due to the speculative nature of cryptocurrencies (Section 3) and outline a generic solution for such problems (Section 4) and a specific solution called the *Integrative Blockchain Provenance Analyzer* (Section 5) that leverages graph analytics on blockchain transactions. In our generic solution, we propose how to leverage a blockchain-based deployment option for AI models designed for misinformation detection. We outline how this solution could be further fortified with crowd-sourced techniques for verifying information in a trustworthy and transparent manner. Finally, we discuss the limitations of the proposed solutions and provide a brief outline of how to address them (Section 6).

2 BACKGROUND

This section provides the background for this research highlighting the rise of misinformation on the Web, followed by failings of prominent social media companies, and thus the need for responsible AI, which would be augmented with human input for combating misinformation on the Web.

[1] https://www.mturk.com

2.1 The Rise of Misinformation

The pervasiveness of the Web has given rise to the use of social media over the past couple of decades with the advent of websites such as Twitter, Facebook, and YouTube. Two-thirds of Americans supposedly get their news via web-based social media websites [28]. However, social media is a double-edged sword. It has been helpful for quickly disseminating (mostly accurate) information about concerns like COVID-19. However, online users can select the news they want by curating their list of accounts to follow, subjecting themselves to "filter bubbles" [47]. Users who post and share content online do not face requirements to adhere to journalistic standards in traditional news media; thus, we see a prevalence of misinformation on social media. This spread of deceptive or misleading information poses a social, economic, and political threat. Misinformation spreads quickly and inexpensively. This spread is particularly effective if the content resonates with the preconceptions and biases of social groups or communities by creating echo chambers and filter bubbles by using mass-produced machine-generated and manipulated images, video, and audio. Recent research has examined factors associated with the origins of, exposure to, and credibility of misinformation [3, 20, 30, 48]. Research has also begun to address the diffusion of misinformation over online networks [27, 67]. It appears that misinformation spreads farther and faster than true information [62].

2.2 Failings of Large Social Media Companies

Several design decisions in popular social media websites have exacerbated the propagation of misinformation. Social media is increasingly full of noise. Thoughts are constantly shared, tweeted, and broadcasted. Everything is measured by "followers" and "likes." So it is hard to decipher what is accurate and what matters, and it is hard to cut through the noise. For example, in 2011, Google engineers noticed a problem caused by YouTube's focus on views, which was that it encouraged creators to use misleading tactics (like racy thumbnails) to dupe people into clicking the videos to play for them. Even if a viewer immediately stopped viewing the video, the click would increase the view count, boosting the video's recommendations [58]. This led to Google stopping ranking videos based on clicks but instead focusing on "watch time," or how long viewers stayed with a video, arguably a far better metric of genuine interest. However, this too was problematic, as the recommendations for such videos are also likely to be videos of misinformation. With respect to the political misinformation campaigns and radicalization, one of the Google employees has claimed to have said that "the time to fix this was yesterday" [58]. An unfortunate consequence of this inaction by social media websites and regulatory bodies is that the algorithms that govern popular social media platforms have an outsized influence on political discourse worldwide, contributing to polarization, unrest, and hate crimes. Divisive rhetoric distributed by Facebook has been linked to violence in Sri Lanka, Myanmar, and India [26]. Furthermore, the worldwide pandemic and a contentious US election whipped up a storm of automated misinformation in the recent past.

By 2015, with the proliferation of deep learning techniques, large social media companies would also introduce neural-net models to craft recommendations. The model would take user actions, i.e., whether a user had finished a video, say, or hit the "Like" button, and blend that with other information the algorithm had gleaned about the user, such as the search history, geographic region, gender, age, and a user's watch history. Then the model would predict which videos you would be most likely to watch, making the recommendations more personalized than ever. These misguided recommendations resulted in very negative behaviors of social media users. For example, teenage boys followed recommendations to far-right white supremacists and "Gamergate" conspiracies [14]. Anti-vaccine falsehoods became ripe during the COVID-19 pandemic [9]. In Brazil, a marginal lawmaker named Jair Bolsonaro rose from obscurity to prominence by posting YouTube videos that falsely claimed left-wing scholars were using "gay kits" to convert kids to homosexuality [8]. Therefore, it is no secret that recommendations by large social media websites like YouTube are a potent force for spreading misinformation.

Social media companies have started to take an active role in combating these issues. For example, YouTube played a game of social media "whack-a-mole" initially–a video that violated YouTube's rules would emerge and rapidly gain views, and then YouTube would take it down. However, it was unclear that recommendations were crucial for these sudden viral spikes, and taking videos down is only a temporary solution to a much deeper issue. YouTube also developed a classifier to identify so-called marginal content, i.e., videos that comply with its rules against hate speech but promote conspiracy theories, medical misinformation, and other fringe ideas [58]. After discovering hundreds of fake user profiles that included headshots generated by AI, Facebook cracked down on manipulated media it deemed misleading and banned deepfake videos outright [42]. The company continues to develop deep learning tools to detect hate speech, memes that promote bigotry, and misinformation about COVID-19. Facebook and Twitter shut down accounts they considered fronts for state-backed propaganda operations [55]. These companies have realized that we need incentive mechanisms that do not just reward numbers of views but shift incentives toward distributing factual information and rational perspective to the extent they can be determined fairly.

However, the tech companies' various algorithmic and policy fixes stop short of making changes that might seriously threaten their bottom line. For example, in June, the Wall Street Journal reported that some Facebook executives had squelched tools for policing extreme content [32]. Later, the company reversed algorithmic changes made during the election that boosted reputable news sources (perceptions that Facebook's effort was halfhearted even prompted some employees to resign [18]). Similarly, YouTube's algorithmic tweaks targeting misinformation have cut traffic to content creators who promote falsehoods. However, they also boosted traffic to larger entities, like Fox News, that often spread the same dubious information [46].

There is also the danger in any company taking on the arbiter of truth and social benefit, but that does not mean it should not moderate the content it delivers. Realizing the need for regulatory intervention in these matters, the US Congress grilled the companies [51], and so did the European Union [53]. A popular Netflix documentary called The Social Dilemma [2] excoriated large social

[2] https://www.thesocialdilemma.com

media companies, and public opinion polls [15] showed that social media companies had lost the trust of most Americans.

2.3 The Need for Responsible AI for Combating Misinformation

AI technologies can contribute to the growing pains of misinformation, with technologies that can be used to generate deep fakes quickly. For example, adversarial generative deep learning technologies have been demonstrated to make up unreal content (mostly videos) for nefarious purpose [63]. The biggest threat that deepfakes pose is that it empowers a culture of impunity wherein bad actors have been facing little to no consequences for the harm they cause to the victims of non-consensual intimate images. For example, one of the unfortunate consequences of deepfakes is the proliferation of websites that use deep learning algorithms to strip women's clothes in photos without their consent [12]. However, AI can also be used to combat misinformation and detection of deepfakes, as evidenced by a vast body of work where researchers focus on the "detection" of deepfakes using AI-assisted approaches to take on this problem [41, 45, 60].

As the world faces multiple crises, from COVID-19 to climate change to other pressing global and local societal problems, it is more important than ever for social media companies to stanch the flow of misinformation, as the proliferation of misinformation can have real irreversible problems. At the same time, the larger AI community has a responsibility to craft algorithms that support a just society, even as they promote lucrative businesses that utilize state-of-the-art AI algorithms.

2.4 State of the Art in Combating Misinformation on the Web

There are many related works to identify fake news and misinformation. As illustrated below, some of them utilize a blockchain-based system to register reputable news sources and utilize cryptographic techniques to determine the propagation of the information. Additionally, other AI techniques are used to rank and filter. There are also knowledge graph-based techniques such as the ClaimsKG [56], a knowledge graph of fact-checked claims, which facilitates structured queries about their truth values, authors, dates, journalistic reviews, and other kinds of metadata.

Some of the blockchain-based approaches for combating misinformation are as follows. Huckle and White [34] describe an application called the Provenator, which stores provenance metadata in the PREMIS format [10] on a blockchain, thus enabling content creators to prove the origins of their media resources. Because of the immutable property of blockchains, users of Provenator can trust the authenticity of the metadata about those resources. However, a limitation of Provenator is that it assumes the same system was used to document the resource in the first place, which is not a reasonable assumption as it is unclear how Provenator could gain wide-scale adoption. Yazdinejad et al. [65] demonstrated how to use blockchain to navigate deepfake AI. They present various use cases and methods for tackling deepfakes technology using blockchain capabilities and functionality. Qayyum et al. [49] created a blockchain approach based on contracts to combat fake news that includes a publisher management protocol, a news-oriented smart contract, and the establishment of a new blockchain. Hasan et al. [31] propose that the content should be authenticated and managed as a global file system over Ethereum smart contracts. Fraga et al. [25] provides a comprehensive overview by leveraging distributed ledger technologies to combat digital deception. Chen et al. [13] tackle the Internet of Fake Media Things (IoFMT) proactively, as opposed to the more common after-the-incident type of mechanisms, by leveraging a gamification component in the consensus mechanism design. Dwivedi et al. [19] introduces a naive blockchain and watermarking-based social media framework to control fake news propagation. Fan et al. [23] introduce a novel method for detecting Ponzi schemes in blockchain called Anti-leakage Smart Ponzi Schemes Detection (Al-SPSD) model. The New York Times is exploring such a blockchain-based approach through their News Provenance Project [50], which uses blockchain to track metadata such as sources and edits for news photos, providing readers with greater context and transparency into when and how content was created. Several works discuss the possible combination of AI and blockchain in tackling misinformation. Lacity [36] discusses the purposes, proliferation, susceptibility, and consequences of fake news and assesses the efficacy of new interventions that rely on emerging technologies such as blockchain and AI, with particular consideration for ethics. One of the main issues with these techniques mentioned above is that they are not robust in evolving tactics of misinformation-mongers. In our work, we believe that for a sustainable misinformation detection solution, we must have an end-to-end system that leverages state-of-the-art AI/ML techniques, data provenance techniques, and scalable incentives, both to the researchers contributing to the models and crowd-sourced verifiers of information. All these aspects are the cornerstone of the blockchain-based solutions we have proposed in Section 4.

3 USE CASE

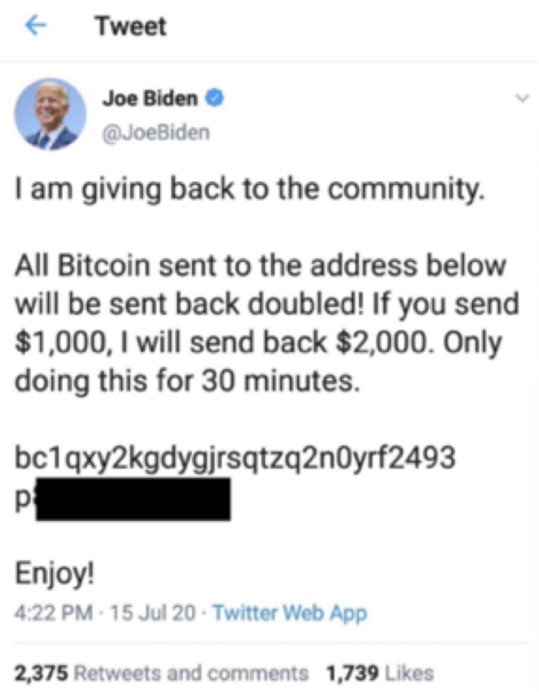
We illustrate a use case with respect to the increasing amounts of cryptocurrency misinformation we see on the Web.

3.1 State of Affairs for Cryptocurrencies

Most of media coverage on cryptocurrencies tend to be negative — focused on lost wallets [59], criminality [1], volatility [37] and energy use [33]. At one extreme, cryptocurrencies are banned, as in China, Qatar, Turkey, Russia, Iraq, and Egypt [22]. On the other, they are mainstream, as in El Salvador [11]. In most other countries, they are permitted, yet not seen as legal tender, since trading is unregulated and treated as just another asset. Most importantly, scams are rife [61]. The anonymity of cryptocurrency transactions appeals to criminals and bad actors. Cryptocurrencies are used to launder money, fund terrorism, and fuel corruption [57]. It has been estimated that up to half of bitcoin transactions support illegal activities [24]. Cryptocurrencies might be used to bypass financial sanctions, such as those currently (as of April 2022) imposed on Russia [40].

3.2 Cryptocurrency Scams

With the rapid growth of cryptocurrency applications, we see many advancements in malicious tactics with increasing levels of sophistication. On Bitcoin, for example, the most apparent scams are Ponzi schemes where the scammers ask victims to send bitcoin to an address, and they promise to double it, often posing as a celebrity on social media. For example, in July of 2020, some prominent accounts, like then US Democratic presidential candidate Joe Biden's account, were hacked using a social engineering attack on Twitter's administrative tools, as can be seen in Figure 1. Twitter reported that about 130 such high-profile accounts, with at least one million followers, got compromised in this attack [44]. The scammer asked unsuspecting users to send bitcoin to an address and mentioned that it would be doubled and returned as a charitable gesture. Twitter took prompt action on this, but not before some damage was done. Within minutes from the initial tweets, more than 320 transactions had already taken place on one of the wallet addresses. US $100,000 worth of bitcoin had been deposited in one account before Twitter removed the scam messages. This attack demonstrates that there is a lot of misinformation and scams like this on the Web now.



← Tweet

Joe Biden
@JoeBiden

I am giving back to the community.

All Bitcoin sent to the address below will be sent back doubled! If you send $1,000, I will send back $2,000. Only doing this for 30 minutes.

bc1qxy2kgdygjrsqtzq2n0yrf2493
p[redacted]

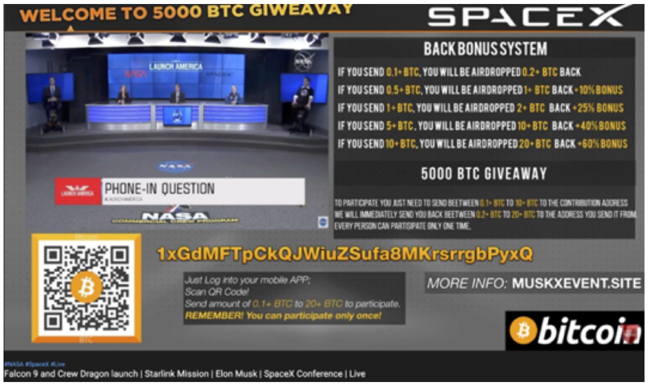
Enjoy!

4:22 PM · 15 Jul 20 · Twitter Web App

2,375 Retweets and comments 1,739 Likes

Figure 1: Twitter Social Engineering Attack in July 2020

Similarly, on YouTube, we see an increasing trend of people getting into cryptocurrency investing, going down a rabbit hole, and stumbling upon cryptocurrency scams. As an example, Figure 2 shows a "live event" on YouTube where scammers target some events like the SpaceX rocket launches, where they say that if you send x BTC, they will send an amount more than that. Even though it is evident that this "Back Bonus System" in Figure 2 appears like a Ponzi scheme, a lot of unsuspecting users have gotten caught in such a scam events, which is exacerbated because of the crypto-related social media activity by many tech celebrity investors like Elon Musk[3]. Of course, such tech celebrities are not behind these scams, but these are social engineering attacks at their best, where scammers ride on the hype around crypto and the activities of the tech celebrities.



WELCOME TO 5000 BTC GIWEAVAY

SPACEX

BACK BONUS SYSTEM

IF YOU SEND 0.1+ BTC, YOU WILL BE AIRDROPPED 0.2+ BTC BACK
IF YOU SEND 0.5+ BTC, YOU WILL BE AIRDROPPED 1+ BTC BACK +10% BONUS
IF YOU SEND 1+ BTC, YOU WILL BE AIRDROPPED 2+ BTC BACK +25% BONUS
IF YOU SEND 5+ BTC, YOU WILL BE AIRDROPPED 10+ BTC BACK +40% BONUS
IF YOU SEND 10+ BTC, YOU WILL BE AIRDROPPED 20+ BTC BACK +60% BONUS

5000 BTC GIVEAWAY

TO PARTICIPATE YOU JUST NEED TO SEND BEETWEEN 0.1+ BTC TO 10+ BTC TO THE CONTRIBUTION ADDRESS
WE WILL IMMEDIATELY SEND YOU BACK BEETWEEN 0.2+ BTC TO 20+ BTC TO THE ADDRESS YOU SEND IT FROM.
EVERY PERSON CAN PARTISIPATE ONLY ONE TIME.

1xGdMFTpCkQJWiuZSufa8MKrsrrgbPyxQ

Just Log into your mobile APP;
Scan QR Code!
Send amount of 0.1+ BTC to 20+ BTC to participate.
REMEMBER! You can participate only once!

MORE INFO: MUSKXEVENT.SITE

#NASA #SpaceX #Live
Falcon 9 and Crew Dragon launch | Starlink Mission | Elon Musk | SpaceX Conference | Live

Figure 2: YouTube Cryptocurrency Scams

4 AI-ASSISTED PROVENANCE CAPTURE AND VERIFICATION ON BLOCKCHAIN

Expanding the scope of identifying misinformation from cryptocurrencies to many other discussion topics on the Web, we propose a secure, efficient, and verifiable cryptographic information dissemination framework that utilizes blockchain and AI technologies. The proposed system architecture is depicted in Figure 3. As shown in the diagram, our proposed solution leverages blockchain and AI along with some human input, which is incentivized. We believe adding the human dimension to be paramount, as AI alone should not be left as the final arbiter of misinformation.

The state-of-the-art secure decentralized techniques such as blockchain do not depend on trusted third parties and enable large-scale secure and auditable computation. The utilization of blockchain technologies could mitigate misinformation as it provides one single source of truth. Leveraging this core principle, we propose developing efficient blockchain architectures to host AI and algorithms sourced from the research community in the form of a trusted and transparent information service. To establish an information service for fake news detection, we are developing a novel framework on blockchain for auditability via publicly-verifiable computation and design incentive mechanisms to elicit truthful information and labels about authenticity and bias in the news and data using a crowd-sourced mechanism. Since the transactions are recorded on the blockchain in an immutable and transparent manner, the transactions can be used to explain the AI models.

As can be seen in Figure 3, first, we utilize web-based content crawlers to discover new and related content. Related content may or may not be misinformation, and we can utilize state-of-the-art AI techniques such as graph analytics [35], semantic similarity methodologies [17] to search for the level of misinformation in such related

[3] https://twitter.com/elonmusk

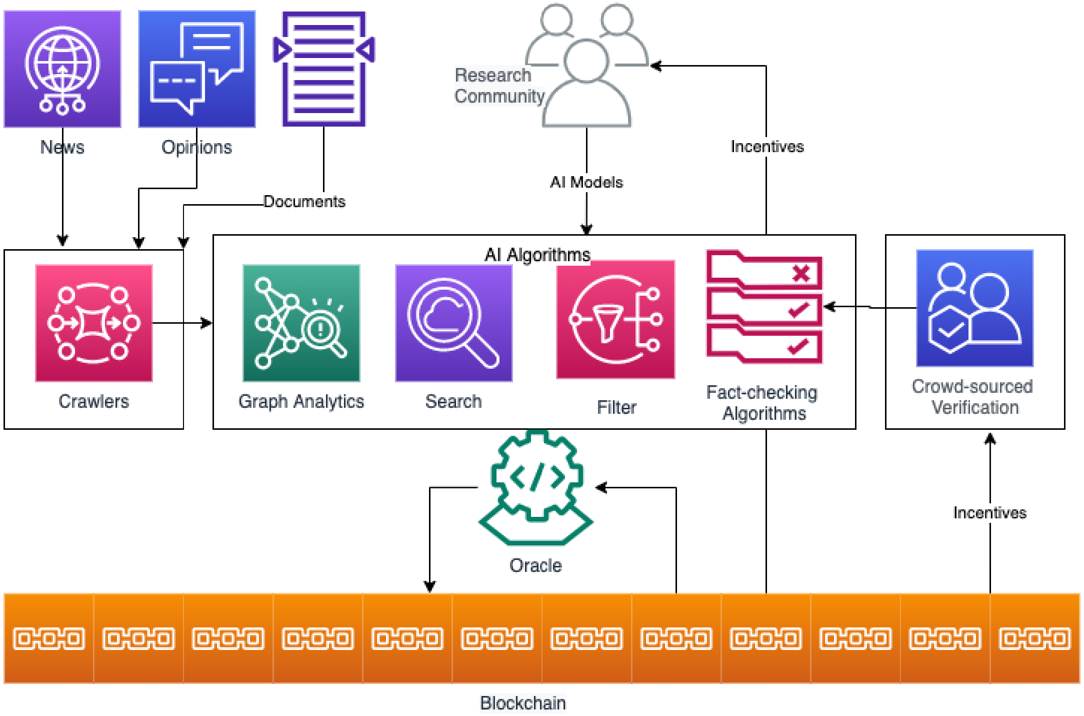


Figure 3: AI-Assisted Blockchain Solution for Identifying Misinformation

content. We can also use data-mining based misinformation detection models [54], supervised ensemble models for automatically filtering fake news [2], and fact-checking algorithms [29] for misinformation detection. We note here that to create an AI classifier that can recognize misinformation, we have to train the AI with many thousands of examples. For example, to detect misinformation on YouTube, the AI model would have to ask hundreds of ordinary humans to decide what looks like objectionable content. However, since news data on the Web is massive and the themes and contents of news are changing dynamically, it is too costly, slow, unrealistic, and non-scalable to rely on human annotation and active supervision to meet the real-time analysis requirements. Therefore, a blockchain-based decentralized incentive scheme could be the driving force to rally humans that need only look at objectionable content filtered by robust AI misinformation detection models.

Furthermore, as misinformation detection is a burgeoning research area, there are hundreds of models being developed by researchers every year, as evident by the research papers published in this space alone[4]. Therefore, we can ask the research community to contribute misinformation detection models to our proposed blockchain-based solution, incentivizing the researchers for their proposed contributions. A blockchain-based decentralized application would keep track of the researchers' model contributions and metadata, such as stated accuracies and third-party verification of their results. If a new model achieves some improvement, blockchain can have its way of attributing the incentive tokens to the researchers who have contributed to this new model. These model update processes would be implemented as blockchain oracles [6]. Oracles act as special-purpose smart contracts because running the misinformation detection models directly on the blockchain may not be feasible due to its computational requirements. Instead, a trusted third party implements the "information service." The results of the computation are relayed to the blockchain using the oracle.

As depicted in Figure 3, crawlers identify new content from the Web from reputable news websites, opinion sites, and documents published by various entities. Then, smart contracts work in tandem with oracles, keep references to the crawled content and others linking to them in an immutable data store, such as the InterPlanetary File System (IPFS) [5]. Unique content identifiers and any re-shares of the content would also be appended into the immutable store. Any application of the AI/ML methods on the newly discovered content and their outcomes, along with the model parameters, would be stored in IPFS containers as well. Creating provenance records this way, utilizing the blockchain properties, enhances the trustworthiness of misinformation detection and reporting.

In addition to models that allow for online verification of information diffused on social or news media, we need a mechanism to provide missing contexts such as author, date, related events, or references of a particular claim shared on social media. For that purpose, we could leverage a global community of professionals and

[4] A quick search on Google scholar for the search terms "misinformation+social+media" yields 162,000 results. See https://scholar.google.com/scholar?q=misinformation+social+media.

digital citizens to augment the automated AI misinformation detection processes in discovering, labeling, and justifying problematic content. With the assistance of advanced AI components making suggestions for reviewing questionable content and automatically linking versions of the same content, this global community would be labeling any problematic content.

Finally, the blockchain features such as digital signatures and immutable records enable anyone to request the provenance trail of a particular piece of information on the Web to ascertain that it is not, in fact, misinformation. All the features mentioned above could be implemented in a decentralized application implemented using smart contract programming languages such as Solidity and deployed on the Ethereum blockchain or a sidechain [16].

5 INTEGRATIVE BLOCKCHAIN PROVENANCE ANALYZER

The Integrative Blockchain Provenance Analyzer (IBPA) [35] is a solution we developed to flag scam addresses before they offload their wallets at an exchange by analyzing the transaction graph. This addresses a specific type of misinformation on the Web introduced in Section 3, that can be easily verified using lightweight AI techniques applied to blockchain transactions.

The IBPA works by analyzing the transaction graph, similar to the one shown in Figure 4, and it looks for some key patterns in the incoming and outgoing transactions. First off, scammers are known to add funds to these scam addresses before starting the scam to make "the fund-raising event" or the "once in a lifetime investment opportunity" seem legitimate. Second, possibly right after the scam, the cryptocurrency is usually subsequently transferred through multiple accounts as a means to obscure their identity. Then finally, there will be transactions to known exchanges to offload their wallets. Therefore, such a solution can be used stop such malicious activity with the coordination of legitimate exchanges, because when given the proof of the scam, these exchanges can easily blacklist the scam wallet addresses, and prevent the scammers from either transferring the tokens out of the exchange or cashing the tokens.

IBPA enables to derive meaning from raw ledger data with the help of graph analytic techniques and leverages patterns in the graph to identify suspicious addresses [35]. It effectively calculates a reasonable suspicion flag for a selected address, and it also creates an easy-to-query graph representation of the surrounding nodes of interest. IBPA extracts raw transaction data from a given blockchain network. Then, it transforms that data, filtering the metrics/keys that matter. These metrics include the final input and output addresses and amounts in the transaction. Next, it loads that data into a graph database for further analysis. Given a suspected scam address (e.g., the node marked in red in Figure 5), there will be a bunch of incoming transactions before and after the scam event. Out of those transactions, IBPA analyzes how many of the earlier transactions (like t_{-1} and t_{-2} in Figure 5) are also from suspected scam addresses. As mentioned above, these transactions may be correlated to scammers injecting money into the scam addresses to make them look legitimate. When the scam commences there could be several unsuspecting users who get caught, like these transactions at t_1, t_2, t_3, and t_4 in Figure 5. IBPA also looks at the scam address's "fan-out" or outgoing transactions. If several transactions happen within a short period of the scam, and some are to an exchange, then that is also a reasonable estimate for this address to be that of a scammer.

Figure 4: Example Cryptocurrency Transaction Graph

This implementation is one of the many implementations possible with the application of AI models. While we have not used human workers in this particular solution, we can assume the initial step of feeding the suspected scam addressed to be coming from a crowd-sourced solution.

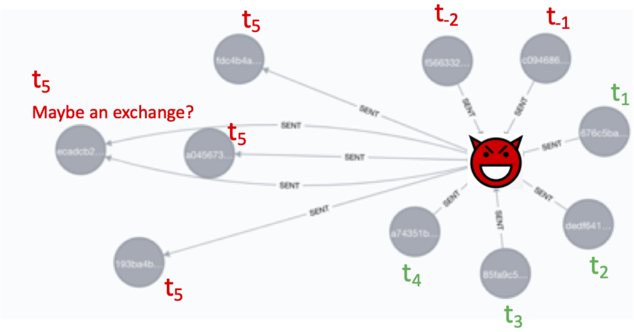


Figure 5: Analyzing Transaction Graph to Detect a Cryptocurrency Scam

6 CONCLUSION

The Web has evolved into a ubiquitous platform for communication, facilitating democratic and bottom-up knowledge construction. The pervasiveness of social media, the emergence of sophisticated digital manipulation tools, and the still outdated, under-resourced, and limited legislative and judicial systems widen the accountability gap between victims of digital harm and their pathway to digital justice. There is a growth of online discourse and data shared on social networks or online news outlets, including claims on controversial topics, their associated stances, sources, and related events. Therefore, there is increasing research interest in developing computational models for claim detection and verification, misinformation detection, and bias detection or propagation. However, in a world where conspiracies are recommended everywhere, even the best AI cannot fix what is broken. We need a solution augmented with strong epistemic security based on robust AI and verifiable provenance and, to some extent, human verification to fix this problem.

An AI-assisted blockchain architecture was introduced in Section 4. This architecture includes a novel privacy-preserving architecture between blockchain and hosts of data and AI/ML models and builds reputation and incentive mechanisms for crowd workers to achieve a more trustworthy information ecology to combat misinformation on the Web. One of the most challenging aspects of promoting accurate information in the current media landscape is that creators and distributors are strongly incentivized to drive clicks at all costs, where the clicks often come from sensationalized content. Therefore, the incentive mechanism must be flipped so that the verification of authentic sources has a higher reward. Technological solutions such as those outlined in this paper aid in this process by providing greater transparency into the lifecycle of the news-worthy content and bridging the accountability gap for victims of digital injustices. When the incentives align, the individuals promote factual information instead of misinformation.

We then outlined how a solution we have developed has been applied in detecting cryptocurrency scams, which is in line with the proposed generalized solution for handling misinformation campaigns that we see on the Web. Even though IBPA, explained in Section 5, is proven to successfully detect scam addresses, solving the ultimate cryptocurrency misinformation problem is next to impossible. It is a tall order because cash has been used to carry out all the crimes for which cryptocurrency is blamed. However, unlike cash, cryptocurrencies could monitor users and introduce "know-your-customer" forms of authentication. However, that would go against the principles of privacy and ownership, lowering the uptake of cryptocurrencies. Examining such trade-offs is an important future research question.

In conclusion, using blockchain and AI for social good is just one piece of the puzzle. Blockchain has the potential to enable greater accuracy and transparency. However, it is not a silver bullet. For blockchain to effectively help in combating misinformation, cohesive partnerships within international communities, including governments, corporations, civil societies, and technical leaders committed to shaping the governance of digital content creation and consumption, are paramount.

REFERENCES

[1] David Adler. 2018. Silk Road: The Dark Side of Cryptocurrency. Retrieved April 7, 2022 from https://news.law.fordham.edu/jcfl/2018/02/21/silk-road-the-dark-side-of-cryptocurrency

[2] Muhammad Pervez Akhter, Jiangbin Zheng, Farkhanda Afzal, Hui Lin, Saleem Riaz, and Atif Mehmood. 2021. Supervised ensemble learning methods towards automatically filtering Urdu fake news within social media. *PeerJ Computer Science* 7 (2021), e425.

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

[4] Albert-László Barabási and Eric Bonabeau. 2003. Scale-free networks. *Scientific american* 288, 5 (2003), 60–69.

[5] Juan Benet. 2014. Ipfs-content addressed, versioned, p2p file system. *arXiv preprint arXiv:1407.3561* (2014).

[6] Abdeljalil Beniiche. 2020. A study of blockchain oracles. *arXiv preprint arXiv:2004.07140* (2020).

[7] Tim Berners-Lee, Robert Cailliau, Ari Luotonen, Henrik Frystyk Nielsen, and Arthur Secret. 1994. The world-wide web. *Commun. ACM* 37, 8 (1994), 76–82.

[8] Ed Bracho-Polanco. 2019. How Jair Bolsonaro used 'fake news' to win power. Retrieved April 17, 2022 from https://theconversation.com/how-jair-bolsonaro-used-fake-news-to-win-power-109343

[9] Talha Burki. 2020. The online anti-vaccine movement in the age of COVID-19. *The Lancet Digital Health* 2, 10 (2020), e504–e505.

[10] Priscilla Caplan. 2009. Understanding premis. Library of Congress Washington DC, USA.

[11] Carrie Khan. 2022. El Salvador's leader wants to go in even bigger on bitcoin. Retrieved April 7, 2022 from https://www.npr.org/2022/03/27/1086851329/el-salvadors-leader-wants-to-go-in-even-bigger-on-bitcoin

[12] Evin Cheikosman, Nadia Hewett, and Karin Gabriel. 2021. Blockchain can help combat the threat of deepfakes. Here's how. Retrieved April 17, 2022 from https://www.weforum.org/agenda/2021/10/how-blockchain-can-help-combat-threat-of-deepfakes/

[13] Qian Chen, Gautam Srivastava, Reza M Parizi, Moayad Aloqaily, and Ismaeel Al Ridhawi. 2020. An incentive-aware blockchain-based solution for internet of fake media things. *Information Processing & Management* 57, 6 (2020), 102370.

[14] Shira Chess and Adrienne Shaw. 2015. A conspiracy of fishes, or, how we learned to stop worrying about# GamerGate and embrace hegemonic masculinity. *Journal of Broadcasting & Electronic Media* 59, 1 (2015), 208–220.

[15] Elizabeth Culliford. 2020. *Social media companies distrusted by most Americans on content decisions: Poll.* Retrieved December 24, 2020 from https://www.reuters.com/article/us-usa-social-media-poll/social-media-companies-distrusted-by-most-americans-on-content-decisions-poll-idUSKBN23N12N

[16] Chris Dannen. 2017. *Introducing Ethereum and solidity.* Vol. 1. Springer.

[17] Ronald Denaux and Jose Manuel Gomez-Perez. 2020. Linked credibility reviews for explainable misinformation detection. In *International Semantic Web Conference.* Springer, 147–163.

[18] Alison Durkee. 2020. Facebook Engineer Resigns, Says Company On 'Wrong Side Of History' As Internal Dissent Grows. Article. Retrieved April 21, 2022 from https://www.forbes.com/sites/alisondurkee/2020/09/08/facebook-engineer-resigns-company-on-wrong-side-of-history-internal-employee-dissent-grows

[19] Ashutosh Dhar Dwivedi, Rajani Singh, Sakshi Dhall, Gautam Srivastava, and Saibal K Pal. 2020. Tracing the source of fake news using a scalable blockchain distributed network. In *2020 IEEE 17th International Conference on Mobile Ad Hoc and Sensor Systems (MASS).* IEEE, 38–43.

[20] Ullrich KH Ecker and Li Chang Ang. 2019. Political attitudes and the processing of misinformation corrections. *Political Psychology* 40, 2 (2019), 241–260.

[21] Enrique Estellés-Arolas and Fernando González-Ladrón-de Guevara. 2012. Towards an integrated crowdsourcing definition. *Journal of Information science* 38, 2 (2012), 189–200.

[22] ET Online. 2022. Countries which have banned or restricted use of cryptocurrency. Retrieved April 7, 2022 from https://economictimes.indiatimes.com/news/web-stories/countries-which-have-banned-or-restricted-use-of-cryptocurrency/slideshow/89153960.cms

[23] Shuhui Fan, Shaojing Fu, Haoran Xu, and Xiaochun Cheng. 2021. Al-SPSD: Anti-leakage smart Ponzi schemes detection in blockchain. *Information Processing & Management* 58, 4 (2021), 102587.

[24] Sean Foley, Jonathan R Karlsen, and Tālis J Putniņš. 2019. Sex, drugs, and bitcoin: How much illegal activity is financed through cryptocurrencies? *The Review of Financial Studies* 32, 5 (2019), 1798–1853.

[25] Paula Fraga-Lamas and Tiago M Fernandez-Carames. 2020. Fake news, disinformation, and deepfakes: Leveraging distributed ledger technologies and blockchain to combat digital deception and counterfeit reality. *IT Professional* 22, 2 (2020), 53–59.

[26] Sheera Frenkel. 2018. *Facebook to Remove Misinformation That Leads to Violence.* Retrieved December 24, 2020 from https://www.nytimes.com/2018/07/18/technology/facebook-to-remove-misinformation-that-leads-to-violence.html

[27] Adrien Friggeri, Lada Adamic, Dean Eckles, and Justin Cheng. 2014. Rumor cascades. In *proceedings of the international AAAI conference on web and social media*, Vol. 8. 101–110.

[28] Jeffrey Gottfried and Elisa Shearer. 2017. Americans' online news use is closing in on TV news use. *Pew Research Center* 7 (2017).

[29] D Graves. 2018. Understanding the promise and limits of automated fact-checking. (2018).

[30] Andrew Guess, Jonathan Nagler, and Joshua Tucker. 2019. Less than you think: Prevalence and predictors of fake news dissemination on Facebook. *Science advances* 5, 1 (2019), eaau4586.

[31] Haya R Hasan and Khaled Salah. 2019. Combating deepfake videos using blockchain and smart contracts. *Ieee Access* 7 (2019), 41596–41606.

[32] Jeff Horwitz and Deepa Seetharaman. 2020. *Facebook Executives Shut Down Efforts to Make the Site Less Divisive: The social-media giant internally studied how it polarizes users, then largely shelved the research.* Retrieved December 24, 2020 from https://www.wsj.com/articles/facebook-knows-it-encourages-division-top-executives-nixed-solutions-11590507499

[33] Jon Huang, Claire O'Neill, and Hiroko Tabuchi. 2021. Bitcoin Uses More Electricity Than Many Countries. How Is That Possible? Retrieved April 7, 2022 from https://www.nytimes.com/interactive/2021/09/03/climate/bitcoin-carbon-footprint-electricity.html

[34] Steve Huckle and Martin White. 2017. Fake news: A technological approach to proving the origins of content, using blockchains. *Big data* 5, 4 (2017), 356–371.

[35] Daniel Kazenoff, Oshani Seneviratne, and Deborah L McGuinness. 2020. Semantic Graph Analysis to Combat Cryptocurrency Misinformation on the Web.. In *ASLD@ ISWC*. 168–176.

[36] Mary C Lacity. 2021. Fake news, technology and ethics: Can AI and blockchains restore integrity? *Journal of Information Technology Teaching Cases* (2021), 2043886921999065.

[37] Nicole Lapin. 2021. Explaining Crypto's Volatility. Retrieved April 7, 2022 from https://www.forbes.com/sites/nicolelapin/2021/12/23/explaining-cryptos-volatility

[38] Ming Li, Jian Weng, Anjia Yang, Wei Lu, Yue Zhang, Lin Hou, Jia-Nan Liu, Yang Xiang, and Robert H Deng. 2018. CrowdBC: A blockchain-based decentralized framework for crowdsourcing. *IEEE Transactions on Parallel and Distributed Systems* 30, 6 (2018), 1251–1266.

[39] Yuan Lu, Qiang Tang, and Guiling Wang. 2018. Zebralancer: Private and anonymous crowdsourcing system atop open blockchain. In *2018 IEEE 38th International Conference on Distributed Computing Systems (ICDCS)*. IEEE, 853–865.

[40] MacKenzie Sigalos. 2022. Russia is considering selling its oil and gas for bitcoin as sanctions intensify from the West. Retrieved April 7, 2022 from https://www.cnbc.com/2022/03/24/russia-might-take-bitcoin-as-payment-for-oil-and-gas-as-sanctions-rise.html

[41] Yisroel Mirsky and Wenke Lee. 2021. The creation and detection of deepfakes: A survey. *ACM Computing Surveys (CSUR)* 54, 1 (2021), 1–41.

[42] Global Policy Management) Monika Bickert (Vice President. 2020. *Enforcing Against Manipulated Media.* Retrieved December 24, 2020 from https://about.fb.com/news/2020/01/enforcing-against-manipulated-media

[43] Satoshi Nakamoto. 2008. Bitcoin: A peer-to-peer electronic cash system. *Decentralized Business Review* (2008), 21260.

[44] New York State Department of Financial Services. 2020. Twitter Investigation Report: Report on Investigation of Twitter's July 15, 2020 Cybersecurity Incident and the Implications for Election Security. Article. Retrieved April 21, 2022 from https://www.dfs.ny.gov/Twitter_Report

[45] Thanh Thi Nguyen, Quoc Viet Hung Nguyen, Cuong M Nguyen, Dung Nguyen, Duc Thanh Nguyen, and Saeid Nahavandi. 2019. Deep learning for deepfakes creation and detection: A survey. *arXiv preprint arXiv:1909.11573* (2019).

[46] Jack Nicas. 2020. *YouTube Cut Down Misinformation. Then It Boosted Fox News.* Retrieved December 24, 2020 from https://www.nytimes.com/2020/11/03/technology/youtube-misinformation-fox-news.htm

[47] Eli Pariser. 2011. *The filter bubble: How the new personalized web is changing what we read and how we think.* Penguin.

[48] Gordon Pennycook and David G Rand. 2019. Lazy, not biased: Susceptibility to partisan fake news is better explained by lack of reasoning than by motivated reasoning. *Cognition* 188 (2019), 39–50.

[49] Adnan Qayyum, Junaid Qadir, Muhammad Umar Janjua, and Falak Sher. 2019. Using blockchain to rein in the new post-truth world and check the spread of fake news. *IT Professional* 21, 4 (2019), 16–24.

[50] Pooja Reddy. 2021. Could We Fight Misinformation With Blockchain Technology? Retrieved April 17, 2022 from https://www.nytimes.com/2020/07/06/insider/could-we-fight-misinformation-with-blockchain-technology.html

[51] Tony Romm, Rachel Lerman, Cat Zakrzewski, Heather Kelly, and Elizabeth Dwoskin. 2020. *Facebook, Google, Twitter CEOs clash with Congress in pre-election showdown.* Retrieved December 24, 2020 from https://www.washingtonpost.com/technology/2020/10/28/twitter-facebook-google-senate-hearing-live-updates

[52] Nejc Rožman, Marko Corn, Gašper Škulj, Janez Diaci, and Lovro Šubelj. 2021. Emergence of a scale-free network topology in a blockchain-based Shared Manufacturing. In *2021 Third International Conference on Blockchain Computing and Applications (BCCA)*. IEEE, 172–178.

[53] Adam Satariano and Milan Schreuer. 2018. Facebook's Mark Zuckerberg Gets an Earful From the E.U. Retrieved April 7, 2022 from https://www.nytimes.com/2018/05/22/technology/facebook-eu-parliament-mark-zuckerberg.html

[54] Kai Shu, Amy Sliva, Suhang Wang, Jiliang Tang, and Huan Liu. 2017. Fake news detection on social media: A data mining perspective. *ACM SIGKDD explorations newsletter* 19, 1 (2017), 22–36.

[55] Jack Stubbs and Christopher Bing. 2020. *Facebook, Twitter dismantle global array of disinformation networks.* Retrieved December 24, 2020 from https://www.reuters.com/article/cyber-disinformation-facebook-twitter/facebook-twitter-dismantle-global-array-of-disinformation-networks-idINKBN26T2XF

[56] Andon Tchechmedjiev, Pavlos Fafalios, Katarina Boland, Malo Gasquet, Matthäus Zloch, Benjamin Zapilko, Stefan Dietze, and Konstantin Todorov. 2019. ClaimsKG: a knowledge graph of fact-checked claims. In *International Semantic Web Conference*. Springer, 309–324.

[57] Fabian Teichmann. 2020. Recent trends in money laundering. *Crime, Law and Social Change* 73, 2 (2020), 237–247.

[58] Clive Thompson. 2020. *YouTube's Plot to Silence Conspiracy Theories: From flat-earthers to QAnon to Covid quackery, the video giant is awash in misinformation. Can AI keep the lunatic fringe from going viral?* Retrieved December 24, 2020 from https://www.wired.com/story/youtube-algorithm-silence-conspiracy-theories

[59] Joe Tidy. 2022. Hackers helped me find my lost Bitcoin fortune. Retrieved April 7, 2022 from https://www.bbc.com/news/technology-60318946

[60] Ruben Tolosana, Ruben Vera-Rodriguez, Julian Fierrez, Aythami Morales, and Javier Ortega-Garcia. 2020. Deepfakes and beyond: A survey of face manipulation and fake detection. *Information Fusion* 64 (2020), 131–148.

[61] Andrew Urquhart and Brian Lucey. 2022. Crypto and digital currencies—nine research priorities.

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

[63] Mika Westerlund. 2019. The emergence of deepfake technology: A review. *Technology Innovation Management Review* 9, 11 (2019).

[64] Xiaolong Xu, Qingxiang Liu, Xuyun Zhang, Jie Zhang, Lianyong Qi, and Wanchun Dou. 2019. A blockchain-powered crowdsourcing method with privacy preservation in mobile environment. *IEEE Transactions on Computational Social Systems* 6, 6 (2019), 1407–1419.

[65] Abbas Yazdinejad, Reza M Parizi, Gautam Srivastava, and Ali Dehghantanha. 2020. Making sense of blockchain for ai deepfakes technology. In *2020 IEEE Globecom Workshops (GC Wkshps*. IEEE, 1–6.

[66] Saide Zhu, Zhipeng Cai, Huafu Hu, Yingshu Li, and Wei Li. 2019. zkCrowd: a hybrid blockchain-based crowdsourcing platform. *IEEE Transactions on Industrial Informatics* 16, 6 (2019), 4196–4205.

[67] Arkaitz Zubiaga, Maria Liakata, Rob Procter, Geraldine Wong Sak Hoi, and Peter Tolmie. 2016. Analysing how people orient to and spread rumours in social media by looking at conversational threads. *PloS one* 11, 3 (2016), e0150989.