A prototype framework for assessing information provenance in decentralised social media: The EUNOMIA Concept

Lazaros Toumanidis[1][0000−1111−2222−3333], Ryan Heartfield[2][0000−0002−3708−1540], Panagiotis Kasnesis[1][0000−0003−3607−8187], George Loukas[2][0000−0003−3559−5182], and Charalampos Patrikakis[1][0000−0003−1921−4466]

[1] University of West Attica, Aegaleo, Attiki, GR
{laztoum,pkasnesis,bpatr}@uniwa.gr
https://www.uniwa.gr
[2] University of Greenwich, London, London, GB
{R.Heartfield,G.Loukas}@greenwich.ac.uk
https://www.gre.ac.uk

Abstract. Users of traditional centralised social media networks have limited knowledge about the original source of information and even less about its trustworthiness and how this information has spread and been modified. Existing media verification tools include websites or browser add-ons that are closed-source or centralised, or they do not include user involvement in the information verification process. In this paper, we introduce EUNOMIA, an open source, decentralised framework that aims at providing information about social media content and context in an intermediary-free approach and in a way that assists users in deriving their own conclusions regarding a social media post's trustworthiness. We present its components, how they interact with each other and how user contribution is key to its concept.

Keywords: decentralised social media · trustworthiness · fake news · human-as-a-trust-sensor

1 Introduction

Traditional Social Media has rapidly become a dominant, direct and highly effective form of news generation and sharing at a global scale, in a manner that influences, enhances, but also challenges and often antagonizes traditional media corporations. However, paradoxically, it has led to the further accumulation of power to a relatively short list of central intermediaries, such as Facebook, Twitter, Instagram and other large companies, whose practices are seen as increasingly invasive of users' privacy. Many users get the bulk of their daily news from social media. As news passes to users, the news passes through the hands of actors whose credibility and goals are unknown. Even less is known about the credibility and quality of the information cascades they trigger.

Ironically, "the most effective forms of censorship today involve meddling with trust and attention, not muzzling speech itself" [**?**], and it is already evident that deliberate misinformation, as exemplified by fake news, is not being tackled effectively by large intermediaries. Relying on large-scale detection of disinformation on third party professional curators (as in, OpenSources, Snopes.com, and politifact.com) would effectively introduce a new type of intermediary, while relying solely on static machine learning (as in, Truthnest and Fakebox), is unsuitable for the dynamic and extremely fast-paced information cascades of social media, especially as fake news adapts and spreads much faster than real news [**?**]. There is a need for an intermediary-free and democratic approach, where what is true and what is false is not left to third party experts or entirely on computer algorithms.

In addition, a series of new decentralised social media networks show that transition away from centralised big companies and other intermediaries to peer-to-peer (P2P) federated social networks is highly practical. For instance, the fast-growing Mastodon platform [1] has similar utility to Twitter, but is open-source and decentralised, and user communities are encouraged to set up their own server domains. Similarly, the Diaspora* platform[2] has a utility akin to Facebook, but is also open-source and decentralised, and the users can choose which server to connect to depending on each country's security and privacy policies, as well as setup their own. Intermediary-free solutions, such as these, are promising in addressing the concerns of users regarding ownership of data and visibility of processes but are not by themselves able to address disinformation. EUNOMIA adopts the same ethos with an intermediary-free decentralised solution, that will help users establish the source of information and the associated information cascade, and evaluate its trustworthiness themselves.

In this paper, we present the architecture and data workflows of the EUNOMIA (Us**e**r-oriented, sec**u**re, trustful & dece**n**tralised s**o**cial **m**ed**ia**) project [3]. EUNOMIA will employ a decentralised architecture and a digital companion providing the user with intuitive indications of the content and context of the sources for defining user-specific trust criteria, and determination of the nodes (users and posts) along an information cascade derived by a machine learning approach.

Towards tackling the challenge of information provenance and veracity in social media, there are three primary questions that EUNOMIA aims to help users answer:

- Which social media user is the original source of a piece of information?
- How has this information had spread and been modified in an information cascade?
- How likely is it to be trustworthy?

[1] https://joinmastodon.org

[2] https://diasporafoundation.org

[3] The EUNOMIA acronym is after the Greek goddess of good order and lawful conduct, associated with the internal stability of a state, including the enactment of good laws and the maintenance of civil order.

At the heart of the EUNOMIA concept is the objective to provide users with the tools and necessary information to help answer these questions, where users are empowered to make informed decisions about whether they can trust a piece of information presented on a social media platform. Importantly, EUNOMIA does not aim to assess trustworthiness itself or to make centralised judgements on what pieces of information are more trustworthy than others. Instead, EUNOMIA aims to enable users to conduct this task themselves, but more efficiently and effectively through the aggregation and formulation of existing information that can be collected or derived from social media platforms. Crucially, EUNOMIA is strictly opt-in, which means that only posts of users that have explicitly opted in are considered. This is rendering the scarcity of information a key challenge in information cascade development.

2 Related Work

In this section, we introduce existing tools (primarily applications, browser addons and websites) that have been developed with the objective of providing answers to at least one of the aforementioned three questions. Human-as-Trust-Sensor (HaTS) is one of the paradigms employed in EUNOMIA, which involves leveraging human sensing capabilities for evaluating trustworthiness. HaTS typically requires two facilities: 1) a way to collect and structure data for human analysis and assessment, and 2) a way for users to respond to information visualise to them, for example by voting. Here, a Reputation Mechanism specifically corresponds to an information cascade that provides data attributes for HaTS analysis to interpret and evaluate the reputation of its content. This evaluation primarily tries to answer the following questions: what is the source of an information cascade and importantly how it might have changed, and thus what metrics might be assessed to score the trustworthiness of an information cascade, its content and source. In the scope of reputation scoring, Web Of Trust is a tool that provides safety and security ratings for visited websites and search engine results. It is mostly based on user ratings, utilizing also some third-party trusted sources, such as phishing directories, and displays reputation icons next to search results, social media, and other popular sites to help users make informed decisions online. Moreover, Microsoft provides a plugin called NewsGuard for its Edge web browser, that uses "Green-Red" ratings to signal if a website is trying to get it right or instead has a hidden agenda or knowingly publishes falsehoods or propaganda, giving readers more context about their news online. To this end, NewsGuard relies on trained analysts, who are experienced journalists, research online news brands and check the validity of the produced news.

Media Bias/Fact Check is a large media bias resource that also provides Firefox and Chrome extensions which displays a color-coded icon denoting the bias of the page one is currently viewing, according to their analysis results. Existing platforms like Twitter are also working towards the use of crowdsourcing tools, as a part of the "battle against rampant abuse on its platform" [**?**]. Further-

more, WhatsApp provides a tip-line to which one can send forwards, rumors, and suspicious-sounding messages and have them verified [**?**].

Regarding text analysis and misinformation detection, TextBox and FakeBox, are two tools created by MachineBox[4] that process text, perform natural language processing, sentiment analysis, entity and keyword extraction and try to assess whether news articles are likely to be real news or not. Users can interact with these services using a web browser, being able to provide the content they want to be analysed and view the results. In table 1, we provide a high-level summary of the existing platforms and services which employ HaTS and Reputation Mechanism functionality to support the assessment of information trustworthiness. This summary evaluates HaTS state-of-the-art by assessing whether the capabilities implement source verification, information cascade, trustworthiness scoring, whether the system requires expert curation (or is crowdsourced) and if users are involved in the assessment of information trustworthiness.

Table 1. The current landscape of social media information verification market and where EUNOMIA sits

Tool name	Type	Verifies source	Verifies cascade	Scores trustw'ness	Users are involved
https://fullfact.org	Website	✓	✗	✗	✗
Truthnest[4]	Software	✓	✗	✓	✗
https://africacheck.org http://politifact.com/ http://factscan.ca/scoring/ http://chequeado.com	Website	✗	✗	✓	✗
aosfatos.org	Website	✓	✗	✓	✗
B.S Detector[6]	Browser add-on	✓	✗	✓	✗
OpenSources[7]	Database resource	✗	✗	✓	✗
http://areyoufakenews.com	Website	✓	✗	✓	✗
Check This by MetaCert[8]	Browser add-on	✓	✗	✓	✗
FiB[9]	Browser add-on	✓	✓	✓	✗
Official Media Bias Fact Check Icon[10]	Browser add-on	✓	✗	✓	✓
Hoaxy	Website	✓	✓	✗	✗
Fakebox	Software	✗	✗	✓	✗
EUNOMIA[11]	**P2P plat. & digital comp.**	✓	✓	✓	✓

Beyond the current state-of-the-art, EUNOMIA addresses a key gap in the landscape of the social media information verification market. Current offerings

[4] https://machinebox.io/
[5] http://www.truthnest.com/
[6] http://bsdetector.tech/
[7] http://opensources.co
[8] https://chrome.google.com/webstore/detail/felmjclcjadopolhjmlbemfekjaojfbn/
[9] https://devpost.com/software/fib
[10] https://chrome.google.com/webstore/detail/official-media-bias-fact/hdcpibgmmcnpjmmenengjgkkfohahegk
[11] https://eunomia.social

cover only a small subset of the requirements that EUNOMIA is addressing. The vast majority are websites, where expert curators analyse a variety of sources to establish the veracity of claims posted on social media, in a manner that is not scalable and generates one more intermediary (the group of expert curators employed). The only alternatives that are available today are browser add-on FiB and software Truthnest, that are not open-source, decentralised or able to involve the users in the information verification challenge.

3 The EUNOMIA Concept

EUNOMIA's concept is based on a circular data-driven "**user involvement →provenance identification →trustworthiness indicators →visualisation**" approach. Data will be collected and shared in accordance with a security and privacy framework, relating to data analytics and the HaTS component, and then will be utilised in a reasoning phase in near real-time evaluation of information trustworthiness supported by machine learning and a user-driven reputation mechanism. Reasoning may be performed locally (on the users' devices) or remotely (on EUNOMIA servers), in a P2P form and in line with the philosophy of Blockchain, avoiding entirely the dependence on third-party centralised cloud servers. A digital companion provides both visualisation of possible indicators of information trustworthiness and a facility for allowing the user to be involved, e.g., by voting or other means.

Decentralised Platform Architecture

EUNOMIA is planned to operate as a system of systems (Figure 1) integrated with existing open-source distributed social networks by extending their application and server software, effectively creating a EUNOMIA-enhanced Mastodon instance and a EUNOMIA-enhanced Diaspora* pod.

The EUNOMIA architecture consists of five core components running in a decentralized manner: the first (1) is a peer-to-peer network between EUNOMIA Services Nodes (ESN) which support and synchronise data and service components across EUNOMIA (and act as service nodes to users' Digital Companion clients). The ESN P2P network enforces (2) a security and privacy framework which supervises a strict GDPR-compliant opt-in post data extraction policy for EUNOMIA users. THe security and privacy framework directly enables (3) a Human-as-Trust-sensor mechanism with an integrated collection and extraction toolkit for analysis of social media post content and context, feeding output into a visualisation interface on (4) the EUNOMIA digital companion, where users are enabled to make their own trustworthiness assessment for post content and metrics. Finally, an immutable record of EUNOMIA user trustworthiness votes (relating to a users own trustworthiness scoring criteria) is stored and tracked in (5) a Blockchain infrastructure, supported and synchronised across the ESN P2P network.

Architecture Components

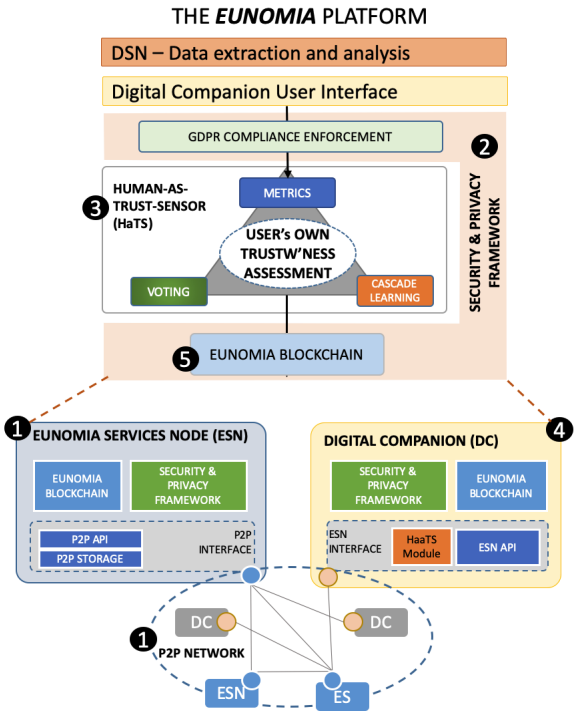

Fig. 1. The overall EUNOMIA architecture

EUNOMIA peer to peer infrastructure

The EUNOMIA infrastructure will be supported by a decentralised P2P overlay network providing access to EUNOMIA services and distributed storage and peer-to-peer communication for two types of peers: service providers (EUNOMIA services) and users (EUNOMIA digital companion). The users will contribute with a tiny part of their storage to the network (i.e. EUNOMIA storage infrastructure). The P2P network will provide:

- **Open decentralised access** – any user will be able to participate in the EUNOMIA P2P network without requiring any central controlling organization
- **Scalability** – as the number of users increases, more users are contributing with their resources, consequently shared resources will be available for the EUNOMIA community
- **Resilience** – due to its highly distributed architecture it is possible to avoid the existence of single points of failure, mitigate and quickly recover from denial of service attacks

Third party EUNOMIA services can run as parallel instances of existing open-source P2P technology (e.g. IPFS). IPv4 and IPv6 internet connections will be the low-level support for the P2P Layer, which will provide:

- **Distributed Storage**: Allows to efficiently and in a distributed way store all the data about media and users. It supports the blockchain as an off-blockchain database, and benefits from its tractability and integrity controls. It manages EUNOMIA data objects inside the P2P network ensuring their isolation.
- **Messaging**: this service provides peer-to-peer messaging, will be used by the Digital Companion peers to access the EUNOMIA services residing on the EUNOMIA Services peers.
- **Network Management**: Will support the management of peers in the EUNOMIA P2P network, providing functions for joining, routing functions, locating, accessing and publishing services.

EUNOMIA Blockchain infrastructure

The Blockchain infrastructure is able to cryptographically link the user posts to a Blockchain and create Blockchain-based signatures that can be used for verification purposes. It includes:

- **Data aggregator**: Collects and stores all posts meant to be published on the Blockchain.
- **Data formatter**: Each post is appropriately formatted. The formatting encodes two basic types of data: (i) the actual content of the post, (ii) any metadata associated with the post (e.g., author, timestamp, assessment labels, etc.).

- **Publisher**: Publishes posts in the Blockchain - this is the permanent storage solution being in accordance with one of the main features of blockchains. Considers a set of formatted posts for creating a single Blockchain transaction. Specifically, the posts are hashed and a single hash code (Merkle tree root) is created. This code can be regarded as cryptographic "summary" of all posts, which is published on the Blockchain. Also Publishes posts in Off-Blockchain storage - this can be regarded as a temporary storage solution that precedes the previous one. For this, the EUNOMIA P2P infrastructure will be utilised.
- **Logger**: Maintains records for each Blockchain transaction.
- **Transaction controller**: Controls the above modules focusing on the number of posts to be included in a single Blockchain transaction, and the frequency of transactions in terms of time.
- **Verifier**: Performs "proof-of-existence" for a given post (i.e., checks if the respective record exists in the Blockchain). In case of successful verification, it returns the associated metadata. This operation is supported for any storage solution (i.e., Blockchain or Off-Blockchain).

Security and privacy component

The starting point for the development of the security and privacy modules will be the privacy and security requirements that will result from the Privacy, social and ethical Impact Assessment (PIA+) and will ensure General Data Protection Regulation (GDPR) compliance. This will be complemented by the security and privacy requirements that should be derived from the overall operational model and exploitation operation model. These security and privacy components will provide the following high-level security properties:

- **Authentication of Users and Devices**: will provide user and device authentication on the decentralised peer-to-peer network, implementing key derivation mechanisms in order to allow a single user to hold multiple related devices and mitigate risks of the hijacking of the "user account".
- **Anonymization and minimization**: functions will be provided to support anonymised, but yet verifiable, voting, and minimization of data recorded on the long-term ledger to the strictly required to achieve the project objectives and in compliance with privacy and security requirements;
- **Confidentiality**: to avoid interception of voting and other user related information
- **Integrity checking**: recording of relevant information and ensuring the transparency of the voting process including integrity forcing mechanisms.

Human-as-Trust-Sensor (HaTS) and user reputation mechanism

EUNOMIA allows the active involvement of social media users, who can act in a Human-as-Trust-Sensor capacity: this is feasible through the use of the digital companion visualization component. A EUNOMIA user can create one or more unique IDs (and do so in an anonymous way or with a public name).

Content and context data collection and analysis

This component focuses on user-oriented and content-oriented analytics. The former relates to information on each user as a unique node of a network exhibiting particular activity, while the latter constitutes a user-agnostic computational analysis dealing with the processing of the posted content. In relation to content, EUNOMIA will focus on the text modality because of the availability of powerful analytic techniques that can be leveraged for trustworthiness evaluation. Yet, the platform will maintain a link between a user and the analytics extracted from the respective content so as to consider other modalities (such as image, video and audio) when similarly powerful techniques become available. Examples include encoding the user's activity in terms of posts-related metrics (number of posts/re-posts/comments, ratio of posts to re-posts, etc.).

The content-oriented analytics will be automatically extracted through the employment of several computational tools, focusing on the processing of textual posts. For this purpose, we aim to incorporate a series of well-established natural language processing and information extraction classifiers for shallow linguistic analysis (language detection, named entity recognition: recognition of main entities mentioned), semantic analysis (categorisation of posts as factual vs. opinionated) and detection of socially deviant language (e.g., offensive statements, hate speech).

Digital Companion

The Digital Companion is conceptualised as an application which will be able to be deployed on all types of devices, (desktop/tablets and personal) and featuring a responsive web-based and a personal (mobile/wearable) app version, allowing for the active involvement of social media users.

Data Workflows

Here we provide early examples of three core data workflows which illustrate the user-driven functionality of EUNOMIA. Specifically, we describe diagrammatically three user functions which EUNOMIA will implement: 1) Creating a new post "EUNOMIA-enriched" post, 2) Information Cascade Query and 3) Trustworthiness Voting.

Creating a new "EUNOMIA-enriched"post

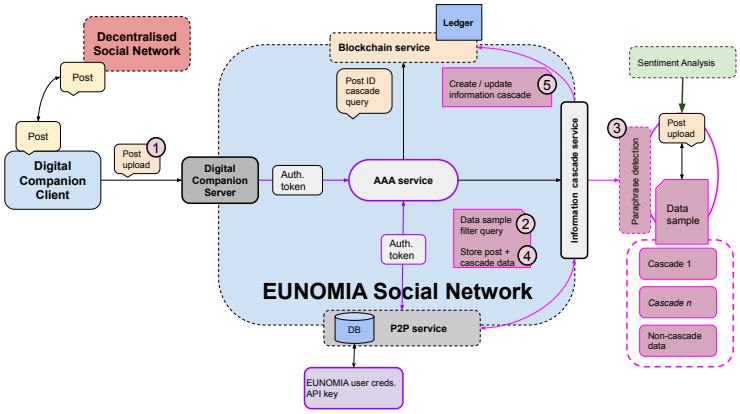

Fig. 2. Creating a new "EUNOMIA-enriched" post

Figure 2 depicts the sequence of actions that occur when a user creates a new post to one of the linked Decentralised Social Networks(DSN), using EUNOMIA Digital Companion (DC). The client, having been authenticated on the DSN, and using its' own Restful API services, forwards the new post to one of the Digital Companion servers (1). Through the Authentication Authorization and Accounting (AAA) Service, this post is added to a cluster group (based on cotent similarity) in the P2P database service (2). At the information cascade module, all the related posts with the same cluster ID are sub-sampled from the P2P service (3) and the results are forwarded, along with the original post, to the paraphrase detection service. The results may contain existing information cascade samples along with their corresponding bloom filters, or samples that do not currently belong to a cascade. In parallel with paraphrase detection, the post is also sentiment analysed, and the classification is added to the post meta-data. The overall result of paraphrase detection and sentiment analysis is stored back to the P2P database whether the post is added to a cascade or not(4). If the post is added to an information cascade, on the Blockchain ledger, a cascade ID is either updated or created if it did not already exist (5).

Information Cascade Query

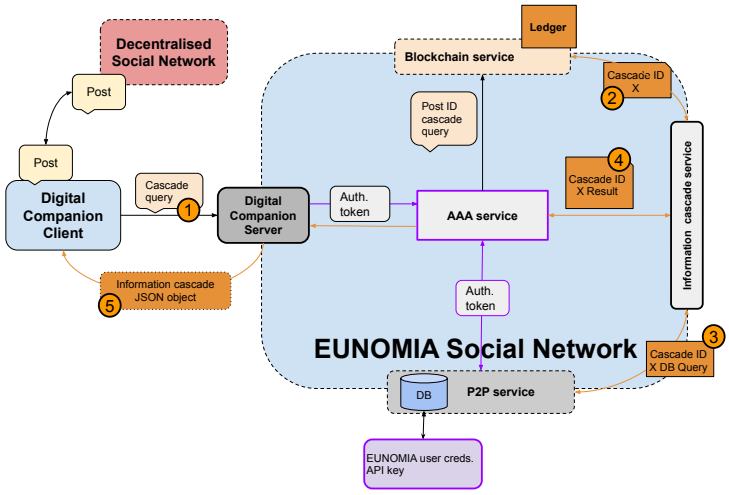

Fig. 3. Information cascade query

Following a similar logic, one can retrieve the information cascade of an existing a post (figure 3). The user, using the DC client, makes a request to the digital companion server regarding a post from the DSN (1). The server, after authentication on the AAA service, forwards the request to the information cascade service. This time, a query on the Blockchain service is made (2) first, and the cascade information results are then used to retrieve the cascade details from the P2P database(3). The JSON representation of the results is sent back to the DC server and they are visualised on the DC client(5). It is worth noting that as EUNOMIA will implement a GDPR-compliant right to be forgotten (and right to privacy) policy, both information cacasde (and non-cascade) posts may be anonymised by users and also deleted. In the case of the latter, deleted posts will be removed from associated cascades, P2P databases and the EUNOMIA Blockchain ledger.

Trustworthiness Voting

Voting will be carried out in a fully decentralised way, where the request is forwarded to several peer nodes inside the EUNOMIA network. The vote results will be saved both on the P2P database and the Blockchain ledger before returning back to the DC client and presented to the user. This same information will be retrieved within an Information Cascade Query, alongside other collected metrics (figure 4).

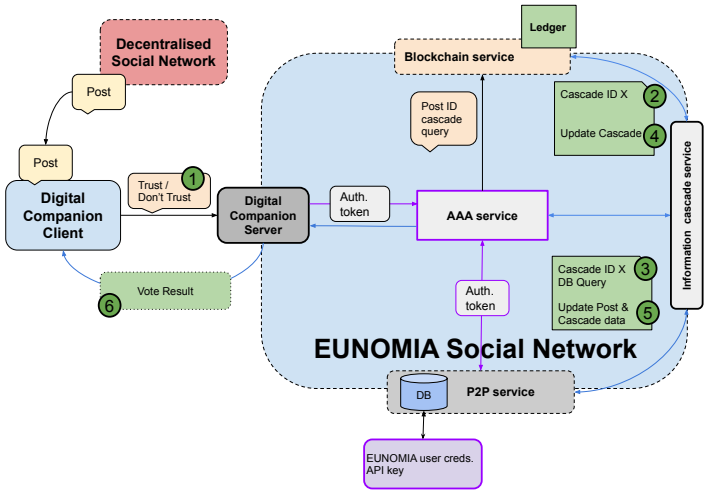

Fig. 4. EUNOMIA Voting Flow

4 Future Work

In a world where misinformation is not only causing a problem in the credible information distribution among users of social media, but is rapidly becoming a serious threat with tangible impact at political and societal level, the adoption and use of tools against misinformation is crucial. The work of EUNOMIA project presented in this paper, in the course of providing tools which can offer the arsenal for fighting misinformation, is at early research stage. What is of high importance here is the easy integration and compliance of the project's results such as the Digital Companion with tools and applications used by the social network users. The next steps in our research are in the direction of working with end users in order to identify the specifications which should define the implementation of these tools, in order to be able to extend the results of EUNOMIA to popular social networks, and provide the basis for implementation of similar tools, or porting of the EUNOMIA tools in these networks. At the same time, future proof design is of high priority for the research teams working in the project. In this course, a promising concept is the use of machine learning not for determining trustworthiness, but for assisting users in accessing information about a post that would not otherwise be available to them in a reasonable amount of time.

5 Conclusion

In this paper we have presented the concept of EUNOMIA, a Horizon 2020 Innovation Action project that aims at providing users information that may be find

useful in assessing themselves the trustworthiness of content they access on social media. It mainly focuses on how information is spread in an information cascade. It is designed to be fully decentralized, utilizing P2P and Blockchain technologies, connecting with the existing open-source distributed social networks Mastodon and Diaspora*. We have provided the structure of the framework's components and the way they connect with each other to provide a modular system of systems. We have also shown the way information flows between these components allowing the end user not only be informed about the results, but also actively participate in the process.

Acknowledgment

Work presented in this paper has received funding from the European Union's H2020 research and innovation programme under EUNOMIA project, grant agreement No. 825171.